

Desarrollo de sistema de Machine Learning para la predicción de vía aérea a partir de imagen facial con dispositivo móvil

Francisco Mendoza Garcés¹, Sofia Irigoyen Miró², Fernando García-García³, Joseba González-García⁴, Susana García-Gutierrez⁵, Sorkunde Telletxea Benguría⁶

¹ *Hospital Universitario Galdakao-Usansolo y Clínica Zorrozaurre, Servicio de Anestesia y Reanimación*

² *Hospital Urduliz, Servicio de Anestesia y Reanimación*

³ *BCAM — Basque Center for Applied Mathematics*

⁴ *Hospital Universitario Basurto, Servicio de Anestesia y Reanimación*

⁵ *Hospital Universitario Galdakao-Usansolo, Unidad de Investigación Clínica*

⁶ *Hospital Universitario Galdakao-Usansolo, Servicio de Anestesia y Reanimación*

Resumen:

El manejo de una vía aérea difícil (VAD) representa aún una causa importante de lesiones relacionadas con la anestesia, cuyas complicaciones son potencialmente mortales. El notable interés en la predicción de VAD ha provocado el desarrollo de modelos de predicción, algunos de los cuales ya incluyen algoritmos de Inteligencia Artificial a partir de imágenes.

Se realizó un estudio observacional, de cohortes prospectivo, en el que se tomaron imágenes de los pacientes sometidos a una anestesia general, recogiendo la información pre-anestésica así como la información post-intubación. Nuestro equipo desarrolló un algoritmo automático de detección de puntos faciales de cara a la toma de medidas de variables ya validadas de predicción de VAD, que se integraron con el modelo predictivo de Naguib.

La incidencia estimada de VAD en nuestra muestra de 503 pacientes fue de un 6,36%. La valoración subjetiva (pre-intervención) de los clínicos obtuvo una sensibilidad de 25.00%, con una especificidad de 93.63%. En comparación, nuestra herramienta alcanzó una sensibilidad del 53.12% y una especificidad del 79.83%. El AUC obtenida, o área bajo la curva ROC, fue de 0.680.

Integrando nuestro sistema de medición IA-ML con el modelo de Naguib, los resultados muestran que estamos cerca de igualar la capacidad predictiva del clínico. El potencial del análisis facial en la predicción de VAD nos anima a seguir investigando y a desarrollar modelos propios. Creemos que proporcionará al anesthesiólogo una herramienta de ayuda en la toma de decisiones automática, objetiva y accesible.

Introducción:

Hoy en día, hasta un 25-33% de las muertes atribuibles a la anestesia se relacionan con la incapacidad de ventilar o intubar al paciente (2). No obstante, predecir con fiabilidad estas situaciones es una tarea compleja. Se estima que aproximadamente el 10% de las intubaciones acaban resultando difíciles (3), y hasta el 93% de estas son no anticipadas. En la práctica clínica habitual, los anesthesiólogos deben estimar la dificultad de intubación mediante una serie de pruebas de cabecera (4). Sin embargo, diferentes revisiones sistemáticas y meta-análisis coinciden en señalar que estas pruebas, y otras similares, sufren de notables problemas de aplicabilidad: con capacidades de discernimiento limitadas, e incluso inconsistencias en la predicción (5).

En la literatura existen propuestas que consideran combinaciones simultáneas de varios de dichos indicadores para la predicción de VAD, como los métodos de regresión logística de Naguib (6); o Cuendet et al (7), quienes elaboraron un algoritmo de Machine Learning -

Inteligencia Artificial (ML-IA) a partir de 4 fotografías del paciente. Estos métodos se han comparado contra la dificultad efectiva de VAD, determinada según criterios IDS [Adnet] (8) y ASA [Apfelbaum] (9).

En este sentido, nuestro trabajo plantea una técnica alternativa en la detección VAD, a partir del análisis de imágenes del paciente, combinando técnicas de ML-IA (10) con el modelo clínico de Naguib (6).

