



Universidad de Concepción  
Dirección de Postgrado  
Facultad de Ingeniería  
Programa de Magíster en Ciencias de la Ingeniería  
con mención en Ingeniería Eléctrica

# **Screening Test para la Identificación del Queratocono**

Tesis para optar al grado de Magíster en Ciencias de la Ingeniería  
con mención en Ingeniería Eléctrica

FELIPE EDGARDO REBOLLEDO WEGNER  
CONCEPCIÓN-CHILE  
2018

Profesor Guía: Luis Arias Parada

UNIVERSIDAD DE CONCEPCIÓN  
Facultad de Ingeniería  
Departamento de Ingeniería Eléctrica

Profesor Guía:  
Dr. Luis E. Arias P.

# Screening Test para la Identificación del Queratocono



Felipe Edgardo Rebolledo Wegner

Tesis para optar al grado de  
Magíster en Ciencias de la Ingeniería con mención en Ingeniería Eléctrica

Octubre 2018

# Resumen

El queratocono es una enfermedad degenerativa de la córnea de difícil diagnóstico en primeras instancias, la cual genera una pérdida sustancial de agudeza visual. Los estudios se han enfocado en la inclusión de algoritmos de Machine Learning para incrementar la detección de la enfermedad, logrando excelentes resultados al ser implementado con datos de pacientes que se someten a exámenes especializados de vanguardia como la topografía corneal Pentacam o equivalentes. No obstante, la detección del queratocono sigue siendo engorrosa para aquellos individuos que por distintos motivos no se han estudiado con estos equipos oftalmológicos avanzados. Es por esto que el propósito del presente trabajo es ayudar a la detección de la patología en lugares donde los equipos de imágenes corneales no son de fácil acceso, desarrollando un screening test.

Para esto, se adquieren mediciones de 123 ojos (42 queratoconos y 81 muestras de control), compuestas de 23 características provenientes de los siguientes equipos oftalmológicos: autorefractómetro, queratómetro, paquímetro y topógrafo corneal. Se propone la utilización de las 7 características más simples y disponibles para entrenar un algoritmo de clasificación supervisada de tipo Support Vector Machine. El modelo propuesto consistirá en la afinación del algoritmo que logre los mejores resultados en términos de sensibilidad, especificidad y precisión. Esta afinación corresponde a la selección de un kernel apropiado y a la selección de valores de penalización 'C' y 'cw'. Luego, se evalúa la funcionalidad como screening test del modelo propuesto al comparar rendimientos con un modelo SVM de referencia que utiliza las 23 características adquiridas, incluyendo las características topográficas.

El mejor resultado para el modelo propuesto de 7 características fue logrado con un kernel lineal, 'C' = 10 y 'cw' = 1.9; alcanzando una sensibilidad de 0.9, especificidad de 0.96 y precisión de 0.93. Comparando rendimientos con el modelo de referencia, el modelo propuesto difiere en solo un 5% en sensibilidad, 3% en especificidad y 5% en precisión.

Los resultados validan la utilización del modelo propuesto como screening test, cumpliendo el objetivo propuesto. El trabajo futuro incluye validar el modelo con un número mayor de muestras, realizar clasificación multiclase para las distintas categorías de queratocono y combinar el modelo con variables categóricas de tamizaje.

# Tabla de Contenidos

<b>RESUMEN</b> .....	<b>III</b>
<b>LISTA DE TABLAS</b> .....	<b>V</b>
<b>LISTA DE FIGURAS</b> .....	<b>VI</b>
<b>CAPÍTULO 1. INTRODUCCIÓN</b> .....	<b>1</b>
1.1. ANTECEDENTES.....	1
1.2. HIPÓTESIS Y OBJETIVOS .....	5
1.2.1 <i>Hipótesis</i> .....	5
1.2.2 <i>Objetivos</i> .....	5
1.3. ALCANCES Y LIMITACIONES .....	6
1.4. TEMARIO.....	7
<b>CAPÍTULO 2. MARCO TEÓRICO</b> .....	<b>8</b>
2.1. CLASIFICACIÓN SUPERVISADA .....	8
2.2. SUPPORT VECTOR MACHINE .....	9
2.2.1 <i>Datos linealmente separables</i> .....	11
2.2.2 <i>Datos no linealmente separables</i> .....	11
2.3. VALIDACIÓN CRUZADA .....	14
<b>CAPÍTULO 3. MÉTODOS</b> .....	<b>16</b>
3.1. MATERIALES .....	16
3.1.1 <i>Base de Datos</i> .....	16
3.1.2 <i>Plataformas</i> .....	17
3.2. METODOLOGÍA.....	17
3.3. ADQUISICIÓN DE DATOS .....	18
3.3.1 <i>Datos de queratocono</i> .....	18
3.3.2 <i>Datos de control</i> .....	18
3.3.3 <i>Ficha única electrónica</i> .....	19
3.3.4 <i>Tratamientos quirúrgicos previos</i> .....	19
3.3.5 <i>Igualdad anatómica de córneas</i> .....	19
3.4. PREPROCESAMIENTO .....	20
3.4.1 <i>Imputación de datos faltantes</i> .....	20
3.4.2 <i>Pruebas de significancia</i> .....	20
3.4.3 <i>Normalización y escalamiento</i> .....	21
3.5. DESARROLLO DEL SISTEMA DE CLASIFICACIÓN .....	21
3.5.1 <i>Selección de características</i> .....	21
3.5.2 <i>Métricas de evaluación</i> .....	21
3.5.3 <i>Validación cruzada Leave One Out</i> .....	24
3.5.4 <i>Desbalance de Clases</i> .....	25
3.5.5 <i>Afinación de parámetros SVM y modelo propuesto</i> .....	26
3.6. EVALUACIÓN DEL MODELO PROPUESTO COMO SCREENING TEST .....	27
3.6.1 <i>Configuración de modelo SVM de referencia</i> .....	27
3.6.2 <i>Comparación de rendimientos</i> .....	27
<b>CAPÍTULO 4. RESULTADOS</b> .....	<b>28</b>
4.1. RESULTADOS DEL TEST DE SIGNIFICANCIA .....	28
4.2. REPRESENTACIÓN DEL DESBALANCE DE CLASES .....	29
4.3. AFINACIÓN DE PARÁMETROS SVM Y MODELO PROPUESTO .....	31
4.4. EVALUACIÓN DEL MODELO PROPUESTO COMO SCREENING TEST .....	34
<b>CAPÍTULO 5. CONCLUSIONES</b> .....	<b>37</b>
5.1. DISCUSIÓN .....	37
5.2. TRABAJO FUTURO .....	41
<b>BIBLIOGRAFÍA</b> .....	<b>42</b>
<b>ANEXO</b> .....	<b>46</b>

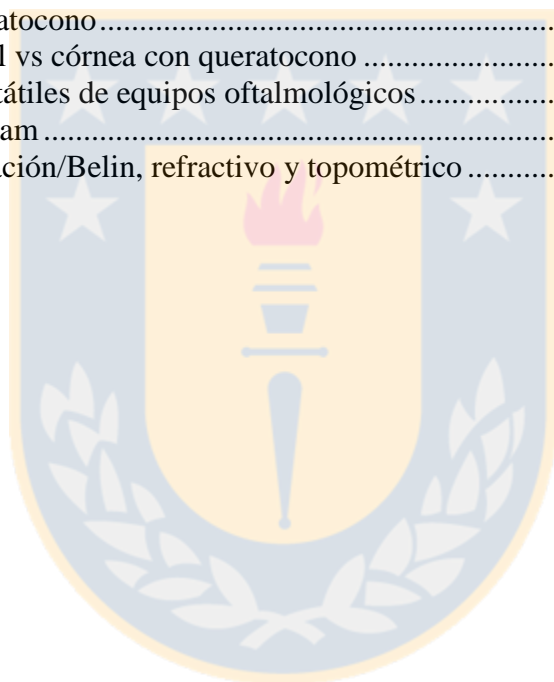
## Lista de Tablas

Tabla 3.1 Características oculares disponibles .....	16
Tabla 3.2 Rango numérico de parámetros SVM .....	26
Tabla 4.1 Test de significancia de características predictoras .....	28
Tabla 4.2 Rango numérico de parámetros SVM .....	31
Tabla 4.3 Afinación de parámetros SVM.....	32
Tabla 4.4 Evaluación del modelo como Screening Test .....	35



