

Escuela Superior Politécnica del Litoral

Facultad de Ciencias Naturales y Matemáticas

Métodos de *machine learning* para clasificar las necesidades de atención de apoyo
en pacientes oncológicos

Proyecto Integrador

Previo la obtención del Título de:

Ingeniero Estadístico

Presentado por:

Moisés Alexander Benítez Soriano

Josías Hazael Chiriboga Azanza

Guayaquil - Ecuador

Año: 2023

DEDICATORIA

Le dedico esta meta y todas las venideras a Dios por Su gracia y dirección en cada paso de este logro académico, a mis padres, hermanas y abuelos que siempre me han brindado consejos y apoyo incondicional en este camino.

Moisés Benítez Soriano

DEDICATORIA

Dedico este logro y todos los demás que haya alcanzado a mi querida madre, quien representa el pilar fundamental en mi vida. Su amor y apoyo constante han sido la fuente de mi fortaleza y determinación a lo largo de este camino.

Josías Chiriboga Azanza

AGRADECIMIENTOS

Quiero agradecer a Dios por haber colmado a mi familia con Su gracia y bendiciones cada día de nuestras vidas. La sabiduría, paciencia y disciplina que ha puesto en mí para poder alcanzar mis metas.

A mis padres, Moisés y Mayra, por su esfuerzo y apoyo inquebrantable, quiénes han sido mi fuente de alegría y amor incondicional. Agradezco todos sus consejos y lecciones transmitidas.

A mis hermanas, Kristhel e Isabella que son los pilares de mi corazón, les agradezco por su ánimo y cariño inmenso que han sido un faro en los momentos más difíciles. A mis abuelitos por ser mi fuente de inspiración, cuya sabiduría ha dejado una huella indeleble en mi vida. Siempre los llevaré en mi corazón.

Agradezco a la ESPOL por su integral formación académica. También agradezco a mi tutora Andrea García por su contribución y enseñanzas que han sido la base para este proyecto. A todos los profesores que han compartido sus conocimientos y experiencias.

Moisés Benítez Soriano

AGRADECIMIENTOS

Quiero expresar mi más profundo agradecimiento a mi madre, Jenny Azanza, cuyo sacrificio ha sido esencial para que yo pudiera llegar hasta aquí. Gracias de todo corazón, mamá.

Agradezco a María, quien fue la inspiración para elegir este camino y me mantuvo enfocado a lo largo de la carrera.

A Gaby por su inestimable apoyo durante todo el proceso, y a todos mis amigos, cuya compañía hizo que esta etapa fuera aún más agradable.

Agradezco a cada uno de los profesores de ESPOL que han contribuido en mi formación académica, especialmente a mi tutora Andrea García, cuya mentoría ha sido imprescindible durante el desarrollo de este proyecto.

Josías Chiriboga Azanza

Declaración Expresa

Nosotros, Moisés Alexander Benítez Soriano y Josías Hazael Chiriboga Azanza acordamos y reconocemos que la titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, información no divulgada y cualquier otro derecho o tipo de Propiedad Intelectual que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada durante el desarrollo de su trabajo de titulación, incluyendo cualquier derecho de participación de beneficios o de valor sobre titularidad de derechos, pertenecerán de forma total, perpetua, exclusiva e indivisible a LA ESPOL, sin limitación de ningún tipo. Se deja además expresa constancia de que lo aquí establecido constituye un “previo acuerdo”, así como de ser posible bajo la normativa vigente de transferencia o cesión a favor de la ESPOL de todo derecho o porcentaje de titularidad que pueda existir. Sin perjuicio de lo anterior los alumnos firmantes de la presente declaración reciben en este acto una licencia de uso gratuita e intransferible de plazo indefinido para el uso no comercial de cualquier investigación, desarrollo tecnológico o invención realizada durante el desarrollo de su trabajo de titulación, sin perjuicio de lo cual deberán contar con una autorización previa expresa de la ESPOL para difundir públicamente el contenido de la investigación, desarrollo tecnológico o invención.

Así también autorizamos expresamente a que la ESPOL realice la comunicación pública de la obra o invento, por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual. Guayaquil, 31 de agosto del 2023.



Moisés Benítez



Josías Chiriboga

EVALUADORES

Ph.D. Sandra García Bustos

Ph.D. Andrea García Angulo

RESUMEN

En el año 2020, se registraron millones de fallecimientos según la OMS. En el ámbito oncológico, los tratamientos paliativos desempeñan un papel fundamental al mejorar la calidad de vida de pacientes en fases avanzadas de la enfermedad, abordando sus aspectos físicos, psicológicos, sociales y espirituales. En este contexto, se busca establecer una conexión entre la calidad de vida y las necesidades particulares de estos pacientes, con el objetivo de brindarles una atención integral. Este estudio empleó diversos enfoques de *machine learning* mediante el lenguaje de programación R para predecir las necesidades en cuatro dominios: psicológico, físico, cuidado y atención, y salud e información. A través de un modelo de regresión logística ajustado, se identificaron las variables significativas en cada dominio, y los modelos de *machine learning* demostraron una capacidad aceptable para clasificar, logrando una exactitud del 65%-70%. La aplicación constante de encuestas especializadas para evaluar las necesidades y la calidad de vida de estos pacientes ofrece información valiosa para mejorar la atención. Además, el desarrollo de los modelos proporciona una predicción que se acerca a las necesidades diarias que enfrentan

Palabras clave: Necesidad, predicción, tratamiento paliativo, calidad de vida, encuestas.

ABSTRACT

In the year 2020, millions of deaths were recorded according to the OMS. In the field of oncology, palliative treatments play a crucial role in improving the quality of life of patients in advanced stages of the disease, addressing their physical, psychological, social, and spiritual aspects. In this context, an effort is made to establish a connection between the quality of life and the specific needs of these patients, with the aim of providing comprehensive care. This study employed various machine learning approaches using the R programming language to predict needs in four domains: psychological, physical, care and attention, and health and information. Through a refined logistic regression model, significant variables were identified in each domain, and the machine learning models demonstrated an acceptable ability to classify, achieving an accuracy of 65%-70%. The continuous application of specialized surveys to assess the needs and quality of life of these patients provides valuable information for improving care. Furthermore, the development of these models offers a prediction that closely aligns with the daily challenges they face.

Keywords: Need, prediction, palliative treatment, quality of life, surveys.

ÍNDICE DE CONTENIDO

ÍNDICE DE CONTENIDO	III
ABREVIATURAS	V
SIMBOLOGÍA.....	VI
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	IX
Capítulo 1	11
Introducción	12
1.1. Descripción del problema.....	12
1.2. Justificación del problema.....	13
1.3. Objetivos.....	14
1.3.1. Objetivo general.....	14
1.3.2. Objetivos específicos	14
1.4. Marco teórico.....	14
1.4.1. Técnicas para medir la calidad de vida.....	14
1.4.2. Técnicas para medir las necesidades atención de apoyo	15
1.4.3. Tamaño Muestral	17
1.4.4. Análisis Factorial Confirmatorio	17
1.4.5. Clasificadores Estadísticos	18
1.4.6. Evaluación de modelos estadísticos.....	21
Capítulo 2	23
Metodología	24

2.1 Recolección de datos.....	24
2.1.2 Muestreo	24
2.1.3 Método de levantamiento.....	25
2.2 Análisis Factorial Confirmatorio	25
2.3 Transformación de variables.....	27
2.4 Caracterización de la Calidad de Vida.....	29
2.5 Ajuste de modelos de clasificación.....	29
Capítulo 3	32
Resultados	33
3.1 Análisis descriptivo de la muestra	33
3.2 Análisis factorial confirmatorio.....	35
3.2.1 Validación de la encuesta QLQ-C15-PAL	35
3.2.2 Validación de la encuesta SCNS-SF34	37
3.3 Caracterización de la calidad de vida.....	40
3.4 Resultados de los modelos de clasificación.....	43
Capítulo 4.....	60
Conclusiones	61
Discusión de los resultados	62
Recomendaciones.....	62
Anexos.....	67

ABREVIATURAS

OMS	Organización Mundial de la Salud
QoL	Quality of Life (Calidad de Vida)
EORTC	European Organization for Research and Treatment of Cancer (Organización Europea para la Investigación y el Tratamiento del Cáncer)
SOLCA	Sociedad de Lucha contra el Cáncer
IRT	Item Response Theory (Teoría de la Respuesta al Ítem)
CNQ	Cancer Needs Questionnaire (Cuestionario de Necesidades de Cáncer)
SCNS	Supportive Cancer Needs Survey (Encuesta de Necesidades de Atención de Apoyo)
AFC	Análisis Factorial Confirmatorio
CFI	Comparative Fit Index (Índice de Ajuste Comparativo)
RMSEA	Root Mean Square Error of Approximation (Error Cuadrático Medio de la Aproximación)
AGFI	The Adjusted Goodness of Fit Index (Índice de Bondad de Ajuste Ajustado)
MAE	Mean Absolute Error (Media Absoluta del Error)
ROC	Receiver Operating Characteristics (Característica Operativa del Receptor)
QLQ	Quality of Life Questionnaire (Cuestionario de Calidad de Vida)
KNN	K- Nearest Neighbors (K- Vecinos más Cercanos)
SVM	Support Vector Machine (Máquinas de Vector Soporte)
TP	True Positives (Verdaderos Positivos)
FP	False Positives (Falsos Positivos)
TN	True Negatives (Verdaderos Negativos)
FN	False Negatives (Falsos Negativos)
AUC	Area Under the Curve (Área Bajo la Curva)

SIMBOLOGÍA

N	Tamaño de la población
e	Margen de error
p	Probabilidad de éxito
z	Estadístico de la distribución normal que depende del nivel de confianza.
k	Cuestionario; $k = 1,2$
η_i	Constructo i ; $i = 1, \dots, n_k$
y_t	Ítem t ; $t = 1, \dots, m_k$
λ_t	Carga del ítem y_t
θ_t	Varianza residual del ítem y_t
ψ_{ij}	Varianza y covarianza del constructo ij
$p(y x)$	Probabilidad de y dado x

ÍNDICE DE FIGURAS

Figura 1: <i>Relaciones de variables y entre factores.</i>	27
Figura 2: <i>Histograma de frecuencia de variable QoL</i>	33
Figura 3: <i>Diagrama de barras por tipo de cáncer.</i>	34
Figura 4: <i>Diagrama de barras por lugar de residencia.</i>	34
Figura 5: <i>Gráfico de densidad del número de medicinas consumidas.</i>	35
Figura 6: <i>Parámetros del modelo ACF del QLQ-C15-PAL.</i>	36
Figura 7: <i>Parámetros del modelo ACF del SCNS-SF34.</i>	39
Figura 8: <i>Boxplot de cada dominio QoL.</i>	40
Figura 9: <i>Matriz de Correlación.</i>	41
Figura 10: <i>Score QoL en función del consumo de opioide.</i>	42
Figura 11: <i>Diagrama de radar de QoL versus Necesidades.</i>	43
Figura 12: <i>Estimación de coeficientes del dominio psicológico del modelo de regresión logística.</i>	44
Figura 13: <i>Accuracy del hiperparámetro mtry del dominio psicológico.</i>	45
Figura 14: <i>Accuracy del hiperparámetro k del dominio psicológico.</i>	45
Figura 15: <i>Accuracy de los hiperparámetros C y sigma del dominio psicológico.</i>	46
Figura 16: <i>Comparación de modelos del dominio psicológico.</i>	47
Figura 17: <i>Estimación de coeficientes del dominio físico del modelo de regresión logística.</i>	48
Figura 18: <i>Accuracy del hiperparámetro mtry del dominio físico.</i>	49
Figura 19: <i>Accuracy del hiperparámetro k del dominio físico.</i>	49
Figura 20: <i>Accuracy de los hiperparámetros C y sigma del dominio físico.</i>	50
Figura 21: <i>Comparación de modelos del dominio físico.</i>	51

Figura 22: <i>Estimación de coeficientes del dominio de cuidado y atención del modelo de regresión logística.</i>	52
Figura 23: <i>Accuracy del hiperparámetro mtry del dominio de cuidado y atención.</i>	53
Figura 24: <i>Accuracy del hiperparámetro k del dominio de cuidado y atención.</i>	53
Figura 25: <i>Accuracy del hiperparámetro C y σ del dominio de cuidado y atención.</i> 54	
Figura 26: <i>Comparación de modelos del dominio de cuidado y atención.</i>	55
Figura 27: <i>Estimación de coeficientes del dominio de salud e información del modelo de regresión logística.</i>	56
Figura 28: <i>Accuracy del hiperparámetro mtry del dominio de salud e información.</i>	57
Figura 29: <i>Accuracy del hiperparámetro k del dominio salud e información.</i>	57
Figura 30: <i>Accuracy de los hiperparámetros C y σ del dominio de salud e información.</i>	58
Figura 31: <i>Comparación de modelos del dominio de salud e información.</i>	59

ÍNDICE DE TABLAS

Tabla 1: <i>Disposición de los ítems con su respectiva escala y constructo del QLQ-C15-PAL</i>	28
Tabla 2: <i>Disposición de los ítems con su respectivo constructo del SCNS-SF34</i>	28
Tabla 3: <i>Tabla de frecuencia para la variable edad</i>	33
Tabla 4: <i>Indicadores de ajuste QLQ-C15-PAL</i>	35
Tabla 5: <i>Indicadores de ajuste SCNS-SF34</i>	37
Tabla 6: <i>Matriz de confusión del dominio psicológico del modelo de regresión logística</i>	44
Tabla 7: <i>Matriz de confusión del dominio psicológico del modelo random forests</i>	45
Tabla 8: <i>Matriz de confusión del dominio psicológico del modelo KNN</i>	46
Tabla 9: <i>Matriz de confusión del dominio psicológico del SVM</i>	46
Tabla 10: <i>Comparación de métricas del dominio psicológico</i>	47
Tabla 11: <i>Matriz de confusión del dominio físico del modelo de regresión logística</i>	48
Tabla 12: <i>Matriz de confusión del dominio físico del modelo random forests</i>	49
Tabla 13: <i>Matriz de confusión del dominio físico del modelo KNN</i>	50
Tabla 14: <i>Comparación de métricas del dominio físico</i>	51
Tabla 15: <i>Matriz de confusión del dominio de cuidado y atención del modelo de regresión logística</i>	52
Tabla 16: <i>Matriz de confusión del dominio de cuidado y atención del modelo random forests</i>	53
Tabla 17: <i>Matriz de confusión del dominio de cuidado y atención del modelo KNN</i>	54
Tabla 18: <i>Matriz de confusión del dominio de cuidado y atención del modelo SVM</i>	54
Tabla 19: <i>Comparación de métricas de cuidado y atención</i>	55

Tabla 20: <i>Matriz de confusión del dominio de salud e información del modelo de regresión logística.</i>	56
Tabla 21: <i>Matriz de confusión del dominio de salud e información del modelo random forests.</i>	57
Tabla 22: <i>Matriz de confusión del dominio de salud e información del modelo KNN.</i>	58
Tabla 23: <i>Matriz de confusión del dominio de salud e información del modelo SVM.</i>	58
Tabla 25: <i>Comparación de métricas del dominio de salud e información.</i>	59

Capítulo 1

Introducción

Alrededor del mundo, el cáncer es una de las principales causas de muerte, provocando 10 millones de fallecidos en el año 2020, según datos de la Organización Mundial de la Salud (OMS, 2021). En el área de la oncología, los tratamientos paliativos juegan un papel fundamental en el cuidado de los pacientes que se encuentran en etapas avanzadas de la enfermedad. Estos tratamientos están diseñados para mejorar la calidad de vida y aliviar los síntomas asociados al cáncer, brindando un enfoque integral que abarca aspectos físicos, psicológicos, sociales y espirituales (NIH, 2021).

Los objetivos de los cuidados paliativos se alinean con las necesidades de atención de apoyo, ya que, permiten abordar los desafíos de mejorar la vida de los pacientes oncológicos. Lidington et al. (2022) utilizaron un cuestionario enfocado a encontrar aquellas necesidades de apoyo, SCNS-SF34. En este estudio, se determinó indicadores que relacionan las necesidades de los pacientes de acuerdo con su calidad de vida.

Para la medición de la calidad de vida, llamada desde ahora QoL (*quality of life*), en pacientes que padecen de cáncer y se encuentran recibiendo tratamientos paliativos, se utiliza el Cuestionario Central de Calidad de Vida de la Organización Europea para la Investigación y el Tratamiento del Cáncer, llamado EORTC QLQ-C15-PAL. Este estudio se centra en la aplicación de técnicas estadísticas para la clasificación de las necesidades de atención de apoyo de estos pacientes, con base en el grado de calidad de vida.