Objetivos:

- Aplicar un algoritmo automático que reconozca puntos de referencia anatómicos relevantes a partir de imágenes del paciente y tome las medidas de las variables morfológicas a estudiar.
- Integrar modelo predictivo de VAD de Naguib (6) con dicho sistema de medición automático desarrollado por nuestro equipo de científicos de datos.

Material y métodos:

Entre los años 2018 a 2020 se realizó un estudio observacional, de cohortes y prospectivo, en el que participaron 503 pacientes intervenidos quirúrgicamente en el Hospital Universitario de Galdakao, Basurto y Clínica Zorrozaurre, en Bilbao, España. A todos ellos se les realizó una anestesia general y fueron intubados mediante laringoscopia directa con la pala de Macintosh.

Los pacientes que tomaron parte en el estudio cumplieron los siguientes criterios obligatorios de inclusión:

- a) Pacientes intervenidos de cualquier tipo de cirugía programada bajo anestesia general e intubación orotraqueal mediante laringoscopia directa.
- b) Edad entre 18 y 99 años de edad.
- c) Pacientes que hayan aceptado formar parte del estudio.
- d) Con el consentimiento informado firmado.

Fueron excluidos del estudio los pacientes con los siguientes criterios:

Cirugía de urgencias.

- a) Menores de 18 años y mayores de 99 años de edad.
- b) Pacientes con deterioro cognitivo o patología que les impida comprender en qué consiste el estudio.
- c) Pacientes con alteración anatómica que implica vía aérea difícil.
- d) Paciente con vía aérea difícil ya conocida en la que será necesario alguna técnica alternativa de intubación.
- e) Embarazadas.
- f) Pacientes con inmovilización cervical o patología cervical.
- g) Contraindicaciones para laringoscopia directa (imposibilidad o riesgo a la hiperextensión cervical, intubación difícil previa conocida, etc.).
- h) Pacientes intubados mediante métodos diferentes al laringoscopio convencional y pala de Macintosh.
- i) Negativa del paciente a formar parte del estudio.

El estudio fue aprobado por el Comité Ético de Investigación Clínica (CEIC). La información obtenida fue considerada confidencial. Los datos recogidos para el estudio estaban identificados mediante un código y solo el investigador principal podía relacionar dichos datos con el paciente y con su historia clínica.

El tratamiento de los datos se realizó con las medidas de seguridad establecidas en cumplimiento de la Ley Orgánica 15/1999 de Protección de Datos de carácter personal

El estudio se llevó a cabo de acuerdo a las recomendaciones para estudios de investigación hecha en seres humanos y siguiendo las Normas de Buena Práctica Clínica y la Declaración de Helsinki, revisada en Tokio, Venecia, Hong-Kong, Sudáfrica, Edimburgo y Seúl (en 2008) y en la actual Legislación Española en materia de investigación biomédica (Ley 14/2007).

La recogida de la imagen se realizó tanto en la unidad prequirúrgica como en el antequirófono el mismo día de la intervención. Las fotografías se tomaron en la aplicación móvil diseñada ex profeso, en la que se recogían igualmente los datos demográficos y antropométricos del paciente, así como la valoración de Vía Aérea realizada previamente en la consulta, que venía reflejada en la Historia Clínica Electrónica, frente a la que comparamos nuestra metodología.

Los pacientes disponían de autonomía para rechazar su participación en cualquier momento. Ningún paciente rechazó la participación en el estudio.

La realización de las imágenes se limitó a 4 investigadores, de cara a homogeneizar en la medida de lo posible la toma de imágenes y para disminuir la variabilidad inter-observador al mínimo.

A los pacientes seleccionados, se les realizaron dos fotografías, una frontal en la que se indicó que abrieran la boca y sacaran la lengua, lo que facilitaba el juicio de Mallampati y Apertura Bucal. Otra en proyección lateral, en la que se colocó a los pacientes con la cabeza en extensión vertical, y la boca cerrada, que es el procedimiento estándar para medir Distancia Tiro Mentoniana (DTM) y Distancia Esterno Mentoniana (DEM). Ambas fotos se realizaron con el paciente sentado, y en ningún caso se modificó el tipo de anestesia que estaba planificada realizar con el fin de incluirlos en el estudio.