## Lista de Figuras

Fig. 2.1	Hiperplanos canónicos .....	9
Fig. 2.2	Margen $\rho$ y vectores de soporte.....	10
Fig. 2.3	Transformación de Kernel gaussiano de $\mathbb{R}^2$ a $\mathbb{R}^3$ .....	12
Fig. 2.4	Kernel Trick .....	12
Fig. 2.5	Variables de error $\xi_i$ .....	13
Fig. 2.6	Validación cruzada <i>4-folds</i> .....	15
Fig. 3.1	Matriz de confusión.....	22
Fig. 3.2	Validación cruzada LOO.....	24
Fig. 4.1	Costo diferenciado por clase .....	30
Fig. 4.2	Evolución en la afinación de parámetros .....	32
Fig. 4.3	Rendimiento al agregar características predictoras .....	35
Fig. A.1	Ojos con queratocono.....	46
Fig. A.2	Córnea normal vs córnea con queratocono .....	46
Fig. A.3	Versiones portátiles de equipos oftalmológicos.....	47
Fig. A.4	Oculus Pentacam .....	48
Fig. A.5	Mapa de elevación/Belin, refractivo y topométrico .....	48



# Capítulo 1. Introducción

---

## 1.1. Antecedentes

El Queratocono es una enfermedad degenerativa de la córnea, en donde esta sufre un adelgazamiento estructural paulatino, perdiendo progresivamente sus propiedades biomecánicas y asumiendo una forma cónica (ver anexo). Esta deformación genera una pérdida constante en la agudeza visual, la que en casos extremos puede llegar a la ceguera y a la necesidad de un trasplante de córnea [1]. Esta enfermedad aparece usualmente en la adolescencia y su prevalencia mundial ha sido estimada entre los 5 y 230 diagnósticos por cada 100.000 habitantes [2]. Sin embargo, la prevalencia varía entre razas y ubicaciones geográficas, aumentando hasta 10 veces en los casos de Israel e India [3], por lo que algunos estudios mencionan la posibilidad de su cálculo de prevalencia haya sido subestimado debido a dificultades en la identificación inicial de la enfermedad y debido a la cohorte de sujetos seleccionados para el estudio, por lo que se cree que esta podría ser mayor [3 - 6].

La detección temprana de esta patología es clave para evitar una mayor degeneración en la córnea. Sin embargo, los signos tempranos de la enfermedad pueden pasar desapercibidos o ser fácilmente atribuidos –tanto por el paciente como por el especialista– a otras patologías comunes como miopía y astigmatismo, desestimando complicaciones mayores [7]. La prescripción del examen de diagnóstico avanzado se realiza solo si el especialista logra identificar correctamente la enfermedad o sospecha de su presencia [8]. Este examen es una topografía corneal, la cual entrega un panorama completo sobre el estado de la córnea (ver anexo). En base a este examen, los médicos especialistas realizan el diagnóstico del queratocono [6], y según el grado de deformación de la córnea lo clasifican en: subclínico, leve, moderado o avanzado. No obstante, aún en esta última etapa del proceso de diagnóstico existen fallos en la detección, especialmente en la detección del queratocono subclínico. Debido a esto, la medicina oftalmológica ha puesto especial énfasis en desarrollar métodos y herramientas que disminuyan los errores en esta última etapa del diagnóstico. Para ello, distintos estudios han propuesto la integración de sistemas automáticos de detección mediante algoritmos de Machine Learning para aumentar la exactitud de los diagnósticos de esta patología.

En estos estudios mencionados, se han utilizado distintos algoritmos de machine learning en la identificación del queratocono, la mayoría aplicando algoritmos de clasificación supervisada. Los primeros estudios ocuparon algoritmos sencillos como en el caso del estudio realizado por Maeda en 1994 [9]. En este trabajo se implementó un clasificador simple de tipo árbol de decisión combinado con un discriminante lineal de 8 índices de la época, derivados de topografía corneal. Se utilizaron 200 córneas con distintos diagnósticos, consiguiendo una sensibilidad de 0.89 para los 28 casos de Qc destinados para evaluar el sistema y un accuracy general de clasificación de 0.96. El sistema combinado se pasó a llamar índice KCI Klyce/Maeda, fue integrado a los topógrafos como método de screening y sentó las bases para la inclusión de herramientas matemáticas y machine learning en la detección del queratocono. Posteriormente, un estudio realizado por Smolek en 1998 [10] utilizó redes neuronales para clasificar 300 córneas en 9 categorías distintas, introduciendo el concepto de clasificación multiclase. El objetivo de este estudio era compararse con los métodos de detección utilizados en la época como los índices K e I-S de Rabinowitz y el Índice KCI de Klyce/Maeda previamente mencionado. El clasificador mejoró los resultados en la detección de una de las clases utilizadas llamada “sospecha de queratocono”, siendo estadísticamente significativa su propuesta al compararlo con los índices anteriores (valor  $p < 0.001$ ), mientras que en el resto de las categorías no tuvo diferencias significativas en términos de sensibilidad.

Siguiendo esta línea investigativa, el estudio de Chastang en 2001 [11] utilizó 208 córneas derivadas de exámenes topográficos, las cuales fueron agrupadas en 8 clases distintas según el diagnóstico, aplicando nuevamente clasificación multiclase. Se entrenó un árbol de decisión que logró como resultado una sensibilidad de 0.88 y especificidad de 0.95 en la detección de la clase queratocono. Luego, en el estudio de Accardo en el 2002 [12] se utilizan nuevamente las redes neuronales, esta vez de forma más robusta. Se entrena y evalúa el sistema con un total de 396 mapas topográficos corneales obtenidos durante un periodo de 3 años, 120 de estos Qc. Una mayor cantidad de datos le permite a este modelo incrementar el rendimiento en la clasificación con respecto a otros estudios, alcanzando una sensibilidad de 0.91 y una especificidad de 0.96.



Posteriormente, el enfoque fue utilizar configuraciones más modernas y robustas de algoritmos de árbol de decisión, como se aprecia en los trabajos de Twa y su equipo investigativo de 2003, 2005 y 2007 [13, 14, 15]. En el primer trabajo [13], utilizan los datos en crudo de elevación de la córnea de un topógrafo Keratron Optikon. Desarrollan una modelación matemática valiéndose de los beneficios en representación de estructuras esféricas de los Polinomios de Zernike (PZ), entrenando y evaluando con 244 datos crudos de elevación de corneas –112 Qc– un clasificador de tipo árbol de decisión en 3 configuraciones distintas. El mejor resultado en la clasificación tuvo una sensibilidad de 0.87 y una especificidad de 0.95 al utilizar PZ de orden 8. Cabe destacar que la propuesta de este trabajo de modelar los datos en bruto de elevación de córnea mediante PZ, fue incorporada posteriormente en los softwares de topógrafos de vanguardia. En la segunda publicación [14] se repite el presente trabajo, ahora comparando la validez de este nuevo modelo con los índices de screening del Qc mencionados previamente en [10]. Finalmente, en el último estudio [15] se utiliza nuevamente el mismo modelo mediante PZ, comparando el rendimiento del árbol de decisión implementado en [13] con el rendimiento de los algoritmos Random Forest, Naive Bayes y Neural Networks. Se midió la exactitud en la clasificación, destacando el modelo de árbol de decisión propuesto en [13] con un accuracy de 0.9 y el modelo Naive Bayes con un accuracy de 0.87.