1.1. DESCRIPCIÓN DEL PROBLEMA

Los pacientes oncológicos que reciben tratamientos paliativos enfrentan una serie de desafíos físicos, emocionales y sociales (Garcia et al., 2023). Por esto, los cuidados paliativos son definidos como un enfoque dirigido a mejorar la QoL del paciente (Faguet, 2016). Por otro

lado, es necesario mencionar que, a menudo los pacientes encuentran dificultades para acceder a servicios de atención de apoyo que se ajusten a sus necesidades específicas. Esto puede deberse a una falta de comprensión o identificación clara de dichas necesidades, que resulta en una atención genérica que no aborda sus preocupaciones individuales (Tieman, 2016). Teniendo en cuenta la importancia y la dificultad que trae consigo la identificación de las necesidades de un paciente, se vuelve vital la aplicación de métodos automáticos basados en datos proporcionados por los mismos pacientes acerca de su estado actual.

La Sociedad de Lucha contra el Cáncer (SOLCA) es una institución sin fines de lucro que se dedica a la prevención, diagnóstico, tratamiento y rehabilitación de pacientes con cáncer en Ecuador. En busca de una mejora en la atención de sus pacientes ofrece los datos necesarios para el desarrollo de los modelos estadísticos.

1.2. JUSTIFICACIÓN DEL PROBLEMA

El propósito de aplicar técnicas innovadoras en el campo de la oncología tiene como metas principales:

1. Permitir el desarrollo de estrategias enfocadas a una buena toma de decisiones clínicas, conociendo las necesidades específicas del paciente en función del grado de calidad de vida;
2. Asegurar una optimización de recursos, priorizando y brindando apoyo adecuado a aquellos que tengan necesidades específicas;
3. Brindar atención individualizada, personalizada y adaptada a la atención de los pacientes.

Por medio de estos procesos se vuelve posible asignar recursos adecuados a cada paciente. En resumen, se busca resolver esta problemática desde la perspectiva del paciente, enfocado a sus necesidades para asegurar un cuidado integral y de mayor calidad.

1.3. OBJETIVOS

1.3.1. OBJETIVO GENERAL.

Desarrollar modelos que permitan relacionar las necesidades de atención de apoyo y calidad de vida de los pacientes oncológicos, en tratamientos paliativos, mediante técnicas estadísticas.

1.3.2. OBJETIVOS ESPECÍFICOS

1. Implementar los cuestionarios EORTC QLQ-C15-PAL y SCNS-SF34 para la medición de la calidad de vida y necesidades de atención de apoyo en pacientes con cáncer en cuidados paliativos.
2. Validar los constructos de las variables de calidad de vida y necesidad de apoyo que permitan la implementación del modelo estadístico en la población objetivo.
3. Caracterizar la calidad de vida en pacientes oncológicos para una mejora en la toma de decisiones clínicas.
4. Implementar modelos de clasificación para relacionar la calidad de vida y necesidad de apoyo.

1.4. MARCO TEÓRICO

1.4.1. TÉCNICAS PARA MEDIR LA CALIDAD DE VIDA

La importancia de medir la condición de un objeto, individuo o evento nace a partir del control de calidad. La necesidad de llevar a cabo este procedimiento a la medicina ha crecido en los últimos años. De este modo, se ha originado la necesidad de evaluar el estado de “vida” de los pacientes (Barofsky, 2011).

La Organización Europea para la Investigación y Tratamiento del Cáncer, EORTC, ha desarrollado el cuestionario QLQ-C30 que es, por mucho, el cuestionario más usado para medir la calidad de vida de pacientes con cáncer. Se emplea para todos sus tipos, como lo mencionan Higgins y Green (2008). La problemática de este cuestionario radica en que no todos los ítems son apropiados para aquellos pacientes que se encuentran en tratamientos paliativos, lo que afecta su efectividad. Por esto, Bjorner et al. (2004) proponen una versión más corta del QLQ-C30, disminuyendo los ítems innecesarios y produciendo el QLQ-C15-PAL. Este cuestionario contiene 15 ítems orientados a evaluar la calidad de vida de pacientes oncológicos en tratamientos paliativos.

Tal como se describe en el manual desarrollado por Groenvold et al. (2006), la versión acortada de las escalas multi-ítem para el QLQ-C15-PAL se construyeron por medio de la Teoría de la Respuesta al Ítem (IRT). Este modelo a su vez emplea algoritmos que evalúan estas escalas en la métrica original. Durante su consolidación, se realizó múltiples entrevistas con expertos de la salud y pacientes para identificar los ítems que no eran apropiados y que se podían eliminar. Los ítems se distribuyen en tres subgrupos: escalas funcionales, escalas sintomáticas y la escala global QoL del ítem 15. En las escalas funcionales, se incluyen los dominios de la parte física y emocional del paciente; mientras que, la escala sintomática se divide en dos dominios de multi-ítem y cinco dominios de ítem único, descrito en el instructivo por McElduff et al. (2004).

1.4.2. TÉCNICAS PARA MEDIR LAS NECESIDADES ATENCIÓN DE APOYO

La evaluación de necesidades es una medida directa de la diferencia entre lo que los pacientes experimentan y esperan. Permite identificar sus problemas más urgentes y priorizar la atención en esas áreas específicas.

A principios de 1990, se desarrolló un cuestionario de necesidades de cáncer (CNQ) que evaluaba la necesidad de los pacientes en los siguientes factores: psicológico, información de salud, vida física y diaria, atención y apoyo al paciente; y comunicación interpersonal (Lattimore-Foot, 1996; Foot y Sanson-Fisher, 1995). Tiempo después, al final de este mismo año, se crea la encuesta de necesidades de atención de apoyo SCNS, que posee tres formatos diferentes, uno de 59 elementos SCNS-LF59, otro de 31 elementos SCNS-SF31, y, por último, la encuesta de 34 ítems, que se creó en el año 2002 SCNS-SF34(Boyes et al, en prensa), siendo ésta la más recomendada para actividades de investigación y control de calidad. Actualmente se cuenta con módulos dirigidos al tratamiento del cáncer de mama, melanoma, cáncer de próstata, colostomía y acceso a atención médica y servicios de apoyo. Sin embargo, se encuentran en proceso de desarrollo módulos destinados a abordar la linfedema posterior al cáncer de mama, brindar apoyo a sobrevivientes de cáncer a largo plazo y atender las necesidades de pacientes con cáncer incurable avanzado.

El SCNS-SF34 solo toma en cuenta las correlaciones con coeficientes mayores a 0.57 y considera la distribución de frecuencia y la importancia clínica de cada ítem usando datos del estudio de validación conforme a la evaluación psicométrica; a su vez realiza un análisis factorial y los 34 elementos son asignados a 5 dominios: psicológico, sistema de salud, vida física y diaria, atención y apoyo, y sexualidad, los mismos que explican el 72.1% de la varianza total.

Asimismo, es necesario mencionar que para la validación del constructo del SCNS-SF34 se realiza un análisis factorial de componentes principales con rotación ortogonal y que la encuesta evalúa la confiabilidad utilizando un alfa de Cronbach fijo de 0.7, aunque la SCNS-LF59 y la SCNS-SF34 superan el 0.8 en todos los dominios.

A diferencia del cuestionario para medir la calidad de vida, que tiene una versión para pacientes en cuidados paliativos QLQ-C15-PAL, el SCNS no tiene una versión específica para dicha población.

1.4.3. TAMAÑO MUESTRAL

Una muestra tiene como objetivo representar las características de una población tomando solo una porción de esta. La población es la totalidad de individuos que se quiere investigar en un estudio. Para el cálculo de la muestra se utiliza la siguiente fórmula:

$$\text{Tamaño de muestra} = \frac{\frac{z^2 p(1-p)}{e^2}}{1 + \frac{z^2 p(1-p)}{e^2 N}}$$

N: Tamaño de la población

e: Margen de error

p: Probabilidad de éxito

z: Estadístico de la distribución normal que depende del nivel de confianza.

1.4.4. ANÁLISIS FACTORIAL CONFIRMATORIO

El análisis factorial confirmatorio, también llamado AFC, es una técnica estadística utilizada para evaluar la validez de un modelo teórico sobre cierta estructura factorial en un conjunto de datos. Se basa en la suposición de que los datos observados o medidos pueden ser explicados por variables latentes o factores intrínsecos (Hair et al., 2010). El AFC es una herramienta valiosa para evaluar la validez de escalas y cuestionarios descrito por Brown (2007).

Las estimaciones de los parámetros se realizan mediante algoritmos de optimización y técnicas de máxima verosimilitud, buscando valores que incrementen la probabilidad de las observaciones mencionado por Muthén (2002).

Por otro lado, para evaluar la concordancia del modelo propuesto con los datos, se usan índices de adecuación como el índice de ajuste comparativo (CFI) (valores cercanos a 1 son un buen ajuste), el índice robusto de Tucker-Lewis (TLI) (valores cercanos a 1 son un buen ajuste), y el error cuadrático medio de la aproximación (RMSEA) (valores cercanos a 0 son un buen ajuste). Idealmente debería ser $RMSEA < 0.08$ o < 0.05 . Se realiza una prueba de hipótesis a un nivel del 5% de significancia cuyo contraste es:

$$H_0: RMSEA \leq 0.05$$

$$H_1: RMSEA > 0.05$$

Kline (2016) indica que los índices de adecuación proporcionan sobre cómo el modelo teorizado se ajusta a las observaciones permitiendo así la validez del modelo.

1.4.5. CLASIFICADORES ESTADÍSTICOS

La clasificación estadística, también llamada aprendizaje supervisado, es una rama importante de los métodos de *machine learning* que se caracteriza por usar etiquetas para ajustar el modelo. Para ello, se asocia los datos de entrada x con los datos de salida y , que son las etiquetas en la variable de respuesta. Muchos modelos de clasificación se basan en estimar la distribución de probabilidad $p(y|x)$ usando técnicas paramétricas, como regresión logística, regresión multinomial, clasificador bayesiano ingenuo, máquinas de vector soporte y árboles de decisión. También se emplea muchos modelos de clasificación no probabilísticos, llamados también no paramétricos, tales como *random forests*, *k*-vecinos más cercanos y redes neuronales profundas. Gran parte de estos modelos son mencionados por Hastie et al. (2013), Goodfellow et al. (2016) y Beysolow (2017).

La regresión logística para clasificación es un método estadístico que se emplea para predecir y asignar observaciones a una de dos categorías posibles, en función de un

conjunto de características. Utiliza una función logística para modelar la probabilidad de pertenencia a una categoría, transformando la suma ponderada de las características en un valor entre 0 y 1. Si la probabilidad calculada es superior a un umbral determinado (por lo general 0.5), la observación se clasifica en una categoría; de lo contrario, se asigna a la otra. Además, permite identificar las variables predictoras que son significativas para la predicción de la variable binaria de respuesta.

Regresión logística

La regresión logística para clasificación es un método estadístico que se emplea para predecir y asignar observaciones a una de dos categorías posibles, en función de un conjunto de características. Utiliza una función logística para modelar la probabilidad de pertenencia a una categoría, transformando la suma ponderada de las características en un valor entre 0 y 1. Si la probabilidad calculada es superior a un umbral determinado (por lo general 0.5), la observación se clasifica en una categoría; de lo contrario, se asigna a la otra. Además, permite identificar las variables predictoras que son significativas para la predicción de la variable binaria de respuesta.

Random forest

Random forest es un algoritmo de aprendizaje automático que se utiliza principalmente para tareas de clasificación y regresión. Se basa en la construcción de múltiples árboles de decisión y combina sus resultados para obtener predicciones más precisas y robustas. Cada árbol en el bosque se construye utilizando una muestra aleatoria del conjunto de datos y un subconjunto aleatorio de características. Las predicciones finales se obtienen promediando las decisiones de todos los árboles (en caso de regresión) o mediante votación (en el caso de clasificación).

El enfoque de construir varios árboles y combinar sus resultados reduce el riesgo de sobreajuste (sobre entrenamiento) y aumenta la generalización del modelo. *Random*

forest es capaz de manejar datos con múltiples características y manejar valores faltantes y ruido en los datos. Además, puede proporcionar información sobre la importancia de las características en el proceso de toma de decisiones.

K-Nearest Neighbors

k-Nearest Neighbors (KNN) es un algoritmo de aprendizaje automático supervisado utilizado tanto para clasificación como para regresión. Se basa en la premisa de que las observaciones similares tienden a compartir características similares. Para hacer una predicción o clasificación, KNN encuentra los "k" ejemplos de entrenamiento más cercanos en función de la distancia (por lo general, la distancia euclidiana) en un espacio de características. En el caso de clasificación, la etiqueta más común entre los vecinos más cercanos se asigna a la observación de prueba. Para regresión, se promedian los valores de los vecinos para predecir un valor numérico.

El valor de "k" influye en la precisión del algoritmo: un valor bajo podría dar lugar a una sensibilidad excesiva al ruido, mientras que un valor alto podría disminuir la capacidad del modelo para capturar patrones locales. KNN es simple de implementar y comprender, y puede manejar datos con varias características.

Support Vector Machines

Support Vector Machines (SVM) son un conjunto de algoritmos de aprendizaje automático utilizados tanto para tareas de clasificación como de regresión. El objetivo principal de SVM es encontrar un hiperplano en un espacio de características que maximice el margen entre diferentes clases de datos. En el caso de clasificación, este hiperplano separa las clases de manera óptima, maximizando la distancia entre los puntos más cercanos de cada clase, llamados vectores de soporte.

SVM es eficaz incluso en espacios de alta dimensionalidad y puede manejar datos no linealmente separables a través de transformaciones no lineales o mediante el uso de

funciones *kernel*. Las funciones *kernel* permiten mapear los datos a un espacio de mayor dimensionalidad donde pueden ser linealmente separables, lo que brinda flexibilidad al algoritmo.

1.4.6. EVALUACIÓN DE MODELOS ESTADÍSTICOS

Según Hastie et al. (2013), el rendimiento de un determinado método de aprendizaje viene dado por la capacidad que tenga para predecir correctamente los datos de evaluación de manera independiente. Dicho esto, la evaluación del modelo varía del tipo de modelo estadístico y de su objetivo al momento de aplicar. Pero en general, el diagnóstico de un modelo tiene que ver con los siguientes tres aspectos:

- **Ajuste del modelo:** Se relaciona con medir qué tan bien el modelo se acopla a la distribución de los datos. De acuerdo con Montgomery et al. (2012), el ajuste del modelo es esencial para evaluar la calidad y la capacidad de descripción del modelo estadístico.
- **Supuestos del modelo:** Gran parte de los modelos estadísticos tienen sus fundamentos en supuestos sobre la distribución de los datos y su relación con las variables. Este aspecto implica evaluar los supuestos y si es apropiado el modelo en cuestión para los datos. En resumen, la evaluación de los supuestos es crucial para garantizar la validez y la confiabilidad de los resultados del modelo (Gelman & Hill, 2007).
- **Capacidad predictiva:** Estimar el desempeño del modelo prediciendo correctamente a partir de nuevos datos, no usados en el ajuste del modelo. Se utilizan medidas como la tasa de error de predicción como el MAE. Según Hastie et al. (2009), la capacidad predictiva es un aspecto fundamental en la evaluación de los modelos estadísticos.

El resultado principal al evaluar un modelo se basa en cometer errores de predicción. Algunas métricas de evaluación para medir esto:

- Tasa de aciertos: Las predicciones correctas entre el total de predicciones.
- Tasa de errores: Las predicciones incorrectas entre el total de predicciones.
- Especificidad: Proporción de la frecuencia de valores correctos negativos y el total de valores negativos observados.
- Sensibilidad: Proporción de la frecuencia de valores correctos positivos y el total de valores positivos observados.
- Curva ROC (*Receiver Operating Characteristic*): La curva representa la relación de la proporción de verdaderos positivos y la proporción de falsos positivos al variar el umbral de clasificación del modelo.
- Validación cruzada: Técnica que se basa en segmentar los datos en prueba y entrenamiento, para así evaluar la tasa de error de predicción independientemente de la partición.

Capítulo 2

Metodología

2.1 RECOLECCIÓN DE DATOS

2.1.1 TAMAÑO DE MUESTRA

La población que se consideró para el estudio tuvo que cumplir 5 criterios principales, pacientes de SOLCA derivados a cuidados paliativos, edad superior o igual a 18 años, pacientes que se encuentren la mayor parte del día encamados, pacientes que acudan a la consulta externa del hospital de SOLCA y pacientes orientados en tiempo y espacio.

