



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Modelamiento del espacio de signos vitales para detectar el deterioro de los pacientes en una unidad de cuidados intensivos

Ledys Maria Izquierdo Borrero

Universidad Nacional de Colombia

Facultad de Medicina

Bogotá, Colombia

2021

Modelamiento del espacio de signos vitales para detectar el deterioro de los pacientes en una unidad de cuidados intensivos

Ledys Maria Izquierdo Borrero

Tesis de investigación presentada como requisito parcial para optar al título de:
Magister en Ingeniería Biomédica

Director (a):
Luis Fernando Niño PhD.

Línea de Investigación: Sistemas Inteligentes
Grupo de Investigación:
Laboratorio de Investigación en Sistemas Inteligentes - LISI

Universidad Nacional de Colombia
Facultad de Medicina
Bogotá, Colombia
2021

Dedicatoria

A Fabio y José David

Por su constante apoyo, paciencia y comprensión en este gran proyecto para nuestras vidas.

No creo en el destino. Creo en las señales.

Elisabet Benave

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Lizquierdo

Ledys Maria Izquierdo

Fecha 16/02/2021

Agradecimientos

A mi profe Luis Fernando Niño, por su mentoría, orientación y motivación. Por mostrarme un camino en el mundo académico. Por su rigor y disciplina. Por estar allí siempre. Para él mi admiración, lealtad y respeto.

A mis compañeros del grupo LISI y semillero, a los docentes y amigos de la maestría de Ingeniería Biomédica, por compartir sus conocimientos a través de los espacios de discusión y su pasión por la investigación.

Al Hospital Militar Central, mi sitio de trabajo, mi casa, que me permitió un horario flexible para poder realizar esta maestría.

Resumen

En el campo de la monitorización continua de los signos vitales en entornos de cuidados intensivos se ha observado que los signos de alerta temprana "de un deterioro fisiológico inminente" pueden no ser detectados a tiempo, hecho que se agrava no solo por la limitación de los recursos médicos, sino también por el "diluvio de datos" causado por la adquisición de información en pacientes cada vez más complejos durante la atención de rutina. El objetivo de este estudio es desarrollar un modelo probabilístico para predecir los episodios clínicos futuros de un paciente utilizando valores de signos vitales observados antes de un evento clínico. Los signos vitales (por ejemplo, frecuencia cardíaca, presión arterial) se utilizan para controlar las funciones fisiológicas de un paciente y sus cambios simultáneos indican las transiciones entre los estados de salud del paciente. Si tales cambios son anormales, puede conducir a un deterioro fisiológico grave. Se utilizó la metodología CRISP-DM (CRoss-Industry Standard Process for Data Mining) como proceso de minería de datos y luego utilizamos cadenas de Márkov para identificar los estados clínicos por los que pasa el paciente. Después, se aplicó un enfoque basado en un modelo oculto de Márkov (*Hidden Márkov Model*, HMM) para la clasificación y predicción del deterioro de un paciente calculando la probabilidad de estados clínicos futuros. Ambos modelos de aprendizaje fueron entrenados y evaluados utilizando seis bioseñales de 90 pacientes para un total de 94.678 instancias, recolectadas de una base de datos de pacientes reales que se encontraban en la Unidad de Cuidados Intensivos Pediátricos del Hospital Militar Central de la ciudad de Bogotá, Colombia. La técnica propuesta basada en el seguimiento de múltiples variables fisiológicas mostró resultados prometedores en la identificación precoz del deterioro de los pacientes críticos.

Palabras clave: Signos vitales, modelo oculto de Márkov, evento clínico, cuidado intensivo pediátrico.

Abstract

In the field of continuous vital-sign monitoring in critical care settings, it has been observed that the “early warning signs” of impending physiological deterioration can fail to be detected timely and sometimes by resource constrained clinical staff. This effect may be escalated by the “data deluge” caused by acquisition of more complex patient data during routine care. The objective of this study is to develop a probabilistic model for predicting the future clinical episodes of a patient using observed vital sign values prior to a clinical event. Vital signs (e.g., heart rate, blood pressure) are used to monitor a patient's physiological functions and their simultaneous changes indicate transitions between patient's health states. If such changes are abnormal then it may lead to serious physiological deterioration. The CRISP-DM (CRoss-Industry Standard Process for Data Mining) methodology was used as a data mining process and then we used Márkov chains to identify the clinical states through which the patient passes. Then, a Hidden Márkov model (HMM) based approach was applied for classification and prediction of patient's deterioration by computing the probability of future clinical states. Both learning models were trained and evaluated using six vital signs data from 94,678 records from 90 patients, collected from the database of real patients who were in the Pediatric Intensive Care Unit of the Central Military Hospital in the city of Bogota, Colombia. The proposed technique based on monitoring multiple physiological variables showed promising results in early identifying the deterioration of critically ill patients.

Keywords: Vital signs, Hidden Márkov model, Clinical event, pediatric critical care

8 Modelamiento del espacio de signos vitales para detectar el deterioro de los pacientes en una unidad de cuidados intensivos

Contenido

Resumen	IX
Lista de figuras	XII
Lista de tablas	XIII
Lista de Símbolos y abreviaturas	XIV
Introducción	1
1. Capítulo 1 Estado del Arte	4
1.1 Clasificación de Riesgo en los Monitores	6
1.1.1. Efectos Secundarios y Riesgos	6
1.1.2. Cómo funcionan los monitores de signos vitales	7
1.1.3. Correlación de Signos Vitales	8
1.2. Trabajos Relacionados	13
1.3. Motivación	20
2. Capítulo 2 Marco conceptual de los Métodos a Utilizar	22
2.1. Análisis de Componentes Principales (ACP)	22
2.2. Análisis de Series de tiempo	29
2.3. Análisis de Agrupación K-medias (<i>K-means</i>)	37
2.3.2. Análisis de Agrupación Bietápico	39
2.4. Métodos de Clasificación	41
2.5. Cadenas de Márkov	46
2.6. Modelos Ocultos de Márkov (HMM)	49
3. Capítulo 3 Metodología	57
3.1. Comprensión del estudio	58
3.2. Comprensión de los datos	58
3.2.1. Caracterización de una muestra de pacientes en estado crítico en la unidad de cuidados intensivos pediátrico	58
3.3. Preparación de los datos	60
3.4. Modelado	63
3.4.1. Transformación del espacio de signos vitales para identificar correlaciones significativas entre los signos vitales y los eventos clínicos	63
3.4.1.1. Análisis de Componentes Principales (ACP)	64
3.4.1.2. Análisis de series de tiempo	70
3.4.1.3. Análisis de Agrupación de datos	76
3.4.1.4. Modelos de clasificación	86
3.4.1.5. Análisis basado en cadenas de Márkov	96
3.4.1.6. Análisis basado en Modelos Ocultos de Márkov (<i>HMM</i>)	98
3.5. Evaluación e implementación del modelo	102
3.5.1. Validación del método propuesto con el diagnóstico proporcionado por los	102

expertos	
4. Capítulo 4. Discusión y análisis de resultados	105
4.1. Futuros trabajos	108
5. Capítulo 5. Conclusiones y recomendaciones	109
5.1. Conclusiones	109
5.2. recomendaciones	111
Bibliografía	112
A. Anexo: MInCir para estudios de pronóstico	121
B. Anexo: Artículo Publicado	122

Lista de Figuras

Figura 1-1: Monitor de signos vitales.	5
Figura 1-2: Adquisición de Señales	8
Figura 1-3: Módulos de signos vitales	8
Figura 1-4: Tarjeta Madre de los Monitores	8
Figura 1-5: Grafica Señales en el Monitor	8
Figura 1-6: Diagrama Estratégico de Búsqueda	16
Figura 2-1 Representación temporal y representación frecuencial de señales	30
Figura 2-2. Relación entre el dominio temporal y frecuencial	30
Figura 2-3 Espectros de señales pasabajos, pasabanda de banda ancha y pasabanda de banda angosta	32
Figura 2-4 Correlograma o función de autocorrelación	35
Figura 2-5 Algoritmo Avance Panel A; Algoritmo Retroceso Panel B	51
Figura 2-6 Algoritmo de Viterbi	52
Figura 2-7 Algoritmo Baum-Welch	53
Figura 2-8 Esquema del funcionamiento de un HMM.	53
Figura 3-1 Esquema de la metodología CRISP MD	57
Figura 3-2 Diagrama de cajas que representan la distribución de las señales de signos vitales para un paciente	62
Figura 3-3. Gráfico de sedimentación número de componentes vs. valores propios	63
Figura 3-4. Gráfico de saturaciones, análisis de componentes principales para un paciente	67
Figura 3-5. Componentes en el espacio rotado en la ventana media hora antes del evento clínico para un paciente.	69
Figura 3-6. Comportamiento de los bioseñales para un paciente	70
Figura 3-7: Análisis Espectral. Transformada de Fourier para un paciente	71
Figura 3-8: Predicción Modelos ARIMA para un paciente	74
Figura 3-9 pronóstico serie de tiempo una hora antes del evento clínico en un paciente	75
Figura 3-10. Pronóstico serie de tiempo media hora antes del evento clínico en un paciente.	75
Figura 3-11. Cálculo del tamaño de cada grupo para la serie de datos de un paciente.	77
Figura 3-12. Listado de los centroides en la serie de datos de un paciente.	78
Figura 3-13. Número óptimo de conglomerados para un paciente, gráfica del codo.	79
Figura 3-14. Coeficiente de silueta para diferentes agrupaciones de los datos de un paciente.	79
Figura 3-15 Resumen del modelo bietápico para la toda la serie de datos con las	83

variables en forma categóricas	
Figura 3-16 Resumen del modelo bietápico para la toda la serie de datos con las variables en forma continua	83
Figura 3-17. Resumen del modelo Bietápico para un paciente	84
Figura 3-18. Análisis de las variables que más contribuyen a la predicción a partir del conglomerado formado.	85
Figura 3-19. Esquema Árbol de decisión para las variables categóricas para un	87
Figura 3-20 Red neuronal generada con tan sólo 4 variables que caracterizan la Clase	93
Figuras 3-21 Cadenas de Márkov que representan las características más frecuentes de los datos en tres pacientes diferentes	97
Figuras 3-22 Cadenas de Márkov con estados absorbentes que representan a los pacientes que fallecieron, aplicado para dos pacientes	98

Lista de tablas

	Pág
Tabla 1-1 Clasificación de los Riesgos en los Monitores	6
Tabla 1-2 Características de los estudios elegidos	17
Tabla 3-1 Descripción de los 6 bioseñales	60
Tabla 3-2 Características demográficas de la población	61
Tabla 3-3 Test Bartlett's y KMO para un paciente	65
Tabla 3-4 Resultado del Análisis de Componentes Principales, varianza total acumulada para los datos de un paciente	65
Tabla 3-5 Varianza total explicada media hora antes del evento clínico para un Paciente	67
Tabla 3-6 Matriz de Componentes rotados media hora antes del evento clínico	68
Tabla 3-7 Modelo ARIMA para un paciente	72
Tabla 3-8 Modelo estadístico ARIMA para un paciente.	73
Tabla 3-9 Método de agrupación (<i>clustering</i>) <i>k-means</i> para un paciente	80
Tabla 3-10 Interpretación Clínica del clúster aplicado a un paciente	80
Tabla 3-11 Agrupación en diferentes estados clínicos de acuerdo a la complejidad arrojada por la interpretación clínica de los clústeres encontrados en un paciente	81
Tabla 3-12. Interpretación de los bioseñales vs. estados clínicos reportados por algoritmo <i>k-means</i> para un paciente.	82
Tabla 3-13. Clasificador basado en árbol de decisión con variables categóricas por clases. Cross-validation 10 particiones para un paciente	87
Tabla 3-14. Matriz de confusión del Árbol de decisiones, usando variables categóricas, para un paciente	88
3-15. Comparaciones de los resultados de los clasificadores Árbol de decisiones vs Bayes ingenuo en clasificar la clase estado crítico en un paciente	89
Tabla 3-16. Matriz costo-beneficio	90
Tabla 3-17. Matriz de confusión del perceptrón multicapa aplicada a toda la base de datos	93
Tabla 3-18. Validación estadística del modelo perceptrón multicapa.	94
Tabla 3-19 Comparación de diferentes técnicas de clasificación para predecir estado crítico, a través de muestreo estratificado proporcional	96
	96
Tabla 3-20. Resultados cálculo de Log verosimilitud HMM para un paciente.	99
Tabla 3-21 Métricas 10-fold cross validation para cada evento clínico usando HMM	100

Tabla 3-22 Comparación de la clasificación de estados críticos entre el método HMM vs otros clasificadores	103
Tabla 3-23 comparación de este modelo (HMM) con los resultados y características obtenidas de trabajos similares	104

Lista de abreviaturas

Símbolos

Símbolo	Término	Unidad SI	Definición
mmHg	Milímetro mercurio		10
<i>Kg</i>	Kilogramo		59
Cm	centímetro		59
E	Número Euler, irracional		29
%	Porcentaje		25
€	Pertenece		37
∀	Cuantificador universal “para todo”		

Símbolos con letras griegas

Símbolo	Término	Unidad SI	Definición
Π	número pi (π),	1	29
Λ	Lambda	1	25
H	mittlere Bettneigungs winkel (Stürzen)	1	45

Superíndices

Superíndice	Término
N	Exponente, potencia
R ²	R cuadrado o coeficiente de determinación

Abreviaturas

Abreviatura Término

Accuracy:	Exactitud
ACM:	Asociaciones para maquinaria de computación
ACP:	Análisis de Componente Principales
AIC:	Criterio de Información de Akaike
ANOVA:	análisis de Varianza
ARIMA:	Modelo Autorregresivo Integrado de media móvil
ARMA:	Modelo Autorregresivo de media móvil
BIC:	Criterio Bayesiano de Schwarz
Bs:	matriz los beneficios
Bvn:	Beneficios verdaderos negativos
Bvp:	Beneficios verdaderos positivos
CF:	características del Clúster
Cfn:	Costos falsos negativos
Cfp:	Costos falsos positivos
Chi-Square:	prueba de distribución χ^2 o de Pearson
CINAHL:	Índice Acumulativo de Enfermería y Literatura de Ciencias de la Salud o, en inglés, Cumulative Index to Nursing and Allied Health Literature
CRISP-MD:	Cross Industry Standard Process for Data Mining
cross validation:	Validación cruzada
Cs:	matriz de costos
DF:	degree freedom en español: grados de libertad
ECG:	electrocardiográfico
EKG	electrocardiograma
Embase:	Excerpta Medica dataBASE
et al:	expresión latina <i>et alii</i> , que significa 'y otros'
F1:	F-score o medida-F
FC:	frecuencia cardíaca
FDA:	Agencia del gobierno de los Estados Unidos que regula la Administración de Drogas y Alimentos
FN:	Falsos Negativos
FORTAN:	contracción del inglés <i>The IBM Mathematical Formula Translating System</i>
FP:	Falsos Positivos
FR:	frecuencia respiratoria
HMM:	sigla en Inglés Modelos Ocultos de Márkov
IA:	Inteligencia Artificial
IEEE:	Instituto de Ingenieros Eléctricos y Electrónicos
K-means:	K- medias. k: número de grupos
K-S:	Kolmogórov-Smirnov
LI:	Ledys Izquierdo

LN:	Luis Fernando Niño
MEDLINE:	Biblioteca Nacional de Medicina de los Estados Unidos
MO:	Kaiser-Meyer-Olkin Measure of Sampling Adequacy.
PA:	presión arterial
Precisión:	Precisión o Valor predictivo positivo
PubMed:	búsqueda de libre acceso de MEDLINE
Recall	Sensibilidad
SCOPUS:	páginas web científicas mediante Scirus, también de Elsevier, y bases de datos de patentes.
SE:	servicios de emergencia
Sig:	significancia
SpO2:	saturación de oxígeno periférica
SaO2	Saturación arterial de Oxígeno
TAD:	tensión arterial diastólica
TAM:	Tensión arterial media
TAS:	Tensión arterial sistólica
TFF:	Transformada de Fourier
ToC:	temperatura periférica o central grados centígrados
UCI:	unidades de cuidado intensivo
UCIN:	Unidad de Cuidado Intensivo neonatal
UCIP:	Unidad de Cuidado Intensivo Pediátrico
VN:	Verdaderos Negativos
VP:	Verdaderos positivos

Introducción

El personal que cuida a los pacientes no siempre tienen todas las habilidades y experiencia requeridas para cuidar a los pacientes cuando están gravemente enfermos [1]. Esto sugiere que incluso en ausencia de fallas de los sistemas de monitorización, los pacientes aún no pueden recibir la atención que requieren porque el personal no posee como prerrequisitos los conocimientos o habilidades. Tradicionalmente, para identificar cambios clínicos en los pacientes en estado crítico, los cuidadores de los pacientes críticos no solo deben saber cómo medir estos signos vitales con precisión, también deben saber cómo interpretarlos y actuar con base en la información sobre ellos. El paciente no puede esperar hasta que el personal tenga experiencia, necesitan ser evaluados por personal experimentado de inmediato.

El futuro de la monitorización pasa por diversas etapas, que no son excluyentes entre sí. En todo tipo de monitor se debe buscar el incremento en la precisión en la detección del parámetro a monitorizar, es decir, hacer aparatos cada vez más fiables, que midan con mejor precisión los parámetros que se quieren vigilar. Por otra parte, está la búsqueda de nuevos parámetros que puedan reflejar de mejor forma la situación clínica del paciente. Pero no hay que olvidar que el objetivo de monitorizar a los pacientes es su bienestar, y hay que cuidar las indicaciones, así como la interpretación de los resultados. Tampoco se debe olvidar que un aparato jamás sustituye a una persona [1, 2].

Desde alrededor del año 2000 ha existido un interés creciente por recuperar datos y extraer conocimiento de grandes volúmenes de datos, mediante la aplicación eficiente de técnicas de análisis de datos y diferentes modelos para intentar evaluar el comportamiento de los signos vitales tanto en pacientes hospitalizados como en los pacientes fuera del hospital, es decir, en casa [3,4]. Los mayores esfuerzos hasta ahora han sido buscar patrones normales en los signos vitales de pacientes sanos, pacientes con patologías crónicas y muy pocos en pacientes críticos [5,6]. Los estudios de ámbito hospitalario se han

desarrollado en las áreas de emergencia, hospitalización y cuidados intensivos. Entre estos, la gran mayoría se han realizado en pacientes adultos y muy pocos en pacientes pediátricos. Estudios anteriores también han intentado relacionar las variables de los signos vitales para buscar predictores del deterioro del paciente. Pero hasta ahora, normalmente solo han establecido la relación entre dos variables, Además, existen muy pocos estudios incluso en pacientes con enfermedades agudas, como los pacientes en unidad de cuidado intensivo pediátrico (UCIP), y solo evalúan una única patología con un solo signo vital [7,8]. Teniendo en cuenta todo lo anterior, es necesario transformar el espacio de signos vitales durante el seguimiento del paciente a través de un modelo que ayude a simplificar el trabajo de los profesionales en UCIP, presentándoles alertas tempranas de forma más comprensible, reduciendo las falsas alarmas y continuar la búsqueda de nuevos parámetros que podrían reflejar de mejor forma la situación clínica del paciente.

La principal contribución del presente estudio es utilizar un modelo predictivo para la detección de anomalías clínicas utilizando tecnología informática generalizada que utiliza varias técnicas de aprendizaje automático. El método propuesto utiliza la agrupación de datos de signos vitales del paciente para reconocer información oculta en estados clínicos utilizando observaciones pasadas, en un intento de identificar el deterioro temprano de pacientes críticamente enfermos.

Este documento contiene cuatro capítulos y se estructura de la siguiente manera:

Capítulo 1: estado de arte y panorama general del proceso de monitorización en el área de UCIP, y descripción de los trabajos relacionados para mejorar el proceso de monitoria en UCIP.

Capítulo 2: se realiza un marco conceptual de los métodos a utilizar

Capítulo 3: se realiza una descripción de la metodología experimental empleada en esta investigación.

Capítulo 4: análisis y discusión de los resultados obtenidos con todas las técnicas de minería de datos y de aprendizaje automático utilizadas en esta investigación, y

Capítulo 5: las principales conclusiones y trabajos futuros propuestos.

Todos los capítulos cuentan con una breve introducción para que el lector tenga una visión general acerca de la temática tratada y al final de cada capítulo observarán las respectivas conclusiones del capítulo.

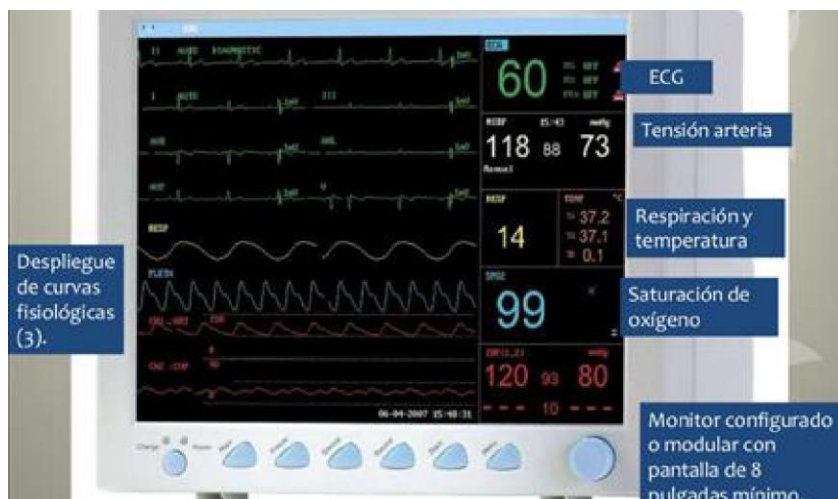
● **Capítulo 1. Estado del Arte**

El término monitor incluye las funciones de observar, informar y ayudar. La monitorización de pacientes es la práctica continua de observaciones para descubrir un hecho específico, para seguir una evolución clínica o para evaluar la respuesta a un tratamiento específico. Los monitores de signos vitales son aparatos complejos que se encargan de obtener, amplificar, procesar, registrar y mostrar señales biológicas. Dichas señales tienen importancia clínica para evaluar a cada instante la evolución del paciente monitorizado [3].

Según el tipo de estado clínico del paciente la monitorización puede ser continua e intermitente, intrahospitalaria o ambulatoria. El control clínico del niño en estado crítico se basa primordialmente en la observación, evaluación directa y continua, y exige una monitorización de los principales parámetros vitales, para detectar precozmente alteraciones hemodinámicas y actuar en consecuencia. La necesidad de cuidados intensivos y monitorización de pacientes ha sido reconocida desde hace muchos años [1], la presencia de los enfermeros o enfermeras durante las 24 horas del día en el caso de pacientes en estado crítico se ha convertido con el tiempo en algo familiar dentro de los hospitales. Pero solo en los últimos años se han diseñado y fabricado equipos suficientemente seguros y precisos para ser empleados extensamente en la monitorización de pacientes.

Las constantes vitales son aquellos parámetros que indican el estado hemodinámico del paciente. Se consideran como principales parámetros fisiológicos: la frecuencia cardíaca (FC), el trazado electrocardiográfico (ECG), la frecuencia respiratoria (FR), la presión arterial medida por medio de manguito neumático o de manera invasiva a través de un catéter arterial (PA), la temperatura periférica o central (ToC), y la saturación de oxígeno (SpO₂) mediante pulsioximetría (ver Figura 1-1).

Figura 1-1. Monitor de signos vitales



Fuente: Tomado y modificado de [9]

La monitorización de signos vitales puede realizarse de manera invasiva y no invasiva y esta diferencia va a depender de si hay o no invasión a los tejidos.

Principios de operaciones de los monitores de signos vitales [9]

Los monitores de signos vitales pueden ser:

- Pre-configurados: los parámetros a configurar son fijados por el proveedor desde la fábrica y no es posible agregarle ningún parámetro adicional.
- Modulares: el usuario puede seleccionar los parámetros a monitorizar adicionando dispositivos conocidos como módulos.
- Mixtos: combinación de los dos anteriores.

Tipos de monitores de signos vitales

- Fijos: Se encuentran colocados en la cabecera del paciente y se sujetan a la pared a través de un soporte diseñado especialmente para colocar el monitor. Estos monitores fijos se encuentran en áreas como máquinas de anestesia o unidades de cuidados intensivos.

- **Móviles:** son los monitores de transporte, utilizados para monitorizar a los pacientes durante su traslado, ya sea intra hospitalario o extrahospitalario.

Tipos de monitores según sus funciones [9]

- Monitor de signos vitales básicos: monitoría no invasiva
- Monitor de signos vitales intermedio: monitoría no invasiva, incluye más variables de signos vitales y puede tener alguna monitoria invasiva
- Monitor de signos vitales avanzado: Incluye monitoria invasiva
- Monitor de signos vitales transporte: Incluye monitoría no invasiva
- Monitor de signos vitales anestesia básico
- Monitor de signos vitales anestesia avanzado

1.1 Clasificación de riesgos en los monitores

Según la Comisión Federal para la Protección de Riesgos Sanitarios, Secretaría de Salud (COFEPRIS) [10] y la *Global Harmonization Task Force* (GHTF) [11], los riesgos en los monitores pueden ser clasificados como se muestra en la Tabla 1.

Tabla 1-1 Clasificación de los Riesgos en los Monitores

ENTIDAD	RIESGO	RAZÓN
COFEPRIS ¹	Clase II	Para aquellos insumos conocidos en la práctica médica y que pueden tener variaciones en el material con el que están elaborados o en su concentración y, generalmente, se introducen en el organismo menos de 30 días.
GHTF ²	C: Riesgo alto moderado	Debido a que estos dispositivos monitorizan y/o procesos fisiológicos de alta importancia para la salud del paciente

Fuente: tomado y modificado de referencia [10,11] 1. Comisión Federal para la Protección de Riesgos Sanitarios, Secretaría de Salud, (www.cofepris.salud.gob.mx) y 2. Global Harmonization Task Force www.ghtf.org

1.1.1 Efectos secundarios y riesgos [9]

Algunos de los problemas asociados a los monitores tienen que ver con el uso inapropiado de los mismos:

- Los electrodos, especialmente asociados a calentamientos deberán cambiarse periódicamente para evitar quemaduras o irritaciones de la piel.
- Los cables de cada uno de los parámetros y, especialmente, los de las derivaciones del ECG deberán ser revisados para detectar rupturas o falsos contactos.
- Se debe tener precaución con el uso de pulsioxímetros, sobre todo cuando se usan en sistemas de electrocirugía, debido a que estos generan señales de alta frecuencia que podrían afectar la operación del sensor de SpO₂.
- Los cables de los diferentes parámetros, así como los cables de ECG deberán estar protegidos como lo indica la FDA (*U.S. Food and Drug Administration*), agencia del gobierno de los Estados Unidos que regula la administración de drogas y alimentos, para evitar descargas en los pacientes.
- Los pacientes en un hospital se encuentran más vulnerables a sufrir un choque eléctrico debido a dos circunstancias: 1. Si el paciente se encuentra conectado con electrodos o catéteres invasivamente en su cuerpo y no cuenta con una integridad en su piel, el paciente se encuentra más expuesto al paso de la corriente de fuga, que para el caso de un monitor de paciente está entre 20 y 350 microamperios. 2. La propia condición de la paciente debida a su propio padecimiento: aplicación de medicamentos, pacientes con baja circulación, entre otros. Por esto, se debe incluir un programa de seguridad eléctrica que incluya la revisión de polaridades, contactos al menos dos veces al mes.

1.1.2 Cómo funcionan los monitores de signos vitales

Un monitor es un instrumento que se encarga de la adquisición, amplificación, procesamiento, visualización y registro de señales biológicas. El monitor puede ser sencillo de un canal, como en los monitores portátiles, o tan complejos como de 16 canales, poseen una computadora y un microprocesador para el procesamiento de las señales. Cuando un monitor cuenta con un microprocesador permite el análisis e interpretación de las señales. Uno de los lenguajes que frecuentemente se utiliza en estos monitores es FORTRAN, (contracción del inglés *The IBM Mathematical Formula Translating System*) [12]. Las señales fisiológicas son recolectadas por los módulos amplificadores (ver Figuras 1-2, 1-3), que luego son enviadas a la tarjeta madre (ver Figura 1-4), la impedancia de la señal que es detectada y amplificada para obtener la señal del signo vital que detecta, luego la

gráfica como una onda o la presenta como un número en la pantalla del monitor (ver Figura 1-5) [9,12].

Figura 1-2: Adquisición de señales



Figura 1-3: Módulos de signos vitales



Fuente: Modificado de referencia bibliográfica [9,12].

Figura 1-4: Tarjeta madre



figura 1-5: Graficador de señales



Fuente: Modificado de [9,12]

1.1.3 Correlación de signos vitales

Los sistemas de monitoreo de signos vitales desempeñan un papel crucial en el ambiente médico, sobre todo en el ámbito de cuidado intensivo. Por medio de estos monitores el paciente puede beneficiarse especialmente de la observación continua. La toma de

decisiones clínicas se basa en la medición o estimación precisa de los signos vitales. Los signos vitales esenciales que se capturan de manera rutinaria son: la frecuencia cardíaca (FC), la frecuencia respiratoria (FR), la presión arterial (PA), la saturación de oxígeno periférica (SpO₂) y la temperatura (ToC).

FRECUENCIA CARDIACA (FC): Es la onda pulsátil de la sangre, originada en la contracción del ventrículo izquierdo del corazón y que resulta en la expansión y contracción regular del calibre de las arterias. La onda pulsátil representa el rendimiento del latido cardíaco, que es la cantidad de sangre que entra en las arterias con cada contracción ventricular y la adaptación de las arterias, o sea su capacidad de contraerse y dilatarse [7]. Las alteraciones de la frecuencia cardíaca se clasifican en *bradicardia* si los valores normales se encuentran por debajo del límite inferior de normalidad para la edad y *taquicardia* si su valor se encuentra por encima del límite superior de normalidad para la edad [13].

Valores normales de Frecuencia Cardíaca: frecuencia cardíaca/minuto [13]

Recién nacido a 1 año de edad: 100 - 160

1 a 3 años de edad: 95 - 150

4 a 5 años de edad: 80 - 140

6 a 12 años de edad: 70 - 120

13 a 18 años de edad: 60- 100

FRECUENCIA RESPIRATORIA (FR): La respiración es el proceso mediante el cual se toma oxígeno del aire ambiente y se expulsa el anhídrido carbónico del organismo. El ciclo respiratorio comprende una fase de inspiración y otra de espiración. La medición de la frecuencia respiratoria consiste simplemente en contar el número de respiraciones durante un minuto. La frecuencia respiratoria puede aumentar con la fiebre, las enfermedades y otras condiciones médicas. Las alteraciones de la respiración se clasifican en:

Bradipnea: es la lentitud en el ritmo respiratorio con una frecuencia inferior a límite inferior del rango normal para la edad por minuto. Se encuentra en pacientes con alteración neurológica o electrolítica, infección respiratoria o pleuritis.

Taquipnea: se presenta cuando la frecuencia respiratoria persistente por encima del límite superior al rango de normalidad para la edad por minuto, es una respiración superficial y rápida. Se observa en pacientes con dolor por fractura costal o pleuritis.

Hiperpnea o hiperventilación: es una respiración profunda y rápida de frecuencia por encima del límite superior al rango de normalidad para la edad por minuto. Es producida por ansiedad, ejercicio, alteraciones metabólicas o del sistema nervioso central.

Apnea: es la ausencia de movimientos respiratorios. Ocurre en el paro cardiorrespiratorio.

Valores normales de Frecuencia Respiratoria: respiraciones/minuto [13]

Recién nacido a 1 año de edad: 30-60

1 a 3 años de edad: 24-40

4 a 5 años de edad: 22-34

6 a 12 años de edad: 18-30

13-18 años de edad: 12-16

PRESIÓN ARTERIAL: Es una medida de la presión que ejerce la sangre sobre las paredes arteriales en su impulso a través de las arterias. Debido a que la sangre se mueve en forma de ondas, existen dos tipos de medidas de presión: la **presión sistólica (TAS)**, que es la presión de la sangre debida a la contracción de los ventrículos, es decir, la presión máxima; y la **presión diastólica (TAD)**, que es la presión que queda cuando los ventrículos se relajan esta es la presión mínima. La **Presión Arterial Media (TAM)** se calcula con la siguiente fórmula:

$$TAM = PAS + 2 PAD/3 \quad (1)$$

Presiones sanguíneas normales por edad [13]

	PAS mmHg	PAD mmHg
Lactante	60 – 90	30 - 62
2 años	78 – 112	48 – 78
8 años	85 – 114	52 – 85

12 años	95 – 135	58 – 88
Adulto	100 - 120	60 – 80

Las alteraciones de la Presión Arterial se clasifican en:

- *Hipertensión arterial:* el diagnóstico de hipertensión debe hacerse después de encontrar cifras tensionales altas, por encima del límite superior para la edad, sistólica o diastólica.
- *Hipotensión arterial:* el diagnóstico de hipotensión debe hacerse después de encontrar cifras tensionales bajas, por debajo del límite inferior para la edad, sistólica o diastólica.