Actualmente, en un trabajo publicado por Souza el 2010 [16], se incorporan algoritmos de clasificación más modernos, robustos y versátiles. Se utilizan 318 mapas topográficos corneales para entrenar y validar un algoritmo Support Vector Machine (SVM), un Radial Basis Function Neural Network y un Multi-Layer Perceptron. El mejor resultado lo logra el algoritmo SVM con un error estándar en la identificación de 0.2 %. Arbeláez en 2012 [17] repite la utilización de SVM entrenando y evaluando el sistema con 2676 mapas corneales – 477 Qc– recopilados por más de 10 años por el National Eye Institute, EEUU. Este trabajo obtiene una sensibilidad de 0.95 al incluir mediciones previamente inaccesibles de la cara posterior de la córnea, mediante topógrafos modernos basados en el principio Scheimpflug.

Finalmente, los investigadores se han enfocado completamente en la identificación del queratocono subclínico, tal como se ve en el trabajo de Smajda de 2013 [18]. En esta investigación se incorporan nuevamente las medidas de la cara posterior de la córnea en 372 mapas corneales utilizados para entrenar y evaluar un Árbol de Decisión que logra una sensibilidad de 0.93 y especificidad de 0.97 en la clasificación de Qc subclínico. De forma similar, el trabajo de Hidalgo en 2014 [19] logra un accuracy de 0.926 en la detección de Qc subclínico implementando un algoritmo Naive Bayes, mientras que la detección del Qc subclínico con un algoritmo SVM logra un accuracy de 0.947.

La inclusión de estos algoritmos como herramienta de identificación en conjunto con las mejoras tecnológicas significativas de los topógrafos corneales de vanguardia ha incrementado sustancialmente la precisión del diagnóstico del queratocono. No obstante, incrementar la detección inicial de la enfermedad no ha tenido un énfasis investigativo similar ni ha tenido mayor evolución en los últimos años. Esto se explica dado que el objetivo de los estudios previamente mencionados se relaciona con evitar complicaciones post quirúrgicas en pacientes sometidos al procedimiento quirúrgico en apogeo LASIK, debido a que el queratocono es la principal contraindicación y pacientes sometidos a esta intervención con falsos negativos pueden desarrollar complicaciones serias, tal como la ectasia post LASIK [16, 20-22]. Además, en el contexto nacional, aparte de los ya mencionados problemas en la detección inicial, se suman factores socioeconómicos y geográficos que hacen que el acceso de la población general a los exámenes especializados necesarios para el diagnóstico sea limitado.

Por los motivos mencionados, el presente trabajo propone desarrollar un método de identificación del queratocono que pueda ayudar a la detección primaria de la enfermedad y que pueda ser utilizado en el futuro como screening test. Para ello, se configurará un algoritmo de aprendizaje automático de tipo Support Vector Machine. Este será entrenado utilizando un set de 123 datos de córneas previamente diagnosticadas y clasificadas como queratocono (42) o control (81), provenientes de exámenes realizados con autorefractómetro, queratómetro, paquímetro y topógrafo corneal Pentacam. Cada dato dispone de un total de 23 características, de las cuales se utilizarán 7 para construir el sistema. Las características elegidas son las que derivan de exámenes oftalmológicos

simples que pueden ser tomados en controles oftalmológicos de rutina, siendo estos exámenes más accesibles para los pacientes especialmente en lugares sin exámenes avanzados de imagenología corneal. Esto responde al objetivo de funcionalidad como test screening de la presente investigación. El proceso de evaluación medirá la sensibilidad, especificidad y precisión en la identificación. Finalmente, el proceso de validación como screening test del modelo construido con las 7 características predictoras comprende la comparación del desempeño en la predicción de estos 7 parámetros, versus el desempeño logrado por un modelo implementado con los 23 parámetros disponibles en el set de datos adquirido, siendo este último modelo el sistema de identificación de referencia.

## **1.2. Hipótesis y Objetivos**

### **1.2.1 Hipótesis**

- Se puede desarrollar un modelo de identificación automática del queratocono utilizando como predictor sólo características de alta disponibilidad no derivadas de exámenes complejos como la topografía corneal.

### **1.2.2 Objetivos**

- Desarrollar un modelo de clasificación supervisada de tipo Support Vector Machine que pueda identificar correctamente la patología construido sólo con características derivadas de exámenes regulares simples
- Desarrollar el núcleo de una herramienta médica de apoyo a la identificación del queratocono que pueda ser utilizado a futuro como un screening test para la patología, especialmente en lugares y personas con poco acceso a equipos avanzados de imagenología corneal.

### 1.3. Alcances y limitaciones

Las mediciones que se utilizarán para implementar el método de identificación propuesto son de difícil acceso, debido a los siguientes puntos:

- *Incidencia del queratocono:* Tal como se mencionó previamente, la prevalencia estimada de este trastorno es de 5 a 230 casos por cada 100.000 habitantes, lo que se implica una incidencia estimada relativamente baja para la población. Por lo mismo, la cantidad de datos disponibles es limitada, especialmente en el contexto nacional en donde la patología es poco conocida, los sistemas de información médica no están unificados y los datos no están al alcance de cualquier investigador.
- *Cantidad de centros oftalmológicos:* A lo largo del territorio nacional, la presencia de los centros oftalmológicos especializados en córnea es reducida, lo que se traduce en un panorama restringido a la hora de establecer vínculos con las instituciones para la recopilación de información.
- *Procesos de acreditación:* En el marco de la acreditación de prestadores del Ministerio de Salud de Chile, existen nuevas normativas con respecto a la información confidencial del paciente, las cuales pueden dificultar el proceso de obtención de datos, o la disposición a colaborar de los centros médicos.
- *Dispersión de pacientes:* Debido a la escasez de centros oftalmológicos especializados en córnea, la mayoría de los registros de pacientes se encuentran dispersos en las consultas oftalmológicas de distintos especialistas a lo largo del país.

## 1.4. Temario

El presente trabajo se compone de 5 capítulos principales: Introducción, Marco Teórico, Materiales y Métodos, Resultados y finalmente Conclusiones. En el segundo capítulo, se profundiza en los conceptos teóricos relacionados a sistemas de clasificación y en la formulación matemática del algoritmo de aprendizaje automático a utilizar en la investigación.

Posteriormente, en el tercer capítulo, se mencionan los materiales necesarios para implementar el modelo propuesto y la metodología seguida en el desarrollo del screening test, profundizando en las etapas del proceso de clasificación: adquisición de datos, pre procesamiento, métricas de evaluación, afinación de parámetros del algoritmo y validación del sistema como screening test.

Luego, en el cuarto capítulo se muestran los resultados en las distintas etapas del sistema de clasificación, poniendo énfasis en los ajustes de los parámetros del algoritmo utilizado y en los resultados comparativos entre el sistema propuesto y un sistema de referencia.

Finalmente, en el último capítulo se desarrollan las ideas principales que deja el desarrollo de la investigación, presentando conclusiones sobre el proceso de extracción de información, los resultados del método de clasificación y mencionar el trabajo futuro que puede complementar y expandir el estudio realizado hacia otros aspectos más tangibles y aplicables.

## Capítulo 2. Marco Teórico

---

El Machine Learning es un área de las ciencias de la computación que se dedica a la creación de algoritmos que permiten a un sistema aprender patrones sobre un conjunto de datos y poder realizar predicciones sobre este. Desde la década de los 80 se ha presentado un apogeo en este tópico, el cual es precursor de mecanismos de inteligencia artificial y también por su aplicación en diversas áreas como la detección de spam, detección de fraudes bancarios, reconocimiento de patrones en imágenes, scanner de retina, motores de búsqueda online y especialmente, en los últimos años, en la detección de diversas enfermedades; entre ellas el queratocono.

### 2.1. Clasificación supervisada

Los algoritmos de Machine Learning mencionados son utilizados, entre otras cosas, para la construcción de sistemas de clasificación supervisada [23]. Estos utilizan datos previamente etiquetados para entrenar un sistema que aprenda a reconocer distintos patrones y entregue como resultado una función que permita clasificar los datos. En el caso de clasificación binaria supervisada, la salida del sistema es una función binaria, mientras que la entrada puede ser de “n” características predictoras.