Dado que el tamaño de población que cumple con estos 5 criterios no está oficialmente determinado, para el cálculo del tamaño muestral, se tomó como tamaño de la población al número promedio de pacientes que tuvieron consultas en el hospital en los tres meses anteriores al inicio del estudio (434 pacientes). Si bien es cierto que las metodologías de Machine Learning tienden a trabajar mejor con la mayor cantidad de datos posibles, existieron limitaciones en el acceso a los pacientes, las cuales son discutidas en el Capítulo 5. Se logró obtener un tamaño muestral de 92 pacientes que corresponde a un margen de error de 9.08% y nivel de confianza al 95%.

2.1.2 MUESTREO

Dada las distintas características de la población y el lugar en que se emplaza, se optó por un muestreo no probabilístico por conveniencia. Los encuestadores se dirigieron por 5 semanas a la consulta externa del hospital durante los horarios asignados por los médicos de cuidados paliativos. Se realizó la encuesta a cada paciente que accedió a participar del estudio durante el tiempo de espera para su consulta.

2.1.3 MÉTODO DE LEVANTAMIENTO

Se usó como instrumento de medición de la calidad de vida de pacientes oncológicos en tratamientos paliativos el QLQ-C15-PAL, mientras que, el cuestionario SCNS-SF34 se lo usó para identificar las necesidades de atención de apoyo. Adicional a esto se añadió a la encuesta la edad y variables de historial médico como tipo de cáncer primario, consumo de opioides, número de medicinas, fecha de derivación al departamento de cuidado paliativos, fecha en que fue diagnosticado el paciente de cáncer y lugar de residencia. Dado que el cuestionario SCNS-SF34 no tiene una versión específica para los pacientes en cuidados paliativos, se obvió aquellos ítems que se consideraron irrelevantes para el estudio, como por ejemplo el quinto dominio de sexualidad. La encuesta utilizada está disponible en el Anexo 1.

Considerando las limitantes en cuanto a calidad de señal de servicio de telefonía que existe en el hospital se optó por usar herramientas digitales como *Google Forms* y también el uso de impresiones para un llenado manual. Por cada paciente se realizó ambas encuestas, una detrás de la otra, durante el tiempo de espera hasta que empiece la consulta del paciente. La estimación del tiempo efectivo para completar ambas encuestas fue de 30 minutos por paciente, realizando una prueba piloto de 5 pacientes.

2.2 ANÁLISIS FACTORIAL CONFIRMATORIO

Tal como es descrito por Groenvold et al. (2006) y McElduff et al. (2004) en los instructivos sobre los cuestionarios QLQ-C15-PAL y SCNS-SF34, las características psicométricas de cada cuestionario permiten identificar los constructos generados y los principales ítems que aportan a su generación, permitiendo generalizar estos factores o

constructos con sus respectivos ítems. Se usó un análisis factorial confirmatorio para validar tal información por cada cuestionario en nuestra población objetivo y además encontrar las interacciones entre cada constructo con su respectivo grupo de ítems, en su escala original. Se realizó un modelo en que se toma en cuenta las relaciones entre todos los constructos que tienen más de un ítem presente. Los datos son ordinales, por lo tanto, fueron analizados a través de matriz de correlación policórica y la estimación del modelo haciendo uso del método Diagonally Weighted Least Squares (DWLS) implementado el modelo en el lenguaje R (versión 4.3.1) a través de la librería “lavaan”.

$k =$ Cuestionario; $k = 1,2$

$\eta_i =$ Constructo i ; $i = 1, \dots, n_k$

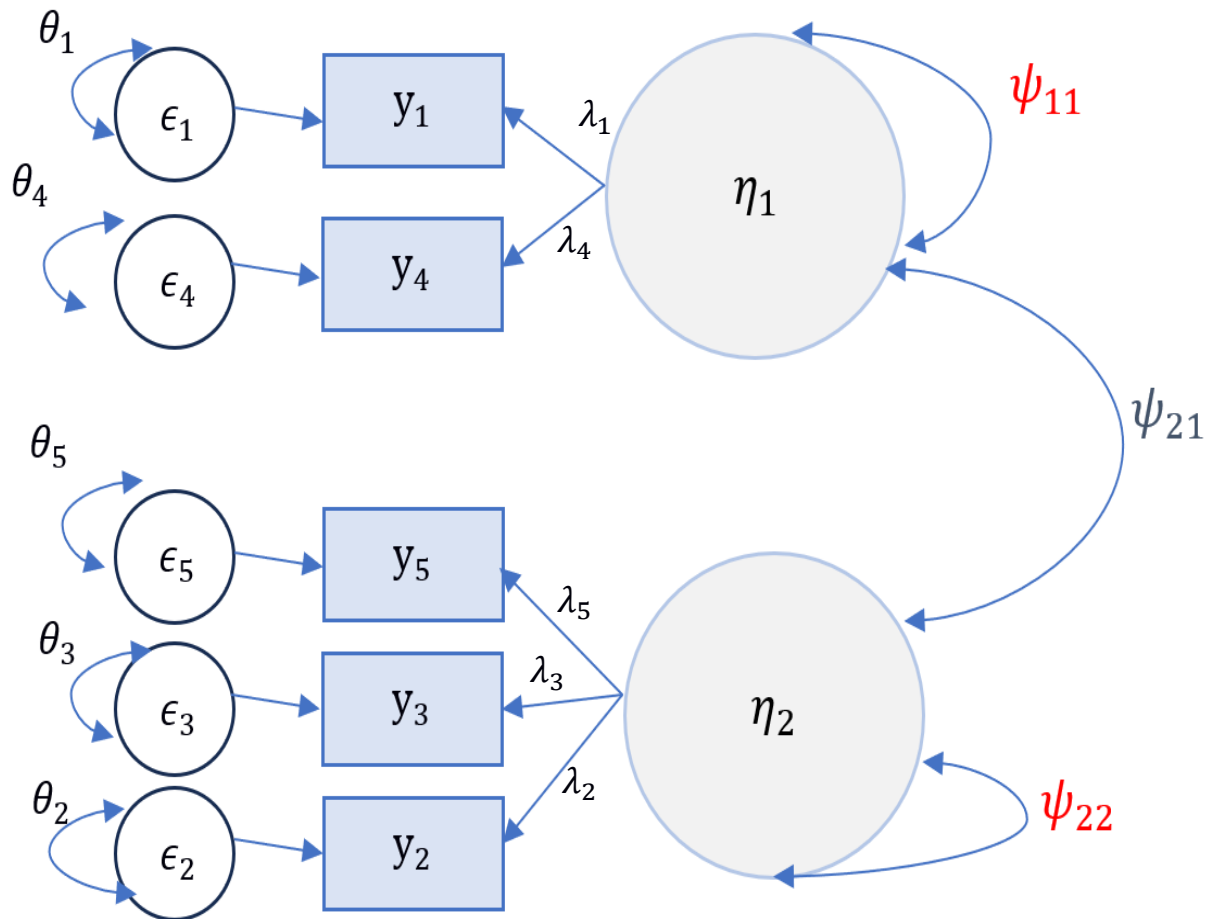
$y_t =$ Ítem t ; $t = 1, \dots, m_k$

$\lambda_t =$ Carga del ítem y_t

$\theta_t =$ Varianza residual del ítem y_t

$\psi_{ij} =$ Varianza y covarianza del constructo ij

Figura 1: Relaciones de variables y entre factores.



2.3 TRANSFORMACIÓN DE VARIABLES

Conforme al instructivo de los cuestionarios QLQ-C15-PAL y SCNS-SF34 realizado por Groenvold et al. (2006) y McElduff et al. (2004) respectivamente, se procedió a realizar transformaciones a cada ítem. En el caso del cuestionario QLQ-C15-PAL se usó una tabla de transformación de respuestas a la escala del QLQ original, para poder obtener scores por cada constructo.

Tabla 1: Disposición de los ítems con su respectiva escala y constructo del QLQ-C15-PAL.

Constructo	Escala	Ítem
Calidad de vida	“QL ítem”	15
Físico	PF2	1-3
Emocional	EF	13, 14
Fatiga	FA	7, 11
Náusea y vómito	NV	9
Dolor	PA	5, 12
Disnea	DY	4
Insomnio	SL	6
Pérdida de apetito	AP	8
Constipación	CO	10

Nota. Cada escala-constructo tiene su propio algoritmo para obtener un *score* entre 0 y 100, como lo especifica Groenvold et al. (2006).

De la misma manera, se usó los dominios de necesidad con los ítems específicos para cada uno, indicado por McElduff et al. (2004) dados en la siguiente tabla:

Tabla 2: Disposición de los ítems con su respectivo constructo del SCNS-SF34.

Constructo	Ítem
Necesidades físicas y de la vida diaria	1-5
Necesidades psicológicas	1-9, 12
Necesidades de atención y apoyo al paciente	13-17
Necesidades del sistema de salud y de información	18-25, 27-28

Fue de interés obtener como resultado de cada dominio una variable dicotómica que indique si el paciente tiene o no necesidad en este dominio o constructo. Para ello, se utilizó la mediana de los ítems, dentro de cada constructo, si esta era mayor o igual a 4 se indicó que el

paciente tiene necesidad en ese constructo y en caso contrario se indicó que el paciente no tiene necesidad (McElduff et al. ,2004).

2.4 CARACTERIZACIÓN DE LA CALIDAD DE VIDA

El procedimiento de describir, analizar descriptivamente y resumir las variables entra en la caracterización, para comprender los datos y realizar un análisis posterior.

Se tomó a consideración para el análisis 14 variables en total y para un mejor apoyo se usó la variable binaria de necesidad de atención de apoyo del paciente. Luego de la transformación realizada, las variables relacionadas a la calidad de vida tuvieron una tipología numérica al igual que la variable de edad y el número de medicinas consumidas por el paciente.

Se usó gráficos de dispersión para comprender el comportamiento de los datos en cada dominio relacionado a la calidad de vida. Para conocer la frecuencia por cada tipo de cáncer se creó un gráfico de columnas. Por último, se realizó un gráfico de radar que nos permita visualmente interpretar cierta relación de los *scores* de calidad de vida con la necesidad del paciente, por cada dominio de necesidad.

2.5 AJUSTE DE MODELOS DE CLASIFICACIÓN

Se emplearon 13 variables predictoras para cada paciente. De estas variables, 10 están relacionadas con la calidad de vida del paciente, las cuales corresponden a las escalas formadas a partir del cuestionario QLQ-C15-PAL (Tabla 1), mientras que las 3 variables restantes son la edad del paciente, el número de medicamentos consumidos por el paciente y una variable binaria que indica si el paciente consume o no opioides. Como variable de respuesta se utilizó por separado cada uno de los dominios formados del cuestionario SCNS-SF34, en donde los valores posibles eran “Si” y “No”, los cuales indican si el paciente tiene o no necesidad dentro de tal dominio.

Se exploraron varios modelos, incluyendo regresión logística, *random forest*, k-vecinos más cercanos (KNN) y *support vector machines* (SVM), con el objetivo de identificar el modelo

que mejor se adapte a cada dominio y permitiera realizar predicciones precisas sobre las necesidades de atención de apoyo de los pacientes. A continuación, se detallan los pasos que se siguieron en el ajuste de los modelos, así como los procedimientos de entrenamiento y evaluación.

2.5.1 ENTRENAMIENTO DE LOS MODELOS

De la muestra total de 92 pacientes, se utilizó el 75% de las observaciones para entrenar los modelos y el 25% restante se utilizó para evaluar el rendimiento de los modelos ajustados. La selección de los subconjuntos de datos fue aleatoria.

Se implementó un enfoque de validación cruzada con 10 particiones y 5 repeticiones para evaluar diferentes combinaciones de hiperparámetros. En específico, se ajustaron los siguientes hiperparámetros para cada modelo:

- **Random forest:** *mtry* (cantidad de variables que se consideran en cada árbol de decisión).
- **KNN:** *k* (número de observaciones más cercanas que se consideran para la clasificación).
- **SVM:** *C* (penalización del error), *sigma* (coeficiente utilizado en el *kernel* radial).

2.5.2 EVALUACIÓN DE LOS MODELOS

Para todos los modelos considerados, se utilizó el subconjunto de prueba para formar una matriz de confusión con el objetivo de evaluar su capacidad de clasificación en términos de verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN).

Además de la matriz de confusión, se calcularon las siguientes métricas de evaluación para cada modelo:

- **Exactitud:** Indica la proporción de predicciones correctas realizadas por el modelo en relación con el total de predicciones.
- **Sensibilidad:** También conocida como tasa de verdaderos positivos, mide la proporción de casos positivos que fueron correctamente identificados por el modelo.
- **Especificidad:** Representa la tasa de verdaderos negativos, es decir, la proporción de casos negativos que fueron correctamente identificados por el modelo.
- **Área bajo la curva (AUC):** Esta métrica proporciona una medida de la capacidad global de discriminación del modelo, considerando varios umbrales de clasificación.
- **Coeficiente Kappa:** Evalúa la concordancia entre las predicciones del modelo y las clases reales, teniendo en cuenta el acuerdo esperado al azar.

El entrenamiento y evaluación de modelos se realizó en el lenguaje R (versión 4.3.1) y las librerías “caret” y “tidyverse”.

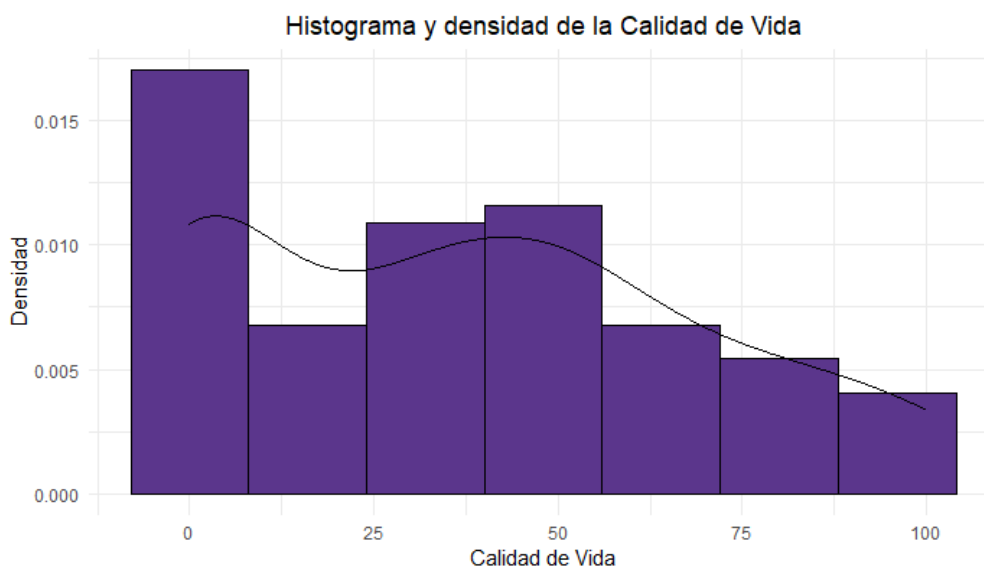
Capítulo 3

Resultados

3.1 ANÁLISIS DESCRIPTIVO DE LA MUESTRA

El total de pacientes oncológicos en tratamientos paliativos encuestados fue de 92, estos se tiene datos demográficos, datos respecto a su calidad de vida y sus necesidades de atención de apoyo. A continuación, se muestra la distribución de frecuencias de la calidad de vida percibida por los pacientes pertenecientes a la muestra.