SATURACIÓN PERIFÉRICA DE OXÍGENO (SpO₂): se mide por medio de la oximetría de pulso o pulsioximetría, que es la medición no invasiva del oxígeno transportado por la hemoglobina en el interior de los vasos sanguíneos. Debido a que el flujo sanguíneo pulsátil es necesario para que un oxímetro de pulso funcione, puede proporcionar resultados inexactos en un niño con mala perfusión periférica (por ejemplo, *shock* o paro cardíaco), en niños con enfermedades crónicas hipoxemia (p. ej., cardiopatía congénita cianótica, pulmonar hipertensión), anemia significativa, carboxihemoglobinemia o metahemoglobinemia. Valor Normal: La saturación de oxígeno debe de ser mayor del 88% a la altura de la ciudad de Bogotá, por debajo de este valor se presenta hipoxemia [13].

TEMPERATURA (ToC): es el equilibrio entre la producción de calor por el cuerpo y su pérdida. El centro termorregulador está situado en el hipotálamo. Cuando la temperatura sobrepasa el nivel normal, se activan mecanismos como vasodilatación, hiperventilación y sudoración que promueven la pérdida de calor. Si por el contrario, la temperatura cae por debajo del nivel normal se activan mecanismos como aumento del metabolismo y contracciones espasmódicas que producen los escalofríos. La temperatura normal del cuerpo de una persona varía dependiendo de su sexo, su actividad reciente, el consumo de alimentos y líquidos, la hora del día y, en las mujeres, de la fase del ciclo menstrual en la que se encuentren, como un mecanismo de defensa ante los procesos infecciosos

[14,2]. Su interpretación depende del sitio en donde se monitorice, ya sea **central**: rectal, esofágica, sublingual, conducto auditivo; o **periférica**: axilar, frontal, inguinal [2].

Anormalidades de la temperatura:

- *Pirexia o hipertermia*: temperatura por encima del límite superior normal. Se presenta aumento de la frecuencia cardiaca, escalofríos, piel pálida y fría y lechos ungueales cianóticos, por vasoconstricción. Se considera que hay fiebre cuando la temperatura corporal es mayor de 37°C en la boca o zona axilar o de 37,6°C en el recto.
- *Hipotermia*: la hipotermia se define como una disminución de la temperatura corporal por debajo de los 35°C.

La fiebre generalmente indica que existe algún proceso anormal en el cuerpo. La gravedad de una condición no se refleja necesariamente en el grado de fiebre. Según la intensidad de la temperatura se clasifica como:

Febrícula: temperatura hasta 38°C.

Fiebre moderada: temperatura entre 38 y 39°C.

Fiebre alta: temperatura superior a 39°C.

Los signos vitales informan y guían a los médicos clínicos sobre el progreso de los pacientes durante las admisiones hospitalarias y los alertan en caso de deterioro del paciente. Una medición de los signos vitales de un paciente es a menudo el primer paso para evaluar cualquier deterioro agudo de su condición clínica [14]. Los signos vitales pueden ayudar a detectar varios problemas, como cardíacos, respiratorios, *shock* y sepsis. Si estos cambios no se detectan y se tratan, existe el riesgo de un paro cardiorrespiratorio [14].

La frecuencia del monitoreo de signos vitales depende de cada política de confianza individual y puede incrementarse en ciertas circunstancias clínicas, como pacientes en estado crítico que requieran un monitoreo más frecuente. Las limitaciones de la práctica de monitoreo actual es que el deterioro entre estos periodos de tiempo de monitoreo puede fácilmente pasarse por alto por los cuidadores del paciente [12].

En ocasiones, el personal que cuida a los pacientes no posee la habilidades y experiencia que se requieren cuando se deteriora su estado clínico [1]. Esto sugiere que incluso en ausencia de fallas de los sistemas de monitorización, los pacientes aún no pueden recibir la atención que requieren porque el personal no posee como prerequisites los conocimientos o habilidades. Tradicionalmente, para identificar cambios clínicos en los pacientes en estado crítico, los (las) enfermeros (as) no solo deben saber cómo medir estos signos vitales con precisión, también deben saber cómo interpretarlos y actuar sobre ellos.

El futuro de la monitorización pasa por diversas etapas, que no son excluyentes entre sí. En todo tipo de monitor se debe buscar el incremento en la precisión en la detección del parámetro a monitorizar. Pero no hay que olvidar que el objetivo de monitorizar a los pacientes es su bienestar y se requiere cuidar las indicaciones, así como la interpretación de los resultados [3,1].

1.2 Trabajos Relacionados

El monitoreo continuo de los signos vitales a través de los últimos avances puede ayudar en la identificación precoz del deterioro del paciente. Afortunadamente, para hacer frente a los desafíos anteriores ha habido desarrollos significativos de sistemas inteligentes de monitoreo para detectar anomalías clínicas en los pacientes confinados a camas de hospital [15,4].

Los signos vitales, tales como la FR, la SpO₂, la TA, la FC y la ToC, son la información más simple, barata y probablemente más importante recopilada en pacientes hospitalizados [16]. Sin embargo, a pesar de que se introdujo en la práctica clínica hace más de un siglo, sorprendentemente se han realizado relativamente pocos intentos para interpretarlos y relacionarlos con el estado clínico del paciente [17,18]. En las últimas décadas, los signos vitales se han convertido en un área de investigación activa [17] y

numerosos estudios han informado que los cambios en los signos vitales se producen varias horas antes de un evento adverso grave [2].

Para este trabajo se desarrolló una búsqueda sistemática de la literatura, siguiendo la Lista de Verificación del Programa de Habilidades de Evaluación Crítica [18,19], para evaluar la calidad de los estudios revisados. En la evaluación, aunque 110 de los estudios cumplieron con un mínimo del 80% de los criterios de evaluación, la mayoría de los estudios incluidos fueron de calidad moderada.

Criterio de elegibilidad: incluimos todos los tipos de estudios sin restricciones de lenguaje o fecha sugeridos por expertos temáticos y metodológicos, que analizarán el monitoreo y correlación de los signos vitales: FC, TA, FR y SpO₂, se incluyeron los siguientes términos: (1) "sensor" o "monitor" o 'signos vitales' "aprendizaje maquinas" "predicción" se incluyeron las investigaciones que se realizaron específicamente en el área de la salud", "en cuidados intensivos pediátrico (UCIP) y neonatal (UCIN)". Limitamos nuestra búsqueda a niños menores de 18 años utilizando los títulos de temas médicos pertinentes. Todos los estudios no publicados encontrados se incluyeron siempre que fue posible para minimizar el sesgo de publicación. Se excluyeron los estudios que reportaban puntaje (*score*) relacionados con los signos vitales, series de casos e informes de casos, estudios en pacientes con afecciones específicas, estudios realizados con pacientes fuera del área hospitalaria. Con estos criterios de elegibilidad se encontraron muy pocos estudios, por lo que tuvimos que ampliar la búsqueda a estudios de pacientes con afecciones específicas.

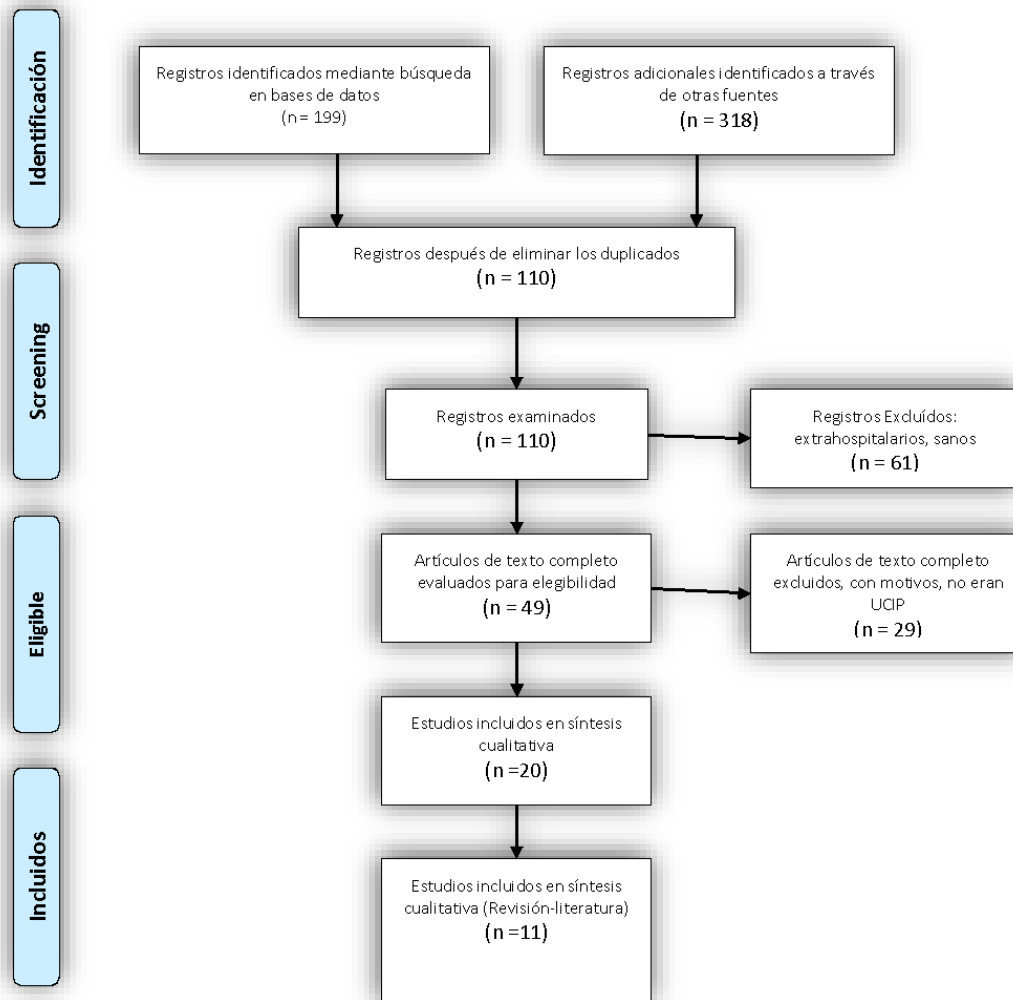
Fuentes de datos: realizamos búsquedas en MEDLINE, -PubMed, Embase, CINAHL, Biblioteca Cochrane, OpenGrey, Biblioteca Digital Xplore del Instituto de Ingenieros Eléctricos y Electrónicos (IEEE), y la Biblioteca Digital de la Association for Computing Machinery (ACM), SCOPUS. Los detalles de la estrategia de búsqueda utilizada se muestran en la Tabla 1-2. La búsqueda final se completó el 12 de marzo de 2021. Cualquier discrepancia sobre la inclusión se resolvió mediante la conversación entre dos miembros del equipo (LI y LN) utilizando la Lista de Verificación del Programa de Habilidades de Evaluación Crítica [18,19] se utilizó la escala MinCir Pr para la evaluación de artículos de predicción [19], que tiene un total de 25 ítems, 6 dominios y un coeficiente de alfa de Cronbach de 0,89. Su puntuación es dicotómica y utiliza una puntuación de 1

para el ítem que no fue cumplido por el artículo y una puntuación de 5 para el ítem que fue cumplido por el artículo. El valor más alto de esta escala es de 125 y el valor más bajo es de 25. Los 6 dominios son: desenlace, exposición, seguimiento, análisis, diseño y variable de confusión, tamaño de la muestra (ver Anexo A) [19]. Cabe anotar que se excluyeron 2 preguntas orientadoras que hacían parte del dominio “tamaño de la muestra”, dado que, para el propósito evaluador de esta revisión, no incluye un número de muestra a calcular.

Resultados: la búsqueda final arrojó 110 estudios después de la eliminación de duplicados. Sin embargo, 61 se consideraron irrelevantes (ver Figura 1-2), porque fueron realizados en el ámbito extrahospitalarios y con pacientes sanos. Se excluyeron 29 estudios evaluados en texto completo, ya que no cumplieron con nuestro criterio de elegibilidad. Estos 29 estudios correspondían a estudios realizados en el servicio de urgencias (14 estudios), en salas de cirugía (9 estudios), y al área de hospitalización o “pisos” (6 estudios). Posteriormente, 9 estudios no se incluyeron en la evaluación síntesis cualitativa en la revisión de la literatura, debido a que se observó que no enunciaron claramente el diseño del estudio y no definían claramente las variables de confusión (ver Figura 1-6).

De los 11 estudios seleccionados, 3 se realizaron en UCIN [17,25,28] y 8 en UCIP [22,23,24,25,26, 27, 30,31]; los países en donde se realizaron los estudios fueron: 5 estudios en Estados Unidos de América [21,24, 37,29,31], 1 estudio en Ruanda África subsahariana [22], 1 estudio en Reino Unido [23], 1 estudio en Japón [25], 1 estudio en Canadá [26], 1 estudio en Corea [30], 1 estudio en países colaborativos USA-India [28]. Todos los estudios correlacionaron los signos vitales FC, FR, TAS, TAD, TAM, SaO₂; algunos incluyeron otros signos como ToC, EKG, llenado capilar [21,22,29]; 3 estudios realizaron predicción en tiempo real de sepsis en UCI correlacionado signos vitales, sin hacer mediciones estáticas como puntuaciones (*score*) [22,25, 28]; y 4 estudios correlacionaron los signos vitales en tiempo real para predecir el riesgo de arresto cardiaco, el tiempo de estancia en UCI pos-intubación, mortalidad [23,24, 29,30].

Figura 1-6. Diagrama de la estrategia de búsqueda



Fuente: Modificada de [20].

Todos los estudios utilizaron herramientas de aprendizaje de máquina para realizar predicciones; los estudios elegidos no realizaron mediciones estáticas como puntuaciones (*score*) para predecir el deterioro de los pacientes. Todos los estudios mostraron métricas de validación de los modelos predictores, con aceptable y buenos resultados; 3 estudios utilizaron curva ROC con área bajo la curva (AUC) > 0,8 [22,24, 30]; 4 estudios utilizaron métricas VPP y VPN con valores > 60% [23,25,27, 31]; y 2 estudios utilizaron regresión logística con un buen rendimiento [28,29]. En el estudio de Ha Uk Chung.et al [21] utilizaron gráfico de Bland-Altman para evaluar el rendimiento del dispositivo diseñado no invasivo

vs. la monitoria con cables que es la utilizada actualmente, obteniendo con el gráfico de Bland-Altman una diferencia media muy cercana a cero (0,16, -0,17) con aceptables resultados. La mayoría de los estudios tuvieron una muestra mayor a 90 pacientes, sólo dos estudios tuvieron menos pacientes como los fueron los estudios de Ha Uk Chung.et al [21], que solo tuvieron 18 pacientes, y el estudio de Ying Zhang et al [31], que contó con 16 pacientes. En cambio, el estudio de BR Matam et al [23] contó con un gran número de muestras, 3836 pacientes (ver Tabla 1-2).

Tabla 1-2 Características de los estudios elegidos

Número	Estudio	Diseño y configuración	Participantes	Intervenciones	Resultados
1	Ha Uk Chung.et al [21]. Science. 2019. USA	Binodal, wireless epidermal electronic systems with in-sensor analytics for neonatal intensive care.	18 neonatos en UCIN	Sensor inalámbrico que captura los signos vitales: FC, FR, SpO ₂ , TA; EKG, ver tendencias de integración signos vitales	La monitorización es igual con cables, y es menos invasiva. Puede reemplazar al sistema de monitorización actual en la UCIN Bland-Altman media diferencia - 0,17; 0,16
2	Arthur Kwizera, et al [22]. Pediatr Crit Care Med. 2019. Ruanda África subsahariana	A Machine Learning-Based Triage Tool for Children with Acute Infection in a Low Resource Setting	949 niños menores de 18 años UCIP	Implementó una herramienta de aprendizaje automático (bosques aleatorios) para desarrollar un modelo que prediga de manera confiable la mortalidad hospitalaria en niños con infecciones agudas. Se correlacionó la edad, los signos vitales frecuencia cardíaca, frecuencia respiratoria y temperatura, el tiempo de llenado capilar, el estado mental alterado al ingreso hospitalario, así como el estado de supervivencia al alta hospitalaria.	Se probaron cinco modelos que podrían predecir la mortalidad hospitalaria con un área bajo la curva característica operativa del receptor que oscila entre 0,69 y 0,8. El modelo que incluyó la edad, la frecuencia respiratoria, el tiempo de llenado capilar y el estado mental alterado exhibió el área de valor predictivo más alto bajo la curva característica de operación del receptor 0.8 (IC del 95%, 0.78-0.8) con el menor número posible de variables.
3	B R Matam. Et al [23]. J Clin Monit Comput. 2019. UK	Machine learning based framework to predict cardiac arrests in a paediatric intensive care unit: Prediction of cardiac arrests	3836 niños menores de 18 años UCIP	Buscar los patrones cambiantes causalmente relacionados en señales como la frecuencia cardíaca, la frecuencia respiratoria, la presión arterial sistólica y la saturación de oxígeno cutáneo periférico para evaluar la predictibilidad de los paros cardíacos en pacientes pediátricos críticamente enfermos en cuidados intensivos.	Se obtuvo una sensibilidad del 71% y una especificidad del 69% un valor predictivo positivo del 11% y un valor predictivo negativo del 98% con una prevalencia del 5% del método de predicción. Mientras que los médicos predijeron 4 de los 69 paros cardíacos (6%), el sistema de predicción predijo 63 (91%) paros cardíacos. Queda la validación prospectiva del sistema automatizado

4	Castiñeira, D. et al [24]. Respiratory Care. 2020. USA	Adding continuous vital sign information to static clinical data improves the prediction of length of stay after intubation: A data-driven machine learning approach	284 pacientes en UCIP	Metodología diseñada para extraer automáticamente información de los datos de signos vitales continuos en el tiempo recopilados de los monitores de cabecera, se aplicaron modelos de aprendizaje automático para predecir la probabilidad de que cada uno de estos pacientes experimente una estadía prolongada definida como > 4 días.	rendimiento del modelo mayor de 83% (área bajo la curva) utilizando sólo información de signos vitales como entrada, y rendimientos del 90% (área bajo la curva) al combinar información de signos vitales con datos clínicos estáticos de los sujetos fácilmente disponibles en formato electrónico.
5	Dagdanpurev S. et al [25]. J Infect. 2019. Japón.	A novel machine-learning-based infection screening system via 2013-2017 seasonal influenza patients vital signs as training datasets	378 pacientes UCIP	Sistemas de detección de infecciones basado en aprendizaje automático que incorpora un algoritmo de árbol aleatorio que utilizan conjuntos de datos de entrenamiento para la detección futura rápida y precisa de la influenza en pacientes con una variedad de antecedentes clínicos, relacionados con los signos vitales	El sistema de detección logró una sensibilidad del 96,2% y un VPN del 96,0% para los pacientes con una variedad de antecedentes clínicos. El sistema de detección mejora a medida que se entrena con más datos.
6	Eytan D. et al [26]. Pediatr Crit Care Med. 2019. Canadá.	Temporal Variability in the Sampling of Vital Sign Data Limits the Accuracy of Patient State Estimation*	747 pacientes menores 18 años UCIP	caracterizar cómo el aumento el tamaño de la ventana temporal de observación en segundos a horas modifica la variabilidad medida y la complejidad de los signos vitales básicos	diferencias Significativas en el oxígeno, pero positivamente con la presión arterial sistólica. Cambios en la variabilidad y complejidad de la frecuencia cardíaca y la presión arterial sistólica desde el momento del ingreso para descargar se encontraron.
7	Georg Seidel. Et al [27]. Computing in Cardiology (CinC), Singapore 2019. Publisher: IEEE	Feasibility of Automated Vital Sign Instability Detection in Children Admitted to the Pediatric Intensive Care Unit	96 pacientes UCIP	Se desarrolló un algoritmo basado en reglas (RBA) para detectar inestabilidades de signos vitales	El mejor resultado del algoritmo basado en reglas generó un 91,6% de respuestas correctas, un 6% de falsos negativos y un 3% de falsos positivos en los datos de la prueba (n = 29)

8	HARPREET SINGH, et al. [28] AI. IEEE Access. 2019. India-USA	Neo-Bedside Monitoring Device for Integrated Neonatal Intensive Care Unit (NICU)	92 pacientes UCIN	Plataforma llamada Neo que automatiza el estado de los signos vitales inmediatos y las tendencias pasadas como gráficos o tablas que el médico y las enfermeras pueden ver desde cualquier lugar a través de Internet. La señal fisiológica y los parámetros clínicos de Neo se utilizan para puntuar diferentes enfermedades como sepsis, síndrome de dificultad respiratoria, enterocolitis necrotizante y retinopatía de precocidad. mediante teorema de Bayes se calcula la probabilidad de deterioro	Regresión logística Se encontró que la proporción de la etiqueta de resultado para los casos de alta morbilidad: baja morbilidad era 40: 30. los datos adquiridos junto con Las características clínicas y de laboratorio específicas para el cuidado del recién nacido pueden ser utilizado para el diseño de predicciones en tiempo real mejores y más rápidas algoritmos
9	Matthew T. Clark, et.al [29]. Pediatric Research. 2012. USA	Predictive monitoring for respiratory decompensation leading to urgent unplanned intubation in the neonatal intensive care unit	287 pacientes UCIN	Probaron la hipótesis de que la respuesta sistémica asociada con la descompensación respiratoria se puede detectar a partir de la monitorización fisiológica de los signos vitales, y que los modelos estadísticos de datos de monitorización a la cabecera pueden identificar a los bebés con mayor riesgo de intubación urgente no planificada.	El modelo predictivo (regresión logística), validado internamente por Bootstrap, tenía un receptor operando área característica de 0.84 ± 0.04 .
10	Soo Yeon Kim, et al [30]. Critical Care. 2019. Korea	A deep learning model for real-time mortality prediction in critically ill children	1445 pacientes UCIP	modelo basado en aprendizaje automático, la Herramienta de predicción del riesgo de mortalidad pediátrica (PROMPT), para tiempo real predicción de la mortalidad por todas las causas en las unidades de cuidados intensivos pediátricos	el modelo logró un área bajo la curva característica de funcionamiento del receptor en el rango de 0,89 a 0,97 para la predicción de mortalidad 6 a 60 h antes de la muerte
11	Ying Zhang, et al [31]. Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France. 2007. USA	Real-Time Development of Patient-Specific Alarm Algorithms for Critical Care	16 pacientes UCIP	Esta la investigación examina la viabilidad de desarrollar algoritmos de alarma en tiempo real junto a la cama del paciente y evalúa el potencial de estos algoritmos para ayudar a mejorar la vigilancia en (UCI), se realizó árboles de clasificación y redes neuronales para generar alarmas clínicas para pacientes de forma individual.	El algoritmo redes neuronales Su especificidad, valor predictivo positivo y La precisión supera la de la alarma de umbral estándar algoritmo.

UCIP: Unidad de Cuidado Intensivo Pediátrico; UCIN: Unidad de Cuidado Intensivo Neonatal; UCI: Unidad Cuidado Intensivo

Los resultados de esta revisión de la literatura sugieren que la combinación del uso generalizado de sistemas electrónicos en el área de la salud, utilizados para registrar los signos vitales monitoreados de forma continua con análisis de tendencias y utilizando modelos predictivos podría mejorar la predicción del deterioro del paciente antes de un evento adverso grave y ayudar a dirigir recursos limitados hacia los pacientes en riesgo.

Los grandes esfuerzos hasta el momento en el área crítica se han desarrollado para evaluar los signos vitales con medidas estáticas, como lo son las puntuaciones (*score*), y este tipo de estudios en revisiones sistemáticas de la literatura han mostrado que tienen una aparente falta de evidencia de alta calidad con respecto a las tendencias en los signos vitales monitoreados de manera continua [32,33]. Los estudios incluidos son análisis retrospectivos de cohortes preexistentes, sin grupos de control y con completa heterogeneidad. Por lo tanto, tienen una certeza de evidencia baja (o muy baja) [32, 33]. Los estudios que incluyen monitoria continua de signos vitales con modelos de predicción en la cabecera del paciente se han desarrollado más frecuentemente en pacientes adultos, con patologías crónicas y pacientes sanos [34,35,36]. Los estudios con medidas estáticas realizados hasta el momento gozan de una buena precisión para la detección del deterioro clínico en las salas generales y en los servicios de urgencias [32,33]. Sin embargo, es difícil probar la validez externa de estos hallazgos. Además, los resultados muestran que no hay consenso sobre cómo analizar mejor las tendencias [32]. Dado que los modelos de tendencias se validan externamente a través de estudios de cohortes multicéntricos prospectivos bien controlados, siguen siendo estudios prometedores y bienvenidos como una valiosa adición al apoyo a las decisiones clínicas.

1.3 Motivación

Estamos motivados por estudios previos que demuestran que el deterioro agudo en los pacientes en UCIP a menudo están precedidos por cambios sutiles en los parámetros fisiológicos [5,7,36,37,38,39,40]. Por lo tanto, este trabajo se enfocó en el descubrimiento del conocimiento clínico basados en interacciones entre múltiples signos vitales por medio de un modelo con inclinación probabilística para la predicción de eventos clínicos anormales en un seguimiento personalizado en tiempo real.

El modelo de monitorización en UCIP de atención tradicional actual descuida la interpretación individual de los pacientes. Una serie de sistemas inteligentes para predicción de deterioro de los pacientes existen, pero son más bien sistemas estáticos como lo son las puntuaciones de *score*, la mayoría de los sistemas de predicción utilizan muestras de un pequeño grupo de pacientes y trabajan sólo en un único parámetro como la TA o FC, estos modelos solo predicen el futuro comportamiento de algunas condiciones clínicas especiales [35,36], dichos sistemas sufren una alta tasa de falsas alertas cuando la alta incertidumbre en los datos desaparece.

El análisis en este trabajo mejora la capacidad de generalización del modelo aprendido al considerar multiparámetros de muchos pacientes, utiliza la agrupación de casos clínicos y luego correlaciona las características para reconocer la información oculta en estados futuros utilizando observaciones pasadas. Es por esto que el objetivo principal de este estudio es proponer una transformación del espacio de signos vitales para detectar el deterioro de los pacientes de una unidad de cuidados intensivos pediátricos

● **Capítulo 2. Marco Conceptual de los Métodos a utilizar**

En este capítulo se revisan algunos conceptos fundamentales de los métodos de minería de datos utilizados, que soportan la metodología empleada en este trabajo. Se describen técnicas como Análisis de Componentes Principales, Análisis de Series de Tiempo, Análisis de Conglomerados, Métodos de Clasificación, Cadenas de Márkov, y Modelos Ocultos de Márkov.

2.1. Análisis de Componentes Principales (ACP)

Es una técnica de análisis multivariante (o multivariada) de reducción de datos. Este método tiene como objetivo transformar un conjunto de variables originales en un nuevo conjunto de variables, sin perder información, siendo una combinación lineal de las originales; se caracteriza porque las variables generadas no están correlacionadas entre sí, y otra particularidad es que estas variables tienen una importancia diferencial y conocida en la explicación de la varianza total, estas nuevas variables son denominadas componentes principales. El ACP también se puede emplear como parte de un análisis exploratorio y para construir modelos predictivos [41].

Las ventajas de este método se pueden resumir como: [41,42]

- Genera información a través de componentes que estructuran nuevas variables
- Reduce la dimensión del espacio en el que están contenidos los datos
- Elimina las variables que aportan poca información
- Los nuevos componentes facilitan la interpretación de la información
- La determinación de unos pocos factores (componentes principales) retienen la mayor variabilidad contenida en los datos. Las variables generadas cumplen con

las características de independencia (bajo el supuesto de normalidad) y no correlación.

El proceso de ACP tiene cinco etapas: [41,42]

- a. Cálculo de los ejes de los componentes.
- b. Cálculo de los valores propios o de las varianzas incorporadas a cada uno de los ejes retenidos.
- c. Reorganización de la matriz de datos en los nuevos ejes retenidos.
- d. Cálculo de la correlación de los componentes con las variables primitivas, comunalidades (proporciones de varianzas compartidas) y recomposición de la matriz de correlaciones.
- e. Interpretación de las componentes en el espacio transformado.

a. Cálculo de los ejes de los componentes

se trata de encontrar las dimensiones latentes del campo de variables que se ha considerado, con relación a los datos introducidos. Es decir, el espacio vectorial generado a partir de las variables iniciales y en donde los datos configuran una nube de N puntos. Las condiciones impuestas para la construcción de dichos vectores o dimensiones son:

Que sean base del espacio vectorial, esto es, que sean linealmente independientes, que acumulen o expliquen la máxima varianza de la inercia total del sistema, que dicha varianza extraída en cada uno de los ejes se realice de manera jerárquica, y, como condición complementaria, que sean unitarios. Todo lo anteriormente expuesto es lo mismo que hallar los vectores propios de la matriz de varianza/covarianza o de correlaciones.

Cuando el método está basado en la matriz de correlación, los datos no son dimensionalmente homogéneos. y cuando el método está basado en la matriz de covarianza, los datos son dimensionalmente homogéneos y presentan valores medios similares.

Una limitación del ACP es la suposición de linealidad, que asume que los datos observados son una combinación lineal de una cierta base.

Es necesario siempre consultar los índices sobre la conveniencia de utilizar el método de ACP. A continuación, presentaremos los más frecuentemente utilizados [41,42].

El Test de Barlett: es un indicador del grado de esfericidad de las variables iniciales, se utiliza para probar si un conjunto de muestras proviene de poblaciones con la misma varianza. A las varianzas iguales a través de las muestras se llama homocedasticidad u homogeneidad de varianzas. La significación da la probabilidad de que se cumpla la hipótesis de no existencia de correlación entre las variables introducidas, o lo que es equivalente de independencia entre ellas. Esta significancia usualmente debe ser menor que 0.05.

Índice de Kaiser-Meyer-Olkin (KMO): se basa precisamente en la relación existente entre el conjunto de coeficientes de correlación simple y de correlación parcial. Se entiende que, si el conjunto de los coeficientes de correlación parcial es de valores reducidos, la parte específica de las variables será menor con relación a la parte común, lo que hará más pertinente la realización del ACP. Este índice oscila entre 0 y 1; la tendencia a 1 es indicativa de la validez de la aplicación al caso del ACP.

b. Cálculo de los valores propios o de las varianzas incorporadas a cada uno de los ejes y del número de los mismos a retener

Consiste en reducir el espacio de atributos inicial. Esta reducción implica normalmente una pérdida de la inercia total o, dicho con otras palabras, si se extraen menos dimensiones de las que configuran el espacio inicial no se podrá dar cuenta de la inercia total del sistema por ejes linealmente independientes. Pero la pérdida también puede implicar ganancia en cuanto a sencillez de la estructura adoptada por el método y en cuanto a la interpretación

de los ejes. De la observación de esta reducción se deduce el rápido decrecimiento de dichos valores y su acumulación en los primeros valores, facilitando la reducción.

Debemos decidir entre la simplificación que supone la reducción de la dimensión y la pérdida de información resultante de la variabilidad no explicada. Como criterios para tomar esta decisión, se suelen emplear los siguientes:

- *Criterio de la varianza explicada.* consiste en retener el número de componentes que conjuntamente explican una proporción de varianza establecida, habitualmente debe ser mayor al 80% del total y significa que el número de factores es suficiente.
- *Criterio del valor propio (Kaiser).* consiste en retener sólo las componentes cuyos valores propios sean mayores que la unidad.
- *Gráfico de sedimentación (scree plot).* representar en un gráfico los valores propios $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ en orden decreciente, y buscar un "codo" en el gráfico, entendiendo por codo un punto a partir del cual los valores propios son claramente más pequeños que los anteriores, y muy similares entre sí y *retener un número preestablecido de componentes principales.* Por ejemplo, es costumbre retener dos componentes, pues se pueden representar fácilmente en el plano.

c. Reorganización de la matriz de componentes en los nuevos ejes retenidos

Una vez reducidas las variables a un número, en principio, considerablemente inferior es posible calcular las puntuaciones factoriales para cada dato o muestra. Los componentes o factores, como nuevas variables, tendrán un coeficiente determinado que podrán ser reutilizados posteriormente en otros tipos de análisis. Estas componentes quedan guardadas y pueden recuperarse en tanto que la nueva matriz continúa ubicando así los datos en el nuevo espacio.

d. Cálculo de la correlación de las componentes con las variables primitivas, comunalidades, recomposición de la matriz de correlaciones.