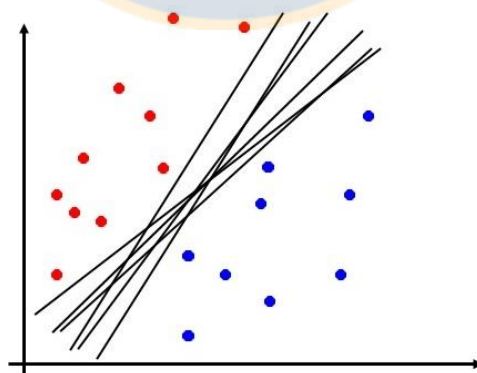
El conjunto de datos disponible para el aprendizaje automático del sistema se divide en un conjunto de entrenamiento y en un conjunto de prueba. Con el conjunto de entrenamiento; el sistema, de acuerdo a los algoritmos con los que trabaja, aprende patrones presentes en las características de entrada que permiten definir una función de decisión que separa los datos en las categorías requeridas. Con el conjunto de prueba, el modelo predice la pertenencia de los nuevos datos a alguna de las categorías estipuladas de acuerdo a la función de decisión construida mediante el conjunto de entrenamiento. Esto se evalúa con distintos indicadores que evidencian el rendimiento del sistema de predicción. Este conjunto de datos de prueba debe ser necesariamente distinto al conjunto de datos de entrenamiento.

Una correcta elección de características de entrada del sistema determina el rendimiento a posteriori del modelo de clasificación. Entre los algoritmos de aprendizaje automático ampliamente utilizados en clasificación supervisada destacan: Neural Networks, Perceptron, Naive Bayes, Bayesian Network, Decision Tree, Random Forest y Support Vector Machine.

## 2.2. Support Vector Machine

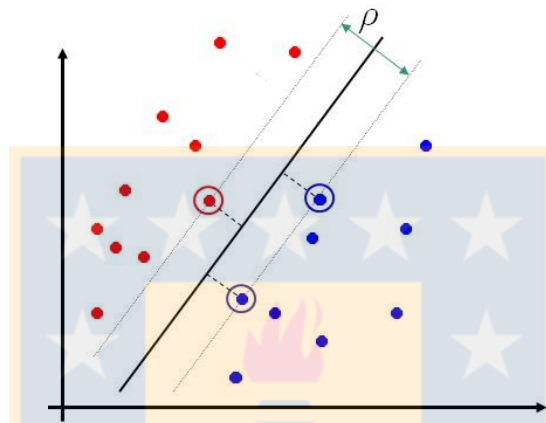
Los Support Vector Machines [24] son algoritmos de clasificación supervisada en donde el sistema, del total de los datos de entrenamiento, sólo utiliza un subconjunto de los datos de distinta clase más próximos entre sí para generar un plano de decisión, la cual se utiliza como criterio para lograr la clasificación de los datos.

En los sistemas de clasificación binaria, los datos pueden ser representados como puntos en el espacio de las características. De esta forma, la clasificación puede ser vista como una tarea de separar linealmente las clases en este espacio descrito. Si el conjunto de datos con el que se entrena el sistema es linealmente separable, existen infinitos hiperplanos que cumplen con separar correctamente las clases llamados hiperplanos canónicos, como se muestra en la Fig. 2.1, por lo cual es necesario encontrar la función de decisión óptima.



**Fig. 2.1 Hiperplanos canónicos [25]**

Los algoritmos SVM buscan construir un hiperplano de decisión que sea capaz de separar los datos de entrenamiento en dos clases y que a la vez maximice la distancia existente entre ellas [25]. Para esto, los SVM utilizan los datos más cercanos entre las categorías, los cuales se denominan vectores de soporte (support vector), y maximizan el margen “ $\rho$ ” el cual es la distancia entre estos vectores de soporte, tal como se muestra en la Fig. 2.2.



**Fig. 2.2 Margen  $\rho$  y vectores de soporte [25]**

En otras palabras, primeramente se busca encontrar un vector  $\mathbf{w}$  y un parámetro  $b$  que sean capaces de construir un hiperplano que separe linealmente el conjunto de datos  $\mathbf{X}$  en las dos categorías; clase “ $y_0$ ” o clase “ $y_1$ ”. A los planos que se pueden generar con las combinaciones de parámetros  $w$  y  $b$  que cumplen la condición de separar las clases, se les llama hiperplanos canónicos. En segundo lugar, de los hiperplanos canónicos, se selecciona el hiperplano que sea capaz de maximizar la distancia o margen  $\rho$  entre los vectores de soporte de cada clase, lo que es equivalente a buscar el hiperplano que minimice la norma “ $\|\mathbf{w}\|^2$ ”. Este hiperplano se transforma entonces en la función de decisión del algoritmo SVM, representada por el funcional “ $\Phi(\mathbf{w})$ ”, el cual dice por definición:

Buscar  $\mathbf{w}$  y  $b$ , tal que: (2.1)

$$\Phi(\mathbf{w}) = \|\mathbf{w}\|^2 = \mathbf{w}^T \mathbf{w}, \text{ sea minimizado, } \forall (x_i, y_i): y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1$$



Dependiendo de las características del conjunto de datos, pueden existir dos situaciones de estudio. La primera de ellas se refiere a cuando el conjunto de datos presenta la capacidad de ser separable linealmente, es decir, existe al menos un hiperplano que divide la totalidad de las muestras en las dos clases preestablecidas. Por otro lado, la segunda situación se refiere a cuando el conjunto de datos no es separable linealmente, es decir, no existe un hiperplano que divida la totalidad de las muestras en las dos clases. Estas situaciones se describen a continuación.

### **2.2.1 Datos linealmente separables**

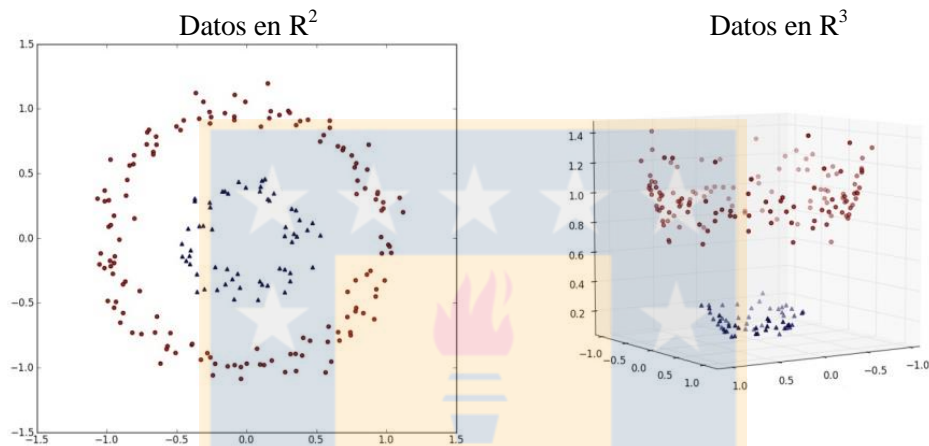
Para el caso de un conjunto de datos linealmente separable, tal como se ha explicado hasta ahora, se busca encontrar hiperplanos canónicos que separen las muestras de datos en las dos clases definidas, y de estos hiperplanos seleccionar el que maximice la distancia  $p$  entre los vectores de soporte. Si se encuentra una función de decisión que separe linealmente el set de datos, implica que no existen errores de clasificación en este escenario, por lo que el algoritmo tiene una exactitud del 100%. Desafortunadamente, en la realidad es poco común trabajar con un conjunto de datos de estas propiedades ideales, dado que frecuentemente existe un traslape entre las clases, por lo que el conjunto de datos no es linealmente separable.