La imagen fue adquirida de manera que se mostraran los puntos previamente definidos por el equipo clínico. Ya que no es un procedimiento rutinario en la práctica clínica habitual, a todos los pacientes se les explicó el procedimiento y pidió consentimiento informado, que firmaron previa a la toma de imágenes.

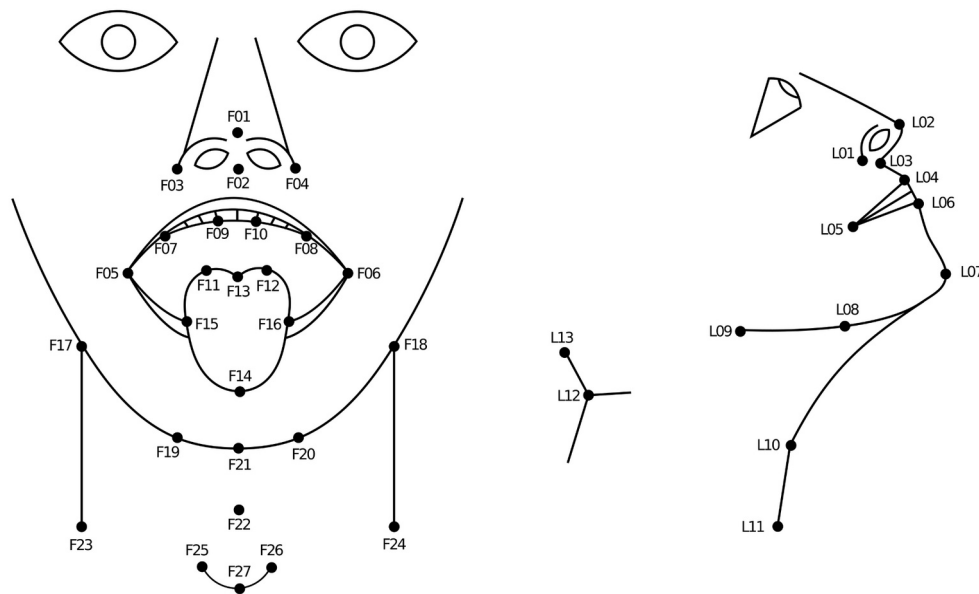


Figura 1: Definición de nuestros 'landmarks' orofaciales, para caracterizar la morfología de la vía aérea en el contexto de anestesia. 27 puntos en la vista frontal (izquierda) y 13 puntos en la vista lateral (derecha).

Se utilizaron teléfonos inteligentes con cámaras de uso general para tomar las fotos, y se agregó una tarjeta de señal al escenario (círculo negro de 25 mm de diámetro sobre fondo blanco), con el fin de que sirviera de referencia para las dimensiones físicas.

Después de la toma de imágenes, se trasladaba a los pacientes a quirófono según el circuito habitual, donde otro anestesiólogo experimentado inducía la anestesia general y tras la intubación, transmitía al IP y colaboradores las variables de ventilación HAN y los criterios de IDS-ASA de intubación difícil que se recogían igualmente en programa integrado en el dispositivo móvil. Puntualmente, los datos recogidos se volcaban en una base de datos para

su posterior análisis, de tal forma que la información de los pacientes siempre estaba almacenada en un disco duro.

En un trabajo previo (10), desarrollamos y evaluamos una herramienta basada en técnicas de Inteligencia Artificial (IA) - Machine Learning (ML), para la identificación automática de 'landmarks' (es decir, puntos de particular relevancia anatómica; como en la Figura 1) a partir de dos fotografías: una frontal y una lateral por cada paciente. Para ello, empleamos algoritmos de tipo 'Deep Learning', entrenados con la referencia del consenso entre dos anesthesiólogos [FMG, SIM], a modo de anotadores expertos. El método se programó mediante la funcionalidad Keras, dentro de la librería de 'Deep Learning' TensorFlow, versión 2.3.0.

En dicho trabajo previo, sobre un conjunto de $n=317$ individuos, se evaluó la capacidad de nuestros algoritmos IA-ML para determinar la visibilidad y localización de los 'landmarks' orofaciales en las imágenes, y se calculó su rendimiento (en modo validación cruzada) frente a las discrepancias inter-anotador, con el consenso entre ambos como referencia patrón. Véase como ejemplo la Figura 2, para más detalle.