El interés de esta fase es encontrar la relación que las variables primitivas tienen con las componentes. Dicha relación permite cubrir varios objetivos simultáneamente:

- Transformar las variables originales en los nuevos ejes lo que nos mostrará la estructura del primer espacio de características, en otras palabras se realiza un cambio del sistema de referencia, pasamos de las coordenadas de las variables originales observadas a un nuevo sistema que es el espacio de los factores y esta transformación de la expresión de las componentes en función de las variables iniciales se formará en la matriz de saturación, y su importancia consiste en que, a partir del significado de sus coeficientes, puede interpretarse como un nuevo espacio vectorial en el que las variables primitivas son puntos en las nuevas dimensiones elegidas, Los valores de la matriz de saturaciones son los coeficientes de correlación entre las variables originales y los factores. Cuanto más elevado es el coeficiente más relevante es la variable para configurar el factor y también mayor será su proyección sobre el eje. Si elevamos al cuadrado y sumamos los coeficientes de cada columna de la matriz de saturaciones obtenemos los valores propios, es decir, la varianza incorporada en cada eje o varianza explicada por el factor o componente. Al conocerse estos nuevos componentes se procede a conocer la comunalidad de estos nuevos componentes, que no es más que la suma de los pesos factoriales al cuadrado en cada una de las filas de la matriz de saturaciones o la proporción de la varianza explicada por los componentes o factores comunes en una variable. [41,42].
- Interpretación geométrica de los componentes: geoméricamente las comunalidades se interpretan como la magnitud de dicho vector propio en el espacio de las componentes. Tomadas todas las componentes, sin reducciones, la comunalidad de cada variable es la misma e igual a 1, y pone de manifiesto el hecho de que todas intervienen comúnmente en la estructuración de la totalidad o contribuyen igualmente a la inercia total. La comunalidad inicial no es idéntica para todas las variables puesto que se parte de hipótesis previas en las que la parte

común de cada variable que interviene en la explicación de la totalidad, y por tanto su parte específica, es diferente. Por consiguiente, comunalidades altas serán interesantes en el sentido de que tiene importancia en la creación de los ejes del sistema o a estructurar el espacio. Dichas variables estarán alejadas en la representación del centro de la misma. Variables con comunalidad baja contribuyen poco a la formación del sistema o a estructurar el espacio, lo que significa que son variables que contribuyen poco a dispersar la nube de puntos: se situarán en puntos próximos al centro de masas del sistema, contrariamente a las primeras que intervienen fuertemente en la dispersión de la población. La rotación de los ejes no afectará a la comunalidad de cada variable con relación a sus nuevos componentes que se tornan. Por ello las reflexiones que se hacen aquí continuarán siendo válidas una vez rotados los ejes iniciales por el procedimiento de rotación que se elija.

e. Interpretación de las componentes, en el espacio rotado

Habitualmente la extracción inicial de los componentes o factores no permite identificar con toda claridad la relación o el modelo subyacente que se establece entre los componentes o factores y las variables. Con el objeto de evidenciar esta relación se procede a la llamada rotación de los factores que consiste en una transformación de la matriz factorial original en otra más simple que adecúa mejor los ejes al aproximarlos a las variables correlacionadas. Facilitando la interpretación de la estructura de los datos no se altera la bondad de ajuste de la solución factorial, las comunalidades y los porcentajes de varianza explicada se mantienen inalterados, simplemente se redistribuye la varianza explicada entre los componentes o factores. Los procedimientos empleados son de dos tipos: a) la rotación rectangular y b) la rotación oblicua [41,42].

- a) La rotación rectangular: la rotación se efectúa haciendo que los ejes permanezcan perpendiculares. Dicha transformación, al conservar las distancias, deja inalterable la comunalidad de cada variable lo que hace que las interpretaciones encontradas a partir de ellas sean las mismas. No así el valor propio de cada componente, siendo diferente

su importancia relativa en la explicación de la varianza total. Los procedimientos más frecuentemente utilizados en rotación rectangular son tres:

Varimax: la más frecuentemente utilizada en ACP, esta técnica se caracteriza porque minimiza el número de variables que tienen un factor o componente de saturación sobre una variable, acentuando los que lo tienen más elevado. Esto hace que se obtengan factores con correlaciones altas con un número pequeño de variables y correlaciones nulas en el resto, quedando así redistribuida la varianza de los factores.

Quartimax: es una técnica que minimiza el número de factores que corresponden a una variable. Consiste en que cada variable se proyecte al máximo sobre factores o componentes diferentes, dentro de los límites del método. El supuesto más adecuado para su aplicación consiste en que las variables introducidas tuvieran determinados grados de independencia supuesta.

Equimax, biquartimax: es una combinación de las dos técnicas anteriores.

b) La rotación oblicua: se utiliza cuando los ejes rotados a partir de los primeros factores o componentes no conservan la ortogonalidad, lo que tiene como consecuencia que tampoco conserven la comunalidad de cada variable, y desde luego rompe con uno de los objetivos que consiste en buscar la no correlación de los ejes. La técnica es más útil y utilizada con análisis factoriales confirmatorios.

Junto con las representaciones gráficas se pueden sugerir varios criterios para encontrar los ejes factoriales: [41,42].

a) Comenzar a identificar el eje de más peso teniendo este dato presente para dar a dicho eje la importancia que le corresponde.

b) Dentro de las variables de máxima comunalidad elegir las de mayor valor sobre el eje analizado y mínima en los otros, lo que equivale a elegir del eje estudiado aquellas variables que mayor coeficiente de saturación tienen sobre él y mínimo en los otros ejes.

c) Establecer una escala de variables que recorran todo el eje, independientemente de su comunalidad: primero, para las variables más próximas a él, y luego para otras siempre que se mantengan constantes con relación a los otros ejes.

d) Agrupar las variables por proximidad, sea en todos los ejes, sea en pares de ellos. Así obtendremos una agrupación de las mismas. Se ubican estos grupos sobre la representación de los ejes y contribuirá a comprender mejor la estructura de la realidad analizada.

e) Establecer agrupaciones de variables según comunalidades.

2.2. Análisis de Series de tiempo

Una serie de tiempo se define como el conjunto de mediciones producidas en determinados momentos durante un periodo de tiempo que puede ser semanal, mensual anual, segundos, minutos, horas, etc. El principal objetivo de analizar una serie de tiempo es poder realizar pronósticos. A continuación, se exponen algunos métodos utilizados para analizar las series de tiempo en este trabajo.

2.2.1. Análisis espectral (*Transformada de Fourier*)

El objetivo de la Transformada de Fourier (TF) se basa en el análisis de señales aperiódicas y periódicas. La TF permite descubrir aspectos de la señal que serían muy difíciles o imposibles de observar a partir de su representación temporal. Para las señales las siguientes ecuaciones representan [43]

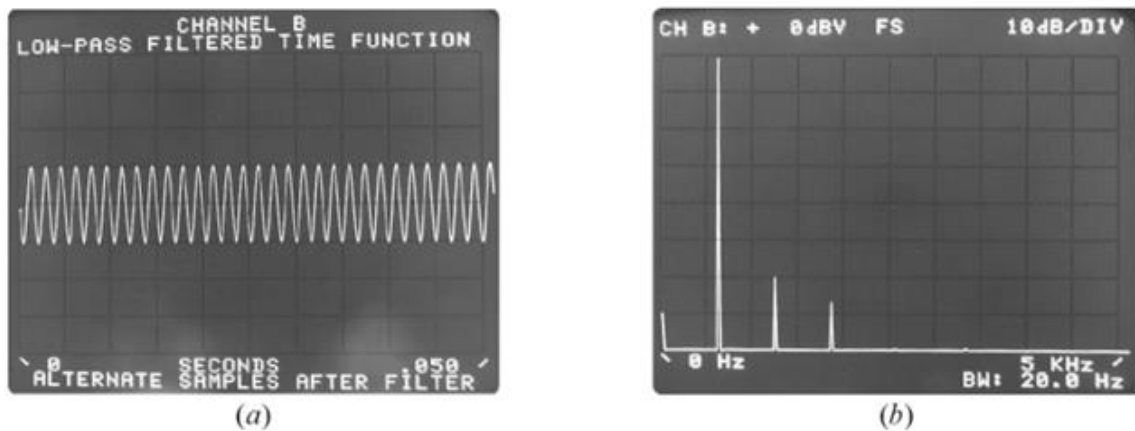
$$\begin{aligned}x(t) &= \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df, \\X(f) &= \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt,\end{aligned}$$

ecuación 1,2

la ecuación 1 se denomina transformada de Fourier de $x(t)$, o ecuación de análisis, mientras que la ecuación 2 es la transformada inversa de Fourier de $X(f)$, o también

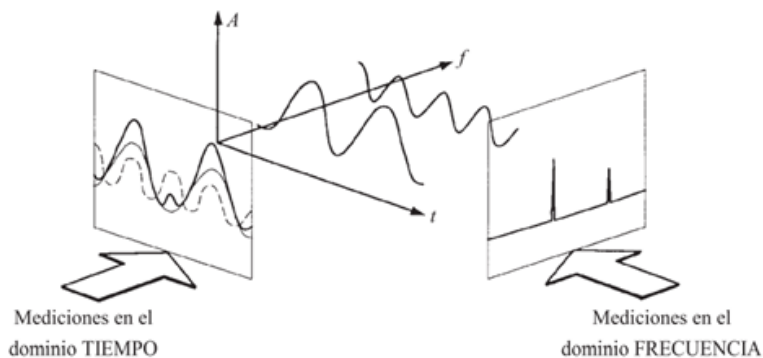
ecuación de síntesis. ambas tratan la descomposición de una señal como una combinación lineal de exponenciales complejas. La transformada $X(f)$ de la señal periódica $x(t)$ se denomina espectro de $x(t)$, ya que brinda información referida a cómo se combinan ondas sinusoidales de distinta frecuencia para componer la señal $x(t)$.

Figura 2-1 (a) representación temporal; (b) representación frecuencial



Fuente: tomado de [43]

Figura 2-2. Relación entre el dominio temporal y frecuencial



Fuente: tomado de [43].

El análisis de señales en el campo transformado permite descubrir aspectos de la señal que serían difíciles o imposibles de observar a partir de su presentación temporal (ver

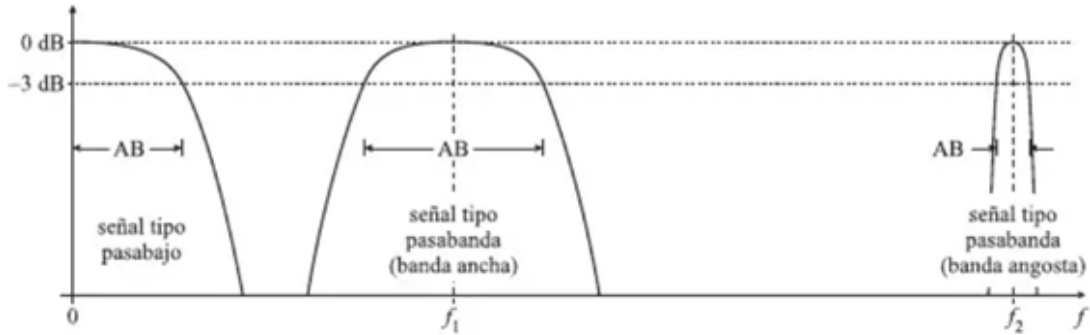
figuras 2-1 y 2-2). En las figuras se representan las señales en el espacio temporal y luego se transforman en espacio frecuencial, es decir, se puede apreciar su relación entre el dominio temporal y el dominio de la frecuencia de una manera más clara. Ciertas observaciones como el tiempo de crecimiento rápido, la caída de un pulso, el sobrepico, y las oscilaciones amortiguadas (*ringing*) de una señal resultan más conveniente de observar en el dominio temporal y no en el dominio de la frecuencia [43, 44].

Teorema del Ancho de Banda

El teorema del ancho de banda aplicado a la TF indica que una señal no puede estar limitada simultáneamente en tiempo y frecuencia. El teorema del ancho de banda, o principio de incertidumbre de Fourier formaliza esta idea. En este teorema se describe que señales extremadamente concentradas en un dominio pierden su localidad en el otro. Esta falta de localidad conjunta es causa de varios problemas cuando se trabaja con aplicaciones en el dominio frecuencial. Por este motivo, cuando hay altibajos en las señales, se puede filtrar la señal en el dominio de la frecuencia para eliminar altas o bajas frecuencias de las señales. Para evaluar mejor estas variaciones de mayor o menor intensidad, a las que también se les puede llamar *ruido*, el procedimiento para reducirlo o eliminarlo se realiza a través de lo que se conoce comúnmente como filtrado. Al aplicar un filtro a una señal lo que se hace es discriminar una determinada frecuencia o gama de frecuencias de una señal y pasarla por el filtro pudiendo modificar tanto su amplitud como su fase [43, 44].

Si el contenido espectral de una señal está concentrado cerca del origen, como el espectro de la izquierda de la Figura 2-3, se dice que la señal es de tipo pasa-bajos. La señal es de tipo pasabanda si el contenido espectral está centrado alrededor de $\pm f_c$, donde f_c , es una frecuencia no nula, como los espectros del centro y de la derecha de la Figura 2-3.

Figura 2-3. Espectros de señales pasabajos, pasabanda de banda ancha y pasabanda de banda angosta.



Fuente: Tomado de [43].

Otra definición existente en cuanto a los tipos de filtros consiste en que al filtrar la señal con un filtro pasabajo deja inalterables las bajas frecuencias (por ejemplo, menor a 200 Hz) y atenúa o elimina las altas frecuencias. Por otro lado, los filtros *pasan alto* dejan inalterables las altas frecuencias (mayor a 2000Hz) y atenúa o elimina las bajas frecuencias. Los filtros de pasabanda atenúan las altas y las bajas frecuencias, pero mantiene intactas las frecuencias que se encuentran en una banda determinada [43,44].

2.2.2. Análisis estadístico de series de tiempo

Las observaciones de una serie de tiempo pueden ser denotadas como: [45,46]

$$Y_1, Y_2, Y_3, Y_4, \dots, Y_{t-1}, Y_t$$

Y_t es el valor tomado por el proceso en el instante t

$$t=1,2,3,\dots,\text{etc.}$$

Habitualmente para realizar el análisis de series temporales, se supone que los valores que toma la variable de observación es la consecuencia de tres componentes, como lo son: [45,46].

a.- Componente tendencia. Se puede definir como un cambio a largo plazo que se produce en la relación al nivel medio, o el cambio a largo plazo de la media.

b.- Componente estacional. Es un comportamiento o patrón que consiste, por ejemplo, en subidas o bajadas periódicas que se presentan en forma regular en la serie de tiempo. Estos efectos son fáciles de entender y se pueden medir explícitamente o incluso se pueden eliminar de la serie de datos, a este proceso se le llama desestacionalización de la serie.

c.- Componente aleatoria. Este componente no responde a ningún patrón de comportamiento, sino que es el resultado de factores fortuitos o aleatorios que inciden de forma aislada en una serie de tiempo.

Las series temporales también se pueden clasificar de manera descriptiva como:

a.- Estacionarias. Una serie es estacionaria cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo.

b.- No estacionarias. Son series en las cuales la tendencia o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

Las series temporales también se definen como un caso particular de los procesos estocásticos. También puede presentarse como un proceso estocástico estacionario, si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende solamente de la distancia o rezago (retraso o residuo) entre estos dos periodos de tiempo y no del tiempo en el cual se ha calculado la covarianza. Pero también se puede presentar un caso simple de los procesos estocásticos como lo es el Ruido Blanco, donde los valores son independientes e idénticamente distribuidos a lo largo del tiempo con media cero, la varianza es constante y es incorrelacionada.

Funciones de autocorrelación

La autocorrelación se define como la relación mutua existente entre valores de una serie de tiempo en diferentes periodos y describe la relación entre un valor con respecto al otro.

Hay dos tipos de autocorrelación que sirven para conocer la existencia de correlación entre los datos de la serie como el grado o intensidad o el tipo de la misma.

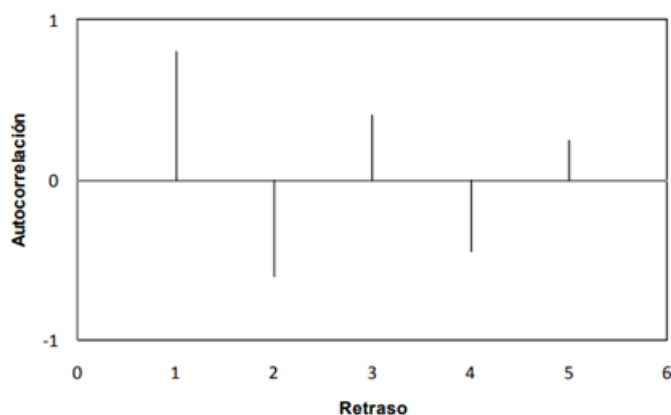
Autocorrelación simple (ACS): mide la relación lineal entre las observaciones de una serie de datos Y_t , distanciados en un lapso de tiempo. A este lapso de tiempo k se le conoce como retardo o retraso. El retardo k , es la autocorrelación entre los valores de las series que se encuentran a k intervalos de distancia [44, 45, 46].

Autocorrelación parcial (ACP): es una medida asociada a la autocorrelación simple. Es la estimación de la autocorrelación simple, para el mismo retardo k , con la eliminación del efecto producido por las autocorrelaciones para retardos menores, ajustada por el efecto de los retardos intermedios, lo que quiere decir, no se considera la dependencia creada por los retardos intermedios existentes entre ambas [44, 45].

El grado de relación entre autocorrelación simple y parcial se mide mediante el coeficiente de autocorrelación tanto para ACS y ACP, ambos coeficientes varían entre +1 y -1. Un valor positivo +1 indica una fuerte relación entre dos observaciones separadas k unidades de tiempo. Y la relación es positiva, es decir, si una observación aumenta o disminuye la otra también lo hará. Un coeficiente de autocorrelación cercano a -1 indica una relación inversa entre las observaciones. Y un coeficiente de 0 indica que las dos observaciones de una misma serie de tiempo no están relacionadas [44, 45, 46,47].

La gráfica de ACS y ACP se realiza mediante una gráfica que se llama correlograma como se observa en la Figura 2-4, en donde se grafican las autocorrelaciones para varios retrasos de tiempo de una serie de tiempo.

Figura 2-4 Correlograma o función de autocorrelación



Fuente: Tomado de [45].

En esta figura 2-4 la escala horizontal muestra cada uno de los retrasos de tiempo que son de interés, mientras que la escala vertical del lado izquierdo muestra el posible rango de un coeficiente de autocorrelación. La línea horizontal en el centro de la gráfica representa autocorrelaciones de cero.

Prueba de Ljung-Box

Esta prueba permite probar en forma conjunta que todos los coeficientes de autocorrelación son simultáneamente iguales a cero, esto es, que son independientes. Esta prueba estadística se requiere para demostrar si una serie de observaciones en un período de tiempo específico son aleatorias e independientes, y está definida como [44,45, 48]:

$$LB = n(n + 2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n - k} \right) \sim \chi^2_{(m)} \quad \text{ecuación 3}$$

donde n es el tamaño de la muestra y m la longitud del rezago.

Hipótesis nula (Ho): las autocorrelaciones son independientes.

Hipótesis alterna (Ha): las autocorrelaciones no son independientes.

En una aplicación, si el cálculo excede el valor crítico de la tabla cuadrada al nivel de significancia seleccionado, no se acepta la hipótesis nula de que todos los coeficientes de autocorrelación son iguales a cero; por lo menos, algunos de ellos deben ser diferentes de cero.

Modelos Autorregresivos: ARIMA

Los modelos autorregresivos se basan en la idea de que el valor actual de la serie, X_t puede explicarse en función de p valores pasados X_{t-1} , X_{t-2} , X_{t-p} , es decir, las evaluaciones futuras vienen explicadas por los datos del pasado. [44, 45, 47, 48]

La estructura de ARIMA combina los modelos de autorregresión y se define según sus componentes AR, I, MA de la siguiente manera:

AR: se refiere a los periodos (año, mes, segundos, minutos) o número de estados anteriores que predicen el modelo, con esta componente se intentan predecir el comportamiento de una variable. Un modelo AR1 puede ser igual a una cadena de Márkov, porque es un modelo simple.

I: se refiere a la integración, quiere decir que el modelo realiza un proceso en aquellas series que requieren ser estacionarias, para hacerlas estacionarias en relación con la media, si es de orden 1, significa que requirió 1 proceso de diferenciación para ser estacionaria, con orden dos, dos procesos de diferenciación, si es de orden 0, significa que son estacionarias y no se requiere proceso de diferenciación.

MA: se refiere a el número de medias móviles que aplica el modelo y se emplea para series de una variable, con este proceso se marca el cambio de las variables en función del tiempo y de su relación aleatoria. Es decir, con este componente se puede predecir una variable con base a los errores pasados observados.

2.3. Análisis de Agrupamiento

2.3.1. K-medias (*k-means*)

K-means es un método de agrupamiento, que consiste en un aprendizaje automático no supervisado y tiene como objetivo la partición de un conjunto de n observaciones en k grupos y cada observación pertenece al grupo cuyo valor medio es el más cercano. El algoritmo utiliza un proceso iterativo en el que se van ajustando los grupos para producir unos conglomerados finales. Para ejecutar el algoritmo se pasa como entrada el conjunto de datos numéricos y un valor del número de grupos deseados, que se conoce como el valor de k . El conjunto de datos serán las características (*features*) para cada punto, luego las posiciones iniciales de los k centroides se asignan de manera aleatoria de cualquier punto de los datos de entrada, aunque puede hacerse otra escogencia aleatoria de los centroides, y luego se itera en dos pasos [49,50]:

- **Fase de asignación** continúa ubicando cada observación al centroide más cercano basado en la distancia Euclidiana, como lo indica la siguiente fórmula, ver ecuación 4 [51]:

$$\arg \min_s \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg \min_s \sum_{i=1}^k |S_i| \text{Var } S_i$$

ecuación 4

donde μ_i es la media de puntos en el grupo S_i

Debido a que la varianza total es constante, esto también equivale a maximizar las desviaciones al cuadrado entre puntos de diferente clúster (suma de cuadrados entre clúster) [51], como se observa en la ecuación 5

$$\sum_{\text{Cluster } C_i} \sum_{\text{Dimensión } d} \sum_{x,y \in C_i} (x_d - y_d)^2$$

ecuación 5

- **Fase de actualización:** calcula nuevos centroides para las observaciones de grupo. Esto se hace tomando la media de todos los puntos asignados en el paso anterior, ver ecuación 6

$$C_i = \frac{1}{|S_i|} \sum x_j \in \text{ecuación 6}$$

El proceso finaliza cuando las asignaciones no presentan cambios. [42,46, 51] y presenta criterios de detención tales como:

- * si no hay cambios en los puntos asignados a los grupos
- * si la suma de las distancias se minimiza, o
- * se alcanza un máximo de iteraciones.

Como se desconoce el valor de k adecuado, se debe realizar el proceso para elegir el valor de k óptimo, y para esto se utilizan frecuentemente las siguientes técnicas [49, 50, 51]:

Método del codo

Al aplicar los métodos de conglomeración puede ocurrir que se agrupen datos heterogéneos en pocos grupos, o en su defecto, que datos muy cercanos entre sí acaben asignados a grupos diferentes. Para reducir cualquiera de estas posibilidades, se aplica a la serie de datos el método del codo (*elbow method*, en inglés), que permite la obtención de la cantidad óptima de clúster s, y representa el punto donde la inercia, luego de cambios bruscos, resulta casi invariable [49,51].

Coficiente de Silhouette (silueta)

Como complemento del método del codo (EM), se emplea una técnica para determinar el número más preciso de grupos a construir, como lo es el coeficiente de silueta. Es una medida de similitud que permite conocer el grado de cohesión y calidad de la agrupación de los datos dentro de un conglomerado, al tiempo que permite reconocer la distancia de un dato dentro de un grupo en relación con los datos de un grupo vecino. El rango del coeficiente se encuentra entre (-1, 1). Los valores cercanos a 1 indican que la observación

está agrupada eficientemente, y, los valores cercanos a (-1) o negativos indican que los datos fueron asignados al conglomerado erróneo [49, 51].

2.3.2. Método de agrupación Bietápico

Los conglomerados obtenidos por el método de agrupación bietápico o de dos fases ofrecen otros criterios de análisis. El proceso del análisis bietápico consiste en segmentar de forma natural y automática un número óptimo de grupos. Utiliza una prueba de hipótesis por medio de la cual puede identificar las diferencias significativas entre los conglomerados y capturar la heterogeneidad en los comportamientos combinados de los conglomerados. Y, por último, se pueden identificar diferentes funciones de distribución de probabilidad de parámetros relacionados mediante el análisis de los conglomerados. En los conglomerados bietápicos se pueden procesar datos en forma continua y categórica [47]. El análisis se realiza en dos pasos: agrupamiento previo (pre clúster) y agrupamiento [52, 53].

En el **primer paso** de agrupamiento previo (*pre-agrupamiento o en su sigla en inglés pre-clúster ing*), se escanean todos los casos de los datos y se mide la distancia logarítmica de probabilidad entre ellos para determinar si van a formar agrupamientos basándose en algún criterio de distancia umbral [45]. Se utiliza la distancia logarítmica de verosimilitud que es una distancia basada en la probabilidad y la distancia entre dos conglomerados está relacionada con la disminución de la probabilidad logarítmica a medida que se combinan en un grupo [53].

En el **segundo paso**, mediante un algoritmo de agrupación aglomerativa, los subgrupos resultantes de la primera etapa se agrupan en un número óptimo de conglomerados. Los subgrupos se consideran casos únicos y se fusionan en un grupo al satisfacer la distancia mínima, hasta que todos los datos se colocan dentro de un solo grupo. Si un número deseado de grupos no está predeterminado, el algoritmo puede determinar

automáticamente el número óptimo de grupos utilizando el Criterio Bayesiano de Schwarz (BIC) o el Criterio de información de Akaike (AIC), ver ecuaciones 7 y 8 [53, 54].

$$AIC(J) = -2 \sum_{j=1}^J \epsilon_j + 2m_j \quad \text{ecuación 7}$$

$$BIC(J) = -2 \sum_{j=1}^J \epsilon_j + m_j \log(N) \quad \text{ecuación 8}$$

donde J es el número de conglomerados; K es el número total de conglomerados de las variables continuas; N es el número total de observaciones; $m_j = 2KJ$.

Criterio de Información de Akaike (AIC): El criterio de Akaike combina la teoría de la máxima verosimilitud, información teórica y la entropía de información. Este criterio tiene en cuenta los cambios en la bondad de ajuste del modelo y también las diferencias en el número de parámetros entre dos modelos, siendo los mejores modelos los que tengan un valor de AIC más bajo, y el modelo que mejor explica los datos con el mínimo número de parámetros es el que presenta más bajo valor de AIC [53, 54].

Criterio Bayesiano de Schwarz (BIC): En este sentido Schwartz estableció que la solución de Bayes consistía en seleccionar el modelo con una alta probabilidad a posteriori, por lo que en este modelo se aplica esta probabilidad a posteriori para grandes volúmenes de datos. El modelo con el valor BIC más bajo se considera el mejor en explicar los datos del análisis con el mínimo número de parámetros [53, 54].

Es importante mencionar que el método de agrupación bietápico durante el segundo paso, realiza un análisis sobre la importancia de la variable contribuye a la predicción a partir del conjunto de datos que se ha agrupado. Este proceso indica el orden que han tomado las variables de la matriz de datos para diferenciar los diferentes clústeres que se han formado, de manera que cuanto mayor es la medida de importancia que tiene la variable, menos probable será la variación para una variable entre clúster s debido al azar y más probable será debido a alguna diferencia subyacente de las propias variables [53].

2.4. Métodos de Clasificación

Una técnica de clasificación (o clasificador) es un enfoque sistemático basado en un análisis supervisado para construir modelos de clasificación a partir de un conjunto de datos de entrada. Se emplea un algoritmo de aprendizaje para identificar el modelo que mejor se ajusta a la relación entre el conjunto de atributos y la etiqueta de clase de los datos de entrada. El modelo generado por un algoritmo de aprendizaje debe ajustarse bien a los datos de entrada y predecir correctamente las etiquetas de clase de los registros que nunca antes ha visto. Por tanto, un objetivo clave del algoritmo de aprendizaje es construir modelos con buena capacidad de generalización; es decir, modelos que predicen con precisión las etiquetas de clase de registros desconocidos y descubrir patrones que sean válidos (precisos), interesantes (útiles y novedosos) y comprensibles [49].

Entre los clasificadores más conocidos están los árboles de decisión, clasificadores basados en reglas, redes neuronales, máquinas de vectores de soporte y clasificadores Bayes ingenuos. Cada técnica emplea un algoritmo de aprendizaje para identificar el modelo que mejor se ajusta a la relación entre el conjunto de atributos y la etiqueta de clase de los datos de entrada [49].

En este trabajo se aplicaron los modelos de clasificación: Bayes ingenuo (*Naive Bayes*), árboles de decisión y redes neuronales, los cuales se explicarán a continuación.

Bayes ingenuo: es un método de predicción basado en el aprendizaje de Bayes, en ellos se asume que las variables predictoras son independientes entre sí. El algoritmo básico frecuentemente utilizado es el método Bayes ingenuo. La clasificación mediante el algoritmo Bayesiano ofrece una forma de calcular la probabilidad a posteriori de que ocurra un cierto evento A, dadas algunas probabilidades de eventos anteriores [55].

Árboles de decisión (*decision trees*): En este caso se construye un árbol desde la raíz hasta las hojas, de tal manera que las ramas se dividen en función de los valores que toman los atributos. El árbol de decisión se construye partiendo el conjunto de datos en

dos o más subconjuntos de observaciones, después estos subconjuntos se vuelven a particionar empleando el mismo algoritmo. La raíz del árbol es el conjunto de datos inicial, los subconjuntos y sub subconjuntos conforman las ramas del árbol. El conjunto en el que se realiza una partición se llama nodo y permite bifurcar en función de los atributos y sus valores. Las hojas del árbol proporcionan predicciones. Entre todos ellos, uno frecuentemente utilizado es el J48 que es una mejora del árbol inicial C4.5 [49,55, 56].

Redes Neuronales Artificiales (ANN): Las redes neuronales artificiales son modelos diseñados para imitar el cerebro humano con el fin de emular el funcionamiento humano y suelen contener desde unos miles a unos pocos millones de unidades neuronales. Los neurólogos han descubierto que el cerebro humano aprende cambiando la fuerza de la conexión sináptica entre neuronas tras la estimulación repetida por el mismo impulso. De forma análoga a la estructura del cerebro humano, una ANN se compone de una estructura interconectada a través de nodos o neuronas y cada neurona está conectada con otras neuronas mediante enlaces de comunicación, cada uno de los cuales tiene asociado un peso. Los pesos representan la información que será usada por la red neuronal para resolver un problema determinado [55, 56].

Las ANN son sistemas adaptativos que aprenden de la experiencia, aprenden a llevar a cabo ciertas tareas mediante un proceso de entrenamiento. Mediante este entrenamiento o aprendizaje, las ANN crean su propia representación interna del problema, por tal motivo se dice que son autoorganizadas. Posteriormente, pueden responder adecuadamente cuando se les presentan situaciones a las que no habían sido expuestas anteriormente, es decir, las ANN son capaces de generalizar de casos anteriores a casos nuevos. En este trabajo, usamos un tipo de red neuronal conocida como los perceptrones multicapa.

Perceptrón multicapa: su aprendizaje se basa en el método de retropropagación, que consiste en que se parte de unos pesos iniciales en las conexiones interneuronales. Para un conjunto de entradas se obtiene una cierta salida. Basándose en que se conoce la salida que se debería haber obtenido (aprendizaje supervisado), se calcula el error. A partir de este error se modifican los pesos siguiendo el sentido inverso, es decir, se parte de la salida hasta llegar a la entrada. De la misma manera se opera con el resto de entradas de entrenamiento. Se puede observar que el error disminuye a medida que se aplica el

algoritmo. Sin embargo, un entrenamiento iterativo con las mismas entradas acaba provocando un sobre entrenamiento de la red neuronal, memorizando características de un conjunto impidiendo así que aprenda a generalizar. Por eso tras cada iteración es necesario realizar una evaluación de los resultados, introducir nuevos datos de entrenamientos y calcular el error de salida. De esta manera se obtiene una función (error de evaluación) de la que interesa hallar su mínimo absoluto, vale mencionar que la función de error puede tener mínimos locales. Determinando el número de iteraciones con que se espera alcanzar un bajo error para el conjunto de datos de entrada. Después, se puede aplicar una prueba (*test*) con un conjunto nuevo de entradas que no han sido presentadas a la red neuronal y que permitirá obtener una medida de la capacidad de discriminación de la red [55, 56].

Evaluación de los métodos de clasificación [49, 55, 56]:

Se realizan las etapas de entrenamiento y datos de prueba para medir la calidad de los modelos predictivos. De esta manera se asegura que las predicciones sean precisas y robustas como se sigue a continuación:

- Datos de entrenamiento (*training dataset*): conjunto de datos que se emplea para construir el modelo.
- Datos de prueba (*test dataset*): es el conjunto de datos que nos reservamos y se usa para comprobar si el modelo obtenido en la etapa de entrenamiento funciona. osea si las respuestas predichas por el modelo para un caso totalmente nuevo son acertadas o no.

Técnicas de evaluación [49].

