



DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL PARA LA
EVALUACIÓN DE LA CALIDAD DE VOZ CON BASE EN LA ESCALA GRBAS

Jesus Andrés Ruano Bolaños – 20561712655

Cristian David Alegría Cardona – 20561713335

Universidad Antonio Nariño

Programa de Ingeniería Biomédica

Facultad de Ingeniería Mecánica, Electrónica y Biomédica

Popayán, Colombia

2021

DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL PARA LA
EVALUACIÓN DE LA CALIDAD DE VOZ CON BASE EN LA ESCALA GRBAS

Jesus Andrés Ruano Bolaños – 20561712655

Cristian David Alegría Cardona – 20561713335

Proyecto de grado presentado como requisito parcial para optar al título de:

Ingeniero Biomédico

Director:

PhD. Julián Antonio Villamarín Muñoz

Línea de Investigación:

Bioingeniería

Universidad Antonio Nariño

Programa de Ingeniería Biomédica

Facultad de Ingeniería Mecánica, Electrónica y Biomédica

Popayán, Colombia

2021

NOTA DE ACEPTACIÓN

El trabajo de grado titulado “**DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL PARA LA EVALUACIÓN DE LA CALIDAD DE VOZ CON BASE EN LA ESCALA GRBAS**”, Cumple con los requisitos para optar al título de **Ingeniero Biomédico**.

Firma del Tutor

Firma del Jurado

Firma del Jurado

RESUMEN

El análisis de la calidad de voz se ha convertido en una actividad rutinaria en clínicas y hospitales, donde son realizadas por profesionales de la voz (logopedas); estos análisis generalmente se realizan con base en la escala GRBAS, y presentan características subjetivas altamente influenciadas por la experiencia, nivel de educación del personal, entre otros (*Gordillo, 2018*).

Para este proyecto se implementó la base de datos de voces sintéticas desarrollada por la Universidad Antonio Nariño por medio de la aplicación Evaper, con el propósito de desarrollar una herramienta computacional para la evaluación de la calidad de voz, mediante la extracción de características vocales tanto acústicas como estadísticas, y la implementación de sistemas de machine learning para dar un respectivo diagnóstico con base en la escala GRBAS. Como primer resultado se encontró que el algoritmo encargado de realizar la extracción de las características vocales presentó un fuerte nivel de correlación con respecto al software Praat, software que se consideró como sistema estándar debido a su muy significativa trayectoria en el campo de la logopedia; llegando a obtener una correlación de Spearman $\rho > 0.85$, validando de esta manera el algoritmo dedicado a la extracción de características y el proceso metodológico implementado. Como segundo resultado, se encontró que los modelos de clasificación implementados en este proyecto presentaron un alto nivel de exactitud, a excepción de uno de los parámetros del género femenino (Ronquera), debido a la existencia de un error en la base de datos, ya que presentaba una falta de información para este género; los resultados obtenidos en escala

porcentual del nivel de exactitud de los modelos que conforman la herramienta son: Modelos de Ronquera para género masculino =71,2% , Modelos de Soplosidad para género masculino = 84,8% , Modelos de Astenia Tensión para género masculino = 93,3%, Modelos de Grado para género masculino = 93,6%, Modelos de Ronquera para género femenino = 40,5% , Modelos de Soplosidad para género femenino = 90,5% , Modelos de Astenia Tensión para género femenino = 97,7%, Modelos de Grado para género femenino = 95,3%. Como último resultado, fruto de realizar un avance preliminar de la fase 2, la cual consistió en evaluar el desempeño de la herramienta implementando 10 voces reales proporcionada por la Universidad del Valle, se obtuvieron bajos niveles de exactitud y correlación, siendo estos valores iguales a: Exactitud $\leq 50\%$ y $-0,2 < \text{Correlación} \leq 0,5$.

Los resultados obtenidos mediante la evaluación de los modelos desarrollados con voces sintéticas, permitieron validar el funcionamiento de la herramienta computacional, sin embargo, los resultados tras realizar el avance preliminar, a pesar de no ser muy significativos por la baja cantidad de datos, demostraron que es necesario llevar a cabo un análisis de la herramienta con el fin de realizar las modificaciones pertinentes que mejoren la eficacia en su funcionamiento al momento de implementarla con voces reales.

Palabras clave: Subjetividad, Herramienta computacional, Calidad vocal, GRBAS, Procesamiento de señales, Machine learning.

ABSTRACT

Voice quality analysis has become a routine activity in clinics and hospitals, where they are performed by voice professionals (speech therapists); These analyses are generally performed based on the GRBAS scale, and present subjective characteristics highly influenced by experience, level of education of staff, among others (Gordillo, 2018).

For this project, the database of synthetic voices developed by the Antonio Nariño University was implemented through the Evaper application, with the purpose of developing a computational tool for the evaluation of voice quality, through the extraction of both acoustic and statistical vocal characteristics, and the implementation of machine learning systems to give a respective diagnosis based on the GRBAS scale. As a first result, it was found that the algorithm in charge of performing the extraction of vocal features presented a strong level of correlation with respect to the Praat software, software that was considered as a standard system due to its very significant trajectory in the field of speech therapy; reaching a Spearman rho correlation > 0.85 , thus validating the algorithm dedicated to the extraction of features, and the implemented methodological process. As a second result, it was found that the classification models implemented in this project presented a high level of accuracy, with the exception of one of the parameters of the female gender (Roughness), due to the existence of an error in the database, since it presented a lack of information for this gender; the results obtained in percentage scale of the level of accuracy of the models that make up the tool are: Hoarseness Models for male gender =71.2% , Tension Asthenia Models for male gender = 84.8% , Tension Asthenia

Models for male gender = 93.3%, Grade Models for male gender = 93.6%, Hoarseness Models for female gender = 40.5% , Tension Asthenia Models for female gender = 90.5% , Tension Asthenia Models for female gender = 97.7%, Grade Models for female gender = 95.3%. As a last result, after performing a preliminary preview of phase 2, which consisted of evaluating the performance of the tool implemented 10 real voices provided by the Universidad del Valle, low levels of accuracy and correlation were obtained, being these values equal to: Accuracy $\leq 50\%$ and $-0,2 < \text{Correlation} \leq 0,5$.

The results obtained through the evaluation of the models developed with synthetic voices, allowed validating the performance of the computational tool, however, the results after performing the preliminary preview, despite not being very significant due to the low amount of data, showed that it is necessary to carry out an analysis of the tool in order to make the appropriate modifications to improve the effectiveness of its operation when implemented with real voices.

Keywords: Subjectivity, Computational tool, Vocal quality, GRBAS, Signal processing, Machine learning.

CONTENIDO

	<i>Pág.</i>
INTRODUCCIÓN	15
CAPÍTULO 1	18
1.1. PLANTEAMIENTO DEL PROBLEMA	18
1.2. JUSTIFICACIÓN	19
1.3. OBJETIVO GENERAL	20
1.4. OBJETIVO ESPECÍFICO	21
1.5. LÍNEA DE INVESTIGACIÓN	21
CAPÍTULO 2	22
2.1. MARCO TEÓRICO	22
2.1.1. La voz	22
2.1.2. Fisiología de la voz	23
2.1.3. Voz Patológica	25
2.1.4. Sobrecarga vocal	25
2.1.5. Voz profesional	26
2.1.6. Escala GRBAS	27
2.1.7. Evaluación acústica de la voz	27
2.1.8. Evaluación perceptual de la calidad de voz	28
2.1.9. Características vocales	29
2.1.10. Praat	33
2.1.11. EVAPER	34
2.1.12. Coeficiente de Spearman	35
2.1.13. Machine learning	36
2.1.14. SVM	37
2.1.15. Kernels	38
2.1.16. Knn	40

2.1.17.	Naive bayes	41
2.1.18.	Cross validation	42
2.1.19.	Matriz de confusión	43
2.1.20.	Curva ROC	44
CAPÍTULO 3		46
3.1. DISEÑO METODOLÓGICO		46
3.1.1.	Fase 1: Establecer el diseño técnico y teórico de los parámetros acústicos y estadísticos implementados en el proyecto para el análisis de la voz.	46
3.1.2.	Fase 2. Implementar el diseño técnico, desarrollando el algoritmo encargado de extraer las características de las señales y los modelos de clasificación para cada parámetro.	48
3.1.3.	Fase 3: Validar el funcionamiento de la herramienta desarrollada en comparación con señales sintéticas proporcionadas por la Universidad Antonio Nariño, y señales reales proporcionadas por la Universidad del Valle.	64
4. RESULTADOS		69
5. CONCLUSIONES		104
REFERENCIAS BIBLIOGRÁFICAS		106
A. Anexo 1: Diagrama de flujo el algoritmo encargado de la extracción de las características vocales.		109
B. Anexo 2: Curvas ROC de los modelos para el género masculinos		110
C. Anexo 3: Curvas ROC de los modelos para el género femenino		122
D. Anexo 4: Curvas ROC de los modelos no dedicados		129
E. Anexo 5: Manual de usuario		141
F. Anexo 6: Manual técnico		151

LISTA DE FIGURAS

	Pág
Figura 1.	
Figura 1 a. Esquema de la fisiología de la voz.	24
Figura 1 b. Anatomía del aparato fonatorio.	24
Figura 2. Medidas de perturbación de Jitter y Shimmer en la señal de voz.	29
Figura 3. Interfaz del software Praat.	34
Figura 4. Interfaz del software Evaper.	35
Figura 5. Valores del coeficiente de Spearman.	36
Figura 6.	
Figura 6 a. Machine learning no supervisado	37
Figura 6 b. Machine learning supervisado	37
Figura 7. Tipos de kernels.	39
Figura 8. Representación gráfica del modelo de clasificación Knn.	41
Figura 9. Representación gráfica del modelo de clasificación Naive Bayes.	42
Figura 10. Explicación del funcionamiento del método de validación cruzada.	43
Figura 11. Representación de la matriz de confusión.	44
Figura 12. Modelo de la curva ROC.	45
Figura 13. Representación gráfica de la señal estéreo segmentada por canal.	49
Figura 14. Diagrama de flujo del preprocesamiento de las señales.	50
Figura 15. Incapacidad del algoritmo para realizar la detección de picos principales con un umbral determinado.	51
Figura 16.	
Figura 16 a. Presencia de errores de subventaneo.	54
Figura 16 b. Reducción de errores de subventaneo.	54
Figura 17. Subventana del grupo Shimmer.	55

Figura 18. Gráfica de la densidad espectral de potencia.	57
Figura 19. Gráfica de detección de armónicos en la señal.	57
Figura 20.	
Figura 20 a. Gráfica del espectrograma 3d.	58
Figura 20 b. Gráfica del espectrograma en 2d.	58
Figura 21. Reporte generado por Praat para ciertas señales.	60
Figura 22. Interfaz gráfica de la aplicación de Matlab “classification learnings”.	62
Figura 23. Matriz de confusión SVM cúbico de ronquera para el género masculino.	72
Figura 24. Matriz de confusión SVM cúbico de soplosidad para el género masculino.	74
Figura 25. Matriz de confusión SVM cuadrático astenia y tensión género masculino.	75
Figura 26. Matriz de confusión SVM cuadrático de grado para el género masculino.	77
Figura 27. Matriz de confusión SVM cúbico de grado para el género masculino.	77
Figura 28. Matriz de confusión SVM cúbico de soplosidad para el género femenino.	79
Figura 29. Matriz de confusión SVM cúbico de astenia y tensión género femenino.	80
Figura 30. Matriz de confusión del modelo no dedicado SVM cúbico de ronquera.	82
Figura 31. Matriz de confusión del modelo no dedicado SVM cúbico de soplosidad.	83
Figura 32. Matriz de confusión del modelo no dedicado SVM cúbico de astenia y tensión.	85
Figura 33. Matriz de confusión modelo no dedicado SVM con kernel gaussiano medio del parámetro grado.	86
Figura 34. Matriz de confusión del modelo dedicado SVM cúbico para ronquera.	88
Figura 35. Matriz de confusión del modelo no dedicado SVM cúbico para ronquera.	88
Figura 36. Matriz de confusión del modelo dedicado SVM cúbico para soplosidad.	89
Figura 37. Matriz de confusión del modelo no dedicado SVM cúbico para soplosidad.	90
Figura 38. Matriz de confusión del modelo no dedicado SVM cuadrático astenia y tensión.	91
Figura 39. Matriz de confusión del modelo no dedicado SVM cúbico para astenia y tensión.	91
Figura 40. Matriz de confusión SVM cúbico para grado.	92

Figura 41. Matriz de confusión SVM cuadrático para grado.	93
Figura 42. Matriz de confusión SVM cúbico para grado.	93
Figura 43. Matriz de confusión masculino SVM cúbico para ronquera.	94
Figura 44. Matriz de confusión del modelo no dedicado SVM cúbico para ronquera.	95
Figura 45. Matriz de confusión del modelo dedicado SVM cúbico para soplosidad.	96
Figura 46. Matriz de confusión del modelo no dedicado SVM cúbico para soplosidad.	96
Figura 47. Matriz de confusión del modelo dedicado SVM cúbico de astenia y tensión.	97
Figura 48. Matriz de confusión del modelo no dedicado SVM cúbico para astenia y tensión.	98
Figura 49. Matriz de confusión del modelo masculino dedicado SVM cúbico para grado.	99
Figura 50. Matriz de confusión del modelo masculino dedicado SVM cuadrático para grado.	99
Figura 51. Matriz de confusión del modelo no dedicado SVM cúbico para grado.	99
Figura 52. Interfaz gráfica del programa.	102

LISTA DE TABLAS

	Pág
Tabla 1. Resultados de los modelos dedicados para el parámetro de Ronquera en género masculino.	72
Tabla 2. Lista de resultados de los modelos dedicados para el parámetro de Soplosidad en género masculino.	73
Tabla 3. Lista de resultados de los modelos dedicados para los parámetros de Astenia y Tensión en género masculino.	74
Tabla 4. Lista de resultados de los modelos dedicados para el parámetro Grado en género masculino.	76
Tabla 5. Lista de resultados de los modelos dedicados para el parámetro de Soplosidad en género femenino.	78
Tabla 6. Lista de resultados de los modelos dedicados para el parámetro de Astenia y Tensión en género femenino.	79
Tabla 7. Lista de resultados de los modelos no dedicados para el parámetro de Ronquera.	82
Tabla 8. Lista de resultados de los modelos no dedicados para el parámetro de Soplosidad.	82
Tabla 9. Lista de resultados de los modelos no dedicados para los parámetros de Astenia y Tensión.	84
Tabla 10. Lista de resultados de los modelos no dedicados para el parámetro Grado.	85
Tabla 11. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro de Ronquera en el género masculino.	87
Tabla 12. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro de Soplosidad en el género masculino.	89
Tabla 13. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para los parámetros de Astenia y Tensión en el género masculino.	90
Tabla 14. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro de Grado en el género masculino.	92
Tabla 15. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro de ronquera en el género femenino.	94

Tabla 16. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro de Soplosidad en el género femenino.	96
Tabla 17. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para los parámetros de Astenia y Tensión en el género femenino.	97
Tabla 18. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro de Grado en el género femenino.	98
Tabla 19. Lista de los modelos para el género masculino que formarán parte de la herramienta final.	100
Tabla 20. Lista de los modelos para el género femenino que formarán parte de la herramienta final.	101
Tabla 21. Resultados obtenidos en la evaluación de la herramienta con voces reales.	103

INTRODUCCIÓN

A lo largo de los últimos años, la evaluación logopédica de la calidad vocal, ha alcanzado un desarrollo bastante importante debido principalmente al crecimiento exponencial de la tecnología, aprovechando esto se han desarrollado diversos softwares que mediante el uso de un sistema computarizado y análisis de señales logran obtener las diversas características acústicas de las señales de voz (método objetivo), para que a partir de ahí, el logopeda profesional pueda evaluar la calidad de voz y/o detectar la existencia de una patología vocal (método subjetivo). En estos softwares tenemos, por ejemplo: Praat (*Boersma & Weenink, 2011*) , VoxMetria (*Behlau, 2014*) , LingWAVES (*Wevosys, 2009*) y WaveSurfer (*Centre for speech technology, 2019*), los cuales fueron diseñados para otorgar una ayuda al logopeda para un correcto seguimiento y análisis de la voz del paciente mediante la extracción de parámetros acústicos.

También podemos encontrar patentes que implementan los sistemas computacionales para crear o diseñar nuevas metodologías que aumenten la eficiencia y exactitud de los parámetros extraídos, por ejemplo, tenemos la patente (*Gomez et al., 2011*) “Método y sistema para la estimación de parámetros fisiológicos de la fonación”, el cual es el desarrollo de una metodología que permite el cálculo de varios parámetros de voz, ya que éstos llevan intrínsecamente la información sobre efectos orgánicos o neurogénicos causantes de una patología; patente (*LevanonLan & Shifrin, 2005*) “*Method and system for diagnosing pathological phenomena using a voice signal*”, ésta invención hace referencia a

un método y sistema para diagnosticar fenómenos patológicos mediante la implementación de una función de intensidad media cuadrada o la intensidad máxima cuadrada de la señal de la voz; por último se detalla la patente (Clavbo, 2006) “*Method and device for speech analysis*” donde se desarrolla la invención de un dispositivo y un método para el análisis del habla, que comprende la medición de las notas fundamentales de una secuencia de habla identificando la frecuencia entre al menos algunas de dichas notas fundamentales, y posteriormente se evalúa la frecuencia con la que se producen al menos algunos de estos intervalos identificados en la secuencia de habla que se va a analizar.

Por último, se recopilaron artículos los cuales realizan una investigación en base al desarrollo de herramientas computacionales para la detección de patologías vocales o comparación de métodos de Machine Learning para la evaluación de la calidad vocal, entre los cuales destacan: “*Classification of voice disorders using a one-dimensional convolutional Neural network*” (Shintaro et al., 2020) en donde los autores desarrollaron una red neuronal unidimensional convolucional para evaluar la calidad de voz con base en la escala GRBAS, para esto implementan una base de datos propia la cual cuenta con 1377 muestras de voces, donde cada una fue diagnosticada por 3 expertos, en este artículo logran obtener una precisión de clasificación alrededor del 70%, y además concluyen que el resultado de la evaluación de la red neuronal es comparable a los resultados de los expertos humanos; También se tomó como referencia el artículo “Red neuronal multiescala para clasificación de la calidad vocal” (García et al., 2021), que, mediante la implementación de una base de datos previamente desarrollada por la universidad politécnica de madrid,

desarrollan un clasificador neuronal multiescala que mediante el análisis del Power Cepstrum de las señales de voz logran realizar una clasificación binaria del grado general de disfonía (G) entre normal (grado 0) y patológico (grado 1, 2, 3), además, cabe resaltar que realizan una comparación de la eficiencia de clasificación cuando a la red neuronal se le entrega la señal cruda o suavizada mediante varios kernels de filtro gaussiano en donde obtienen los resultados de 76,98% y 80,69% respectivamente.

Con base en lo anterior, se puede concluir que uno de los problemas que se presentan en la mayoría de las pruebas tradicionales de la escala GRBAS, es la dependencia de un profesional para la interpretación de los resultados, lo que representa un alto grado de subjetividad y por ende posibles errores en el diagnóstico, esto provocado por diversos factores subjetivos casi imposibles de erradicar (*Gordillo, 2018*), lo que podría implicar para el usuario un impacto de manera negativa en las áreas personal, profesional, social y/o económica.

CAPÍTULO 1

1.1. PLANTEAMIENTO DEL PROBLEMA

Actualmente la evaluación de la calidad vocal suele realizarse mediante métodos de percepción auditiva, los cuales involucran una gran cantidad de tiempo para el logopeda y un gran costo para el usuario, y ya que depende estrictamente de la experiencia del personal logopédico, los resultados que se obtendrán presentarán un alto grado de subjetividad, el cual puede afectar negativamente a la hora de escoger el mejor tratamiento (*Zheng et al., 2018*).

Un estudio publicado en la revista *ARETÉ* (*Gordillo, 2018*), donde hablan de los hitos de la evaluación perceptual auditiva de la voz, muestra los factores que contribuyen a la mala correlación entre calificadores, los cuales son: el oyente, haciendo referencia a los factores intrínsecos del evaluador como la experiencia, la profesión, entre otras; el estímulo, en donde especifica las dificultades de la clasificación perceptual y la capacidad retentiva a un estímulo en específico; y la escala, refiriéndose a los diferentes métodos de evaluación que puede seleccionar el logopeda, ya que existen por ejemplo: el método visual-analógico, que cuenta con un 88% de precisión en los resultados, pero con una concordancia entre

evaluadores considerablemente baja, y el método ordinal, con un 64% de precisión en los resultados, y una mayor correlación entre evaluadores.

Teniendo en cuenta la información anterior surge la necesidad de crear una herramienta computacional que de manera automática analice y evalúe la calidad de la voz con base en la escala GRBAS, pero, y debido a las complicaciones logísticas y salubres ocasionados por la emergencia sanitaria de Covid-19, el desarrollo de ésta herramienta se ha dividido en 2 etapas: construcción y validación de la herramienta mediante voces sintéticas, y validación del desempeño con voces reales, de las cuales en este documento solo se desarrolla la primera etapa, y adicional a ello, se realiza un avance preliminar de la segunda.

1.2. JUSTIFICACIÓN

Este proyecto tiene como finalidad el aportar a la necesidad histórica que ha tenido el diagnóstico de la condición de la voz, sobre todo cuando se habla de la condición en relación a estados patológicos, de manera que se logre reducir la subjetividad en los métodos tradicionales de diagnóstico, ya que estos dependen únicamente de la interpretación del logopeda (*Gordillo, 2018*), incluso en los pocos casos donde se implementan algunos softwares existentes en la actualidad, ya que éstos han sido diseñados como sistemas de *apoyo*. Teniendo en cuenta lo anterior se puede observar claramente la existencia de una gran brecha de subjetividad presente en cada examen, esto sin mencionar que el diagnóstico puede variar no solo de un logopeda a otro, sino que también entre el

mismo profesional examinando la misma muestra en tiempos distintos, lo que podría terminar en diagnósticos que no sean del todo apropiados, quedando esto reflejado en un plan de rehabilitación que no sea el óptimo para las condiciones reales del paciente y en un impacto negativo en las diferentes áreas como: personal, profesional, social y/o económica.

Además, cabe resaltar los beneficios intrínsecos que se pueden obtener como consecuencia de la implementación de una herramienta de este tipo, como por ejemplo: la objetividad en el análisis de las señales y el tiempo de evaluación, ya que al tratarse de un software, el procesamiento de los datos se realiza de manera más rigurosa y siguiendo únicamente modelos lógicos y matemáticos, teniendo además un tiempo de procesamiento considerablemente corto, reduciendo de ésta manera las condiciones negativas relacionadas al proceso tradicional de la toma de exámenes para la evaluación de la calidad de la voz, e incluso la reducción de la subjetividad en los diagnósticos ya que los métodos tradicionales están sujetos a la posible existencia de errores.

1.3. OBJETIVO GENERAL

Desarrollar una herramienta computacional para la evaluación de la calidad de voz con base en la escala GRBAS a partir de voces sintéticas, empleando técnicas de procesamiento digital de señales y Machine Learning.

1.4. OBJETIVO ESPECÍFICO

- Establecer el diseño técnico y teórico de los parámetros acústicos y estadísticos implementados en el proyecto para el análisis de la voz.
- Implementar el diseño técnico, desarrollando el algoritmo encargado de extraer las características de las señales y los modelos de clasificación para cada parámetro.
- Validar el funcionamiento de la herramienta desarrollada en comparación con señales sintéticas proporcionadas por la Universidad Antonio Nariño, y señales reales proporcionadas por la Universidad del Valle.

1.5. LÍNEA DE INVESTIGACIÓN

La facultad de ingeniería Mecánica, Electrónica y Biomédica (FIMEB) de la Universidad Antonio Nariño ofrece el programa de ingeniería biomédica. Este programa se define como la aplicación de conocimientos de ingeniería y ciencias exactas para solucionar problemas de medicina y biología, dentro del programa se encuentra el grupo de investigación Bio-Ingeniería que cuenta con seis líneas de investigación; el presente anteproyecto se encuentra situado en la línea de análisis de información biomédica, debido a que la herramienta propuesta cumple con el segundo objetivo de la línea del grupo de bioingeniería, ya que por medio de ésta se busca implementar técnicas de procesamiento digital de señales con el fin de presentar un proyecto con áreas interdisciplinarias, entre las cuales están biomedicina y fonoaudiología.

CAPÍTULO 2

2.1. MARCO TEÓRICO

2.1.1. La voz

La voz se puede explicar como el sonido que se produce por causa de la vibración de las cuerdas vocales ubicadas en la zona de la laringe a través de la conducción del aire exhalado desde la caja torácica. Este es un fenómeno fisiológico, y su resultado es de naturaleza acústica, además la voz es un medio de comunicación entre las personas, y es el medio de comunicación más esencial para expresar y transmitir conocimientos, pensamientos, ideas y sentimientos, además, influye significativamente en nuestras relaciones tanto personales como profesionales.

Las voces tienen diferentes cualidades como: timbre, volumen, tono, duración o velocidad y ritmo, estas cualidades, aunque están directamente relacionadas con la postura corporal, el tono muscular y el óptimo manejo emocional, también pueden relacionarse al estado de salud de la persona (*Nonó et al., 2014*).

2.1.2. Fisiología de la voz

La voz humana es producida voluntariamente a través del aparato fonatorio, más específicamente por la glotis, aunque para la formación de sonidos y las propiedades de la voz intervienen en general tres partes del cuerpo humano, las cuales son: el aparato respiratorio, el aparato fonador y el aparato resonador (Figura 1.a) (*Preparadores de Oposiciones para la Enseñanza, 2005*).

- **Aparato respiratorio:** Este aparato proporciona el aire necesario para producir sonidos, está conformado principalmente por la nariz, tráquea, pulmones y el diafragma; siendo este último el encargado de realizar el proceso de respiración mediante la producción en los pulmones de una diferencia de presión con respecto al ambiente (presión negativa) que induce a la introducción del aire y viceversa.
- **El aparato fonador:** Este aparato lo conforma principalmente la laringe, la cual está situada a continuación de la tráquea, y está formada por varios cartílagos (tiroides, cricoides, aritenoides y epiglotis) y por un hueso (hioides); y las cuerdas vocales, las cuales son membranas formadas por ligamentos y músculos con un espacio entre sí el cual se ha denominado glotis, las cuales son las encargadas de generar sonidos mediante su vibración, ésto debido al flujo de aire que existe entre ellas (Figura 1.b); En la fonación, la respiración es activa y es producida por el aparato respiratorio, ésta respiración se denomina “soplo fonatorio”.

- Aparato resonador:** Este aparato es aquel que amplifica y transforma el sonido, dándole brillo y redondez a la voz, funcionando de la misma manera que la caja de resonancia acústica de una guitarra; En la voz humana los resonadores son los huesos del cuerpo que vibran como consecuencia del canto o habla, influyendo con mayor medida los resonadores faciales (paladar óseo, cavum, región de la laringe) y los senos (cavidades óseas), estos resonadores se enfatizan en determinadas bandas de frecuencia generadas por las cuerdas vocales, que si se analizan el espectro de frecuencia de las señales, se pueden observar la presencia de formantes y/o armónicos.

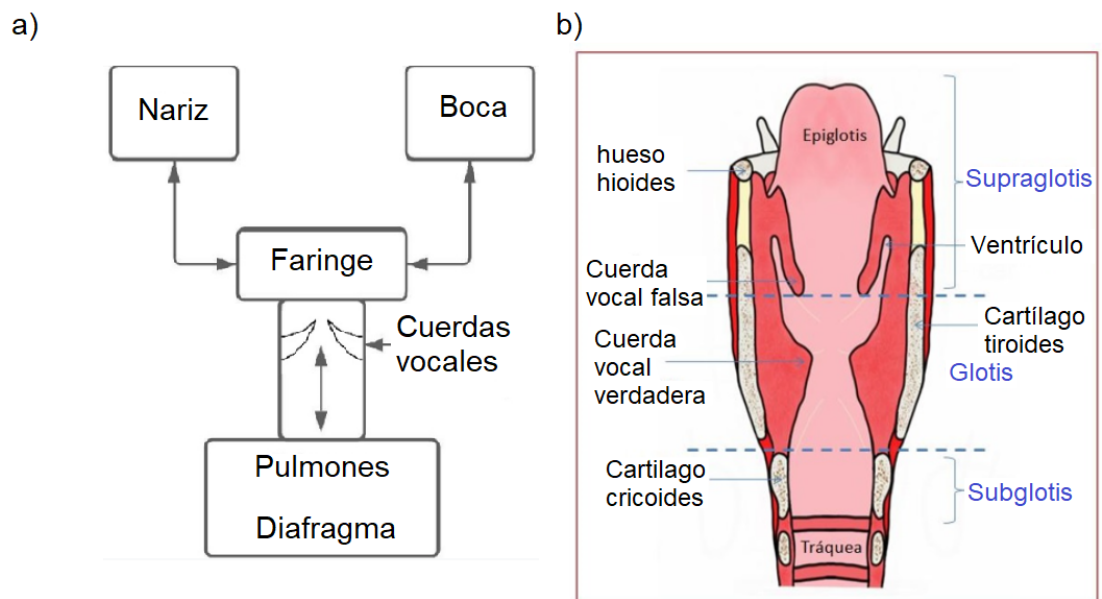


Figura 1 a) Esquema de la fisiología de la voz, **b)** Anatomía del aparato fonatorio, Adaptado de *Estenosis laríngea no neoplásica: cómo reconocerla y qué sospechar*, Madureira ,2014, Researchgate: <https://epos.mysr.org/poster/esr/seram2014/S-0784>

2.1.3. Voz Patológica

Una voz se considera patológica cuando existe un trastorno de la voz el cual ocasiona que el tono, timbre, intensidad o flexibilidad difiera con relación a las voces de las demás personas del mismo sexo, edad y grupo cultural. Las alteraciones de la voz tienen factores desencadenantes que pueden generar fatiga vocal y/o que las alteraciones logren mantenerse o agravarse, lo que puede concluir en un menor rendimiento vocal (*Aguilar & Vélez Julia, 2016*); estas anomalías pueden ser de carácter temporal, por ejemplo la disfonía ocasionada por el uso exhaustivo de la voz durante un breve periodo de tiempo, o de carácter permanente, ocasionado por enfermedades o patologías que comprometan algunos de los sistemas fundamentales para la producción del sonido vocal.

2.1.4. Sobrecarga vocal

Cuando hablamos de sobrecarga vocal, nos referimos a los efectos provocados en las cuerdas vocales a causa del abuso y uso excesivo de la función fonadora, el cual se manifiesta en forma de lesiones y síntomas funcionales que pueden afectar de manera leve o grave el funcionamiento y capacidad normal de la voz (*Mora & Cobeta, 2013*); Esta sobrecarga vocal se clasifica en 3 grados:

- **Grado I:** Se caracteriza por un aumento en el nivel de secreciones en el área de las cuerdas vocales, la voz puede sonar más oscura y se presentan molestias generales para hablar.

- **Grado II:** Esta suele ser la forma de sobrecarga más habitual, además del aumento del nivel de secreciones que se presenta en el grado I, se caracteriza por una limitación en el alcance de agudos en la voz, por lo que los más afectados serían los cantantes, es frecuente que se produzca luego de extensas y exigentes jornadas de ensayos
- **Grado III:** En este grado ya se pueden identificar inflamaciones leves y concentradas en zonas de ambas cuerdas vocales, estas lesiones pueden desaparecer si se conserva reposo vocal el tiempo suficiente.

2.1.5 Voz profesional

Los profesionales de la voz son todas aquellas personas que utilizan la voz como herramienta de trabajo y como medio principal para la ejecución de su actividad laboral.

Los profesionales que implementan la voz como herramienta de trabajo, requieren ciertas necesidades en comparación a las personas comunes como por ejemplo: adecuando niveles de funcionamiento y coordinación de los diversos sistemas (nervioso, fonatorio, respiratorio, resonador, auditivos, postural, hormonal entre otros), con el fin de que la producción vocal se realice de manera satisfactoria y cumpla con sus necesidades comunicativas y laborales (*Aguilar & Vélez Julia, 2016; Nonó et al., 2014*).

2.1.6. Escala GRBAS

La escala GRBAS propuesta por Hirano (1981) y desarrollada por la sociedad japonesa de logopedia y foniatría es una de las escalas más confiables, ampliamente utilizada y más aún para la evaluación perceptiva de la calidad de la voz, ésta se compone por cinco parámetros que a su vez son calificados en 4 parámetros que varían en función de la gravedad de la alteración, los parámetros evaluados son: G(grado), este parámetro califica el grado general de disfonía que posee el paciente, R(ronquera o rugosidad) ésta contempla en mayor medida la aspereza de la voz, A(astenia) representa el grado de fatiga o cansancio de la voz, B(soplosidad) corresponde al carácter de voz aireada, velada y soplada y S(tensión) califica la voz espástica, constreñida y tensa. Esta escala permite la mitigación de las variaciones en los análisis inter e intra evaluación que se realiza de manera subjetiva, ya que tiene una alta calidad diagnóstica, y su uso es importante en entornos clínicos ya que es una prueba no invasiva con facilidad de uso, y no requiere de costos excesivos (Sotres et al., 2015).

2.1.7. Evaluación acústica de la voz

El análisis acústico de la voz es un método de evaluación vocal objetivo y no invasivo, utilizado con fines diagnósticos, terapéuticos e investigativos; Este análisis utiliza registros acústicos obtenidos directamente de la voz del paciente por medio de un micrófono de carácter profesional, y mediante el uso de diversos programas dedicados a la extracción de las características acústicas, se puede detallar mayor cantidad de información ya sea de manera gráfica o numérica, favoreciendo la versatilidad y especificidad de la evaluación

vocal. Un detalle a resaltar es que estos programas han sido desarrollados como sistemas de apoyo al logopeda, debido a que estos presentan simplemente las características acústicas de la voz (*Droguett, 2017*).

2.1.8. Evaluación perceptual de la calidad de voz

En los orígenes de la Fonoaudiología la evaluación se centró en la información perceptual, aunque más adelante se introdujera el desarrollo de tecnologías acústicas y fisiológicas para la recolección de información sobre la voz. Tradicionalmente la voz humana ha sido descrita contemplando las características de un sonido como: duración, tono, intensidad y timbre, y para caracterizar este último se emplearon términos como voz ronca, áspera, raspada, metálica, con escape de aire/soplada, hipernasal, hiponasal y desnasal; Este enfoque subjetivo usa calificaciones perceptivas de los oyentes en una dimensión dada, generalmente en una escala predeterminada, en contraposición, en el enfoque objetivo de la evaluación del timbre empleando tecnologías acústicas, se miden algunos parámetros de la forma de onda acústica, con la creencia de que estos parámetros de alguna manera influyen en la respuesta provocada, pero sea cual fuese el enfoque, se reconoce que la calidad vocal es multidimensional y recibe influencia de factores tales como el tono, la intensidad, la precisión articulatoria, la velocidad ritmo del habla y características de resonancia (*Gordillo, 2018*).

2.1.9. Características vocales

Los parámetros de perturbación de la frecuencia de la voz (Jitter) y de la amplitud (Shimmer) se utilizan habitualmente como parte de un examen exhaustivo de la voz. El Jitter es la medida de las variaciones ciclo a ciclo del periodo glótico fundamental y el Shimmer es la variación ciclo a ciclo de las amplitudes del periodo glótico fundamental, como se representa en la siguiente figura (Teixeira & Gonçalves, 2014).

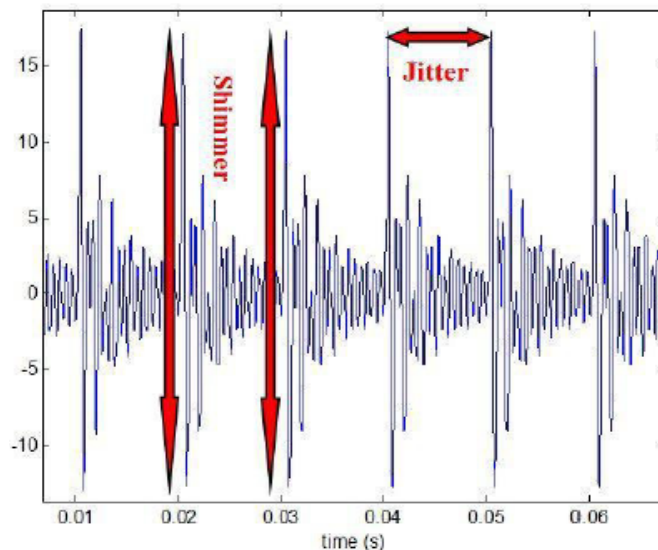


Figura 2 Medidas de perturbación de Jitter y Shimmer en la señal de voz, Adaptado de: *Accuracy of Jitter and Shimmer Measurements*, Teixeira & Gonçalves, 2014.

- **Jitter:** las perturbaciones de Jitter pueden definirse mediante 4 parámetros como: Jitter local, Jitter local absoluto, Jitter rap y Jitter ppq 5; pero para el desarrollo de este proyecto se ha implementado únicamente el Jitter local y Jitter local absoluto (Teixeira & Gonçalves, 2014).

- **Jitter local:** es la diferencia absoluta promedio entre períodos consecutivos, dividida por el período promedio, y este valor se da en porcentaje (Ecuación 1).

$$Jitter\ local = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100$$

Ecuación 1

Donde T_i es la longitud de los periodos glotales extraídos y N es el número de periodos glotales extraídos.

- **Jitter local absoluto:** Es la diferencia absoluta promedio entre períodos consecutivos, dividida por el período promedio, este valor está dado en segundos o microsegundos (Ecuación 2).

$$Jitter\ local\ absoluto = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}|$$

Ecuación 2

Donde T_i es la longitud de los periodos glotales extraídos y N es el número de periodos glotales extraídos.

- **Shimmer:** En la característica Shimmer también podemos encontrar 4 medidas relacionadas: Shimmer local, Shimmer local dado en dB, Shimmer apq 3 y Shimmer apq 5; pero para el desarrollo de este proyecto se ha implementado únicamente el Shimmer local dado en dB (Teixeira & Gonçalves, 2014).

- **Shimmer local dB:** Es el logaritmo absoluto medio en base 10 de la diferencia entre las amplitudes de períodos consecutivos, multiplicado por 20 y dado en una escala de decibelios (dB) (Ecuación 3).

$$ShdB = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 * \log\left(\frac{A_{i+1}}{A_i}\right) \right|$$

Ecuación 3

Donde A_i es la amplitud extraída y N es el número de periodos de frecuencia fundamental extraídos.

- **Curtosis:** La curtosis es una medida de forma que mide cuán escarpada o achatada está una curva o distribución, indicando de esta manera, la cantidad de datos que hay cercanos a la media (*Matlab, 2021*).

La curtosis de una distribución está definida como:

$$K = \frac{E(x - \mu)^4}{\sigma^4}$$

Ecuación 4

Donde μ corresponde a la media de x , σ es la desviación estándar de x , y $E(t)$ representa el valor esperado de la cantidad t .

- **Asimetría:** La asimetría es una medida de la asimetría de los datos en torno a la media de la muestra. Si la asimetría es negativa, los datos se extienden más a la

izquierda de la media que a la derecha y viceversa, y para una distribución perfectamente simétrica, el valor esperado es igual a 0 (*Matlab, 2021*).