### **2.2.2 Datos no linealmente separables**

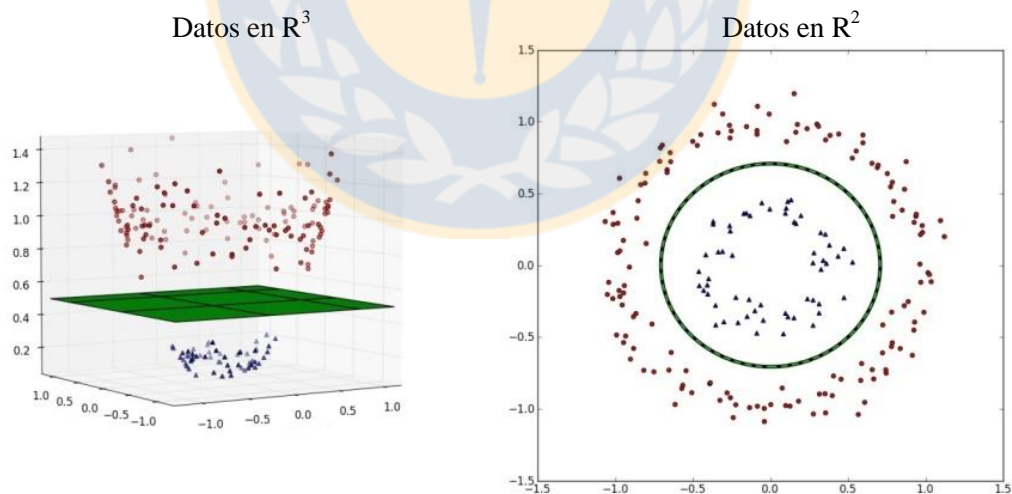
En el caso de datos no linealmente separable, es decir, que no existe un hiperplano que separe la totalidad de las muestras en las clases establecidas debido a un traslape entre las categorías, los algoritmos Support Vector Machine ocupan dos mecanismos para realizar la clasificación de la mejor forma posible.

El primer mecanismo [26] se denomina “*kernel trick*”. Este consiste en mapear los datos existentes del espacio de características original a un espacio de mayor dimensión. Con esto, se busca poder definir la función de decisión del algoritmo en una dimensión en donde las clases sí sean linealmente separables. Para ello, se necesita de una función de transformación llamada Kernel, la cual se utiliza para generar este espacio dimensional. Luego de obtener el hiperplano óptimo, se realiza la proyección inversa de la intersección entre el plano de decisión generado y la transformación kernel utilizada, obteniendo como resultado una función de decisión no lineal en el espacio de características original.

Las funciones Kernel a usar dependen de las propiedades del set de datos. Las transformaciones más comunes presentan kernels lineal, gaussiano, polinomial y exponencial. A modo de ejemplo, las Fig. 2.3 y 2.4 muestran la secuencia del mapeo de datos desde el espacio  $R^2$  a  $R^3$  aplicando una transformación kernel de tipo gaussiana. Luego, se muestra la obtención del hiperplano óptimo en  $R^3$  y la intersección entre el hiperplano generado y la función kernel utilizada, dando como resultado una función de clasificación no lineal.



**Fig. 2.3 Transformación de Kernel gaussiano de  $R^2$  a  $R^3$  [26]**



**Fig. 2.4 Kernel Trick [26]**

El segundo mecanismo [25] para realizar clasificación en conjuntos de datos no linealmente separables, se basa en la incorporación de variables de error “ $\xi_i$ ”, las cuales

definen un “soft margin”, en donde son permitidos los desaciertos en la clasificación, tal como se muestra en la Fig. 2.5. Es decir, se busca el hiperplano que maximice la separación entre clases, minimizando los errores de clasificación, tal como se muestra a continuación:

Buscar  $\mathbf{w}$  y  $b$ , tal que:

(2.2)

$$\Phi(\mathbf{w}) = \mathbf{w}^T \mathbf{w} + C \sum \xi_i, \text{ sea minimizado. } \forall (x_i, y_i): y_i(\mathbf{w}^T x_i + b) \geq 1$$

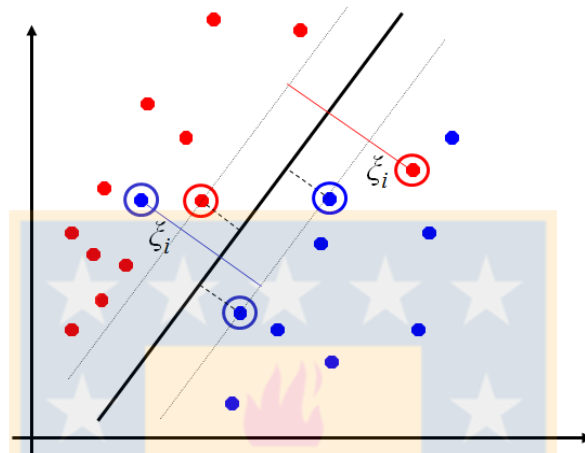


Fig. 2.5 Variables de error  $\xi_i$  [25]

El parámetro  $C$  representa un factor de costo o penalización, el cual permuta la exactitud de la clasificación con la complejidad de la función de decisión lograda, variando el *soft margin* a conveniencia.

La inclusión de las variables mencionadas permite a los modelos lineales proporcionar resultados que no son 100% exactos. Esto presenta la ventaja de disminuir considerable el costo computacional al evitar un uso en exceso de las dimensiones utilizadas por el kernel trick, lo que disminuye también el tiempo necesario para la solución computacional del problema. Además, existen casos en donde el problema no tiene solución en ninguna dimensión cercana al espacio original, y el algoritmo itera múltiples veces sin dar resultados, aumentando cada vez las dimensiones del espacio de características debido al kernel trick. Mediante la incorporación de estas variables, se puede llegar un resultado óptimo en dimensiones cercanas al espacio original al permitir controladamente los desaciertos.

La utilización conjunta de los dos mecanismos para realizar clasificación no lineal hace que los clasificadores SVM sean algoritmos versátiles y de buen desempeño en la clasificación de enfermedades, y particularmente en la detección del queratocono. Esto gracias a los distintos tipos de afinaciones existentes para lograr un buen performance y un buen ajuste del algoritmo al conjunto de datos. También, la utilización de vectores de soporte evita que la dispersión y la media de los datos de cada clase incidan en la función de decisión del clasificador, disminuyendo el efecto de datos ruidosos o dispersos y permite manejar de mejor forma conjuntos de datos desbalanceados. Todo esto puede resultar particularmente útil para el trabajo propuesto.

### 2.3. Validación cruzada

La validación cruzada [27] o “cross-validation” es un mecanismo de evaluación de modelos de clasificación que permite validar la consistencia de los datos utilizados para entrenar el modelo. Se utiliza para garantizar que los análisis de los resultados de la clasificación sean independientes de la partición de datos utilizada para entrenar y evaluar el modelo. Es decir, se garantiza que fenómenos o resultados que podrían ser aleatorios y vinculados a la partición del conjunto o set de datos no sean tomados como conclusiones relacionadas a los resultados de la clasificación per se, sino que se asocien a frutos del azar.

En las implementaciones de modelos de clasificación sin validación cruzada, la partición del conjunto de datos en un set de entrenamiento y un set de prueba se realiza según un porcentaje arbitrario elegido por el usuario de acuerdo a las características del problema y de las variables. Esta partición en dos subconjuntos de entrenamiento y de prueba se llama *hold out* y es el método más simple de validación de modelos estadísticos o de clasificación. En la actualidad, este método no es recomendado, ya que es muy sensible a posibles inconsistencias de los datos, a valores de ruido del sistema y a outliers. Además, en este tipo de validación se desaprovecha por completo el porcentaje de los datos utilizados para evaluar el modelo, ya que estos no pueden ser usados en la implementación del sistema, lo que genera una gran desventaja en los casos donde las muestras para implementar el sistema son limitadas.

En el caso de validación cruzada, se realizan distintas iteraciones, en donde cada una de ellas tiene un subconjunto de entrenamiento y prueba distinto. El tipo de validación cruzada más común es la validación cruzada  $k$ -folds. En esta, el conjunto de muestras se divide en  $k$  particiones, donde una de ellas se utiliza para evaluar el modelo y el resto se utiliza para realizar el entrenamiento. Este proceso se repite  $k$  veces con tal de utilizar la totalidad de los datos para entrenar y para evaluar el modelo, en instancias distintas cada vez, tal como muestra la Fig. 2.6. Los resultados finales son los promedios de las métricas obtenidas en cada iteración, los cuales son más representativos de la realidad que los modelos sin este tipo de validación. Esto genera resultados más robustos y menos sensibles a inconsistencias y outliers. La validación cruzada más utilizada es el método 10-folds.