Figura 2: *Histograma de frecuencia de variable QoL*



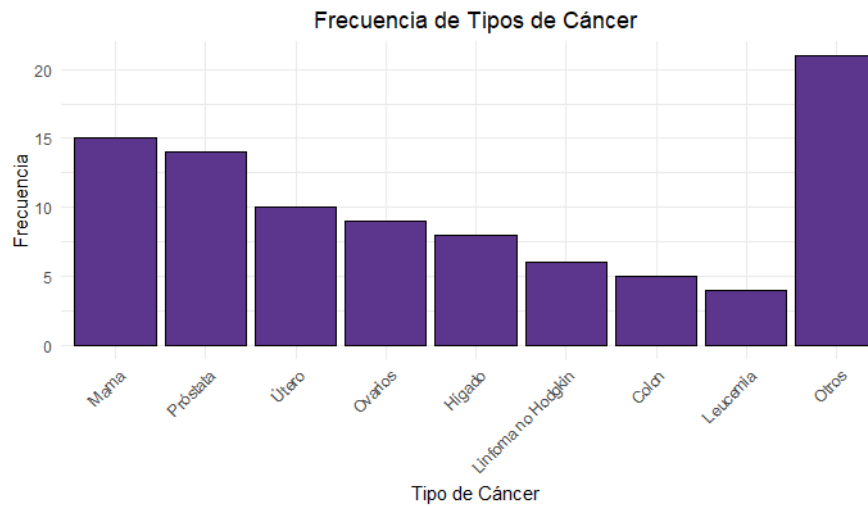
En la siguiente tabla tenemos la distribución de la edad de los pacientes encuestados:

Tabla 3: *Tabla de frecuencia para la variable edad.*

Edad	Marca de clase	Frecuencia absoluta	Frecuencia acumulada	Frecuencia relativa	Frec. relativa acumulada
[20,30)	25	18	18	20%	20%
[30,40)	35	21	39	23%	42%
[40,50)	45	16	55	17%	60%
[50,60)	55	14	69	15%	75%
[60,70)	65	11	80	12%	87%
[70,80)	75	8	88	9%	96%
[80,90]	85	4	92	4%	100%
Total		92		100%	

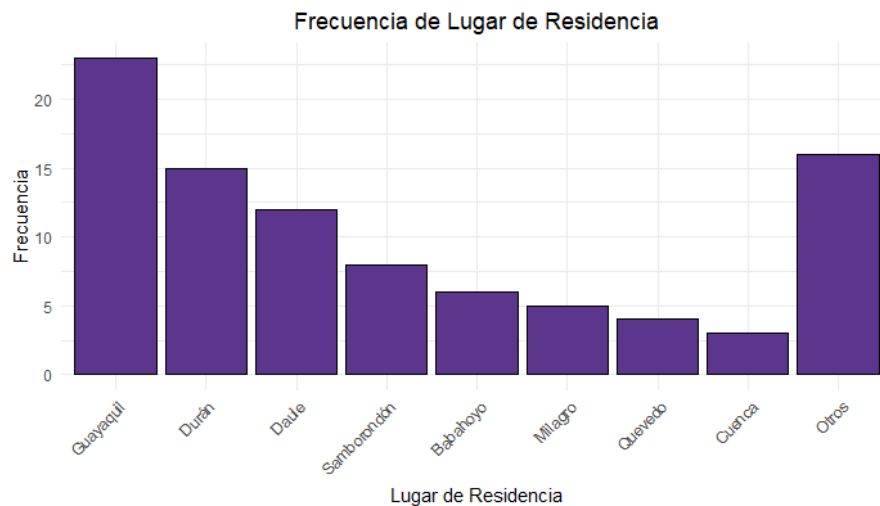
En el siguiente gráfico se puede observar la frecuencia por cáncer primario que presenta cada paciente encuestado:

Figura 3: *Diagrama de barras por tipo de cáncer.*



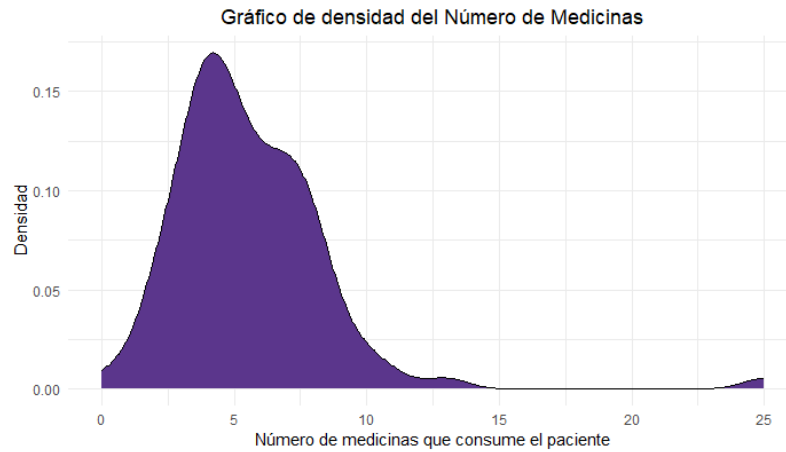
Otra de las variables demográficas de los pacientes es su lugar de residencia, en la figura 4 podemos observar la frecuencia de la ciudad en la que residen:

Figura 4: *Diagrama de barras por lugar de residencia.*



También se les preguntó a los pacientes la cantidad de medicinas que consumen diariamente y el siguiente gráfico muestra la densidad de la variable:

Figura 5: Gráfico de densidad del número de medicinas consumidas.



3.2 ANÁLISIS FACTORIAL CONFIRMATORIO

3.2.1 VALIDACIÓN DE LA ENCUESTA QLQ-C15-PAL

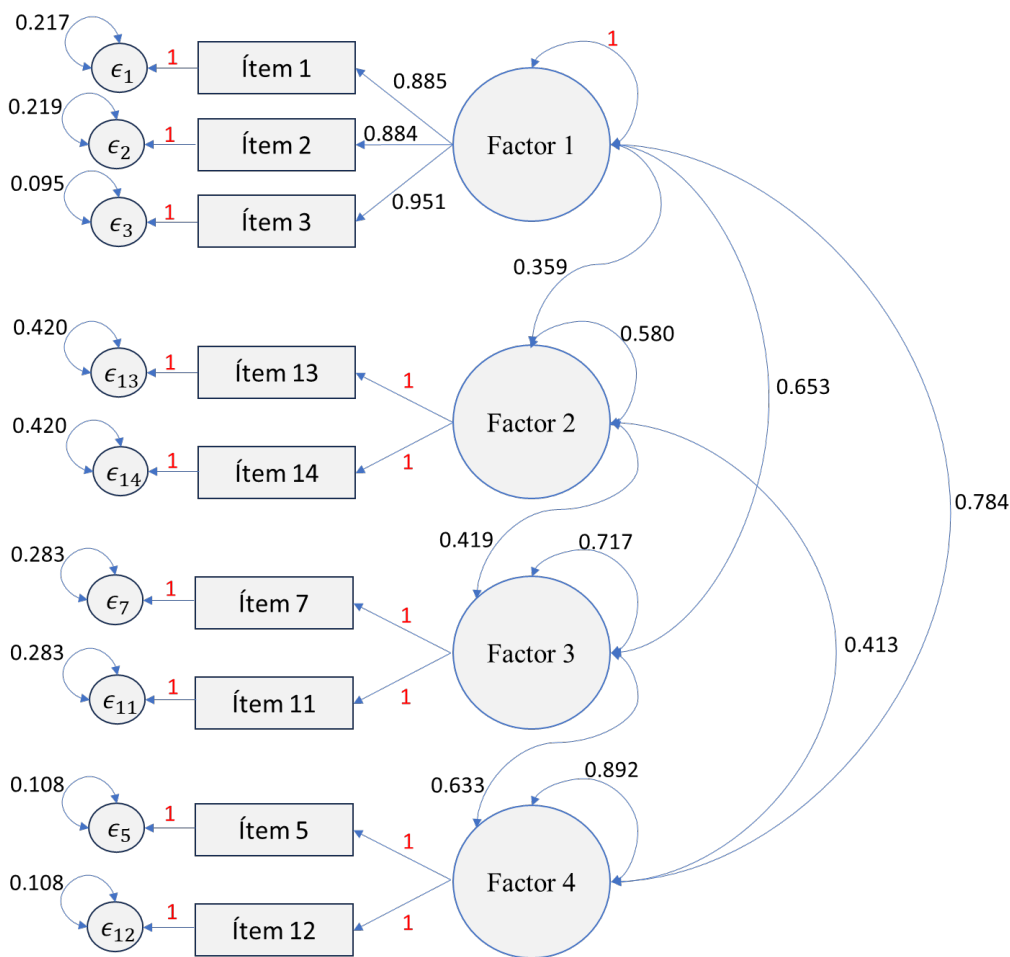
Los indicadores utilizados para medir el ajuste del modelo ACF muestran que existe consistencia de los factores o dominios del modelo teórico del cuestionario QLQ-C15-PAL, en base a los datos observados (Tabla 4). El CFI junto con el TLI sugirió un buen ajuste; de la misma manera, el valor-p de que $RMSEA < 0.05$ fue 0.456, que nos permitió rechazar la hipótesis nula de que el modelo teórico es un mal ajuste. Además, con todas las variables estadísticamente significativas.

Tabla 4: Indicadores de ajuste QLQ-C15-PAL.

Indicador	Valor
CFI	0.997
TLI	0.944
RMSEA	0.051

En la figura 6 se muestra la estimación de los parámetros del modelo. Para el Factor 1 (dominio físico), el ítem 3 tiene la mayor carga factorial y el ítem 2 la menor, aunque los tres ítems del Factor 1 tienen una alta carga y, por lo tanto, una fuerte relación. Por otro lado, el Factor 4 (dominio de dolor) muestra una mayor variabilidad estimada, de 0.892, y también éste junto al Factor 1 tienen una covarianza estimada de 0.784, siendo los dominios mayormente correlacionados.

Figura 6: Parámetros del modelo ACF del QLQ-C15-PAL.



Nota. Factor 1: Dominio Físico; Factor 2: Dominio Emocional; Factor 3: Dominio de Fatiga; Factor 4: Dominio de Dolor.

Las covarianzas del Dominio Físico entre los demás factores fueron significativas; es decir, los pacientes que tienen un valor alto en el Dominio Físico tienden también a tener valores altos en los demás dominios. De la misma manera las covarianzas entre los demás factores. Por último, las varianzas de los factores latentes dados por Dominio Físico, Emocional, Fatiga y Dolor se encuentran fijadas en 1, que indica la estandarización para poder tomar esto como punto de referencia. Las varianzas residuales que fueron estimadas para cada ítem/variable observada indica la variabilidad del ítem después de considerar el efecto de las variables latentes o factores. Los ítems del Dominio Emocional tienen mayor varianza residual.

3.2.2 VALIDACIÓN DE LA ENCUESTA SCNS-SF34

Para la comparación del ajuste del modelo ACF en la encuesta SCNS-SF34 en los datos de la muestra, se hizo uso de los índices de la Tabla 5.

Tabla 5: *Indicadores de ajuste SCNS-SF34.*

Indicador	Valor
CFI	0.984
TLI	0.983
RMSEA	0.078

El CFI indicó un ajuste bueno del modelo al igual que el indicador TLI, el indicador RMSEA es medianamente aceptable, pero el p-valor nos indica que no se puede rechazar la hipótesis nula de que el RMSEA sea menor que 0.05. Esto sugiere que el modelo podría mejorar.

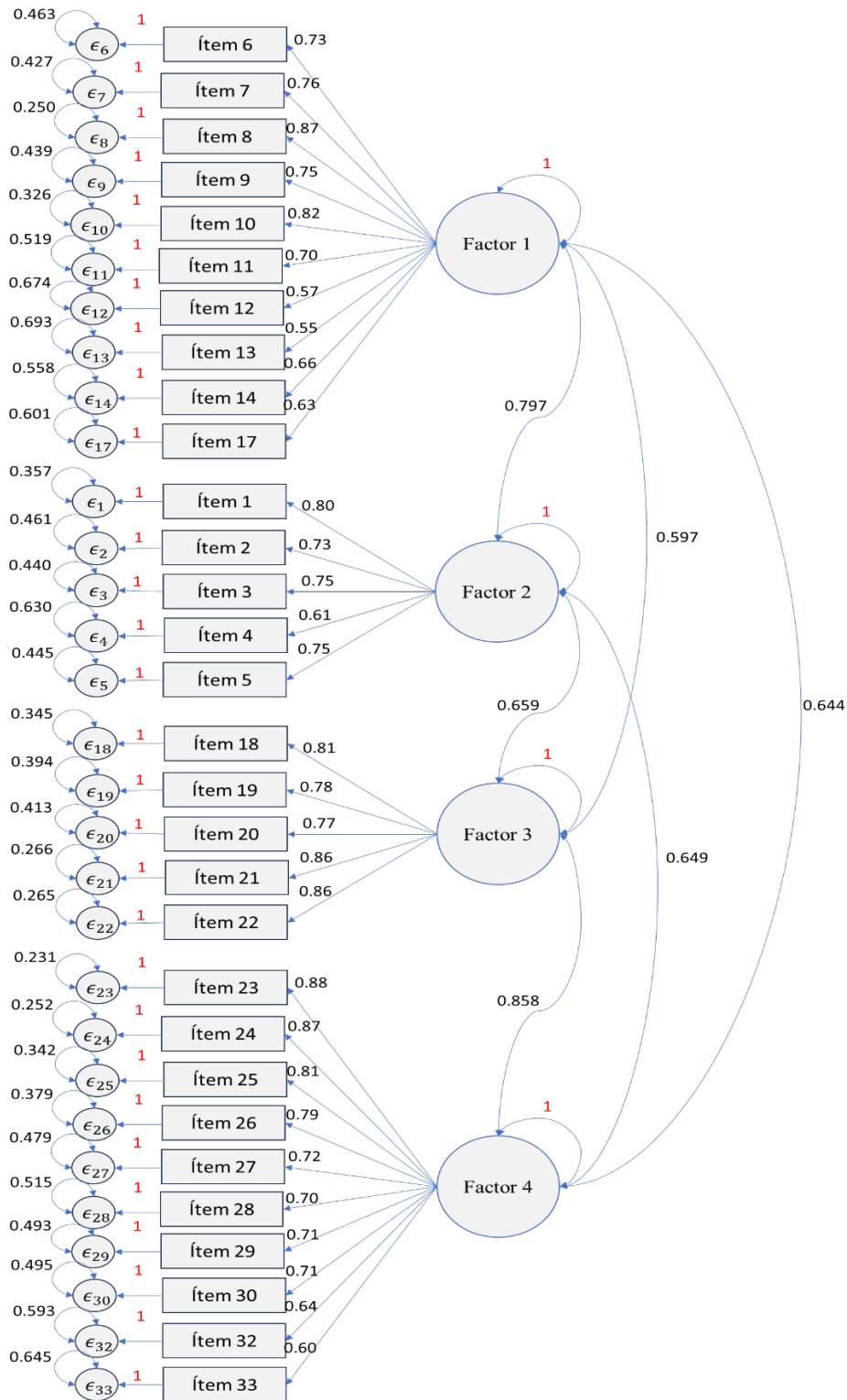
El resumen de las estimaciones del modelo ACF dónde se estima las relaciones de los factores o variables latentes y los ítems o variables observadas se muestra en la figura 7.

Las cargas factoriales, también llamado *Factor Loadings*, representan la relación de los ítems con el factor teorizado; entre más alto el valor indica una mayor relación. Pudimos notar que para el Factor 1 (Necesidades psicológicas) el ítem 8 es aquel que tiene la relación más fuerte, explicando el 82% de la varianza del ítem 8; para el Factor 2 (Necesidades físicas y de la vida diaria), fue el ítem 1 el de mayor valor; el Factor 3 (Necesidades de atención y apoyo al paciente) tiene la relación más fuerte para el ítem 21 y el ítem 22; para el Factor 4 (Necesidades del sistema de salud y de información), el ítem 28 tuvo el mayor vínculo entre todos sus ítems.

Las covarianzas estimadas por el modelo ACF entre todos los factores sugirieron que existe una relación fuerte y significativa entre todas las variables latentes. La covarianza más alta del modelo indica también la relación más fuerte, y esta fue entre el Factor 3 y Factor 4; es decir, los pacientes que tuvieron un mayor nivel de necesidad de atención de apoyo tienden también a tener un mayor grado de necesidad del sistema de salud y de información.

Las varianzas de los dominios son 1, ya que es parte de la estandarización y estas no se interpretan en términos de magnitud. La varianza estimada de los ítems 12 y 13 que se refieren a aprender a sentir que los pacientes tienen el control de la situación y mantener una perspectiva positiva, lo que significa que después de considerar el efecto del Factor 1, estos ítems tiene una suma sustancial de varianza propia del ítem, además de que la carga factorial de estos ítems fue la más baja de todo su dominio.

Figura 7: *Parámetros del modelo ACF del SCNS-SF34.*



Nota. Factor 1: Necesidades psicológicas; Factor 2: Necesidades físicas y de la vida diaria; Factor 3: Necesidades de atención y apoyo al paciente; Factor 4: Necesidades del sistema de salud y de información.