La asimetría de una distribución está definido como:

$$S = \frac{E(x - \mu)^3}{\sigma^3}$$

Ecuación 5

Donde μ corresponde a la media de x , σ es la desviación típica de x , y $E(t)$ representa el valor esperado de la cantidad t .

- **RMS:** La raíz de la media cuadrática o RMS es una medida estadística de la magnitud de una cantidad variable, la cual se define como la suma de los promedios al cuadrado dividido entre el número de elementos (*Belletti, 2018*).

El valor Rms está definido como:

$$Rms = \sqrt{\frac{1}{N} \sum_{n=1}^N |X_n|^2}$$

Ecuación 6

Donde N el número de elementos, y X_n representa los elementos de un vector que varían desde $n=1$ hasta el número total de elementos.

- **Energía de la señal:** La energía de una señal finita se puede representar como el área bajo la curva del cuadrado de una señal continua, como se puede observar en la

ecuación 7, pero para este caso, en donde se hace uso de señales discretas, se ha implementado la ecuación 8 (Rosario, 2017).

$$E = \int_{-\infty}^{\infty} |x(t)|^2 dt$$

Ecuación 7

$$E = \sum_{n=-\infty}^{\infty} |x[n]|^2$$

Ecuación 8

2.1.10. Praat

PRAAT es un programa diseñado especialmente para hacer investigaciones en fonética. Es un programa de libre distribución, de código abierto, multiplataforma y, además es gratuito. Fue desarrollado en la Universidad de Amsterdam por Paul Boersma y David Weenink a partir del año 1992. Este software posee una amplia gama de propósitos, por ejemplo: permite hacer análisis acústico, síntesis articulatoria, procesamiento estadístico de los datos, edición, manipulación de señales de audio, etc; (Boersma & Weenink, 2011).

En la Figura 3 se puede detallar una de las ventanas de la interfaz de Praat, mostrando a simple vista las señales cargadas en la parte superior, y algunas de las características extraídas como el espectrograma y formantes en la parte inferior, además cuenta con un menú bastante robusto en donde almacena varias de sus funciones dedicadas al análisis de las señales.

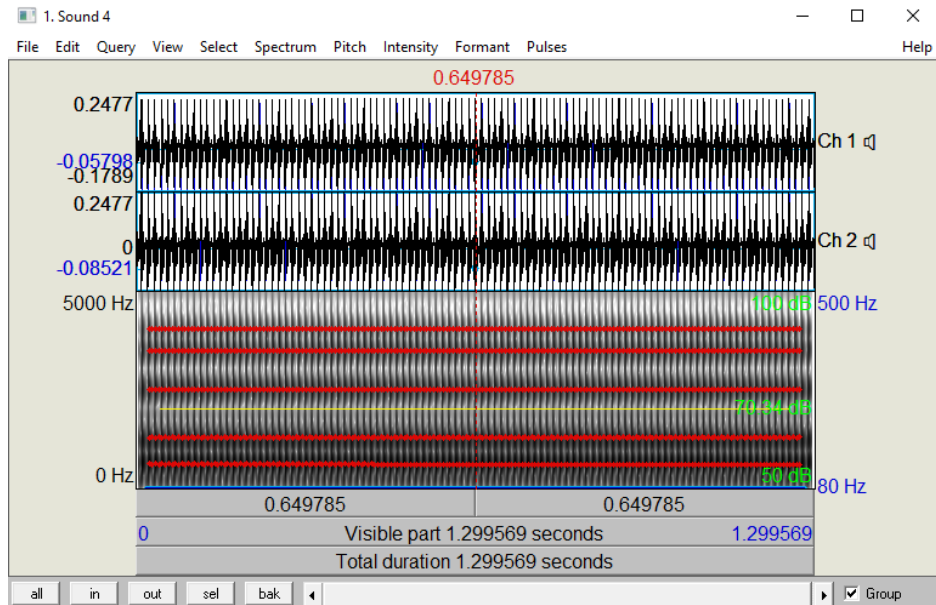


Figura 3 Interfaz del software Praat.

2.1.11. EVAPER

EVAPER es un sistema de evaluación perceptual basado en el método psico-acústico de producción de la magnitud y “matching intramodal”. Este método consiste en evaluar perceptualmente el grado de aspereza de la voz de un paciente cuando se contrastan los desvíos de las respuestas obtenidos con el método clásico de estimaciones numéricas. La evaluación se realiza ajustando los atributos como soplo, astenia, tensión y temblor; y una vez obtenida la señal de referencia, el usuario puede definir su estrategia identificando las señales de referencia más próximas para finalmente realizar la comparación por pares para cada rasgo, (Gurlekian & Torres, 2016) (Figura 4) .



Figura 4 Interfaz del software Evaper, Adaptado de: Informe Técnico, Gurlekian & Torres, 2016.

2.1.12. Coeficiente de Spearman

El coeficiente de correlación de Spearman es un índice que examina la fuerza y la dirección de la relación monótona entre dos variables continuas u ordinales (*Figura 5*). Este coeficiente al igual que el coeficiente de correlación de Pearson, mide su precisión en una escala de -1 a 1, pero éste índice no es tan sensible a los valores que están muy distantes a los valores reales (*Matlab, 2021; Minitab 18, 2019*), el coeficiente de Spearman lo podemos expresar siguiendo la siguiente ecuación:

$$Rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Ecuación 9

Donde d_i es la diferencia entre los rangos de las dos columnas y n es la longitud de cada columna.

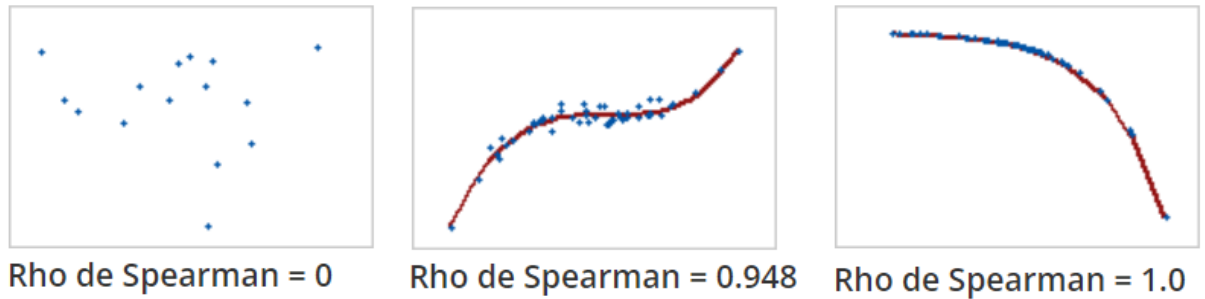


Figura 5 Valores del coeficiente de Spearman, adaptado de: Interpretar los resultados clave para Correlación, Minitab 18, 2019, Researchgate: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/correlation/interpret-the-results/key-results/#:~:text=Relaci%C3%B3n%20positiva%20fuerte%3A%20Rho%20de,variables%20aumentan%20al%20mismo%20tiempo.>

2.1.13. Machine learning

El machine learning o también conocido como aprendizaje automático, es una disciplina científica en el campo de las ciencias de la computación que, de acuerdo a Arthur Samuel en 1959, les da a las computadoras la habilidad de aprender sin ser explícitamente programadas, aprender en este contexto quiere decir identificar patrones complejos en millones de datos. Una vez se haya realizado el procedimiento de aprendizaje, las máquinas son capaces de predecir comportamientos futuros, e incluso de mejorar de manera automática con el tiempo sin intervención humana. A lo largo de la literatura podemos encontrar varios métodos de machine learning como por ejemplo: machine learning supervisado, y no supervisado, la diferencia de estos métodos radica en la intervención

humana para su aprendizaje ya que para el método supervisado, se debe etiquetar y clasificar los datos para el entrenamiento, mientras que en el método no supervisado, el algoritmo no necesita ninguna intervención, ya que realiza un análisis de patrones en búsqueda de alguna relación entre los datos (González, 2021; Prieto, 2017) (Figura 6).

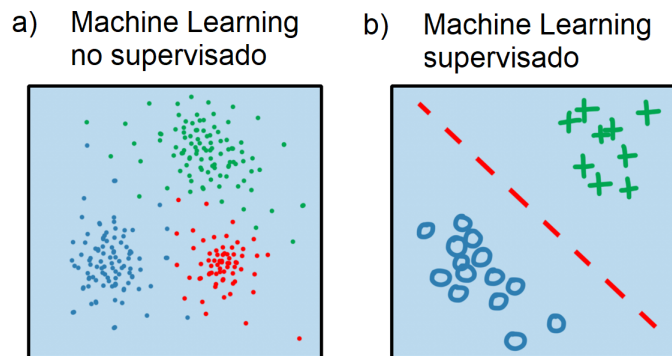


Figura 6. a) Metodología machine learning no supervisada, **b)** Metodología machine learning supervisada; adaptado de Reinforcement learning, MathWorks, 2021, Researchgate:<https://la.mathworks.com/discovery/reinforcement-learning.html>

2.1.14. SVM

El método de clasificación-regresión Máquinas de Soporte Vectorial (Support Vector Machines, SVMs) fue desarrollado en la década de los 90, y aunque originalmente fue desarrollada como un método de clasificación binaria, su aplicación se ha extendido a problemas de clasificación múltiple y regresión. En la actualidad se considera como uno de los mejores métodos de clasificación automática para amplios casos de estudio.

El SVM está fundamentado en el Maximal Margin Classifier, que a su vez se basa en el concepto de hiperplano, el cual nos dice que el hiperplano es un subespacio de una dimensión menos que su espacio ambiental, siendo este el que divide las diferentes clases.

2.1.15. Kernels

Un truco conocido en estos clasificadores es el uso de kernels, el cual consiste en inventar una dimensión nueva en la que podamos realizar la clasificación de las clases con una mayor precisión; en la siguiente figura (Figura 7) se logra observar cómo al añadir una nueva dimensión, se puede separar fácilmente las dos clases con una superficie de decisión (*Gonzalez, 2019; Martinez, 2019; Vaerenbergh & Santamaría, 2018*).

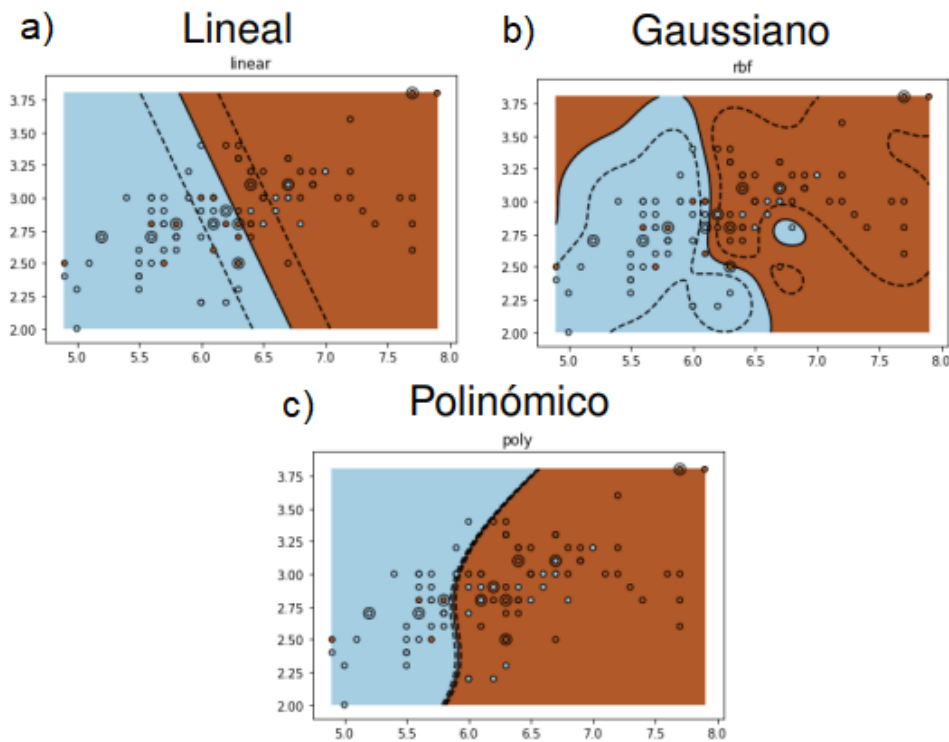


Figura 7 En esta gráfica se muestra el funcionamiento de algunos de los tipos de kernels que son implementados en los clasificadores supervisados, a) Kernel lineal, b) Kernel Gaussiano, c) Polinómico Adaptado de: *Métodos kernel para clasificación*, S. Van Vaerenbergh, I. Santamaría, 2018, Universidad de Cantabria: https://gtas.unican.es/files/docencia/APS/apuntes/07_svm_kernel.pdf.

Existen diferentes tipos de kernel, a continuación se explican brevemente algunos de ellos:

- **Kernel lineal:** puede ser utilizado como el producto normal de dos observaciones dadas. El producto entre dos vectores es la suma de la multiplicación de cada par de valores de entrada.
- **Kernel polinomial:** este es una forma más generalizada del núcleo lineal. El kernel polinomial no solo observa las características dadas por las muestras de entrada para

determinar su similitud, sino también las combinaciones de éstas. Con “n” características originales y “d” grados de polinomio produce n^d características expandidas.

- **Kernel gaussiano:** puede mapear un espacio de entrada en un espacio dimensional infinito.

2.1.16. Knn

“Dime con quién andas y te diré quién eres”, este viejísimo refrán resume a la perfección la forma en la que funciona el algoritmo de knn (K Nearest Neighbours) o K vecinos más cercanos por su traducción al español; Este es un algoritmo de aprendizaje supervisado y su concepto es realmente sencillo, el funcionamiento de este algoritmo se basa en clasificar cada nuevo dato en el grupo que corresponda acorde a la cantidad de vecinos más cercanos a un grupo o a otro (Figura 8). explicándolo de un modo más completo, el algoritmo calcula las distancias del nuevo elemento a cada uno de los elementos ya existentes, luego ordena estas distancias de menor a mayor para seleccionar el grupo al que pertenece el nuevo elemento, el cual será el que presente mayor frecuencia con distancias menores (*MERKLE, 2020*).

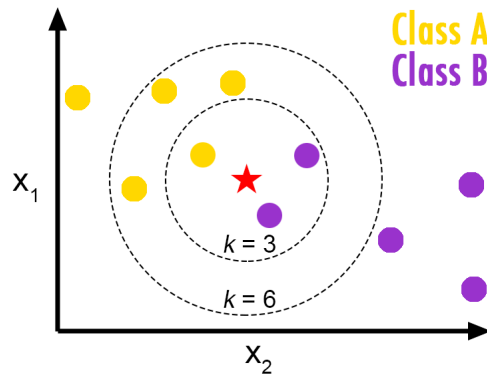


Figura 8. Representación gráfica del modelo de clasificación Knn, Adaptado de: *Introducción al Machine Learning #9 - K Vecinos más cercanos (Clasificación y Regresión)*, Salcedo L, 2018, Researchgate: <https://pythondiario.com/2018/01/introduccion-al-machine-learning-9-k.html>.

2.1.17. Naive bayes

Los modelos de Naive Bayes se pueden considerar como una clase especial de algoritmos de inteligencia artificial, específicamente de machine learning que se basan en las técnicas estadísticas basadas en el teorema de Bayes (Ecuación 10).

Se llaman “Naive” o lo que en español sería “Inocente” debido a que en ellos se asume que las variables que van a ser empleadas como predictoras son independientes cada una entre sí, o explicado de otra forma, se asume que la presencia de una cierta característica en un conjunto de datos no está relacionada en lo absoluto con la presencia de cualquier otra característica, (Roman Victor, 2019) (Figura 9).

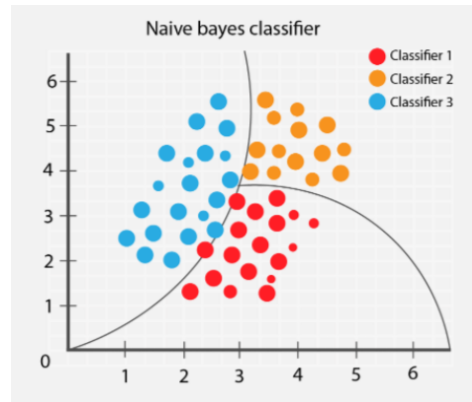


Figura 9. Representación gráfica del modelo de clasificación Naive Bayes, Adaptado de: *Introduction to Naïve Bayes Classifier*, Yang, 2019, Researchgate: <https://towardsdatascience.com/introduction-to-naive-bayes-classifier-fa59e3e24aaf>.

$$P(A | B) = \frac{P(B | A) * P(A)}{P(B)} = \frac{P(B | A) * P(A)}{P(A) * P(B | A) + P(-A) * P(B | -A)}$$

Ecuación 10 Modelo matemático del teorema de Bayes

El modelo matemático del teorema de Bayes (Ecuación 10), se entiende como la probabilidad a priori de que la hipótesis A sea verdadera ($P(A)$), multiplicado por la probabilidad del evento dado que dicha hipótesis sea verdad ($P(B | A)$); este resultado es dividido por la probabilidad total de que ocurra ese evento ($P(B)$), es decir, la combinación de que la hipótesis sea verdad dado que el evento sea verdadero, sumada la probabilidad de que la hipótesis sea falsa dado que el evento sea verdadero ($P(A) * P(B | A) + P(-A) * P(B | -A)$). (Muller, 2017).

2.1.18. Cross validation

La validación cruzada, o K-fold cross validation por su traducción en inglés, es un proceso de validación iterativo, el cual consiste en dividir los datos disponibles de forma aleatoria

en K grupos que tengan un tamaño similar entre ellos, $k-1$ grupos son utilizados para entrenar el modelo, y el grupo faltante es utilizado como validación; este proceso se repite k veces variando el grupo que se emplea como validación en cada ciclo (Figura 10). El proceso genera K estimaciones del error cuyo promedio es empleado como estimación final (Rodrigo, 2016).

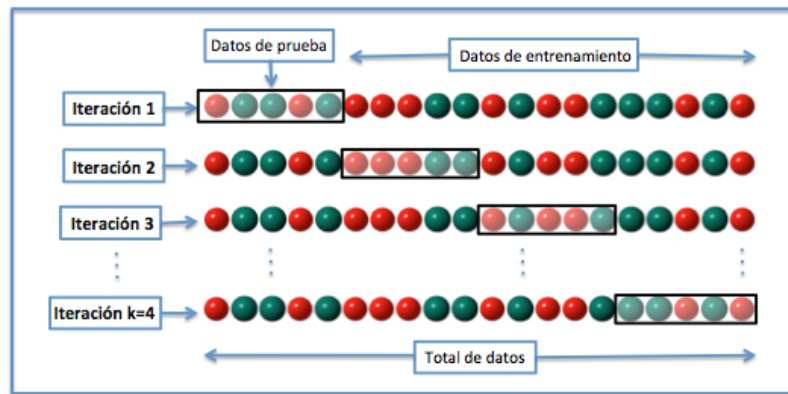


Figura 10 Explicación del funcionamiento del método de validación cruzada (Cross Validation), adaptado de: Validación cruzada de K iteraciones con $K=4$, Joanneum, 2006, Researchgate:
https://www.researchgate.net/figure/Figura-2-Validacion-cruzada-de-K-iteraciones-con-K4-Joanneum-2005-2006_fig2_325766285.

2.1.19. Matriz de confusión

En el popular campo de la inteligencia artificial, y más específicamente en el del machine learning, una matriz de confusión cumple el papel de una herramienta que nos permite visualizar el desempeño obtenido por un algoritmo de aprendizaje supervisado, cada columna de esta matriz vendría a representar cada el número de predicciones que realiza el algoritmo sobre clase dada, a su vez el número de filas representa las instancias en la clase

real, resumiendo, se tiene que la matriz de confusión nos permite visualizar cuáles y cuántos errores y aciertos está produciendo nuestro modelo al momento de ser entrenado con los datos dados (Barrios, 2019) (Figura 11).

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Figura 11. Representación de la matriz de confusión, adaptado de Evaluación de modelos de clasificación, Zelada, 2017, Researchgate: <https://rpubs.com/chzelada/275494>.

2.1.20. Curva ROC

En el espectro de los test diagnósticos, un gran número de estos reportan sus resultados de una forma cuantitativa, utilizando escalas de valores continuas, las curvas ROC (receiver operating characteristic curve) se constituye como un método para determinar la exactitud diagnóstica de estos test, siendo utilizadas específicamente para 3 propósitos: evaluar la capacidad discriminativa del test diagnóstico, determinar el punto de corte de una escala continua en el que se alcanza la sensibilidad y especificidad más alta (Cerde & Cifuentes, 2012) (Figura 12).

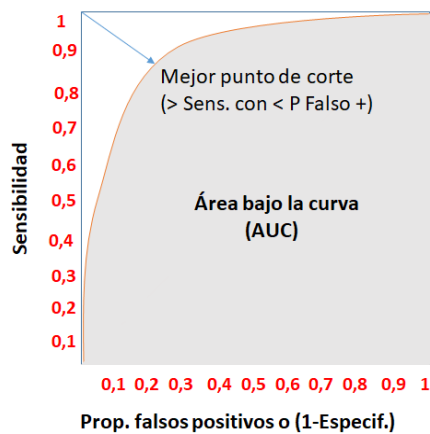


Figura 12. Modelo de la Curva ROC, adaptado de Curva Operador-Receptor (ROC), por Obero C. 2015 Módulo de Bioestadística Researchgate: <https://modulodeestadistica.wordpress.com/curva-operador-receptor-roc/>

Donde el mejor punto de corte de la prueba, es aquel, que arroja una mayor proporción de resultados positivos verdaderos (sensibilidad) con el menor número de resultados falsos positivos (1 – especificidad). Entonces el mejor punto de corte se calcula gráficamente, al proyectar una línea recta cuyo origen está en la esquina superior izquierda de la gráfica y su final interseca la curva ROC donde resulta la línea más corta entre ambos puntos (ángulo superior izquierdo y curva ROC).

CAPÍTULO 3

3.1. DISEÑO METODOLÓGICO

El desarrollo de este proyecto se ha dividido en 3 fases, estructurado bajo el marco de referencia CDIO y siguiendo los objetivos específicos planteados anteriormente.

3.1.1. Fase 1: Establecer el diseño técnico y teórico de los parámetros acústicos y estadísticos implementados en el proyecto para el análisis de la voz.

3.1.1.1. Selección de las características acústicas y estadísticas para este proyecto mediante la búsqueda de información en revistas, artículos científicos y en base a herramientas virtuales de análisis de voz estandarizadas.

La selección de características se llevó a cabo mediante una revisión bibliográfica de artículos científicos y softwares presentes en el mercado, donde se detallan aquellas características más implementadas y conocidas en el campo de la logopedia para el análisis de voz y detección de patologías en general; en la revisión bibliográfica realizada se pueden destacar el software Praat (*Boersma & Weenink, 2011*), y los siguientes artículos científicos (Dodero et al., 2005; Delgado et al., 2017; Teixeira & Gonçalves, 2014).

Una vez finalizada la revisión bibliográfica, las características seleccionadas para llevar a cabo el desarrollo de esta herramienta fueron las siguientes: Jitter local, Jitter local

absoluto, Shimmer local dB, frecuencia fundamental, potencia de armónicos, densidad espectral de potencia, energía de la señal, curtosis, asimetría y RMS.

3.1.1.2. Adquisición de la base de datos destinada a la extracción de parámetros, la cual está conformada por voces sintéticas con diversos grados de afectación.

Para este proyecto, la universidad Antonio Nariño nos brindó la base de datos de las voces sintéticas creadas por medio de la aplicación EVAPER, la cual contiene 500 señales con los diferentes grados de afectación de la voz para la escala GRBAS; de esas 500 señales, 294 señales eran del género masculino y 206 señales del género femenino, además, cabe destacar que se encontraban en formato WAV, con una duración aproximada de 1 segundo, y con 2 canales por señal, es decir, cada señal está en modo estéreo.

3.1.1.3. Reacondicionamiento de la base de datos por la presencia de errores en los diagnósticos del grado de afectación para cada voz: debido a inconsistencias producidas en la generación original de la base de datos, por lo que se debió producir cada voz de nuevo para corregir el grado diagnosticado por el software.

La base de datos previamente descrita presentaba un error en el parámetro G, ya que la información presente en ese grado no coincidía con los diversos niveles de los parámetros RBAS (para todas las señales el parámetro G marcaba la señal como “Normal”), debido a este particular error, se tuvo que corregir ingresando a la aplicación EVAPER y determinando de nuevo el grado para cada una de las 500 señales.

Para el funcionamiento del software Evaper es necesario portar una llave física, dando acceso a los usuarios únicamente autorizados, y debido a las diversas complicaciones de transporte ocasionadas por la actual de pandemia por Covid-19, se tuvo que implementar la plataforma Teamviewer para llevar a cabo el uso del software de voz, ésta plataforma de conectividad remota basada en la nube nos permite realizar una conexión a cualquier dispositivo a través de la red de internet, TeamViewer conecta computadoras, smartphones, servidores, con conexiones rápidas y de alto rendimiento, lo que para este caso, resultó óptimo para realizar la respectiva conexión y por ende la corrección del problema. Una vez solucionado el problema presente en el parámetro G, ya fue posible continuar con el desarrollo de la fase 2.

3.1.2. Fase 2. Implementar el diseño técnico, desarrollando el algoritmo encargado de extraer las características de las señales y los modelos de clasificación para cada parámetro.

Para llevar a cabo esta fase, se debe recordar que las señales implementadas se encuentran en modo estéreo (2 canales), y con el fin de aumentar la cantidad de datos de entrenamiento, se trató cada canal de la señal (mono) de manera independiente, ya que, aunque representen la misma señal de voz, poseen variaciones que las vuelven numéricamente diferente (Figura 13).

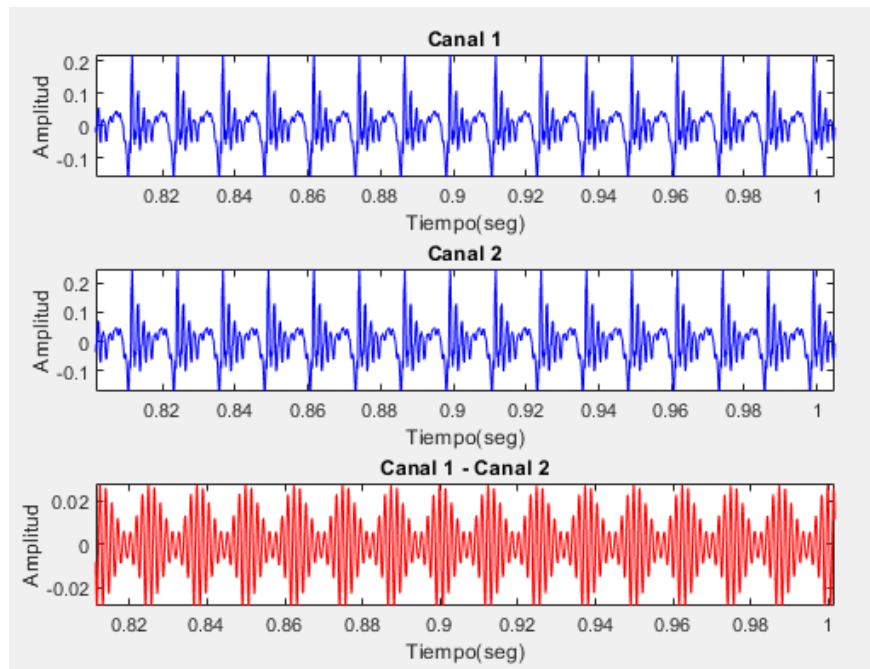


Figura 13 En esta figura se muestra la representación gráfica de la señal estéreo segmentada por cada canal (señales azules), y la resta llevada a cabo dato por dato entre los 2 canales (señal roja), mostrando una cierta diferencia numérica entre cada uno.

3.1.2.1 Estructuración del algoritmo dedicado a la extracción de estas características acústicas, estadísticas y espectrales, realizado a partir de sus respectivos modelos matemáticos.

La estructuración de este algoritmo se llevó a cabo en el software Matlab y se implementaron los modelos matemáticos de las características a extraer; algunas de éstas como: Jitter local, Jitter local absoluto y Shimmer dB, se realizaron manualmente, es decir, se llevó a cabo el desarrollo de cada línea presente en el algoritmo para la obtención de estos datos, y para obtener las características como: Curtosis, Asimetría, RSM y power Spectrum Density, se implementaron funciones predeterminadas de Matlab.

A continuación, se puede observar la estructuración del algoritmo de manera gráfica, mediante un diagrama de flujo (Figura 14, Anexo 1).

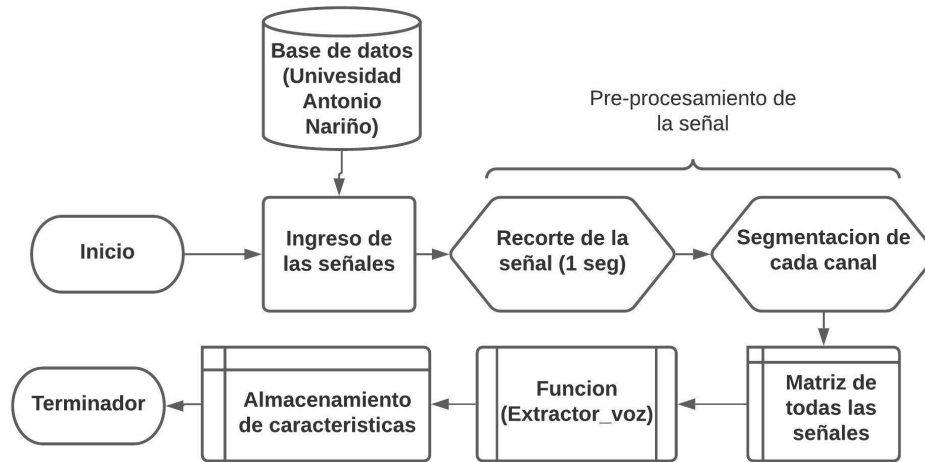


Figura 14. Diagrama de flujo del Preprocesamiento de las señales.

Para realizar la extracción de características, se han realizado dos algoritmos, uno de ellos (figura 13) es el encargado de leer los datos de la base de datos, realizar el preprocesamiento de las señales como la estandarización de los tiempos de reproducción a 1 segundo y la segmentación de cada canal de la señal, posteriormente, este algoritmo concatena las señales de la base de datos en una única matriz para que mediante la función propia “Extractor_voz” realice la extracción de las respectivas características.

El segundo algoritmo (Anexo 1), es el encargado de realizar el análisis de cada señal ingresada; Este algoritmo comienza realizando la eliminación de la componente DC y la normalización de la señal, a continuación determina la frecuencia fundamental y realiza una segmentación de la señal con una ventana de 400 milisegundos con un traslapamiento de

200 milisegundos, una vez se realiza este ventaneo se lleva a cabo la extracción de las características, las cuales se ha dividido en distintos grupos como: “Grupo Jitter”, “Grupo Shimmer”, “Grupo Espectral” y “Grupo estadístico + Energía”.

- Grupo Jitter

Para calcular las características de Jitter, las cuales consisten en el análisis de la diferencia temporal entre cada pico de la señal, se implementó la función de detección de pico mediante un umbral determinado (umbral = 0.9), pero, un problema en ésta metodología es que por las diversas variaciones de amplitud de cada pico existente en las señales, se generan ciertas discrepancias, debido a que el umbral en ciertos casos detectaba valores que se pueden interpretar como ruido, o no lograba detectar algunos de los picos necesarios para el análisis (figure 15).

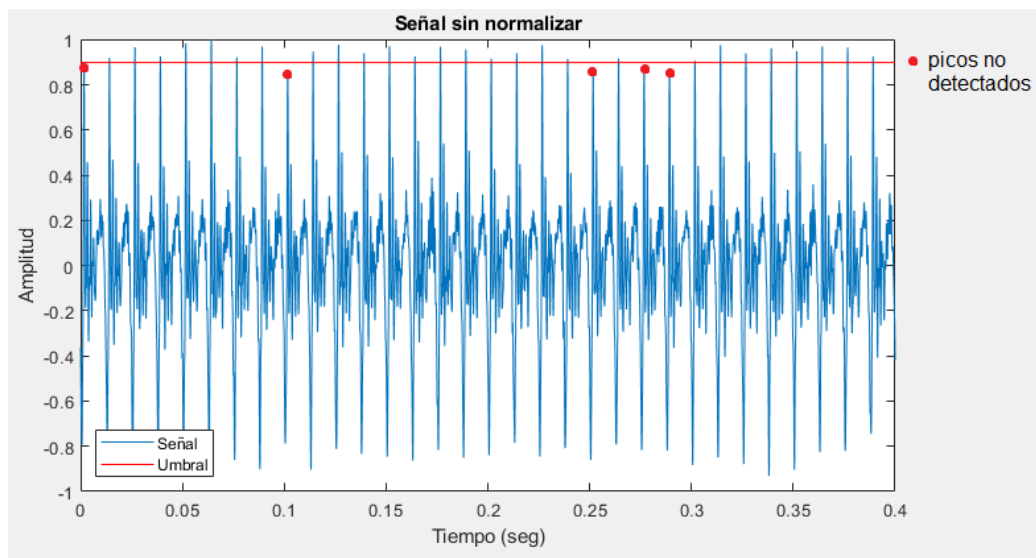


Figura 15 En esta figura se muestra la incapacidad del algoritmo para realizar la detección de picos principales de la señal con un umbral determinado igual a 0.9.

Debido a este problema, se tuvo que incorporar al algoritmo un sistema encargado de reducir este error a su mínima expresión, la cual consta de 3 pasos:

1. Para cada ventana de la señal se realiza una subventana con una duración igual a 12 milisegundos, la cual es multiplicada por la frecuencia de muestreo (**Ecuación 11**), para determinar el número de puntos existentes en ese periodo de tiempo.

$$\text{Subventaneo} = 0.012 * F_m$$

Ecuación 11

Se realizaron varias pruebas para determinar el periodo de subventaneo óptimo que permita conseguir los mejores resultados, obteniendo que ese valor debe ser aproximadamente 12 milisegundos para aquellas señales que poseen una frecuencia de muestreo igual a 44.100, por ende, la herramienta computacional ajusta la frecuencia de muestreo de la señal para cumplir con este requerimiento.

2. Posteriormente se realiza la normalización y eliminación de las componentes DC de cada subventana, y por último, su concatenación formar nuevamente la ventana de 400 milisegundos, y es en este punto, donde ya es posible implementar un umbral de detección con un valor fijo, esto se puede ver en la Figura 16.a, pero como se puede observar, varios de los picos no fundamentales, que para este caso se los considera como ruido, logran alcanzar el umbral establecido, lo que podría desencadenar posible errores en los resultados, y esto debido al paso anterior de subventaneo.

3. Para corregir los errores presentes en el sub ventaneo, se realiza la detección de solo los picos principales; Para este paso, el algoritmo realiza un análisis de la cantidad de picos presentes en la ventana de 400 milisegundos y mediante la frecuencia fundamental de la señal, determina la cantidad de picos que deben existir en ese segmento y la separación promedio que deben tener; esta comparación se realiza con el fin de eliminar aquel conjunto de picos cercanos que representan ruidos en la señal, dejando simplemente el que cumpla con la condición de distancia entre picos. Los resultados obtenidos durante este tercer paso se pueden observar en la figura 16b (la condición de distancia entre picos está dada por la **ecuación 13**, la cual está en relación con la **ecuación 12** en donde se deja un margen de error de $\frac{1}{6}$ veces la separación promedio de picos).

$$Y = Fm/F0$$

Ecuación 12

Donde Y equivale a la cantidad de picos que debería tener cada señal.

$$Cond = Y - (Y/6)$$

Ecuación 13

Donde Cond, equivale a la condición de distancia entre picos para la eliminación de los conjuntos de datos que representan ruido en la señal.

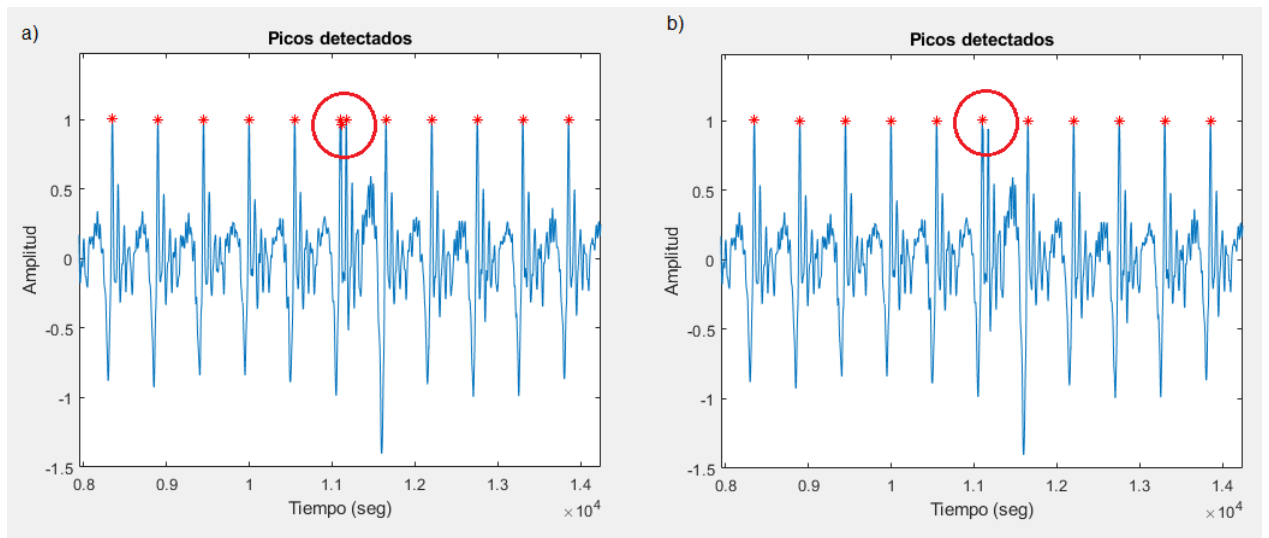


Figura 16, a) Se puede observar en esta figura la presencia del error de subventaneo, ya que el algoritmo detecta varios picos de ruido cercanos al pico principal, **b)** Se puede detallar que el error previamente mencionado se ha reducido significativamente gracias a la ejecución del tercer paso del sistema corrección de errores en el grupo Jitter (ecuación 13).

Una vez se ha realizado el anterior procesamiento de señales, se procede a la extracción del Jitter local y el Jitter local absoluto implementando sus respectivos modelos matemáticos (Ecuación 1, 2), este proceso se realiza para cada ventana de 400 milisegundos, y los resultados obtenidos son almacenados en una tabla.

- Grupo Shimmer

Para calcular las características del grupo Shimmer, las cuales consisten en el análisis de la amplitud total de cada pico de la señal, se implementó la función de detección de picos mediante un umbral determinado al igual que en el grupo Jitter (umbral = 0.9), logrando generar los mismos problemas en la detección de picos producido por sus diversas

variaciones de amplitudes, para dar solución a estos problemas específicamente en el grupo Shimmer, se adicionaron al algoritmo 3 pasos los cuales son:

1. Para cada ventana de la señal se realiza un subventana en relación a la frecuencia de muestreo (**Ecuación 11**), esto con el fin de realizar una segmentación uniforme a lo largo de la ventana y aislar cada pico presente en ésta.
2. Determinación de los valores máximos y mínimos de cada pico principal presente en cada subventana (Figura 17).
3. Concatenación de los anteriores valores en vectores, destinados para el cálculo del Shimmer dB.