- Validación simple: se divide aleatoriamente el conjunto de datos en dos subconjuntos: uno para el entrenamiento y otro para la validación. Con el primero se construye el modelo que luego es validado con los datos del segundo. El conjunto de validación suele constar del 5% al 50% de los datos iniciales.

- Validación cruzada (cross-validation): se divide aleatoriamente el conjunto de datos en dos subconjuntos. Se construye el modelo con el primer subconjunto y se valida con el segundo. Esta técnica es utilizada para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición entre datos de entrenamiento y prueba.
- Validación cruzada con k pliegues: similar a la validación cruzada, pero dividiendo el conjunto de datos iniciales en k subconjuntos. Se toma el primer subconjunto para validar y se construye el modelo con la unión del resto de subconjuntos. Se repite el paso anterior con cada uno de los (k-1) subconjuntos restantes. Finalmente, se construye el modelo con todos los datos.

Medidas de evaluación

La medida de evaluación empleada depende, entre otros factores, del contexto de aplicación o de la tarea a evaluar.

- Clasificación: normalmente, se emplea como medida la precisión, que se obtiene como la proporción de instancias clasificadas correctamente. No obstante, pueden considerarse otras medidas o estadísticos como los que se explican a continuación.
- Matriz de confusión: es una medida que permite a través de una tabla la visualización del desempeño de un algoritmo. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real. Uno de los beneficios de las matrices de confusión es que facilitan ver si el sistema está confundiendo dos clases. Por ejemplo, si en los datos de entrada el número de muestras de clases diferentes cambia mucho la tasa de error del clasificador no es representativa de lo bien que realiza la tarea el clasificador. Si por ejemplo hay 990 muestras de la clase 1 y sólo 10 de la clase 2, el clasificador puede tener fácilmente un sesgo hacia la clase 1. Si el clasificador clasifica todas las muestras como clase 1 su precisión será del 99%. Esto no significa que sea un buen clasificador, pues tuvo un 100% de error en la clasificación de las muestras de la clase 2.

Estadísticos [49]:

- Verdadero Positivo (VP): también conocida como Recall, sensibilidad, o tasa de verdaderos positivos. Es un resultado en el que el modelo predice correctamente la clase positiva. osea aquellos que verdaderamente pertenecen a esa clase.
- Verdadero Negativo (VN): es un resultado en el que el modelo predice *correctamente* la clase *negativa*
- Falso Positivo (FP): es un error en el resultado de una medición o cálculo, indicando la presencia de una condición o característica cuando la misma no está presente en realidad.
- Falso Negativo (FN): es la ausencia de una condición o característica en el resultado de una medición o cálculo cuando en realidad sí que está presente.
- Precisión: Esta métrica representa el número de verdaderos positivos que son realmente positivos en comparación con el número total de valores positivos predichos. ver ecuación 9

$$\frac{VP}{VP+FP} \quad \text{ecuación 9}$$

- Medida-F: es una medida que combina los valores de la precisión y *recall* y se utiliza para ver el compromiso entre ellas. Ver ecuación 10

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad \text{ecuación 10}$$

- Exactitud: se refiere a que tan cerca del valor real se encuentra el valor medido. La exactitud está relacionada con el sesgo de una estimación. Cuanto menor es el sesgo más exacto es una estimación. Cuando se expresa la exactitud de un resultado, se expresa mediante el error absoluto que es la diferencia entre el valor experimental y el valor verdadero.

- Índice Kappa: se trata de un índice de concordancia propuesto por Cohen que, en clasificación, mide la concordancia existente entre las frecuencias de ocurrencia de cada clase y las frecuencias obtenidas al usar el clasificador.
- Curva ROC (acrónimo de *Receiver Operating Characteristic*, en español, Característica Operativa del Receptor): es una representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario según se varía el umbral de discriminación.
- Error cuadrático medio (MSE, *Mean Squared Error*): cuanto mayor sea este valor, peor es el modelo. Nunca es negativo, ya que los errores de predicción individuales se elevan al cuadrado antes de sumarlos, pero sería cero para un modelo perfecto.
- Raíz cuadrada del error cuadrático medio (RMSE, *Root Mean Squared Error*): es solo la raíz cuadrada del MSE. Mide el error cuadrático medio entre el verdadero valor de la clase y el valor predicho.
- Error absoluto medio (MAE, *Mean Absolute Error*): mide el error absoluto medio entre el verdadero valor de la clase y el valor predicho.
- R al cuadrado (R^2): el coeficiente de determinación o determinación múltiple (en la regresión lineal múltiple). El R^2 indica la variación de la variable de respuesta que explica su relación con una o más variables predictoras

2.5 Cadenas de Márkov

Los modelos de Márkov son un sistema matemático que experimenta transiciones de un estado a otro de acuerdo a un conjunto de reglas probabilísticas. Las cadenas de Márkov hacen parte de los procesos estocásticos, este método se caracteriza porque la probabilidad de que ocurra un evento depende *solamente* del evento inmediatamente anterior. Esta característica de falta de memoria recibe el nombre de *propiedad de Márkov* [57].

Se considera que un conjunto de datos cumple las propiedades de Márkov si presenta las siguientes condiciones [57, 58, 59, 60]:

- a. Intuitivamente, se interpreta como que, dado el *presente* del proceso, el *futuro* es independiente del *pasado*. Es decir, el valor en el instante n -ésimo depende solamente de la última observación. Ecuación 11.

$$P(X_n = j \mid X_{n-1} = i) \quad \text{ecuación 11}$$

La probabilidad de que $X_n = j$ sólo depende del estado inmediatamente anterior del sistema: X_{n-1}

- b. La evolución de una cadena de Márkov viene descrita por sus probabilidades de transición. Se define como la probabilidad de que la cadena esté en el estado j después de n pasos, dado que la cadena empezó en el estado i .
- c. Diremos que una cadena de Márkov X es homogénea si dada una cadena de Márkov X finita, con m posibilidades de estados, se puede introducir la notación, ver ecuación 12:

$$P_{ij} = P(X_n = j \mid X_{n-1} = i) \quad \text{ecuación 12}$$

donde $i, j = 1, 2, \dots, m$, si $P_{ij} > 0$ entonces se dice que el estado E_i puede comunicarse con el estado E_j . La comunicación puede ser mutua si también $P_{ji} > 0$

La clasificación de las cadenas Márkov se realiza también según su temporalidad, el desarrollo de una cadena se realiza simulando el movimiento de una cadena que salta entre los estados de S en cada momento. También se tiene en cuenta el número (posiblemente infinito) de instantes que tarda la cadena en regresar a su punto de origen. Basado en lo anterior las cadenas de Márkov se clasifican teniendo en cuenta lo siguiente [57, 60]:

- a. **Espacios:** pueden ser discretos o continuos
- b. **Según la probabilidad de recurrencia los estados** pueden ser
- *Recurrente:* un estado *recurrente* significa la probabilidad que el proceso comenzando en el estado i , se tenga la certeza de volver a i en algún momento del tiempo en una determinada cantidad de pasos, si $F(i, i) = 1$.
 - *No recurrente o transiente* significa lo contrario a un estado recurrente es la probabilidad que el proceso comenzando en el estado i no se tenga la certeza de volver a i en un determinado número de pasos.
 - *Recurrente nulo:* Un estado recurrente se dice que es nulo, cuando el número de pasos necesarios para regresar a dicho estado es infinito.
 - *Transitorios:* Se decide que un estado es transitorio, si la probabilidad que, habiendo empezado en él, regrese a él es igual 1.
- c. **Cadenas de Márkov** pueden ser:
- *Absorbente:* una cadena de Márkov es absorbente cuando una vez que se accede a un estado, la probabilidad de seguir en ese estado es de 1. Un estado absorbente define una clase de estados por sí mismo.
 - *Irreducibles:* cuando en una cadena de Márkov todos sus estados son accesibles entre sí y por tanto se comunican, se dice que es irreducible, es decir, que existe una única clase de estados.
- d. **Periodicidad:** se refiere a la probabilidad que tiene un estado de volver a sí mismo después de haber pasado por un paseo aleatorio pasando por un número de estados. usualmente se clasifican en:
- *Periódico:* si partiendo de ese estado, sólo es posible volver a él en un número de pasos que sea múltiplo de un cierto número entero mayor que 1.
 - *Aperiódico:* Si un estado no es periódico se dice que es aperiódico.

2.6. Modelos Ocultos de Márkov (HMM)

Hay muchos escenarios reales que no se logran modelar mediante las cadenas de Márkov, porque las señales observadas corresponde a los estados que son visibles directamente para el observador, por lo que las probabilidades de transición entre estados son los únicos parámetros que pueden conocerse, pero cuando el estado no es visible para el observador, se puede recurrir a los modelos ocultos de Márkov (HMM, *Hidden Márkov Models*). En los HMM, la cadena de Márkov subyace tras las observaciones, y así al correlacionar las observaciones y las transiciones de estado lo que se busca es encontrar el estado de secuencias más probable [62, 63].

Un HMM puede ser entendido como una especie de doble proceso estocástico [62, 63, 64]:

- El primer proceso estocástico es un conjunto finito de estados, cada uno de ellos generalmente asociado a una distribución de probabilidad multidimensional. Mediante la denominada matriz de transición controla las transiciones entre estados.
- El segundo proceso estocástico es aquel en que cualquier estado puede ser observado, es decir, analizaremos lo observado sin ver en qué estado está ocurriendo, de aquí el epíteto *oculto* que define a este modelo.

Componentes de un HMM

Un HMM, se define por cinco elementos [62, 64, 65, 66]:

1. Los N estados del modelo $S = \{S_1, \dots, S_N\}$
2. Los M diferentes valores $V = \{V_1, \dots, V_M\}$ que pueden observarse en cada estado.

3. La matriz de transición $A = \{a_{ij}\}$, donde $a_{ij} = P(q_{t+1} = j | q_t = i)$, siendo Q_t el estado actual. Esta matriz es equivalente a la vista en la definición de las cadenas de Márkov.

4. La distribución de probabilidad de las observaciones en cada estado, $B = \{b_j(v_k)\}$,

$$B = \{b_j(v_k)\} = P(o_t = v_k | q_t = j), \text{ ecuación 13}$$

la probabilidad de observar v_k cuando se está en j en el instante t . La secuencia observable se denota como $O = (o_1, o_2, \dots, o_T)$.

5. La distribución inicial de estados $\pi = \{\pi_i\}$, donde π_i es la probabilidad de que el modelo está en el estado S_i en el tiempo inicial $t = 0$ [66].

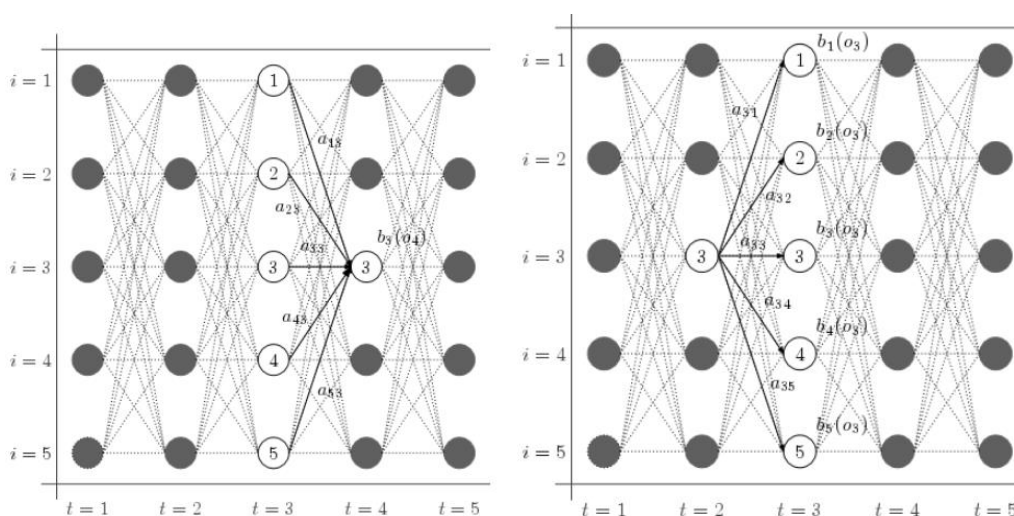
El objetivo de un HMM, es aprender a clasificar los datos proporcionados. Un HMM tiene que ser capaz de discernir las diferencias de comportamiento que poseen los diferentes estados a partir del flujo (*stream en inglés*) de observaciones con el que lo alimentamos.

En la historia de los HMM se han destacado tres problemas que se enuncian a continuación y los algoritmos para resolver estos problemas son muy conocidos, en nuestro caso usaremos la versión dada por Przemyslaw Dymarski.[66].

1. **Problema de Evaluación:** dada la secuencia de observaciones $O = \{o_1, \dots, o_m\}$, ¿Cuál es la probabilidad $P(O|\lambda)$ de que O haya sido generada por el modelo λ , dado un λ ? Se utiliza frecuentemente el algoritmo de avance (*Forward Algorithm en su sigla en inglés*).
2. **Problema de Decodificación:** dada la secuencia de observaciones $O = \{o_1, \dots, o_m\}$, ¿Cuál es la secuencia de estados más probable dado el modelo λ ? Se utiliza frecuente Algoritmos de Viterbi
3. **Problema de Aprendizaje:** dada la secuencia de observaciones $O = \{o_1, \dots, o_m\}$, ¿Cómo se deben ajustar los parámetros de λ para maximizar $P(O|\lambda)$? Se utiliza frecuentemente el algoritmo Baum-Welch [65, 66].

Algoritmos de avance-retroceso (*Forward-Backward Algorithms*): hacia adelante: es la probabilidad de un estado dadas las evidencias hasta ese estado. Hacia atrás: la probabilidad que se den las evidencias que hay hasta el estado definido k , dado el estado k , ver Figura 2-5.

Figura 2-5 Algoritmo de Avance Panel A; Algoritmo de Retroceso Panel B



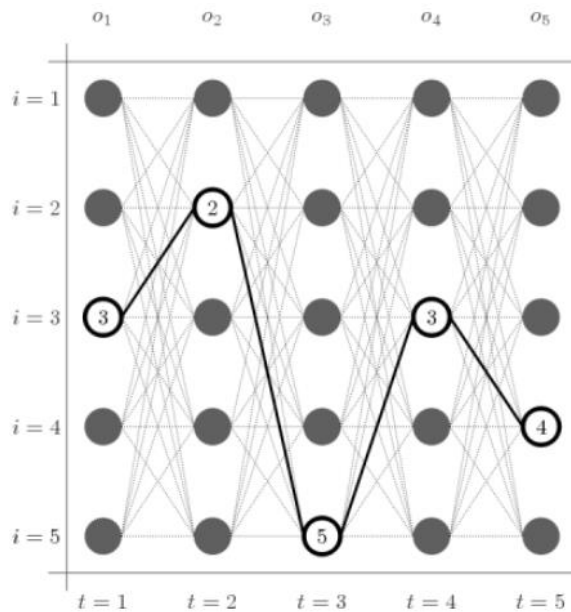
Panel A: Algoritmo de Avance
 Fuente: Tomado de [64]

Panel B: Algoritmo de Retroceso

Algoritmo de Viterbi: en la Figura 2-6 se ilustra gráficamente el Algoritmo de Viterbi, el cual se basa en dos principios [64, 65, 66]:

- Existe una relación recursiva entre los caminos más probables hasta X_t y los caminos más probables hasta X_{t+1}
- En el filtrado (algoritmo hacia adelante) se calcula la suma de los caminos, pero en el algoritmo de Viterbi se maximiza esta suma en la búsqueda del máximo probable, luego se sustituye la sumatoria por una función máxima probable. Esto nos permite encontrar el estado de secuencias que tiene mayor verosimilitud, ver Figura 2-6.

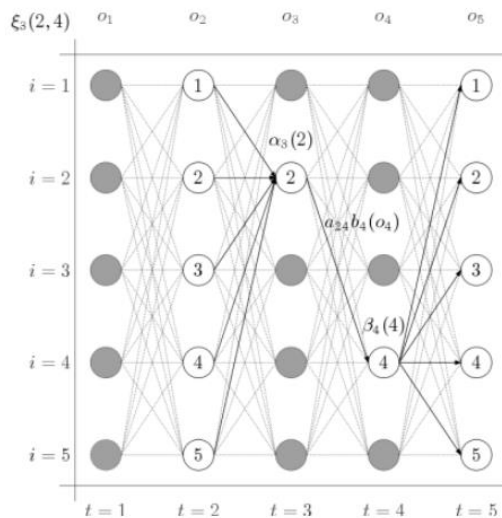
Figura 2-6 Algoritmo de Viterbi



Fuente: Tomado de [64]

Algoritmo de Baum-Welch: El segundo proceso estocástico es aquel en que cualquier estado puede ser observado, es decir, se analiza lo observado sin ver en qué estado está ocurriendo, el algoritmo de Baum-Welch también define nuevas variables a través del algoritmo avance-retroceso definido anteriormente. Dada una secuencia de salida o un conjunto de tales secuencias, se debe encontrar el conjunto de estados de transición y probabilidades de salida más probables. Primero determina un modelo inicial, que se puede seleccionar aleatoriamente, para facilitar la convergencia del algoritmo. Luego realiza el cálculo de transiciones y símbolos de emisión más probables del modelo inicial. Y, por último, se redefinen los parámetros a partir de lo calculado en el paso anterior para crear un nuevo modelo que mejore en verosimilitud al modelo inicial, ver Figura 2-7 [63, 64, 65,66].

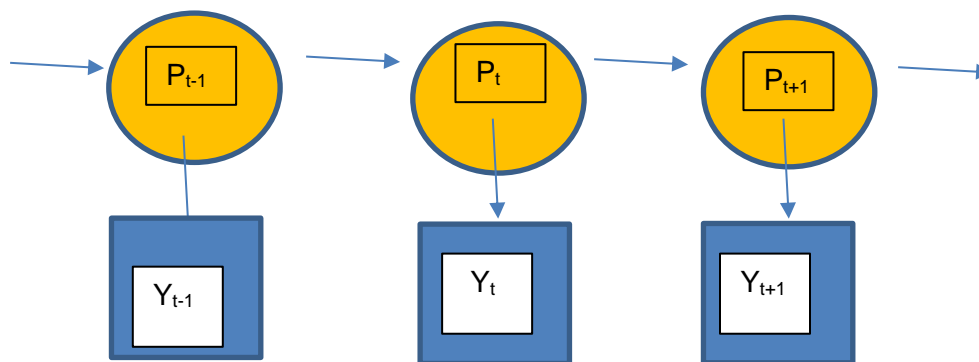
Figura 2-7. Algoritmo de Baum-Welch.



Fuente: Tomado de [64]

Con este algoritmo el objetivo es encontrar el modelo que mejor *explica* la secuencia observada. Ahora, se pueden construir HMM de 1er, 2do, 3er orden y así sucesivamente teniendo en cuenta lo ilustrado en la Figura 2-8.

Figura 2-8 Esquema del funcionamiento de un HMM.



Fuente: Autores

La Figura 2-8 se muestra un diagrama correspondiente a un HMM de primer orden, donde los nodos representan las variables aleatorias, en este ejemplo los círculos representan los estados ocultos y los cuadrados las observaciones. Las variables aleatorias pueden

tomar diferentes valores. La variable aleatoria $P(t)$ es el valor de la variable oculta en el mismo instante de tiempo t . La variable aleatoria $Y(t)$ es el valor de la variable observada en el mismo instante de tiempo. El valor de la variable oculta $P(t)$ (en el instante t) sólo depende del valor de la variable oculta $P(t-1)$ en el instante $(t-1)$. De forma similar, el valor de la variable observada $Y(t)$ sólo depende del valor de la variable oculta $X(t)$ (ambas en el instante (t)) [64, 65].

Entrenamiento del modelo HMM

El entrenamiento del modelo de un HMM consiste en ajustar los parámetros del modelo para maximizar la probabilidad de generación de una secuencia de observaciones dado el modelo. Se pueden elegir los parámetros del modelo de forma que se maximice localmente la probabilidad $P(O|X)$ mediante un procedimiento iterativo como el de Baum-Welch o equivalentemente el método Maximización de la Esperanza (*EM Expectation-Maximization en inglés*) o técnicas de gradiente [64,66].

Baum-Welch: Es el más utilizado, consiste en que a partir de un modelo $X = (n, A, B)$, utilizando estas fórmulas de re estimación, se obtiene un nuevo modelo $A' = (n_1, A', B')$, la probabilidad de generación de la secuencia de observaciones dado el modelo X' es siempre mayor que la obtenida para el modelo inicial X , excepto cuando se alcanza un valor crítico de la función probabilidad, en cuyo caso las dos probabilidades coinciden. Esta prueba garantiza una convergencia uniforme de este método de re estimación hacia este punto límite, llamado Estimación de Máxima Probabilidad [66].

EM: Alternativamente, se pueden obtener también las mismas relaciones maximizando directamente la función probabilidad, sujeta a las restricciones estocásticas de los parámetros, mediante un método tradicional como el de los multiplicadores de Lagrange que es un procedimiento para encontrar los máximos y mínimos de funciones de múltiples variables sujetas a restricciones

Técnicas clásicas de gradiente: Esta aproximación permite utilizar otros criterios de optimización como el MMI (o de Máxima Información Mutua), que se basa en maximizar la información mutua promedio entre el conjunto de secuencias de entrenamiento y el

conjunto de modelos a diseñar, cuando se desea diseñar conjuntamente una serie de modelos que se utilizarán con propósitos discriminatorios [66].

También se han utilizado para la estimación de los parámetros de los modelos de Márkov algoritmos como el llamado entrenamiento correctivo utilizados en los métodos de clasificación anteriormente descritos.

Insuficiencia en los datos de entrenamiento

En muchas situaciones no se dispone de una secuencia de observaciones de la suficiente duración como para poder estimar adecuadamente los parámetros del modelo. También suele ocurrir que existe una gran variabilidad entre las secuencias de observaciones que se desean hacer corresponder con el mismo modelo. En cualquiera de estos casos la naturaleza del proceso a modelar hace necesario utilizar varias secuencias de observaciones en la estimación de dichos parámetros. Por tanto, es necesario modificar las fórmulas de estimación, que están desarrolladas para una secuencia simple de observaciones, para tener en cuenta un conjunto de secuencias. Dado que estas fórmulas están basadas en términos de frecuencias de ocurrencia de determinados sucesos y suponiendo independencia estadística y equi-probabilidad en las diferentes secuencias de observaciones, se puede simplemente acumular los valores de dichas frecuencias para todas las secuencias de conjunto de entrenamiento [66].

La secuencia de observaciones usadas en el entrenamiento es necesariamente finita, lo cual da lugar a menudo a un número insuficiente de ocurrencias de los diferentes eventos del modelo para proporcionar buenas estimaciones de los parámetros, especialmente las probabilidades de observación. En el caso de los modelos discretos como en este trabajo, aunque los símbolos más frecuentes serán bien entrenados, si un símbolo no aparece nunca en un estado durante el entrenamiento se dará valor nulo a su probabilidad en la distribución correspondiente a dicho estado [66].

Se puede resolver este problema aumentando el conjunto de entrenamiento, reduciendo el número de parámetros del modelo.

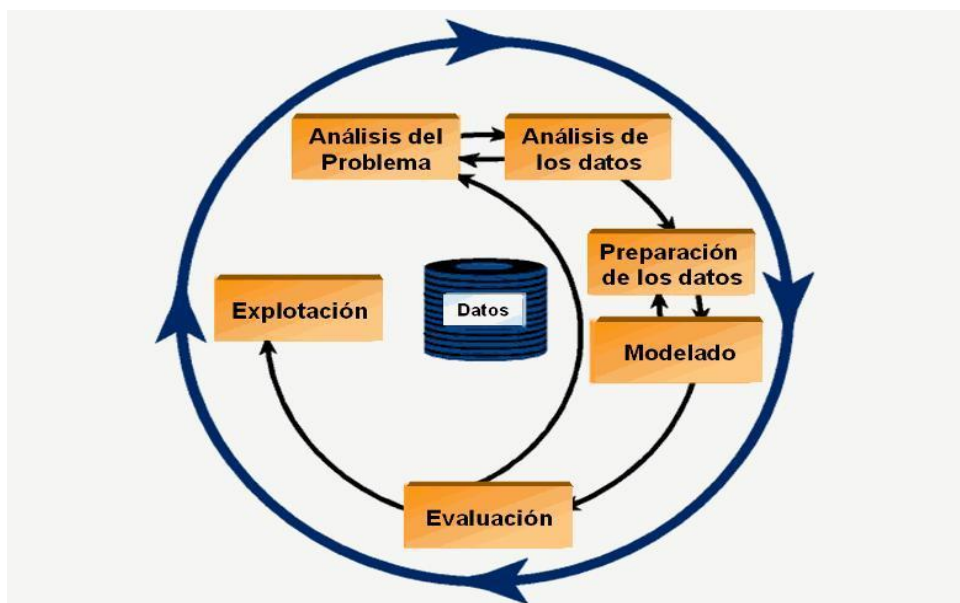
Validación del HMM

Función de verosimilitud: Es necesario identificar cual es la función de verosimilitud de T observaciones x_1, \dots, x_T suponiendo que son generadas por un HMM de m -estados. El objetivo es encontrar la probabilidad de observar la secuencia de observaciones, calculadas bajo el HMM encontrado, para una distribución inicial y una matriz de transición de probabilidad para la cadena de Márkov asociada. Se calcula a través del algoritmo de avance (forward algorithm). Su cálculo en forma recursiva juega un papel clave no solo en la estimación y evaluación de la función de máxima verosimilitud, también en el pronóstico y validación del modelo [64, 65, 66].

● Capítulo 3. Metodología

Para el desarrollo de este estudio, se utilizó como guía la metodología CRISP-MD, siendo una guía de referencia ampliamente utilizada para el desarrollo de proyectos de minería de datos. La metodología CRISP-MD está estructurada en 4 niveles de abstracción organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos y organiza el desarrollo de un proyecto de minería de datos en una secuencia de seis fases. Esta sucesión de fases no es necesariamente rígida y las tareas generales se proyectan a tareas específicas (ver Figura 3-1). Así pues, las fases de esta metodología son: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación del modelo. Aplicando esta metodología CRISP-MD [67, 68] (ver Figura 3-1), en este trabajo se llevaron a cabo las fases como se detalla a continuación.

Figura 3-1 Esquema de la metodología CRISP-MD



Fuente: Modificado y adaptado de Ordoñez, Yoanni & Horacio, Darian & grass boada, Darian. (2011). HERMINWEB: Herramienta de Minería de Uso de la Web Aplicado a los Registros del Proxy

3.1. Comprensión del estudio:

El objetivo principal de este estudio es proponer una transformación del espacio de signos vitales para detectar el deterioro de los pacientes de una unidad de cuidados intensivos pediátricos.

Objetivos Secundarios

- Caracterizar una muestra de pacientes en estado crítico en la unidad de cuidados intensivos pediátricos
- Construir modelos que permitan seleccionar signos significativos que se correlacionen con diferentes eventos clínicos
- Validar método propuesto con el diagnóstico proporcionado por los expertos

3.2. Comprensión de los datos:

Se complemento esta etapa con la metodología por objetivos:

3.2.1. Caracterización de una muestra de pacientes en estado crítico en la unidad de cuidados intensivos pediátrico

Se realizó un estudio de cohorte transversal, durante dos años (enero 2018 a enero 2020). Se ejecutó una descarga y almacenamiento de la información de signos vitales de los pacientes menores de 18 años que se encontraban hospitalizados en la UCIP del Hospital Militar Central de Bogotá, Colombia, previa autorización por parte del comité de ética en investigación institucional, acta número 013-164-19, recopilando los datos que arrojaban los monitores, discriminando cuando los signos vitales correspondían a eventos clínicos normales y cuando eran anormales. Para esto, solo se consideraron los registros que contenían al menos 1 día de datos de tendencias numéricas de seis bio-señales de signos vitales para la detección de anomalías: TAS, TAD, TAM, FC, FR y SpO₂. Todas las muestras se tomaron por un muestreo segundo a segundo. Los datos con valores perdidos consecutivos durante un largo período fueron filtrados y eliminados.

Se realizó la caracterización demográfica de los pacientes seleccionados, recopilando datos epidemiológicos y demográficos como son: edad en meses, peso en kilogramos (kg), talla en centímetros (cm), diagnóstico de ingreso, complicaciones durante el tiempo de monitorización morbilidad y mortalidad. Con esta información se clasificaron los pacientes por categorías etarias porque los valores de los signos vitales son específicos del paciente, así como y su normalidad, además y el rango varía según la edad, el sexo y la condición médica.

Los rangos esperados de signos vitales son específicos para cada paciente, aunque hay algunas reglas médicas que definen el límite normal y anormal de diferentes signos vitales, de acuerdo a cada situación clínica del paciente [13]. En este trabajo sólo se utilizaron los rangos normales generales de signos vitales definidos por la literatura [13] y, la experiencia del clínico para definir el grado de anormalidad de la normalidad, de acuerdo a la situación clínica individual de cada paciente. Se definieron como “eventos clínicos” cuando los seis parámetros vitales están por encima o por debajo de los rangos normales generales, previamente estipulados con el clínico experto, y solo si persistían por un período de tiempo fijo (más de 3 segundos a 5 minutos), que será la duración de la ventana de observación.

Se identificó que los datos recopilados tenían entonces 7 atributos: valorados para periodos de tiempo en segundos, FC, FR SpO2, TAS, TAD y TAM. Vale anotar que se trata de una base de datos cuantitativa que correspondía a un conjunto de series de tiempo, como la mostrada en la Tabla 3-1.

Tabla 3-1 Descripción de los 6 bioseñales.

hora	FC	SpO2	FR	TAS	TAD	TAM
18:52	138	92	36	100	56	73
18:53	142	92	31	99	53	69
18:54	139	92	31	99	52	68
18:55	142	92	29	97	51	66
18:56	144	92	25	102	56	71
18:57	145	94	34	105	57	73
18:58	142	94	30	108	60	76
18:59	142	93	32	108	59	75
19:00	142	91	34	97	50	66
19:01	147	92	30	101	56	71
19:02	137	93	32	103	55	71
19:03	138	93	29	94	50	64
19:04	144	91	30	104	60	75
19:05	141	94	32	103	57	72
19:06	141	91	33	95	50	65
19:07	141	92	28	95	52	66
19:08	140	92	29	95	51	66
.....

FC: frecuencia cardíaca, SPO2: saturación de oxígeno; FR: frecuencia respiratoria; TAS: tensión arterial sistólica; TAD: tensión arterial; TAM: tensión arterial media

3.3. Preparación de los datos

Se realizó una estadística descriptiva utilizando estimaciones de proporciones y frecuencias relativas para las variables categóricas y, para las variables continuas se utilizaron medidas de tendencia central y de dispersión (mediana, media, moda, desviación estándar). Se realizó una prueba de normalidad de Kolmogórov-Smirnov (K-S) para cada uno de los pacientes, la descripción demográfica de la población se muestra en la tabla 3-2 y para la totalidad de los datos se observó que la distribución de los mismos fue normal en variables como TAS TAD, TAM, y en variables como FR y FC tenía distribuciones diferentes. La significancia, de mayor a menor valor, se presentó en el siguiente orden: TAS, TAM, TAD, SaO2, FC, siendo última la FR, lo que indicaba que la distribución de los datos es normal en variables como TAS TAD, TAM, y en variables como FR y FC no lo era.