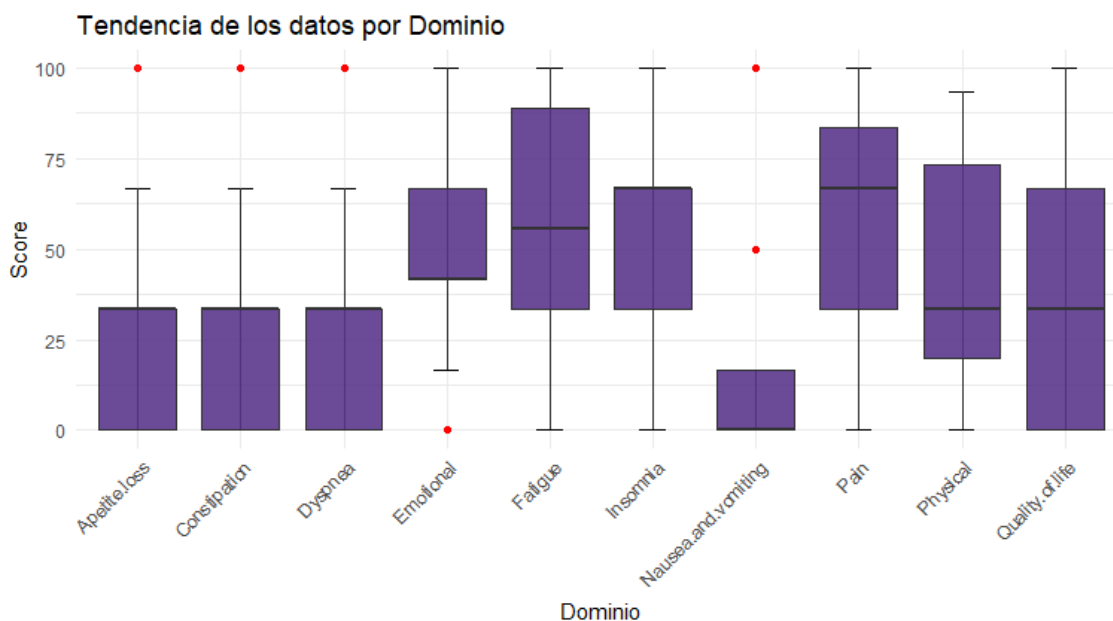
3.3 CARACTERIZACIÓN DE LA CALIDAD DE VIDA

En una primera instancia se analizó la dispersión de las variables numéricas (Figura 8) Valores altos en los dominios de *Quality of Life*, Físico y Emocional indican un alto nivel de funcionamiento y buena calidad de vida. Mientras que, puntajes altos en los constructos sintomáticos, que son Fatiga, Nausea y vómito, Dolor, Disnea, Insomnio, Pérdida de apetito y Constipación, representan alto grado de problemas o sintomatología.

El promedio de *Quality of Life* es alrededor de 39.8. Dicho número representa cierto valor regular de la percepción propia que tiene en promedio un paciente con cáncer en tratamientos paliativos. Más del 25% de los encuestados reportaron QoL de cero y 75% menor a 66.67.

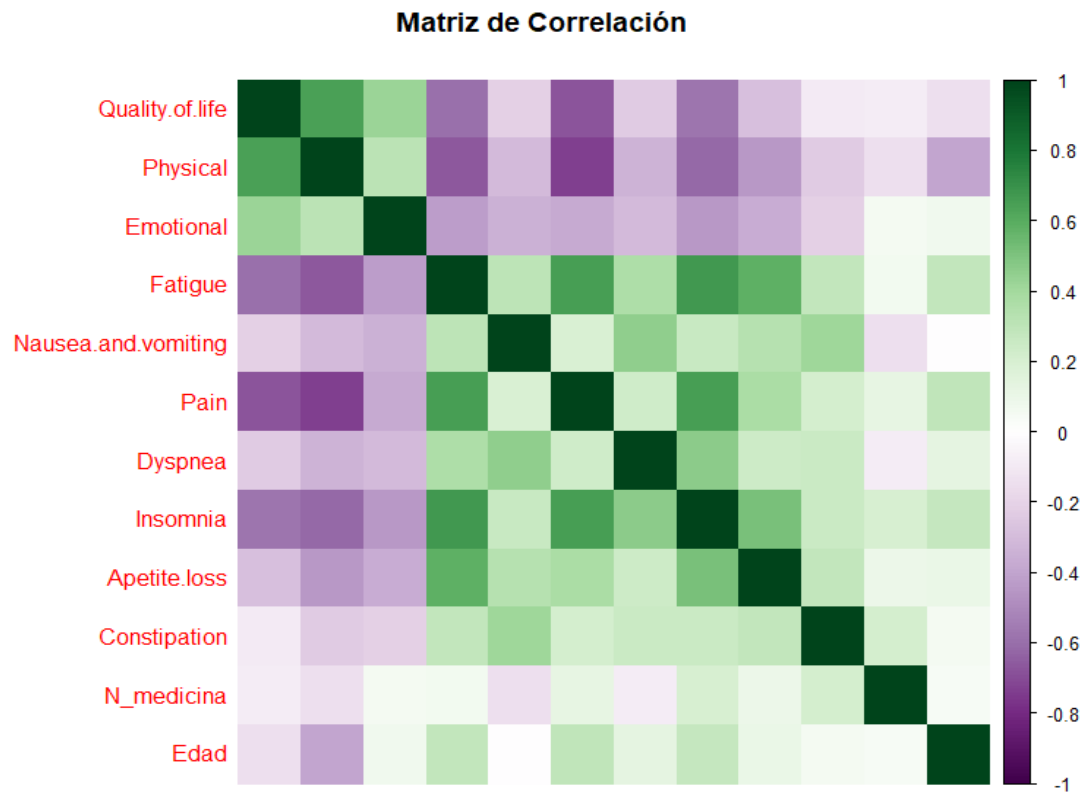
Los síntomas con mayor frecuencia son fatiga y dolor, pero también son los de mayor variabilidad en los encuestados.

Figura 8: *Boxplot de cada dominio QoL.*



Para identificar y explorar visualmente si existe alguna relación lineal entre pares de variables en un conjunto de datos se hace uso de la matriz de correlación (Figura 9).

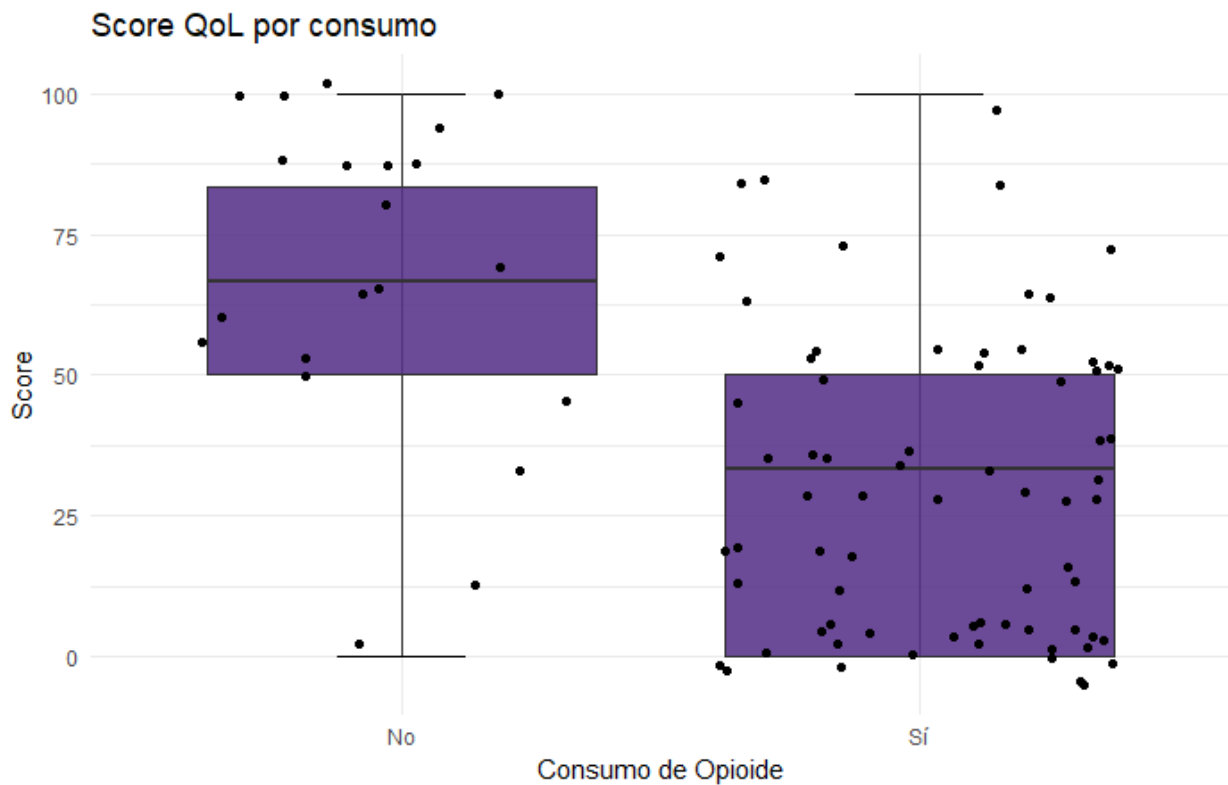
Figura 9: Matriz de Correlación.



En esta matriz de correlación claramente se vio aquellas variables (Edad, Número de medicinas “N_medicina”) que aportan en menor medida redundancia, lo que evita llegar a tener problemas de multicolinealidad y a simplificar el modelo. Lo mencionado anteriormente sobre la interpretación de los dominios *Quality of Life*, Físico y Emocional se notó también en la matriz de correlación, dado que estas 3 variables tienen una correlación negativa con respecto al resto de variables o dominios.

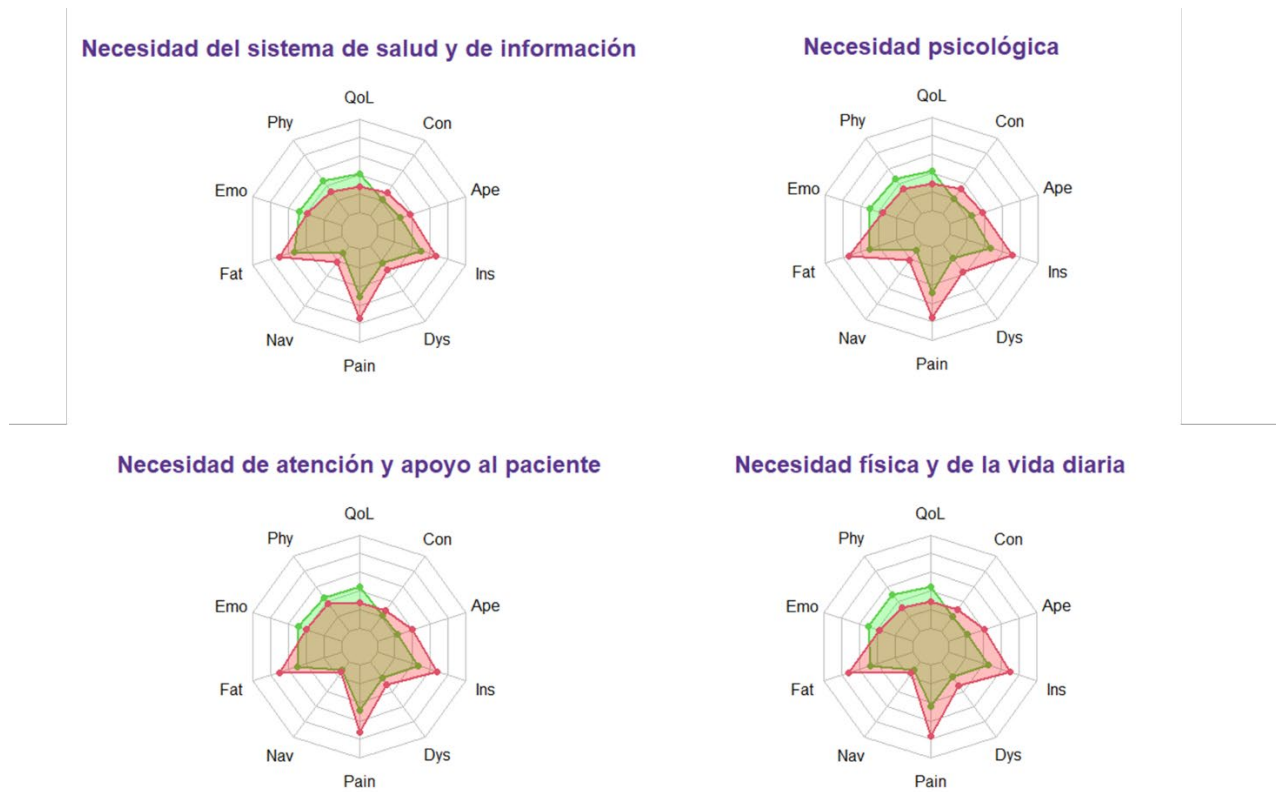
El tema de los opioides suele ser regular al momento de llegar al departamento de cuidados paliativos como parte del plan de alivio del dolor (ACS, 2019). En la fig. 10 se mostró la importante diferencia que existe en el promedio de *Score Quality of Life* en función de su consumo de opioide para tratar el dolor.

Figura 10: *Score QoL en función del consumo de opiode.*



Finalmente, para visualizar la variación de los puntajes en cada constructo, de la tabla 1, relacionado a la calidad de vida en función de la existencia o no de necesidad en cada dominio en la fig. 10. Se muestra que la mayor diferencia existe en las necesidades físicas y de la vida diaria y las necesidades psicológicas; es decir, cuando existe una necesidad del paciente en ese dominio los puntajes de los constructos de calidad de vida tienen una variación sustancial versus aquellos puntajes cuando no existe tal necesidad. Es de notar que hay un comportamiento similar en todos los dominios de necesidad del paciente.

Figura 11: Diagrama de radar de QoL versus Necesidades.



Nota. Color verde: el paciente no tiene necesidad. Color rojo: el paciente tiene necesidad.

3.4 RESULTADOS DE LOS MODELOS DE CLASIFICACIÓN

En esta sección se muestran los resultados obtenidos de todos los modelos de clasificación aplicados en cada dominio: psicológico, físico, cuidado y apoyo, salud e información.

3.4.1 DOMINIO PSICOLÓGICO

Regresión logística

Se ajustó un modelo de regresión logística para clasificar la necesidad psicológica del paciente, a partir de las 13 variables predictoras y se alcanzó una exactitud de 0.65 en el conjunto de entrenamiento. De acuerdo con el criterio del valor p, con un nivel de significancia del 5%, se

obtuvo que la variable más significativa es la que corresponde a la fatiga del paciente (valor $p = 0.04$). Además, la variable del componente emocional también se acerca mucho al nivel de significancia (valor $p = 0.1$).

Figura 12: Estimación de coeficientes del dominio psicológico del modelo de regresión logística.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6914 -0.6106 -0.2524  0.5118  2.1219

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.283e+00  3.754e+00 -2.473  0.0134 *
Quality.of.life  2.562e-02  2.043e-02  1.254  0.2098
Physical       2.355e-02  2.555e-02  0.922  0.3567
Emotional     -3.467e-02  2.166e-02 -1.601  0.1094
Fatigue       4.907e-02  2.338e-02  2.099  0.0358 *
Nausea.and.vomiting 1.579e-02  2.411e-02  0.655  0.5126
Pain          2.739e-02  2.209e-02  1.240  0.2151
Dyspnea       8.873e-03  1.788e-02  0.496  0.6198
Insomnia      6.486e-05  2.146e-02  0.003  0.9976
Apetite.Loss -1.670e-02  2.264e-02 -0.738  0.4608
Constipation  1.451e-02  1.561e-02  0.929  0.3527
N_medicina    -2.046e-02  1.374e-01 -0.149  0.8816
Edad          2.346e-02  2.503e-02  0.937  0.3487
Opiode1       2.217e+00  2.076e+00  1.068  0.2857
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 84.801 on 68 degrees of freedom
Residual deviance: 53.770 on 55 degrees of freedom
AIC: 81.77