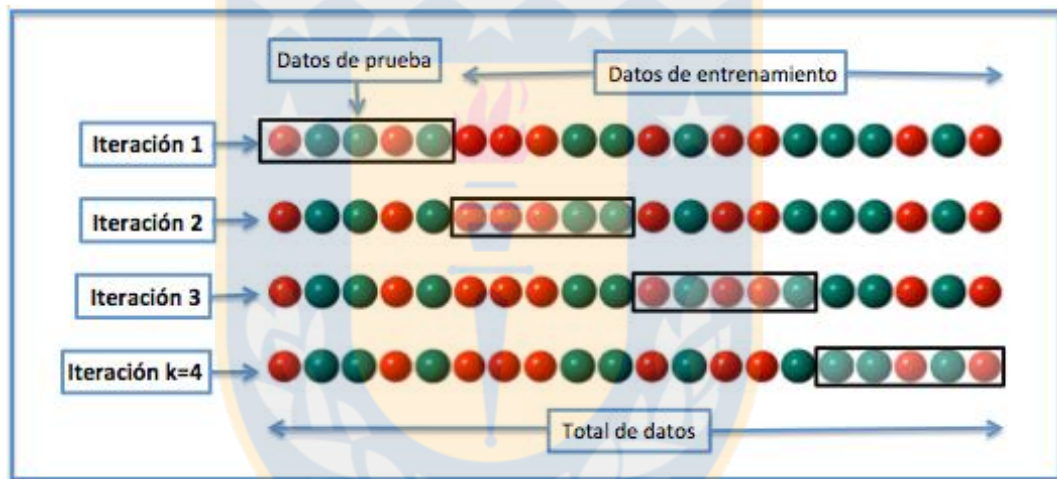


Fig. 2.6 Validación cruzada 4-folds [28]

Otros tipos de validación cruzada utilizados son: Random Permutation, Stratified K-fold, Leave P Out y Leave One Out. Este último tipo de validación representa un gasto computacional mayor, y usualmente no se recomienda utilizarlos en conjuntos de muestras demasiado extensas.

## Capítulo 3. Métodos

El presente estudio transversal y modelo de identificación del queratocono fue desarrollado en la Universidad de Concepción, Concepción, Chile; y la examinación de pacientes tomó lugar durante el periodo diciembre – 2016 a marzo – 2018 en las dependencias del “Instituto Oftalmológico Profesor Artensen”, Santiago, Chile.

### 3.1. Materiales

Los materiales e instrumentos utilizados para el desarrollo del modelo implementado se describen a continuación

#### 3.1.1 Base de Datos

Se examinaron un total de 140 córneas; 43 correspondientes a pacientes diagnosticados con queratocono y 97 datos de control, utilizando un autorefractómetro, queratómetro, paquímetro y topógrafo corneal. Los datos fueron previamente diagnosticados por los especialistas mediante examinación clínica y topografía corneal, y fueron clasificados entre pacientes con queratocono y pacientes normales para la clase de control. Por cada córnea examinada se recopilaban 28 mediciones distintas, las cuales se presentan en la Tabla 3.1

**Tabla 3.1 Características oculares disponibles. Fuente: Elaboración propia**

Métrica	Descripción	Métrica	Descripción
Sph	Esfera	RPIavg	Índice de progresión paquimétrico medio
Cyl	Cilindro	RPImin	Índice de progresión paquimétrico mínimo
Axis	Eje del cilindro	RPImax	Índice de progresión paquimétrico máximo
K1	Queratometría más plana	IVA	Índice de asimetría vertical
K2	Queratometría perpendicular	ISV	Índice de variación superficial
Km	Queratometría media	KI	Índice queratométrico
AP	Grosor corneal en el ápex	CKI	Índice queratométrico central
TP	Grosor corneal mínimo	IHD	Índice de descentración
DAP-TP	Distancia entre AP y TP	IHA	Índice de asimetría
Kmax	Queratometría máxima	CV	Volumen de la córnea
Rmin	Curvatura sagital mínima	ChV	Volumen de la cámara
Aemax	Elevación máxima anterior	AChD	Profundidad de la cámara anterior
Aedif	Diferencia entre elevación máxima y mínima en mapa anterior	Pemax	Elevación máxima posterior
D	Belin Ambrosio ectasia aumentada desviación total	Pedif	Diferencia entre elevación máxima y mínima en mapa posterior

Las medidas de Esfera, Cilindro, Axis, K1, K2 y Km son aportadas por un autorefractómetro-queratómetro; mientras que las medidas AP, TP y D.AP-TP son aportadas por la funcionalidad paquimétrica del topógrafo avanzado Pentacam. De todas formas, las medidas paquimétricas también pueden ser obtenidas de un paquímetro de ultrasonido. El resto de las medidas presentadas en la tabla 3.1 son otorgadas por Pentacam.

### **3.1.2 Plataformas**

Debido a restricciones impuestas por la fundación oftalmológica, para el tratamiento de estos datos fue necesario anotar manualmente las métricas de cada muestra desde los equipos oftalmológicos. Luego, estos datos fueron traspasados a una planilla Excel para ser cargados posteriormente al ambiente de desarrollo Spyder del software open source Python. En esta última plataforma es donde se realiza el pre procesamiento y se implementa el modelo de identificación de la patología. Este programa tiene la ventaja de poder utilizar bibliotecas especializadas para machine learning y tratamiento de datos.

## **3.2. Metodología**

Para desarrollar el método de identificación propuesto se definen los criterios de inclusión, de exclusión y la cantidad de datos necesarios para el modelo. Los datos se obtienen mediante exámenes realizados a pacientes de interés en las dependencias del centro oftalmológico especializado acordado. Estos datos son extraídos de forma unitaria y anónima desde el centro especializado y anotados manualmente en planillas de datos.

Con los datos ya recopilados, se realiza el pre procesamiento de la información. Se eliminan todos los casos perdidos de la clase de control y se imputan datos faltantes en la clase de queratocono si es que los hubiera. Se realiza una regresión mediante la cual se implementan pruebas de significancia para las características de los datos recopilados. Mediante los resultados de estas pruebas de significancia, se descartan las características que no sean estadísticamente significativas para esta cohorte de datos. Se realiza también un escalamiento, randomizado y normalizado de las muestras.

Posteriormente, se definen las características más disponibles en los controles oftalmológicos de rutina y con estas se construye el sistema de identificación del queratocono propuesto. Se seleccionan las métricas de evaluación pertinentes para la

afinación de parámetros del modelo y para la medición del desempeño en la clasificación de este. Se entrena y evalúa el sistema de forma recursiva utilizando validación cruzada y se afinan los parámetros del algoritmo para completar el modelo propuesto.

Luego, se realiza la evaluación del sistema como screening test. Para esto, se configura un nuevo algoritmo que utiliza las características seleccionadas originalmente, y añade todo el resto de características recopiladas por dato. Este modelo representa a la herramienta de identificación del queratocono más completa disponible, ya que agrega mediciones de topógrafos corneales de vanguardia y, para efectos del presente trabajo, se toma como sistema de identificación de referencia. Finalmente, se compara el rendimiento en la clasificación del sistema propuesto versus el sistema de referencia y se evalúa la funcionalidad del sistema propuesto como screening test.

### **3.3. Adquisición de datos**

Para la adquisición de muestras desde los centros oftalmológicos especializados se siguen los criterios que se explican a continuación.

#### **3.3.1 Datos de queratocono**

Para la obtención de datos de córneas que presenten la enfermedad, es necesario que estos sean diagnosticados previamente mediante examinación clínica y topografía corneal avanzada como Pentacam o equivalente, llevado a cabo por médicos especialistas en el centro oftalmológico acordado. Usualmente, estas mediciones se obtienen con mayor facilidad en candidatos a procedimientos correctivos del queratocono como anillos intraoculares o cross-linkin, previo a cualquier intervención.