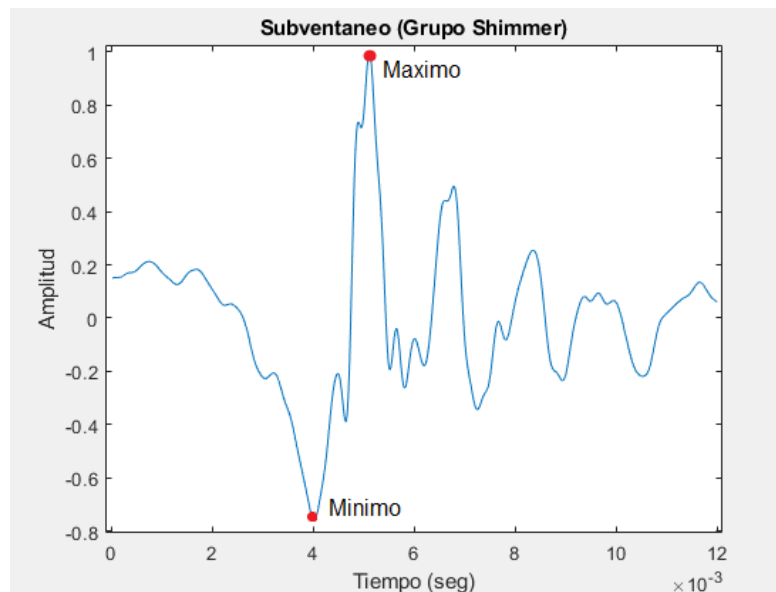


Figura 17 Ejemplo de una subventana del grupo Shimmer, donde se evidencia la detección de los valores máximos y mínimos de cada pico principal.

Una vez se ha realizado el anterior procesamiento de señales, se procede a la extracción del Shimmer dB implementando su respectivo modelo matemático, este proceso se realiza para cada ventana de 400 milisegundos, y los resultados obtenidos son almacenados en una tabla.

- Grupo Espectral

Para calcular las características Espectrales de cada ventana de la señal, se realiza una normalización de la ventana y se calcula su densidad espectral de potencia mediante una función intrínseca en Matlab, y a partir de ahí, se obtienen los valores de frecuencia fundamental, armónicos y su respectiva potencia (figura 18, 20); para detectar estos valores se implementa la función de detección de picos sobre un umbral determinado para la señal correspondiente a la densidad espectral de potencia de la ventana (umbral = 0.01), como se puede ver en la figura 19, posteriormente cuando se obtiene la posición de los picos detectados que corresponden a cada armónico, se lleva estos datos a la figura 18 para determinar la frecuencia de los picos detectados y su respectiva potencia dada en deciBelios (dB), además, como dato complementario, se calcula la media de la densidad espectral de potencia, y se suma este dato a las características de voz extraídas.

Una vez obtenidas estas características espectrales para cada ventana de 400 milisegundos, los resultados son almacenados en una tabla específica.

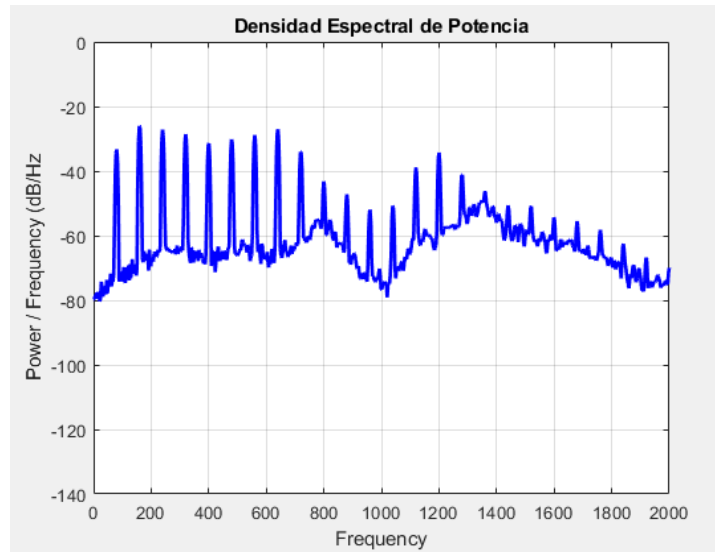


Figura 18 Gráfica de la densidad espectral de potencia obtenida para la ventana de cada señal de voz.

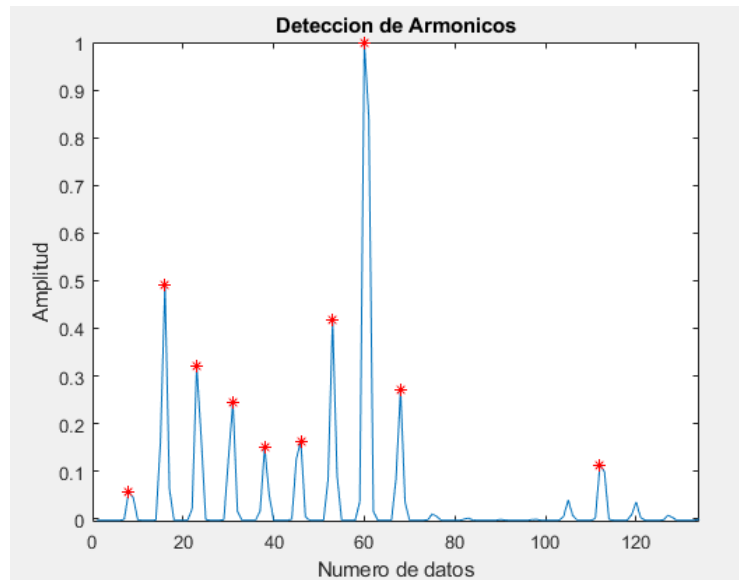


Figura 19 Ejemplo de la detección de armónicos que realiza el algoritmo, mediante la señal normalizada de densidad espectral de potencia.

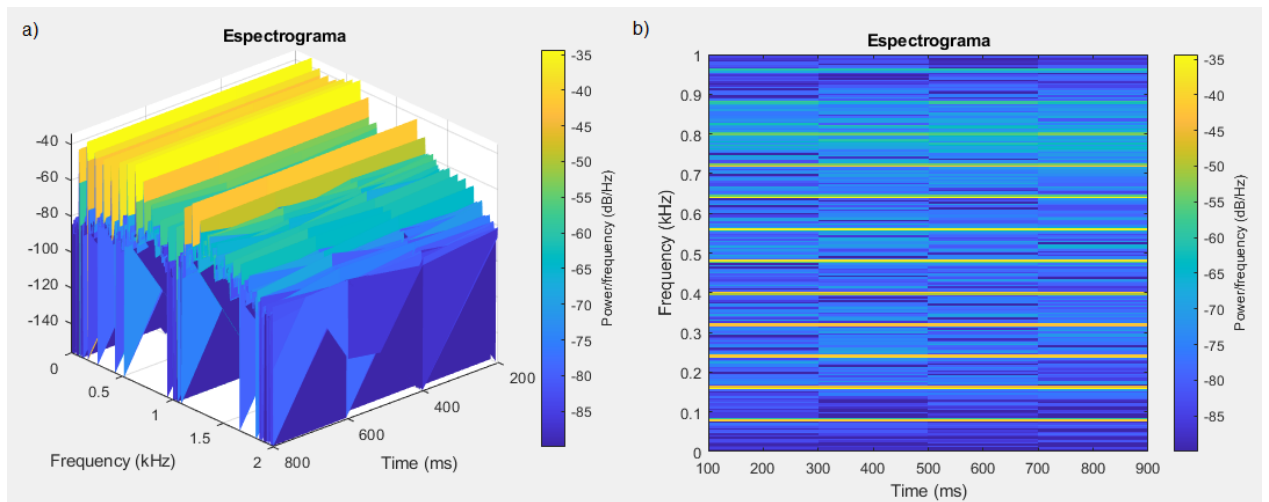


Figura 20 a) Gráfica de espectrograma 3D obtenida para la señal de voz donde se muestran los ejes de tiempo, frecuencia, y potencia, **b)** Gráfica de espectrograma 2D obtenida para la señal de voz donde se muestran los ejes de tiempo y frecuencia, y en donde gracias a la variación de colores se puede identificar la potencia de cada segmento.

- Grupo estadístico + Energía

Para calcular las características del grupo estadístico de cada ventana de la señal, se implementaron los comandos predefinidos por Matlab que llevan a cabo esta tarea, logrando obtener los valores de asimetría, curtosis y RMS para cada ventana de 400 milisegundos; por otro lado, Matlab no disponía de un comando dedicado a la extracción de la energía de una señal, así que se llevó a cabo la implementación de la ecuación 8, la cual consiste en realizar una sumatoria del cuadrado de los valores de cada punto (x,y), en el rango de -infinito hasta +infinito.

Una vez obtenidas estas características estadísticas y la energía de la señal para cada ventana de 400 milisegundos, los resultados son almacenados en una nueva tabla.

3.1.2.2. Validación de las características acústicas obtenidas por el algoritmo (Jitter local, Jitter local absoluto y Shimmer dB) en comparación con el software Praat al implementar el 10% de las señales de la base de datos en modo de prueba.

Esta fase consiste en realizar la validación del algoritmo, más específicamente de los valores obtenidos de Jitter local, Jitter local absoluto y Shimmer dB, en comparación con el software Praat, ya que la intersección de las características entre Praat y el algoritmo desarrollado son las anteriormente mencionadas.

Esta comparación consiste en obtener los valores de correlación de Spearman, RMS y el error absoluto de cada característica, para ello se implementó el 10% de la base de datos de manera aleatoria, la cual, el 50% de las señales corresponden al género masculino y el 50% restantes al género femenino; cabe destacar que se utilizó los dos canales de cada señal como un canal mono independiente y una única ventana de 400 milisegundos para cada señal mono; además, un dato a destacar es que el software Praat con algunas de las señales no logró obtener un resultado para las características anteriormente mencionadas “undefined”, cómo se logra observar en la Figura 21.

a)

```
Praat Info
File Edit Search Convert Font Help

-- Voice report for 1. Sound 36
Time range of SELECTION
  From 0 to 0 seconds (duration: 0.400000 seconds)
Pitch:
  Median pitch: --undefined-- Hz
  Mean pitch: --undefined-- Hz
  Standard deviation: --undefined-- Hz
  Minimum pitch: --undefined-- Hz
  Maximum pitch: --undefined-- Hz
Jitter:
  Jitter (local): --undefined--
  Jitter (local, absolute): --undefined-- seconds
Shimmer:
  Shimmer (local): --undefined--
  Shimmer (local, dB): --undefined-- dB
```

b)

```
Praat Info
File Edit Search Convert Font Help

-- Voice report for 2. Sound 40 --
Time range of SELECTION
  From 0 to 0 seconds (duration: 0.400000 seconds)
Pitch:
  Median pitch: --undefined-- Hz
  Mean pitch: --undefined-- Hz
  Standard deviation: --undefined-- Hz
  Minimum pitch: --undefined-- Hz
  Maximum pitch: --undefined-- Hz
Jitter:
  Jitter (local): --undefined--
  Jitter (local, absolute): --undefined-- seconds
Shimmer:
  Shimmer (local): --undefined--
  Shimmer (local, dB): --undefined-- dB
```

Figura 21 Reporte generado por Praat, donde se logra apreciar que las señales 35 (a), y 40 (b), no obtuvieron ningún valor debido a que Praat no pudo reconocerlas. (Los resultados obtenidos durante este proceso de correlación (fase 2.2) se muestran en la sección de resultados).

3.1.2.3. Extracción y registro de las características acústicas y estadísticas para el 80% de la base de datos.

Una vez validado el algoritmo encargado de realizar la extracción de las características vocálicas de cada señal, el paso siguiente es reunir el 80% de los datos de la base de datos, tomados de manera aleatoria para cada género, y mediante la función “Extractor Voz”, recolectar todas las características tales como: Jitter local, Jitter local absoluto, Shimmer local dB, frecuencia fundamental, potencia de armónicos, densidad espectral de potencia, energía de la señal, curtosis, asimetría y RMS, para posteriormente guardarlos en dos tablas, uno para las voces masculinas y otro para las voces femeninas, para posteriormente proceder con la siguiente etapa, la cuales consiste en el desarrollo de los modelos de clasificación.

3.1.2.4. Desarrollo de los diversos modelos de clasificación para cada parámetro GRBAS discriminado por género, implementando para ello el 80% de la base de datos mediante el método de entrenamiento - validación “cross validation” utilizando 10 folders.

Esta fase se llevó a cabo mediante la implementación de la aplicación de Matlab “Classification Learning” (Figura 22), la cual nos permite de manera gráfica e intuitiva la configuración del método de entrenamiento y la generación de varios modelos de clasificación, mostrando además el valor porcentual de precisión, matriz de confusión y la curva ROC de cada modelo. Para este proyecto se llevó a cabo el método de entrenamiento - validación “cross validation” con un 10 folders, y se llevó a cabo la construcción de 14

modelos de clasificación, los cuales son: SVM Lineal, SVM Cuadrático, SVM Cúbico, SVM con Kernel Gaussiano Fino, SVM con Kernel Gaussiano Medio, SVM con Kernel Gaussiano Grueso, Naive Bayes con Kernel Gaussiano, Naive Bayes Kernel, Knn Fino, Knn Medio, Knn Grueso, Knn Coseno, Knn Cúbico y Knn Ponderado.

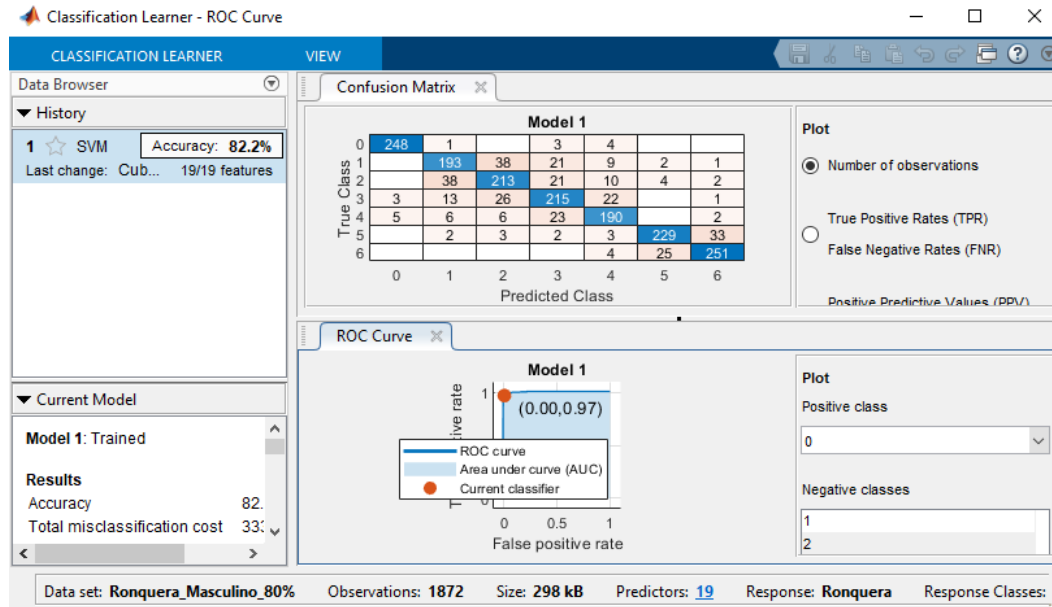


Figura 22 Interfaz gráfica de la aplicación de Matlab “Classification Learning”.

Para la determinación de los modelos de clasificación se implementó las características vocálicas del 80% de los datos de la base de datos segmentados por cada género, es decir, se separó la base de datos por género, y se eligió de manera aleatoria el 80% de las voces masculinas y el 80% de las voces femeninas, en donde se extrajeron sus parámetros previamente en la etapa 2.3, e implementando las tablas de las características de las señales de masculino y femenino, y mediante la aplicación “Classification Learning”, se realizó el entrenamiento de los diferentes modelos de clasificación para cada parámetro RBAS;

posteriormente los datos arrojados por cada modelo como su precisión y la matriz de confusión fueron almacenados para su futuro análisis y selección del mejor modelo. Para el parámetro G, se realizó el mismo procedimiento anterior, pero su diferencia radica en los datos de entrada para el clasificador, ya que los modelos entrenados para la determinación del parámetro G reciben los parámetros RBAS de cada señal.

Cabe destacar sobre la presencia de un error en el desarrollo de los modelos de clasificación para los parámetros de Ronquera y Grado en el género femenino, ya que por un error en la base de datos, no se presentaban la información suficiente para llevar a cabo el desarrollo de sus respectivos modelos, ya que: según la aplicación EVAPER la ronquera se puede variar entre un rango de 0 a 6, y el grado se muestra en un rango de 0 a 3, pero la base de datos para el género femenino solo posee la ronquera en un rango de 0 a 4 y el grado en un rango de 1 a 3.

Una solución para el anterior inconveniente fue el generar los modelos de clasificación de ronquera y grado implementando el 80% de los datos totales, donde se encuentran mezclados las señales del género masculino y femenino, además, también se generaron los modelos para el resto de parámetros siguiendo las anteriores especificación de los datos, con el fin de realizar una comparativa entre los modelos dedicados para género (**modelos dedicados**) y aquellos que no discrimina el género (**modelos no dedicados**).

3.1.3. Fase 3: Validar el funcionamiento de la herramienta desarrollada en comparación con señales sintéticas proporcionadas por la Universidad Antonio Nariño, y señales reales proporcionadas por la Universidad del Valle.

3.1.3.1. Comparación de los resultados obtenidos entre los diversos modelos de clasificación, en donde se seleccionará para cada parámetro aquellos modelos que presenten los mejores resultados de clasificación.

Una vez obtenido el modelo de clasificación para cada parámetro según su género, es tiempo de comparar los modelos de clasificación entre los modelos dedicados y los modelos no dedicados, en busca de aquel que presente los mejores resultados de clasificación. Este nuevo proceso se realiza para cada parámetro e implementando el 20% de los datos restantes dedicados para la validación, los cuales han sido discriminados según su género, es decir, se obtuvo el 20% de los datos para el género masculino y el 20% de los datos para el género femenino; Para realizar la selección del modelo de clasificación que obtenga el mejor desempeño, se llevará a cabo el análisis del nivel porcentual de la exactitud y sus matrices de confusión. Cabe destacar que por la existencia del error anteriormente nombrado para el género femenino, la comparación de los parámetros de ronquera y grado se realizaron entre los modelos dedicados para el género masculino y los modelos no dedicados.

3.1.3.2. Validación de la herramienta implementando el 20% de los datos restantes y comparando los resultados con respecto al informe oficial de diagnóstico de la base de datos de la Universidad Antonio Nariño, en donde se examine el error y el porcentaje de exactitud de la herramienta.

Una vez se tienen los modelos destinados para la conformación de la herramienta, se procede a crear un nuevo script donde según su género se concatenan todos los modelos de clasificación de la escala GRBAS, y se realiza nuevamente una validación final que describa la eficiencia total de la herramienta para la evaluación y clasificación de la calidad de voz, para ello se implementará nuevamente el 20% de los datos dedicados para la validación y se realizará por cada género, obteniendo su respectivo nivel de exactitud y matriz de confusión.

3.1.3.3. Re-escalamiento de los resultados obtenidos por parte de la herramienta desarrollada

El objetivo de este apartado es re-escalar los valores generados por la herramienta computacional, a valores los cuales los logopedas han sido entrenados para interpretar (rango (0, 3)), este re-escalamiento es necesario ya que el software de donde se obtuvieron las voces sintéticas posee una mayor resolución para los parámetros de ronquera y soplosidad, presentando un rango entre los valores (0, 6), y para el parámetro de astenia, los valores se encuentran en un rango de (-3, 0) debido a la contraposición de este parámetro con la tensión cuyos valores se encuentran en un rango de (0, 3).

Este re-escalamiento se realizó siguiendo un método lineal (*para futuras explicaciones se ha denominado E para aquel rango que varía entre (0, 6) y L para aquel rango que varía entre (0, 3)*), es decir, se le asignó el valor de $L = 0$ para $E = 0$, $L = 1$ para $E = (1, 2)$, $L = 2$ para $E = (3, 4)$ y $L = 3$ para $E = (5, 6)$; y con respecto a los valores presentes en el parámetro de astenia, únicamente se obtuvo el valor absoluto de cada valor antes de ser presentado en la interfaz gráfica.

3.1.3.4. Desarrollo de una interfaz gráfica que permita visualizar los resultados obtenidos por la herramienta.

Concluyendo con el desarrollo de la herramienta, se lleva a cabo la construcción de su interfaz gráfica, la cual cuenta con diversas funciones que permiten cargar la señal deseada, seleccionar el género del paciente, es decir, si se trata de un paciente masculino o femenino, y sobre todo visualizar los resultados obtenidos por la herramienta, la cual posee espacios delimitados para presentar la información de forma cómoda, contando con secciones específicas que muestra mediante una tabla y un gráfico de barras el diagnóstico generado para cada parámetro GRBAS, además, posee otro panel en el que podemos visualizar los valores obtenidos de las diferentes características vocales extraídas como Shimmer, Jitter, asimetría etc. y como función adicional se habilitó una ventana en la que se puede ver la representación gráfica de la señal, su espectrograma y su densidad espectral de potencia (Figura 52).

Una vez se obtiene la versión final de la herramienta computacional, se llevó a cabo un avance preliminar de la segunda etapa, la cual consiste en evaluar el desempeño de la herramienta implementando un banco de voces reales suministrado por la Universidad del Valle; Esta base de datos consta de 12 señales, con una duración de 3 segundos cada una, estas señales fueron suministradas con sus respectivo reporte de diagnóstico, el cual fue desarrollado bajo el criterio de 3 expertos.

3.1.3.5. Segmentación de las señales vocales del banco de voces de la Universidad del Valle.

En esta sección se realiza la segmentación de las señales vocales, con el fin de aumentar el número de datos destinados para realizar la evaluación de la herramienta.

Esta segmentación se llevó a cabo bajo el criterio de la duración mínima permitida para cada señal en la herramienta desarrollada, la cual es de 1,2 segundos, lo que permitió realizar la segmentación de las señales en 3 secciones con una duración de 1,2 segundos cada una.

3.1.3.6. Evaluación de la calidad de voz con base en la escala GRBAS para las señales segmentadas.

En este apartado se ingresa de manera individual las señales anteriormente segmentadas, y se registra en una tabla los resultados obtenidos, cabe destacar que, no se implementaron las

señales del paciente 7 y paciente 11, debido a que éstas señales entraron en conflicto con el algoritmo, generando un error lógico producido por la incapacidad del algoritmo en extraer la totalidad de las características necesarias para su correcto funcionamiento; las características que el algoritmo no logró determinar son: Armónico 1, Armónico 2, Armónico 3, Armónico 4, Armónico 5, Potencia del armónico 1, Potencia del armónico 2, Potencia del armónico 3, Potencia del armónico 4 y Potencia del armónico 5.

3.1.3.7. Comparación de resultados entre los diagnósticos reportados por expertos y los generados por la herramienta computacional.

Para realizar la comparación de los resultados, primero se realizó la ponderación y aproximación al número entero inferior más cercano para cada parámetro GRBAS entre los distintos valores obtenidos por la herramienta al evaluar los diferentes segmentos de una misma señal, posteriormente se obtienen los valores de exactitud y correlación con respecto al diagnóstico de los expertos (correlación de Spearman), y a continuación se determina la eficacia de la herramienta para realizar la evaluación de la calidad vocal con base en la escala GRBAS para las voces reales.

4. RESULTADOS

1. Validación de las características Jitter local, Jitter local absoluto y Shimmer dB en comparación con el software Praat.

En esta etapa se realiza la validación de las características acústicas con respecto al software Praat, el cual se ha utilizado en este trabajo como sistema estandarizado de medida.

Esta actividad se realizó implementando el 10% de la base de datos e implementando únicamente los primeros 400 milisegundos de cada una.

La validación llevada a cabo consiste en obtener la correlación lineal de Spearman, el error cuadrático medio (RMS) y error absoluto medio entre las variables obtenidas por medio de Matlab y el software Praat; los resultados obtenidos se muestran a continuación:

- Jitter local absoluto
 - Correlación de Spearman= 87,82 %
 - Error Absoluto Medio = $4,1331 \times 10^{-4}$ segundos
 - Error Cuadrático Medio = $4,759 \times 10^{-7}$ segundos²
- Jitter local
 - Correlación de Spearman= 87,38 %
 - Error Absoluto Medio = 3,3672 %

- Error Cuadrático Medio = 0,301016 %
- Shimmer dB
 - Correlación de Spearman= 88,18 %
 - Error Absoluto Medio = 0,3399 dB
 - Error Cuadrático Medio = $15,02 \times 10^{-2} \text{ dB}^2$

Con los resultados obtenidos se logra evidenciar un fuerte nivel de correlación de Spearman, ya que se obtuvo una $Rho > 85\%$, además de encontrar bajos niveles en los valores del error cuadrático medio y error absoluto medio, a pesar de las diferencias metodológicas implementadas en el algoritmo y el software Praat, ya que, por ejemplo Praat se desarrolló bajo la metodología de adaptación de onda (*Boersma, 2009*), y el algoritmo de este proyecto bajo la metodología de detección de picos en la señal.

A pesar de las diferencias en cada uno de estos algoritmos, los resultados obtenidos en esta sección logran avalar el funcionamiento de la herramienta con respecto a la extracción de las características acústicas: Jitter local, Jitter local absoluto y Shimmer dB.

Cabe destacar que la validación de las características vocales obtenidas por medio del algoritmo desarrollado, se llevó a cabo únicamente para aquellas características que no cuentan con funciones predeterminadas de Matlab encargadas de encontrar sus respectivos valores.

2. Comparación de los resultados obtenidos entre los diversos modelos de clasificación.

Los resultados obtenidos en esta etapa se pueden dividir en 2 subetapas las cuales son: Comparación entre los diversos modelos entrenados y Comparación entre los modelos dedicados según su género y no dedicados.

2.1. Comparación entre los diversos modelos entrenados.

Esta subetapa se puede describir como la comparación que se llevó a cabo entre los modelos entrenados para cada parámetro de la escala GRBAS, tanto los modelos dedicados como los modelos no dedicados; entre los modelos entrenados podemos encontrar: SVM Lineal, SVM Cuadrático, SVM Cúbico, SVM con Kernel Gaussiano Fino, SVM con Kernel Gaussiano Medio, SVM con Kernel Gaussiano Grueso, Naive Bayes con Kernel Gaussiano, Naive Bayes Kernel, Knn Fino, Knn Medio, Knn Grueso, Knn Coseno, Knn Cúbico y Knn Ponderado.

Para dar inicio a esta etapa se ha dividido los resultados por género y por parámetros GRBAS.

Masculino

- Ronquera

Modelos	Exactitud (%)
SVM Lineal	52,5 %
SVM Cuadrático	79,2 %
SVM Cúbico	83,7 %
SVM con Kernel Gaussiano Fino	74,1 %
SVM con Kernel Gaussiano Medio	52,8 %

SVM con Kernel Gaussiano Grueso	34,8 %
Naive Bayes con Kernel Gaussiano	NaN*
Naive Bayes Kernel	42 %
Knn Fino	72 %
Knn Medio	50,5 %
Knn Grueso	36,6 %
Knn Coseno	47,9 %
Knn Cúbico	45,5 %
Knn Ponderado	67,9 %

Tabla 1 Lista de resultados de los modelos dedicados para el parámetro de ronquera en el género masculino. *Modelo donde no se obtuvieron resultados válidos

Modelo Ronquera Masculino (SVM Cúbico)

0	246	1		2	7		
1	1	204	30	18	8	2	1
2		29	221	25	7	3	3
3	2	12	23	222	20		1
4	4	4	4	26	190		4
5		2	2	2	1	236	29
6		1	1	2	5	24	247
	0	1	2	3	4	5	6

True Class

Predicted Class

Figura 23 Matriz de confusión del modelo SVM Cúbico de ronquera para el género masculino.

Con base en la Tabla 1, se puede observar que el modelo que presenta mayor nivel de exactitud, es el modelo SVM Cúbico, ya que logró obtener un 83,7 % de exactitud; este

nivel puede ser confirmado por medio de su matriz de confusión la cual corresponde a la figura 23, su curva ROC y el valor de AUC para cada clase (Anexo 2 - 2.1).

Una vez ha sido seleccionado el modelo con el mejor desempeño, que para este caso (ronquera en género masculino) es el modelo SVM Cúbico, se continuará con la siguiente subetapa, la cual es la comparación con el modelo no dedicado implementando para ello el 20% de los datos de validación.

- Soplosidad

Modelos	Exactitud (%)
SVM Lineal	75,1 %
SVM Cuadrático	87,7 %
SVM Cúbico	90,5 %
SVM con Kernel Gaussiano Fino	88,2 %
SVM con Kernel Gaussiano Medio	83,1 %
SVM con Kernel Gaussiano Grueso	71,4 %
Naive Bayes con Kernel Gaussiano	58,3 %
Naive Bayes Kernel	78,4 %
Knn Fino	86,4 %
Knn Medio	80,2 %
Knn Grueso	64,6 %
Knn Coseno	79,2 %
Knn Cúbico	78,2 %
Knn Ponderado	87,4 %

Tabla 2. Lista de resultados de los modelos dedicados para el parámetro de soplosidad en el género masculino.

Modelo Solposidad Masculino (SVM Cúbico)

0	271	1					
1	1	235	4				8
2		2	257	13			
3			8	228	27	1	
4				22	231	14	5
5				2	19	208	27
6					4	19	265
	0	1	2	3	4	5	6

Predicted Class

***Figura 24** Matriz de confusión del modelo SVM Cúbico de soplosidad para el género masculino.*

Con base en la Tabla 2, se puede observar que el modelo que presenta mayor nivel de exactitud, es el modelo SVM Cúbico, ya que logró obtener un 90,5 % de exactitud; este nivel puede ser confirmado por medio de su matriz de confusión la cual corresponde a la figura 24, su curva ROC y el valor de AUC para cada clase (Anexo 2 - 2.2).

Una vez ha sido seleccionado el modelo con el mejor desempeño, que para este caso (soplosicada en género masculino) es el modelo SVM Cúbico, se continuará con la siguiente subetapa, la cual es la comparación con el modelo no dedicado implementando para ello el 20% de los datos de validación.

- Astenia y Tensión

Modelos	Exactitud (%)
SVM Lineal	92,6 %
SVM Cuadrático	95,8 %

SVM Cúbico	95,6 %
SVM con Kernel Gaussiano Fino	92,1 %
SVM con Kernel Gaussiano Medio	92,9 %
SVM con Kernel Gaussiano Grueso	85,7 %
Naive Bayes con Kernel Gaussiano	NaN*
Naive Bayes Kernel	66,3 %
Knn Fino	94 %
Knn Medio	89,7 %
Knn Grueso	80,7 %
Knn Coseno	88 %
Knn Cúbico	89 %
Knn Ponderado	93,2 %

Tabla 3. Lista de resultados de los modelos dedicados para los parámetro de astenia y tensión en el género masculino. *Modelo donde no se obtuvieron resultados válidos.

Modelo Astenia Tension Masculino (SVM Cuadrático)

True Class	-3	305	15				
	-2	12	295	13			
	-1		18	268	2		
	1			1	319	8	
	2				7	302	3
	3						304
		Predicted Class	-3	-2	-1	1	2

Figura 25 Matriz de confusión del modelo SVM Cuadrático de Astenia y Tensión para el género masculino.

Con base en la Tabla 3, se puede observar que el modelo que presenta mayor nivel de exactitud, es el modelo SVM Cuadrático, ya que logró obtener un 95,8 % de exactitud; este

nivel puede ser confirmado por medio de su matriz de confusión la cual corresponde a la figura 25, su curva ROC y el valor de AUC para cada clase (Anexo 2 - 2.3).

Una vez ha sido seleccionado el modelo con el mejor desempeño, que para este caso (astenia tensión en género masculino) es el modelo SVM Cuadrático, se continuará con la siguiente subetapa, la cual es la comparación con el modelo no dedicado implementando para ello el 20% de los datos de validación.

- Grado

Modelos	Exactitud (%)
SVM Lineal	74,4 %
SVM Cuadrático	100 %
SVM Cúbico	100 %
SVM con Kernel Gaussiano Fino	94,9 %
SVM con Kernel Gaussiano Medio	97 %
SVM con Kernel Gaussiano Grueso	79,9 %
Naive Bayes con Kernel Gaussiano	NaN*
Naive Bayes Kernel	98,3 %
Knn Fino	96,6 %
Knn Medio	97 %
Knn Grueso	59,7 %
Knn Coseno	88,5 %
Knn Cúbico	96,6 %
Knn Ponderado	97,9 %

Tabla 4. Lista de resultados de los modelos dedicados para el parámetro grado en el género masculino. *Modelo donde no se obtuvieron resultados válidos.

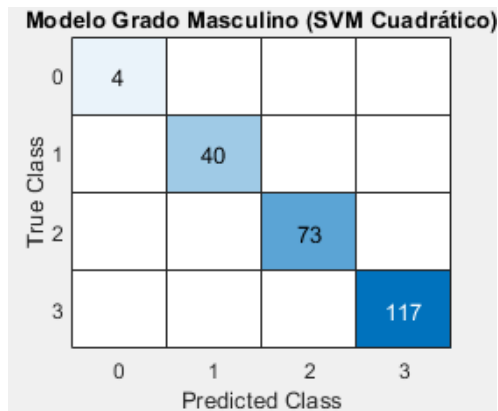


Figura 26 Matriz de confusión del modelo SVM Cuadrático de Grado para el género masculino.

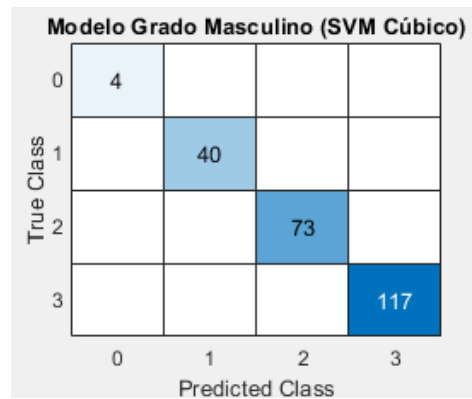


Figura 27 Matriz de confusión del modelo SVM Cúbico de Grado para el género masculino.

Con base en la Tabla 4, se puede observar que existen 2 modelos cuya precisión resultante es igual al 100%, logrando detallar en sus matrices de confusión (figura 26 y 27), curvas ROC y AUC (Anexo 2 - 2.4) que su clasificación es perfecta, por ende, cada uno es llevado a la siguiente fase en donde serán comparados entre sí y con el modelo no dedicado, con fin de seleccionar aquel modelo que se logre desempeñar de la mejor manera.

Femenino

- Soplosidad

Modelos	Exactitud (%)
SVM Lineal	92,7 %
SVM Cuadrático	97,2 %
SVM Cúbico	98,3 %
SVM con Kernel Gaussiano Fino	97,2 %
SVM con Kernel Gaussiano Medio	95,2 %
SVM con Kernel Gaussiano Grueso	83,8 %
Naive Bayes con Kernel Gaussiano	NaN*
Naive Bayes Kernel	77,7 %
Knn Fino	96,6 %
Knn Medio	88,4 %
Knn Grueso	56,7 %
Knn Coseno	87,8 %
Knn Cúbico	85,6 %
Knn Ponderado	94,2 %

Tabla 5. Lista de resultados de los modelos dedicados para el parámetro de soplosidad en el género femenino. *Modelo donde no se obtuvieron resultados válidos.

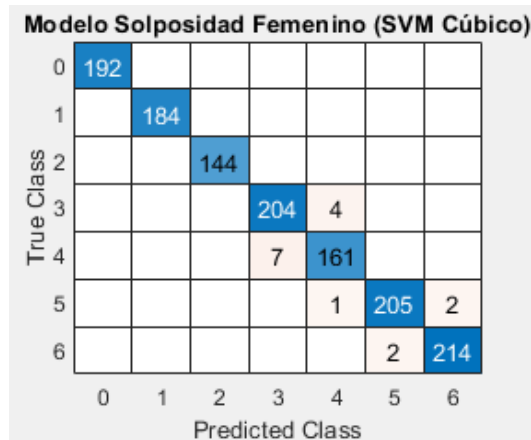


Figura 28 Matriz de confusión del modelo SVM Cúbico de soplosidad para el género femenino.

Con base en la Tabla 5, se puede observar que el modelo que presenta mayor nivel de exactitud, es el modelo SVM Cúbico, ya que logró obtener un 98,3 % de exactitud; este nivel puede ser confirmado por medio de su matriz de confusión la cual corresponde a la figura 28, su curva ROC y el valor de AUC para cada clase (Anexo 3 - 3.1).

Una vez ha sido seleccionado el modelo con el mejor desempeño, que para este caso (soplosidad en género femenino) es el modelo SVM Cúbico, se continuará con la siguiente subetapa, la cual es la comparación con el modelo no dedicado implementando para ello el 20% de los datos de validación.

- Astenia y Tensión

Modelos	Exactitud (%)
SVM Lineal	96,8 %
SVM Cuadrático	98,2 %
SVM Cúbico	98,6 %

SVM con Kernel Gaussiano Fino	96,7 %
SVM con Kernel Gaussiano Medio	96,3 %
SVM con Kernel Gaussiano Grueso	95 %
Naive Bayes con Kernel Gaussiano	NaN*
Naive Bayes Kernel	90,8 %
Knn Fino	95,9 %
Knn Medio	92,5 %
Knn Grueso	95,1 %
Knn Coseno	91,9 %
Knn Cúbico	92,1 %
Knn Ponderado	94,8 %

Tabla 6. Lista de resultados de los modelos dedicados para los parámetros de astenia y tensión en el género femenino. *Modelo donde no se obtuvieron resultados válidos.

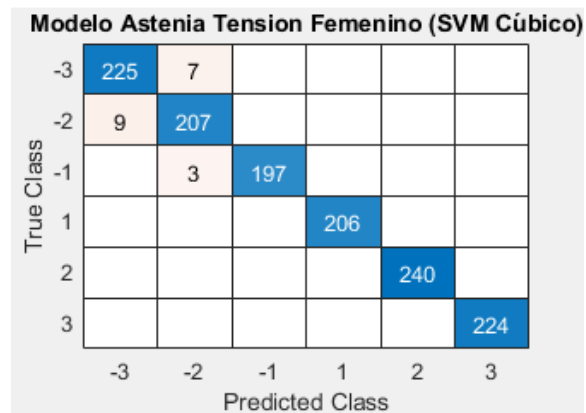


Figura 29 Matriz de confusión del modelo SVM Cúbico de Astenia y Tensión para el género femenino.

Con base en la Tabla 6, se puede observar que el modelo que presenta mayor nivel de exactitud, es el modelo SVM Cúbico, ya que logró obtener un 98,6 % de exactitud; este

nivel puede ser confirmado por medio de su matriz de confusión la cual corresponde a la figura 29, su curva ROC y el valor de AUC para cada clase (Anexo 3 - 3.2).

Una vez ha sido seleccionado el modelo con el mejor desempeño, que para este caso (astenia tensión en género femenino) es el modelo SVM Cúbico, se continuará con la siguiente subetapa, la cual es la comparación con el modelo no dedicado implementando para ello el 20% de los datos de validación.