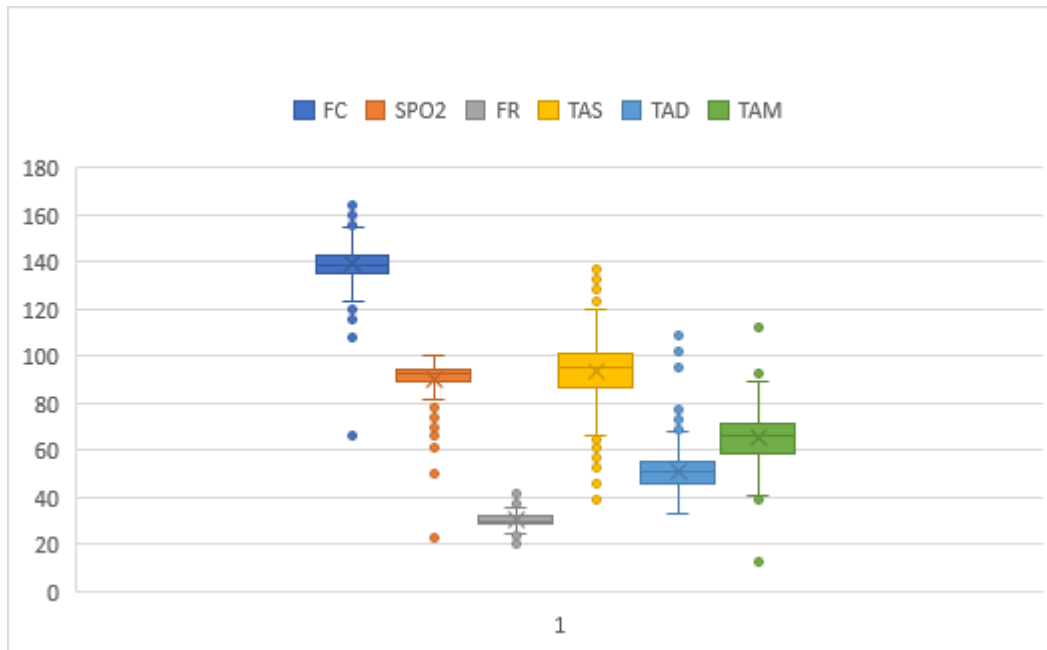
Tabla 3-2 Características demográficas de la población

Total, población N=90	sobrevivieron 65 (72,2%)	No sobrevivieron 25 (27,7%)
Edad (meses)	2 - 204	4 - 144
Peso (Kg)	4 a 65	5 a 40
Talla (cms)	45 a 160	55 a 145
Género		
Hombre	34	20
Mujer	31	5
Diagnóstico		
Oncológicos	5	4
SDRA	0	2
bronquiolitis	8	4
HTP-neumonía	0	5
neumonía	11	6
DBP	0	2
sepsis abdominal	15	0
cardiomiopatías descompensadas	13	0
poli traumatismo	3	0
Otros	10	2

Kg: kilogramos; cms: centímetros; SDRA: síndrome de dificultad respiratoria aguda; DBP: Displasia broncopulmonar; HTP: hipertensión pulmonar

En la Figura 3-2 se muestra un ejemplo del gráfico de cajas de un paciente, en donde se observó que, para la FC, TAD y FR los datos tienden a estar centralizados, no así los datos de las variables como SpO₂, TAS y TAM, que tienden a estar más concentrados hacia el percentil 75. El análisis descriptivo anterior realizado para cada paciente sirvió para observar la presencia de valores atípicos en todas las variables. Al analizarlos en la fase de preprocesamiento, se observó que correspondían a los deterioros presentados por los pacientes, o sea a los *eventos clínicos*. Después del preprocesamiento se obtuvieron datos para 90 pacientes, y en total 94,678 instancias. Este conjunto de datos se puede consultar en el siguiente dominio (consult database critically ill pediatric patients in PICU.csv. " Kaggle, doi: 10.34740 / KAGGLE / DSV / 1518232).

Figura 3-2 Diagrama de cajas que representan la distribución de las señales de signos vitales para un paciente



Fuente: los autores.

En el análisis relacional de los datos se realizó una comparación entre variables continuas mediante la prueba t de Student para muestras independientes y/o la U de Mann-Whitney y K-S, de acuerdo a la naturaleza de las variables.

En cuanto al análisis explicativo, las variables que resultaron significativas en el análisis bivariado se ingresaron en un modelo de predicción. Para la comparación de las variables de resultado de naturaleza cuantitativa entre los grupos, se utilizó el análisis de varianza a través de un análisis ANOVA de un factor. En este caso, se verificó el cumplimiento de los supuestos inherentes al uso de dicho estadístico de prueba: independencia de las observaciones (que es común de demostrar), homocedasticidad (verificada mediante la prueba de Bartlett) y normalidad (verificada por la prueba de Shapiro-Wilk). En caso de no verificarse el cumplimiento de los supuestos descritos anteriormente, se utilizó el análogo no paramétrico respectivo. En variables cualitativas, se utilizó una prueba Z de diferencia proporcional y en politómicas una prueba χ^2 de Pearson cuando los valores esperados en cada celda eran mayores o iguales a 0,05; de lo contrario, se utilizó una prueba exacta de Fisher.

3.4. Modelado

3.4.1. Transformación del espacio de signos vitales para identificar correlaciones significativas entre los signos vitales y los eventos clínicos

La limpieza y consolidación de los datos, y la aplicación de pruebas paramétricas y no paramétricas para conocer su distribución, amplía el conocimiento sobre el comportamiento de cada variable y su relación con los eventos clínicos. Desde este punto se llevó a cabo la selección y aplicación de técnicas de modelamiento y análisis predictivo, con el propósito de construir una métrica que permita seleccionar signos vitales significativos que se correlacionen con diferentes eventos clínicos, y conocer un conjunto de patrones característicos en dicho espacio que también puedan correlacionarse y a su vez interpretarse.

Las variables continuas se analizaron en su naturaleza y, según los resultados, en casos detallados, se construyeron variables categóricas, que permitieron ampliar el conocimiento sobre el comportamiento predictivo de los datos a través de agrupaciones. Las técnicas de minería de datos empleadas para la fase de modelamiento (siguiendo la metodología CRISP-MD) [67,68] se describen a continuación:

- **Análisis de componentes principales:** se empleó esta técnica como parte de un análisis exploratorio, y una reducción en la dimensión de los datos.
- **Análisis de series de tiempo:** la principal motivación fue obtener el espectro de una señal dada que revelara el contenido frecuencial de la misma, para luego aplicar modelos autorregresivos que permitan la identificación de patrones predictivos en los seis bioseñales.

- **Análisis de agrupación de los datos:** se realizó esta técnica con la intención de encontrar agrupaciones según los criterios de distancias, que permitan distinguir características comunes de cada grupo y relacionarla con su clasificación original.
- **Análisis de clasificación:** el propósito de la clasificación es categorizar las variables para determinar comportamientos clínicos frecuentes y sucesivos, que faciliten la toma de decisiones frente a los eventos clínicos de interés.
- **Análisis basados en cadenas de Márkov:** se utilizaron estos modelos probabilísticos para predecir la evolución y el comportamiento a corto plazo de los 6 bioseñales.
- **Análisis basado en Modelos Ocultos de Márkov** se empleó este modelo estadístico con la intención de determinar a través de un proceso de Márkov los parámetros desconocidos de dicha cadena a partir de los parámetros observables en los 6 bioseñales.

A continuación, se explican los detalles de cada uno de los tipos de análisis mencionados anteriormente.

3.4.1.1. Análisis de Componentes Principales (ACP)

Se realizó una reducción de dimensiones basado en el método de extracción de componentes, conocido como Análisis de Componente Principales (ACP). Para la construcción de este modelo se trabajó con las variables continuas originales, tratando de hallar las variables fuertemente relacionadas entre sí, que posteriormente formarán las componentes principales. La idea fue reducir las variables originales a un conjunto de variables que pudieran explicar la mayor proporción de la variabilidad entre ellos. Se trabajó paciente por paciente, y toda la serie de datos combinando la información de todos los pacientes. Al realizar el análisis de todos los datos, el resultado de los componentes obtenidos se consideró no satisfactorio porque sólo explicaba el 40% de la variabilidad entre datos y se prefirió trabajar paciente por paciente. A continuación, se realiza una demostración de los procesos realizados con cada uno de los pacientes, mostrando como ejemplo el proceso que se llevó a cabo con un paciente.

En la tabla 3-3, se muestra la aplicación de la prueba Kaiser-Meyer-Olkin que indicaba un valor de 0.686, y la prueba Bartlett's Test of Sphericity cuyo valor fue menor a 0,05, lo que indicaba que se rechaza la H_0 , indicando que hay diferencias entre los datos, y por lo tanto resulta conveniente aplicar la técnica ACP. era adecuada.

Tabla 3-3 Test Bartlett's y KMO para un paciente

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,686
Bartlett's Test of Sphericity	Approx. Chi-Square	5189,079
	df	15
	Sig.	0,000

KMO: prueba Kaiser-Meyer-Olkin

Posteriormente, se realizaron las tablas de comunalidades, extracción y, finalmente, rotación. Al aplicar el método de extracción como se observa en la tabla 3-4 en el cuadro de la varianza total explicada de cada componente, se indica cuáles son las componentes que han sido extraídas (aquellas cuyos autovalores superan la unidad), entre las dos primeras componentes extraídas se acumula el 66,59% de la variabilidad de las variables originales.

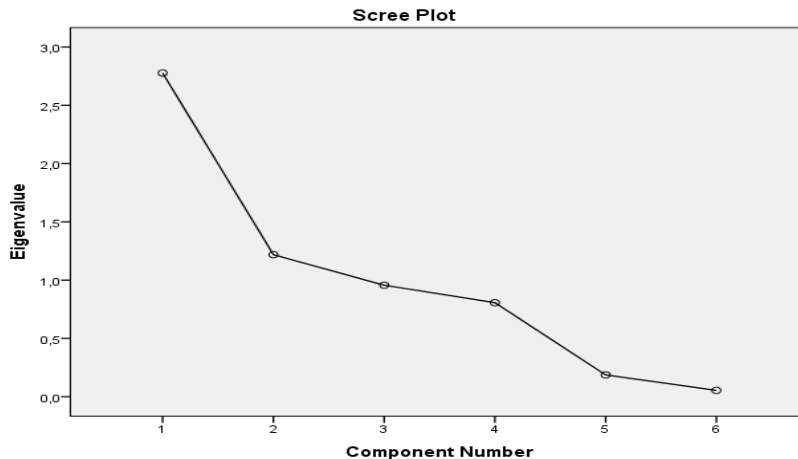
Tabla 3-4. Resultado del Análisis de Componentes Principales, varianza total acumulada para los datos de un paciente.

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,777	46,283	46,283	2,777	46,283	46,283	2,737	45,615	45,615
2	1,219	20,309	66,592	1,219	20,309	66,592	1,259	20,977	66,592
3	,957	15,943	82,535						
4	,807	13,443	95,978						
5	,187	3,118	99,096						
6	,054	,904	100,000						

Extraction Method: Principal Component Analysis.

Luego en el gráfico de sedimentación (herramienta gráfica para la decisión del número de componentes que hay que seleccionar) ver figura 3-3, se visualiza que la selección de dos primeras componentes parece ser adecuada, pues a partir de la tercera componente no es muy acusada la pendiente de la representación gráfica de los autovalores.

Figura 3-3. Gráfico de sedimentación número de componentes vs. valores propios.

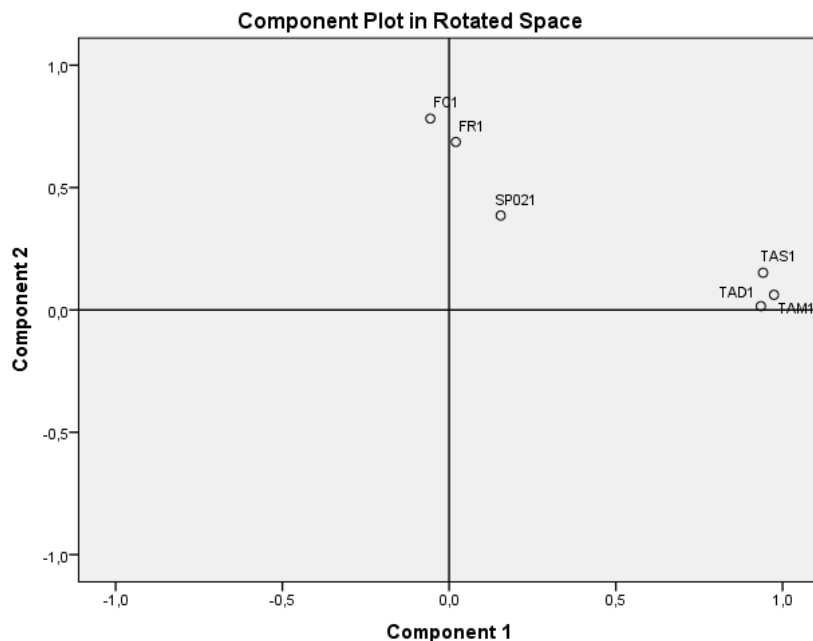


Fuente: los autores

En el gráfico de saturaciones (gráfico de componentes principales), como se muestra en la figura 3-4, se visualiza la representación gráfica de la matriz de componentes analizados. De la representación se extrae la explicación que el primer componente lo conforman las variables TAS, TAD, TAM, que se opone a las variables FC, FR, y SpO2. El segundo componente está conformado por FC, FR, y SpO2. El componente 1 es más positivo, porque se alejaban de 0, por lo tanto, era el componente que mejor explica el modelo, porque el segundo componente se ubicaba casi en el origen como se observa en la Figura 3-4.

Del análisis anterior se concluye que los resultados no fueron satisfactorios porque el modelo con dos componentes sólo explicaba la variabilidad de los datos en un 66,5%. Por esta razón, se analizaron varias ventanas para evaluar en donde se explicaba mejor la variabilidad en los datos y con cuántos componentes.

Figura 3-4. Gráfico de saturaciones, análisis de componentes principales para un paciente.



La ventana más explicativa se fijó media hora antes que ocurriera el evento clínico; ésta mejoró la explicación de la variabilidad entre las variables, ya que con las tres primeras componentes extraídas acumulo el 78,94% de la variabilidad de las variables originales, como se observa en la tabla 3-5 correspondiente a la varianza total explicada de cada componente.

Tabla 3-5 Varianza total explicada media hora antes del evento clínico para un paciente

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,631	43,856	43,856	2,631	43,856	43,856	2,535	42,243	42,243
2	1,075	17,912	61,768	1,075	17,912	61,768	1,134	18,907	61,150
3	1,033	17,216	78,984	1,033	17,216	78,984	1,070	17,834	78,984
4	,880	14,661	93,645						
5	,345	5,758	99,403						
6	,036	,597	100,000						

Extraction Method: Principal Component Analysis.

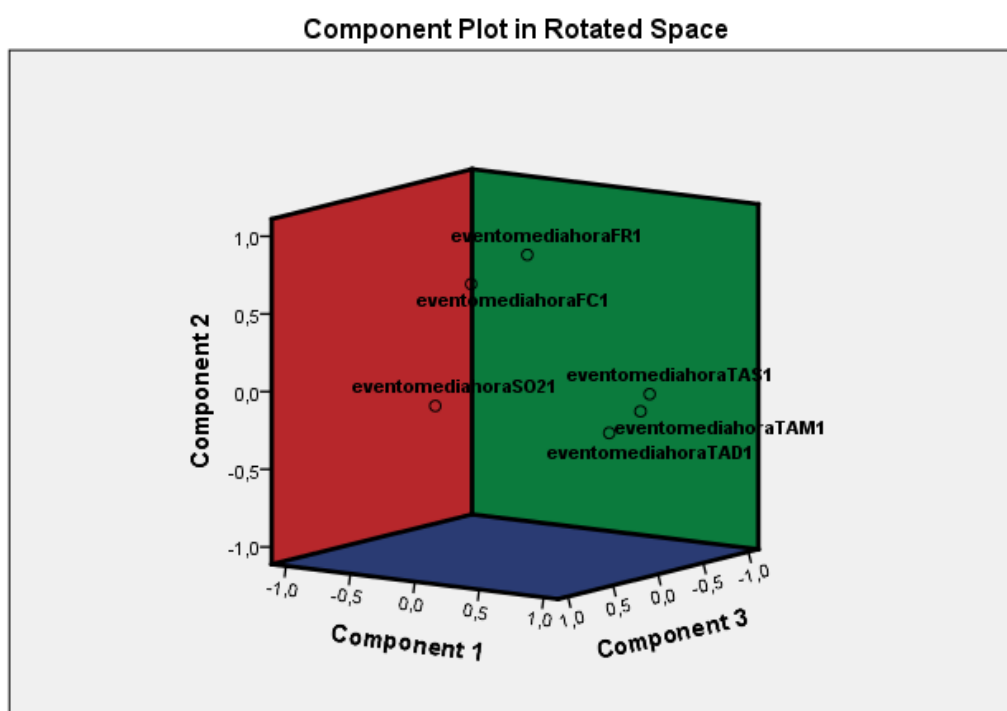
En la Tabla 3-6 se observa la matriz de Componentes o matriz de cargas o saturaciones factoriales, indica la carga de cada variable en cada factor, de forma que los factores con pesos factoriales más elevados en términos absolutos indican una relación estrecha con las variables. Además, al observar estas puntuaciones se afirma que las variables quedan muy bien explicadas a través de las componentes extraídas. siendo las variables TAS, TAD, TAM las que tienen los coeficientes más altos, cercanos a 1 y forman la primera componente.

Tabla 3-6 Matriz de Componentes rotados media hora antes del evento clínico			
	Component		
	1	2	3
evento anormal 1/2 hora antes TAM1	,986		
evento anormal 1/2 hora antes TAS 1	,899		-,209
evento anormal 1/2 hora antes TAD1	,860	-,201	,184
evento anormal 1/2 hora antes FR1		,783	-,276
evento anormal 1/2 hora antes FC1		,690	,371
evento anormal 1/2 hora antes SO21			,882
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. a. Rotation converged in 4 iterations. comunalidades			

también se observa que las variables TAD, FR, FC son las que más contribuyen con sus ponderaciones en la construcción de la segunda componente, y en la tercera componente las variables con las ponderaciones más altas que contribuyen a formar esta componente son las variables TAS, TAD, FR, FC, SO2. En la Figura 3-5 se ve esquematizada la importancia de cada una de las componentes, bajo una representación tridimensional de las saturaciones factoriales para las tres primeras componentes.

El ACP anteriormente realizado fue útil para entender el conjunto de datos, evidenciar que hay variables dentro del conjunto de datos que pueden explicar la variabilidad entre ellas, reduciendo la dimensionalidad de los datos.

Figura 3-5. Componentes en el espacio rotado en la ventana media hora antes del evento clínico para un paciente.



Fuente: Los autores

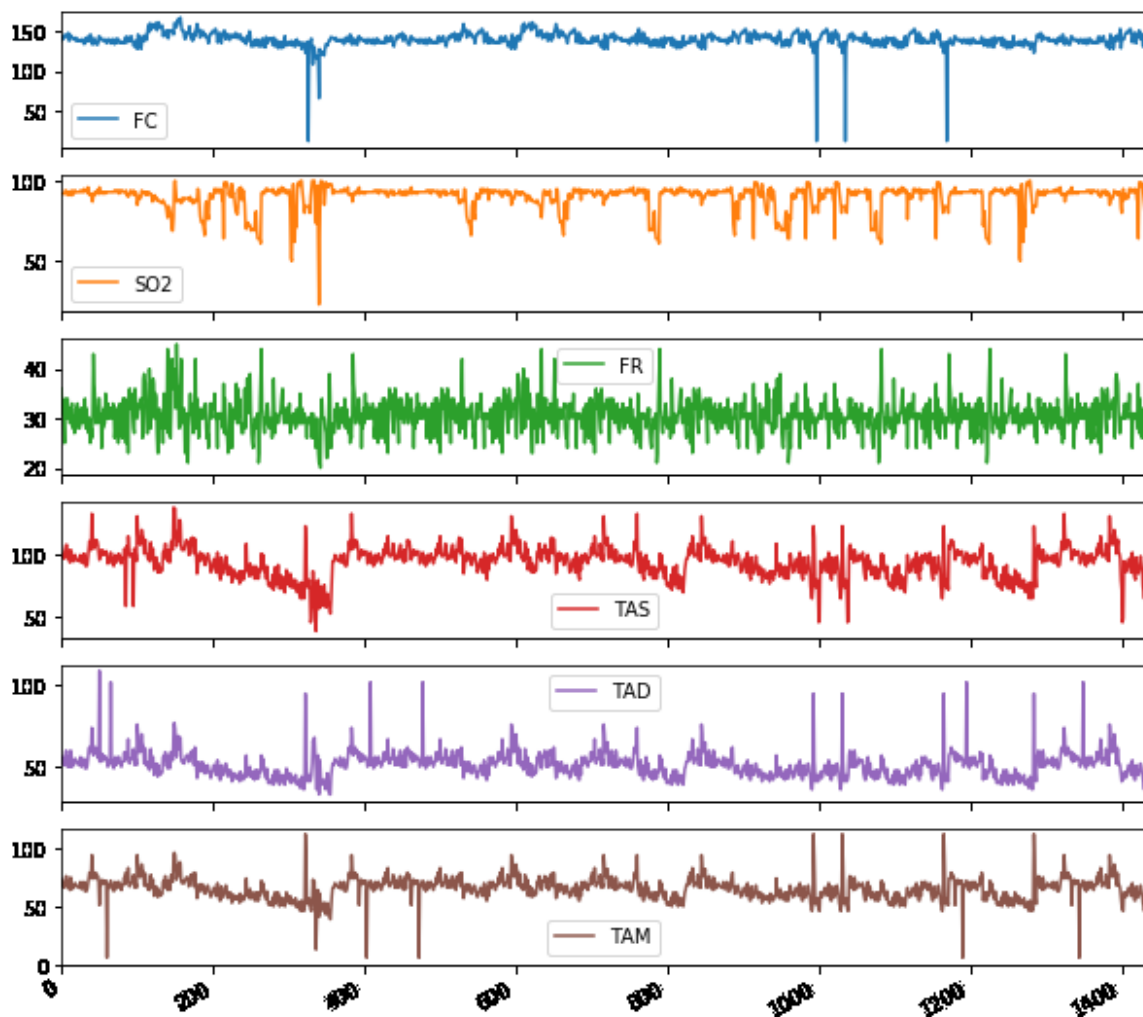
Sin embargo, al trabajar solamente con variables cuantitativas que en ocasiones tienen valores en cero, que en este conjunto de datos caracterizan a eventos clínicos de pacientes críticamente enfermos, el procesamiento de datos utilizando esta técnica de PCA se trunca. Por lo anterior se debe identificar la ventana adecuada para cada caso, resultando un proceso altamente dispendioso.

Por lo anterior se continuaron explorando otras técnicas de minería de datos para lograr buscar el modelo que mejor pudiera predecir el comportamiento basado en los datos.

3.4.1.2. Análisis de series de tiempo

Para evaluar si los resultados futuros estaban influenciados por el comportamiento pasado [46], se tomaron las bioseñales descritas en la Figura 3-6, que muestra el comportamiento de un paciente; cada señal es una variable individualmente considerada como serie de tiempo y se estudia con técnicas de análisis espectral, que permiten identificar las ventanas significativas dentro de las señales, evaluar su comportamiento y analizar cada uno de sus componentes como estacionariedad, tendencia, aleatoriedad y estacionalidad.

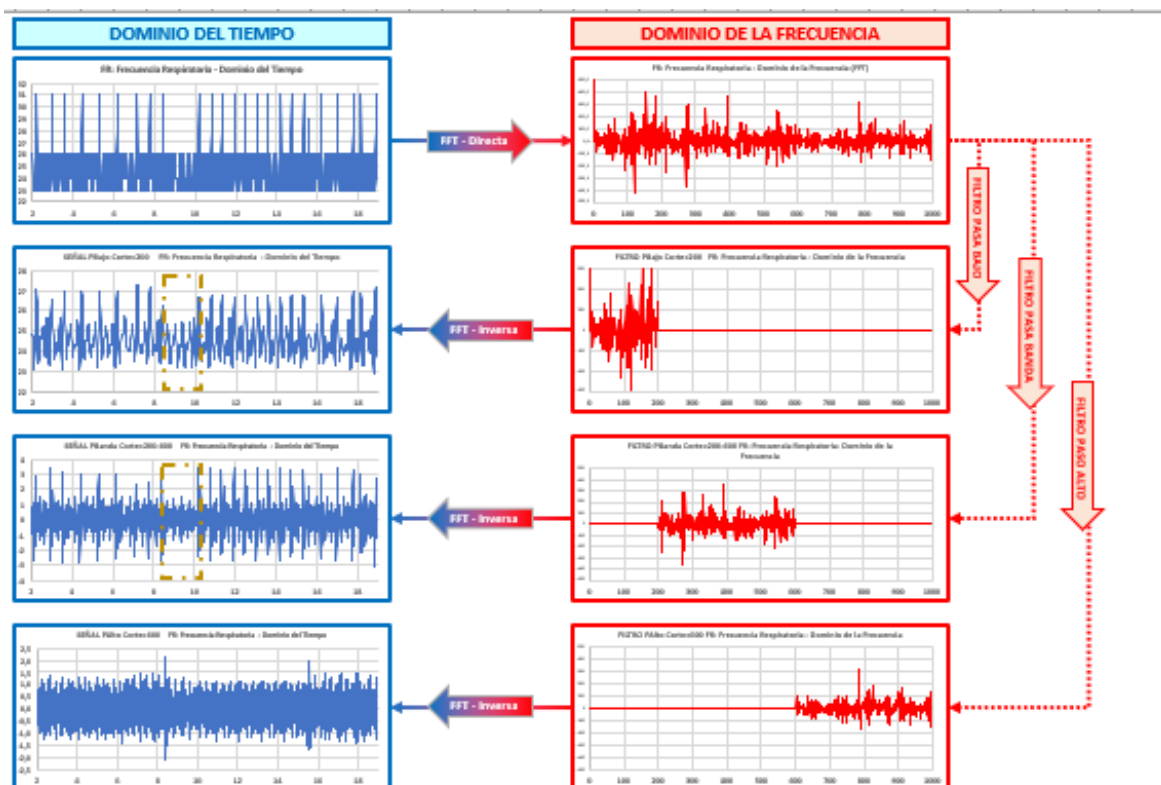
Figura 3-6. Comportamiento de los bioseñales para un paciente



Fuente: los autores

Para realizar el análisis espectral (ver Figura 3-7) se muestra el proceso realizado con los 6 bioseñales de un paciente; se usó la Transformada de Fourier (TF) y posteriormente se aplicaron filtros pasa bajo para frecuencias bajas menores a 200 Hz, filtro pasa alto para frecuencias altas mayores a 2000 Hz; y filtros pasa banda, para frecuencias que se encuentran dentro de unos límites superior e inferior particulares.

Figura 3-7. Análisis espectral. Transformada de Fourier para un paciente.



Fuente: Los autores

La TF determinó puntos de variación en el dominio del tiempo, para que dichas posiciones puedan ser ampliadas y analizadas en el dominio de la frecuencia como se muestra el proceso la Figura 3-7. En tales espacios de frecuencia también se aplicaron los modelos autorregresivos que podían predecir el comportamiento de la serie de tiempo, explicados más adelante.

En las series de tiempo, observadas en cada paciente y para cada variable, también se realizó una clasificación descriptiva de la serie, para determinar cuáles eran estacionarias y cuáles no. Para tal fin se evaluaron diferentes ventanas en la mayoría de los datos, se aplicaron diferentes métodos de suavizamiento para evaluar pronóstico, teniendo en cuenta si la serie tenía un patrón estacionario se aplicaron técnicas como: móvil simple, suavización exponencial simple; si la serie tenía una tendencia se aplicó una regresión lineal simple, suavización exponencial doble y si la serie tenía una tendencia y demanda estacional: se aplicó la técnica de suavización exponencial triple. todas las técnicas de suavizamiento anteriormente mencionadas, fueron aplicadas con el software SPSS versión 23, y el software R.

Por otro lado, se aplicaron modelos autorregresivos ARIMA. Tomando una de las ventanas seleccionadas en la TF, el resultado para la serie de tiempo al aplicar los modelos ARIMA para un paciente, como se muestra en la Tabla 3-7, se puede apreciar que los parámetros de los modelos ARIMA para cada una de las variables son diversos y se observa los ajustes que realiza este modelo para llevarlas a la estacionariedad.

Tabla 3-7. Modelo ARIMA para un paciente.

Modelo Descripción			
Model ID	Descripción		Modelo Tipo
	FC1	Modelo 1	ARIMA (1,1,9)
	porcentaje de oxigenación de la sangre	Modelo 2	ARIMA (1,0,5)
	FR1	Modelo 3	ARIMA (2,0,8)
	TAS 1 tensión sistólica	Modelo 4	ARIMA (0,1,10)
	TAD1 presión diastólica arterial	Modelo 5	ARIMA (0,1,9)
	TAM1 la presión de perfusión de los órganos corporales	Modelo 6	ARIMA (2,0,14)

FC: frecuencia cardíaca, FR: frecuencia respiratoria, TAS: tensión arterial sistólica, TAD: tensión arterial diastólica, TAM: tensión arterial media

Luego al aplicar el supuesto estadístico de Ljung-Box Q a este modelo ARIMA, como lo muestra la Tabla 3-8, se observan variables con una significancia menor a 0,05 como lo fueron las variables FC FR, SO2 en las ventanas media hora y toda la serie antes del

evento clínico, y las otras variables como TAS, TAD, TAM obtuvieron un nivel de significancia en todas las ventanas consideradas mayor a 0,05.

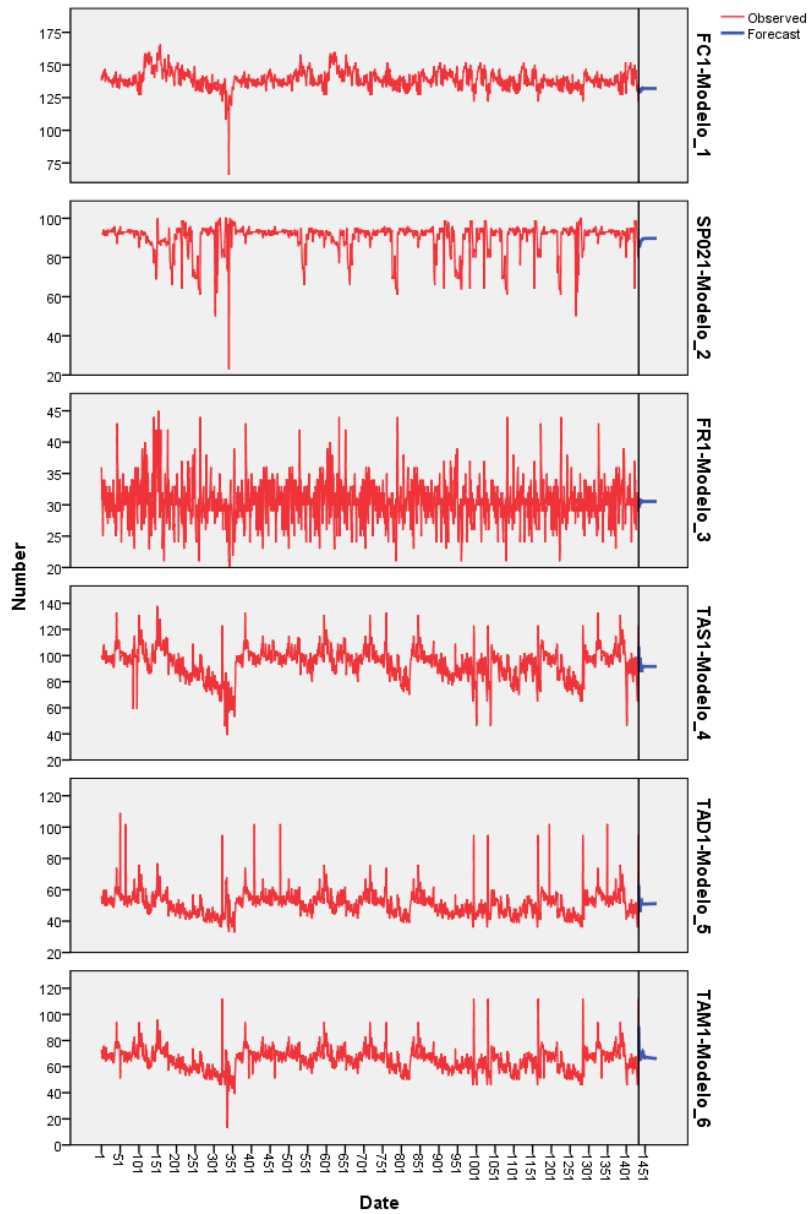
Tabla 3-8. Modelo estadístico ARIMA para un paciente.