Entre las dos metodologías 'Deep Learning' propuestas, la más compleja de ambas fue la que mejor rendimiento global obtuvo (10). Esta se comportó con una precisión estadísticamente comparable a la humana en la vista frontal; así como con una precisión ligeramente inferior, aunque con un tamaño de efecto no-significativo, en la vista lateral.

A partir de la ubicación de dichos 'landmarks' orofaciales, en este trabajo hemos aplicado métodos de posicionamiento de cámara y medición 3D; a fin de calcular distancias físicas relevantes de la morfología de la vía aérea, a saber: DTM, DEM, una estimación de la distancia interincisivos y del perímetro cervical. Para ello, se utilizó el lenguaje de programación Python, con la librería OpenCV, versión 4.5.4.

A continuación, estas mediciones individualizadas por paciente se integraron en un modelo de regresión logística ya validado, como el de Naguib et al. 2006, que calcula el siguiente puntaje predictivo de gravedad (6):

$$\text{score Naguib 2006} = 0.2262 - 0.4621 \cdot \text{thyromental distance} + 2.5516 \cdot \text{Mallampati score} - 1.1461 \cdot \text{interincisor gap} + 0.0433 \cdot \text{height}$$

estando la distancia tiro-mentoniana DTM ('thyromental distance'), la interincisivos ('interincisor gap') y la altura del paciente ('height') expresadas en centímetros, y el grado de Mallampati dicotomizado a valor 0 (grados I y II) o 1 (grados III y IV).

Finalmente, el sistema ajustó automáticamente el umbral óptimo de decisión, para maximizar la capacidad predictiva de la función. De nuevo, se implementó en Python, esta vez con la librería *scikit learn*, versión 0.24.2.

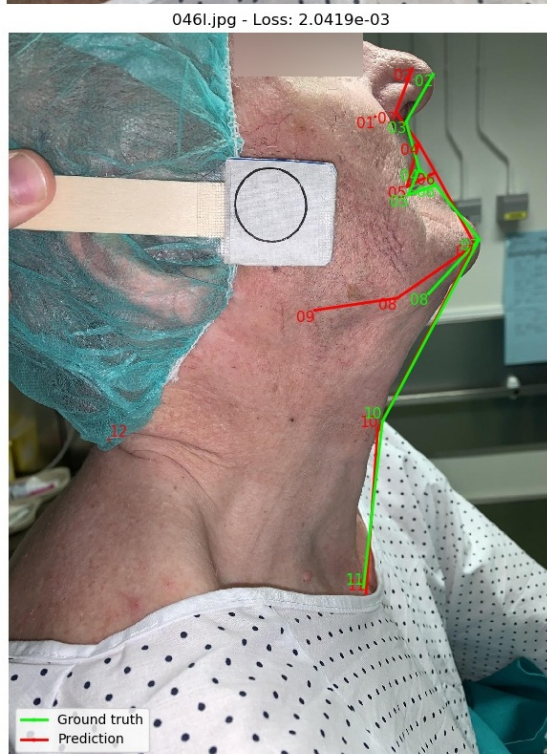
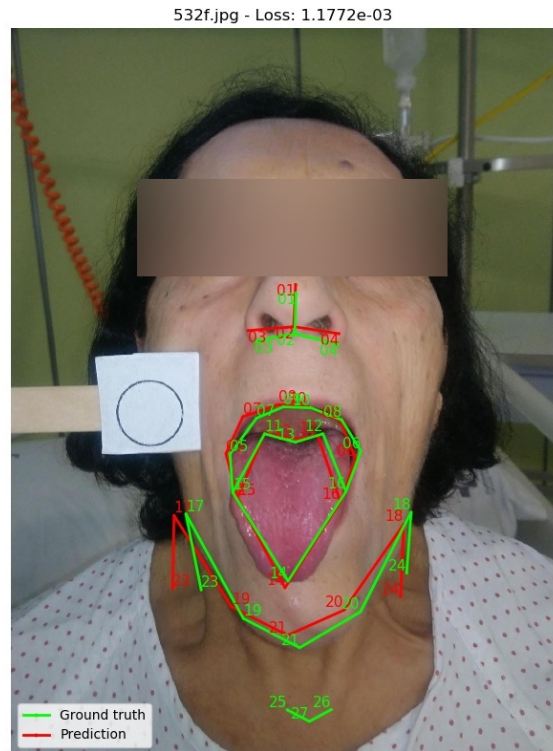
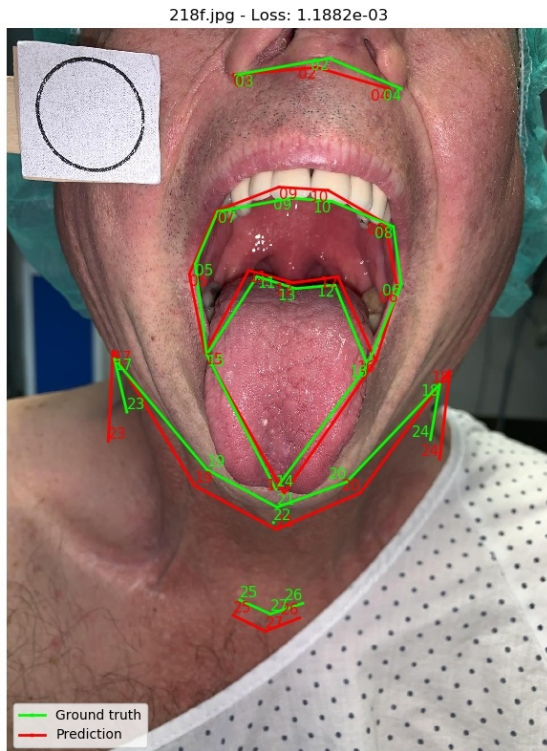