Number of Fisher Scoring iterations: 7
```

Aplicando el modelo de regresión logística en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

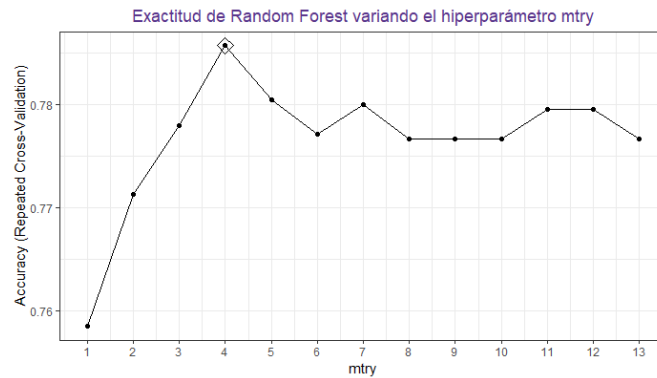
Tabla 6: Matriz de confusión del dominio psicológico del modelo de regresión logística.

		Actual	
		No	Si
Predicción	No	12	4
	Si	4	3

Random forests

Ajustando el modelo de *random forests* para clasificar la necesidad psicológica, se alcanzó una exactitud de 0.79 en el conjunto de entrenamiento con un hiperparámetro *mtry* de 4, como se muestra en la figura.

Figura 13: *Accuracy del hiperparámetro mtry del dominio psicológico.*



Aplicando el modelo de *random forests* en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

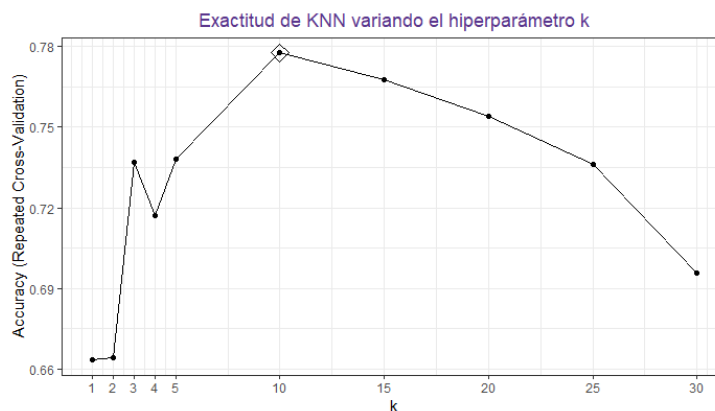
Tabla 7: *Matriz de confusión del dominio psicológico del modelo random forests.*

		Actual	
		No	Si
Predicción	No	14	5
	Si	2	2

KNN

Ajustando el modelo de KNN para clasificar la necesidad psicológica, se alcanzó una exactitud de 0.78 en el conjunto de entrenamiento con un hiperparámetro k de 10, como se muestra en la figura.

Figura 14: *Accuracy del hiperparámetro k del dominio psicológico.*



Aplicando el modelo de KNN en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

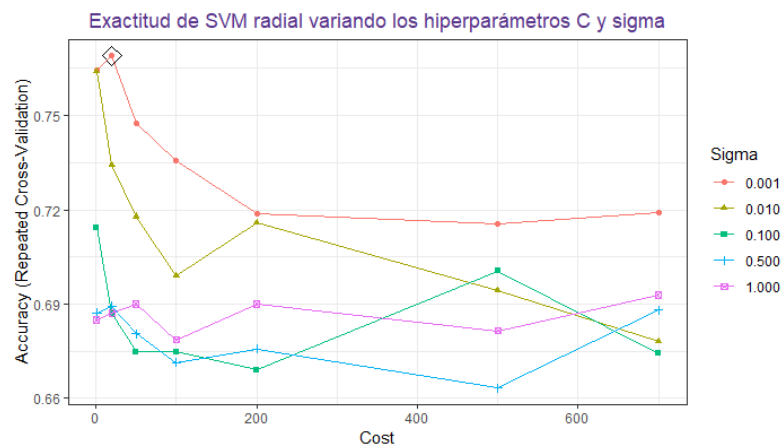
Tabla 8: Matriz de confusión del dominio psicológico del modelo KNN.

		Actual	
		No	Si
Predicción	No	13	4
	Si	3	3

SVM

Ajustando el modelo de SVM con *kernel* radial para clasificar la necesidad psicológica, se alcanzó una exactitud de 0.77 en el conjunto de entrenamiento con los siguientes hiperparámetros: sigma de 0.001 y C de 20, como se muestra en la figura.

Figura 15: Accuracy de los hiperparámetros C y sigma del dominio psicológico.



Aplicando el modelo de SVM en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

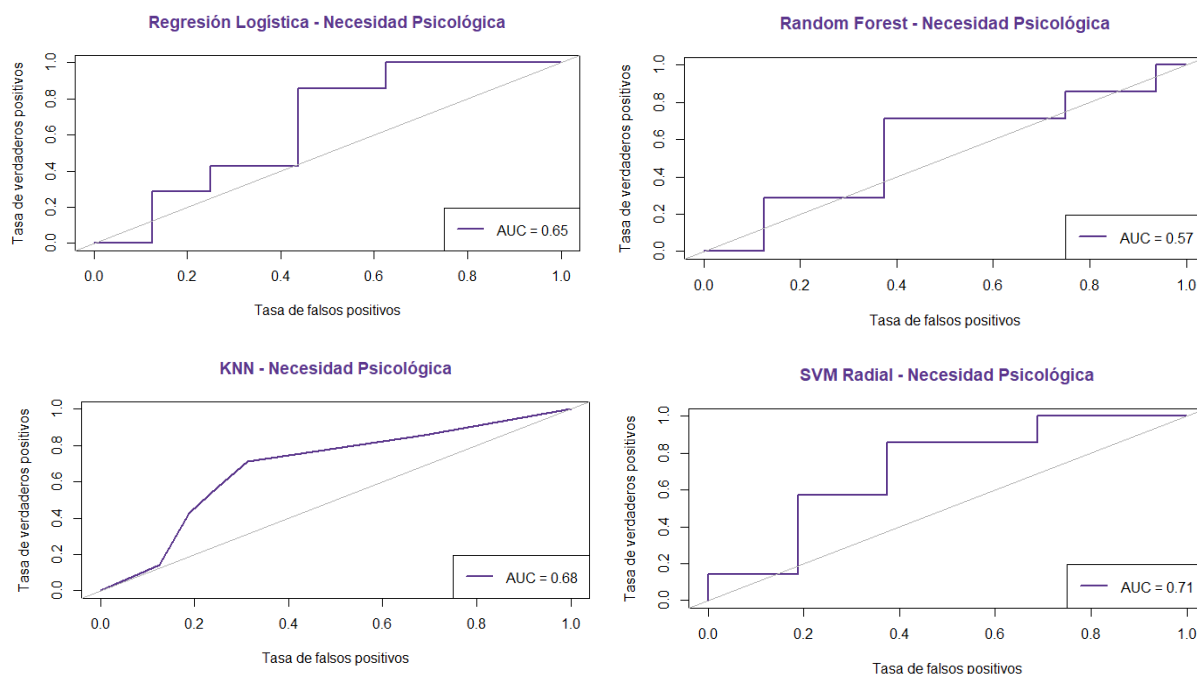
Tabla 9: Matriz de confusión del dominio psicológico del SVM.

		Actual	
		No	Si
Predicción	No	13	5
	Si	3	2

Comparación de modelos

Se realizó una comparación de rendimiento de los modelos ajustados a través de la curva ROC construida con datos del conjunto de prueba, como resultado se observa que el modelo SVM presenta una mayor área bajo la curva (AUC de 0.71).

Figura 16: Comparación de modelos del dominio psicológico.



A continuación, se tienen las métricas de evaluación obtenidas para cada modelo:

Tabla 10: Comparación de métricas del dominio psicológico.

Modelo	Exactitud	Sensibilidad	Especificidad	Kappa
Regresión logística	0,65	0,43	0,75	0,18
<i>Random forest</i>	0,70	0,29	0,88	0,18
KNN	0,70	0,43	0,81	0,25
SVM	0,65	0,29	0,81	0,11

3.4.2 DOMINIO FÍSICO

Regresión logística

Se ajustó un modelo de regresión logística para clasificar la necesidad física del paciente, a partir de las 13 variables predictoras y se alcanzó una exactitud de 0.57 en el conjunto de

entrenamiento. De acuerdo con el criterio del valor p, a un nivel de significancia del 5%, se obtuvo que la variable más significativa es la que corresponde al dolor del paciente (valor p = 0.03).

Figura 17: Estimación de coeficientes del dominio físico del modelo de regresión logística.

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5889  -0.7872  -0.3813   0.8768   2.1194

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.461081   2.843884  -2.272  0.0231 *
Quality.of.life  0.016083   0.017001   0.946  0.3442
Physical      0.026336   0.018727   1.406  0.1596
Emotional    -0.018904   0.016232  -1.165  0.2442
Fatigue      0.027412   0.018548   1.478  0.1394
Nausea.and.vomiting 0.004756   0.020879   0.228  0.8198
Pain         0.042821   0.019516   2.194  0.0282 *
Dyspnea     0.002717   0.016711   0.163  0.8708
Insomnia    -0.008073   0.017201  -0.469  0.6388
Apetite.loss 0.015210   0.014715   1.034  0.3013
Constipation 0.000884   0.013872   0.064  0.9492
N_medicina  0.161809   0.122961   1.316  0.1882
Edad        -0.006024   0.019786  -0.304  0.7608
Opiode1    -0.166928   1.133137  -0.147  0.8829
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 93.351  on 69  degrees of freedom
Residual deviance: 70.959  on 56  degrees of freedom
AIC: 98.959

Number of Fisher Scoring iterations: 5

```

Aplicando el modelo de regresión logística en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

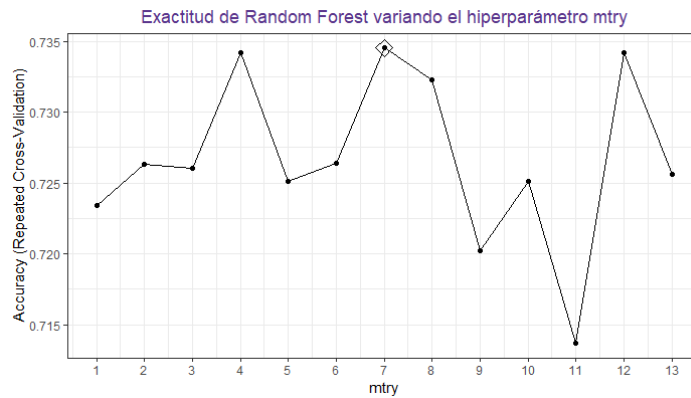
Tabla 11: Matriz de confusión del dominio físico del modelo de regresión logística.

		Actual	
		No	Si
Predicción	No	14	6
	Si	0	2

Random forest

Ajustando el modelo de *random forest* para clasificar la necesidad física, se alcanzó una exactitud de 0.73 en el conjunto de entrenamiento con un hiperparámetro *mtry* de 7, como se muestra en la figura.

Figura 18: *Accuracy del hiperparámetro mtry del dominio físico.*



Aplicando el modelo de *random forest* en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

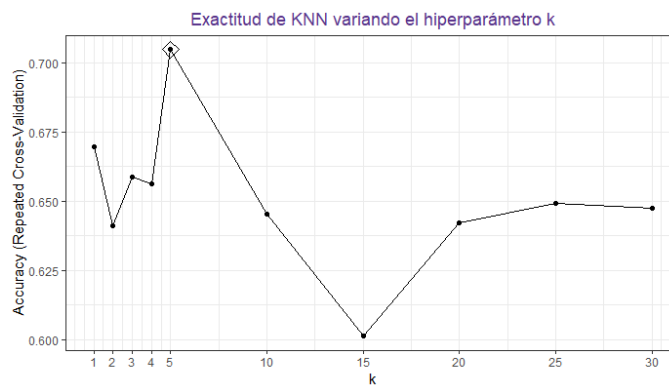
Tabla 12: *Matriz de confusión del dominio físico del modelo random forests.*

		Actual	
		No	Si
Predicción	No	14	7
	Si	0	1

KNN

Ajustando el modelo de KNN para clasificar la necesidad psicológica, se alcanzó una exactitud de 0.70 en el conjunto de entrenamiento con un hiperparámetro k de 5, como se muestra en la figura 15.

Figura 19: *Accuracy del hiperparámetro k del dominio físico.*



Aplicando el modelo de KNN en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

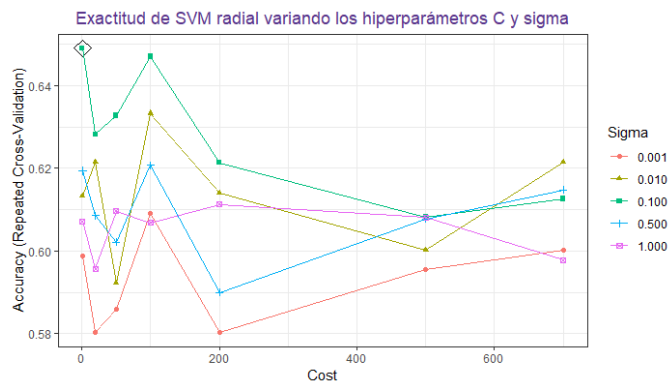
Tabla 13: Matriz de confusión del dominio físico del modelo KNN.

		Actual	
		No	Si
Predicción	No	13	6
	Si	1	2

SVM

Ajustando el modelo de SVM con *kernel* radial para clasificar la necesidad física, se alcanzó una exactitud de 0.65 en el conjunto de entrenamiento con los siguientes hiperparámetros: sigma de 0.1 y C de 1, como se muestra en la figura.

Figura 20: Accuracy de los hiperparámetros C y sigma del dominio físico.



Aplicando el modelo de SVM en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

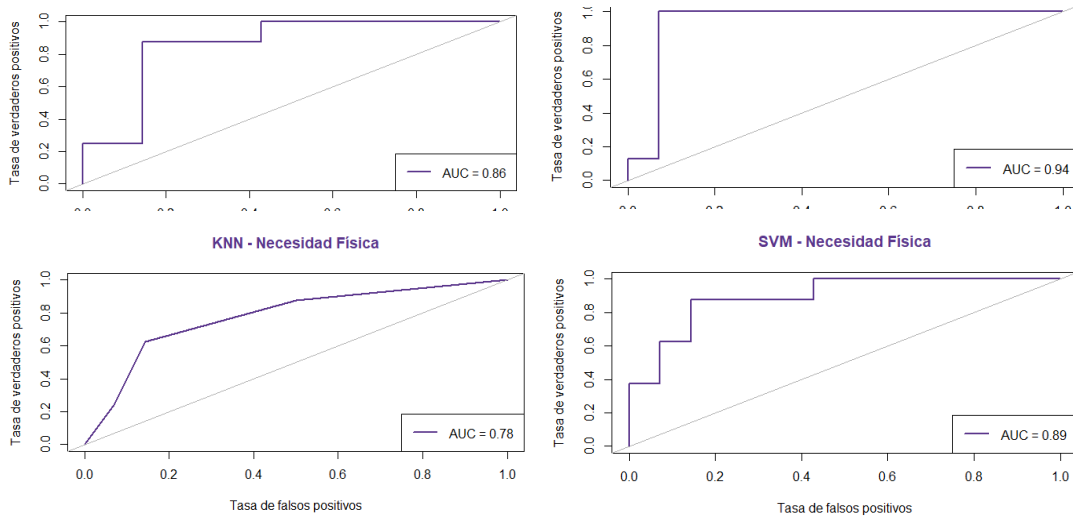
Tabla 13: Matriz de confusión del dominio físico del modelo SVM.

		Actual	
		No	Si
Predicción	No	13	5
	Si	1	3

Comparación de modelos

Se realizó una comparación de rendimiento de los modelos ajustados a través de la curva ROC construida con datos del conjunto de prueba, como resultado se observa que el modelo *random forest* presenta una mayor área bajo la curva (AUC de 0.94).

Figura 21: Comparación de modelos del dominio físico.



A continuación, se tienen las métricas de evaluación obtenidas para cada modelo:

Tabla 14: Comparación de métricas del dominio físico.

Modelo	Exactitud	Sensibilidad	Especificidad	Kappa
Regresión logística	0,73	0,25	1,00	0,30
<i>Random forest</i>	0,68	0,13	1,00	0,15
KNN	0,68	0,25	0,93	0,21
SVM	0,73	0,38	0,93	0,34

3.4.3 DOMINIO DE CUIDADO Y ATENCIÓN

Regresión logística

Se ajustó un modelo de regresión logística para clasificar la necesidad de cuidado y atención del paciente, a partir de las 13 variables predictoras y se alcanzó una exactitud de 0.62 en el conjunto de entrenamiento. De acuerdo con el criterio del valor p, con un nivel de significancia del 5%, se obtuvo que las variables más significativas son las que corresponden al estado físico del paciente (valor $p = 0.005$) y el dolor (0.02).

Figura 22: Estimación de coeficientes del dominio de cuidado y atención del modelo de regresión logística.

