

1. Contenido

2. ÍNDICE DE TABLAS	5
3. ÍNDICE DE ILUSTRACIONES.....	6
4. RESUMEN	7
5. ABSTRACT	9
6. INTRODUCCIÓN	10
6.1. Características y transmisión del SARS-CoV2.....	10
6.2. Factores de riesgo al cursar COVID-19.....	12
6.2.1. Obesidad	12
6.2.2. Diabetes.....	12
6.2.3. Afecciones cardíacas	13
6.2.4. Enfermedades mentales	13
6.3. Minería de Datos e Inteligencia artificial	13
6.3.1. Aprendizaje Automático (Machine Learning)	14
6.3.2. Selección de características	15
6.3.3. Modelos de predicción.....	16
6.4. Aplicación de Inteligencia Artificial en estudios de COVID-19.....	17
6.5. Desafío planteado.....	18
7. HIPÓTESIS.....	21
8. OBJETIVOS	21
8.1. Objetivo General.....	21
8.2. Objetivos Específicos	21
9. MATERIALES.....	22
9.1. Diseño del Estudio	22
9.2. Conjunto de datos	24
9.3. Herramientas de análisis de datos	25
10. MÉTODOS.....	26
10.1. Metodología objetivo específico 1: Configurar un set de datos a partir de información clínica, experimental y demográfica obtenida desde 60 pacientes que cursaron COVID-19 e individuos control.	26
10.1.1. Preparación de datos	26
10.1.2. Exploración de datos	27

10.2.	Metodología objetivo específico 2: Determinar las características más relevantes asociadas a los diferentes niveles de severidad de pacientes que cursaron COVID-19.....	28
10.2.1.	Selección de características	28
	Feature Selection	28
	Mutual information	28
	ANOVA (Analysis of Variance)	29
	SHapley Additive exPlanation	29
10.3.	Metodología objetivo específico 3: Evaluar diferentes modelos de Inteligencia Artificial para clasificar grado de severidad y nivel de secuela en pacientes que cursaron COVID-19.	30
10.3.1.	Modelos de aprendizaje automático a implementar.	30
	Modelos de clasificación	30
	Optimización de hiperparámetros	32
10.3.2.	Entrenamiento de algoritmos de aprendizaje automático supervisado.	33
10.3.3.	Validación de algoritmos de aprendizaje supervisado.	33
	Métricas de clasificación	33
11.	RESULTADOS.....	34
11.1.	Resultados objetivo específico 1: Configurar un set de datos a partir de información clínica, experimental y demográfica obtenida desde 60 pacientes que cursaron COVID-19.	34
11.1.1.	Sets de datos	35
11.1.2.	Clase objetivo	35
11.2.	Resultados objetivo específico 2: Determinar las características más relevantes asociadas a los diferentes grados de severidad y nivel de secuela en pacientes que cursaron COVID-19.	36
11.2.1.	Set completo	36
11.2.2.	Subsets.....	38
11.3.	Resultados objetivo específico 3: Evaluar diferentes modelos de Inteligencia Artificial para clasificar grado de severidad y nivel de secuela en pacientes que cursaron COVID-19.	45
11.3.1.	Grado de severidad	45
	Set Completo.....	45
	Subset TAC	47
	Nivel de secuela	50
	Set Completo.....	50
	Set Espirometría.....	52
12.	DISCUSIÓN.....	56
12.1.	Discusión resultados objetivo específico 1: Configurar un set de datos a partir de información clínica, experimental y demográfica obtenida desde 60 pacientes que cursaron COVID-19.	56
12.2.	Discusión resultados objetivo específico 2: Determinar las características más relevantes asociadas a los diferentes grados de severidad y nivel de secuela en pacientes que cursaron COVID-19.	57
12.3.	Discusión resultados objetivo específico 3: Evaluar diferentes modelos de Inteligencia Artificial para clasificar grado de severidad y nivel de secuela en pacientes que cursaron COVID-19.	59

13.	CONCLUSIÓN.....	61
14.	BIBLIOGRAFÍA.....	62
15.	ANEXO.....	71
1.	Consentimiento informado.....	71
2.	Identificadores de pacientes	73
3.	Identificadores de características	74
4.	Características por subset de datos	76

2. ÍNDICE DE TABLAS

Tabla 1. Manifestaciones clínicas en tres niveles diferentes de severidad en COVID-19	11
Tabla 2. Resumen áreas de aplicación Inteligencia Artificial ³⁵	14
Tabla 3. Breve descripción de técnicas de predicción ⁵⁵	16
Tabla 4. Grado de severidad de los pacientes que padecieron la enfermedad COVID-19	27
Tabla 5. Subsets de datos y su dimensión.	35
Tabla 6. Clases objetivos	36
Tabla 7. Número de características reducidas por algoritmos de selección de características... ..	43
Tabla 8. LOOCV accuracy – Set Completo, Random Forest para grado de severidad.....	45
Tabla 9. Métricas de evaluación por clase	46
Tabla 10. LOOCV accuracy. Set completo, XGBoost, grado de severidad.....	47
Tabla 11. Métricas de evaluación por clase	47
Tabla 12. LOOCV accuracy, Set TAC, modelo de clasificación XGBoost para grado de severidad	48
Tabla 13. Métricas de evaluación modelo de clasificación de secuela	49
Tabla 14. LOOCV accuracy – Set Completo, Random Forest para nivel de secuela	50
Tabla 15. Métricas de evaluación modelos de clasificación de secuela	51
Tabla 16. LOOCV accuracy – SubSet Espirometria, Random Forest para nivel de secuela.....	52
Tabla 17. LOOCV accuracy – SubSet Espirometria, XGBoost para nivel de secuela.....	52
Tabla 18. Métricas de evaluación modelos de clasificación de secuela	52
Tabla 19. Métricas de evaluación modelos de clasificación de secuela	53
Tabla 20. LOOCV accuracy – SubSet Completo, excluye TAC y Espirometria, modelo de clasificación Random Forest para nivel de secuela	55
Tabla 21. Métricas de evaluación modelos de clasificación de secuela	55

3. ÍNDICE DE ILUSTRACIONES

Ilustración 1. Aplicación de inteligencia artificial y aprendizaje automático en la lucha contra el COVID-19.	19
Ilustración 2. Diagrama de flujo de la metodología utilizada.	26
Ilustración 3. Gráfico Feature Importance - Set Completo	37
Ilustración 4. Feature importance subsets de datos para grado de severidad.....	41
Ilustración 5. Feature importance subsets de datos para nivel de secuela.....	42
Ilustración 6. Gráfico de SHAP – Set Completo.....	44
Ilustración 7. <i>Gráfico de curvas ROC, modelo de clasificación Random Forest para grado de severidad, Set Completo.</i>	<i>46</i>
Ilustración 8. <i>Gráfico de curvas ROC modelo de clasificación XGBoost para grado de severidad, Set completo.</i>	<i>48</i>
Ilustración 9. Gráfico de curvas ROC modelo de clasificación XGBoost para grado de severidad, subset TAC.	49
Ilustración 10. Gráfico de curvas ROC modelo de clasificación Random Forest para nivel de secuela..	51
Ilustración 11. Gráfico de curvas ROC modelo de clasificación Random Forest para nivel de secuela, Subset espirometria.	53
Ilustración 12. Gráfico de curvas ROC modelo de clasificación XGBoost para nivel de secuela, Subset espirometria..	54