Model Statistics					
Model	Number of Predictors	Ljung-Box Q(18)			Number of Outliers
		Statistics	DF	Sig.	
evento anormal 1 hora antesFC1-Modelo_1	0	15.286	15	.431	0
evento anormal 1 hora antesFR1-Modelo_2	0	15.298	15	.430	0
evento anormal 1 hora antes SO21-Modelo_3	0	7.313	15	.948	0
evento anormal 1 hora antes TAS1-Modelo_4	0	15.142	15	.441	0
evento anormal 1 hora antes TAD1-Modelo_5	0	14.808	15	.465	0
evento anormal 1/2 hora antesFC1-Modelo_6	0	36.921	16	.002	0
evento anormal 1/2 hora antesFR1-Modelo_7	0	29.159	16	.023	0
evento anormal 1/2 hora antes SO21-Modelo_8	0	8.104	17	.964	0
evento anormal 1/2 hora antes TAS1-Modelo_9	0	19.311	17	.311	0
evento anormal 1/2 hora antes TAD1-Modelo_10	0	16.458	16	.421	0
Evento antes evento antes del evento FC1-Modelo_11	0	41.367	16	.000	0
Evento antes evento antes del evento FR1-Modelo_12	0	29.631	16	.020	0
Evento antes evento antes del evento SO21-Modelo_13	0	31.466	16	.012	0
Evento antes evento antes del evento TAS1-Modelo_14	0	46.483	16	.000	0
Evento antes evento antes del evento TAD1-Modelo_15	0	52.799	16	.000	0

FC: frecuencia cardíaca, FR: frecuencia respiratoria, TAS: tensión arterial sistólica, TAD: tensión arterial diastólica, TAM: tensión arterial media

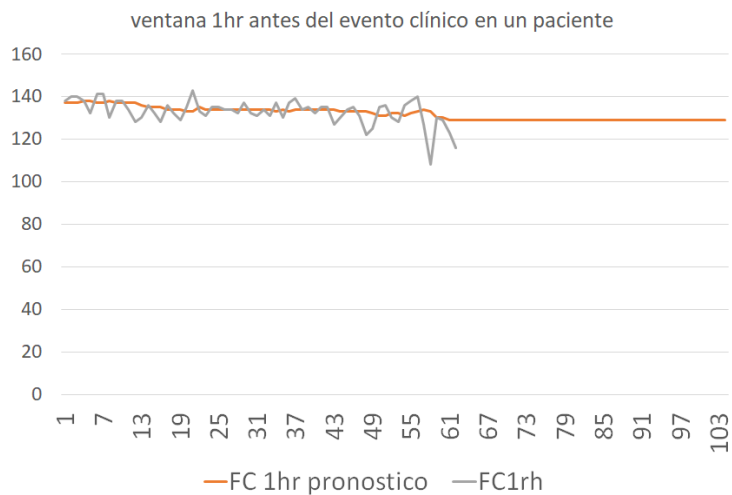
Es difícil con este modelo predecir el comportamiento en cada variable, para cada paciente y en cada ventana de tiempo seleccionada. Al realizar el pronóstico de cada variable con las diferentes ventanas seleccionadas en el análisis espectral, no se observó que se tuviera un adecuado pronóstico de acuerdo a la serie original, como se puede observar posteriormente en la Figura 3-8 y se amplían dos ventanas correspondientes a la ventana una hora antes del evento clínico y la ventana media hora antes del evento clínico, como se observa en las figuras 3-9 y 3-10 en donde se puede apreciar que la señal pronosticada se aleja de manera importante de la serie original.

Figura 3-8: Predicción Modelos ARIMA para un paciente



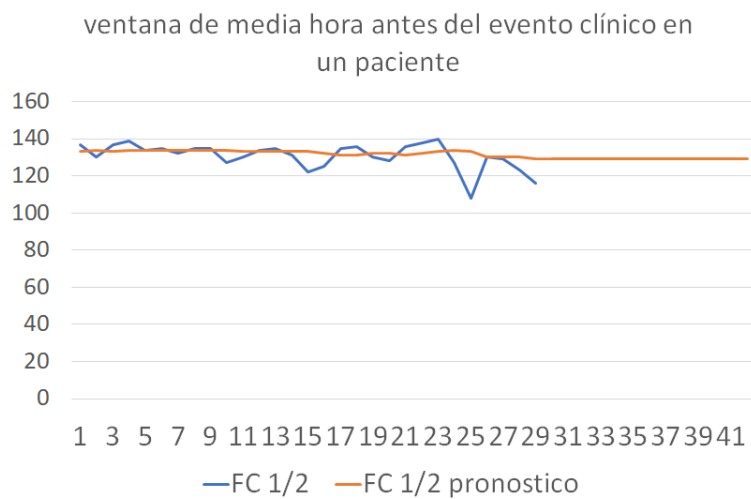
Fuente: Los autores

Figura 3-9. Pronóstico serie de tiempo una hora antes del evento clínico en un paciente.



Fuente: Los autores

Figura 3-10. Pronóstico serie de tiempo media hora antes del evento clínico en un paciente.



Fuente: Los autores

Conclusiones del análisis de series de tiempo:

- El análisis espectral (TF) resultó ser de valiosa ayuda para la elección de las ventanas más significativas, ya que por medio de TF se obtuvo el espectro de las señales en el dominio frecuencial, para aplicar modelos autorregresivos que permitieron relacionar los eventos clínicos con el comportamiento predictivo de las variables.
- Los resultados obtenidos en el modelo autorregresivo son ampliamente variables, y en algunas ventanas los datos son más eficientes para predecir y en otras no, lo que impide la realización de un adecuado pronóstico general o aplicable a un gran número de pacientes.
- Se encontró que esta serie de datos tenían un componente de estacionariedad, aleatoriedad, no toda la serie tenía estacionalidad y son datos en general sin tendencia.
- La aplicación de modelos autorregresivos demanda tiempo, alta capacidad operativa en términos del volumen de información que se procesa y los resultados predictivos de las variables son bajos.

Continuamos explorando otras técnicas de minería de datos para lograr buscar el modelo que mejor prediga el comportamiento de los datos.

3.4.1.3. Análisis de Agrupación de los Datos

Para conocer mejor la estructura de los datos se emplearon métodos de aprendizaje automático no supervisados que permitieron organizar los datos en conglomerados (*clústers*, en inglés) relativamente homogéneos. Se hizo un análisis de las técnicas de agrupación (*clústering* en inglés) y así poder visualizar los conglomerados y relacionarlos con las características de los 6 bioseñales.

Al realizar un análisis deductivo a partir de los clústeres encontrados y el problema planteado a solucionar en el desarrollo del estudio se observó que resultaron cuatro grupos que muestran la tendencia de los datos, esto para las seis variables de análisis y paciente por paciente. En el caso de la combinación de datos para cada patología, para las mismas

variables, pero, con la serie total de datos de pacientes, los datos pierden homogeneidad, resultando más grupos y un número elevado de patrones que hace más compleja la interpretación clínica del experto. En consideración a lo anterior, se adoptó el análisis paciente por paciente y los agrupamientos en la clasificación no supervisada se basaron en los algoritmos de *k-means* y de agrupación bietápico. La explicación de la aplicación de los modelos utilizados y los resultados para un paciente se presentan a continuación.

Técnica de agrupamiento *k-means*: se realizó el proceso en un programa implementado en el lenguaje Python, en donde primero se calculó el tamaño de los grupos (*clúster* s), luego se listan los centroides de cada uno de los grupos. Posteriormente, se graficaron los centroides de los clústeres s como se muestra cada uno de estos pasos en las figuras 3-11; 3-12.

Figura 3-11. Cálculo del tamaño de cada grupo para la serie de datos de un paciente.

```
[ ] y = km.predict(X)
    X.shape

↳ (20, 2)

[ ] #Se calcula el tamaño de cada cluster
    pd.Series(y).value_counts()

↳ 1    5
    6    4
    7    2
    5    2
    3    2
    9    1
    8    1
    4    1
    2    1
    0    1
    dtype: int64
```

Fuente: los autores

Posteriormente, el programa continúa con la fase de iteración, en donde realiza la inicialización y aleatoriamente asigna los demás puntos al grupo (*clúster*) más cercano en donde se realiza la conformación de los clústeres s finales.

Figura 3-12. Listado de los centroides en la serie de datos de un paciente.

```
[ ] #Listar los centroides de los clusters
    km.cluster_centers_

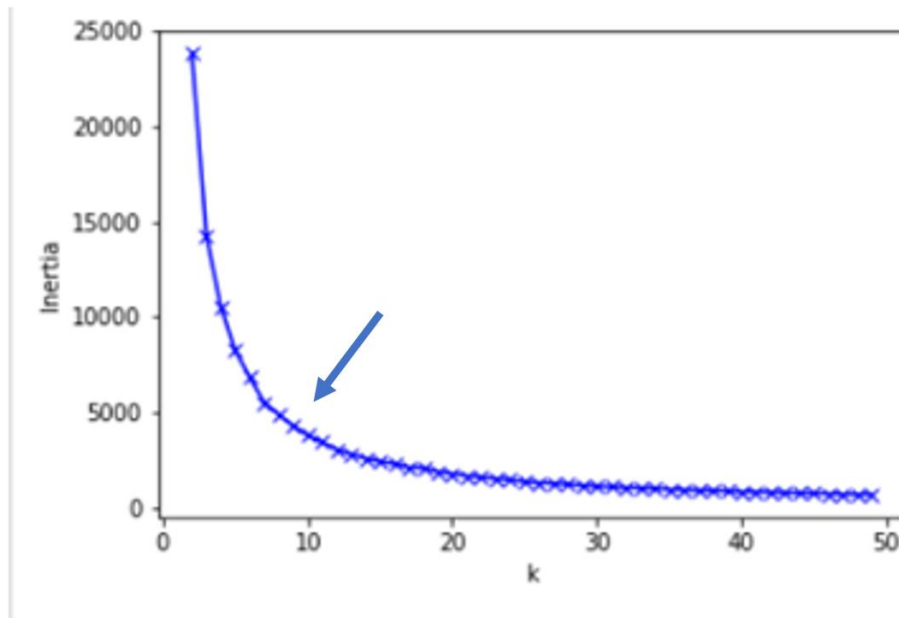
array([[131. ,  90. ],
       [153.4, 108.2],
       [ 81. ,  38. ],
       [163.5, 105.5],
       [123. , 115. ],
       [154.5,  87. ],
       [147.5, 107.5],
       [138.5,  99. ],
       [177. , 107. ],
```

Fuente: los autores

Para encontrar el número adecuado de grupos, se aplicó a la serie de datos el método del codo (*elbow method* en inglés), en este método del codo se grafica en el eje de las Y la inercia, en el eje de las X el número de grupos, y este valor corresponde el punto donde la inercia, luego de cambios bruscos, resulta casi invariable. Los datos agrupados por conglomerados para la serie de datos y su número óptimo para un paciente se muestran en la Figura 3-13, considerándose un número adecuado de conglomerados entre 8-9 grupos.

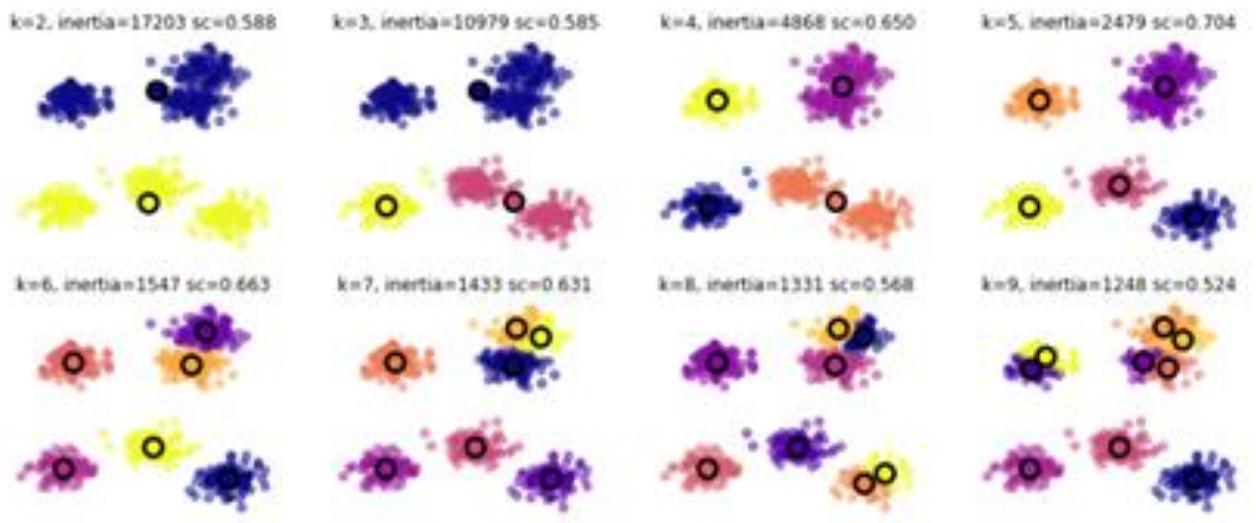
Como complemento al método del codo, se calcularon los valores del coeficiente de silueta (cs). Esta métrica permite conocer el grado de cohesión y calidad de la agrupación de los datos dentro de un conglomerado. El rango del coeficiente se encuentra entre (-1,1). Los valores cercanos a 1 indican que la observación está agrupada eficientemente y los valores cercanos a -1 o negativos indican que los datos fueron agrupados de manera errónea. En la Figura 3-14, el punto donde la inercia es más baja y se muestra un mayor coeficiente de silueta, corresponde a valores de 8 o 9 grupos (k=8 inercia:1331 cs=0,568; k=9. inercia:1248 cs=0,524)

Figura 3-13. Número óptimo de conglomerados para un paciente, gráfica del codo.



Fuente: autores

Figura 3-14. Coeficiente de silueta para diferentes agrupaciones de los datos de un paciente.



Fuente: autores

La interpretación de los grupos seleccionados se realizó por el clínico experto, en donde se observó que cada conglomerado reflejaba un estado clínico del paciente, esto se puede observar en la Tabla 3-9, donde primero se observan los valores de cada bioseñal en cada clúster y en la Tabla 3-10 se observa la interpretación clínica dada por el médico experto al evaluar cada grupo, de acuerdo al diagnóstico clínico, la edad y sexo, aplicada para un paciente.

Tabla 3-9. Método de agrupación (*clúster ing*) *k-means* para un paciente.

	Initial Cluster Centers								
	Cluster								
	1	2	3	4	5	6	7	8	9
FC1	132	143	140	159	66	125	143	122	116
porcentaje de oxigenación de la sangre	94	96	98	99	23	84	96	84	75
FR1	28	34	33	40	21	29	34	30	24
TAS1 tension sistolica	59	96	53	136	57	114	96	123	39
TAD1 presion diastoica arterial	61	59	33	77	33	54	109	95	48
TAM1 la presión de perfusión de los órganos corporales	77	51	39	96	41	89	51	112	13

FC: frecuencia cardíaca, FR: frecuencia respiratoria, TAS: tensión arterial sistólica, TAD: tensión arterial diastólica, TAM: tensión arterial media

Tabla 3-10 Interpretación clínica de cada uno de los grupos aplicada a un paciente

Interpretación clínica clúster	FC	SPO2	FR	TAS	TAD	TAM
clúster 1: hipotensión	133	93	30	59	40	46
clúster 2: Hipertensión	132	89	30	130	74	92
clúster 3: normal	141	95	33	97	58	52
clúster 4: normal	138	93	31	101	55	71
clúster 5: normal	146	91	30	86	45	59

clúster 6: hipoxemia	125	84	29	106	54	89
clúster 7: normal	138	93	31	95	51	65
clúster 8: hipertensión arterial	162	92	36	129	71	90
clúster 9: bradicardia extrema, hipoxemia, hipotensión	66	23	21	57	33	41

Posteriormente a la interpretación de los grupos, se pudo observar que los 9 grupos seleccionados por el programa, se podían finalmente reagrupar en 4 estadios clínicos, de acuerdo al número de valores anormales encontrados en cada bioseñal que tenía cada grupo, según lo indicado en la literatura médica para las condiciones clínicas, edad, y sexo de cada paciente. Esta reagrupación se puede ver en la tabla 3-10, donde se representan con colores de acuerdo a la complejidad del estado clínico que representaba cada grupo. Luego, en la Tabla 3-11 se observa la clasificación de los estados clínicos obtenidos por la reagrupación de los grupos iniciales arrojados por el algoritmo de *k-means*, para un paciente.

Tabla 3-11. Agrupación en diferentes estados clínicos de acuerdo a la complejidad arrojada por la interpretación clínica de los grupos encontrados por *k-means* para un paciente.

	estable
	potencialmente inestable
	inestable
	crítico

Tabla 3-12. Interpretación de los bioseñales vs. estados clínicos reportados por algoritmo *k-means* para un paciente.

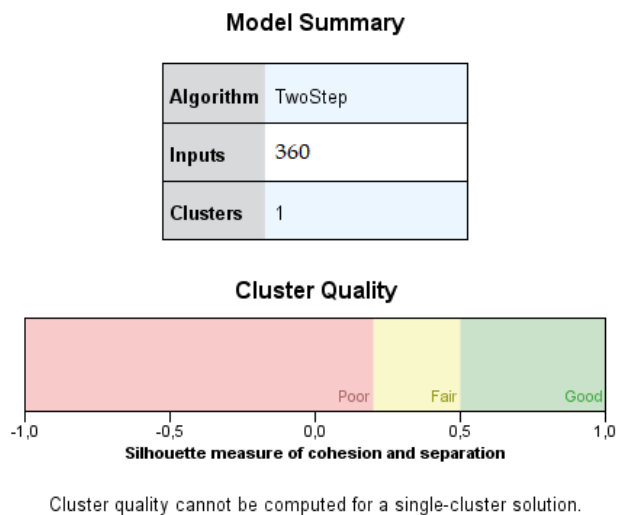
Descripción de los valores bioseñales	Evento clínico	Código
Todos los bioseñales en rangos normales	Estable	1
1-2 bio señales anormales	Potencialmente Inestable	2
3-4 bio señales anormales	Inestable	3
>4 bio señales anormales	Crítico	4

Con esta interpretación clínica, se convirtieron las variables continuas a categóricas como se muestra en la Tabla 3-12. Vale anotar que esta interpretación fue fácil de reproducir en todos los pacientes de la base de datos.

Análisis de conglomerados bietápico: con la aplicación de esta técnica se pretendía identificar las diferencias significativas entre los conglomerados y capturar la heterogeneidad en los comportamientos combinados.

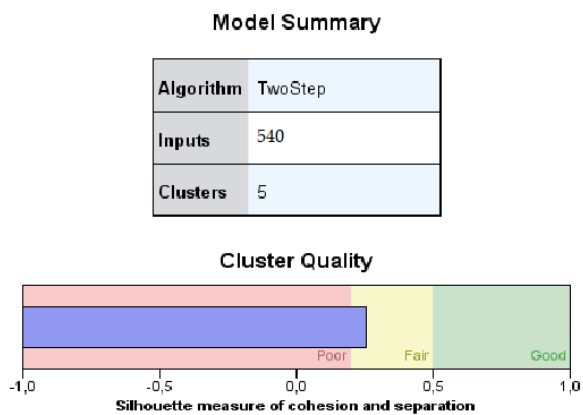
Los datos fueron procesados en el programa estadístico SPSS versión 23.0, y la base de datos completa de pacientes se ingresó con las variables categorizadas con las cuatro tendencias clínicas obtenidas por el método de *k-means* y también se realizó un análisis con las variables en su naturaleza continua. En la Figura 3-15 se describe el procedimiento para la serie total de los datos con las variables de forma categórica, en donde se observa que el modelo no es capaz de asignar los grupos y en la Figura 3-16 se describe el procedimiento para la serie total de datos con las variables de forma cuantitativa, en donde se observa poca cohesión en los grupos, el coeficiente de silueta es cercano a 0, indicando que no hay una adecuada medida de cohesión y separación entre los grupos.

Figura 3-15. Resumen de la aplicación del método de agrupación bietápico para toda la serie de datos con las variables en forma categóricas.



Fuente. Los autores

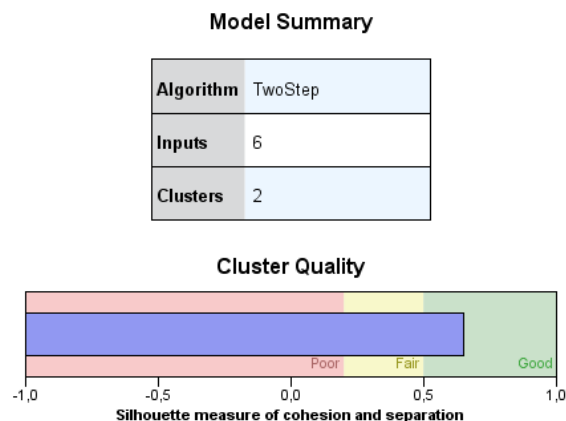
Figura 3-16. Resumen del método de agrupación bietápico para toda la serie de datos con las variables en forma continua.



Fuente: Los autores.

Adoptando el análisis paciente por paciente, el algoritmo construye de 2 a 3 conglomerados, dependiendo del paciente y, muestra que el grado de cohesión de los datos es adecuado puesto que el coeficiente de silueta de cada clúster es cercano a 1 (ver figuras 3-17 y 3-18). De otro lado, se presenta la tendencia clínica de los datos agrupados en cada grupo para el ejemplo de uno de los pacientes analizados. En la Figura 3-17 se puede observar el resumen del modelo utilizando variables cuantitativas. El modelo elaborado fue adecuado en el primer paso porque el coeficiente de silueta es cercano a 1. Luego observamos que en el segundo paso del método bietápico figura 3-18, la variable predictora, que es la que más le aporta a la elaboración del modelo para un paciente particular es la TAS, que corresponde con la información clínica de este paciente el cual tenía como diagnóstico de base Hipertensión Pulmonar.

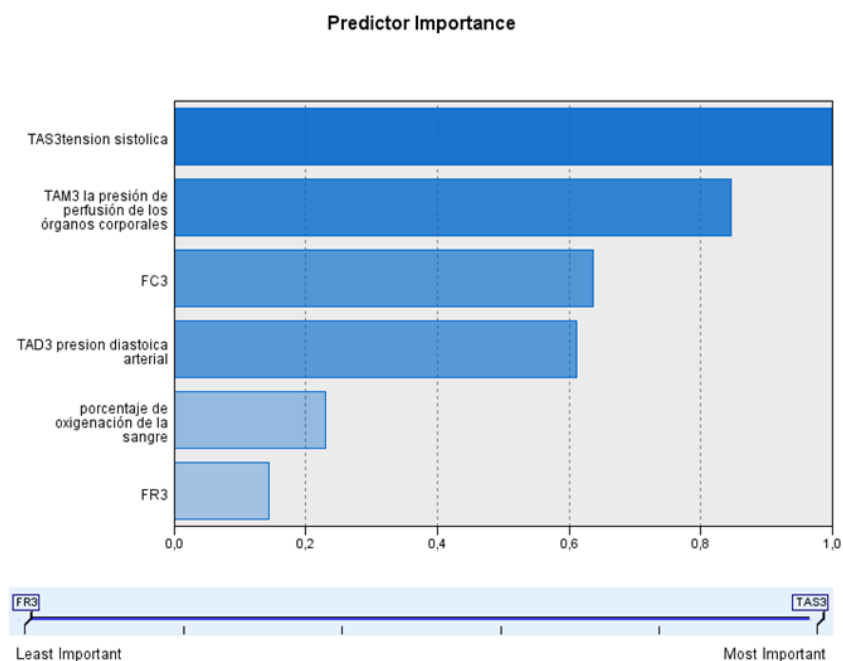
Figura 3-17. Resumen del método de agrupación bietápico para un paciente.



Fuente: autores

La interpretación de los conglomerados obtenidos por el método bietápico se realizó con el clínico experto, en donde se observó que el modelo bietápico reduce de manera muy significativa máximo el número de los conglomerados, obteniendo en la mayoría de los casos sólo dos estados clínicos en comparación a la interpretación de los estados clínicos interpretados por *k-means* que estaba más acorde al comportamiento real de los pacientes en UCIP.

Figura 3-18. Análisis de las variables que más contribuyen a la predicción a partir del conglomerado formado.



Fuente: Los autores.

Pero el agrupamiento bietápico mostró algo importante en la construcción del segundo paso del modelo para el clínico experto y fue la importancia en la predicción de la variable que más le aporta a la construcción del modelo, cuando se relacionaba la variable que más y la que menos le aportaba al modelo en contraste con el diagnóstico del paciente, se relacionaba con su respuesta fisiológica de compensación ante la patología que presentaba el paciente.

Continuamos entonces explorando otras técnicas de minería de datos para lograr buscar el modelo que mejor prediga el comportamiento de los datos, y tomando la descripción de los eventos clínicos arrojada en los modelos de agrupación.

3.4.1.4. Modelos de clasificación

Se realizó un análisis supervisado con clasificadores como: Bayes Ingenuo, árboles de decisión C4.5 (J48) utilizando la técnica J48 que es una mejora del árbol inicial C4.5 y redes neuronales (perceptrones multicapa), se trabajó con la variable de respuesta de manera categórica [42]. Se realizaron las anteriores técnicas en programas estadísticos como SPSS versión 23, WEKA, y en lenguaje Python.

Los árboles de decisión y Bayes ingenuo se realizaron para clasificación con variables categóricas (regresión logística) [42]. Cuando se realizó un análisis con variables categóricas se pretendía resolver un problema de tipo clasificación, intentando predecir los valores de una variable dependiente categórica (salida) a partir de una o más variables predictoras continuas (entradas). Por ejemplo, los atributos que eran las entradas (variables de signos vitales: FC, FR, SpO2, TAS, TAD y TAM) y las salidas corresponden a la clasificación de estados clínicos, que se habían obtenidos previamente en el análisis por conglomerados *k-means* (estado estable, potencialmente estable, inestable y crítico).

Se realizó una partición del conjunto de datos, usando muestreo estratificado: 70 % para entrenamiento y 30 % para *test*. Asimismo, se aplicó al conjunto de *test*, 10 *particiones cross-validation* con diferentes valores, tratando de obtener los mejores resultados.

En la Tabla 3-13 se observa que la clasificación de los estados clínicos (estable, potencialmente estable, inestable y crítico) mediante la metodología árbol de decisiones, se realizó una clasificación correcta de las instancias (98%), con una precisión adecuada para discriminar las clases estable y potencialmente inestable e inestable, pero no tuvieron buena precisión para determinar la clase crítica por observarse un valor de falsos positivos (FP) =0, ver Tabla 3-13. Los mismos resultados en la clasificación de los diferentes estados clínicos se observó con el clasificador Bayes ingenuo.

En el análisis de clasificación, se pudo observar que los dos modelos (Bayes ingenuo y árbol de clasificación J48) arrojaron iguales resultados, obteniendo un adecuado rendimiento al clasificar los datos cuando se trabajó con datos categóricos (más del 95%); sin embargo, ambos modelos tuvieron limitaciones en realizar la clasificación adecuada de los estados *críticos*, cuando se evaluaba la matriz de confusión, ver Tabla 3-14, y en la tabla 3-15 se realiza una comparación entre los dos clasificadores en relación a la clasificación de la clase estado crítico en un paciente, en donde se observa un inadecuado desempeño e iguales resultados en los dos clasificadores.

Tabla 3-14. Matriz de confusión del Árbol de decisión, usando variables categóricas, para un paciente.

```
=== Confusion Matrix ===
  a  b  c  d  <-- classified as
1120 0  0  0 |  a = POTENCIAL INESTABLE
  3 92  1  1 |  b = INESTABLE
  0  0 210  0 |  c = ESTABLE
  4  2  0  0 |  d = CRITICO
```

Fuente: los autores

Debido a los resultados anteriores, en donde los clasificadores tenían limitaciones para la clasificación de los estados críticos, se decidió realizar una matriz costo-beneficio [42].

Tabla 3-15. comparación de los resultados de los clasificadores Árbol de decisiones vs Bayes ingenuo en clasificar la clase estado crítico para un paciente

	Naive Bayes	Árbol de decisiones
cross-validation 10 fold		
% Instancias correctamente clasificadas	98,3	98,3
% Instancias incorrectamente clasificadas	1,67	1,67
TP Rate	0,5	0,5
FP Rate	0,011	0,011
Precision	0,158	0,158
Recall	0,5	0,5
F-Measure	0,240	0,240

TP: verdaderos positivos; FP: falsos positivos; Recall: sensibilidad

Bajo estos escenarios es posible incorporar los costos para cada tipo de error. Esto se realizó integrando en una matriz los beneficios (Bs) y costos (Cs), ver Tabla 3-16. La integración de una matriz de confusión y una matriz de costos-beneficios brindó las siguientes apreciaciones del modelo obtenido, como se muestra en las ecuaciones 17 y 18:

$$\text{Beneficios (Bs)} = VP * Bvp + VN * Bvn \quad \text{ecuación 17}$$

$$\text{Costo (Cs)} = FP * Cfp + FN * Cfn \quad \text{ecuación 18}$$

donde los beneficios (Bs) de la matriz de costos están dados por VP que corresponde a los verdaderos positivos, multiplicado por los beneficios de los verdaderos positivos (Bvp), sumado a VN que corresponde a los verdaderos negativos multiplicado por los beneficios de los verdaderos negativos (Bvn).

Los costos (Cs) están dados por los FP que corresponden a los falsos positivos, multiplicados por los costos de los falsos positivos (Cfp) sumado a FN que corresponden a los falsos negativos multiplicados por los costos de los falsos negativos (Cfn).

Tabla 3-16. Matriz costo-beneficio.

		predicción	
		Clase 1	Clase 2
Clase Real	Clase 1	Beneficio VP	Costos FN
	Clase 2	Costo FP	Beneficio VN

VP: verdaderos positivos; FN: falsos negativos; FP: falsos positivos; VN: verdaderos negativos

Se aplicaron las siguientes reglas para el análisis del clasificador costo-beneficio

- se asignaron valores a los errores de clasificación, estos valores estaban en un rango de 0 a 1,
- se minimizó la probabilidad de los costos y el modelo se volvía sensible al costo,
- se realizó un análisis de costo de aprendizaje tomando la clase minoritaria, compensando los costos mediante los pesos de los ejemplos y el costo se volvía sensible a la clasificación,
- Se realizó un análisis de elevar los costos a valores diferentes

Los resultados obtenidos en la matriz costo-beneficio de los diferentes clasificadores, tomando en consideración qué tan costoso resulta uno con respecto a los demás, mostró que clasificar a los falsos negativos (FN) (es decir, que la enfermedad esté presente y sea clasificada como ausente), era más costoso en todos los clasificadores. Si bien es cierto, que un diagnóstico FN puede traer consecuencias fatales, los falsos positivos (FP) también tienen un costo, pero no tan elevado o tan significativo, porque no conlleva a las mismas consecuencias que un FN. Se ha observado que, en los sistemas de diagnósticos asistidos

por computadora, se prefiere implementar un clasificador que cueste menos a uno que clasifique mejor [42]. Pero esta consideración en el área de UCIP, no se consideró pertinente, razón por la cual no se eligió ningún clasificador como modelo, porque todos tenían limitaciones en identificar un estado menos frecuente, que en este trabajo correspondía a los estados críticos.

Conclusiones del análisis de los métodos de clasificación:

- La reducción de características utilizando métodos de clasificación no fue satisfactorio para las series de datos aplicada.
- Los modelos son débiles en clasificar adecuadamente los estados “menos activos”, es decir, aquellos en los que los pacientes de la serie de datos permanecieron por menor tiempo. Siendo modelos costosos para identificar estos estados críticos, como se observó en la matriz de costos los FN.
- Estos modelos tuvieron un mejor desempeño cuando la serie se discretiza con datos categóricos más que con datos cuantitativos.

Continuamos explorando otras técnicas de minería de datos para lograr buscar el modelo que mejor prediga el comportamiento de los datos.

Perceptrón Multicapa

Esta técnica tiene la intención de estudiar e interpretar las clasificaciones y relaciones con el etiquetado original de estados clínicos encontrados con el experto en el análisis de agrupación *k-means* (salidas), de los datos correspondientes a la base de datos de signos vitales (entradas), y compararlos con los resultados obtenidos con los otros métodos de clasificación realizados anteriormente. El procesamiento de datos se practicó mediante el software WEKA (entorno para análisis del conocimiento de la Universidad de Waikato).

Utilizando el algoritmo de aprendizaje supervisado retropropagación (Backpropagation), la red aprende la relación entre los ejemplos de entrada, valores de signos vitales, y su correspondiente clase, definida por los diferentes estados clínicos (fase de entrenamiento), para luego poder categorizar un ejemplo de entrada del cual se desconoce su salida (fase

de aprendizaje). Antes de realizar la fase de entrenamiento, se escogió el método utilizado para la evaluación de este algoritmo. Los métodos escogidos para este estudio fueron: uso del conjunto de entrenamiento (Using training set), validación cruzada (Cross-validation), porcentaje dividido (Percentage Split).

Los valores de los parámetros frecuencia de aprendizaje (Learning Rate) y momento, en la mayoría de los casos donde se aplicó este algoritmo de aprendizaje, para algunas pruebas satisfactorias, estos parámetros fueron entre 0,1-0,5 el número de épocas fue entre 500-1000.

Se trabajó con datos de paciente por paciente y toda la combinación de toda la base de datos, pero este algoritmo arrojó mejores resultados cuando se trabajó con toda la base de datos que correspondió a las 94,678 instancias. Esto hizo que se mejorará la clasificación de los estados clínicos, también se aumentó el número de neuronas en la capa oculta al mismo número de entradas (signos vitales), pero cabe destacar que al aumentar el número de neuronas por encima del número de entradas no se obtuvieron mejores resultados, simplemente no variaron.

Utilizando el método de evaluación Validación Cruzada (Cross-Validation), se obtuvo mejores resultados, se validó sobre las 94,678 muestras, con diez particiones (folds). El valor promedio obtenido fue de 96,7% instancias bien clasificadas y 3,2% de instancias mal clasificadas de la muestra total. esta evaluación se realizó con un valor de

frecuencia de aprendizaje: 0,4

Momento:0.5

época: 500

Cross validation 10 folds

Lográndose un error más pequeño y una mejor clasificación de las clases.