Modelos no dedicados

- Ronquera

Modelos	Exactitud (%)
SVM Lineal	43 %
SVM Cuadrático	69,9 %
SVM Cúbico	79,5 %
SVM con Kernel Gaussiano Fino	68,4 %
SVM con Kernel Gaussiano Medio	46,3 %
SVM con Kernel Gaussiano Grueso	26,4 %
Naive Bayes con Kernel Gaussiano	21,4 %
Naive Bayes Kernel	36,7 %
Knn Fino	75,4 %
Knn Medio	53,5 %
Knn Grueso	30,3 %
Knn Coseno	52,7 %

Knn Cúbico	50,6 %
Knn Ponderado	70,8 %

Tabla 7. Lista de resultados de los modelos no dedicados para el parámetro de ronquera.

Modelo no dedicado Ronquera (SVM Cúbico)

0	462	5	26	8	17	1	1
1	12	417	55	21	10	12	9
2	12	64	376	38	20	4	6
3	17	24	58	415	34	6	6
4	13	14	12	39	415	6	5
5	2	9	4	6	4	218	29
6	1	3	7	3	8	24	234
	0	1	2	3	4	5	6

True Class

Predicted Class

Figura 30 Matriz de confusión del modelo no dedicado SVM Cúbico de ronquera.

Con base en la Tabla 7, se puede observar que el modelo que presenta mayor nivel de exactitud, es el modelo SVM Cúbico, ya que logró obtener un 79,5 % de exactitud; este nivel puede ser confirmado por medio de su matriz de confusión la cual corresponde a la figura 30, su curva ROC y el valor de AUC para cada clase (Anexo 4 - 4.1).

Una vez ha sido seleccionado el modelo con el mejor desempeño, que para este caso (ronquera) es el modelo SVM Cúbico, se continuará con el desarrollo de la siguiente subetapa.

- Soplosidad

Modelos	Exactitud (%)
SVM Lineal	74,3 %

SVM Cuadrático	88,3 %
SVM Cúbico	93,7 %
SVM con Kernel Gaussiano Fino	92,4 %
SVM con Kernel Gaussiano Medio	81,9 %
SVM con Kernel Gaussiano Grueso	68 %
Naive Bayes con Kernel Gaussiano	57,4 %
Naive Bayes Kernel	73,6 %
Knn Fino	91,5 %
Knn Medio	83,5 %
Knn Grueso	60,5 %
Knn Coseno	82,5 %
Knn Cúbico	80,8 %
Knn Ponderado	90,4 %

Tabla 8. Lista de resultados de los modelos no dedicados para el parámetro de soplosidad.

Modelo no dedicado Solposidad (SVM Cúbico)

0	463	1					
1	2	420	2				8
2		3	396	16	1		
3			8	435	28	1	
4	1			29	388	17	5
5				2	23	414	25
6					2	26	476
	0	1	2	3	4	5	6
	Predicted Class						

Figura 31 Matriz de confusión del modelo no dedicado SVM Cúbico de soplosidad.

Con base en la Tabla 8, se puede observar que el modelo que presenta mayor nivel de exactitud, es el modelo SVM Cúbico, ya que logró obtener un 93,7 % de exactitud; este

nivel puede ser confirmado por medio de su matriz de confusión la cual corresponde a la figura 31, su curva ROC y el valor de AUC para cada clase (Anexo 4 - 4.2).

Una vez ha sido seleccionado el modelo con el mejor desempeño, que para este caso (soplosidad) es el modelo SVM Cúbico, se continuará con el desarrollo de la siguiente subetapa.

- Astenia y Tensión

Modelos	Exactitud (%)
SVM Lineal	92,8 %
SVM Cuadrático	96,1 %
SVM Cúbico	96,5 %
SVM con Kernel Gaussiano Fino	94,7 %
SVM con Kernel Gaussiano Medio	93,8 %
SVM con Kernel Gaussiano Grueso	86,3 %
Naive Bayes con Kernel Gaussiano	NaN*
Naive Bayes Kernel	97,8 %
Knn Fino	94,3 %
Knn Medio	90,1 %
Knn Grueso	79,1 %
Knn Coseno	89,5 %
Knn Cúbico	89,9 %
Knn Ponderado	93,4 %

Tabla 9. Lista de resultados de los modelos no dedicados para los parámetros de astenia y tensión. *Modelo donde no se obtuvieron resultados válidos.

Modelo no dedicado Astenia Tension (SVM Cuadrático)

-3	527	22	3			
-2	20	478	18			
-1		21	467			
1			1	525	10	
2				14	535	3
3						528
	-3	-2	-1	1	2	3
True Class						
	Predicted Class					

Figura 32 Matriz de confusión del modelo no dedicado SVM Cúbico de Astenia y Tensión.

Con base en la Tabla 8, se puede observar que el modelo que presenta mayor nivel de exactitud, es el modelo SVM Cúbico, ya que logró obtener un 96,5 % de exactitud; este nivel puede ser confirmado por medio de su matriz de confusión la cual corresponde a la figura 32, su curva ROC y el valor de AUC para cada clase (Anexo 4 - 4.3).

Una vez ha sido seleccionado el modelo con el mejor desempeño, que para este caso (astenia tensión) es el modelo SVM Cúbico, se continuará con el desarrollo de la siguiente subetapa.

- Grado

Modelos	Exactitud (%)
SVM Lineal	74,4 %
SVM Cuadrático	94,2 %
SVM Cúbico	92,2 %
SVM con Kernel Gaussiano Fino	91,5 %
SVM con Kernel Gaussiano Medio	95 %

SVM con Kernel Gaussiano Grueso	79,7 %
Naive Bayes con Kernel Gaussiano	NaN*
Naive Bayes Kernel	94,2 %
Knn Fino	90,7 %
Knn Medio	93,2 %
Knn Grueso	70,4 %
Knn Coseno	85,7 %
Knn Cúbico	93,2 %
Knn Ponderado	90,7 %

Tabla 10. Lista de resultados de los modelos no dedicados para el parámetro grado.
*Modelo donde no se obtuvieron resultados válidos.

Modelo no dedicado Grado (SVM Gaussiano Medio)

0		4		
1		75	2	
2			148	13
3			1	156
	0	1	2	3

True Class

Predicted Class

Figura 33 Matriz de confusión del modelo no dedicado SVM con Kernel Gaussiano Medio del parámetro Grado.

Con base en la Tabla 10, se puede observar que el modelo que presenta mayor nivel de exactitud, es el modelo SVM con Kernel Gaussiano Medio, ya que logró obtener un 95 % de exactitud; este nivel puede ser confirmado por medio de su matriz de confusión la cual

corresponde a la figura 33, su curva ROC y el valor de AUC para cada clase (Anexo 4 - 4.4).

Una vez ha sido seleccionado el modelo con el mejor desempeño, que para este caso (grado) es el modelo SVM con Kernel Gaussiano Medio, se continuará con el desarrollo de la siguiente subetapa.

2.2. Comparación entre los modelos dedicados según su género y no dedicados

Esta subetapa se puede describir como la comparación que se llevó a cabo entre los modelos dedicados según su género y los no dedicados, los cuales fueron seleccionados en la subetapa anterior (2.1.), esto con el fin de determinar los modelos con los mejores niveles de exactitud para la conformación de la herramienta; Este proceso se llevó a cabo implementando el 20% de los datos restantes dedicados para la validación, los cuales han sido discriminados según su género, es decir, se implementó el 20% de los datos del género masculino y el 20% de los datos del género femenino.

Masculino

- Ronquera

Modelos	Exactitud (%)
Modelo dedicado (SVM Cúbico)	71,18 %
Modelo no dedicado (SVM Cúbico)	61,01 %

Tabla 11. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro de ronquera en el género masculino.

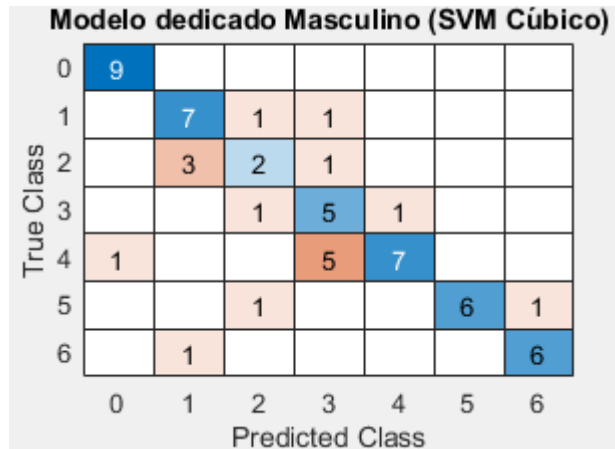


Figura 34 Matriz de confusión del modelo dedicado SVM Cúbico para Ronquera.

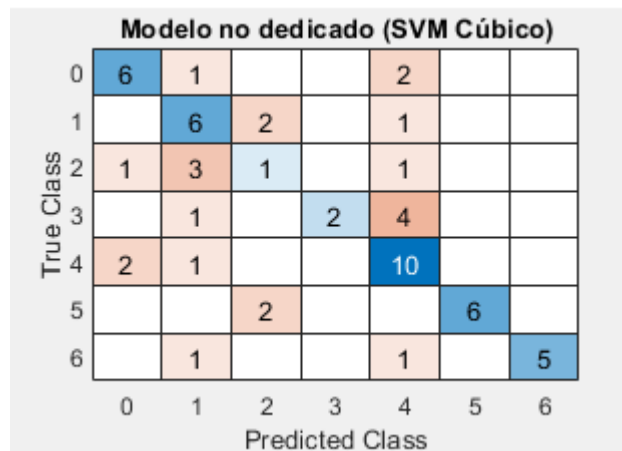


Figura 35 Matriz de confusión del modelo no dedicado SVM Cúbico para Ronquera.

Con base en la Tabla 11 y sus representaciones gráficas de su respectiva matriz de confusión (figura 34, 35), se puede apreciar que el modelo dedicado, presenta un mayor porcentaje de exactitud al realizar la clasificación de los datos (71,18%), obteniendo una superioridad del 10,17%, ya que este modelo posee un de exactitud, por ende, ha sido

seleccionado para conformar la herramienta en su versión final; Cabe destacar que este modelo a pesar de tener una exactitud relativamente alta, si se compara es resultado con respecto a los demás clasificadores del género masculino, este es el que presenta los menores niveles de exactitud.

- Soplosidad

Modelos	Exactitud (%)
Modelo dedicado (SVM Cúbico)	84,74 %
Modelo no dedicado (SVM Cúbico)	84,74 %

Tabla 12. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro de soplosidad en el género masculino.

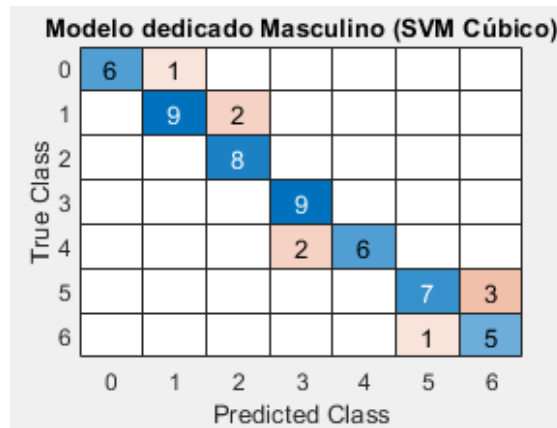


Figura 36 Matriz de confusión del modelo dedicado SVM Cúbico para Soplosidad.

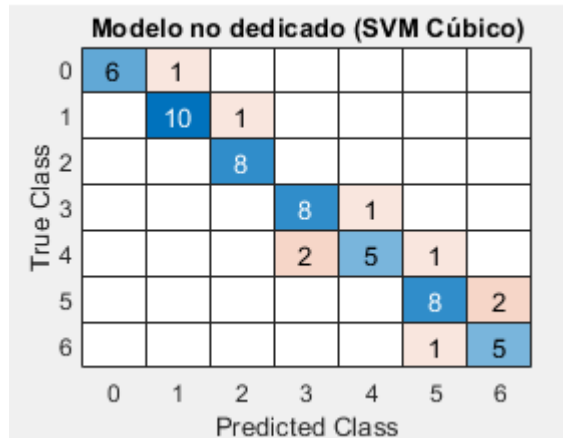


Figura 37 Matriz de confusión del modelo no dedicado SVM Cúbico para Soplosidad.

Con base en la Tabla 12, se puede apreciar los 2 modelos presentan el mismo valor de exactitud, así que se realizó el análisis de sus respectivas matrices de confusión (figura 36, 37), en donde se encontró que en el modelo no dedicado, el error de clasificación presenta una mayor dispersión de datos para las clases, además, también se tomó cuenta de manera estadísticamente el tipo modelo cuyo desempeño ha sido superior para los anteriores parámetros, de tal manera que se seleccionó el modelo de clasificación dedicado para para la conformación de la herramienta final.

- Astenia Tensión

Modelos	Exactitud (%)
Modelo dedicado (SVM Cuadrático)	86,44 %
Modelo no dedicado (SVM Cúbico)	93,22 %

Tabla 13. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para los parámetros de astenia y tensión en el género masculino.

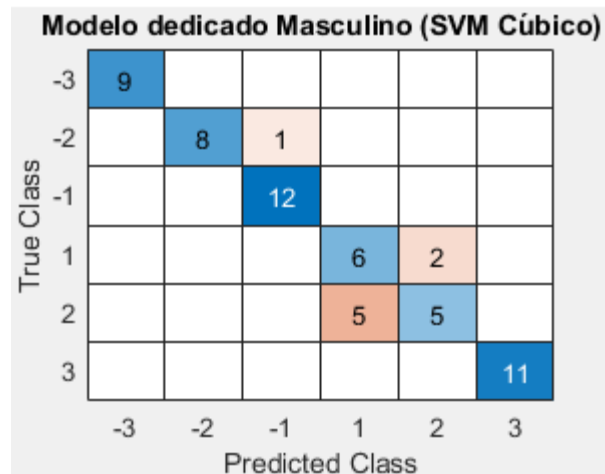


Figura 38 Matriz de confusión del modelo dedicado SVM Cuadrático para Astenia y Tensión.

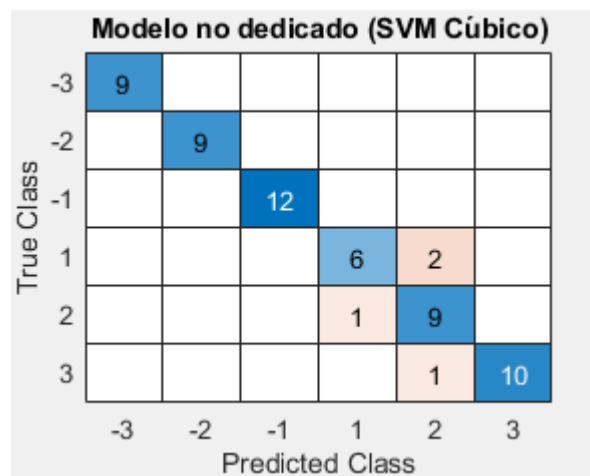


Figura 39 Matriz de confusión del modelo no dedicado SVM Cúbico para Astenia y Tensión.

Con base en la tabla 13 y sus representaciones gráficas de su respectiva matriz de confusión (figura 38, 39), se puede apreciar que el modelo no dedicado, presenta un mayor porcentaje de exactitud al realizar la clasificación de los datos (93,22%), obteniendo una superioridad

del 6,78 %, por ende, ha sido seleccionado para conformar la herramienta en su versión final.

- Grado

Modelos	Exactitud (%)
Modelo dedicado (SVM Cúbico)	96,61 %
Modelo dedicado (SVM Cuadrático)	96,61 %
Modelo no dedicado (SVM Cúbico)	94,91 %

Tabla 14. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro grado en el género masculino.

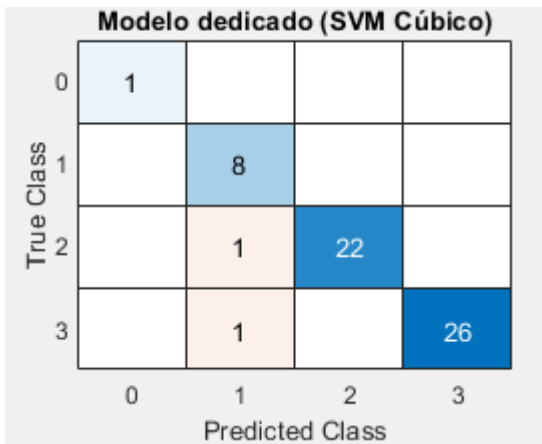


Figura 40 Matriz de confusión del modelo dedicado SVM Cúbico para Grado.

Modelo dedicado (SVM Cuadrático)

0	1			
1		8		
2		1	22	
3		1		26
	0	1	2	3

Predicted Class

Figura 41 Matriz de confusión del modelo dedicado SVM Cuadrático para Grado.

Modelo no dedicado (SVM Cúbico)

0		1		
1		8		
2		1	22	
3		1		26
	0	1	2	3

Predicted Class

Figura 42 Matriz de confusión del modelo no dedicado SVM Cúbico para Grado.

Con base en la tabla 14 y la representación gráficas de sus respectivas matrices de confusión, se puede observar que este caso presenta la particularidad de tener 2 de los modelos dedicados que presentan un igual porcentaje de exactitud y con una superioridad con respecto al modelo no dedicado, y con el fin de determinar un único modelo para llevar a cabo el desarrollo de la herramienta final, se realizó el análisis de sus matrices de confusión (figura 41, 42), pero al no poseer diferencia alguna en su tipo de clasificación, el

proceso de selección se llevó a cabo de manera estadística, tras observar con anterioridad que el modelo de SVM cúbico ha tenido un mejor desempeño en la clasificación de los demás parámetros, así que para la clasificación del grado en el género masculino será realizado por el modelo SVM Cúbico.

Femenino

- Ronquera

Modelos	Exactitud (%)
Modelo dedicado_Masculino (SVM Cúbico)	2,38 %
Modelo no dedicado (SVM Cúbico)	40,47 %

Tabla 15. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro de ronquera en el género femenino.

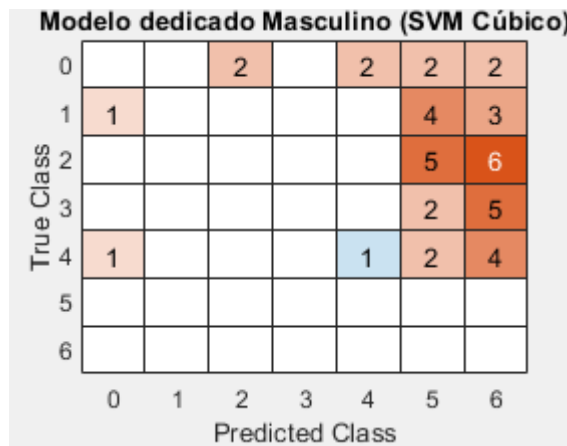


Figura 43 Matriz de confusión del modelo masculino dedicado SVM Cúbico para Ronquera.

1	4	1	2	1		
2	1	3	4			
3	1	4	4	2		
4		1		4	2	
5				5	2	1
6						
	1	2	3	4	5	6

Figura 44 Matriz de confusión del modelo no dedicado SVM Cúbico para Ronquera.

Este caso presenta la particularidad de no poseer un modelo dedicado para el parámetro de ronquera en el género femenino, debido a la existencia de un error anteriormente mencionado, así que la comparación de los modelos se llevó a cabo entre el modelo dedicado para el género masculino y el modelo no dedicado, esto se puede detallar con base en la tabla 15, además, y junto las figuras 43 y 44, se logra evidenciar que el modelo dedicado para el género masculino presenta un desempeño muy deficiente, así que para la clasificación de la ronquera en el género femenino se ha implementado el modelo no dedicado, aunque, cabe resaltar que a pesar de poseer el mayor nivel de exactitud, su desempeño se ha considerado como moderadamente bajo, obteniendo un nivel de exactitud menor al 50% (40,47%).

- Soplosidad

Modelos	Exactitud (%)
Modelo dedicado (SVM Cúbico)	90,47 %
Modelo no dedicado (SVM Cúbico)	88,09 %

Tabla 16. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro de soplosidad en el género femenino.

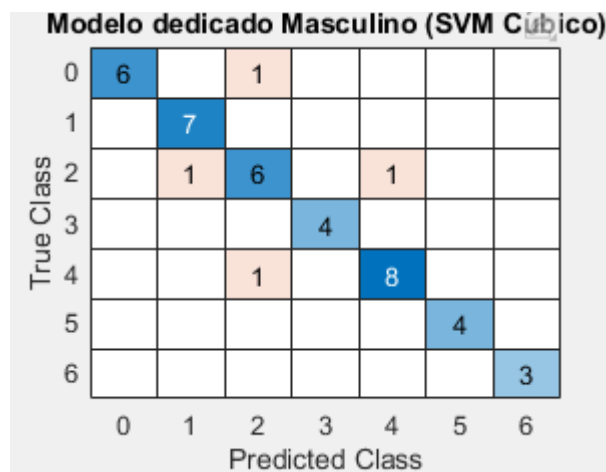


Figura 45 Matriz de confusión del modelo dedicado SVM Cúbico para Soplosidad.

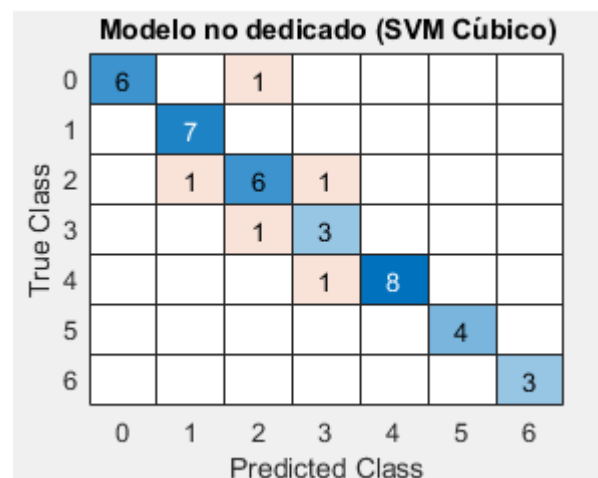


Figura 46 Matriz de confusión del modelo no dedicado SVM Cúbico para Soplosidad.

Con base en la Tabla 16 y sus representaciones gráficas de su respectiva matriz de confusión (figura 45, 46), se puede apreciar que el modelo dedicado, presenta un mayor porcentaje de exactitud al realizar la clasificación de los datos (71,18%), obteniendo una superioridad del 2,38%, por ende, ha sido seleccionado para conformar la herramienta en su versión final.

- Astenia Tensión

Modelos	Exactitud (%)
Modelo dedicado (SVM Cúbico)	97,61 %
Modelo no dedicado (SVM Cúbico)	97,61 %

Tabla 17. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para los parámetros de astenia y tensión en el género femenino.

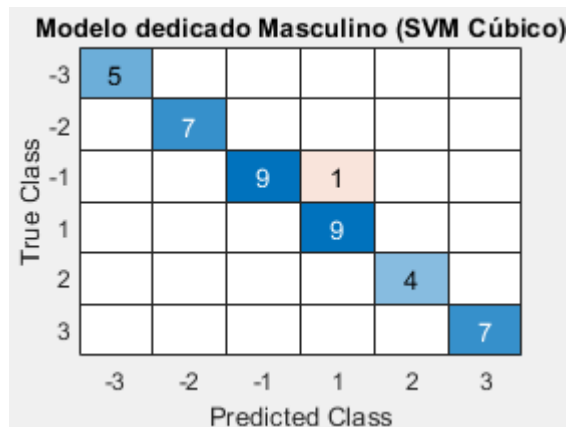


Figura 47 Matriz de confusión del modelo dedicado SVM Cúbico para Astenia y Tensión.

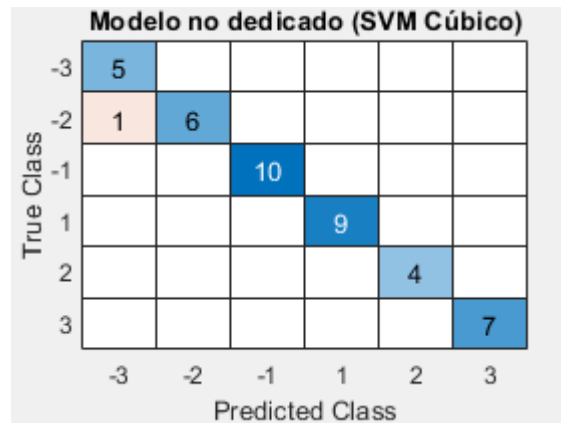


Figura 48 Matriz de confusión del modelo no dedicado SVM Cúbico para Astenia y Tensión.

Con base en la Tabla 17, se puede apreciar los 2 modelos presentan el mismo valor de exactitud, así que se realizó el análisis de sus respectivas matrices de confusión (figura 47, 48), pero al no poseer una variación significativa, el proceso de selección se llevó a cabo de manera estadística, tras observar con anterioridad que el modelo dedicado ha tenido un mejor desempeño en la clasificación de los demás parámetros, así que para la clasificación de la Astenia Tensión en el género femenino será realizado por el modelo dedicado.

- Grado

Modelos	Exactitud (%)
Modelo dedicado_Masculino (SVM Cúbico)	92,85 %
Modelo dedicado_Masculino (SVM Cuadrático)	92,85 %
Modelo no dedicado (SVM con Kernel Gaussiano Medio)	95,23 %

Tabla 18. Lista de resultados de la comparación entre el modelo dedicado y no dedicado para el parámetro grado en el género femenino.

Modelo dedicado (SVM Cúbico)

0				
1	1	15		
2		1	18	1
3				6
	0	1	2	3

True Class

Predicted Class

Figura 49 Matriz de confusión del modelo masculino dedicado SVM Cúbico para Grado.

Modelo dedicado (SVM Cuadrático)

0				
1	1	15		
2		1	18	1
3				6
	0	1	2	3

True Class

Predicted Class

Figura 50 Matriz de confusión del modelo masculino dedicado SVM Cuadrático para Grado.

Modelo no dedicado (SVM Cúbico)

1	16		
2	1	18	1
3			6
	1	2	3

True Class

Predicted Class

Figura 51 Matriz de confusión del modelo no dedicado SVM Cúbico para Grado.

Este caso, al igual que para el parámetro de ronquera en el género femenino, presenta la misma particularidad de no poseer un modelo dedicado para el parámetro de grado en este género, debido al error ya anteriormente mencionado, así que la comparación de los modelos se llevó a cabo entre los modelos dedicados para el género masculino y el modelo no dedicado, esto se puede detallar con base en la tabla 18, además, y junto las figuras 49 , 50 y 51, se logra evidenciar que el modelo dedicado para el género masculino presenta un desempeño bastante prometedor, pero, el modelo no dedicado presenta una superioridad del 2,38% en la exactitud de clasificación, ya que obtuvo un porcentaje de 95,23%, por ende, para llevar a cabo la clasificación del grado en el género femenino se ha implementado el modelo no dedicado.

Para concluir la etapa 2 de la sección de resultados, se presenta a continuación 2 tablas en donde se muestran los clasificadores que harán parte de la versión final de la herramienta junto con su nivel porcentual de exactitud, esto además, se realizará por cada género:

Modelos seleccionados (Género Masculino)	Exactitud
Modelo SVM cúbico dedicado para el parámetro Ronquera	71,18 %
Modelo SVM cúbico dedicado para el parámetro Soplosidad	84,74 %
Modelo SVM cúbico no dedicado para el parámetro Astenia Tensión	93,22 %
Modelo SVM cúbico dedicado para el parámetro Grado	96,61 %

Tabla 19 Lista de los modelos para el género masculino que formarán parte de la herramienta final.

Modelos seleccionados (Género Femenino)	Exactitud
Modelo SVM cúbico no dedicado para el parámetro Ronquera	40,47 %
Modelo SVM cúbico dedicado para el parámetro Soplosidad	90,47 %
Modelo SVM cúbico dedicado para el parámetro Astenia Tensión	97,61 %
Modelo SVM con kernel Gaussiano medio no dedicado para el parámetro Grado	95,23 %

Tabla 20 Lista de los modelos para el género femenino que formarán parte de la herramienta final.

3. Desarrollo de una interfaz gráfica que permita visualizar los resultados obtenidos por la herramienta

En la figura 52 es posible evidenciar las diferentes pantallas desarrolladas para la interfaz gráfica, básicamente son consta 3 ventanas en las que se pueden visualizar: a) la opción de cargar un nuevo audio, la gráfica de la señal y dos botones capaces de abrir nuevas ventanas en donde se observe el espectrograma y la densidad espectral de potencia, b) En la segunda ventana se pueden observar las características vocales que se extrajeron de la señal ingresada, tales como: Jitter local, Shimmer dB, frecuencia fundamental, entre otras, c) Por último, la tercera ventana nos permite visualizar el diagnóstico dictaminado por el algoritmo tras realizar la evaluación de la voz, es decir el valor de cada parámetro GRBAS tanto de forma numérica presente en una tabla, como en forma gráfica, siendo ésta representada en una gráfico de barras.

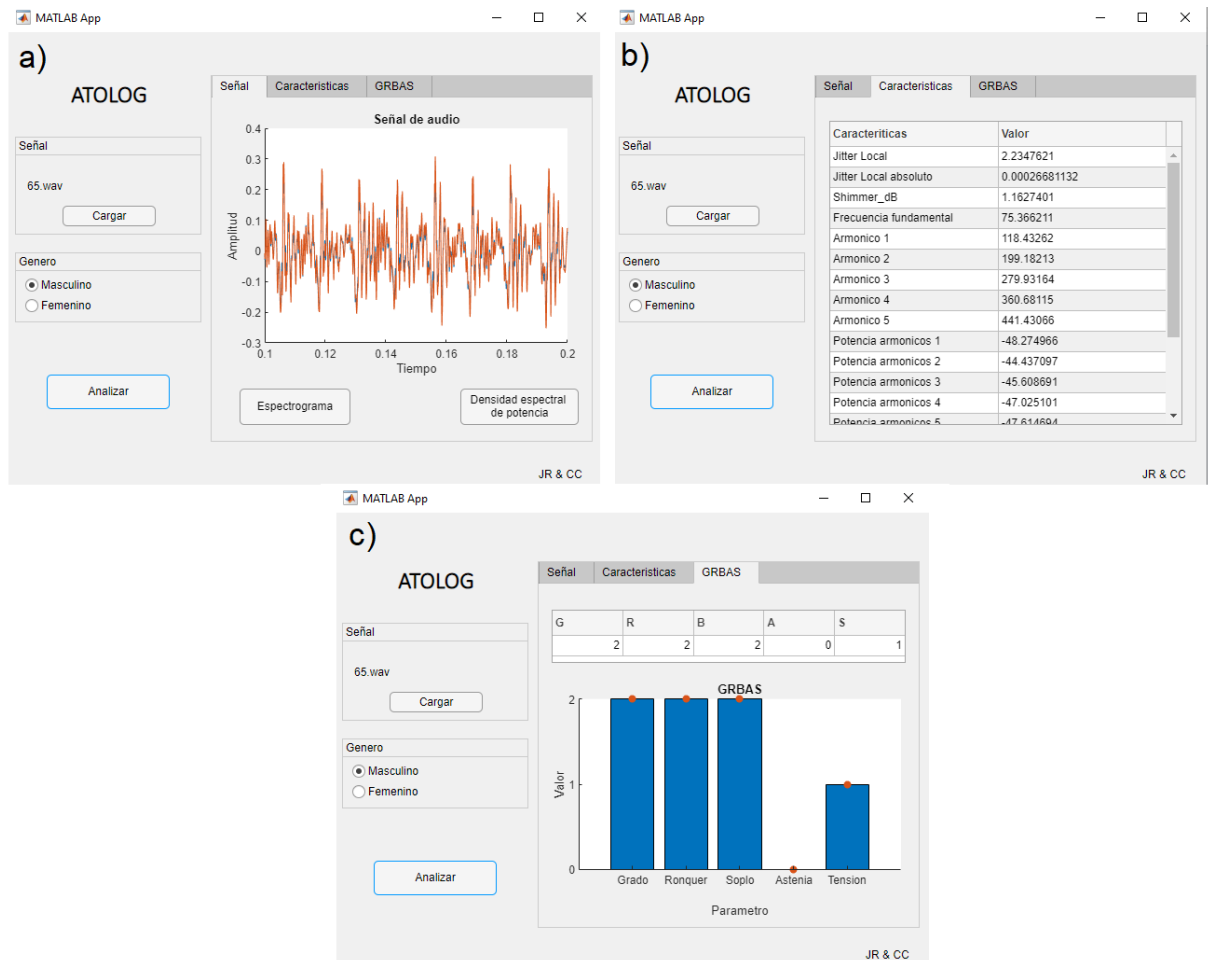


Figura 52 Interfaz gráfica del programa.

4. Evaluación preliminar de la herramienta al trabajar con voces reales

4.1. Análisis de los niveles de exactitud y correlación para los resultados obtenidos en la evaluación preliminar

Los resultados obtenidos en esta evaluación preliminar al implementar voces reales, demostraron que la herramienta posee algunas deficiencias al realizar la clasificación en

comparación con el análisis y diagnóstico realizado por personal experto, ya que se obtuvieron resultados con un bajo nivel de exactitud y correlación; estos resultados se presentan en la tabla 21.

<i>Parámetros</i>	<i>Exactitud</i>	<i>Correlación de Spearman</i>
Grado	20%	0,49
Ronquera	30%	0,32
Soplosidad	10%	0,25
Astenia	50%	0,36
Tensión	10%	-0,18

Tabla 21 Resultados obtenidos en la evaluación de la herramienta con voces reales.

Los bajos valores obtenidos en esta evaluación, pueden explicarse debido al método empleado para el entrenamiento del algoritmo, el cual se llevó a cabo en su totalidad con voces sintéticas provenientes del software Evaper, lo que significa que la herramienta posee un mejor desempeño en el análisis de voces sintéticas, lo cual queda demostrado con sus porcentajes de exactitud donde se obtuvo en su mayoría valores por encima del 80%.

5. CONCLUSIONES

- Con base en los resultados obtenidos en la validación preliminar con voces reales, se concluye que la herramienta desarrollada obtuvo un gran nivel de exactitud en la evaluación de la calidad vocal únicamente en voces sintéticas (Exactitud > 71%), debido a que los datos con los que se llevó a cabo el entrenamiento de la herramienta poseen esta misma característica; cabe destacar que, el desempeño al realizar la evaluación con voces reales, decae significativamente (Exactitud < 50%), siendo esto consecuencia de la diferencia técnica existente entre voces real y sintética, además, por la falta de voces reales en la fase de entrenamiento de la herramienta, ésto a causa de las limitaciones de tiempo y recursos generados por la emergencia sanitaria de Covid - 19.
- Esta primera aproximación de la herramienta desarrollada, a pesar de presentar resultados contundentes únicamente con señales sintéticas de voz, ha permitido cimentar el camino para un futuro proyecto en donde se realice las correcciones pertinentes para que esta herramienta aumente de manera significativa los niveles de exactitud en la clasificación de las voces reales.

- Esta herramienta representa un acercamiento a la sistematización de la evaluación de la calidad vocal, además de poseer un gran potencial en el área de la logopedia, ya que cierra la brecha de subjetividad existente en un análisis de voz convencional, debido a que todos los análisis llevados a cabo por la herramienta se realizan de manera cuantitativa, siguiendo únicamente argumentos lógicos y matemáticos, además de realizarlos en tiempos considerablemente cortos.
- Se hace necesario en un futuro cercano implementar la etapa 2 que permitirá evidenciar con mayor soporte la calidad algorítmica de la herramienta implementada; esta actividad podrá ser llevada a cabo por los autores en etapas posteriores a esta etapa de sustentación de trabajo de investigación en cuestión.

REFERENCIAS BIBLIOGRÁFICAS

- Aguilar, H., & Vélez Julia. (2016). Relación entre las pruebas aerodinámicas de la fonación con la escala GRBAS para alertar riesgo de disodias en cantantes en formación de la Universidad del Valle, Santiago de Cali - 2016. Universidad del Valle.
- Barrios, J. (2019, July 26). La matriz de confusión y sus métricas. Health Big Data.
- Behlau, M. (2014). Voxmetria (3.3.). CTS informática.
- Belletti, A. (2018). Valor Cuadrático Medio o RMS. SCRIBD.
- Boersma, P. (2009, October 10). Should Jitter Be Measured by Peak Picking or by Waveform Matching? Folia Phoniatica et Logopaedica.
- Boersma, P., & Weenink, D. (2011). Praat (6.1.52).
- Camacho, C. (2007). COEFICIENTE DE CORRELACIÓN LINEAL DE PEARSON. In CORRELACIÓN LINEAL DE PEARSON.
- Centre for speech technology. (2019). WaveSurfer (1.8.8). speech.kth.se.
- Cerda, J., & Cifuentes, L. (2012). Uso de curvas ROC en investigación clínica. Aspectos teórico-prácticos. Revista Chilena de Infectología.
- Clavbo, B. (2006). Method and device for speech analysis (Patent No. US7092874).
- Delgado, J., Leon, N., Jimenez, A., & Izquierdo, L. (2017). Análisis acústico de la voz: medidas temporales, espectrales y cepstrales en la voz normal con el Praat en una muestra de hablantes de español. Revista de Investigación En Logopedia, 108–127.
- DELSOL Software. (2021, October 26). Teorema de bayes. Software DELSOL.
- Dodero, A., Hortas, A., & Wilder, F. (2005). METODOLOGÍA DE ESTUDIO DE ALTERACIONES DE LA VOZ Y BAREMOS DE INCAPACIDAD VOCAL. In Cuadernos de Medicina Forense (1st ed., Vol. 4, pp. 41–51). Cuadernos de medicina forense.
- Droguett, Y. (2017). Aplicaciones clínicas del análisis acústico de la voz. Otorrinolaringología y Cirugía de Cabeza y Cuello, 77(4), 474–483.
- García, M., Rosset, A., & Destefanis, E. (2021). Red neuronal multiescala para clasificación de la calidad vocal. XXIII Workshop de Investigadores En Ciencias de La Computación, 144–148.

- Gomez, P., Rodellar, V., Nieto, V., Alvarez, A., & Martinez Rafael. (2011). Método y sistema para la estimación de parámetros fisiológicos de la fonación (Patent No. WO2013001109A1).
- González, A. (2021). ¿Qué es Machine Learning? Cleverdata.
- Gordillo, A. (2018). Hitos de la evaluación perceptual auditiva de la voz. *ARETE*, 18(2), 65–74.
- Gurlekian, J., & Torres, H. (2016). Informe Técnico. EVAPER.
- Gonzalez, L. (2019, August 23). Kernel. AprendeIA.
- Hernández, J., Franklenn, J., Penaloza, M., Camargo, E., Garzón, M., Chacon, O., Suarez, K., Alvarez, D., & Pirela, V. (2018). Sobre el uso adecuado del coeficiente de correlación de Pearson: verificación de supuestos mediante un ejemplo aplicado a las ciencias de la salud. In *Archivos Venezolanos de Farmacología y Terapéutica* (Vol. 38, pp. 552–561).
- LevanonLan, Y., & Shifrin, L. (2005). Method and system for diagnosing pathological phenomena using a voice signal (Patent No. US7398213B1).
- Martinez, J. (2019, May 28). Máquinas de Vectores de Soporte (SVM). IArtificial.Net.
- Matlab. (2021). Help Center. MathWorks.
- MERKLE. (2020, September 1). El algoritmo K-NN y su importancia en el modelado de datos. MERKLE.
- Minitab 18. (2019). Interpretar los resultados clave para Correlación. Soporte de Minitab® 18.
- Montes de Oca, D. (2014). Manual básico de PRAAT. Pontificia Universidad Católica de Chile.
- Mora, E., & Cobeta, I. (2013). Voz profesional. In I. Cobeta, F. Núñez, & S. Fernández (Eds.), *Patología de la voz* (pp. 516–545). SEORL PCF.
- Muller, D. (2017, April 5). How to update your beliefs systematically - Bayes Theorem. Youtube. <https://www.youtube.com/watch?v=R13BD8qKeTg&t>
- Nonó, J., Plaja, A., Pagès, E., Corbella, C., & Santamaria, M. (2014). El uso profesional de la voz.
- ORACLE. (21 C.E., October 26). ¿Qué es la inteligencia artificial—IA? ORACLE.