#### **3.3.2 Datos de control**

Para corroborar que las muestras de control no tengan indicios de la patología estudiada, los pacientes deben tener un diagnóstico que respalde la ausencia de queratocono en sus córneas, en base a examinación clínica y a topografía corneal mediante el sistema Pentacam o equivalente por parte del centro oftalmológico especializado. Los datos de control se extraen de exámenes que se aplican a candidatos a cirugía refractiva LASIK, previo al tratamiento correctivo.



### **3.3.3 Ficha única electrónica**

Dado que los casos de queratocono son escasos y, generalmente, los pacientes que padecen la afección acuden a distintos especialistas, se prefieren centros oftalmológicos especializados que tengan un sistema de ficha electrónica. De esta forma, se unifica el formato utilizado en la anotación y medición de parámetros de la enfermedad, eliminando así posibles ambigüedades.

### **3.3.4 Tratamientos quirúrgicos previos**

Tanto las muestras de pacientes que poseen la afección como las mediciones de pacientes de control no deben presentar ningún tipo de tratamiento correctivo. En especial, no deben presentar intervenciones quirúrgicas. Las intervenciones quirúrgicas en la córnea, ya sea intervenciones correctivas de ametropías o debido a enfermedades de otra índole, modifican los parámetros normales de refracción de la córnea, la superficie, el espesor o su curvatura. Por estos motivos, los parámetros que se puedan obtener de estos pacientes no clasifican como pacientes de control o pacientes “normales”.

### **3.3.5 Igualdad anatómica de córneas**

Debido a la potencial participación de centros oftalmológicos extranjeros al inicio de la investigación, fue necesario establecer el criterio de similitud anatómica de córneas y evaluar que las estructuras oculares de las muestras que fueron utilizadas por los estudios extranjeros sean compatible con la estructura ocular de la población chilena. Esto debido a que las distintas razas presentan pequeñas variaciones en las medidas de sus corneas [4, 6, 29]. En general, las curvaturas de la córnea no varían significativamente entre las etnias, a excepción de las estructuras oculares de rasgos orientales marcados y puros, en donde las diferencias podrían llegar a ser significativas. Debido a esto, finalmente se acordó descartar conjuntos de datos provenientes de colaboraciones internacionales. Cabe destacar que el proceso de clasificación –que es lo que se desarrolla en el presente trabajo- no se vería afectado por la utilización de estos tipos de registros. No obstante, la aplicación como screening test en el contexto nacional podría verse afectado.

### 3.4. Preprocesamiento

Las 140 muestras provenientes de exámenes de autorefractómetro, queratómetro, paquímetro y topógrafo corneal fueron traspasadas manualmente desde los equipos oftalmológicos mencionados hacia una planilla Excel dispuesta. Esta planilla fue ingresada a la vez al ambiente de desarrollo Spyder para su posterior preprocesamiento.

#### 3.4.1 Imputación de datos faltantes

De las 140 muestras recopiladas en el proceso de adquisición, un total de 17 muestras presentaban pérdida de al menos una de las 28 características por muestra. De la clase de control se analizaron 15 casos con datos perdidos, los cuales fueron eliminados de la base de datos final. Por otra parte, existían 2 muestras con datos perdidos en la clase de queratocono. Estos datos fueron sometidos a un proceso de imputación de datos, el cual puede reemplazar los datos faltantes por un cálculo de la media, mediana o moda. Se seleccionó reemplazar los datos faltantes por la media de cada categoría respectiva.

#### 3.4.2 Pruebas de significancia

Para corroborar que las mediciones oftalmológicas sean estadísticamente significativas como predictoras del modelo de clasificación, se realiza un test de significancia mediante una regresión por mínimo cuadrados ordinarios, y calcula el  $p$  – value y  $f$  – value para verificar que cada una de las características del sistema esté aportando a la predicción de la variable de interés y en qué grado están aportando, respectivamente. Las características predictoras que presenten un valor  $p < 0.001$  son altamente significativas, las que presenten un valor  $p < 0.05$  son mediana o bajamente significativas, mientras que las características que presenten un valor  $p > 0.05$  no aportan como predictores y solo incrementan el ruido del sistema, siendo estas eliminadas del set de datos final. También, este test prueba la independencia entre las características predictoras, siendo eliminadas las características que sean linealmente dependientes de otra característica presente en el set de datos, y no necesariamente porque no sean informativas como características predictoras de forma singular. Un total de 23 características son seleccionadas como significativas y se muestran en el capítulo de resultados

### **3.4.3 Normalización y escalamiento**

Para evitar la aparición de ruido en el sistema de clasificación y para una mejor performance por parte del clasificador SVM, se escalan todas las características del set de datos a varianza unitaria, se le resta a cada conjunto de características sus medias con tal de centrar cada distribución en cero y se escalan los valores entre el rango  $[-1, 1]$ , siendo -1 y 1 los mínimos y máximos valores para cada característica respectivamente.

## **3.5. Desarrollo del sistema de clasificación**

El desarrollo del sistema de clasificación contempla definir las características oculares con las que se construirá el modelo de clasificación, seleccionar las métricas con las cuales se evaluará el desempeño en la clasificación, afinar los parámetros internos del algoritmo aplicando recursivamente validación cruzada y validar el modelo final comparándolo con un modelo de referencia.

### **3.5.1 Selección de características**

Para elegir las características predictoras del screening test, se privilegia la utilización de todas las mediciones disponibles que no tengan relación con topografía corneal avanzada. En este caso, los predictores que cumplen dicho criterio y que son estadísticamente significativos para nuestro set de datos son: esfera, astigmatismo, eje del astigmatismo, queratometría plana, queratometría perpendicular, grosor en el ápex y grosor mínimo corneal. Estas métricas derivan de los equipos oftalmológicos más accesibles como autorefractómetro, queratómetro y paquímetro. Cabe destacar, que los actuales sistemas de topografía corneal incluyen y a la vez mejoran los exámenes de queratometría y paquimetría convencionales.

### **3.5.2 Métricas de evaluación**

Las métricas utilizadas para evaluar el proceso de clasificación se desprenden de la Matriz de Confusión, la cual agrupa todas las mediciones necesarias para evaluar distintos ámbitos del rendimiento en sistemas de predicción.

La matriz de confusión es un cuadro específico que permite visualizar el desempeño de un clasificador en distintos aspectos y para las distintas categorías implementadas. Esta matriz, que se muestra en la Fig. 3.1, compara los resultados de la predicción de pertenencia de las muestras de evaluación en cada clase (queratocono y control), versus la etiqueta real de pertenencia de cada muestra. En clasificación binaria, esta matriz tiene 4 componentes:

- **True Positive (TP):** Cantidad de registros pertenecientes a la clase de interés –en este caso queratocono– que han sido correctamente clasificados por el modelo
- **True Negative (TN):** Cantidad de registros pertenecientes a la clase de control correctamente clasificados por el modelo.
- **False Positive (FP):** Muestras pertenecientes a la clase de control, erróneamente clasificados como casos de queratocono.
- **False Negative (FN):** Datos pertenecientes a la clase queratocono, clasificados deficientemente como casos de control.

		Predicción	
		Qc	Control
Real	Qc	<b>TP</b>	<b>FN</b>
	Control	<b>FP</b>	<b>TN</b>

**Fig. 3.1 Matriz de Confusión.** Fuente: Elaboración propia

Esta matriz permite obtener un vasto número de métricas del rendimiento de un clasificador. De estas métricas, para el presente estudio son de interés la sensibilidad, la especificidad y la precisión. Las primeras dos métricas permiten visualizar el desempeño del clasificador para cada clase por separado, mientras que la tercera relaciona ambas clases. Estas fórmulas permiten representar correctamente este modelo en particular, debido a que el desbalance de clases presente en el conjunto de datos afecta la significancia de otras métricas comúnmente utilizadas que no fueron seleccionadas para este trabajo.