Figura 2: Cuatro casos de ejemplos de resultados alcanzados por nuestro algoritmo automático AI-ML de identificación de 'landmarks' [líneas y puntos rojos], respecto al consenso de anesthesiólogos como referencia [en verde]: dos imágenes frontales (arriba), dos laterales (abajo). Estos 4 casos presentados aquí no son escogidos como especialmente favorables, sino que recogen desempeños intermedios; es decir, grados de prestación similares al rendimiento mediano total. [Imagen adaptada desde (10)]

Resultados:

	Total n=503	Por sexo		Binomial 0.0013
		Mujeres n=215	Hombres n=288	
Edad	62.0 (51.0, 72.0)	57.0 (45.0, 69.5)	65.0 (54.0, 73.0)	Mann-Whitney <0.001
Peso [kg]	75.0 (65.0, 87.5)	67.0 (58.0, 80.5)	78.5 (70.0, 92.0)	<0.001
Altura [m]	1.67 (1.60, 1.73)	1.60 (1.56, 1.65)	1.72 (1.67, 1.77)	<0.001
IMC [kg m⁻²]	26.81 (24.02, 30.44)	26.08 (23.13, 31.23)	27.05 (24.27, 30.27)	0.2626
Mallampati				χ^2
<i>Grado I</i>	235 (46.72%)	114 (53.02%)	121 (42.01%)	0.0422
<i>Grado II</i>	172 (34.19%)	70 (32.56%)	102 (35.42%)	
<i>Grado III</i>	80 (15.90%)	27 (12.56%)	53 (18.40%)	
<i>Grado IV</i>	16 (3.18%)	4 (1.86%)	12 (4.17%)	
Cormack-Lehane				χ^2
<i>Clase I</i>	367 (72.96%)	167 (77.67%)	200 (69.44%)	0.1712
<i>Clase II</i>	103 (20.48%)	38 (17.67%)	65 (22.57%)	
<i>Clase III</i>	31 (6.16%)	9 (4.19%)	22 (7.64%)	
<i>Clase IV</i>	2 (0.40%)	1 (0.47%)	1 (0.35%)	
Vía aérea difícil				χ^2
<i>Criterio IDS</i>				0.1892
<i>No</i>	474 (94.23%)	206 (95.81%)	268 (93.06%)	
<i>Sí</i>	29 (5.77%)	9 (4.19%)	20 (6.94%)	
<i>Criterio ASA</i>				χ^2
<i>No</i>	479 (95.23%)	209 (97.21%)	270 (93.75%)	0.0718
<i>Sí</i>	24 (4.74%)	6 (2.79%)	18 (6.25%)	
<i>IDS o ASA</i>				χ^2
<i>No</i>	471 (93.64%)	205 (95.35%)	266 (92.36%)	0.1744
<i>Sí</i>	32 (6.36%)	10 (4.65%)	22 (7.64%)	