Preparadores de Oposiciones para la Enseñanza. (2005). La voz humana y su fisiología. <https://www.preparadores.eu/secundaria/Musica/Musica-Tema.pdf>

Prieto, G. (2017). ¿Qué es machine learning? Platzi.

Rodrigo, J. (2016, November). Validación de modelos predictivos: Cross-validation, OneLeaveOut, Bootstrapping. Ciencia de Datos.

Roman Victor. (2019, April 25). Algoritmos Naive Bayes: Fundamentos e Implementación. Ciencia & Datos.

Rosario, E. (2017, February 27). Señales de Energía y Potencia. Señales y Sistemas.

Shintaro, F., Tsuyoshi, K., Yusuke, O., Kazuhiko, S., Masato, I., Koichi, O., & Ryusuke Hori. (2020). Classification of voice disorders using a one-dimensional convolutional Neural network. Journal of Voice.

Sotres, Y., Fernández, R., & Portuondo, M. (2015). USING THE GRABS SCALE IN THE PERCEPTUAL EVALUATION OF VOICE DYSPHONIC PATIENTS. Revista Cubana de Tecnología de La Salud, 6(4), 78–87.

Teixeira, J., & Gonçalves, A. (2014). Accuracy of Jitter and Shimmer Measurements. Procedia Technology.

Vaerenbergh, S., & Santamaría, I. (2018). Métodos kernel para clasificación.

Wevosys. (2009). LingWAVES. Wevosys.

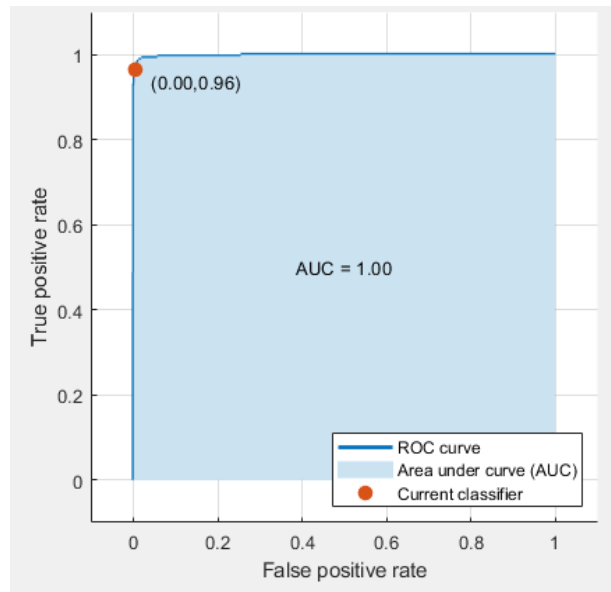
Zheng, X., Chaitanya, G., Farideh, J., Barry, C., & Jarrod, H. (2018). Machine Learning Applied to GRBAS Voice Quality Assessment. Astes, 3(6), 329–338.

B. Anexo 2: Curvas ROC de los modelos para el género masculinos

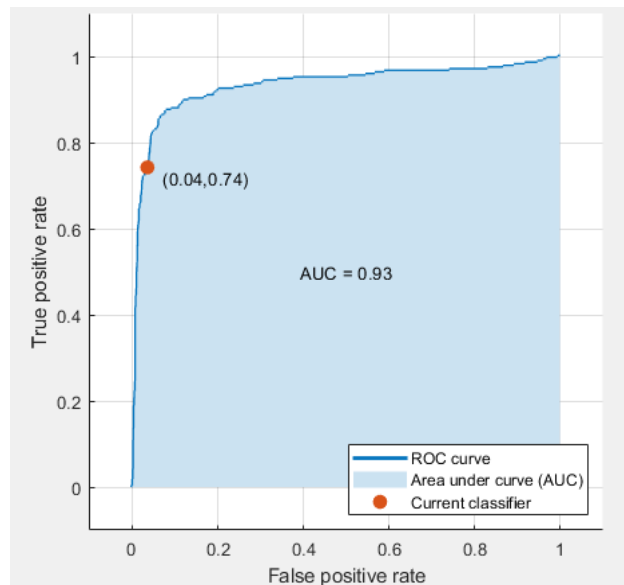
En esta sección, se muestra la curva Roc de cada clase para cada parámetro GRBAS, de los modelos seleccionados para el género masculino, donde muestran la sensibilidad frente a la especificidad de la clasificación binaria para cada clase, y la capacidad discriminatoria diagnóstica representada por el valor de AUC.

Anexo 2.1. Modelo SVM Cúbico para el parámetro de ronquera.

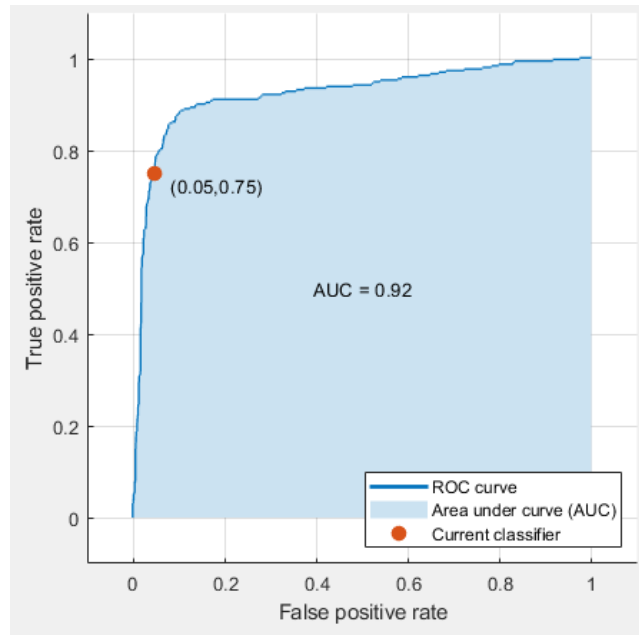
- Clase 0



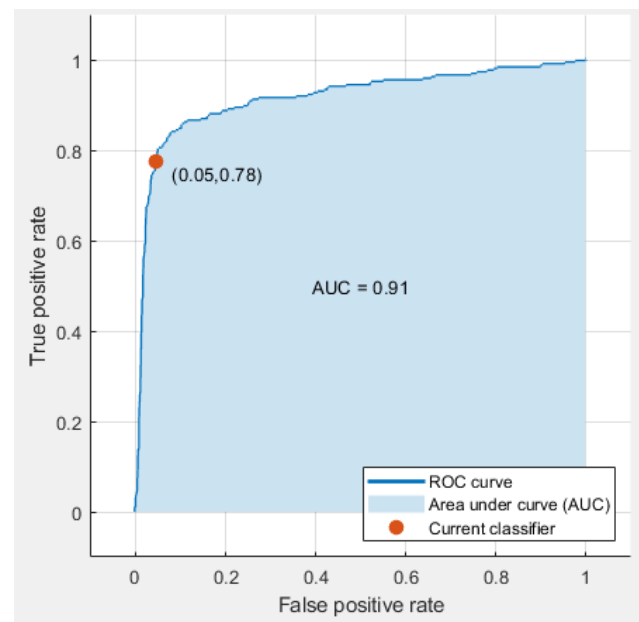
- Clase 1



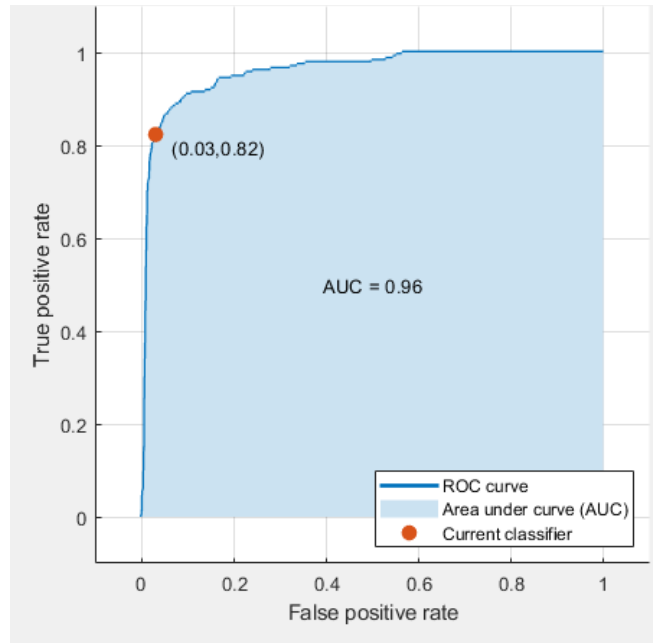
- Clase 2



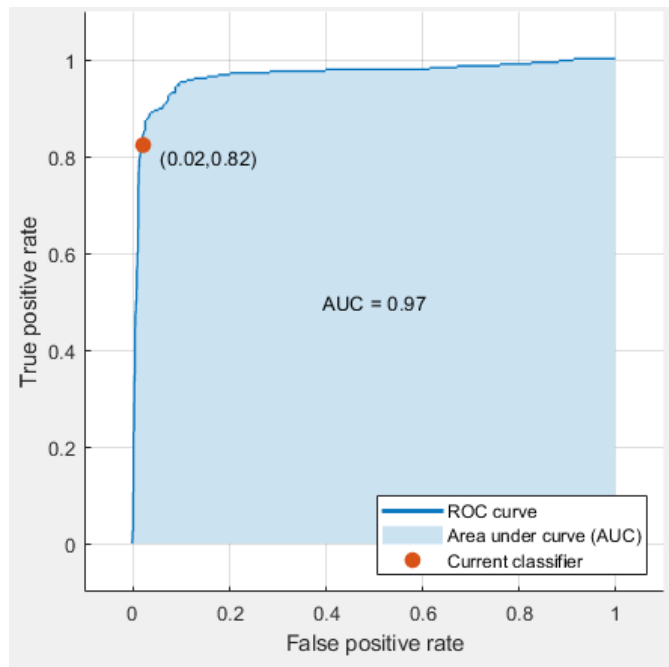
- Clase 3



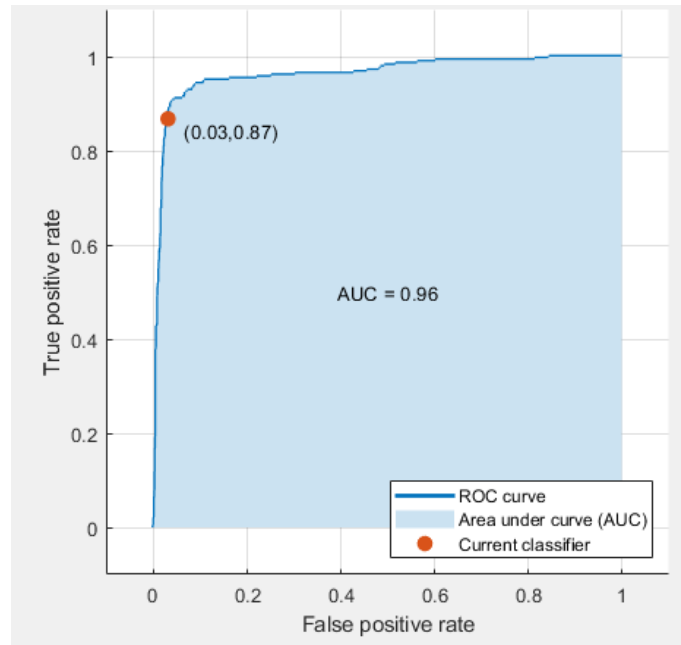
- Clase 4



- Clase 5

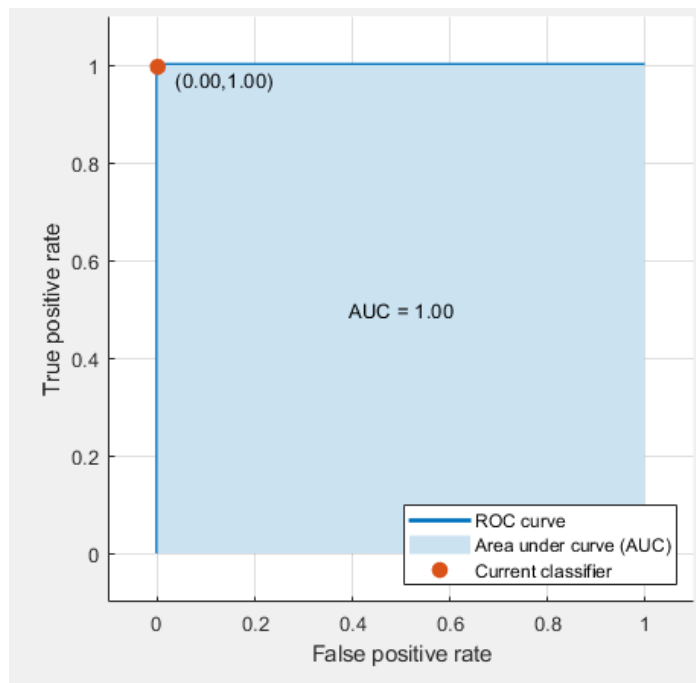


- Clase 6

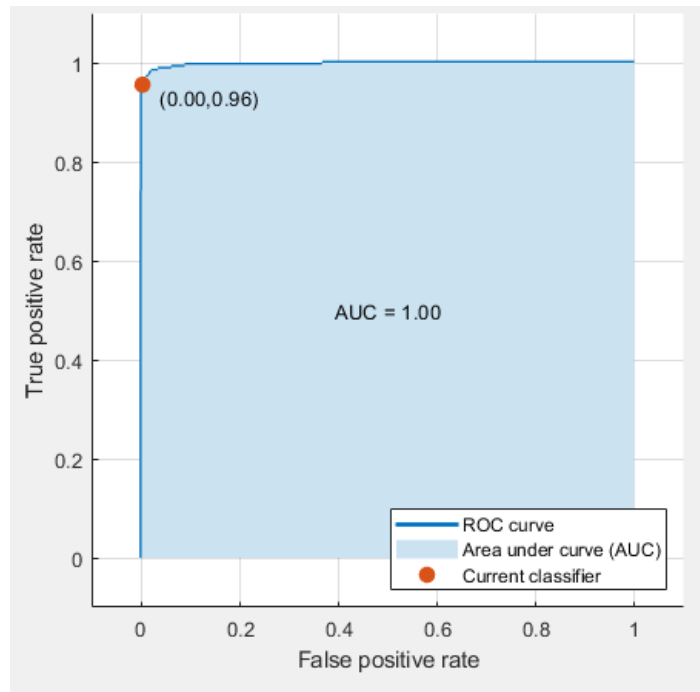


Anexo 2.2. Modelo SVM Cúbico para el parámetro de soplosidad.

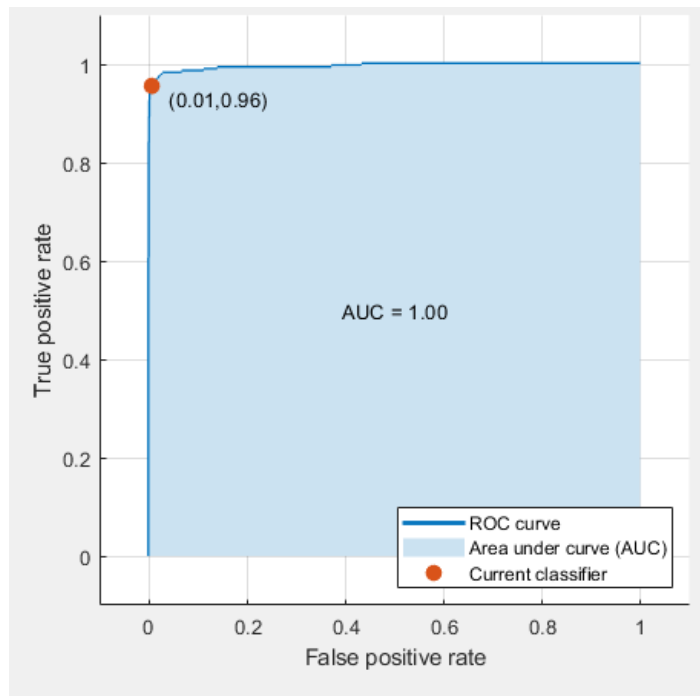
- Clase 0



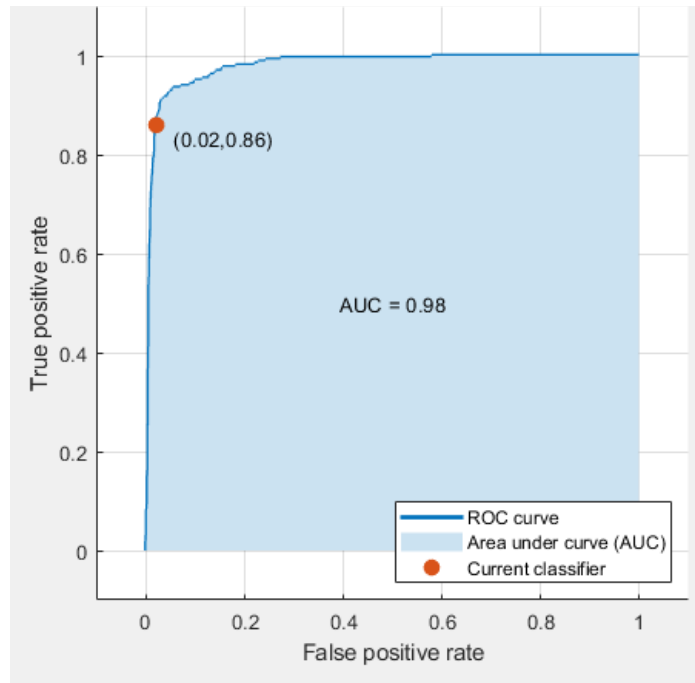
- Class 1



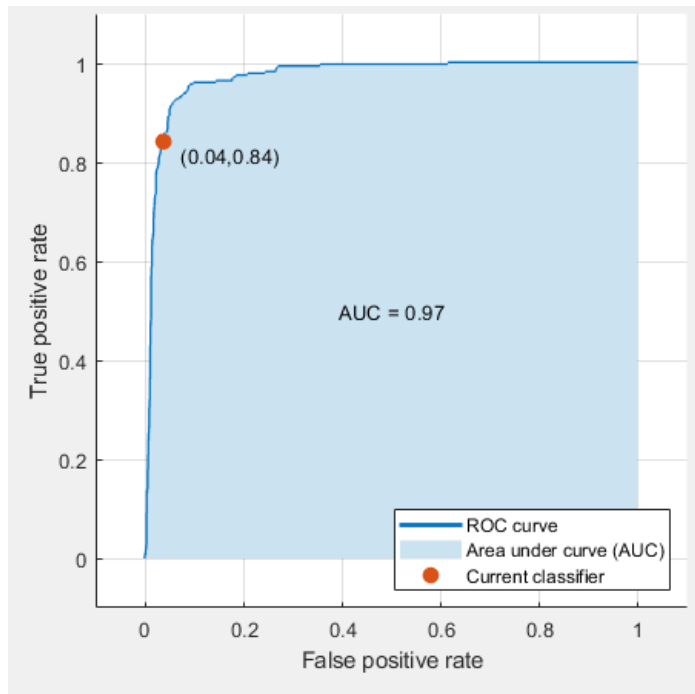
- Class 2



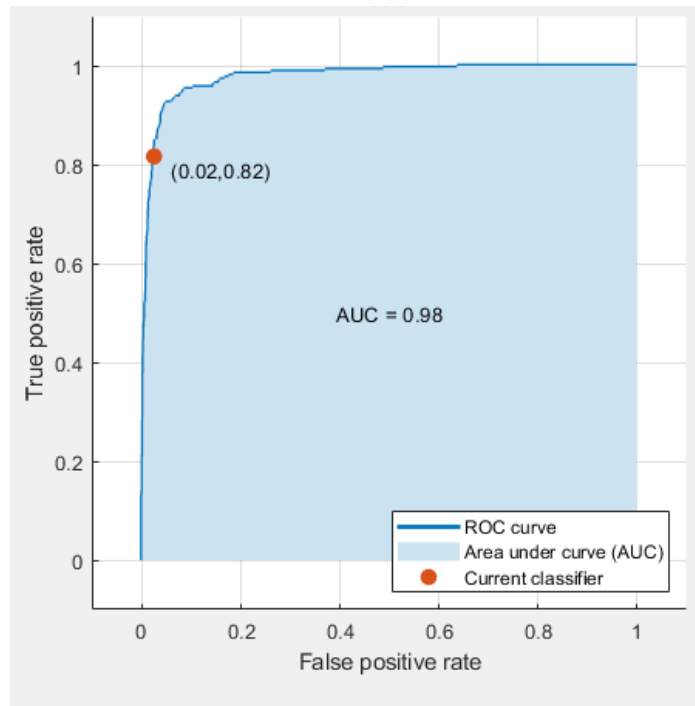
- Clase 3



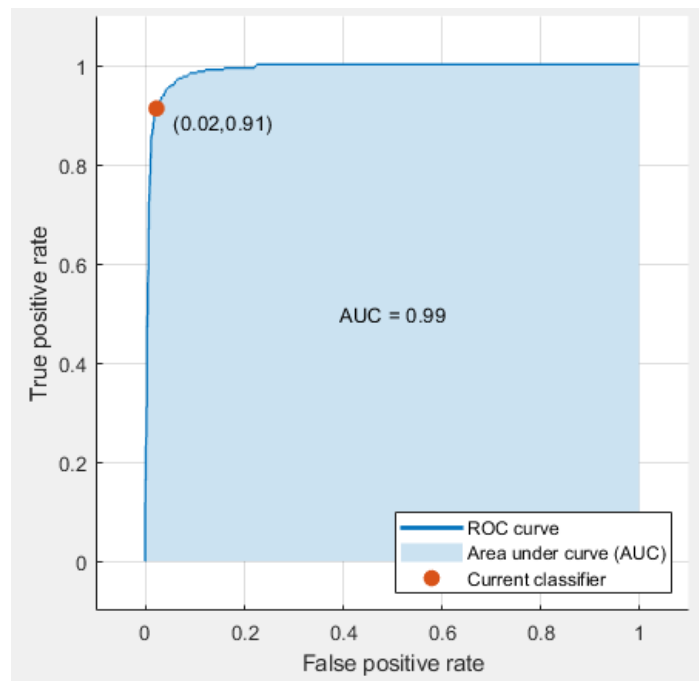
- Clase 4



- Clase 5

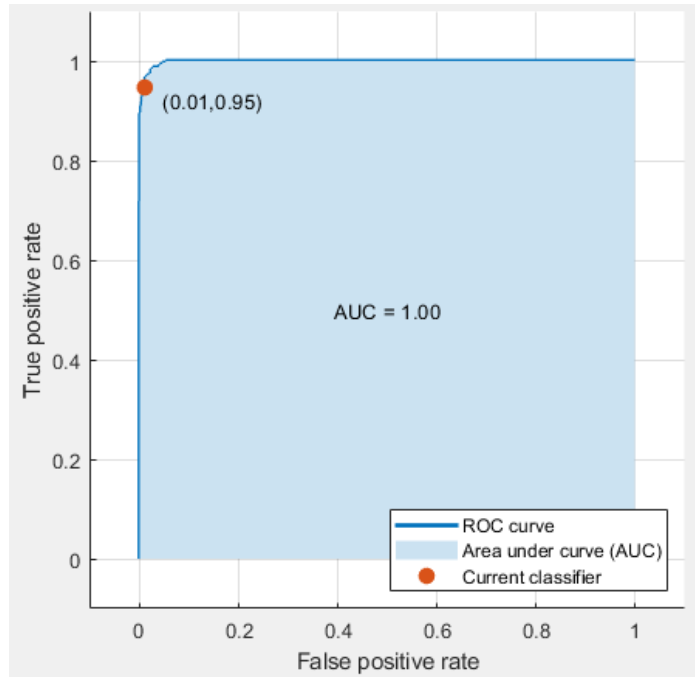


- Clase 6

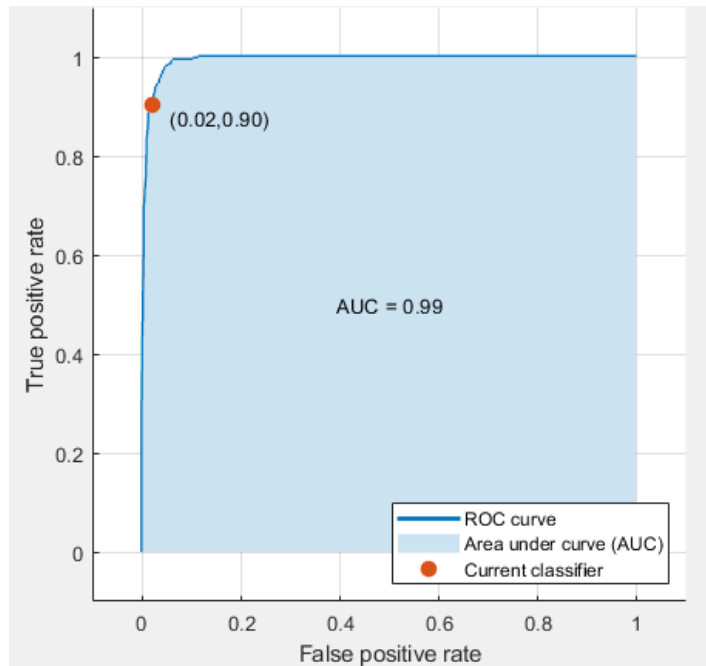


Anexo 2.3. Modelo SVM Cuadrático para el parámetro de astenia y tensión.

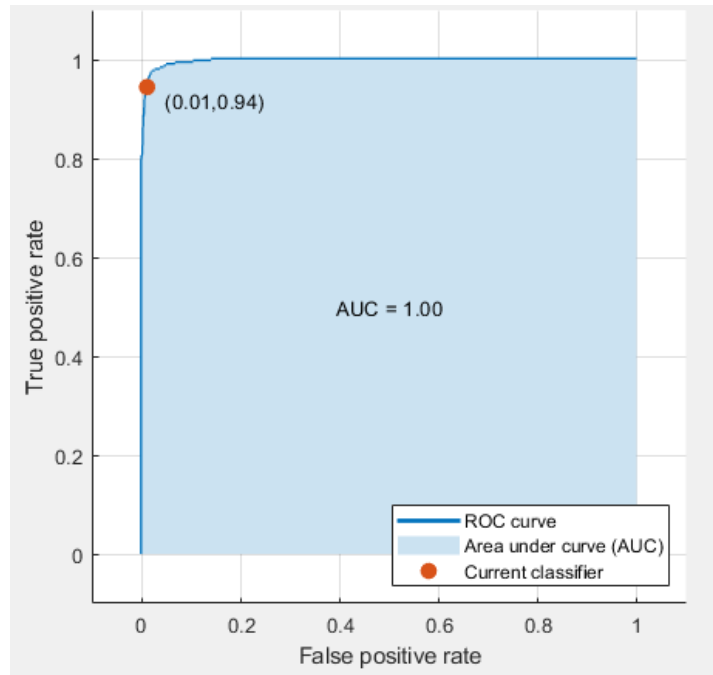
- Clase -3



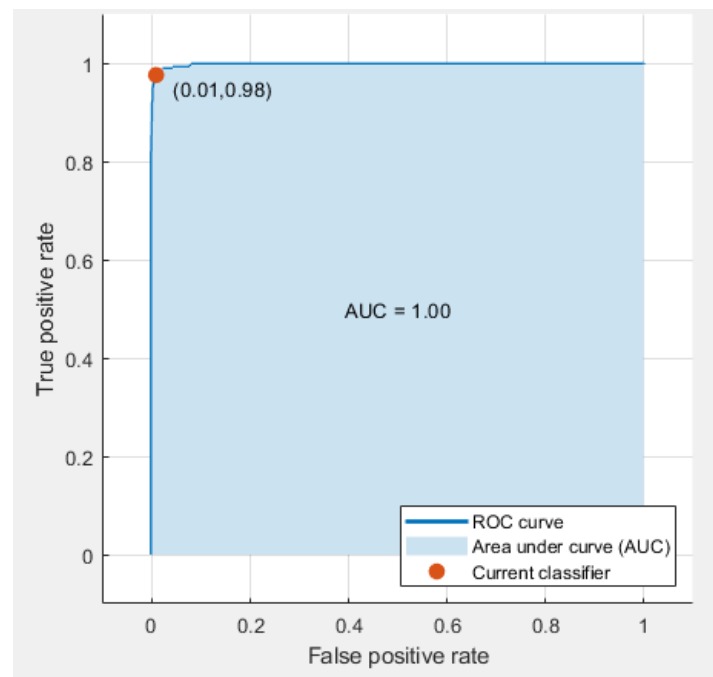
- Clase -2



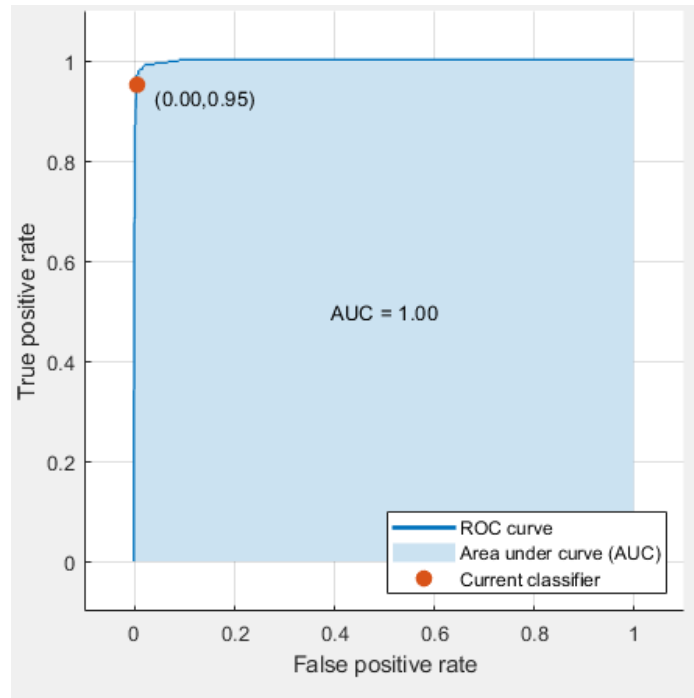
- Clase -1



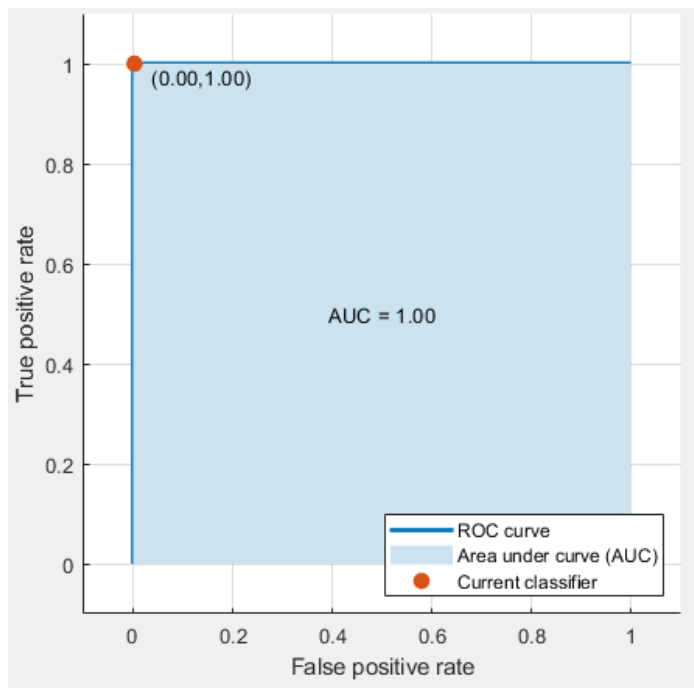
- Clase 1



- Clase 2

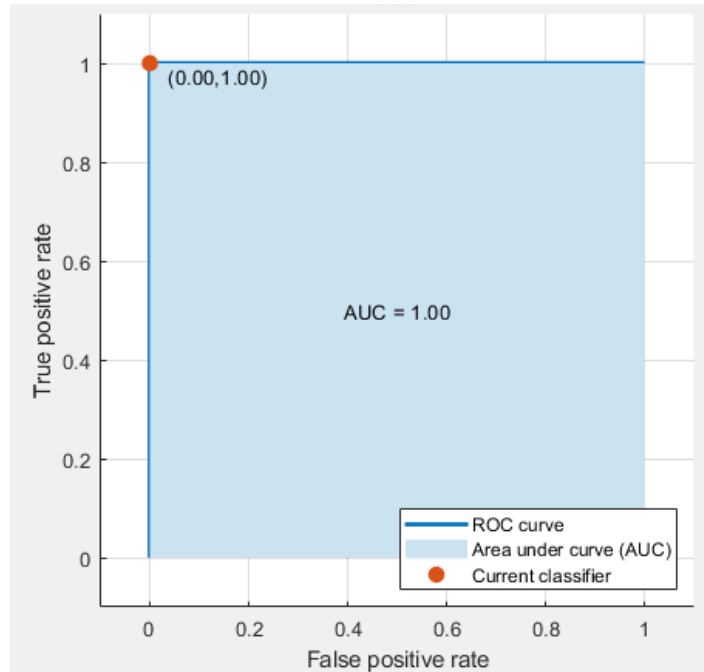


- Clase 3

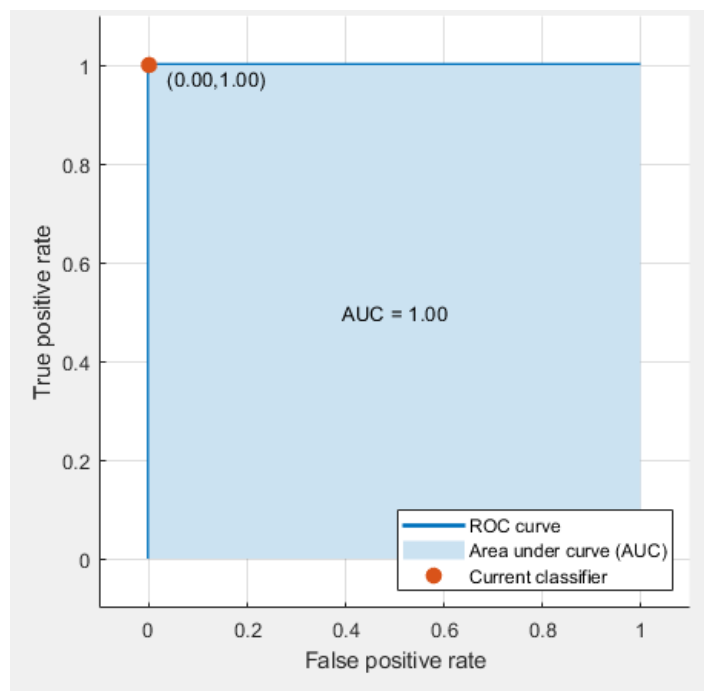


Anexo 2.4. Modelo SVM Cúbico para el parámetro grado.

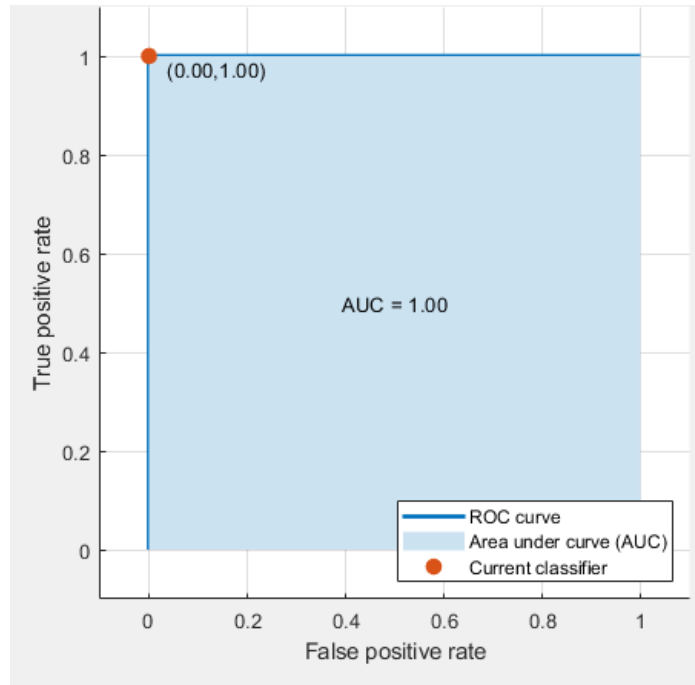
- Clase 0



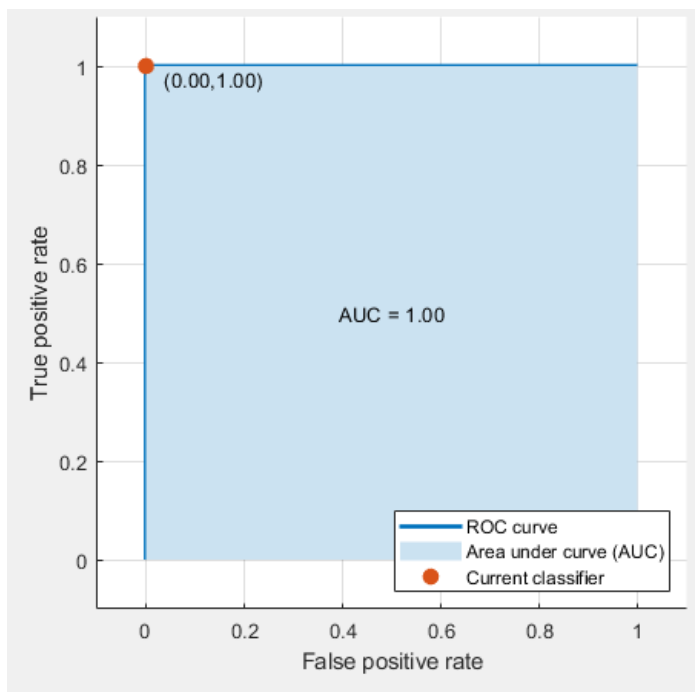
- Clase 1



- Clase 2



- Clase 3

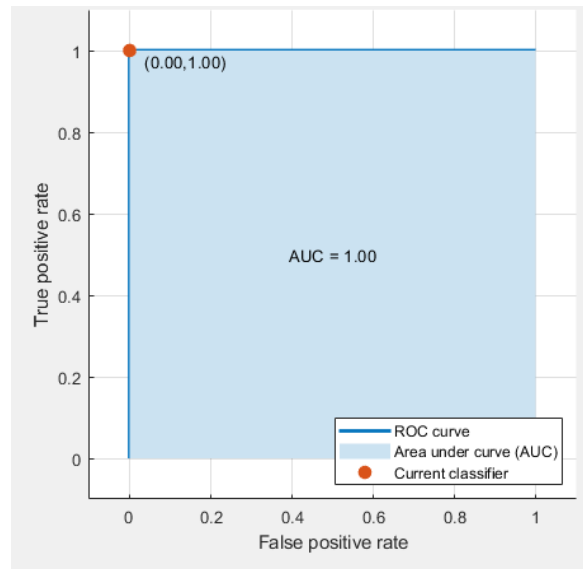


C. Anexo 3: Curvas ROC de los modelos para el género femenino

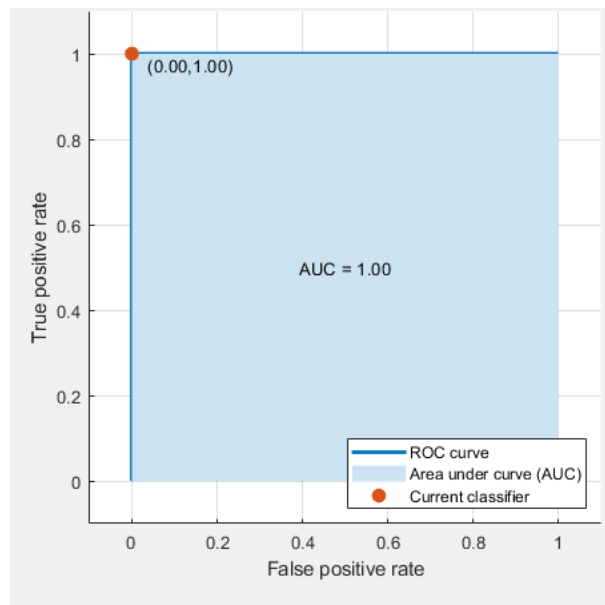
En esta sección, se muestra la curva Roc de cada clase para cada parámetro GRBAS, de los modelos seleccionados para el género femenino, donde muestran la sensibilidad frente a la especificidad de la clasificación binaria para cada clase, y la capacidad discriminatoria diagnóstica representada por el valor de AUC.