La matriz de confusión que se muestra en la tabla 3-17 y por último la red neuronal generada por el modelo que se observa en la figura 3-20 muestra que aún hay estados mal clasificados, ya que no es detectado por el clasificador. Como se puede observar, se

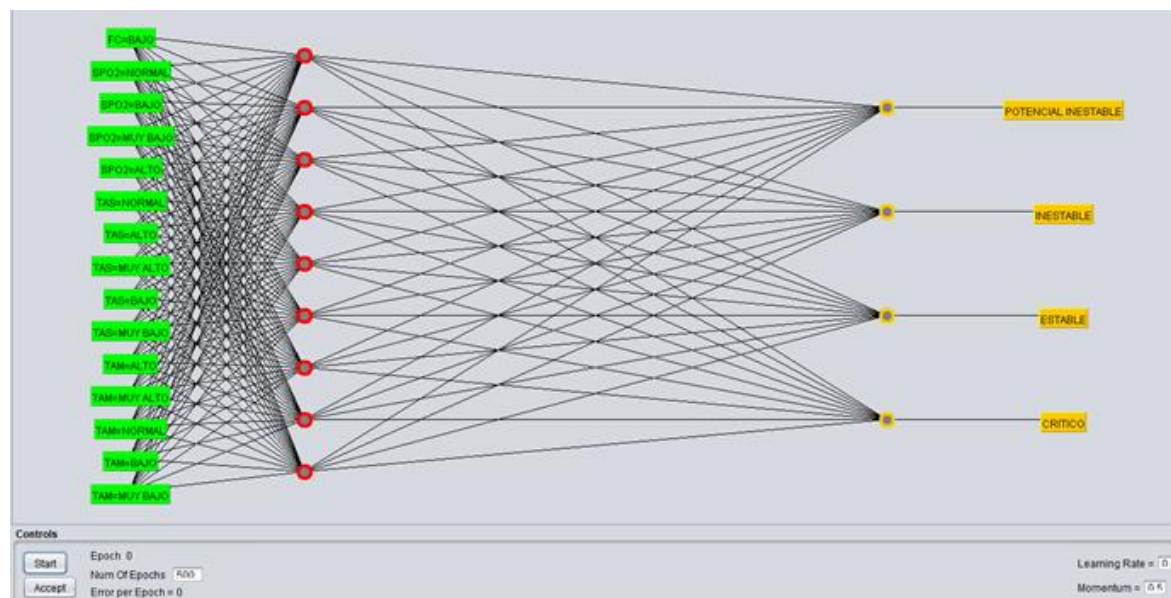
obtuvo una disminución de valores que no influyen mayormente en los resultados de clasificación y por consiguiente se obtiene un modelo más simple y acotado. Por otro lado, ya se tenía indicios con los otros clasificadores utilizados sobre la limitación de estos clasificadores en clasificar los estados críticos.

Tabla 3-17. Matriz de confusión del perceptrón multicapa aplicada a toda la base de datos.

```

=== Confusion Matrix ===
      a      b      c      d  <-- classified as
47334 1783      0      0 |   a = POTENCIALMENTE INESTABLE
 969 32281      0    107 |   b = INESTABLE
  0      0 9623      0 |   c = ESTABLE
  8    148      0 1106 |   d = CRITICO
  
```

Figura 3-20 Red neuronal generada con tan sólo 4 variables que caracterizan la clase.



Fuente: los autores

En la Tabla 3-18 se observa que las métricas que evalúan el desempeño del modelo no son confiables por el desbalance de las clases, no se puede estimar un valor en las métricas precisión, medida F, y el área ROC no fue satisfactorio con valor de 0.139, reflejado también en el puntaje del área PRC (curva precision-Recall), que es útil en estos casos cuando tenemos datos no balanceados y la clase positiva ocurre pocas veces, su valor está alejado de uno (0.003).

Tabla 3-18. Validación estadística del modelo perceptrón multicapa.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,924	0,032	0,990	0,924	0,956	0,830	0,953	0,977	POTENCIAL INESTABLE
	0,918	0,002	0,967	0,918	0,942	0,938	0,953	0,909	INESTABLE
	1,000	0,070	0,709	1,000	0,830	0,812	0,965	0,721	ESTABLE
	0,000	0,000	?	0,000	?	?	0,139	0,003	CRITICO
Weighted Avg.	0,931	0,035	?	0,931	?	?	0,951	0,931	

TP: verdaderos positivos; FP: falsos positivos; Recall: sensibilidad; F-measure: medida F; MCC: coeficiente de correlación Matthews; PRC: área curva entre precisión-Recall

Sin embargo, al evaluar toda la base de datos, en lo que respecta a la comparación de los resultados obtenidos tanto con el árbol de decisión como con el clasificador Bayes ingenuo y el perceptrón multicapa, se pudo verificar que éste último permite la resolución del problema en la búsqueda de los atributos que mejor caracteriza a una clase, cuando se evalúa toda la serie de datos con las 94,678 instancias, con resultados adecuados como si la evaluación se realizará de manera individual con cada paciente. Este efecto no se observó con las otras técnicas de clasificación. Como se observa en la tabla 3-18, en donde se observa la matriz de confusión aplicada a toda la serie de datos, aunque se observa que aún presenta las mismas limitaciones para clasificar correctamente a la clase menos frecuentes.

Justificación sobre las técnicas de clasificación utilizadas

Se realizó una serie de pruebas sobre el conjunto de datos con el fin de caracterizar patrones para la detección de deterioro de los pacientes en estado crítico que ayuden al clínico experto a examinar y verificar más fácilmente los resultados obtenidos para apoyar la toma de decisiones. Estos experimentos realizados fueron orientados a evaluar la

exactitud de clasificación de los estados clínicos menos frecuentes (estados críticos), basándonos en técnicas como arboles de decisión (J48), Naive Bayes y redes neuronales (Multicapa Perceptrón). Estas técnicas de aprendizaje empleadas son los métodos fáciles de utilizar, entender y que han demostrado dar buenos resultados. Se utilizan comúnmente cuando se necesitan detectar reglas que puedan ser fácilmente traducidas a lenguaje natural o en la construcción de modelos de clasificación. Para ello se empleó el análisis de matrices de confusión y una vez que el modelo de clasificación fue construido, fue necesario evaluar su rendimiento respecto a modelos similares a fin de determinar, si el modelo propuesto es competitivo en comparación con los demás. En este caso, una prueba de significancia fue realizada, en la cual, el modelo de red neuronal (perceptrón multicapa) fue comparado con un modelo de árbol de decisión (J48) y un modelo probabilístico como Naive Bayes. La metodología para realizar las comparaciones se basó en seleccionar el mejor modelo encontrado, teniendo como parámetro de desempeño la medida de sensibilidad, F1 score exactitud (*Accuracy*), tasa de falsos positivos (FP Rate) y tasa de falsos negativos (FN Rate). Las tablas 3-13, 3-14, 3-15, muestran estos resultados teniendo en cuenta la clase etiquetada como inusual o menos frecuente (estados críticos) las cuales son el interés principal de este trabajo.

Analizando los resultados anteriores, dado el conjunto de datos de validación, se evidencian dificultades para obtener una buena exactitud en todas las técnicas en la detección de clase inusual (estado crítico). Los anteriores resultados se pueden justificar dado el reducido número de ejemplos de la clase inusual utilizados en el conjunto de entrenamiento, lo cual ocasionó que los algoritmos se inclinaron por la clase con mayor presencia. Dado los malos resultados en el proceso de predicción para las clases etiquetadas como menos frecuentes del muestreo anterior, se decidió hacer un muestreo estratificado proporcional y pruebas (validación cruzada de 10 pliegues) como se muestra en la Tabla 3-19, para el modelo de red neuronal, arboles de decisión y Bayes. pero el error de predicción de la clase menos frecuente sigue siendo alto en comparación con las clases más frecuentes, el valor de ROC para Naive Bayes fue 0,59, para el árbol de decisión de 0,32 y para el perceptrón multicapa de 0,4. Lo que indica que no hay una diferencia estadísticamente significativa para decir que un algoritmo es mejor que el otro.

Tabla 3-19 Comparación de diferentes técnicas de clasificación para predecir estado crítico, a través de muestreo estratificado proporcional

	Naive Bayes	Árbol de decisiones	Multicapa Perceptrón
cross-validation 10 fold			
% Instancias correctamente clasificadas	33	45	45
% ERA	86,8	100,10	115,8
TP Rate	0,5	0,5	0,41
FP Rate	0,011	0,011	0,04
Precision	0,158	0,158	0,4
Recall	0,5	0,5	0,45
F-Measure	0,240	0,240	0,42
ROC	0,59	0,32	0,4

cross-validation 10 fold: validación cruzada 10 veces; %ERA: porcentaje error relativo absoluto; TP Rate: verdaderos positivos; FP Rate: falsos positivos; Recall: sensibilidad; F-Measure: medida F; ROC: Característica Operativa del Receptor

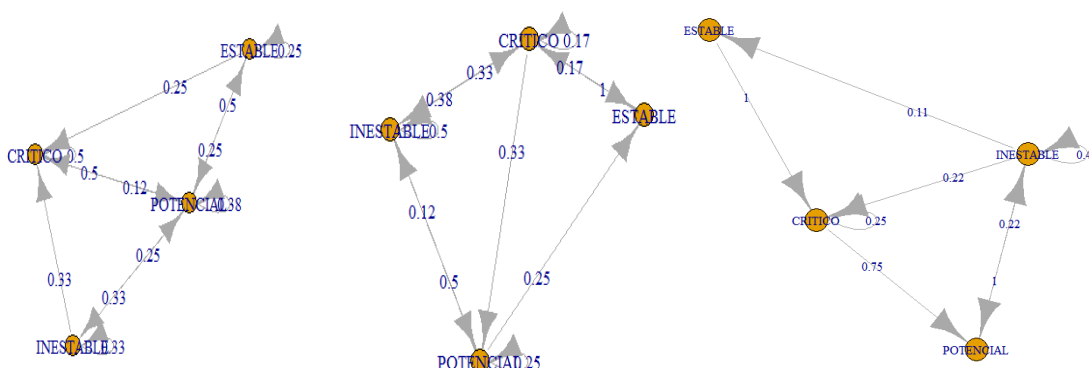
3.4.1.5. Análisis basado en cadenas de Márkov

Los modelos de Márkov son un sistema matemático que experimenta transiciones de un estado a otro de acuerdo a un conjunto de reglas probabilísticas. Para evaluar si la serie de datos cumplía con las propiedades de Márkov, se llevó a cabo la función: *verificar* propiedad de Márkov con el lenguaje Python, el cual prueba si una secuencia dada de eventos sigue la propiedad de Márkov realizando pruebas de Chi-cuadrado en una serie de tablas de contingencia derivadas de la secuencia de eventos. Los valores p con una significancia mayor a 0.05 indican que la hipótesis nula de una secuencia que sigue a la propiedad de Márkov no debe ser rechazada.

Se trabajó con las variables categóricas, en donde se obtuvieron buenos resultados. Sin embargo, tomando la base de datos transformada en categóricas según los estados clínicos definidos en los conglomerados encontrados en *k-means*, y evaluando paciente por paciente, se logró obtener los mejores resultados, y estos fueron los siguientes:

En la Figura 3-21 se muestra el ejemplo de 3 pacientes con patologías diferentes, en donde se observan cadenas de Márkov con estados recurrentes positivos, accesibles, irreducibles, transitorios.

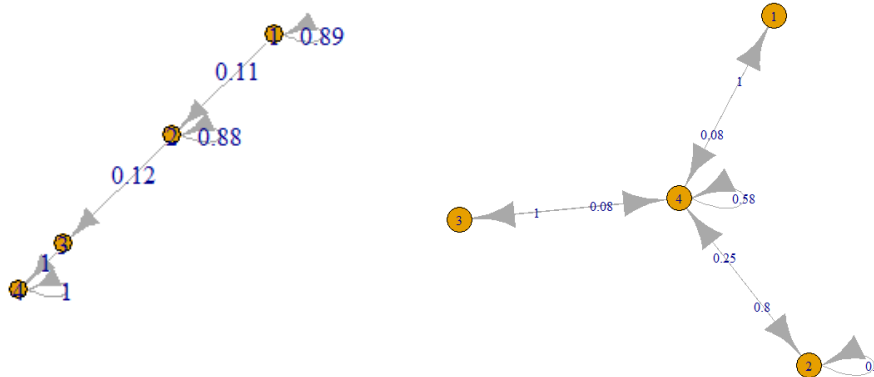
Figuras 3-21 Cadenas de Márkov que representan las características más frecuentes de los datos en tres pacientes diferentes



Fuente: los Autores.

En la Figura 3-22 se observan comportamientos de pacientes con cadenas no recurrentes, con estados absorbentes. Este comportamiento fue más frecuentemente observado en los pacientes que fallecieron.

Figura 3-22. Cadenas de Márkov con estados absorbentes que representan a los pacientes que fallecieron, aplicado para dos pacientes.



Fuente: Los autores.

Conclusiones del análisis basado en cadenas de Márkov:

- El modelo basado en cadenas de Márkov es satisfactorio para abordar el comportamiento de los pacientes en UCIP de manera individual.
- Pareciera que las cadenas de Márkov con características: con estado absorbente, reducibles, aperiódicas se encuentran más frecuentemente en los pacientes que fallecen en comparación con los que no; vale anotar que no hubo diferencias cuando se agrupaban enfermedades o edades similares.
- Con las cadenas de Márkov se puede observar que los estados recurrentes pueden corresponder a 1 solo estado.

3.4.1.6. Análisis basado en Modelos Ocultos de Márkov (HMM)

Para el análisis basado en HMM se trabajó con la base de datos transformada en datos categóricos, se utilizaron seis signos vitales para la detección de anomalías. Los valores de normalidad están descritos en la Tabla 3-12 que se utilizaron para definir los diferentes estados clínicos. Los 4 estados clínicos elegidos fueron los obtenidos por el análisis *k-means* y se utilizaron para construir los modelos HMM. La duración de la ventana de predicción se consideró durante un tiempo mayor de 3 segundos a 5 minutos, basados en el concepto definido por el médico experto.

Primero se exploró la base de datos tomando paciente por paciente, usando más de 1000 segundos en un paciente, observando cuatro tipos de eventos clínicos (estable, potencialmente estable, inestable, crítico). para este paciente las anomalías ocurrieron después de 321 segundos. Durante los primeros 321 segundos, no se observó variabilidad en los datos, su comportamiento hasta ese momento era constante, durante este tiempo el estado evidenciado fue potencialmente inestable, en total se evidenciaron 610 episodios de estados estables, 97 episodios inestables y 6 episodios en estado crítico. Los eventos menos frecuentes, como en este caso era el estado crítico, siendo el estado que menos frecuentemente se detectó por el sistema de reconocimiento basado en HMM, al calcular el log verosimilitud se obtuvieron valores entre -749.312425 y -756.528002 como se puede observar en la tabla 3-20, este mismo comportamiento se pudo observar para los datos de otros pacientes.

Tabla 3-20. Resultados cálculo de Log verosimilitud HMM para un paciente.

Mejor log likelihood: -736.7266170235812 con estados	
	log likelihood
Log de la probabilidad que el siguiente estado sea crítico	-756.528002
Log de la probabilidad que el siguiente estado sea estable	-752.661404
Log de la probabilidad que el siguiente estado sea inestable	-750.035935
Log de la probabilidad que el siguiente estado sea potencialmente inestable	-749.312425
Siguiente observación más probable	INESTABLE

log likelihood: logaritmo verosimilitud

Con base a los resultados obtenidos, se combinaron todos los pacientes y se realizó un nuevo análisis con un total de 90 pacientes y 94,678 instancias. Para ello fue necesario

montar una única base de datos con todos los registros para separar el conjunto de entrenamiento, prueba y validación. En los modelos de clasificación se realizó una partición del conjunto de datos, utilizando un muestreo estratificado: 70% para entrenamiento y 30% para pruebas y prueba de validación cruzada. Luego se entrenó el modelo y se aplicó el modelo al conjunto de prueba para luego validar el desempeño del modelo calculando la matriz de confusión, exactitud, error de clasificación, precisión y medida F1. Se prefieren estos criterios de evaluación porque dan cuenta del desequilibrio de clases en nuestro conjunto de datos, como se observa en la tabla 3-19, en donde se obtuvo una exactitud para la clasificación de todos los estados clínicos de 0,7; precisión de 0,75; medida F1: 0,61; y sensibilidad de 0,75. siendo resultados aceptables para esta prueba, teniendo en cuenta las limitantes encontradas en los modelos de clasificación anteriormente expuestos. Después de realizar lo anterior, en la Tabla 3-21 se muestran los resultados obtenidos

Tabla 3-21 Métricas 10-fold cross validation para cada evento clínico usando HMM

	Exactitud	Precisión	Medida-F1	Sensibilidad
ESTABLE	0,81	0,9	0,7	0,9
POTENCIALMENTE INESTABLE	0,8	0,7	0,7	0,7
INESTABLE	0,6	0,93	0,7	0,93
CRÍTICO	0,5	0,6	0,5	0,6
TOTAL	0,7	0,75	0,65	0,75

cross validation: validación cruzada

El mejor desempeño se obtuvo con *10-fold cross validation* cuando se observa un desbalance en los datos como es el caso en este trabajo, es importante el resultado que nos muestra la medida F1, que en este análisis fue de 0,65, ver Tabla 3-21.

Conclusiones del análisis HMM:

- HMM es un estimador probabilístico que puede servir para predecir anomalías en los pacientes críticamente enfermos
- Los HMM presentan limitaciones al igual que otros clasificadores como lo es el detectar aquellos estados que se presentan con menor frecuencia
- Los HMM pueden ser utilizados con variables en forma categórica, y los resultados de predicción se pueden reproducir a todos los pacientes.

Discusión sobre la integración de los diferentes modelos propuestos

En este estudio la combinación de diferentes técnicas de minería de datos tanto descriptivas como predictivas ayudaron a identificar patrones en los datos y que se relacionaban con el comportamiento de los pacientes de este conjunto de datos. Estos patrones fueron visibles con el análisis de conglomerados *k means*, en donde se pudieron determinar cuatro estados clínicos (estado estable, potencialmente estable, inestable, crítico), que resumían el comportamiento individual de los pacientes, estas salidas arrojadas por *K means* fueron muy útiles para ser posteriormente utilizadas en modelos predictivos como lo son los modelos markovianos.

Por medio de las cadenas de Márkov, se pudieron observar las transiciones por las que cada paciente pasaba de un estado a otro y la probabilidad de permanecer en un estado absorbente. Con el análisis de cadenas de Márkov observamos que el 27% de los pacientes de la base de datos habían fallecido y en este grupo de pacientes se observó que era más frecuente la presencia de estados absorbentes en comparación a los que no fallecieron.

Las salidas que se obtuvieron en el análisis de *k means*, correspondientes a 4 estados clínicos (estado estable, potencialmente estable, inestable, crítico), también contribuyeron a la realización de HMM y además la realización de HMM complementan los hallazgos encontrados en las cadenas de Márkov, puesto que con las cadenas de Márkov las señales observadas corresponden a los estados que son visibles directamente para el observador, por lo que las probabilidades de transición entre estados son los únicos parámetros que pueden conocerse, pero cuando el estado no es visible para el observador, se puede recurrir a los HMM. En los HMM, la cadena de Márkov subyace tras las observaciones, y así al correlacionar las observaciones y las transiciones de estado lo que se busca es encontrar el estado de secuencias más probable. Estos estados más probables fueron observados con HMM, sin embargo, se observó que los HMM tienen limitaciones al igual que los modelos de clasificación en predecir los estados inusuales o menos frecuentes por los que pasa un paciente, correspondiendo a los estados críticos. Aunque el desempeño

de HMM en comparación con los métodos de clasificación fue mejor como se observa en la Tabla 3-20 y 3-21, se convierte en un método prometedor para predecir el deterioro de los pacientes en estado crítico.

3.5. Evaluación e implementación del modelo

3.5.1. Validación del método propuesto con el diagnóstico proporcionado por los expertos

El clasificador desarrollado basado en HMM tiene múltiples propósitos. En primer lugar, puede decir cuál de los posibles eventos clínicos estará un paciente en la ventana de predicción, utilizando T segundos de datos observados inmediatamente antes de que ocurra dicho evento clínico. En segundo lugar, una vez que conoce la posible clase de una observación, también puede estimar la secuencia de estado más probable para esa observación utilizando el algoritmo de Viterbi [62,65]. Este proceso también ayuda a estimar estados futuros para posibles observaciones y, por tanto, a detectar anomalías futuras. En tercer lugar, se puede utilizar en la monitorización de pacientes en tiempo real estimando la probabilidad de cada uno de los estados clínicos. En el momento t , la secuencia observada se genera a partir de los datos en forma permanente y se calcula $\log P(O | \lambda)$ para cada uno de los estados, lo que genera las estimaciones de probabilidad, y se clasifica continuamente la condición de los pacientes en uno de los posibles estados clínicos.

Se comparó el desempeño de este clasificador basado en HMM con clasificadores como árbol de decisiones, J48 y redes neuronales como perceptrón multicapa. Por tanto, podemos ver que se entrenó un modelo probabilístico dinámico como el HMM y puede identificar estados clínicos menos frecuentes, como el estado "crítico" con una buena precisión en comparación con un modelo predictor como una red neuronal, cuando se utilizó el mismo conjunto de características. La comparación entre las diferentes medidas de precisión para detectar el estado crítico usando varios modelos de clasificación se presentan en la Tabla 3-22. De esta observación podemos ver que la clasificación basada en el modelo probabilístico HMM se comportó mejor que un modelo predictor por otros

clasificadores (árbol de decisión, Bayes ingenuo, ANN). Cuando se discrimino por estados clínicos, se observó que a pesar que todos los modelos gozaban de un aceptable desempeño, y una buena exactitud, el modelo HMM mostró un mejor desempeño en la predicción de los estados clínicos menos frecuentes. En la tabla 3-22 tabla podemos ver que la clasificación de los estados críticos basada en HMM tuvo un mejor resultado en las métricas de clasificación en comparación con los demás clasificadores.

Tabla 3-22. Comparación de la clasificación de estados críticos entre el método HMM vs otros clasificadores

Evento Crítico	Recall	Precisión	F-medida	Exactitud
Bayes Ingenuo	0,5	0,158	0,24	0,33
árbol de decisiones	0,5	0,15	0,24	0,45
Multicapa Perceptrón	0,45	0,4	0,42	0,45
HMM	0,6	0,6	0,5	0,5

Recall: sensibilidad, Precisión: precisión; F-Measure: medida F; Accuracy: Exactitud; HMM: Hidden Márkov model; Multilayer Perceptron: perceptrón multicapa

También, se realizó una comparación entre los estudios realizados similares a éste como lo fueron los trabajos de Wongeun Song [15], en donde se describe un modelo de predicción basado en métodos de clasificación para predecir el deterioro clínico de pacientes neonatos con sepsis; además, se incluyó el trabajo realizado por el grupo de Ioan Stanculescu et al [58], ellos también realizaron un modelo predictivo basado en HMM, detectando el deterioro de pacientes pediátricos con sepsis. Sayandeep Acharya et al, [67] mediante filtros de Kalman evaluaron series de tiempo conformada por signos vitales el deterioro de los pacientes, y el grupo de Aditya Nagori [68] construyeron un modelo predictivo para detectar estados de choque mediante imágenes térmicas comparándolas con signos vitales. Estas comparaciones se presentan en la Tabla 3-23, con esta comparación se puede concluir que el modelo propuesto en este trabajo tiene un desempeño adecuado, en comparación con otros trabajos (medida F =0,65, exactitud

=0,7), estos resultados se obtuvieron con un significativo número de muestras, el modelo propuesto puede predecir los eventos clínicos menos frecuentes en referencia a otros trabajos similares y otros clasificadores. Ver Tabla 3-23

Tabla 3-23 comparación de este modelo (HMM) con los resultados y características obtenidas de trabajos similares

	nuestro modelo [40]	trabajo Wongeun Song [15]	Trabajo Ioan Stanculescu[69]	Trabajo Sayandeep Acharya[67]	Trabajo Aditya Nagori[68]
número de bioseñales	6	7	3	2	6
niños críticos	si	si	si	no	si
evento clínico	si	si	si	si	si
número de muestra	90 pacientes (94,678 instancias)	7870	18 pacientes (25920 instancias)		539
Accuracy	0,7	0,56		0,76	0,69
F1 score	0,65	0,3	0,65	0,56	AUC: 0,7
capacidad de predicción	SI HMM	si	HMM	ARMA-Kallman filtro	si

Accuracy: exactitud; F1 score: medida F1; HMM: Hidden Márkov chain (modelos ocultos de Márkov);AUC: área debajo de la curva

● Capítulo 4. Discusión y análisis de resultados

En este trabajo se obtuvo con un resultado satisfactorio la implementación de un modelo que apoya el seguimiento y ayuda al diagnóstico médico y refleja la situación clínica de los pacientes de una UCIP. Se verificó que el uso de técnicas de reducción de dimensionalidad, análisis de agrupamiento, análisis de clasificación en aplicaciones médicas puede ser de gran utilidad, si se considera como una herramienta auxiliar para profesionales de la salud.

El análisis espectral, específicamente usando la transformada de Fourier y la aplicación de filtros de paso de banda que se emplearon, son útiles para identificar las ventanas más significativas en donde ocurrieron los eventos clínicos, y estos hallazgos asociados con la estadística descriptiva, pueden ayudar a identificar el comportamiento de cada paciente [13].

La utilización de análisis de agrupamiento, como en este estudio que se basó en *k-means* o por el método de agrupación bietápico basados en análisis no supervisado fue importante, porque también simplifica el trabajo de los médicos, se observó que por medio de ellos también se obtenían los límites de normalidad en los signos vitales, de acuerdo a la edad, y también se correlacionaron con los utilizados actualmente descritos en la literatura y a la vez la tendencia del comportamiento de los signos vitales de manera individualizada, que al comparar estos límites de normalidad arrojados por esta técnica en comparación con los estipulados por la literatura, se puede concluir que fue otra herramienta valiosa para simplificar el trabajo del personal de salud que cuida a los pacientes, porque no tendrían que estar realizando ajustes a los monitores de manera continua de acuerdo a las metas que se vayan trazando en la evolución continua del paciente críticamente enfermo. el conglomerado Bietápico proporciona en la segunda etapa de su construcción, otra información valiosa a los clínicos, y es el mostrarle la variable que más influye en la construcción del modelo, siendo la evaluación de esta

variable por el médico experto muy importante porque justamente se asoció al mecanismo de compensación individual que hace cada paciente de manera fisiológica para subsanar las alteraciones hemodinámicas que presenta cada paciente. Estos hallazgos también han sido demostrados en es estudios como Guoxian Yu et al [70], en donde utilizaron la técnica de bi clúster para detectar expresión génica del cáncer a gran escala y lograron descubrir subtipos de cáncer con mayor precisión que otros algoritmos relacionados, siendo otra herramienta que puede ser aplicada en el escenario de UCIP.

El análisis de agrupamiento también mostró un clasificación de estados clínicos que correspondían a los eventos clínicos por lo que cursaron los pacientes durante su estancia en la UCIP, y esta manera de clasificación de los eventos clínicos fue fácilmente reproducible en toda la serie de datos, siendo una base de datos muy heterogénea, el análisis de agrupamiento por *k-means*, dio una información sobre el comportamiento individual de los paciente, teniendo en cuenta que este tipo de aprendizaje no supervisado es usado más frecuentemente par la identificación de grandes grupos de pacientes con ciertas características de enfermedad, para estudios de grupos de patologías específicas, reduciendo errores médicos y costos y mejorando la eficiencia del personal y los recursos [71,72]. En este estudio fue útil para un grupo de pacientes con enfermedades heterogéneas.

Cuando se analizó las cadenas de Márkov, también mostró ser una herramienta útil revelando una mejor aproximación diagnóstica para dirigir la terapéutica que se le brinda a los pacientes en estos ambientes críticos. Las cadenas de Márkov mostraron las diferentes transiciones por las que pasa un paciente en estado crítico, y la evaluación por el médico experto lo asocio a la variabilidad fisiológica de cada paciente ante una situación de inestabilidad clínica, en este estudio se encontró que los pacientes que tenían cadenas de Márkov no recurrentes, con estados absorbentes, fueron los que tuvieron un desenlace fatal, esto se correlaciona con lo descrito en la literatura, en donde se observa que aquellos pacientes que no tienen variabilidad en la respuesta fisiológica de signos vitales, mucho más estudiado con la FC [73,74,75], tiene más riesgo de muerte en comparación con aquellos pacientes que tienen variabilidad fisiológica es decir, que tienen un descenso y aumento de FC en situaciones de stress, y esta respuesta les da la capacidad de entrar y

lograr salir de los estados más inestables, eso significa buena respuesta o tener variabilidad fisiológica [75].

Ahora bien, aún es difícil predecir un comportamiento de un paciente en UCIP, estas respuestas fisiológicas en los estados críticos, en donde estos escenarios se caracterizan por ser muy heterogéneos, con respuesta fisiológicas distintas, aún no es fácil de reproducir con técnicas actuales de IA, con técnicas de redes neuronales, técnicas de clasificación, porque vemos que tanto este estudio como los estudios similares encontrados hasta el momento [15,69], las métricas no muestran aún un excelente rendimiento. Y esto puede ser debido a que reproducir un comportamiento fisiológico aún depende de una serie de factores que son de respuesta individual, como lo son la patología de base, antecedentes, capacidad de respuesta ante el estrés, entre otros [67,68, 76], que aún están en proceso de desarrollo en las técnicas de IA actuales.

A pesar de que nuestro modelo predictor tuvo un desempeño aceptable en comparación con los otros trabajos, este sigue siendo un terreno fértil para explorar nuevos modelos matemáticos que puedan reproducir la fisiología de cada individuo, y anexar a los modelos de redes neuronales o de clasificación otras variables fisiológicas, como el modelar la interacción de las neuronas con el estado vascular o el estado gasométrico, entre otros [15,74].

La fortaleza de este trabajo es que, con las técnicas de IA, se puede simplificar el trabajo de los clínicos en UCIP, mostrándoles el comportamiento individual de cada paciente, con diferentes patologías en estado agudo, para que los clínicos puedan tener un panorama más amplio de la respuesta fisiológica de los pacientes críticos y, de esta manera, realizar intervenciones más oportunas. Se deben combinar varios modelos de IA, porque lo que está claro es que la confianza en una sola prueba puede dar una representación engañosa de la complejidad fisiológica. Vale resaltar también que hasta el momento la interpretación de estos resultados por el clínico experto es importante para contribuir con el modelamiento de la predicción.

4.1 FUTUROS TRABAJOS

Integrar los diferentes resultados que fueron satisfactorios encontrados en estudio en una máquina local o en un servidor hospitalario, o también podría implementarse en la nube, se sabe que la implementación en la nube para el entrenamiento de modelos puede ayudar a mejorar el rendimiento del modelo, además de reducir el tiempo de entrenamiento [67].

La unidad de cuidados intensivos es un departamento complejo y rico en información. Los pacientes ingresados en la UCI requieren un seguimiento estrecho y continuo debido a la alta gravedad de la enfermedad y al potencial de progresión rápida de la enfermedad. Los pacientes de la UCIP son heterogéneos, a menudo padecen múltiples problemas concurrentes y son menos numerosos que los pacientes que acuden a especialidades de un solo sistema. Por lo tanto, la investigación está necesariamente limitada tanto por la heterogeneidad como por el pequeño número de muestras. Por esta y otras razones, la base de pruebas para la práctica asistencial está menos desarrollada que para otras especialidades agudas. La combinación única de ricas fuentes de datos de la monitorización y una población de pacientes compleja y heterogénea hace que el entorno de la UCIP sea particularmente adecuado para la implementación de un sistema asistido basado en datos que analiza grandes cantidades de datos sin procesar que los expertos humanos podrían pasar por alto [76]. El uso de datos de la UCIP en la predicción temprana de la mortalidad es un área abierta atractiva para la investigación.

● 5. Conclusiones y recomendaciones

5.1. Conclusiones

1. Las técnicas de minería de datos empleadas como ACP, análisis de series de tiempo y la aplicación de modelos autorregresivos fueron útiles en la etapa de preprocesamiento de los datos, como parte de un análisis exploratorio de los datos. Pero cuando estas técnicas fueron utilizadas como modelos predictivos, se observó que en esta serie de datos los resultados predictivos eran bajos.
2. Del análisis de las diferentes técnicas de clasificación empleadas en esta serie de datos, se concluye que tuvieron limitaciones para identificar las clases inusuales, que en esta base de datos fueron los estados críticos, las cuales son el interés principal de este trabajo. No presentaron una diferencia estadística significativa para decir que un algoritmo es mejor que el otro.
3. El análisis con técnicas de conglomerados, resulta ser una herramienta valiosa para determinar perfiles, relaciones indirectas entre los datos, encontrar registros con valores normales o anormales, reducción de dimensiones.
4. La combinación de diferentes técnicas de minería de datos es de gran ayuda para poder hallar mejores modelos que nos ayuden a entender el comportamiento individual de los pacientes en estado crítico.
5. Hemos descrito la importancia del análisis de conglomerados para identificar los diferentes estados clínicos por los que atraviesan los pacientes críticos de esta base de datos y con este análisis se encontraron las salidas para continuar construyendo otros modelos, con la utilización de estas salidas describimos la

importancia de evaluar el comportamiento clínico integrando los signos vitales individualmente con cadenas de Márkov. Hemos demostrado la ventaja de usar HMM y cómo puede servir en caso de predicción de estados clínicos anormales.