Los valores se muestran como mediana (rango inter-cuartil), o número (porcentaje).
La columna de más a la derecha contiene los p-valores de las diferencias poblacionales respecto al sexo, y el test estadístico usado.
IMC: índice de masa corporal, IDS: Intubation Difficulty Scale, ASA: American Society of Anesthesiologists.

Figura 3: Características principales de la cohorte.

La incidencia estimada de VAD documentada fue de un 6.36% (32 de 503 pacientes) según los criterios IDS y ASA: en cada caso, el más estricto de ambos criterios (véase Figura 3).

La valoración subjetiva (pre-intervención) de los clínicos obtuvo una sensibilidad de 25.00%, con una especificidad de 93.63%. En comparación, nuestra herramienta de IA-ML para extraer 'landmarks' y mediciones morfológicas a partir de ellos, combinada después con el modelo de regresión logística de Naguib 2006, alcanzó una sensibilidad del 53.12% y una especificidad del 79.83%. El AUC obtenida, o área bajo la curva ROC, fue de 0.680 (Figura 4).

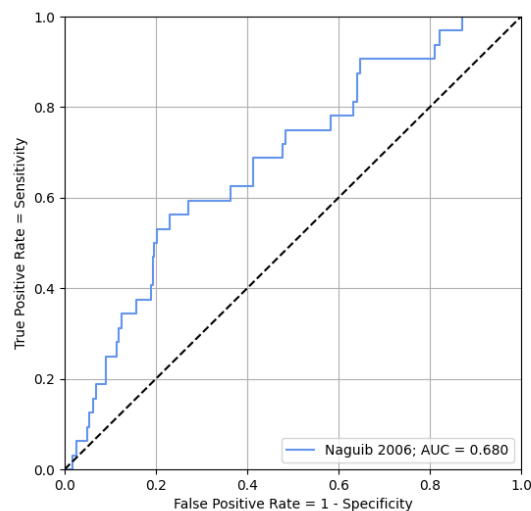


Figura 4: Curva ROC (Receiver Operating Characteristic). Se muestra la relación entre sensibilidad y especificidad del modelo de Naguib 2006, evaluado a partir de las mediciones extraídas automáticamente de las fotografías preoperatorias. El AUC (área bajo la curva ROC) cuantifica la capacidad de discriminación de la metodología.

Conclusiones:

La morbilidad relacionada con el manejo de la vía aérea, como resultado de la incapacidad de anticipar la vía aérea difícil, sigue siendo la principal preocupación del anestesiólogo. La incidencia de laringoscopia difícil es muy variable y oscila entre 1.5% y 13%, acorde con lo observado en nuestra muestra.

En este trabajo proponemos un método basado en Inteligencia Artificial-Machine Learning para evaluar el grado de dificultad de la vía aérea a partir de dos imágenes faciales tomadas desde un dispositivo móvil. La capacidad del anestesiólogo de identificar una VAD en la consulta de anestesia es solo del 25%, lo que nos indica que 3 de cada 4 casos no se identifican. Este hecho constata que, a pesar de los numerosos trabajos dirigidos a desarrollar nuevos modelos de predicción y de las nuevas variables utilizadas por el anestesiólogo, la capacidad de discriminación de VAD en la práctica clínica sigue siendo baja.

Es, por tanto, necesario desarrollar, con la tecnología de que disponemos, nuevas herramientas que nos permitan no solamente una mejora en la capacidad predictiva, sino una automatización de todo el proceso, pudiendo tener una valoración previa a la consulta de anestesia, o incluso prescindir de la misma.