Anexo 3.1. Modelo SVM Cúbico para el parámetro soplosidad.

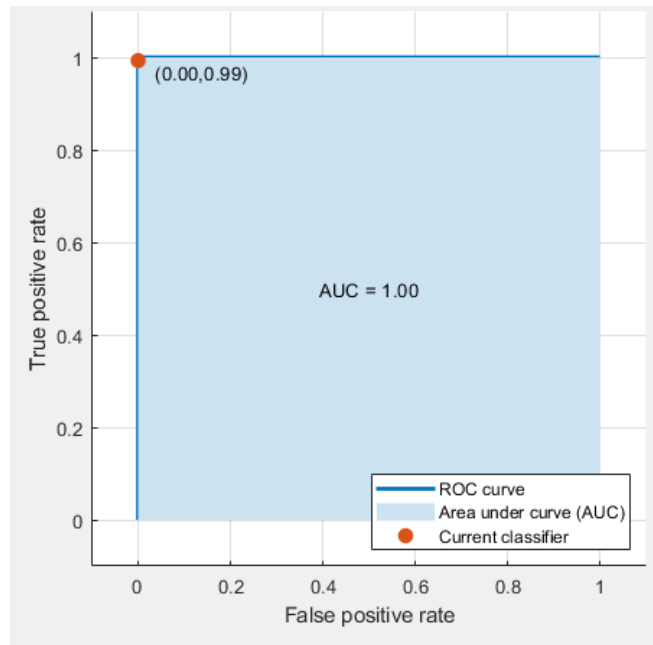
- Clase 0



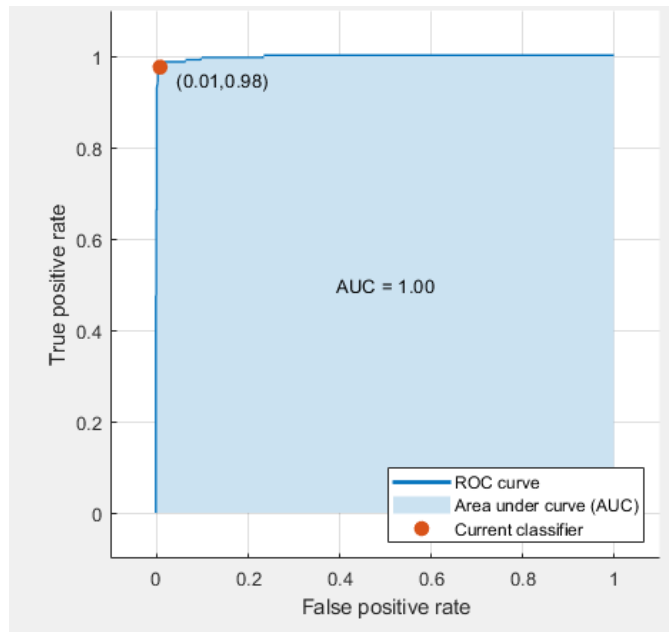
- Clase 1



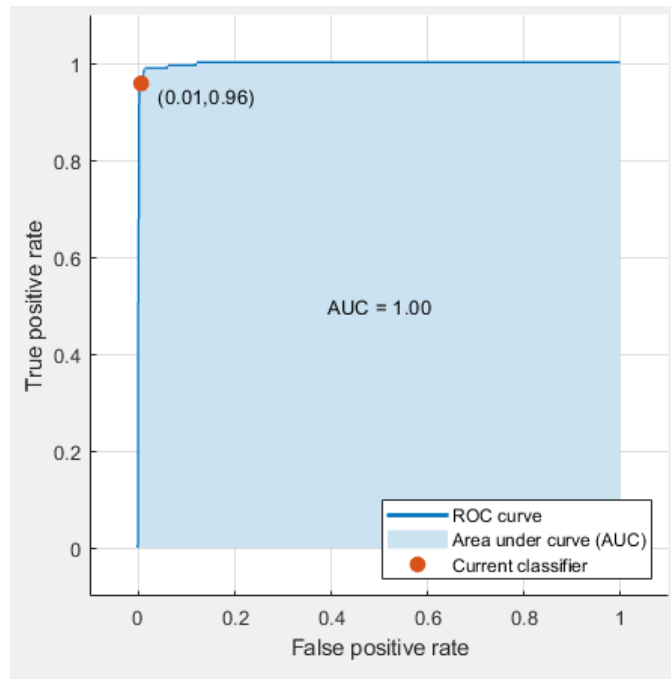
- Clase 2



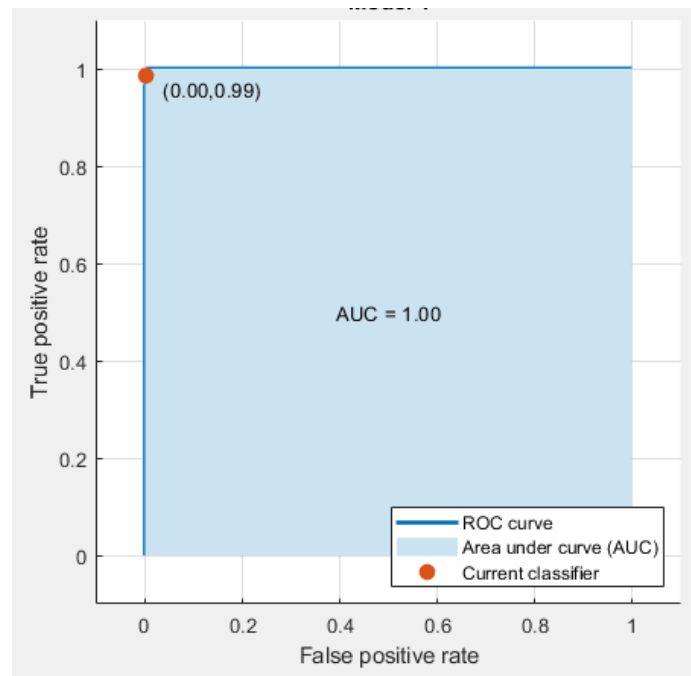
- Clase 3



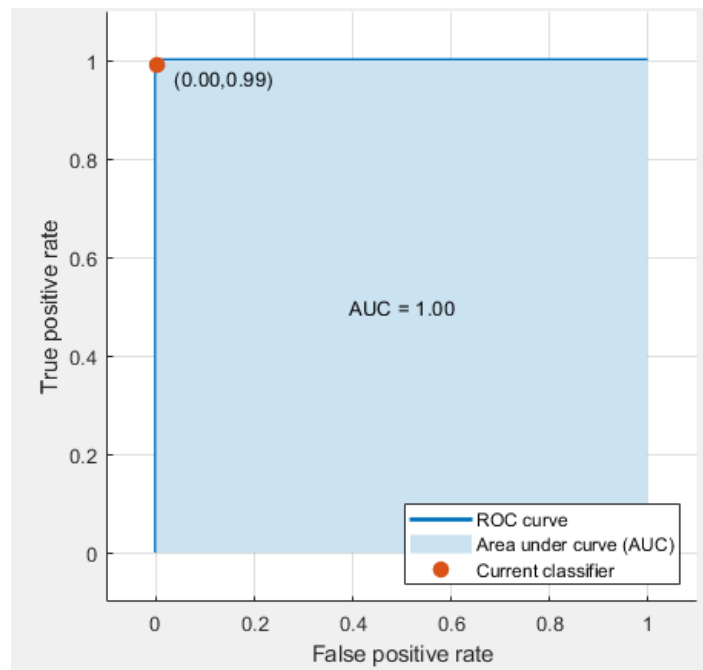
- Clase 4



- Clase 5

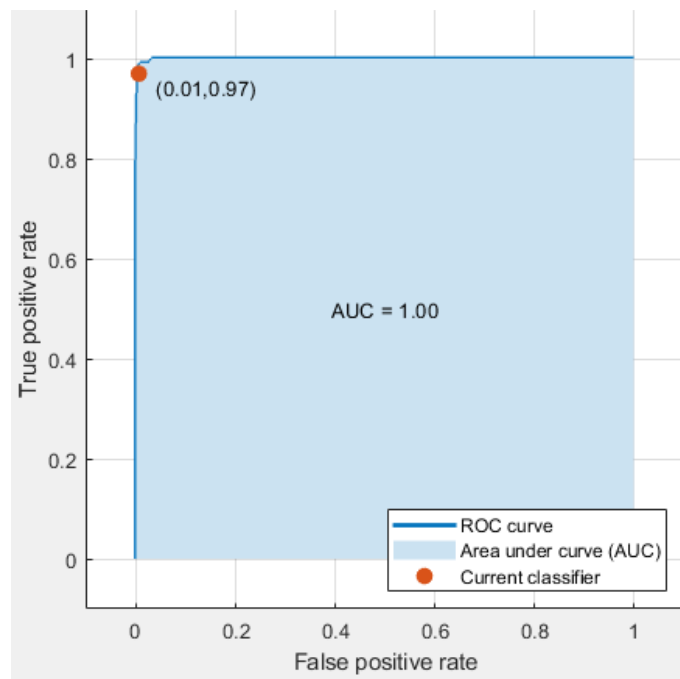


- Clase 6

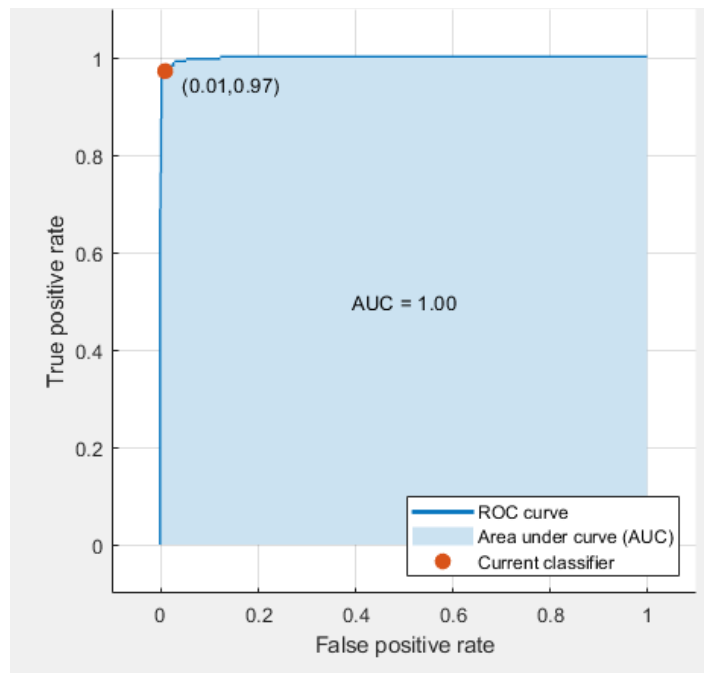


Anexo 3.2. Modelo SVM Cúbico para el parámetro de astenia tensión.

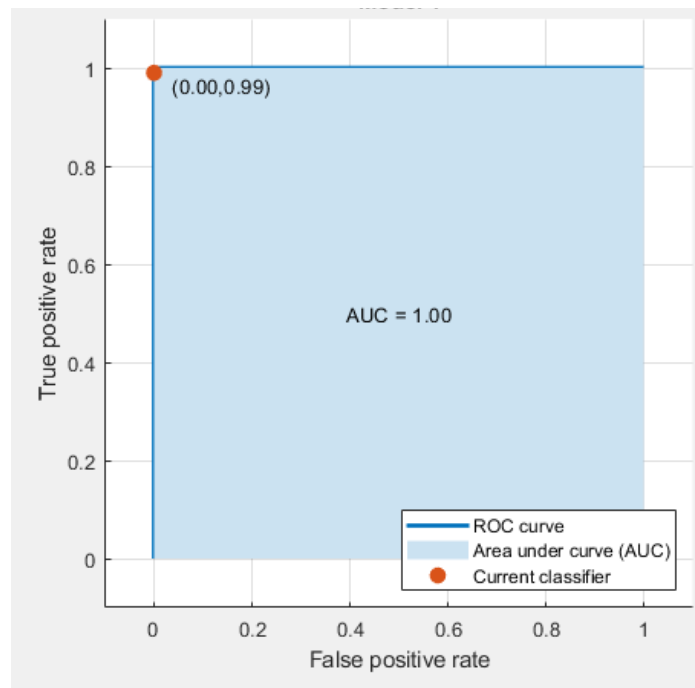
- Clase -3



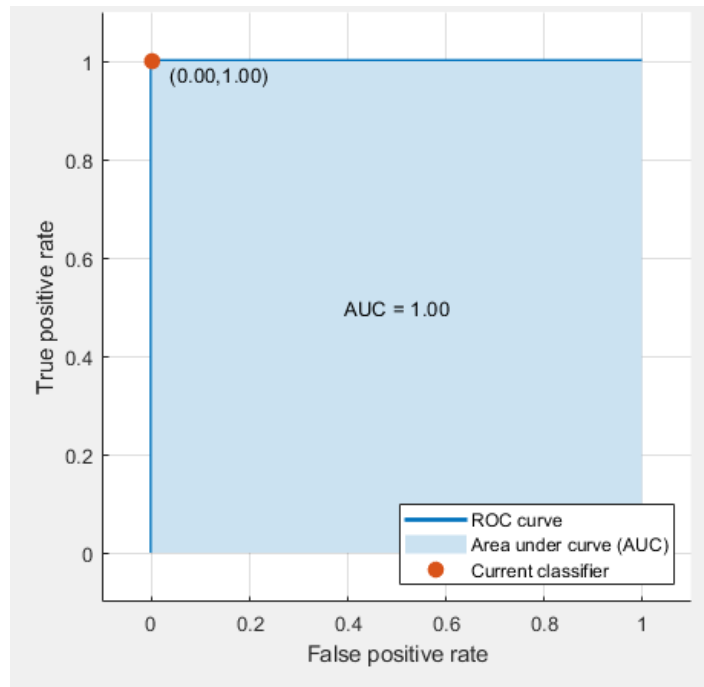
- Clase -2



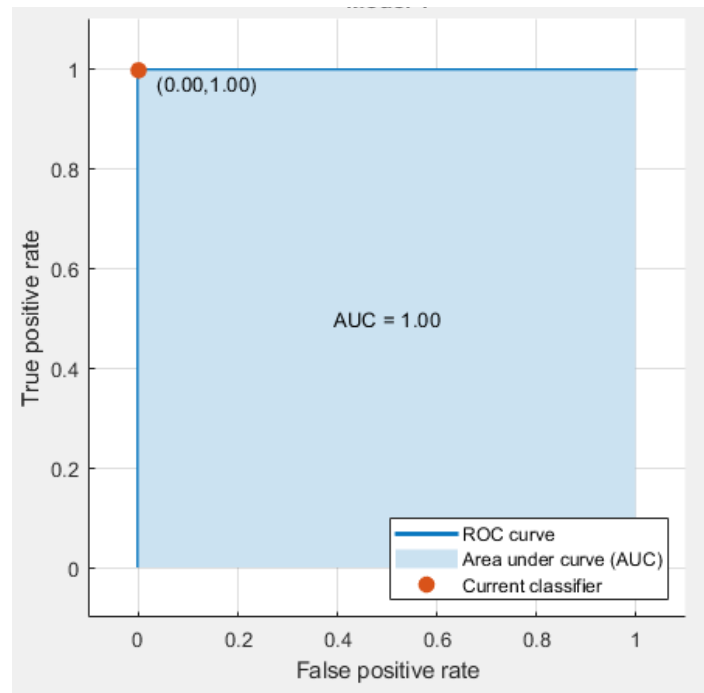
- Clase -1



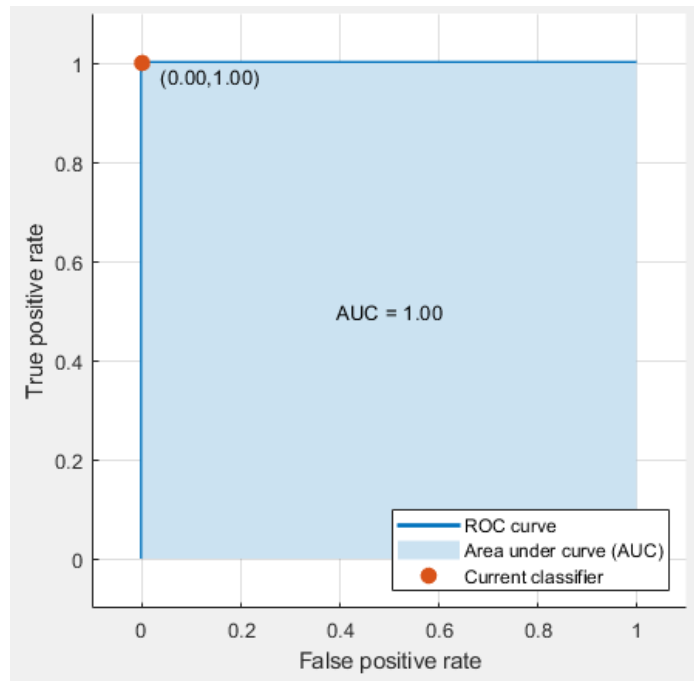
- Class 1



- Class 2



- Class 3

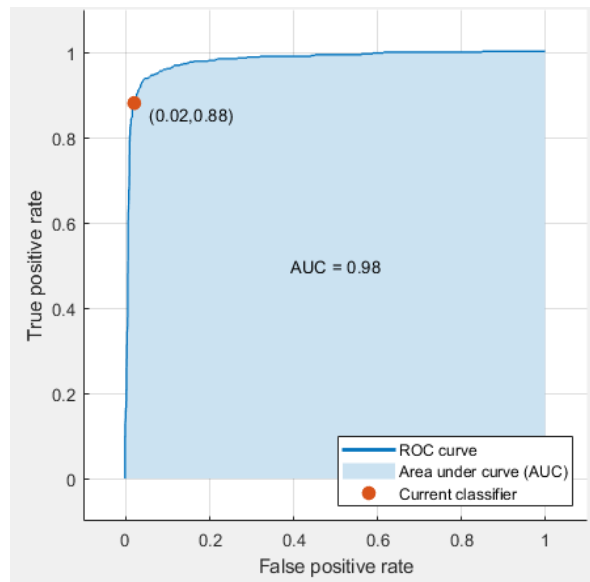


D. Anexo 4: Curvas ROC de los modelos no dedicados

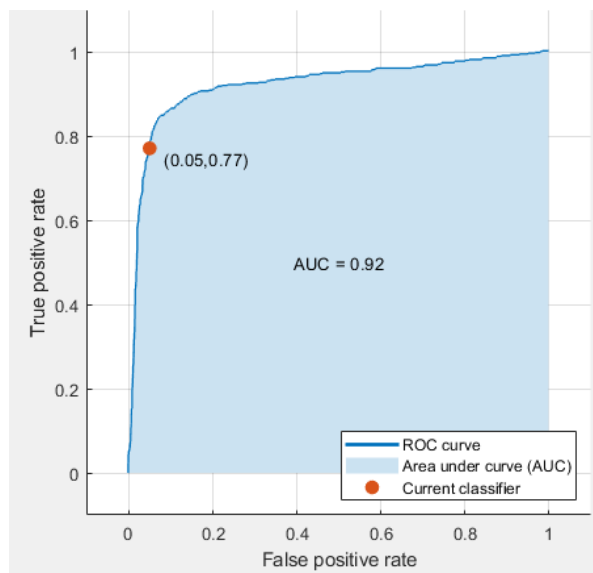
En esta sección, se muestra la curva Roc de cada clase para cada parámetro GRBAS, de los modelos seleccionados no dedicados, donde muestran la sensibilidad frente a la especificidad de la clasificación binaria para cada clase, y la capacidad discriminatoria diagnóstica representada por el valor de AUC.

Anexo 4.1. Modelo SVM Cúbico para el parámetro de ronquera.

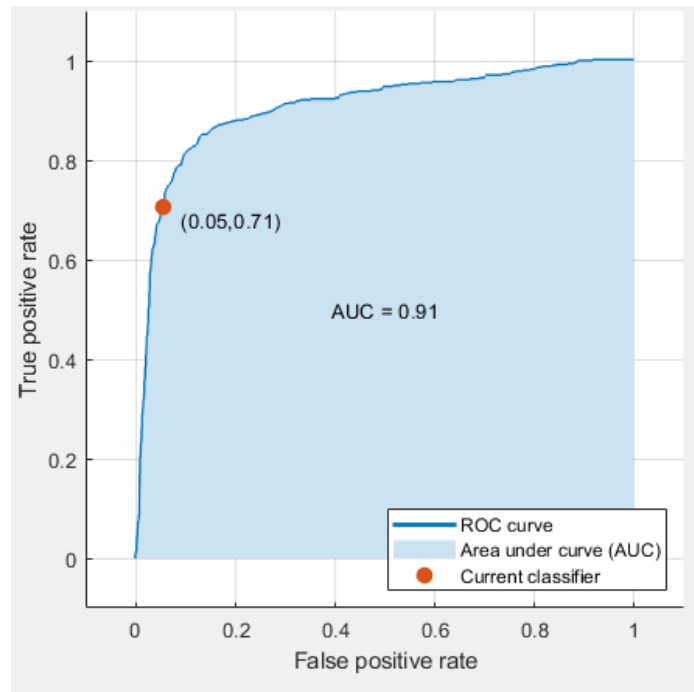
- Clase 0



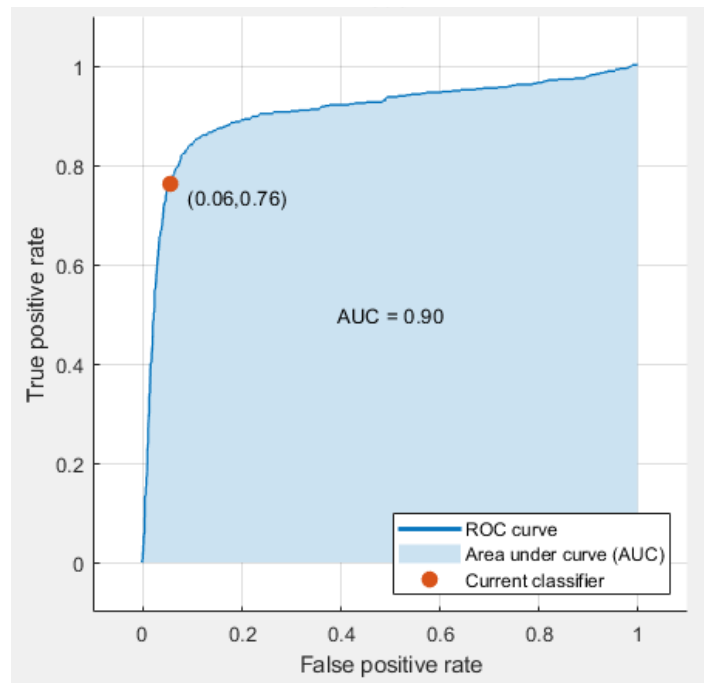
- Clase 1



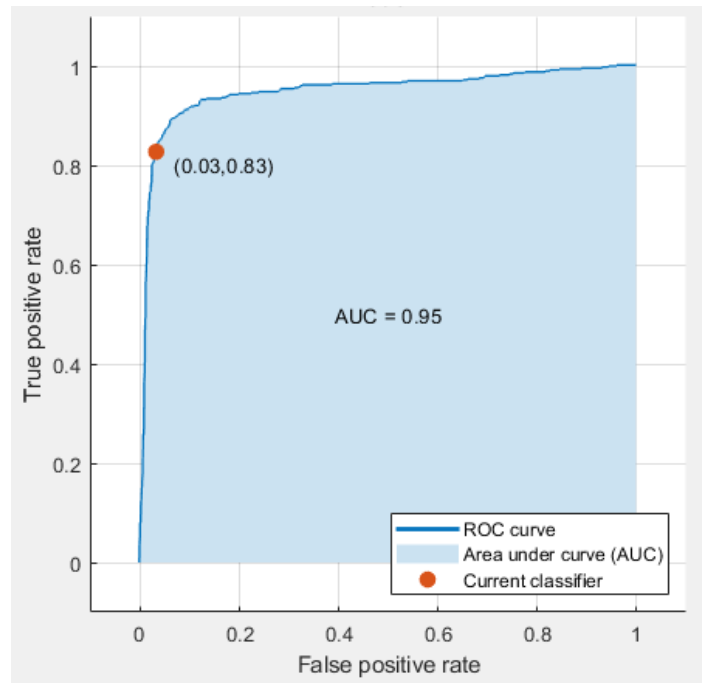
- Clase 2



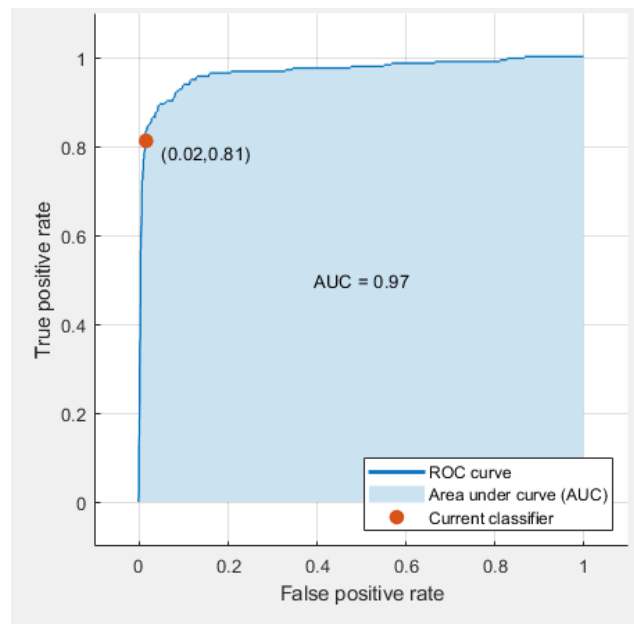
- Clase 3



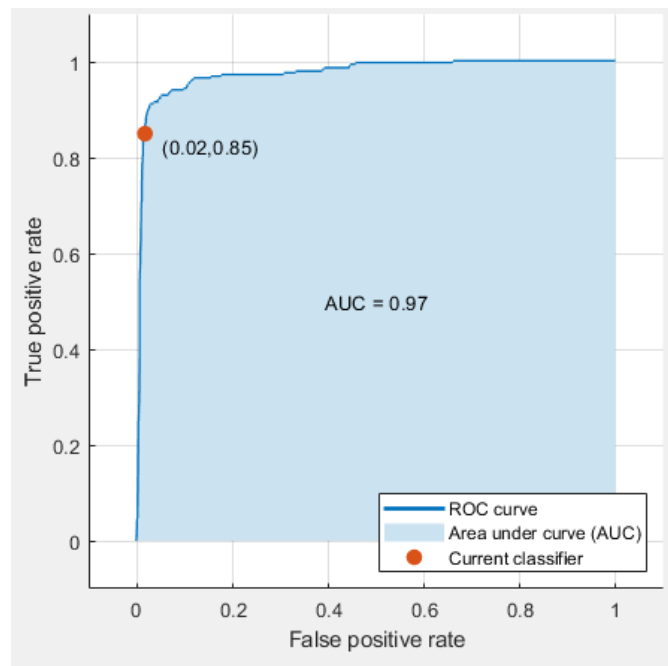
- Clase 4



- Clase 5

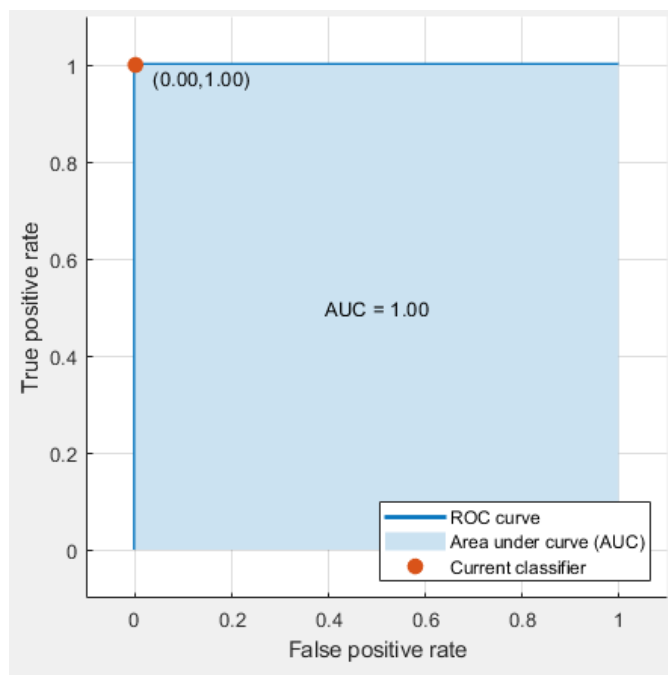


- Clase 6

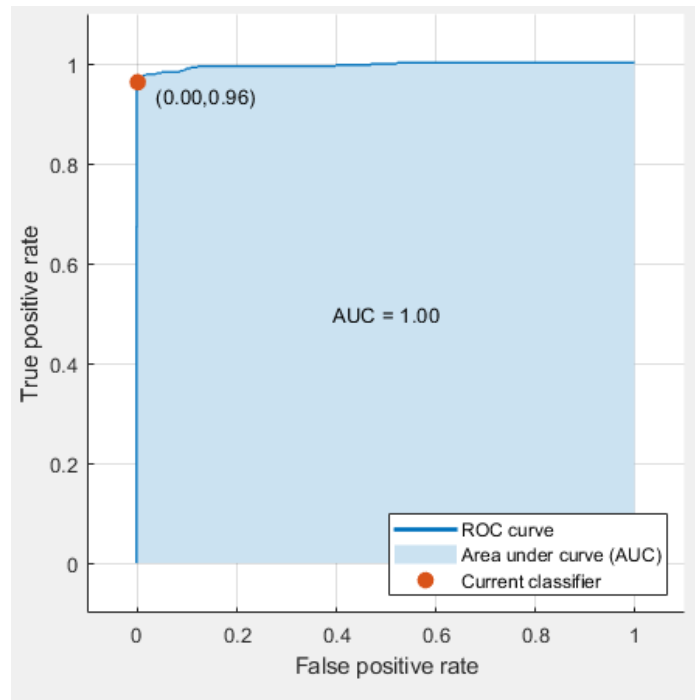


Anexo 4.2. Modelo SVM Cúbico para el parámetro de soplosidad.

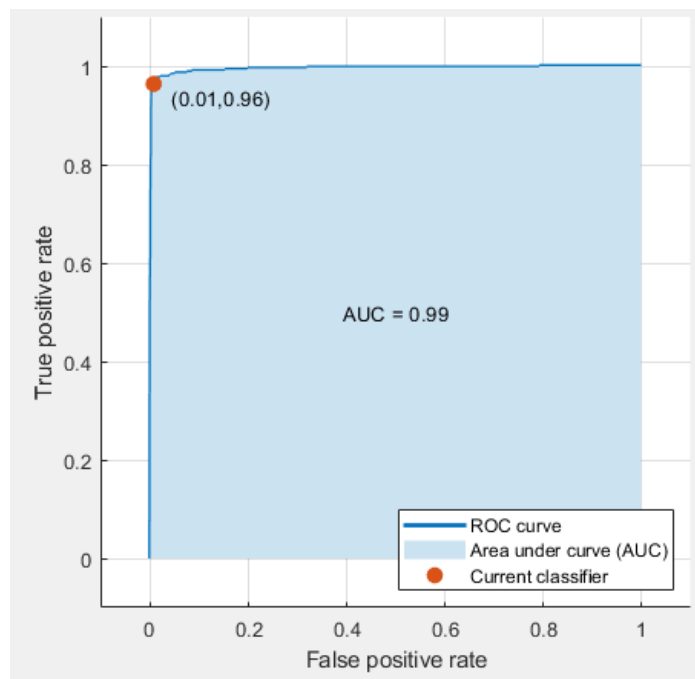
- Clase 0



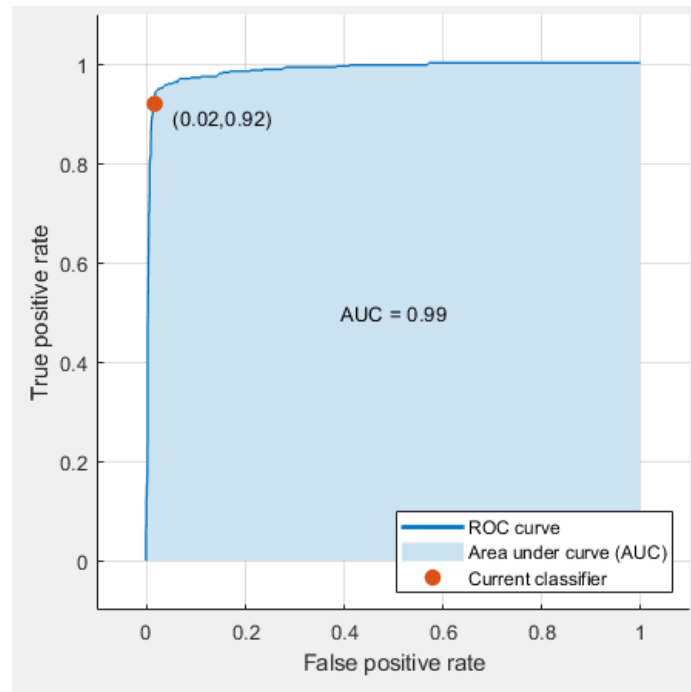
- Clase 1



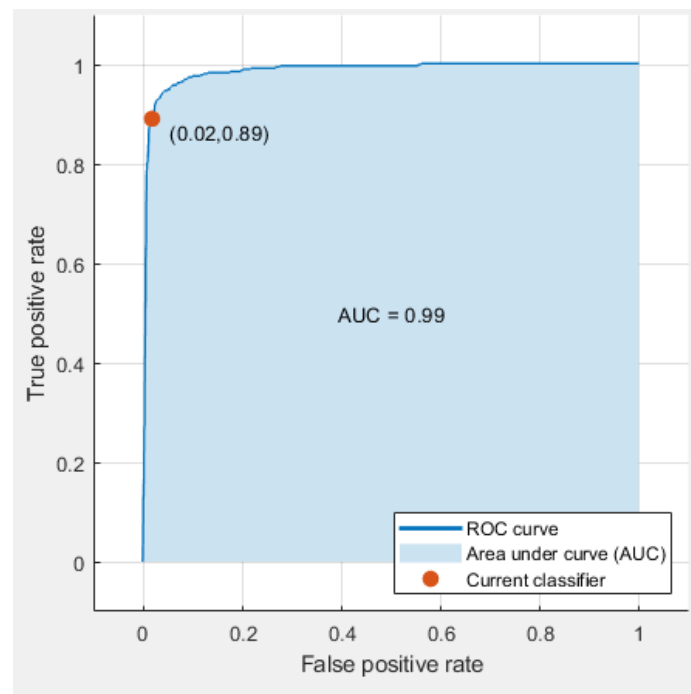
- Clase 2



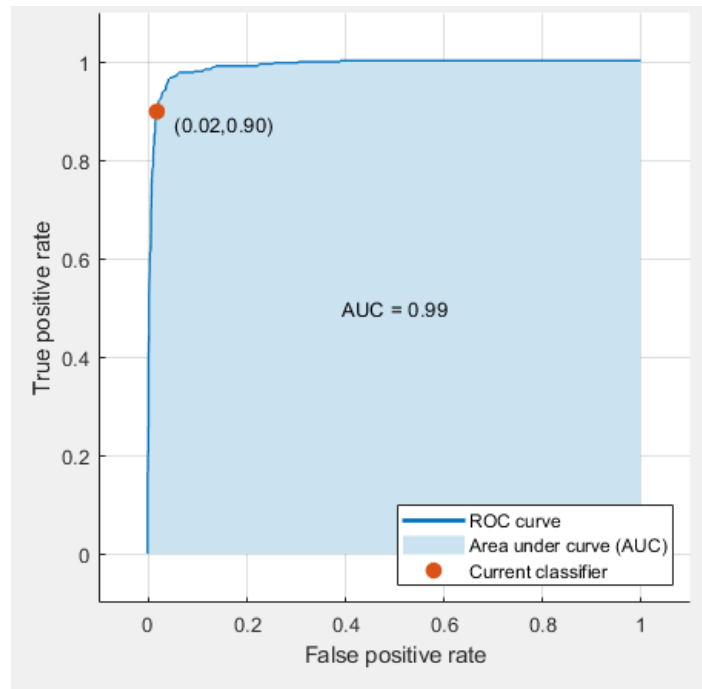
- Clase 3



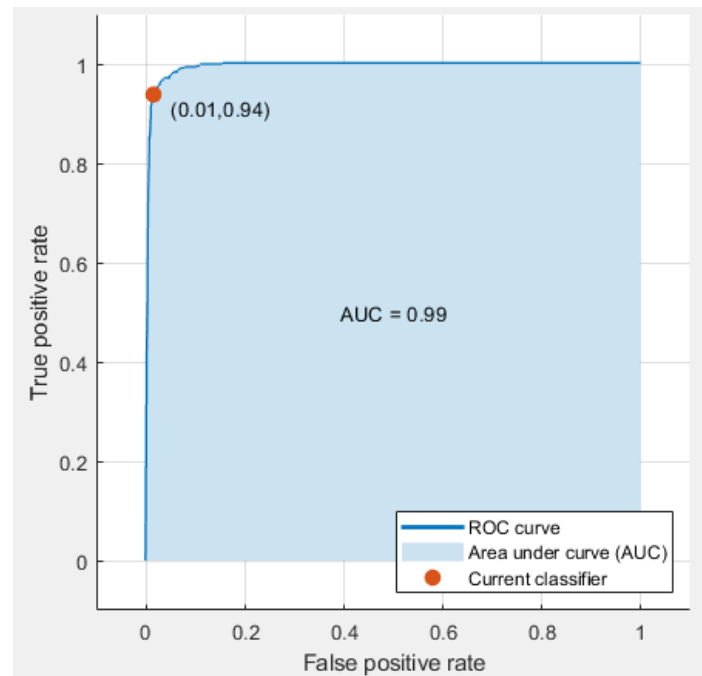
- Clase 4



- Clase 5

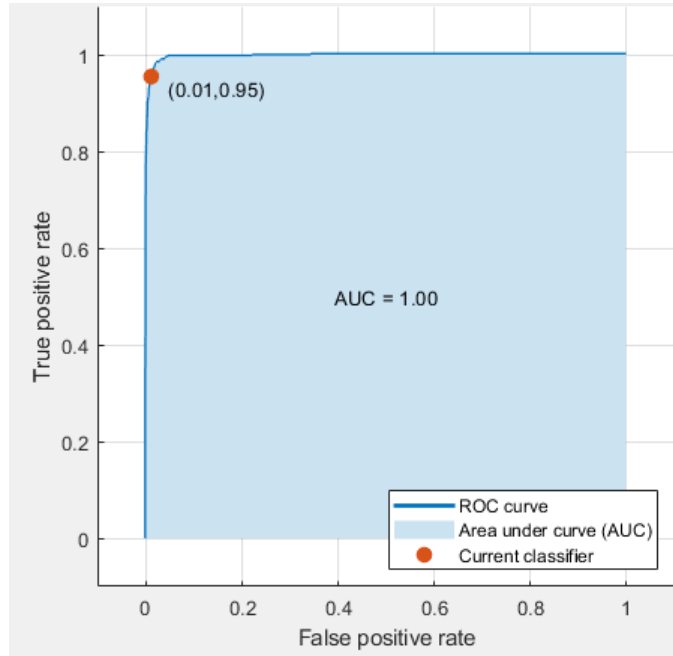


- Clase 6

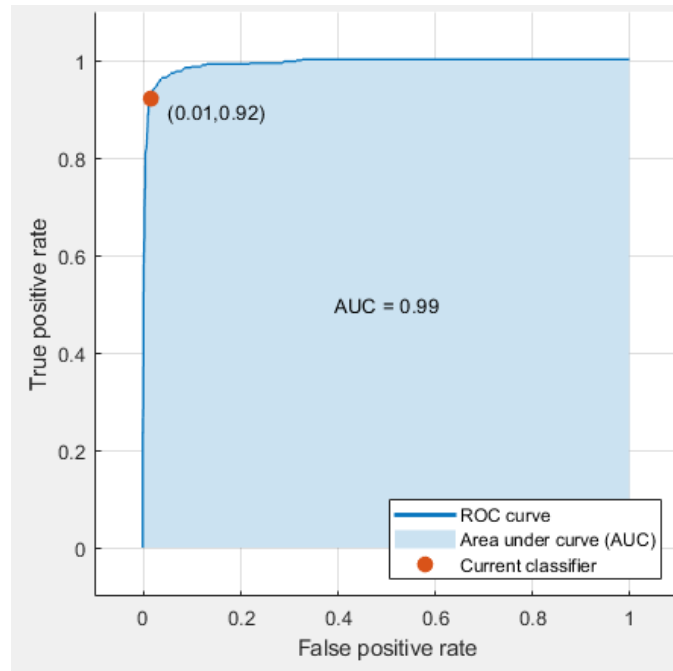


Anexo 4.3. Modelo SVM Cúbico para el parámetro de astenia y tensión.