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5459  -0.8210  -0.4394   0.7999   2.1859

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.957424   2.566745  -2.711  0.00672 **
Quality.of.life -0.010048   0.013728  -0.732  0.46419
Physical      0.060084   0.021195   2.835  0.00459 **
Emotional    -0.009980   0.016182  -0.617  0.53742
Fatigue      0.022511   0.019002   1.185  0.23615
Nausea.and.vomiting -0.001822   0.023030  -0.079  0.93695
Pain         0.045816   0.019426   2.358  0.01835 *
Dyspnea     0.010376   0.014809   0.701  0.48351
Insomnia    -0.005252   0.017410  -0.302  0.76293
Appetite.loss 0.008297   0.017351   0.478  0.63253
Constipation -0.011050   0.015224  -0.726  0.46796
N_medicina  0.151476   0.128653   1.177  0.23904
Edad        0.009141   0.018885   0.484  0.62836
Opioid     -0.772161   1.093692  -0.706  0.48018
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 91.246  on 69  degrees of freedom
Residual deviance: 71.665  on 56  degrees of freedom
AIC: 99.665

Number of Fisher Scoring iterations: 4
    
```

Aplicando el modelo de regresión logística en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

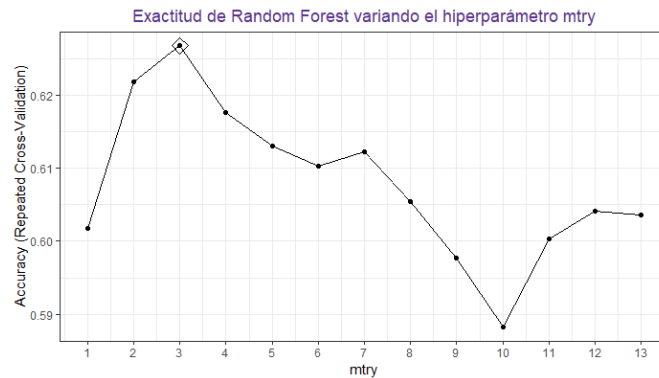
Tabla 15: Matriz de confusión del dominio de cuidado y atención del modelo de regresión logística

		Actual	
		No	Si
Predicción	No	10	5
	Si	4	3

Random forest

Ajustando el modelo de *random forest* para clasificar la necesidad física, se alcanzó una exactitud de 0.63 en el conjunto de entrenamiento con un hiperparámetro *mtry* de 3, como se muestra en la figura.

Figura 23: *Accuracy del hiperparámetro mtry del dominio de cuidado y atención.*



Aplicando el modelo de *random forest* en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

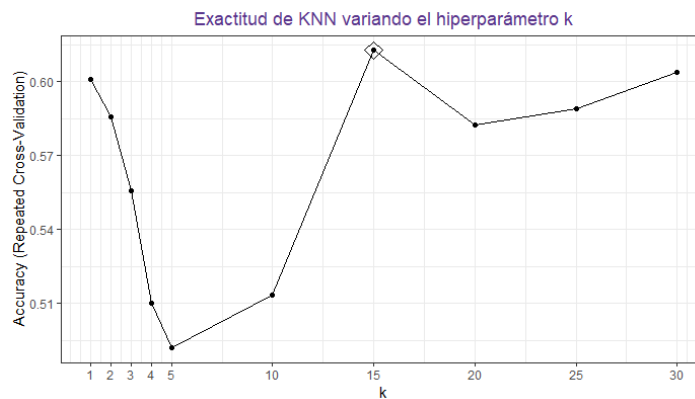
Tabla 16: *Matriz de confusión del dominio de cuidado y atención del modelo random forests.*

		Actual	
		No	Si
Predicción	No	14	5
	Si	0	3

KNN

Ajustando el modelo de KNN para clasificar la necesidad psicológica, se alcanzó una exactitud de 0.61 en el conjunto de entrenamiento con un hiperparámetro k de 15, como se muestra en la figura.

Figura 24: *Accuracy del hiperparámetro k del dominio de cuidado y atención.*



Aplicando el modelo de KNN en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

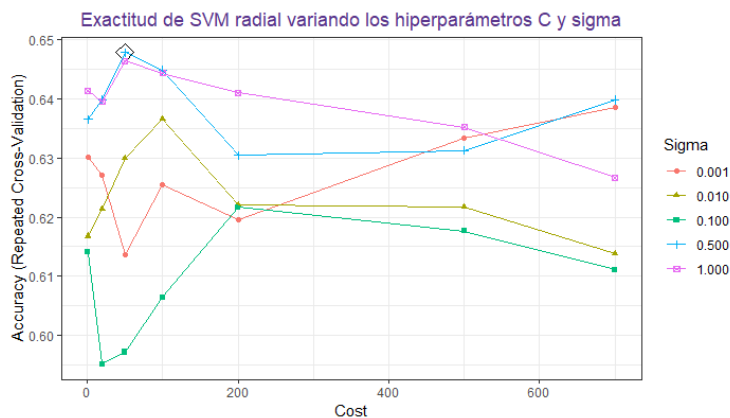
Tabla 17: Matriz de confusión del dominio de cuidado y atención del modelo KNN.

		Actual	
		No	Si
Predicción	No	12	5
	Si	2	3

SVM

Ajustando el modelo de SVM con *kernel* radial para clasificar la necesidad física, se alcanzó una exactitud de 0.65 en el conjunto de entrenamiento con los siguientes hiperparámetros: sigma de 0.5 y C de 50, como se muestra en la figura.

Figura 25: Accuracy del hiperparámetro C y sigma del dominio de cuidado y atención.



Aplicando el modelo de SVM en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

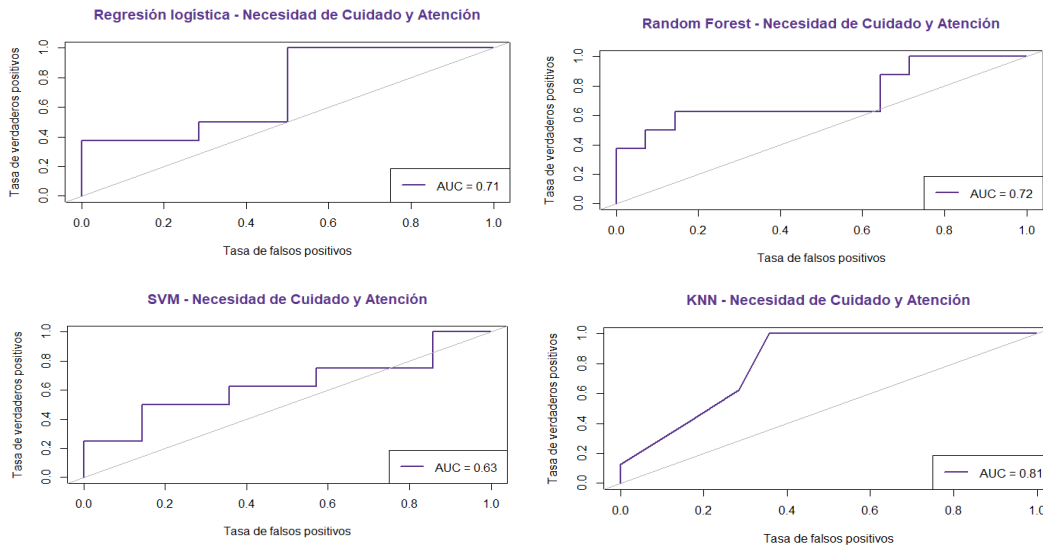
Tabla 18: Matriz de confusión del dominio de cuidado y atención del modelo SVM.

		Actual	
		No	Si
Predicción	No	14	6
	Si	0	2

Comparación de modelos

Se realizó una comparación de rendimiento de los modelos ajustados a través de la curva ROC construida con datos del conjunto de prueba, como resultado se observa que el modelo KNN presenta una mayor área bajo la curva (AUC de 0.81).

Figura 26: Comparación de modelos del dominio de cuidado y atención.



A continuación, se tienen las métricas de evaluación obtenidas para cada modelo:

Tabla 19: Comparación de métricas de cuidado y atención.

Modelo	Exactitud	Sensibilidad	Especificidad	Kappa
Regresión logística	0,59	0,38	0,71	0,09
<i>Random forest</i>	0,77	0,38	1,00	0,43
KNN	0,68	0,38	0,86	0,25
SVM	0,73	0,25	1,00	0,30

3.4.4 DOMINIO DE SALUD E INFORMACIÓN

Regresión logística

Se ajustó un modelo de regresión logística para clasificar la necesidad de salud e información del paciente, a partir de las 13 variables predictoras y se alcanzó una exactitud de 0.65 en el conjunto de entrenamiento. Sin embargo, vemos que ninguna de las variables es significativa para la clasificación de la necesidad.

Figura 27: Estimación de coeficientes del dominio de salud e información del modelo de regresión logística.

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4090 -0.6638 -0.4442  0.2063  2.3212

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.682258   3.016079  -2.216  0.0267 *
Quality.of.life  0.021699   0.018626   1.165  0.2440
Physical     -0.001583   0.021783  -0.073  0.9421
Emotional     0.006521   0.018236   0.358  0.7206
Fatigue       0.010716   0.018739   0.572  0.5674
Nausea.and.vomiting 0.037996   0.024031   1.581  0.1139
Pain          0.008842   0.022138   0.399  0.6896
Dyspnea       0.018924   0.019436   0.974  0.3302
Insomnia     -0.005246   0.021268  -0.247  0.8052
Apetite.loss  0.017325   0.017924   0.967  0.3337
Constipation -0.005440   0.015316  -0.355  0.7225
N_medicina    0.177413   0.144617   1.227  0.2199
Edad          0.013245   0.019393   0.683  0.4946
Opioide1      0.807104   1.439943   0.561  0.5751
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 79.807  on 69  degrees of freedom
Residual deviance: 63.708  on 56  degrees of freedom
AIC: 91.708

Number of Fisher Scoring iterations: 5

```

Aplicando el modelo de regresión logística en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

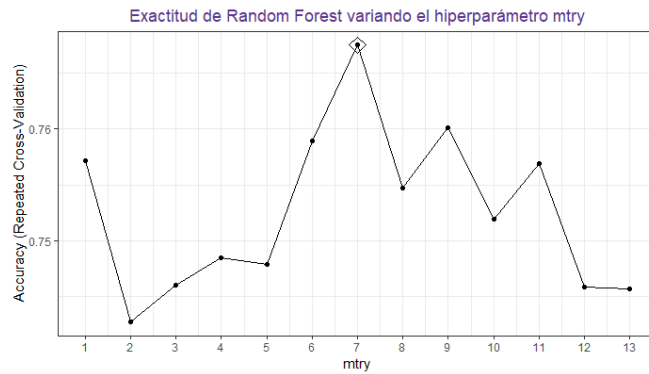
Tabla 20: Matriz de confusión del dominio de salud e información del modelo de regresión logística.

		Actual	
		No	Si
Predicción	No	11	3
	Si	6	2

Random forest

Ajustando el modelo de *random forest* para clasificar la necesidad física, se alcanzó una exactitud de 0.77 en el conjunto de entrenamiento con un hiperparámetro *mtry* de 7, como se muestra en la figura.

Figura 28: *Accuracy del hiperparámetro mtry del dominio de salud e información.*



Aplicando el modelo de *random forest* en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

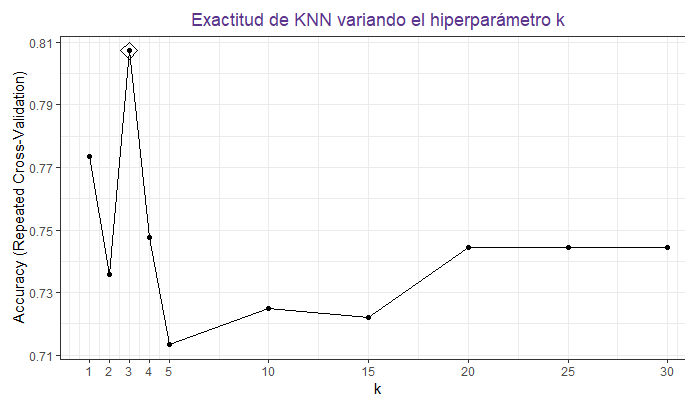
Tabla 21: *Matriz de confusión del dominio de salud e información del modelo random forests.*

		Actual	
		No	Si
Predicción	No	14	5
	Si	3	0

KNN

Ajustando el modelo de KNN para clasificar la necesidad psicológica, se alcanzó una exactitud de 0.81 en el conjunto de entrenamiento con un hiperparámetro k de 3, como se muestra en la figura.

Figura 29: *Accuracy del hiperparámetro k del dominio salud e información.*



Aplicando el modelo de KNN en el conjunto de prueba, se obtuvo la siguiente matriz de confusión:

Tabla 22: Matriz de confusión del dominio de salud e información del modelo KNN.

		Actual	
		No	Si
Predicción	No	12	3
	Si	5	2

SVM

Ajustando el modelo de SVM con *kernel* radial para clasificar la necesidad física, se alcanzó una exactitud de 0.75 en el conjunto de entrenamiento con los siguientes hiperparámetros: sigma de 0.01 y C de 20, como se muestra en la figura.

Figura 30: Accuracy de los hiperparámetros C y sigma del dominio de salud e información.

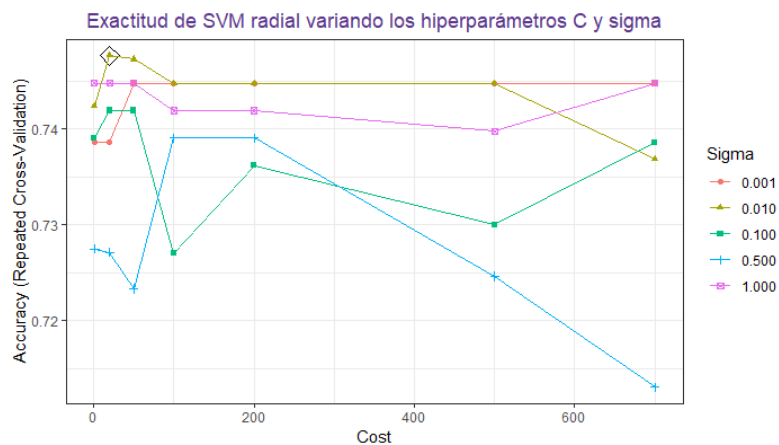


Tabla 23: Matriz de confusión del dominio de salud e información del modelo SVM.

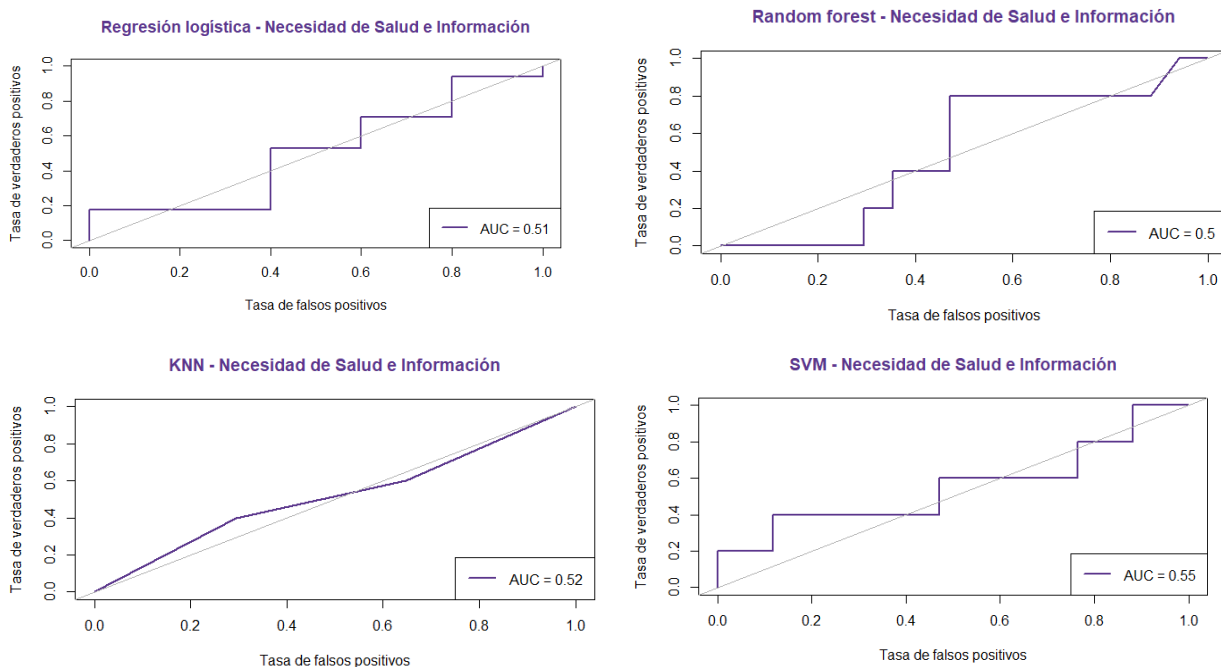
		Actual	
		No	Si
Predicción	No	17	5
	Si	0	0

Comparación de modelos

Se realizó una comparación de rendimiento de los modelos ajustados a través de la curva ROC construida con datos del conjunto de prueba, como se observa en la figura todos los modelos

presenta un área bajo la curva muy cercana a 0.5, lo cual indica una predicción totalmente aleatoria para la necesidad de salud e información.

Figura 31: Comparación de modelos del dominio de salud e información.



A continuación, se tienen las métricas de evaluación obtenidas para cada modelo:

Tabla 24: Comparación de métricas del dominio de salud e información.