Integrando nuestro sistema de medición IA-ML con el modelo de Naguib, los resultados muestran que estamos cerca de igualar la capacidad predictiva del clínico. Cabe resaltar, que si bien es un modelo externo, desarrollado con una metodología y una población distintas, nos sirve como punto de partida de cara a desarrollar nuestros propios modelos de predicción.

Dado que la incidencia de VAD es baja, alcanzar un número razonable de casos es un procedimiento a largo plazo. Por lo tanto, el desarrollo actual y futuro incluye la colección de más datos.

Los resultados mostrados nos alientan pues a seguir investigando en esta línea de análisis de imágenes faciales. Además del enorme potencial que tiene en el ámbito de la telemedicina, creemos que proporcionará al anestesiólogo una herramienta de ayuda en la toma de decisiones automática, objetiva y accesible para la evaluación de intubación difícil.

Bibliografía:

1. T. M. Cook and S. R. Macdougall-Davis, "Complications and failure of airway management," *British J. of anaesthesia*, vol. 109, pp. i68–i85, 2012
2. L. D. Hove et al., "Analysis of deaths related to anesthesia in the period 1996-2004 from closed claims registered by the danish patient insurance association," *Anesthesiology*, vol. 106, no. 4, pp. 675–680, 2007.
3. Karakus O, Kaya C, Emre F, Koksall E. Valor predictivo de los test preoperatorios para estimar la intubación difícil en pacientes sometidos a la laringoscopia directa para la cirugía de oído, nariz y garganta. *Brazilian J Anesthesiol*. 2015;65(2):85-91.
4. Roth D, Pace NL, Lee A, Hovhannisyan K, Warenits AM, Arrich J, Herkner H. Bedside tests for predicting difficult airways: an abridged Cochrane diagnostic test accuracy systematic review. *Anaesthesia*. 2019 Jul;74(7):915-928.
5. Merino M, Marcos JM, García R, Díez F, España L, Bermejo, et al. Evaluación de un protocolo de predicción de vía aérea difícil en la práctica habitual : estudio de concordancia. 2010;473-8.
6. Naguib M, Scamman FL, O'Sullivan C, Aker J, Ross AF, Kosmach S, Ensor JE. Predictive performance of three multivariate difficult tracheal intubation models: a double-blind, case-controlled study. *Anesth Analg*. 2006 Mar;102(3):818-24.
7. Cuendet GL, Schoettker P, Yüce A, Sorci M, Gao H, Perruchoud C, Thiran JP. Facial Image Analysis for Fully Automatic Prediction of Difficult Endotracheal Intubation. *IEEE Trans Biomed Eng*. 2016 Feb;63(2):328-39.
8. Adnet F, Borron SW, Racine SX, Clemessy JL, Fournier JL, Plaisance P, Lapandry C. The intubation difficulty scale (IDS): proposal and evaluation of a new score characterizing the complexity of endotracheal intubation. *Anesthesiology*. 1997 Dec;87(6):1290-7. doi: 10.1097/00000542-199712000-00005. PMID: 9416711.
9. Apfelbaum JL, Hagberg CA, Caplan RA, Blitt CD, Connis RT, Nickinovich DG, Hagberg CA, Caplan RA, Benumof JL, Berry FA, Blitt CD, Bode RH, Cheney FW, Connis RT, Guidry OF, Nickinovich DG, Ovassapian A; American Society of Anesthesiologists Task Force on Management of the Difficult Airway. Practice guidelines for management of the difficult airway: an updated report by the American Society of Anesthesiologists Task Force on Management of the Difficult Airway. *Anesthesiology*. 2013 Feb;118(2):251-70.
10. Fernando García-García, Dae-Jin Lee, Francisco Mendoza-Garcés, Sofía Irigoyen-Miró, María J.Legarreta-Olabarrieta, Susana García-Gutiérrez, Inmaculada Arostegui. (2022). Automated location of orofacial landmarks to characterize airway morphology in anaesthesia via deep learning. Under review.