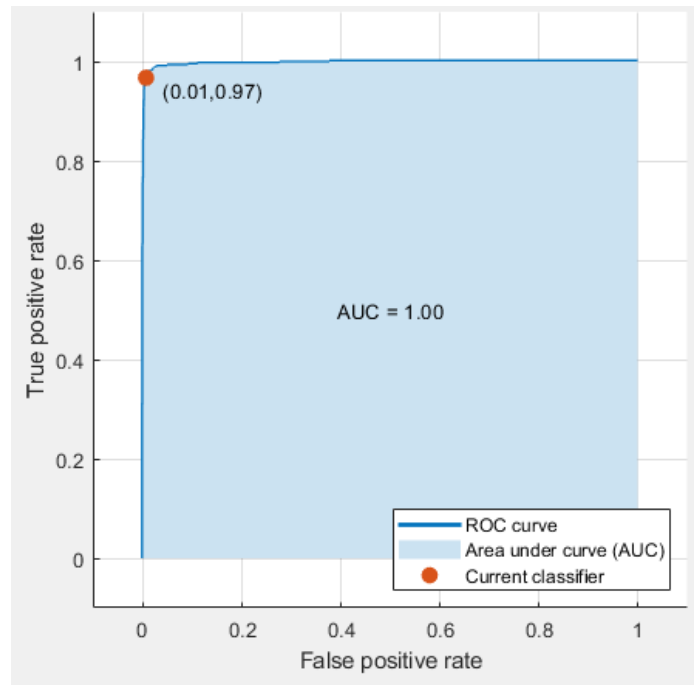
- Clase -3



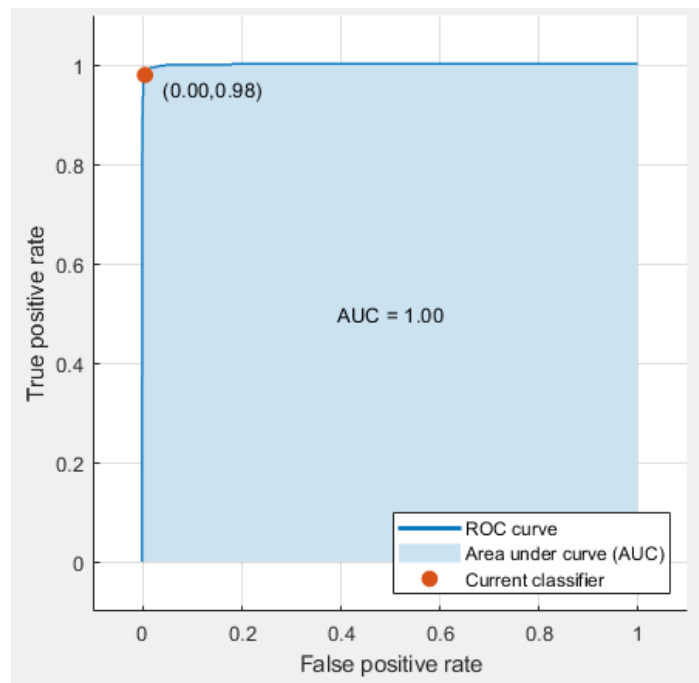
- Clase -2



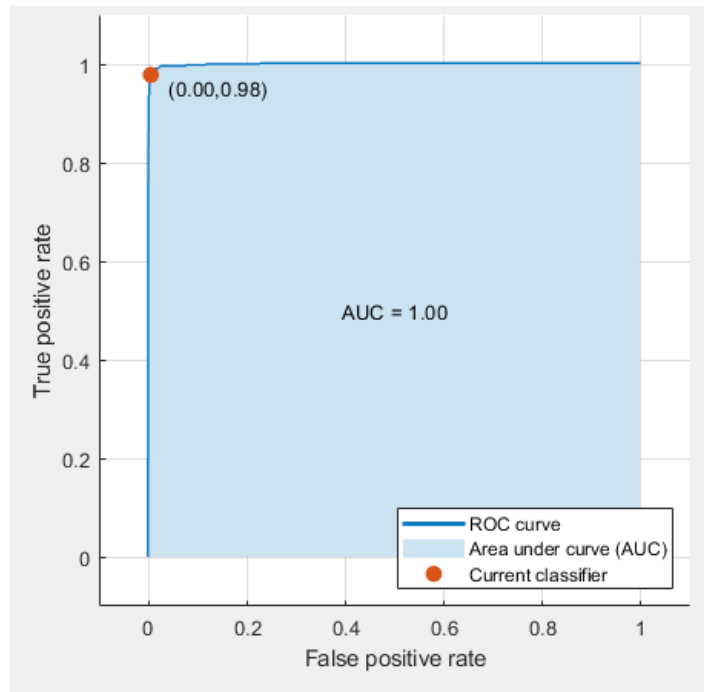
- Clase -1



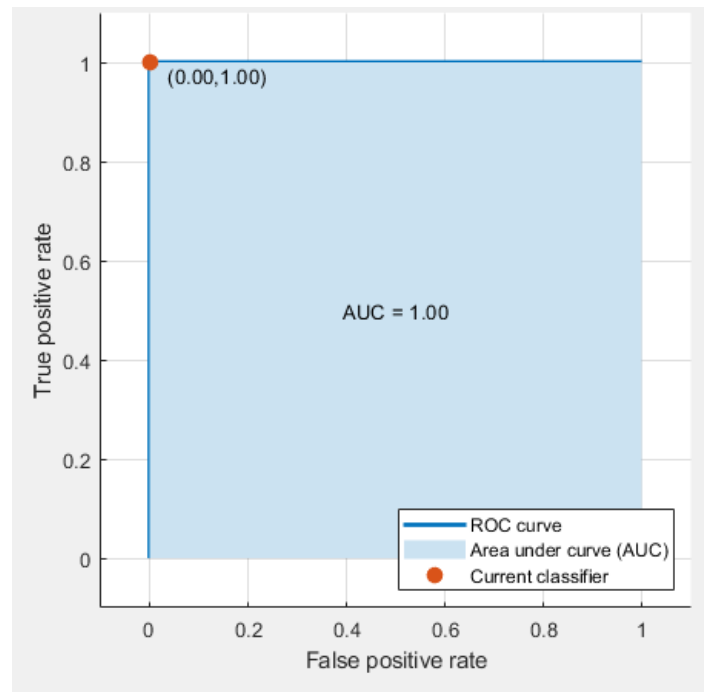
- Clase 1



- Class 2

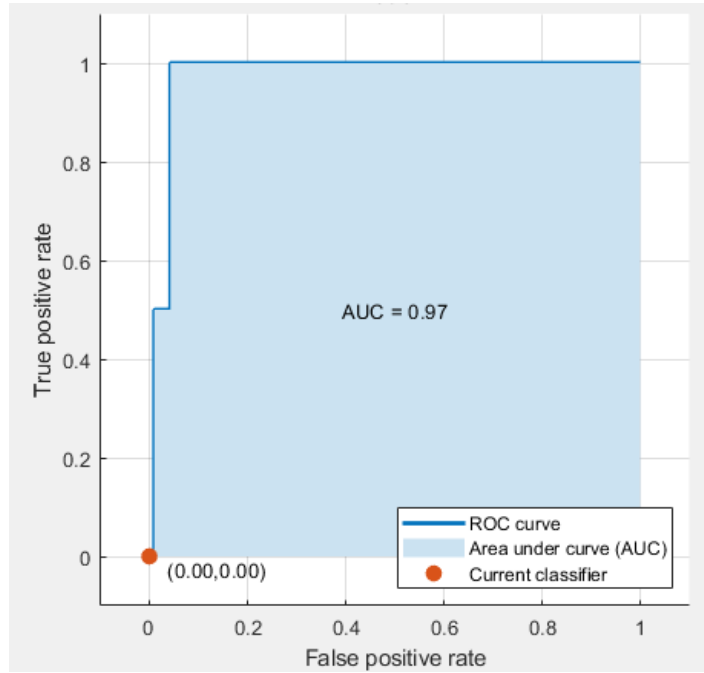


- Class 3

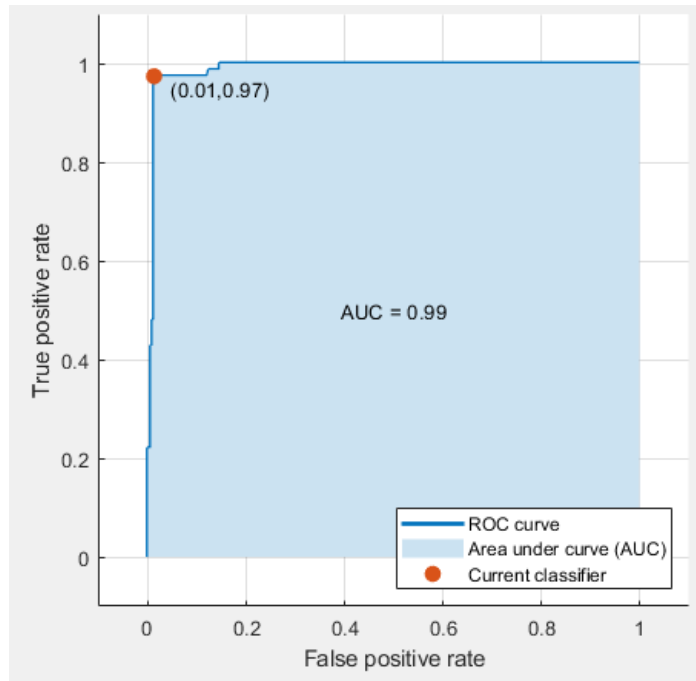


Anexo 4.4. Modelo SVM con Kernel Gaussiano Medio para el parámetro grado.

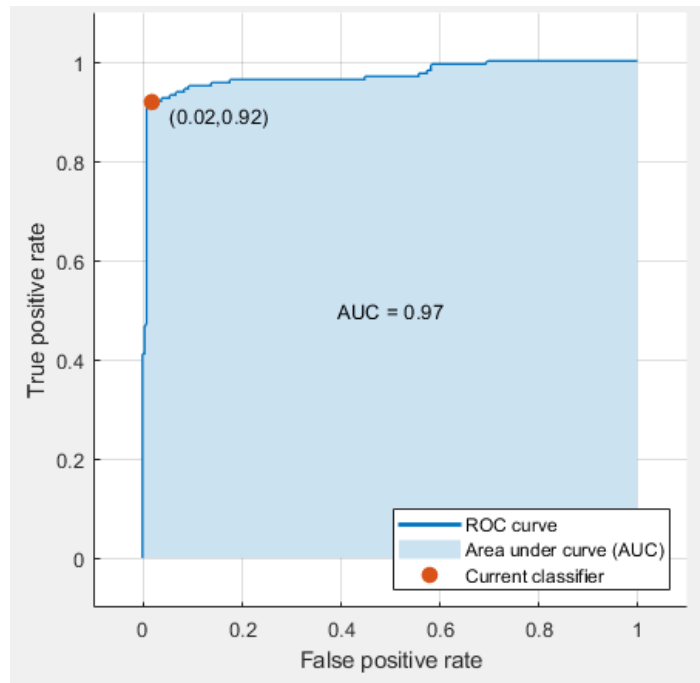
- Clase 0



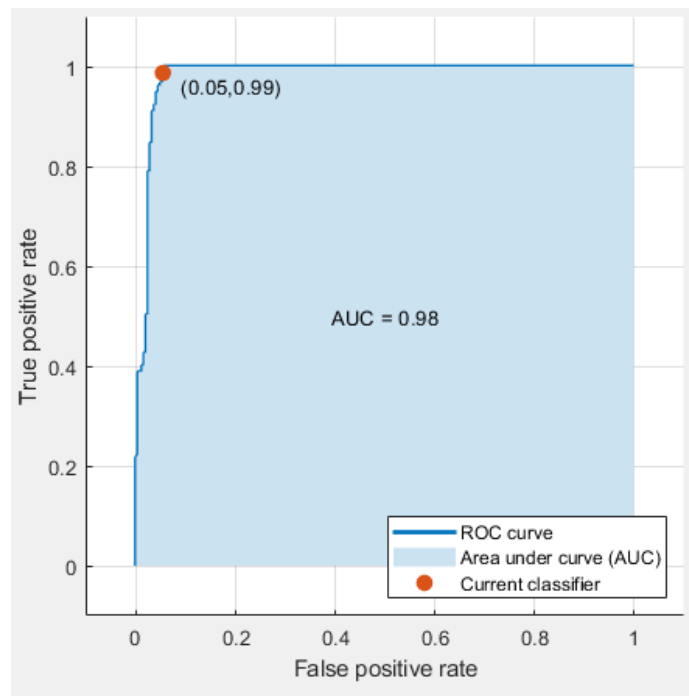
- Clase 1



- Clase 2



- Clase 3





**Desarrollo de una aplicación software para la
evaluación de la calidad vocal con
base en la escala GRBAS.**

Manual de Usuario

Versión: 1.0

Fecha: 24/Noviembre/2021

[Versión 1.0]

Queda prohibido cualquier tipo de explotación y, en particular, la reproducción, distribución, comunicación pública y/o transformación, total o parcial, por cualquier medio, de este documento sin el previo consentimiento expreso por la Universidad Antonio Nariño.

CONTENIDO

1	DESCRIPCIÓN DEL SISTEMA	3
1.1	Objeto.....	3
1.2	Alcance	3
1.3	Funcionalidad – Descripción del Sistema.....	3
1.3.1	Ejecución de la herramienta	4
1.3.2	Selección de audio	6
1.3.3	Análisis de voz	6
1.3.4	Resultados.....	7

1 DESCRIPCIÓN DEL SISTEMA

1.1 Objeto

Herramienta computacional sobre dispositivos con sistema operativo *Windows*, diseñada para realizar la evaluación de la calidad vocal con base en la escala GRBAS, y dirigido al personal logopédico como sistema de apoyo para al análisis e interpretación de las características acústicas, estadísticas y espectrales de la voz.

1.2 Alcance

Herramienta computacional destinada a la evaluación de la calidad vocal con base en la escala GRBAS: Fue desarrollado mediante el proceso de investigación y desarrollo en el Programa de Ingeniería Biomédica (UAN – Popayán). Dirigido hacia el personal logopédico, con el fin de reducir la brecha de subjetividad presente en una evaluación vocal convencional, por esta razón, el siguiente trabajo pretende desarrollar una aplicación software que permita realizar de manera automática e implementando algoritmos computacionales como también sistemas de machine learning, una evaluación que le sirva al logopeda como una segunda opinión obtenida de manera objetiva con el fin de reducir aquellos errores de diagnósticos que se presentan de manera inevitable al realizar una evaluación totalmente subjetiva.

1.3 Funcionalidad – Descripción del Sistema

El presente manual de usuario describe el funcionamiento de la aplicación “ATOLOG” y su uso adecuado para obtener los mejores resultados.

El software se ha diseñado para ser utilizado de manera simple por profesionales de la salud vocal, conocedores de esta rama o las áreas específicas sobre el análisis de la voz. A continuación, se describe de manera consecutiva cada una de las opciones que componen el software “ATOLOG”.

1.3.1 Ejecución de la herramienta

La ejecución o inicialización del archivo “ATOLOG.exe” presenta la interface principal de la aplicación (Figura 1), donde se pueden visualizar las diferentes secciones de las cuales está conformada, siendo la parte lateral izquierda (1), la sección donde se puede seleccionar la señal de audio en formato WAV, además del género de la persona a quien le corresponde la señal de voz; en la parte lateral derecha (2), se puede detallar la gráfica de la señal seleccionada, y en su parte inferior, se ubican 2 botones que le permiten al usuario visualizar en una nueva ventana las gráficas de espectrograma o densidad espectral de potencia; además, en la parte superior de esta interfaz se presentan 3 pestañas entre las cuales el usuario puede navegar para visualizar las diversas clases de información obtenida por la herramienta; la pestaña 2, muestra todas las características acústicas, espectrales y estadísticas, que la herramienta extrae de la señal de voz (Figura 2); y la pestaña 3, presenta el diagnóstico de la señal con base en la escala GRBAS de manera tabular y gráfica, siendo este grafico un diagrama de barras (Figura 3)

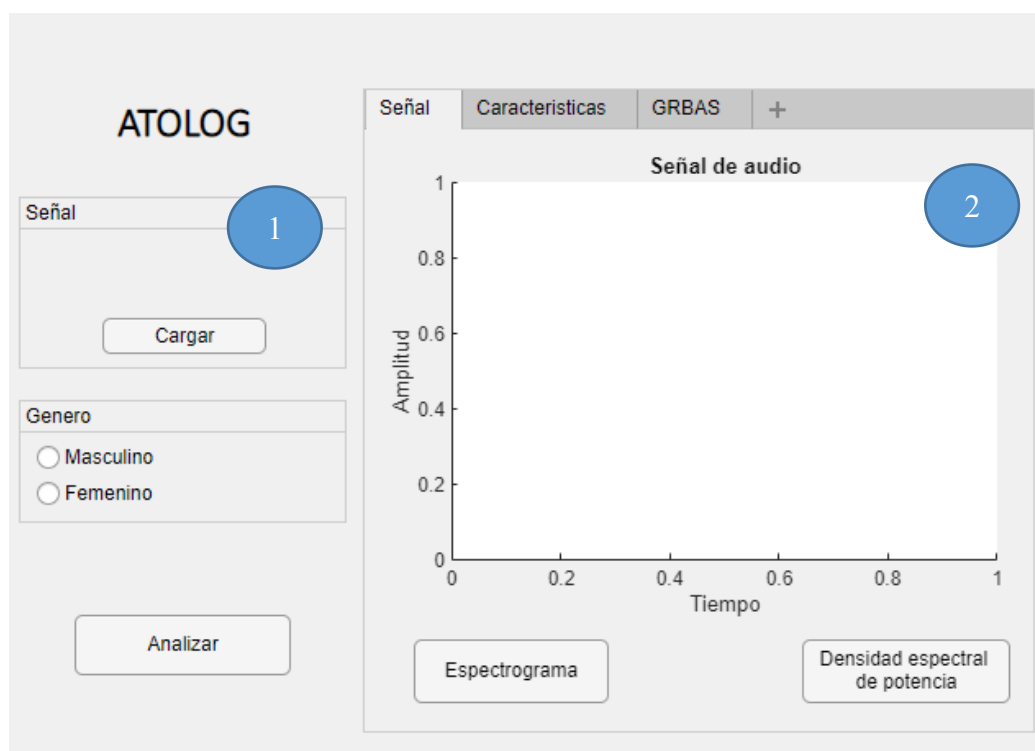


Figura 1: Interface general (Pestaña 1).

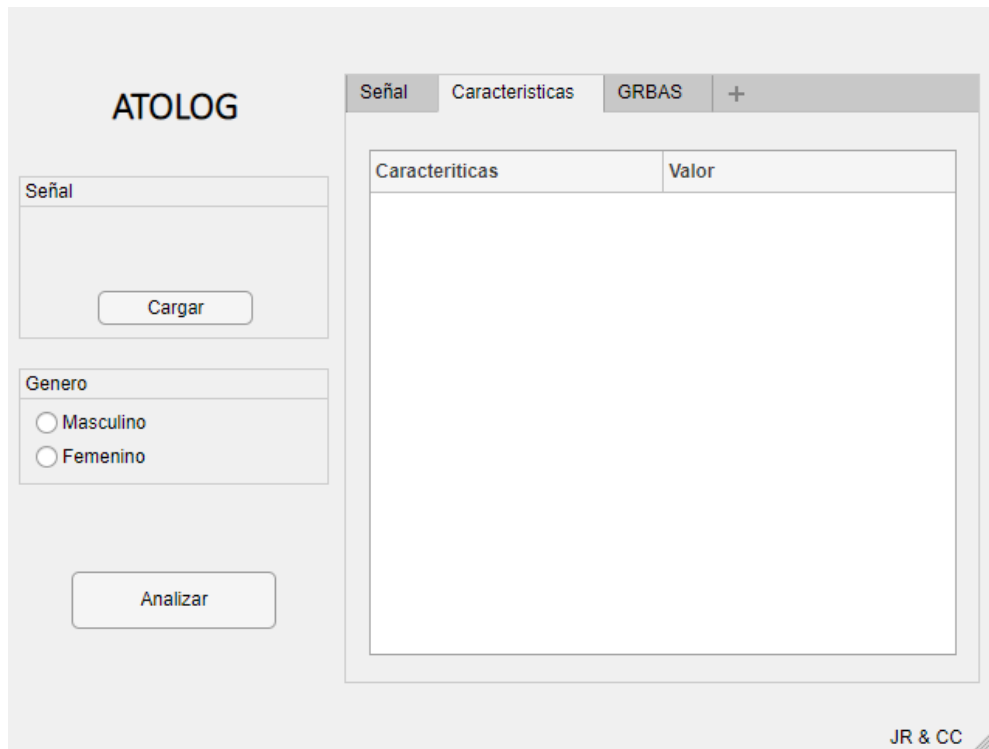


Figura 2: Interface general (Pestaña 2).

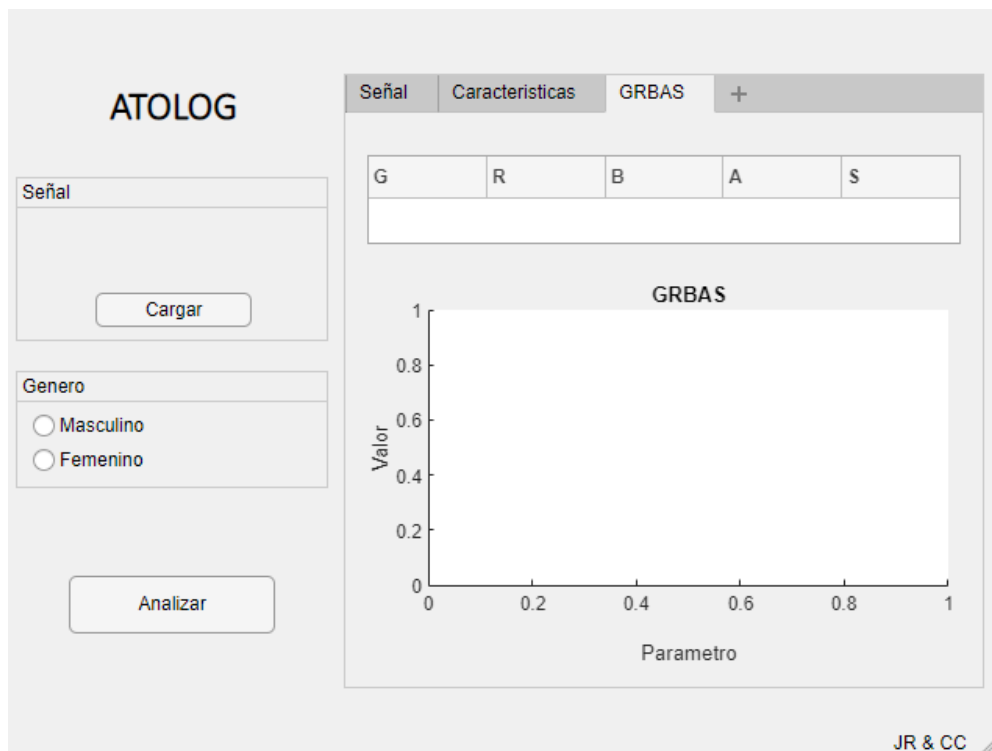


Figura 3: Interface general (Pestaña 3).

1.3.2 Selección de audio

Al presionar el botón “Cargar” ubicado en la parte central de la sección lateral izquierda de la interfaz, se despliega automáticamente una ventana de exploración de los archivos internos (Figura 4), donde el usuario puede seleccionar la señal de audio de interés que se va a ingresar y posteriormente analizar (En formato WAV).

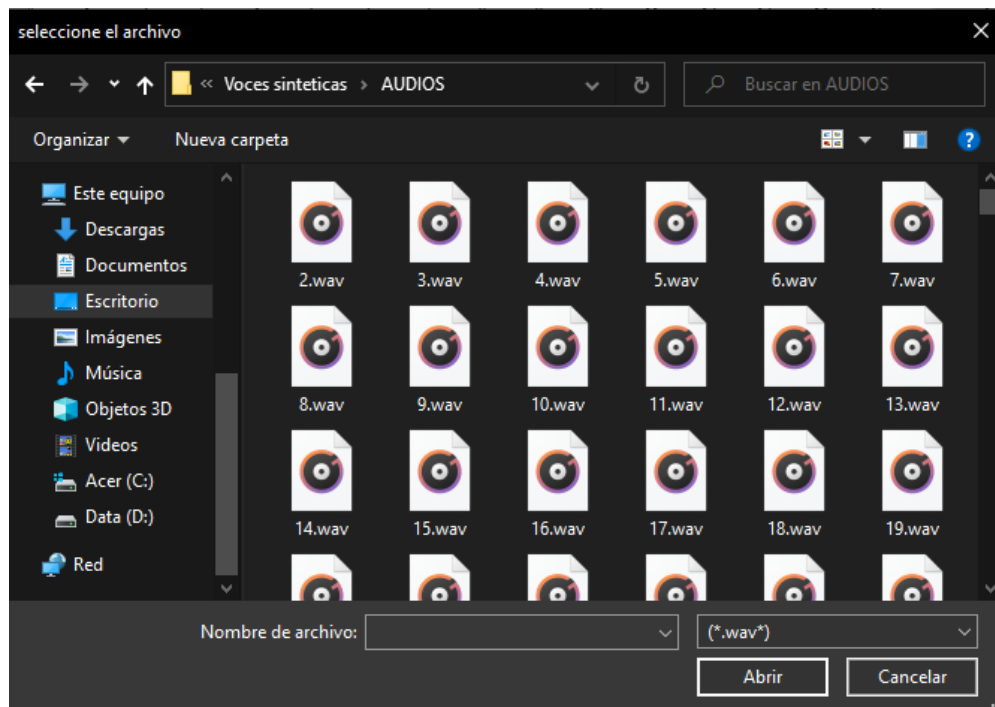


Figura 4: Explorador de archivos.

1.3.3 Análisis de voz

Luego de cargada la señal de audio, se procede a seleccionar el género de la persona de quien corresponde la señal de voz, y a continuación se desbloqueará automáticamente el botón “Analizar”, el cual el usuario deberá presionar para que el algoritmo comience a ejecutarse (Figura 5).

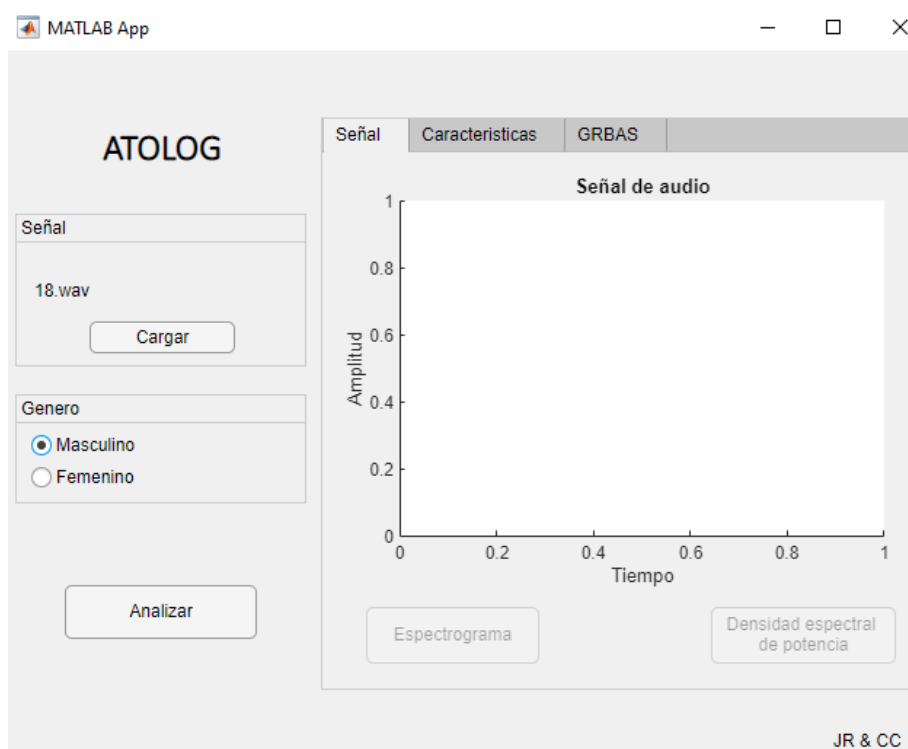


Figura 5: Puesta en marcha de la herramienta

1.3.4 Resultados

Al terminar el análisis, se puede visualizar la gráfica de la señal de audio ubicada en la parte lateral derecha de la primera pestaña (1) (Figura 6) y en la parte inferior (2), se ubican 2 botones que le permiten al usuario visualizar en una nueva ventana las gráficas del espectrograma (Figura 7a) o la densidad espectral de potencia (Figura 7b); en la segunda pestaña, se muestra las características acústicas, estadísticas y espectrales de la señal con sus respectivos valores en una tabla vertical (Figura 8); y por último, en la tercera pestaña se muestra el diagnóstico de la voz con base en la escala GRBAS, este diagnóstico se presenta de manera tabular y grafica (Figura 9).

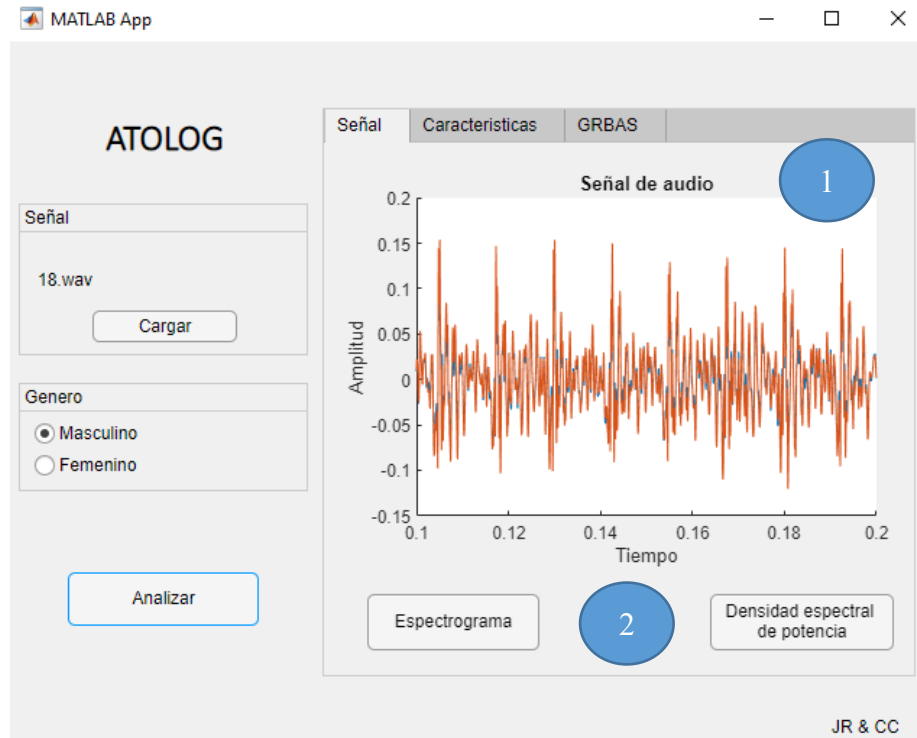


Figura 6: Análisis de la señal (Primera pestaña).

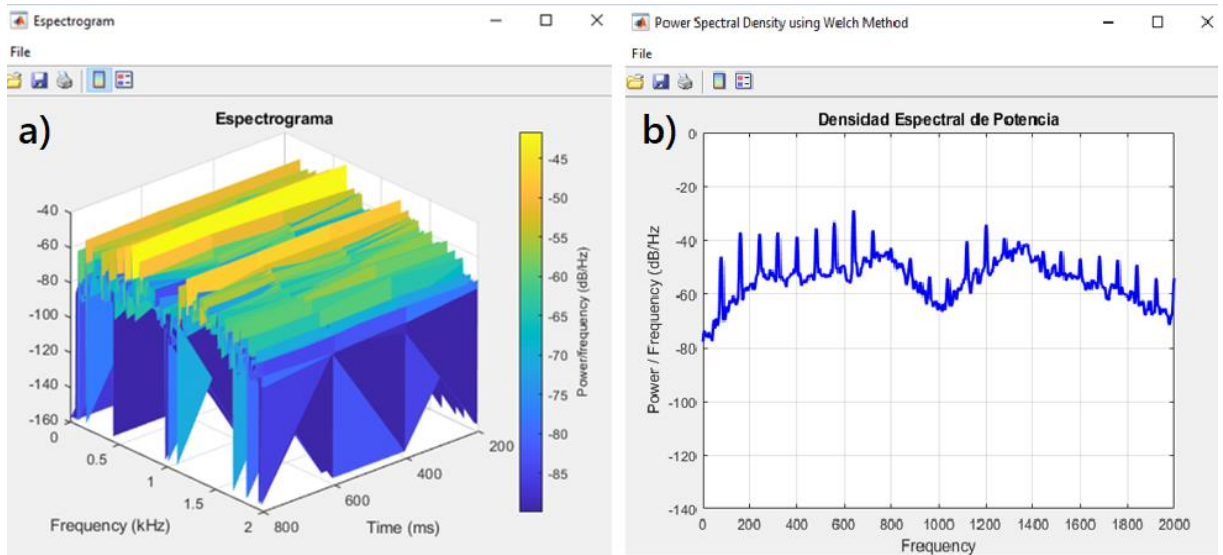


Figura 7: a) Espectrograma, b) Densidad espectral de potencia

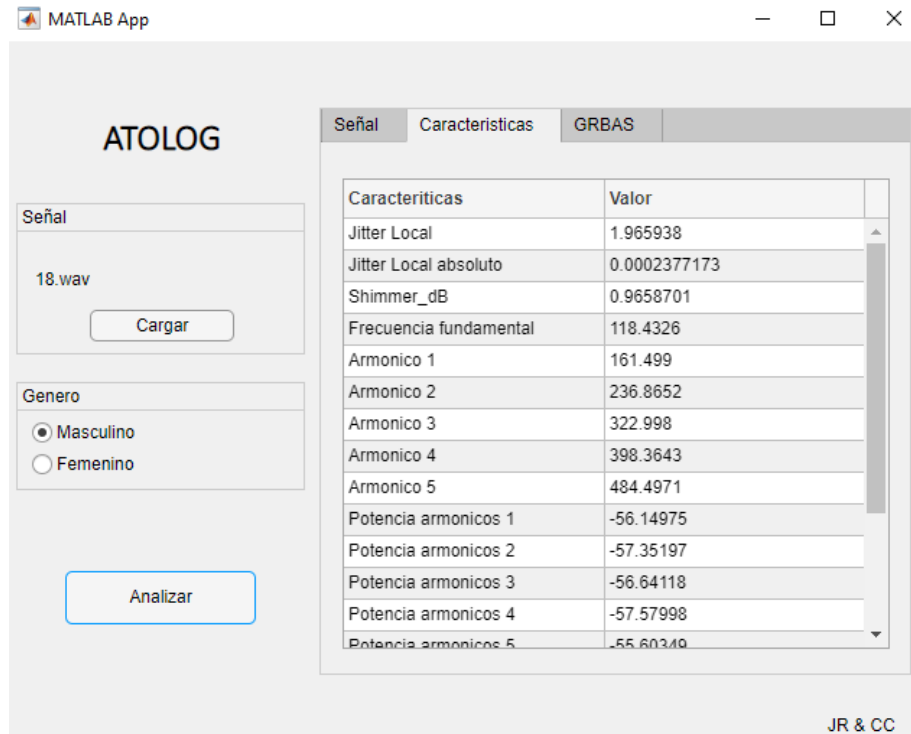


Figura 8: Análisis de la señal (Segunda pestaña).