6. El modelo de predicción utilizado HMM posee aún limitaciones que son detectadas con otros clasificadores y en otros estudios, como lo es el detectar las clases inusuales. Hay que continuar explorando sobre las limitaciones en las técnicas de clasificación para la identificación de los estados inusuales.
7. Las señales fisiológicas en cuidado intensivo necesitan un proceso de preprocesamiento, aún no pueden ser tomadas directamente para ser analizadas, necesitan ser procesadas para luego ser interpretadas. En este trabajo en la fase de preprocesamiento de los datos se perdieron muchas instancias, quedándonos sólo con 94,678.
8. Cualquiera que sea la técnica utilizada para la detección de los diferentes estados clínicos, es necesario hacer una labor de la mano de los clínicos expertos para su correcta interpretación.
9. En la construcción de los modelos, es necesario determinar el conjunto de entrenamiento, validación y pruebas con variables significativas que puedan arrojar un resultado confiable y no elemental.
10. En los estudios con pacientes críticos, es importante que en las métricas de validación de los modelos propuestos no perdamos el objetivo de una adecuada detección de los falsos negativos y limitar falsos positivos.
11. Es importante aportar iniciativas para solucionar el manejo de pacientes cada vez más complejos, ofrecerle a estos pacientes una medicina de precisión que cada vez más está avanzando, pero con muchos retos por delante que requieren la adición de herramientas analíticas útiles, tecnologías, bases de datos y enfoques para aumentar de manera eficiente las redes y la interoperabilidad de los sistemas clínicos, de laboratorio y de salud pública, así como abordar los problemas éticos y

sociales relacionados con la privacidad y la protección de los datos de salud con un equilibrio efectivo.

12. Las técnicas de IA y ML tienen el potencial de revolucionar el campo de la medicina y mejorar la calidad y la transición de la atención médica al analizar de manera inteligente los datos clínicos estructurados disponibles en gran cantidad y volumen, lo que plantea desafíos sin precedentes en el almacenamiento de datos, procesamiento, intercambio y curación, y desarrollo de una mejor comprensión de la fisiología.

5.2. Recomendaciones

Es necesario continuar fomentando el fortalecimiento de estudios encaminados al desarrollo de monitores inteligentes de aprendizaje automático que sean capaces de evaluar continuamente la respuesta humana a una enfermedad crítica con un alto grado de certeza. El desarrollo de estos monitores proporcionará el conocimiento y la experiencia necesaria para la creación de la UCIs semiautónoma, un entorno donde las máquinas inteligentes brindan la mayor parte de la atención que brindan los humanos en la actualidad. La observación de la monitorización continua en el paciente críticamente enfermo, actualmente es una tarea que es difícil para el ser humano, porque se pasan eventos clínicos sutiles para el ser humano, pero también es cierto que el potencial completo de la IA se hará realidad una vez que se convierta en un complemento clínico confiable para los intensivistas. Al ayudarnos a lidiar con la sobrecarga de información, las máquinas dotadas de inteligencia artificial pueden permitir que nuestras habilidades humanas de reflexión, imaginación y compasión pasen a primer plano al cuidar a otros seres humanos en peligro. El futuro de la IA en la UCI es realmente brillante. Como ocurre con todas las nuevas tecnologías, habrá fanáticos y fariseos, altibajos, euforias y desilusiones, así como espinosos dilemas éticos. Sin embargo, no tengo ninguna duda de que la IA llegó para quedarse, y nos corresponde familiarizarnos con esta tecnología para el mejoramiento de nuestros pacientes.

Bibliografía

1. I.Wheatley, «The nursing practice of taking level 1 patient observations, »*Intensive and Critical Care Nursing*, 22(2):115-21. 2006. doi: 10.1016/j.iccn.2005.08.003
2. B. J. Idar. P. H. Lars, P. B. Pedersen. K. John y M. Brabrand, «The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review, » *PLOS ONE*, 14(1): e0210875, pp.1-13. 2019. doi.org/10.1371/journal.pone.0210875
3. A.C. Malcolm Elliott, «Critical care: the eight vital signs of patient monitoring, » *British Journal of Nursing*, 21(10):621-5. 2012. DOI: 10.12968/bjon.2012.21.10.621
4. Glen Wright Colopy, Member, Stephen J. Roberts, Member, and David A. Clifton. Bayesian Optimization of Personalized Models for Patient Vital-Sign Monitoring, *IEEE Journal of Biomedical and Health Informatics*. 22(2), pp. 301-310, 2018. DOI 10.1109/JBHI.2017.2751509.
5. K. Abdur, Rahim Mohammad Forkan. PEACE-Home: Probabilistic estimation of abnormal clinical events using vital sign correlations for reliable home-based monitoring, » *Pervasive and Mobile Computing*, 38(2), pp. 296-311. 2017. dx.doi.org/10.1016/j.pmcj.2016.12.009
6. Sidney Le, Jana Hoffman, Christopher Barton, Julie C Fitzgerald, Angier Allen, Emily Pellegrini et al. Pediatric Severe Sepsis Prediction Using Machine Learning". In: *Front. Pediatr* 7:413. 2019. doi: 10.3389/fped.2019.0041
7. Haoran Xu Peiyao Li, Zhicheng Yang, Xiaoli Liu, Zhao Wang, Wei Yanet al. Construction and Application of a Medical-Grade Wireless Monitoring System for Physiological Signals at General Wards". In: *J Med Syst* 44.10 p. 182. 2020. doi.org/10.1007/s10916-020-01653-z.
8. Ryo Ueno, Liyuan Xu, Wataru Uegami. Hiroki Matsui, Jun Okui, Hiroshi Hayashi et al. Value of laboratory results in addition to vital signs in a machine learning

- algorithm to predict in-hospital cardiac arrest: A single-center retrospective cohort study". In: PLoS One 15: (7). e0235835. 2020. doi.org/10.1371/journal.pone.0235835.
9. Julio Frenk, Enrique Ruelas, Adriana Velázquez Guía tecnológica No.13: Monitor de signos Vitales, Cenetec, México, 2005.
 10. COFEPRIS: Comisión Federal para la protección de riesgos sanitarios [Internet]. México: Secretaría de Salud; 5 julio 2001-2020 [citado: 2020 Nov 5]- disponible: [https:// www.cofepris.salud.gob.mx](https://www.cofepris.salud.gob.mx).
 11. IMDRF: International Medical Device Regulators Forum [Internet]. Global Harmonization Task Force 2011-2020 [citado: 2020 Nov 5]. Disponible: <https://www.ghrf.org>.
 12. FORTRAN: *The IBM Mathematical Formula Translating System [internet]* Softward Preservation Group. Computer History Museum's, 13 08 2017. [citado 5 octubre 2020]. Available: <http://www.softwarepreservation.org/projects/FORTRAN/>.
 13. Ricardo A, Samsom MD, Stephen M et al. Pediatric Advanced Life Support Study Guide, American Heart Association. Texas 75149. Jones & Bartlett Learning, 2018. ISBN:978-1-669-623-8.
 14. J. Meera, A. Hutan, A. Lisa. K. Sadia, A. Sonal y C. Graham, Wearable sensors to improve detection of patient deterioration, *Expert Review of Medical Devices*. 16(2); pp. 145-154, 2019.
 15. Wongeun Song, Se Young Jung, Hyunyoung Baek, Chang Won Choi, Young Hwa Jung, Sooyoung Yoo. A Predictive Model Based on Machine Learning for the Early Detection of Late-Onset Neonatal Sepsis: Development and Observational study. *JMIR Med Inform*. 2020 Jul; 8(7): e15965.doi: 10.2196/15965.
 16. J. Kellett y F. Sebat. Make vital signs great again—A call for action. *Eur J Intern Med*, vol. 45(Supplement), nº C, p. 13–9. 2017.
 17. J. Kellett, M. De Vita, K. Hillman, R. Bellomo y D. Jones. The Assessment and Interpretation of Vital Signs. *Textbook of Rapid Response Systems: Concept*

- and Implementation*, Switzerland, Springer International Publishing, 2017, p. 63–85.
18. J. B. Cabello y V. Ruiz, «Critical Appraisal Skills Programme español, 01 01 1998. [En línea]. Available: <http://www.redcaspe.org/herramientas/instrumentos>. [Último acceso: 6 11 2020].
 19. Oxford Centre for Evidence-based Medicine. Levels of evidence. May 2001. Produced by Phillips B, Ball C, Sackett D, et al., since November 1998. Disponible en: <http://163.1.96.10/docs/levels.html#levels>. Consultado nov 1 de 2019.
 20. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*. 2015; 4(1):1. <https://doi.org/10.1186/2046-4053-4-1>.
 21. Chung HU, Kim BH, Lee JY, Lee J, Xie Z, Ibler EM, et al. Binodal, wireless epidermal electronic systems with in-sensor analytics for neonatal intensive care. *Science*. 2019 Mar 1;363(6430): eaau0780. doi: 10.1126/science.aau0780.
 22. Kwizera A, Kisson N, Musa N, Urayeneza O, Mujyarugamba P, Patterson AJ et al. “Sepsis in Resource-Limited Nations” Task Force of the Surviving Sepsis Campaign. A Machine Learning-Based Triage Tool for Children With Acute Infection in a Low Resource Setting. *Pediatr Crit Care Med*. 2019 Dec;20(12):e524-e530. doi: 10.1097/PCC.0000000000002121.
 23. Matam BR, Duncan H, Lowe D. Machine learning based framework to predict cardiac arrests in a paediatric intensive care unit: Prediction of cardiac arrests. *J Clin Monit Comput*. 2019 Aug;33(4):713-724. doi: 10.1007/s10877-018-0198-0.
 24. David Castiñeira, Katherine R Schlosser, Alon Geva, Amir R Rahmani, Gaston Fiore, Brian K Walsh, et al. Adding Continuous Vital Sign Information to Static Clinical Data Improves the Prediction of Length of Stay After Intubation: A Data-Driven Machine Learning Approach. *Respiratory Care* September 2020, 65 (9) 1367-1377; DOI: <https://doi.org/10.4187/respcare.07561>.

-
25. Dagdanpurev, Sumiyakhand Abe, Shigeto Sun, Guanghao Nishimura, Hidekazu Choimaa, Lodoiravsal Hakozaki, et al. A novel machine-learning-based infection screening system via 2013-2017 seasonal influenza patients' vital signs as training datasets. *Journal of Infection*. 2019. 78. 10.1016/j.jinf.2019.02.008.
 26. Eytan D, Jegatheeswaran A, Mazwi ML, Assadi A, Goodwin AJ, Greer RW, et al. Temporal Variability in the Sampling of Vital Sign Data Limits the Accuracy of Patient State Estimation. *Pediatr Crit Care Med*. 2019 Jul;20(7): e333-e341. doi: 10.1097/PCC.0000000000001984.
 27. G. Seidel, S. Murthy, C. Peters, P. Rostalski and M. Görges. Feasibility of Automated Vital Sign Instability Detection in Children Admitted to the Pediatric Intensive Care Unit. *2019 Computing in Cardiology (CinC)*, Singapore, Singapore, 2019, pp. Page 1-Page 4, doi: 10.23919/CinC49843.2019.9005547.
 28. H. Singh Ravneet Kaur; Abhilash Gangadharan; Ashish Kumar Pandey; Ashray Manur; Yao Sun *et al.*, "Neo-Bedside Monitoring Device for Integrated Neonatal Intensive Care Unit (iNICU)," in *IEEE Access*, vol. 7, pp. 7803-7813, 2019. doi: 10.1109/ACCESS.2018.2886879.
 29. Clark, M., Vergales, B., Paget-Brown, A., Terri J. Smoot, Douglas E. Lake, John L. *et al.* Predictive monitoring for respiratory decompensation leading to urgent unplanned intubation in the neonatal intensive care unit. *Pediatr Res* **73**, 104–110 (2013). <https://doi.org/10.1038/pr.2012.155>.
 30. Kim, S.Y., Kim, S., Cho, J, Young Suh Kim., In Suk Sol., Youngchul Sung, *et al.* A deep learning model for real-time mortality prediction in critically ill children. *Crit Care* **23**, 279 (2019). <https://doi.org/10.1186/s13054-019-2561-z>.
 31. Ying Zhang, MEng. Real-Time Development of Patient-Specific Alarm Algorithms for Critical Care. Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France August 23-26, 2007.
 32. Brekke IJ, Puntervoll LH, Pedersen PB, Kellett J, Brabrand M. The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic

- review. PLOS ONE. 2019. 14(1): e0210875.
<https://doi.org/10.1371/journal.pone.0210875>
33. Medic G, Kosaner Klie M, Atallah L, Louis Atallah, Jochen Weichert, Saswat Panda *et al.* Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review *F1000Research* 2019, **8**:1728
<https://doi.org/10.12688/f1000research.20498.1>
34. Sprogis, SK, Currey, J, Considine, J. Patient acceptability of wearable vital sign monitoring technologies in the acute care setting: A systematic review. *J Clin Nurs*. 2019; 28: 2732– 2744. <https://doi.org/10.1111/jocn.14893>.
35. Muaddi Alharbi, Nicola Straiton, Sidney Smithb, Lis Neubeck, Robyn Gallagher. Data management and wearables in older adults: A systematic review. *Maturitas*;123 pp 100-110- 2019.
<https://doi.org/10.1016/j.maturitas.2019.03.012>.
36. M Harford, J Catherall, S Gerry, JD Young, P Watkinson. Availability and performance of image-based, non-contact methods of monitoring heart rate, blood pressure, respiratory rate, and oxygen saturation: a systematic review. *Physiol. Meas;* **40** (6). 2019.
<https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-017-0615-3>.
37. Acharya S, William M. Mongan, Ilhaan Rasheed, Yuqiao Liu, Genevieve Dion, Adam Fontecchio, et al. Ensemble Learning Approach via Kalman Filtering for a Passive Wearable Respiratory Monitor". In: *IEEE J Biomed Health Inform* 23(3): 1022–1031. 2019. doi:10.1109/JBHI.2018.2857924.
38. Heather P. Duncan, Balazs Fule, Iain Rice, Alice J. Sitch, David Lowe. Wireless monitoring and real-time adaptive predictive indicator of deterioration. In: *Sci Rep* 10: 11366. 2020. doi.org/10.1038/s41598-020-67835-4.
39. Goto T, Carlos A. Camargo, Mohammad Kamal Faridi, Robert J. Freishtat, Kohei Hasegawa. Machine Learning-Based Prediction of Clinical Outcomes for Children During Emergency Department Triage. In: *JAMA Netw Open* 2(1), e186937-e186937. 2019. doi:10.1001/jamanetworkopen.2018.6937-
40. Ledys Izquierdo, Luis Fernando Niño, Jhon Sebastian Rojas. Modeling the vital sign space to detect the deterioration of patients in a pediatric intensive care

- unit. Proc. SPIE 11583, 16th International Symposium on Medical Information Processing and Analysis, 115830Q (3 November 2020); doi: 10.1117/12.2579629.
41. L.Roldán, S. Fachelli. Metodología de la Investigación Social Cuantitativa. 1,Ed Bellaterra: Universitat Autònoma de Barcelona. 2015. <https://ddd.uab.cat/record/129382>. [Consulta: 14 febrero 2021].
 42. S. de la fuente Fernández. Análisis de Componentes Principales. [Internet]. Escrito por S. Fernández, 2011. [citado feb 14 2021]. Disponible <https://docplayer.es/74190975-Santiago-de-la-fuente-fernandez-analisis-componentes-principales.html>.
 43. J.E. Briceño. Principio de las comunicaciones, Procesamiento digital de señales. 3 ed. Universidad de los Andes, Mérida Venezuela. 2012. Disponible https://www.academia.edu/16344758/An%C3%A1lisis_de_Fourier?email_work_card=reading-history.
 44. P. Prandoni, M. Vetterli. Signal processing for communications. 1 Ed. Centre Midi Italy. Springer; 2008. ISBN 978-2-940222-20-9 (EPFL Press). 1-367 p.
 45. Conceptos fundamentals de series de tiempo. [Internet]. Escrito por German Aneiros. 2008-2009. [citado 14 feb 2021]. Disponible <http://www.ptolomeo.unam.mx> › jspui › bitstream. <https://bopdf.info/edoc/3a4dc81/avohk-5k-series-2018-race-4-series-results-xlsx>
 46. Brockwell, P.J, Davis, R.A. Introduction to Time Series and Forecasting. 2da Ed. Madrid: Springer; c2007. ISBN: 978-0-387-21657-7. 1-273 p.
 47. Brockwell, Peter J, Davis, Richard A. Introduction to Time Series and Forecasting. Third ed. Switzerland: Springer; 2016. ISBN 978-3-319-29852-8. 1-419p.
 48. Fournies, Aldo. Modelos ARMA y Box and Jenkins. Universidad Técnica Federico Santa María. Chile. 2015. 10.13140/RG.2.1.2907.0883.
 49. Pang Ning T, Steinbach M, Kumar B. Introduction to Data Mining. 2nd ed. Boston: Pearson Education, Inc; 2018. ISBN-10: 0321321367. 427p.
 50. S. de la fuente Fernández. Análisis conglomerados. [Internet]. Escrito por S. Fernández, 2011. [citado feb 14 2021]. Disponible

https://www.academia.edu/32046069/An%C3%A1lisis_Conglomerados_Santiago_de_la_Fuente_Fern%C3%A1ndez

51. Orellana L. Análisis de Clúster ing K-means para la Base de Datos ZOO. [Master's tesis]. [Chile]: Universidad Santiago de Chile; 2017. 48p.
52. A. Fahad, Najlaa Alshatri, Zahir Tari, Abdullah Alamri, Ibrahim Khalil; Albert Y. Zomaya *et al* A Survey of Clúster ing Algorithms for Big Data: Taxonomy and Empirical Analysis. *IEEE Transactions on Emerging Topics in Computing*; 2(3): pp. 267-279, 2014. doi: 10.1109/TETC.2014.2330519.
53. Mangudo C. Two Step Clúster en SPSS y técnicas relacionadas. [Master's tesis]. [Salamaca]: Universidad de Salamanca; 2015. 161p.
54. Gen Li, Lu Sun. Characterizing Heterogeneity in Drivers Merging Maneuvers Using Two-Step Clúster Analysis. *Journal of Advanced Transportation*, vol. 2018, Article ID 5604375, 15 pages, 2018. <https://doi.org/10.1155/2018/5604375>.
55. Orellana L. Análisis de Redes Neuronales para la Base de Datos ZOO Utilizando la Herramienta de Software "WEKA". [Master's tesis]. [Chile]: Universidad Santiago de Chile; 2017. 38p.
56. H: Rivera, S. Zuñiga, L. Vera, L. Meneses, A. Escudero. Métodos de clasificación; minería de datos; datos meteorológicos, Número 20, Vol.2: 2018) Perfiles Revista Científica. ISSN 2477-9105. <http://dspace.esPOCH.edu.ec/handle/123456789/9395>.
57. Ocaña-Riola, R. Modelos de Márkov aplicados a la investigación en ciencias de la salud. *Interciencia* 34(2009):157-162.
58. Lindsey JK. *Statistical analysis of stochastic processes in time*. Cambridge University Press. Cambridge, 2004. RU. 338 pp.
59. Karatzas I, Schreve SE. *Brownian Motion and Stochastic Calculus*, 2a ed. Springer. Nueva York, NY, EEUU. 1991. 470 pp.
60. Ocaña-Riola R. Márkov processes for biomedical data analysis. En Conn M (Ed.) *Source Book of Models for Biomedical Research*. Humana. Nueva Cork, NY, EEUU. 2008. pp. 739-745.
61. GEO tutoriales: Gestión de Operaciones [Internet]. Ejercicios cadenas de Márkov; 2011-2020 [citado: Noviembre 6 2020] Disponible

- <https://www.gestiondeoperaciones.net/cadenas-de-Márkov/cadenas-de-Márkov-ejercicios-resueltos/>.
62. Juan Tornero Lucas. Machine Learning: Modelos Ocultos de Márkov (HMM) y Redes Neuronales Artificiales (ANN). [Master's thesis]. [Barcelona]. Universitat de Barcelona; 2017. 53p.
 63. Pearl, J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco, Ca, USA. Morgan Kaufmann Publishers Inc: San Mateo. 1988. 552p. ISBN:978-1-55860-479-7.
 64. Rabiner, L.R. 1989. A tutorial on hidden Márkov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2): 257-286. Doi:10.1109/5.18626.
 65. Toolbox for Matlab: Hidden Márkov Model (HMM) Toolbox for Matlab [Internet]. Written by Kevin Murphy, 1998. 2005-2020 [citado nov 6 2020]. Disponible: www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html.
 66. F. Hernando. Técnicas de procesado y representación de la señal de voz para el reconocimiento del habla en ambientes ruidosos con técnicas de Modelos Ocultos de Márkov. [internet] [Master's thesis]. [Barcelona]: Universidad Politécnica de Cataluña 1993. P203.
 67. Sayandeep Acharya, William M. Mongan, Ilhaan Rasheed, Yuqiao Liu, Endla Anday, Genevieve Dion et al. Ensemble Learning Approach via a Kalman Filtered Márkov Model for a Passive Wearable Respiratory Monitor. IEEE J Biomed Health Inform. 23(3): 1022–1031. 2019. doi:10.1109/JBHI.2018.2857924.
 68. Aditya Nagori, Lovedeep Dhingra, Ambika Bhatnagar, Rakesh Lodha, Tavpritesh Sethi. Predicting Hemodynamic Shock from Thermal Images using Machine Learning. Scientific reports. 9:91. 2019. DOI:10.1038/s41598-018-36586-8.
 69. Ioan Stanculescu, Christopher K. I. Williams, and Yvonne Freer Autoregressive Hidden Márkov Models for the Early Detection of Neonatal Sepsis. IEEE journal of Biomedical and Health Informatics; 18(5)2014. Doi: 10.1109/JBHI.2013.2294692.
 70. Guoxian Yu, Xianxue Yu, Jun Wang. Network-aided Bi-Clustering for discovering cancer subtypes. Scientific Reports. 7: 1046. 2017. DOI:10.1038/s41598-017-01064-0.

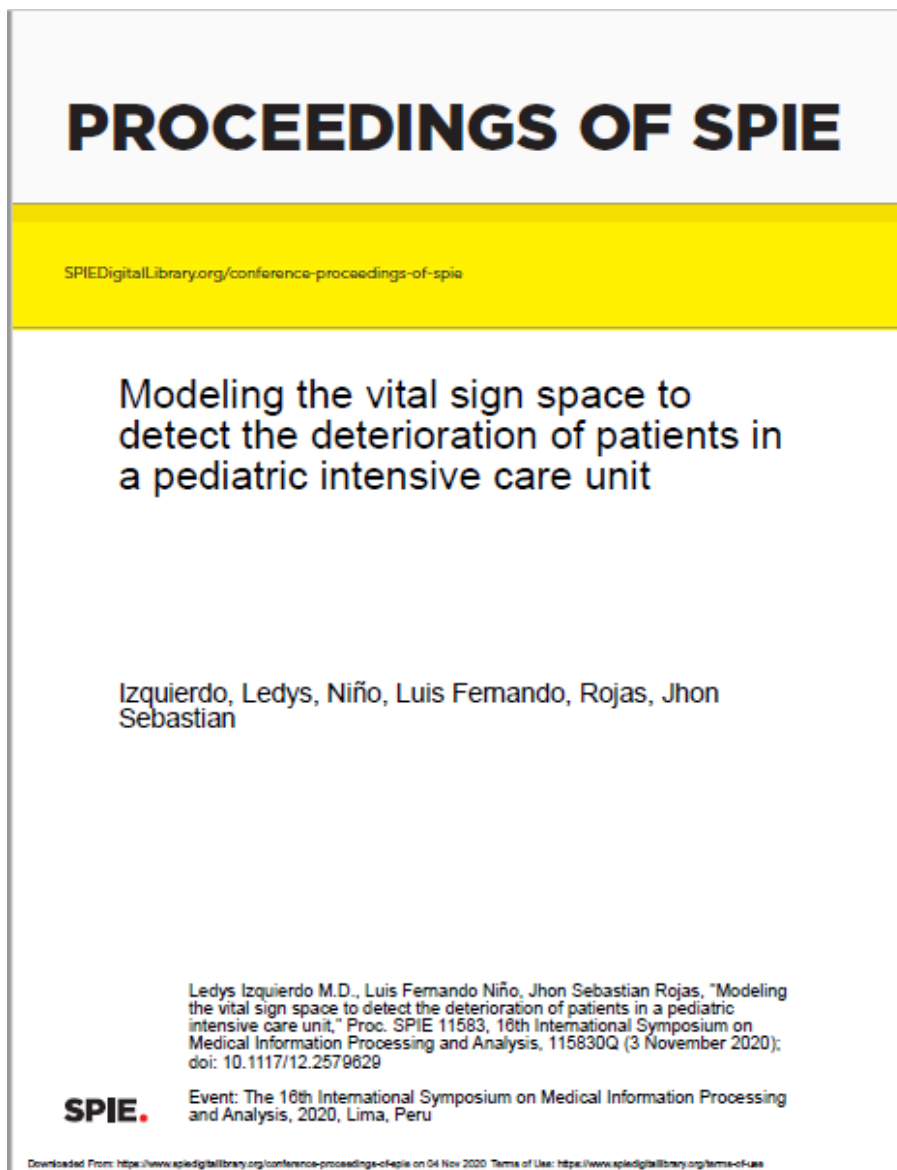
71. G. Tzortzis, A. Likas. The Global Kernel k-Means Clustering Algorithm. IEEE world congress on computational intelligence, neural networks. 2008. July 8. Hong Kong, c2008. pp1977 - 1984. 10.1109/IJCNN.2008.4634069.
72. B. Said, A. Al-Sa'd, M. Tlili, M. Abdellatif, A. Mohamed, A. El-Fouly, et al. A deep learning approach for vital signs compression and energy efficient delivery in mHealth systems. IEEE Access. PP. 1-1. 2018. 10.1109/ACCESS.2018.2844308.
73. U. Acharya, P. Joseph, N. Kannathal, C. Min Lim, J. S. Suri. Heart rate variability: a review. Med Bio Eng Comput; 44:1031–1051. 2006. DOI 10.1007/s11517-006-0119-0.
74. B. Yi Hao Wee, J. H. Lee, Y. H. Mok, S-L. Chong. A narrative review of heart rate and variability in sepsis. Ann Transl Med;8(12):768. 2020. dx.doi.org/10.21037/atm-20-148.
75. I. Badke, L. Marsillio, D. Weese-Mayer, L. Sanchez-Pinto. Autonomic Nervous System Dysfunction in Pediatric Sepsis. Frontiers in Pediatrics; 6:280. 2018. doi: 10.3389/fped.2018.00280
76. IBM Watson Machine Learning. [Internet]. U.S.: IBM Corporation 1994, [Last Updated: October 2020; cited Nov 6 2020] disponible: <https://www.ibm.com/cos/cloud/machine-learning/pricing>

A. Anexo: Escala MInCir para estudios de pronóstico

Tabla 1. Escala MInCir para estudios de pronóstico (MInCir Pr).

	1	5
 dominio 1. (Variable de desenlace)		
Se define la variable de desenlace		
El método para medir la variable de desenlace es válido y confiable		
Se define las unidades de medida de la variable de desenlace		
Una adecuada proporción de la muestra tiene datos completos de la variable de desenlace		
El método y las unidades de medición de la/las variables de desenlace son los mismos para todos los pacientes		
 dominio 2. (Variable de exposición)		
Se define la variable de exposición		
El método para medir la variable de exposición es válido y confiable		
Se define las unidades de medida de la variable de exposición		
Una adecuada proporción de la muestra tiene datos completos de la variable de exposición		
El método y las unidades de medición de la/las variables de exposición son los mismos para todos los pacientes		
 dominio 3. (Seguimiento)		
Se enuncia el porcentaje de pérdida de seguimiento		
Se describen las características de los sujetos perdidos		
Se hicieron esfuerzos por recolectar la información de los sujetos perdidos		
Las causas de pérdida de seguimiento son descritas		
Se analizaron las diferencias entre las características de los sujetos con pérdida de seguimiento y los que no		
 dominio 4. (Análisis y conclusiones)		
Con los datos presentados es posible determinar las medidas de riesgo y asociación		
Se realizan modelos de asociación o predictivos		
El modelo elegido es adecuado según el diseño del estudio		
Las conclusiones del estudio son coherentes con el objetivo y la metodología empleadas		
 dominio 5. (Diseño y variables de confusión)		
Se enuncia claramente el diseño del estudio		
Considera variables de confusión		
Se definen adecuadamente las variables de confusión		
Los métodos para medir las variables de confusión son válidos y confiables		
 dominio 6. (Tamaño de la muestra)		
Hay cálculo del tamaño de la muestra		
Se justifica el tamaño de la muestra		
 Total		

B. Anexo: Artículo Publicado conferencia



Se puede consultar la base de datos en el siguiente dominio (consult database critically ill pediatric patients in PICU.csv. " Kaggle,doi: 10.34740 / KAGGLE / DSV / 1518232)

Modeling the vital sign space to detect the deterioration of patients in a pediatric intensive care unit

Ledy's Izquierdo^a, Luis Fernando Niño^a, and Jhon Sebastian Rojas^a

^aLaboratorio de Investigación en Sistemas Inteligentes (LISI), – Depto. de Ingeniería de Sistemas e Industrial. Facultad de Ingeniería, Universidad Nacional de Colombia, Bogotá, Colombia

ABSTRACT

In the field of continuous vital-sign monitoring in critical care settings, it has been observed that the "early-warning signs" of impending physiological deterioration can fail to be detected timely and sometimes by resource-constrained clinical staff. This effect may be escalated by the "data deluge" caused by acquisition of more complex patient data during routine care. The objective of this study is to develop a probabilistic model for predicting the future clinical episodes of a patient using observed vital sign values prior to a clinical event. Vital signs (e.g. heart rate, blood pressure) are used to monitor a patient's physiological functions and their simultaneous changes indicate transitions between patient's health states. If such changes are abnormal then it may lead to serious physiological deterioration. The CRISP-DM (CRoss-Industry Standard Process for Data Mining) methodology was used as a data mining process and then we used Markov chains to identify the clinical states through which the patient passes. Then, a Hidden Markov model (HMM) based approach was applied for classification and prediction of patient's deterioration by computing the probability of future clinical states. Both learning models were trained and evaluated using six vital signs data from 94,678 records from 90 patient, collected from the database of real patients who were in the Pediatric Intensive Care Unit of the Central Military Hospital in the city of Bogotá, Colombia. The proposed technique based on monitoring multiple physiological variables showed promising results in early identifying the deterioration of critically ill patients

Keywords: Vital signs, Hidden Markov model, Clinical event, pediatric critical care

1. INTRODUCTION

The importance of monitoring vital signs in clinical practice is undisputed, but it is still unclear how best to monitor and interpret them and the frequency with which they should be measured [1]. Thus, this motivated us to search the literature for studies that explicitly attempt to determine and quantify the increase or decrease in risk associated with changes in vital signs, measured intermittently or continuously [2].

Since around 2000 there has been a growing interest in retrieving data and extracting knowledge from large volumes of data, through the efficient application of data analysis techniques and different models to try to assess the behavior of vital signs in both the hospitalized and in the patients outside the hospital, i.e. at home. It can be seen that the greatest efforts so far have been in trying to look for normal patterns in the vital signs of healthy patients [3][4], patients with chronic pathologies [5] and very few in critically ill patients [6][7]. It can also be seen that the studies to date have been carried out with patients to a greater extent in out-of-hospital settings [8], rather than in-hospital settings [9][10]. The in-hospital setting studies have been developed in the areas of emergency, hospitalization and intensive care [11][6][7]. Among these, the vast majority have been performed on adult patients [11] and very few in pediatric patients [12][13].

Today the admission of patients to the Pediatric Intensive Care Unit (PICU) is due to increasingly complex pathologies and nurses must not only know how to measure vital signs accurately, they must also know how to interpret and act on them [2][14]. Previous studies have also attempted to relate the variables of vital signs to search for predictors of patient deterioration. But, so far they have typically only related two variables.

Further author information: Intelligent Systems Research Laboratory (LISI), CR 45 No. 26-85, Bogotá, COL. Ledy's M. Izquierdo: E-mail: Izquierdo@una.edu.co

18th International Symposium on Medical Information Processing and Analysis, edited by Eduardo Romero
Natalia Laporte, Jorge Bileva, Marius Lingureu, Proc. of SPIE Vol. SPIE220, SPIE22000
© 2020 SPIE · CCC code: 0277-786X/20/0221 · doi: 10.1117/12.2579829

Proc. of SPIE Vol. SPIE220 115830Q-1

Downloaded From: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie> on 04 Nov 2020
Terms of Use: <https://www.spiedigitallibrary.org/terms-of-use>