Modelo	Exactitud	Sensibilidad	Especificidad	Kappa
Regresión logística	0,59	0,40	0,35	0,04
Random forest	0,64	0,00	0,21	-0,21
KNN	0,64	0,40	0,71	0,09
SVM	0,77	0,00	1,00	0,00

Capítulo 4

Conclusiones

Con el objetivo de hallar la relación entre la calidad de vida y las necesidades de atención de apoyo de pacientes oncológicos en tratamientos paliativos, a través de diversas técnicas estadísticas, se concluye que:

1. La implementación rutinaria de encuestas especializadas para medir las necesidades de atención de apoyo y la calidad de vida en pacientes oncológicos en tratamientos paliativos proporciona información valiosa sobre las áreas y aspectos de la vida de los pacientes que requieren una mayor atención. Esta información también permitirá conocer las necesidades particulares del paciente. En consecuencia, esta información ayuda a los equipos médicos, trabajadores sociales y enfermeros a mejorar la gestión de los síntomas y problemas funcionales del paciente para garantizar una atención de calidad.
2. La fiabilidad de los constructos teóricos, a través de la validación realizada, asegura que los resultados sobre la calidad de vida y las necesidades de atención de apoyo obtenidos a partir de las encuestas tienen fundamentos sólidos para el uso en la población de pacientes con cáncer en tratamientos paliativos. Además, contribuye al crecimiento del conocimiento científico y de la literatura médica en el Ecuador, asentando las bases para posibles mejoras en los procedimientos y métodos.
3. La caracterización de los datos de calidad de vida mediante métodos validados dará la facultad a los profesionales de la salud priorizar tratamientos e intervenciones que aborden los principales problemas del paciente. No focalizar todos los esfuerzos en aplicar tratamientos de alivio al dolor; de tal manera, que se permita ampliar el abanico de opciones de apoyo.
4. La implementación de estos modelos de clasificación brinda una capacidad moderada para asignar las necesidades particulares a cada paciente según su calidad de vida. La comparación exhaustiva, respaldada por diferentes métricas de evaluación, permite

determinar el modelo más recomendado para utilizar en cada dominio de necesidad.

Además, el estudio permitió identificar las variables asociadas con la calidad de vida que ejercen una influencia significativa en cada ámbito de necesidad, fortaleciendo así la comprensión integral de los factores influyentes en el bienestar de los pacientes.

Discusión de los resultados

- La forma de atención de los pacientes en la clínica, que era por orden de llegada, el filtro de consentimiento de los pacientes y el corto tiempo impidió la aplicación del muestreo probabilístico. Al ser una muestra no probabilística, toda inferencia debe ser interpretada con cautela.
- Debido al corto periodo de tiempo para la implementación de los cuestionarios se obtuvo una cantidad de muestra no muy grande, con un nivel de error mayor al deseado, lo que ocasionó que los modelos tengan una capacidad predictiva limitada, especialmente para los pacientes que se habían clasificado como que sí necesitaban soporte en algunas necesidades. Sin embargo, se considera que con una mayor muestra para entrenar los modelos se podría conseguir una mayor capacidad predictiva de los modelos de *machine learning*, en especial con respecto a la sensibilidad.
- La falta de una pregunta en el cuestionario de necesidades de apoyo puede afectar los resultados obtenidos en el dominio de salud e información, ya que esta pregunta forma parte de la construcción de la necesidad de salud e información.

Recomendaciones

Este estudio se beneficiaría de un estudio confirmatorio donde se subsanen los problemas discutidos en la sección de discusión en términos de representatividad, tamaño de muestra y validación de los dominios. Adicionalmente se realizan las siguientes recomendaciones:

1. Implementar regularmente en los pacientes de cuidados paliativos las encuestas EORTC QLQ-C15-PAL y una modificación del SCNS-SF34 para seguir recolectando información que ayude a mejorar el apoyo que los centros de atención les brindan a los pacientes de acuerdo con sus necesidades.
2. Adaptar el cuestionario SCNS-SF34, para que sus preguntas permitan medir la necesidad del paciente de cuidados paliativos solo en los dominios que se encontró mayor relación con la calidad de vida del paciente: psicológico, físico, cuidado y atención.
3. Realizar una comparativa en el rendimiento con otros modelos de *machine learning* o redes neuronales que no se abarcaron en esta investigación.
4. Utilizar una mayor cantidad de variables demográficas ayudará a fortalecer la caracterización de la calidad de vida en pacientes oncológicos.

BIBLIOGRAFÍA

Barofsky, I. (2011). Why perform a quality or quality-of-life assessment? *Quality of Life Research*, 21(4), 633-636. <https://doi.org/10.1007/s11136-011-9962-z>

Beysolow, T., II. (2017). Introduction to Deep Learning Using R. En *Apress eBooks*.
<https://doi.org/10.1007/978-1-4842-2734-3>

Bjorner, J. B., Petersen, M. D., Groenvold, M., Aaronson, N. K., Ahlner-Elmqvist, M., Arraras, J. I., Brédart, A., Fayers, P., Jordhøy, M. S., Sprangers, M. A. G., Watson, M., & Young, T. L. (2004). Use of item response theory to develop a shortened version of the EORTC QLQ-C30 emotional functioning scale. *Quality of Life Research*, 13(10), 1683-1697.
<https://doi.org/10.1007/s11136-004-7866-x>

Boyes, A., Girgis, A., & Lecathelinais, C. (2009). Brief assessment of adult cancer patients' perceived needs: Development and validation of the 34-item Supportive Care Needs Survey (SCNS-SF34). *Journal of Evaluation in Clinical Practice*, 15(4), 602-606.
<https://doi.org/10.1111/j.1365-2753.2008.01057.x>

Brown, T. A. (2007). Confirmatory factor analysis for applied research. *Choice Reviews Online*, 44(05), 44-2769. <https://doi.org/10.5860/choice.44-2769>

Faguet, G. B. (2016). Quality End-of-Life cancer care: an overdue imperative. *Critical Reviews in Oncology Hematology*, 108, 69-72. <https://doi.org/10.1016/j.critrevonc.2016.10.012>

Garcia, G. S. G., Meira, K. C., De Souza, A. H., & Guimarães, N. S. (2023). Anxiety and depression disorders in oncological patients under palliative care at a hospital service: a cross-sectional study. *BMC Palliative Care*, 22(1). <https://doi.org/10.1186/s12904-023-01233-1>

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Groenvold, M., Petersen, M., Aaronson, N. K., Arraras, J. I., Blazeby, J. M., Bottomley, A., Fayers, P., De Graeff, A., Hammerlid, E., Kaasa, S., Sprangers, M. A. G., & Bjorner, J. B. (2006). The development of the EORTC QLQ-C15-PAL: A shortened questionnaire for cancer patients in palliative care. *European Journal of Cancer*, 42(1), 55-64.
<https://doi.org/10.1016/j.ejca.2005.06.022>

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis (Seventh Edition)*. Upper Saddle River, NJ: Pearson Education.
http://digilib.uhl.ac.id/index.php?p=show_detail&id=15963

Hastie, T., Tibshirani, R., & Friedman, J. (2013). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

Higgins, J. P. T., & Green, S. (2008). *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley.

Kline, R. B. (2016). *Principles and practice of structural equation modeling*, 4th ed. Guilford Press. <https://psycnet.apa.org/record/2015-56948-000>

McElduff, P., Boyes, A., Zucca, A., & Girgis, A. (2004). *The Supportive Care Needs Survey: A guide to administration, scoring and analysis*. Newcastle: Centre for Health Research & Psycho-Oncology.

Muthén, B. (2002). Beyond SEM: General Latent variable modeling. *Behaviormetrika*, 29(1), 81-117. <https://doi.org/10.2333/bhmk.29.81>

Opioides contra el dolor causado por el cáncer. (s. f.). American Cancer Society.
<https://www.cancer.org/es/cancer/como-sobrellevar-el-cancer/efectos-secundarios/dolor/medicamentos-opioides-para-aliviar-el-dolor-causado-por-el-cancer.html>

Tieman, J. (2016). Ensuring quality in online palliative care resources. *Cancers*, 8(12), 113.
<https://doi.org/10.3390/cancers8120113>

What are palliative care and hospice care? (s. f.). National Institute on Aging.

<https://www.nia.nih.gov/health/what-are-palliative-care-and-hospice-care>

Anexos

Cuestionario

Para ayudarnos a planificar mejores servicios para las personas diagnosticadas con cáncer, nos interesa saber si se han satisfecho o no las necesidades que usted puede haber enfrentado como resultado de tener cáncer. Para cada elemento de las siguientes páginas, indique si ha necesitado ayuda con este problema en el último mes como resultado de tener cáncer. Encierre en un círculo el número que mejor describa si ha necesitado ayuda con esto en el último mes. Hay 5 respuestas posibles para elegir.

* Indica que la pregunta es obligatoria

Datos Demográficos

1. Cédula: *

2. Número de celular: *

3. Fecha de nacimiento:

Ejemplo: 7 de enero del 2019

4. ¿Se lo puede contactar vía telefónica para completar un formulario?

Marca solo un óvalo.

Sí

No

Otro: _____

5. Ciudad en la que reside:

6. Fecha de diagnóstico:

Indique la fecha en que fue diagnosticado por primera vez de cáncer.

Ejemplo: 7 de enero del 2019

12. ¿Tiene alguna dificultad para salir a caminar por corto tiempo fuera de la casa?

Marca solo un óvalo.

1

2

3

4

13. ¿Necesita quedarse en cama o en una silla durante el día?

Marca solo un óvalo.

1

2

3

4

7. Fecha de derivación:

Indique la fecha en que fue derivado a "Cuidados paliativos" por primera vez.

Ejemplo: 7 de enero del 2019

8. ¿Con quién o quiénes vive usted?

Marca solo un óvalo.

Solo

Familia

Amigos

Otro

9. Número de medicinas:

Indique el número aproximado de medicinas que ingiere diariamente.

10. Tipo de opiode *

¿Qué tipo de opiode usted consume como parte de su medicina?

11. Cáncer primario *

Cuestionario EORTC QLQ-C15-PAL

Estamos interesados en algunas cosas sobre usted y su salud. Por favor, conteste todas las preguntas usted mismo/a marcando con un círculo el número que mejor se aplique a su caso. No hay respuestas "correctas" ni "incorrectas". La información que nos proporcione se mantendrá estrictamente confidencial.

Preguntas sobre calidad de vida

Descripción de las respuestas:

1. Para nada
2. Un poco
3. Bastante
4. Extremadamente

14. ¿Necesita ayuda para comer, vestirse, bañarse o ir al baño?

Marca solo un óvalo.

1

2

3

4

Durante la última semana

Descripción de las respuestas:

1. Para nada
2. Un poco
3. Bastante
4. Extremadamente

15. ¿Le faltó el aire? *

Marca solo un óvalo.

1

2

3

4

16. ¿Ha tenido dolor? *

Marca solo un óvalo.

1

2

3

4

17. ¿Ha tenido problemas para dormir? *

Marca solo un óvalo.

1

2

3

4

18. ¿Se ha sentido débil? *

Marca solo un óvalo.

1

2

3

4

19. ¿Le ha faltado el apetito? *

Marca solo un óvalo.

1

2

3

4

20. ¿Ha sentido náuseas? *

Marca solo un óvalo.

1

2

3

4

22. ¿Estuvo cansado/a? *

Marca solo un óvalo.

1

2

3

4

21. ¿Ha estado estreñado/a? *

Marca solo un óvalo.

1

2

3

4

23. ¿Interfirió el dolor con sus actividades diarias? *

Marca solo un óvalo.

1

2

3

4

24. ¿Se sintió tenso/a? *

Marca solo un óvalo.

1

2

3

4

—

25. ¿Se sintió deprimido/a? *

Marca solo un óvalo.

1

2

3

4

—

Para la siguiente pregunta, por favor, marque con un círculo el número del 1 al 7 que mejor se aplique a su caso.

26. ¿Cómo calificaría su calidad de vida en general durante la última semana? *

Marca solo un óvalo.

Muy mala

1

2

3

4

5

6

7

Excelente

Preguntas sobre necesidad de atención de apoyo (SUPPORTIVE CARE NEEDS SURVEY SHORT FORM 34)

Descripción de respuestas:

1. **No aplicable** - esto no fue un problema para mí como resultado de tener cáncer.
2. **Satisfecho** - necesitaba ayuda con esto, pero mi necesidad de ayuda quedó satisfecha en ese momento.
3. **Necesidad baja** - este artículo me causó preocupación o incomodidad. Tenía poca necesidad de ayuda adicional.
4. **Necesidad moderada** - este artículo me causó preocupación o incomodidad. Necesitaba ayuda adicional.
5. **Alta necesidad** - este artículo me causó preocupación o incomodidad. Tenía una fuerte necesidad de ayuda adicional.

En el último mes, indique cuál fue su nivel de necesidad de ayuda con:

27. Dolor *

Marca solo un óvalo.

1

2

3

4

5

—

29. Sentirse mal la mayor parte del tiempo *

Marca solo un óvalo.

1

2

3

4

5

—

28. Falta de energía/cansancio *

Marca solo un óvalo.

1

2

3

4

5

—

30. Trabajar en casa *

Marca solo un óvalo.

1

2

3

4

5

—

31. No poder hacer las cosas que solía hacer *

Marca solo un óvalo.

1

2

3

4

5

—

33. Sentirse deprimido o deprimida *

Marca solo un óvalo.

1

2

3

4

5

—

32. Ansiedad *

Marca solo un óvalo.

1

2

3

4

5

—

34. Sentimientos de tristeza *

Marca solo un óvalo.

1

2

3

4

5

—

35. Temores sobre la propagación del cáncer *

Marca solo un óvalo.

1

2

3

4

5

—

37. Incertidumbre sobre el futuro. *

Marca solo un óvalo.

1

2

3

4

5

—

36. Preocupación de que los resultados del tratamiento estén fuera de su control. *

Marca solo un óvalo.

1

2

3

4

5

—

38. Aprender a sentir que tiene el control de su situación. *

Marca solo un óvalo.

1

2

3

4

5

—

39. Mantener una perspectiva positiva.*

Marca solo un óvalo.

1

2

3

4

5

—

41. Cambios en los sentimientos sexuales.*

Marca solo un óvalo.

1

2

3

4

5

—

40. Sentimientos sobre la muerte y el morir.*

Marca solo un óvalo.

1

2

3

4

5

—

42. Cambios en las relaciones sexuales.*

Marca solo un óvalo.

1

2

3

4

5

—

43. Se inquieta por las preocupaciones de sus allegados.*

Marca solo un óvalo.

1

2

3

4

5

—

45. Más opciones sobre el hospital al que acudir.*

Marca solo un óvalo.

1

2

3

4

5

—

44. Más posibilidades de elegir especialistas oncológicos.*

Marca solo un óvalo.

1

2

3

4

5

—

46. Que el personal médico le asegure que lo que siente es normal.*

Marca solo un óvalo.

1

2

3

4

5

—

47. El personal del hospital atiende rápidamente sus necesidades físicas. *

Marca solo un óvalo.

1

2

3

4

5

—

49. Recibir información escrita sobre los aspectos importantes de su asistencia. *

Marca solo un óvalo.

1

2

3

4

5

—

48. Que el personal del hospital reconozca y sea sensible a sus sentimientos y necesidades emocionales. *

Marca solo un óvalo.

1

2

3

4

5

—

50. Recibir información (escrita, diagramas, dibujos) sobre aspectos de la gestión de su enfermedad y efectos secundarios en casa. *

Marca solo un óvalo.

1

2

3

4

5

—

51. Recibir explicaciones de aquellas pruebas para las que desee explicaciones. *

Marca solo un óvalo.

1

2

3

4

5

—

53. Ser informado de los resultados de sus pruebas lo antes posible. *

Marca solo un óvalo.

1

2

3

4

5

—

52. Informarse adecuadamente sobre los beneficios y efectos secundarios de los tratamientos antes de decidir someterse a ellos. *

Marca solo un óvalo.

1

2

3

4

5

—

54. Estar informado sobre el cáncer que está bajo control o disminuyendo (es decir * remisión).

Marca solo un óvalo.

1

2

3

4

5

—

55. **Estar informado sobre lo que puede hacer para *
ayudarse a sí mismo a recuperarse.**

Marca solo un óvalo.

1

2

3

4

5

57. **Recibir información sobre las relaciones sexuales. ***

Marca solo un óvalo.

1

2

3

4

5

56. **Tener acceso a asesoramiento profesional *
(por ejemplo, psicólogo, trabajador social, orientador, enfermero especialista) si
usted, su familia o sus amigos lo necesitan.**

Marca solo un óvalo.

1

2

3

4

5

58. **Ser tratado como una persona y no como un caso más. ***

Marca solo un óvalo.

1

2

3

4

5

59. **Recibir tratamiento en un hospital o clínica que sea lo más agradable físicamente *
posible.**

Marca solo un óvalo.

1

2

3

4

5