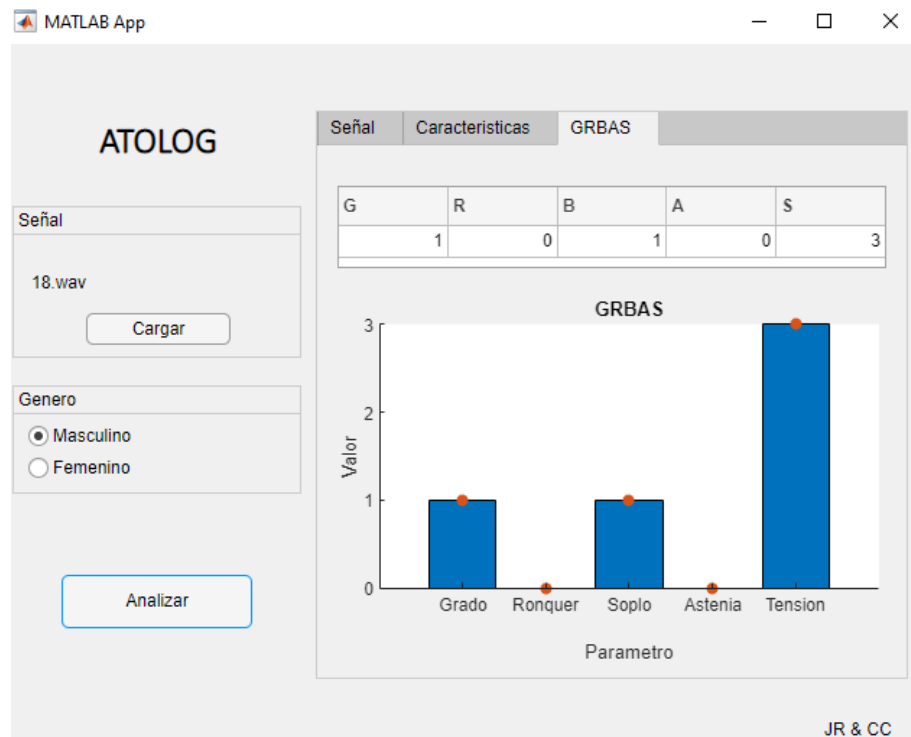


Figura 9: Análisis de la señal (Tercera pestaña).

Listado de resultados integrados de la escala GRBAS.

- Parámetros GRBAS

1. Normal = 0
2. Leve = 1
3. Moderado = 2
4. Severo = 3

Nota: Se debe aclarar que estos resultados no son definitivos, esta herramienta no reemplaza el diagnóstico de un logopeda, ya que su objetivo es proporcionar un apoyo para la interpretación de la señal de audio, debido a que el profesional de la salud es quien toma la decisión final sobre el diagnóstico.



**Desarrollo de una aplicación software para la
evaluación de la calidad vocal con
base en la escala GRBAS.**

Manual de Técnico

Versión: 1.0

Fecha: 24/Noviembre/2021

[Versión 1.0]

Queda prohibido cualquier tipo de explotación y, en particular, la reproducción, distribución, comunicación pública y/o transformación, total o parcial, por cualquier medio, de este documento sin el previo consentimiento expreso por la Universidad Antonio Nariño.

CONTENIDO

1. Sistema	3
1.1. Descripción	3
1.2. Alcance	3
2. Requerimientos técnicos	3
2.1. Hardware	3
3. Requisitos generales	3
3.1. Requerimientos de hardware	3
3.2. Requerimientos de software	4
4. Instalación	4
5. Configuración	4
6. Diseño de la arquitectura	5
7. Guía de uso Rápido	5
7.1. Uso	5
7.2. Variables o control del software	12
7.2.1. Valores fijos	12
8. Instalación de software	12
9. Recomendaciones	15

.....

1. Sistema

1.1. Descripción

Herramienta computacional para la evaluación de la calidad vocal con base en la escala GRBAS, que implementa métodos de procesamiento digital de señales, con el fin de extraer algunas de las características acústicas, espectrales y estadísticas de la voz, además de implementar sistemas de machine learning supervisado para realizar la clasificación del grado de disfonía y sus demás parámetros como: ronquera, soplosidad, aspereza y tensión

1.2. Alcance

Este documento está dirigido a los profesionales de la salud vocal (logopedas), como herramienta de apoyo en la evaluación de calidad vocal, ofreciendo a este personal una opinión de carácter objetivo frente a una señal de voz.

2. Requerimientos técnicos

2.1. Hardware

Se requiere un Computador con sistema operativo Windows x64.

3. Requisitos generales

3.1. Requerimientos de hardware

Procesador: Intel Core i3 de quinta generación, similar o superior

RAM: 4GB o 8GB (recomendada)

Espacio libre: 4GB

3.2. Requerimientos de software

Sistema operativo: Windows 7/8/8.1/10

Herramientas utilizadas:

- Matlab 2020b

4. Instalación

El instalador incluye los paquetes o archivos necesarios para ejecutar la aplicación sin que sea necesario tener el software Matlab instalado.

(se recomienda añadir las señales de audio a utilizar en el directorio del ejecutable .exe)

5. Configuración

No hay configuración general, la aplicación trabaja con parámetros ajustados para operar con señales que presenten cualquier frecuencia de muestreo, aunque se recomienda implementar una frecuencia de muestreo igual a 44100 Hz.

6. Diseño de la arquitectura

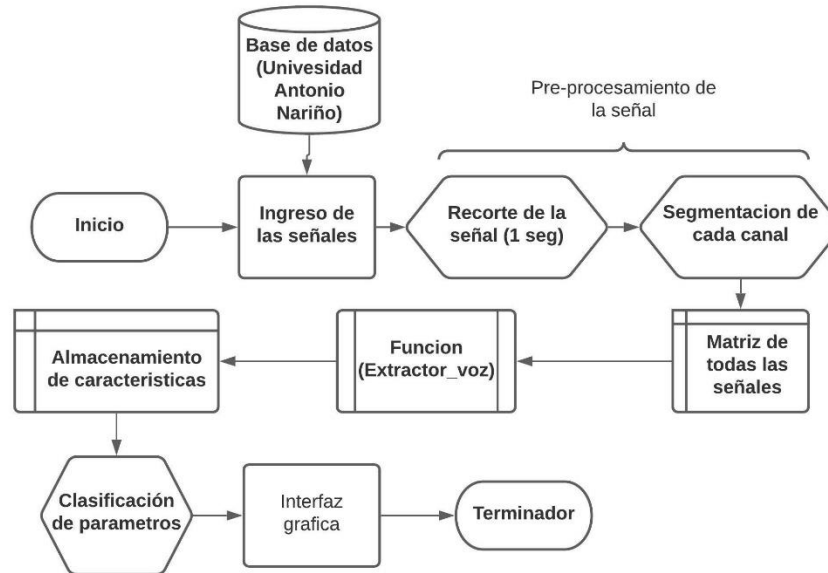


Figura 1 Diagrama de flujo.

7. Guía de uso Rápido

7.1. Uso

Una vez se tiene descargada e instalada la aplicación en cuestión para el ambiente Windows x64, se procede a iniciar la aplicación ATOLOG, la cual dará inicio a su interfaz principal como se ve en la Figura 2 con sus respectivas secciones y pestañas (Figura 3, Figura 4).

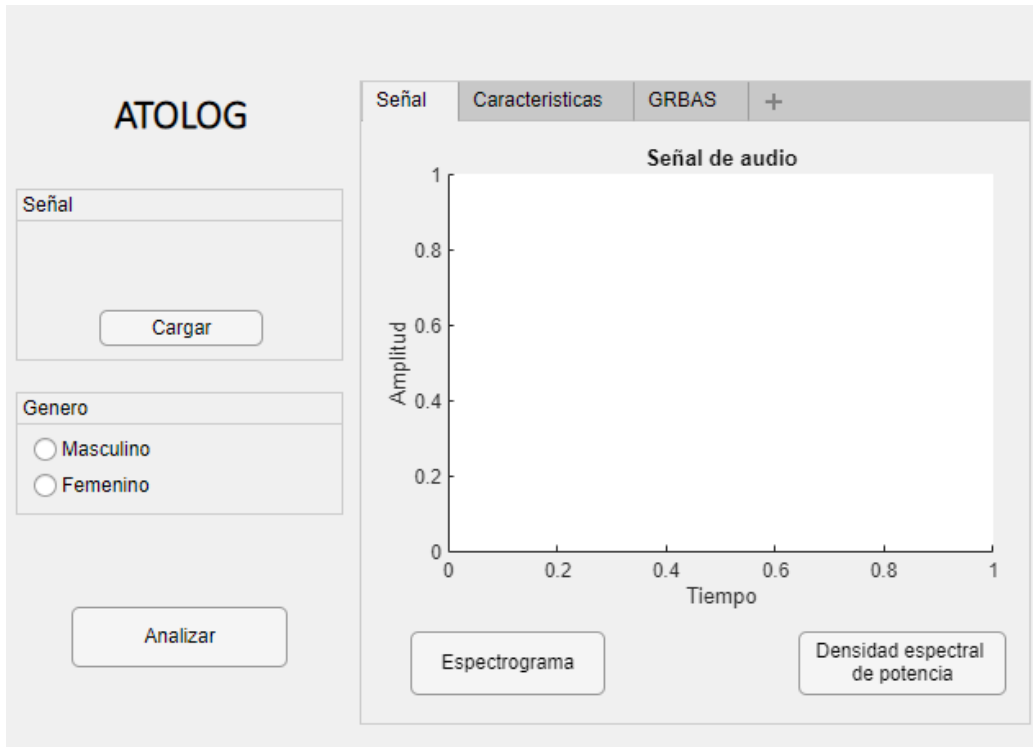


Figura 2. Interface principal ATOLOG (Pestaña 1).



Figura 3. Interface ATOLOG (Pestaña 2).

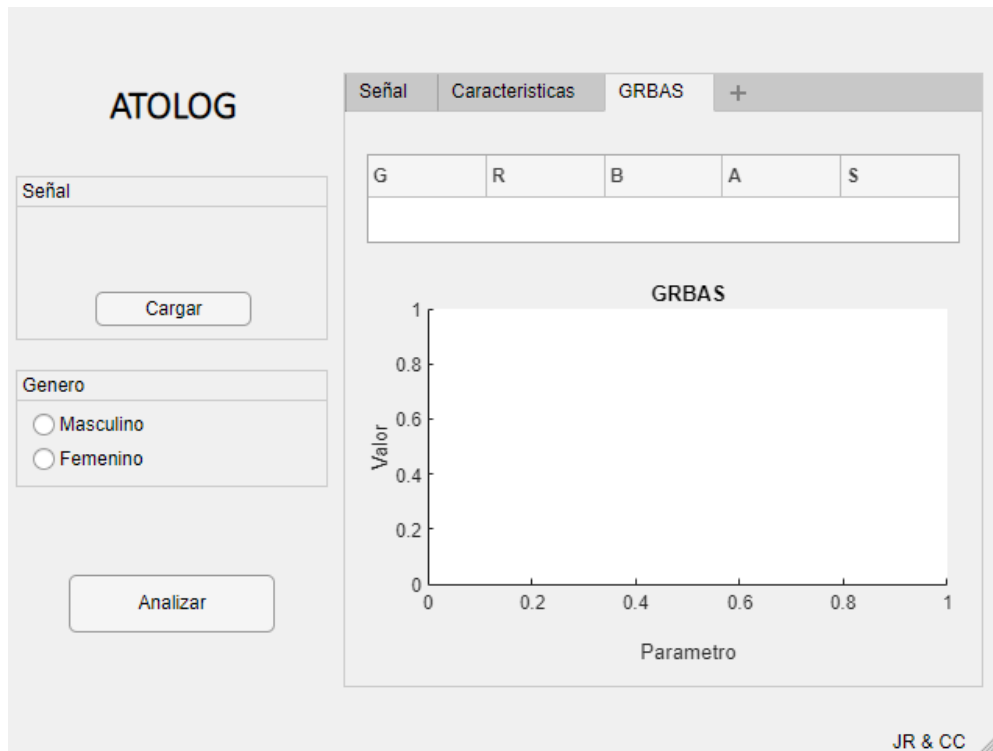


Figura 4. Interface ATOLOG (Pestaña 3).

Al presionar el botón “Cargar” ubicado en la parte central de la sección lateral izquierda de la interfaz, se despliega automáticamente una ventana de exploración de los archivos internos (Figura 5), donde el usuario puede seleccionar la señal de audio de interés que se va a ingresar y posteriormente analizar.

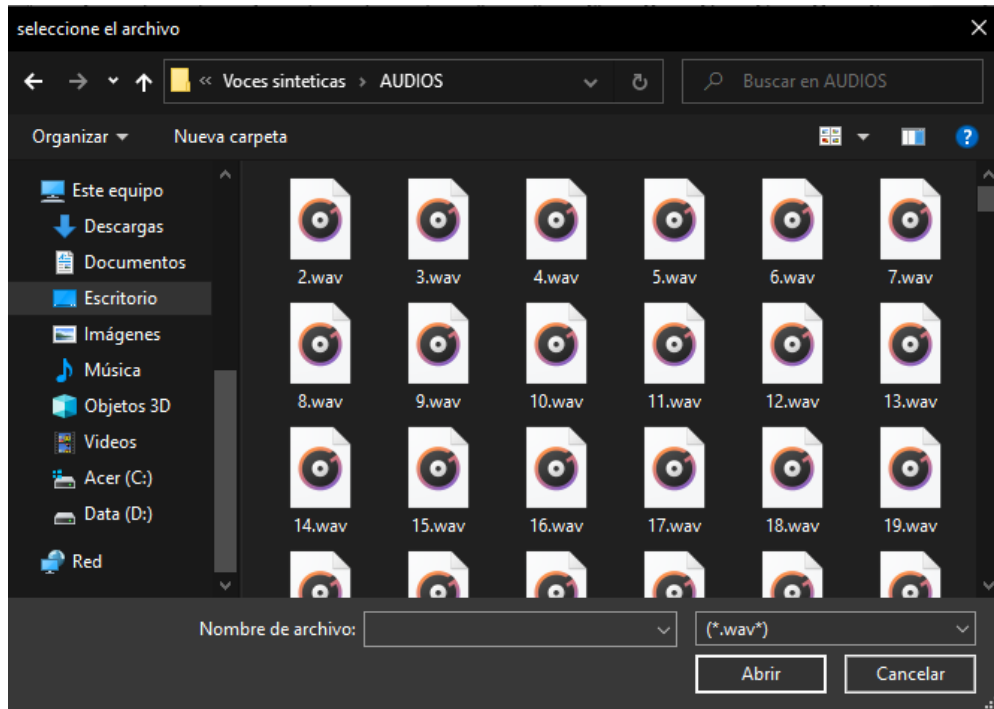


Figura 5. Explorador de archivos

Luego de cargada la señal de audio, se procede a seleccionar el género de la persona de quien corresponde la señal de voz, y a continuación se desbloqueará automáticamente el botón “Analizar”, el cual el usuario deberá presionar para que el algoritmo comience a ejecutarse (Figura 6).

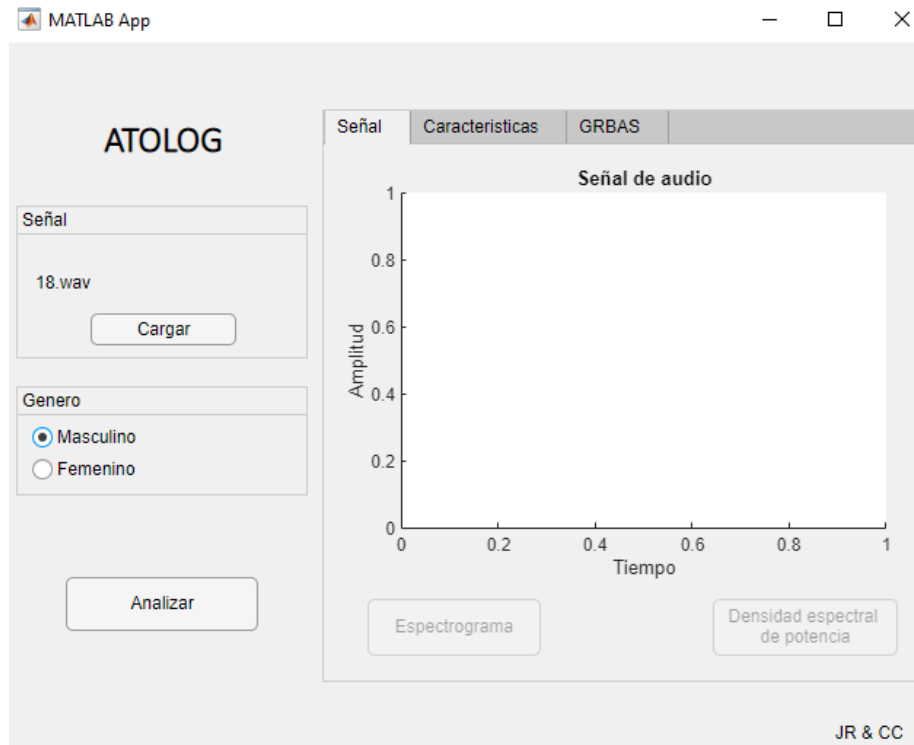


Figura 6. Puesta en marcha de la herramienta.

Al terminar el análisis, se puede visualizar la gráfica de la señal de audio ubicada en la parte lateral derecha de la primera pestaña (1) (Figura 7) y en la parte inferior (2), se ubican 2 botones que le permiten al usuario visualizar en una nueva ventana las gráficas del espectrograma (Figura 8a) o la densidad espectral de potencia (Figura 8b); en la segunda pestaña, se muestra las características acústicas, estadísticas y espectrales de la señal con sus respectivos valores en una tabla vertical (Figura 9); y por último, en la tercera pestaña se muestra el diagnóstico de la voz con base en la escala GRBAS, este diagnóstico se presenta de manera tabular y grafica (Figura 10).

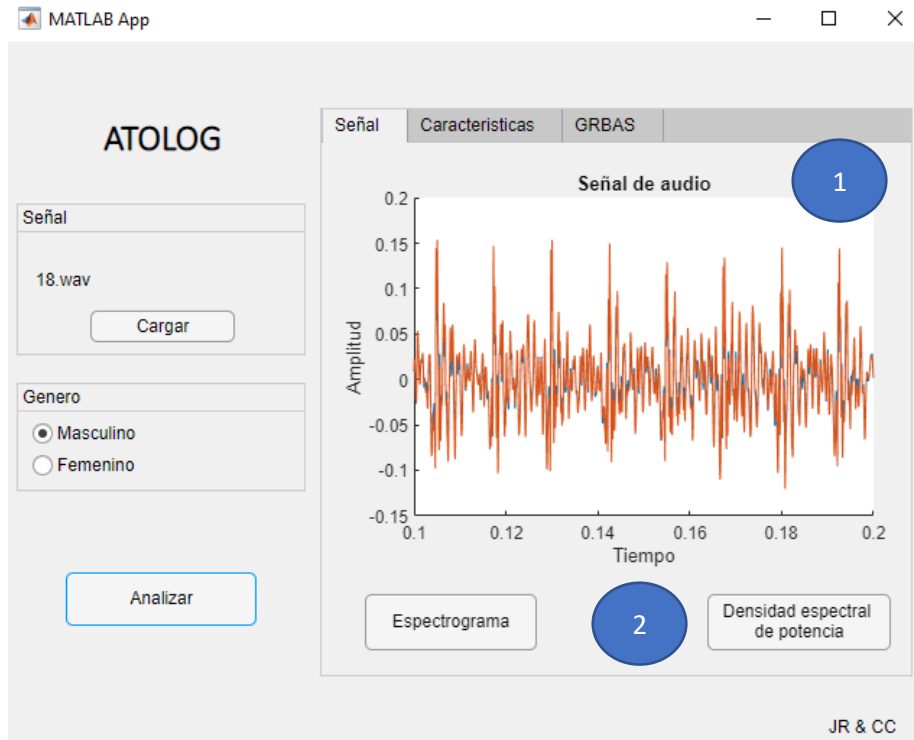


Figura 7 Análisis de la señal (Primera pestaña).

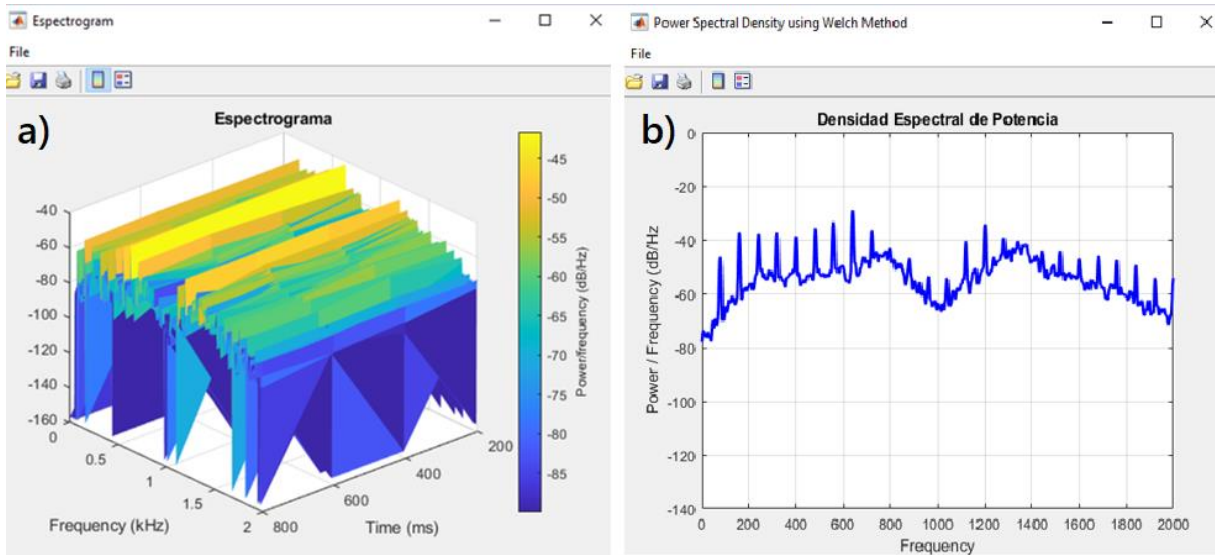


Figura 8 a) Espectrograma, b) Densidad espectral de potencia

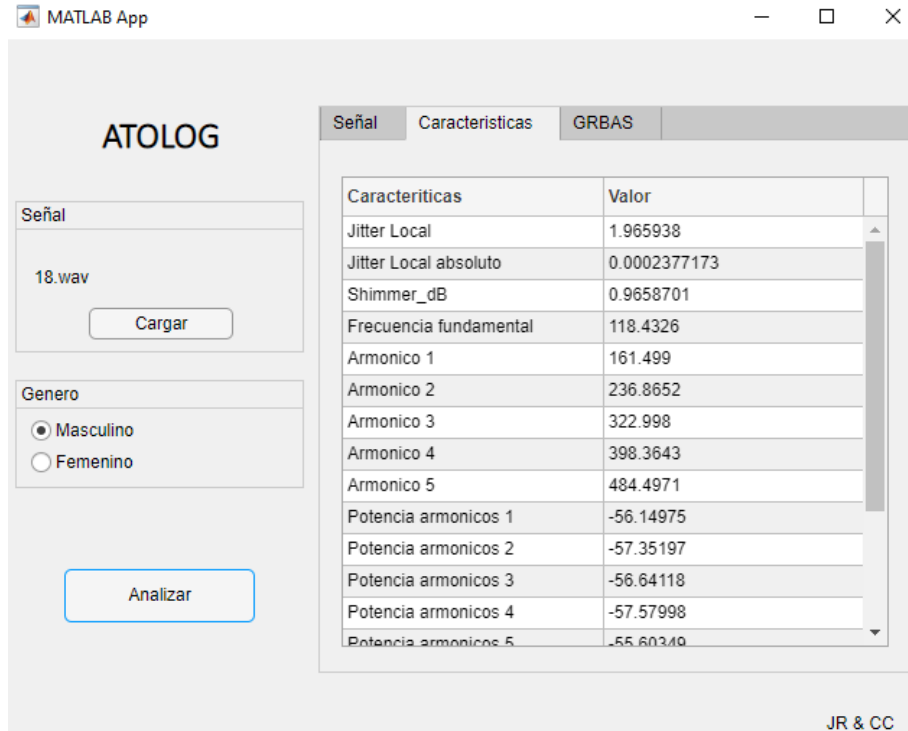


Figura 9 Análisis de la señal (Segunda pestaña).

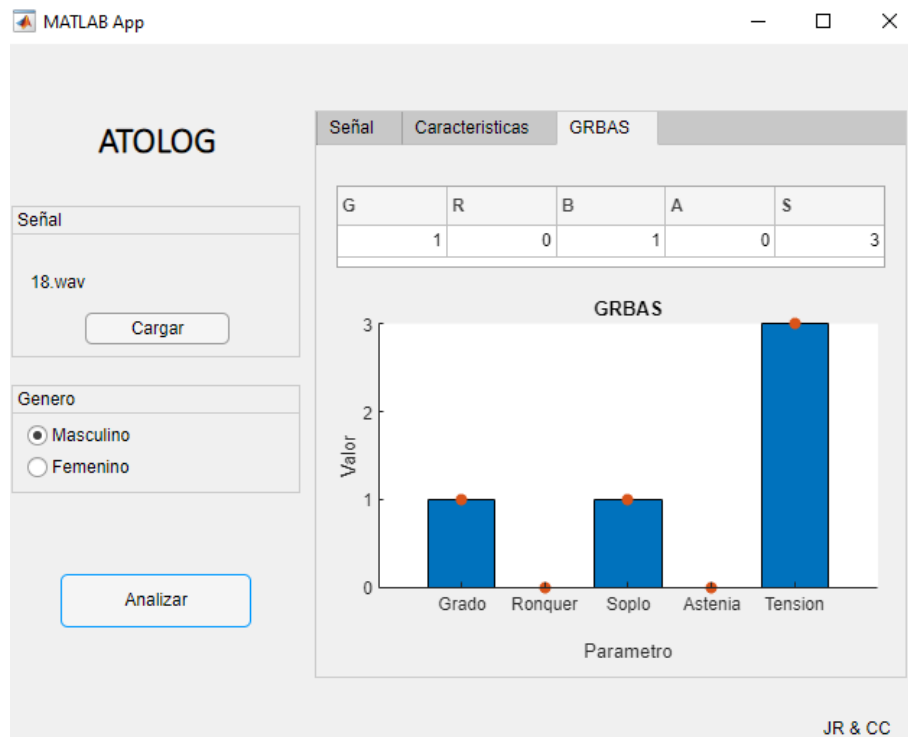


Figura 10 Análisis de la señal (Tercera pestaña).

7.2. Variables o control del software

7.2.1. Valores fijos

El software posee un valor estándar, el cual es un valor fijo e intrínseco del programa, este valor hace referencia a la duración mínima de la señal de audio, la cual equivale a 1,2 segundos, además, este periodo no debe poseer secciones de silencio o un elevado nivel de ruido ambiente.

8. Instalación de software

El proceso de instalación es sencillo, ya que el instalador solo requiere la ruta de instalación para generar la carpeta con el ejecutable, el proceso puede tardar dependiendo si se tienen o no funciones de Matlab preinstaladas en el equipo, ya que deriva su funcionamiento de algunos procesos de este software.

A continuación, se visualiza el seguimiento de una instalación limpia del software:

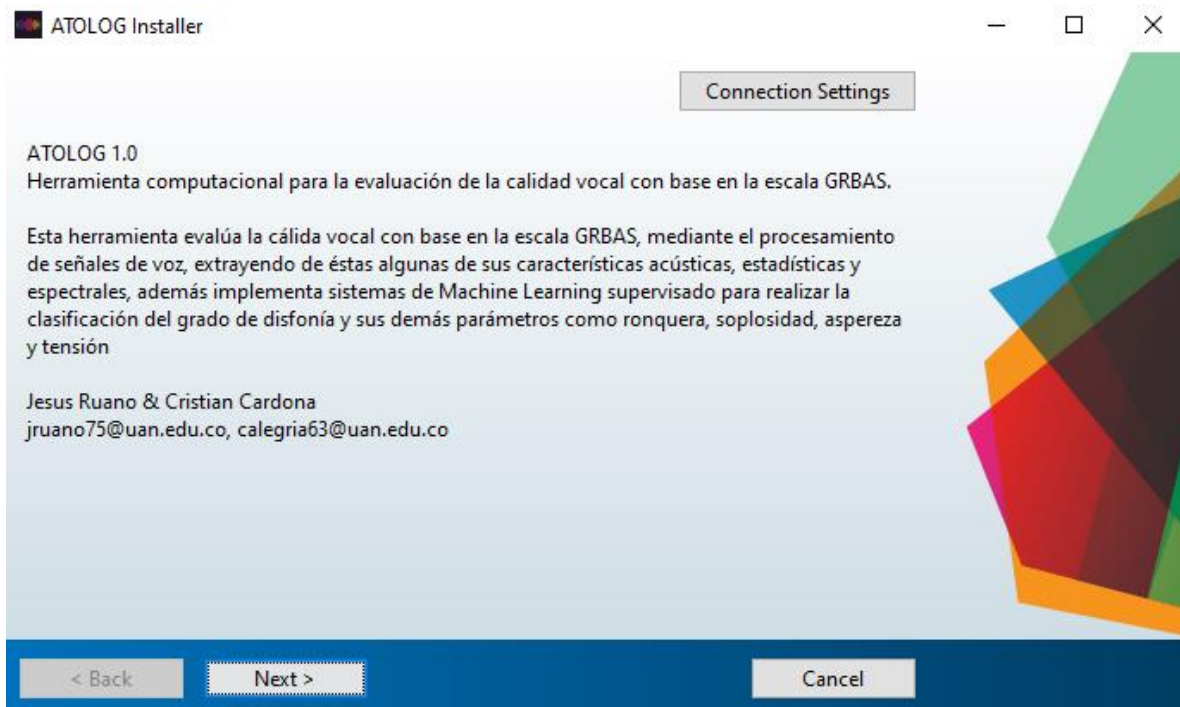


Figura 11 Proceso de instalación 1 de 7.

Figura 12 Proceso de instalación 2 de 7.

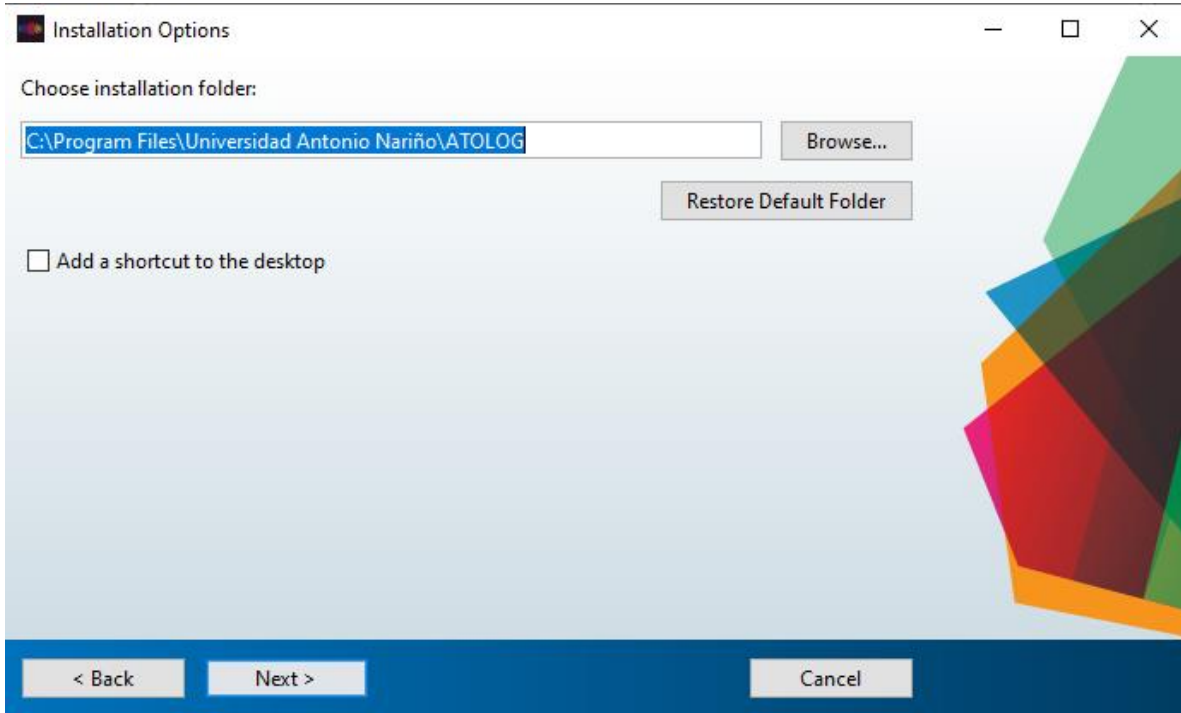


Figura 12 Proceso de instalación 3 de 7.

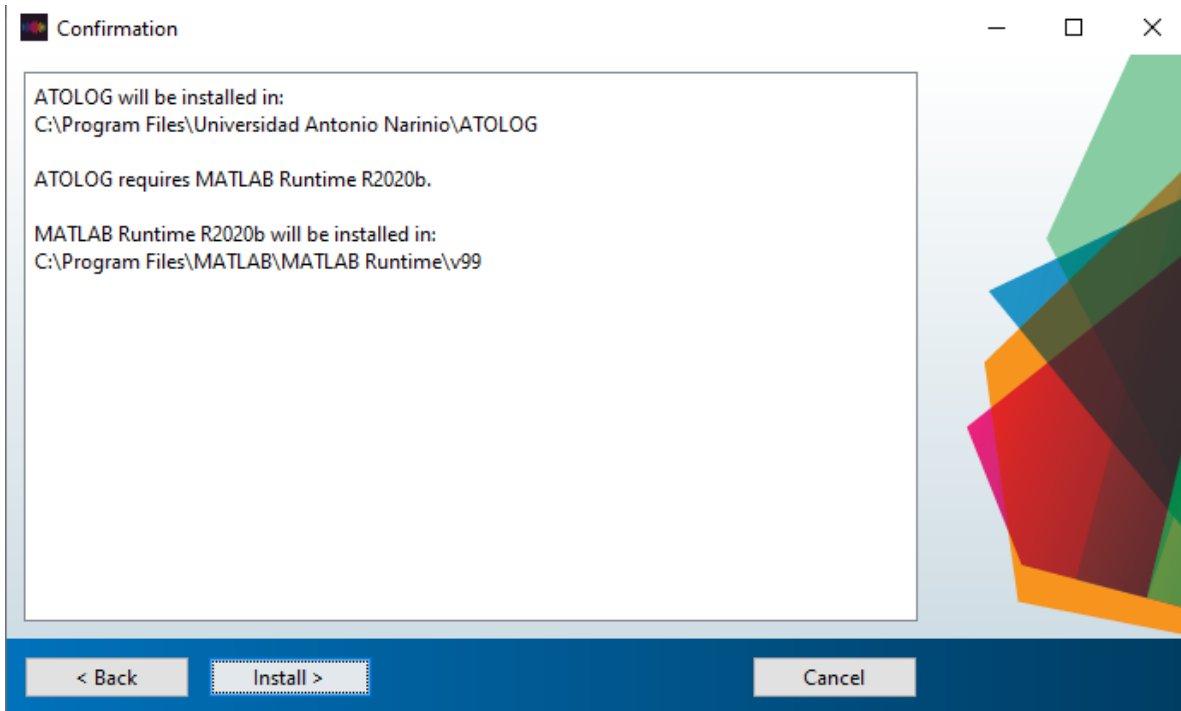


Figura 13 Proceso de instalación 5 de 7.

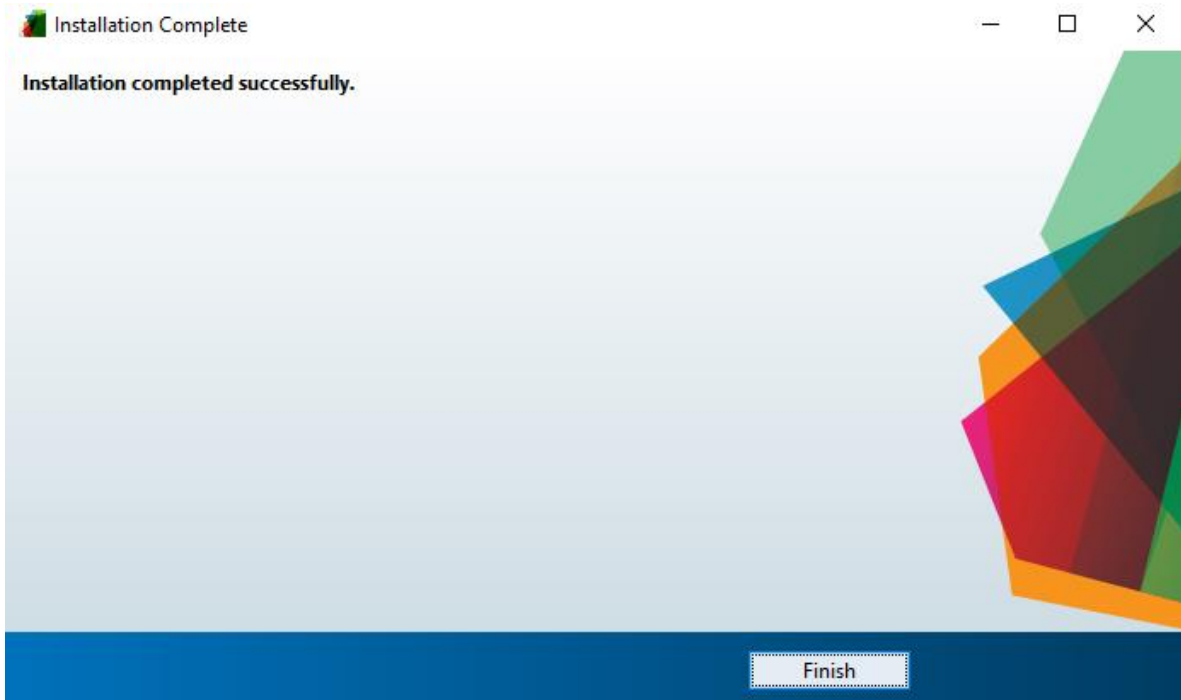


Figura 14 Proceso de instalación 6 de 7.

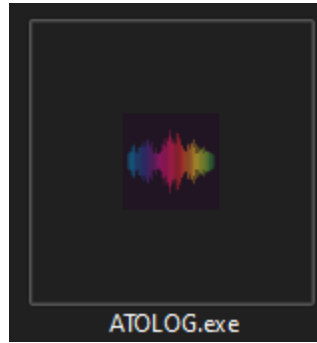


Figura 15 Proceso de instalación 7 de 7.

9. Recomendaciones

Para llevar a cabo la correcta implementación de la herramienta, es necesario que la señal del audio se encuentre en formato WAV, tenga una duración mínima de 1,2 segundos, además de no poseer secciones de silencio y un elevado nivel de ruido ambiental; También es necesario indicar de manera correcta el género de la persona a quien corresponde la señal de voz, ya que el algoritmo dependiendo del género seleccionado, determina de manera interna los clasificadores que deben ser utilizados para llevar a cabo una correcta evaluación y su posterior diagnóstico de la señal.