### A. Sensibilidad

La sensibilidad (Se), también llamado *Recall* o *True Positive Rate*, es una medida que evalúa la capacidad predictiva del modelo en la clase de interés, relacionando las instancias correctamente clasificadas de queratocono (TP) con el total real de muestras de queratocono, compuesto por las mencionadas previamente (TP) más los casos de queratocono erróneamente clasificados (FN), tal como se muestra a continuación:

$$\frac{TP}{TP + FN} \quad (3.1)$$

### B. Especificidad

La Especificidad (Sp), también llamada *True Negative Rate*, es una medida que evalúa la capacidad predictiva del modelo en la clase de control, relacionando las instancias de control correctamente clasificadas, con la suma real de muestras de control, siguiendo la proporción que se muestra a continuación:

$$\frac{TN}{TN + FP} \quad (3.2)$$

### C. Precisión

Esta métrica relaciona los datos verdaderamente positivos con todos los datos clasificados como positivos, tal como se describe en la siguiente ecuación:

$$Pr = \frac{TP}{TP + FP} \quad (3.3)$$

Esta relación indica que la precisión en el proceso de clasificación del queratocono disminuye cuando se incrementa el número de errores en la clasificación de las muestras de control, los cuales son los falsos positivos (FP). De esta forma, se mide que los *trade-off* que se implementen en búsqueda de una mejor sensibilidad no afecten excesivamente en el rendimiento de esta arista del proceso de clasificación, evitando que el porcentaje de falsos positivos entregado por el algoritmo sea demasiado significativo.

### 3.5.3 Validación cruzada Leave One Out

El modelo Leave One Out (LOO) es un método de validación cruzada que se caracteriza por utilizar solo una muestra por iteración para evaluar el modelo, mientras que todo el resto del conjunto de datos es utilizado para entrenar el sistema, tal como muestra la Fig. 3.2. Esto es equivalente al modelo k-folds expuesto en el marco teórico, con  $k$  igual al número de muestras. Para implementar este modelo, necesariamente el número de iteraciones tiene que ser igual al número de muestras utilizadas. De esta forma, el conjunto de entrenamiento y de prueba del modelo son lo más grande posible, entrenando y a la vez evaluando con la totalidad de los datos. De este modo se puede utilizar, mediante iteración, la totalidad de muestras disponibles para entrenar el sistema, sin desperdiciar datos en el proceso de evaluación y atenuando los efectos de la moderada cantidad de registros disponibles en la presente investigación.

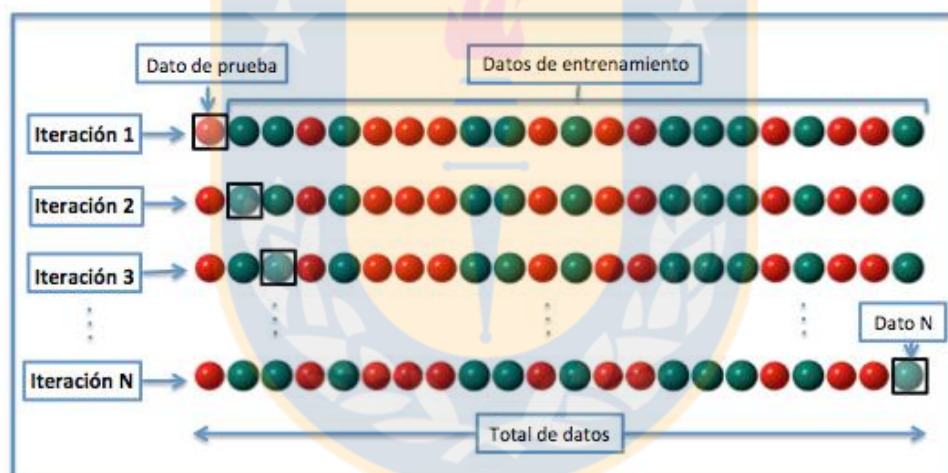


Fig. 3.2 Validación cruzada *LOO* [28]

La principal desventaja de este método de validación cruzada es que presenta un elevado costo computacional, ya que el número de iteraciones tiene que igualar el número de muestras utilizadas. Esta inconveniencia toma relevancia en set de datos numerosos, por lo que no es comúnmente utilizado. No obstante, debido a que la cantidad de muestras obtenidas para el presente estudio es moderada, la desventaja del costo computacional no es significativa.

### 3.5.4 Desbalance de clases

Existen distintas formas para disminuir el impacto del desbalance de clases en la clasificación. Procesos como oversampling o undersampling son comúnmente utilizados para igualar el peso de ambas clases frente al aprendizaje del algoritmo de clasificación. No obstante, para el caso del algoritmo SVM, existe una solución que radica en la afinación de parámetros de la formulación matemática interna del algoritmo. En esta, se busca minimizar los errores de clasificación, representados por las variables de error  $\xi_i$  que conforman el *soft margin* (mencionado en el marco teórico), los cuales tienen un parámetro de penalización  $C$ , tanto para errores de clasificación de la clase  $y_0$ , como para los desaciertos en la identificación de la clase  $y_1$ , como se ve en la ecuación (3.4).

$$\text{Min } \Phi(\mathbf{w}) = \mathbf{w}^T \mathbf{w} + C \sum \xi_i \quad (3.4)$$

El problema descrito se puede solucionar estableciendo dos valores distintos de la variable  $C$  para cada tipo de error;  $C_0$  y  $C_1$ . Esto permite penalizar de forma diferenciada los errores de clasificación para cada clase, moviendo el *soft margin* y permitiendo incluso aumentar los desaciertos en la clase mayoritaria, con tal de disminuir los errores de predicción en la clase de interés, lo que se denomina *trade-off*. De esta forma, la ecuación (3.4), se transforma en la ecuación (3.5).

$$\text{Min } \Phi(\mathbf{w}) = \mathbf{w}^T \mathbf{w} + C_0 \sum \xi_0 + C_1 \sum \xi_1 \quad (3.5)$$

Eligiendo un valor  $C_0 > C_1$ , se penaliza más un error de clasificación de la clase de interés, resultando en la elección de un hiperplano que minimice los desaciertos en la identificación de esta clase. De esta forma se puede, en parte, disminuir los errores de clasificación de la enfermedad de interés relacionados al desbalance de clases. Esto toma especial relevancia en los kernel de tipo lineal, los que se ven comparativamente más afectados a los efectos de la penalización diferenciada por su simplicidad de cómputo. Este tipo de ajuste de parámetros internos del clasificador SVM se implementa en el presente trabajo con el parámetro “C” de penalización, y con el parámetro “cw” que representa la razón  $C_1/C_0$ .

### 3.5.5 Afinación de parámetros SVM y modelo propuesto

Con las siete características seleccionadas para la construcción del modelo propuesto, se procede a entrenar y evaluar recursivamente mediante la validación cruzada LOO implementada. Este proceso tiene como propósito la afinación de distintos parámetros del algoritmo SVM para la obtención del mejor performance en la identificación de la clase de interés, modificando parámetros internos y evitando un sobreajuste.

Los cuatro métodos para afinar manualmente el algoritmo SVM ya han sido mencionados previamente. Estos métodos son: la elección de una función kernel para que el algoritmo pueda implementar el “kernel trick” mencionado en el marco teórico, el parámetro “C” de penalización de errores de clasificación, el parámetro “cw” para tratar con el desbalance de clases y un parámetro gamma que es exclusivo para algunas funciones kernel, el cual define la influencia de cada muestra en la función de decisión del clasificador. Este último parámetro se utiliza en funciones kernel “Radial Basis Function” (rbf), Sigmoide y Polinomial, entre otros. Los kernel seleccionados para la implementación del modelo propuesto son: función kernel lineal y función kernel rbf. El rango de estos factores se muestra en la tabla 3.2 a continuación.

**Tabla 3.2 Rango numérico de parámetros SVM. Fuente: Elaboración propia**

Parámetro	Rango
C	[1, 10, 100]
cw	[1.0 – 3.0]
$\gamma$	[0.001 – 0.009, 0.01 – 0.09, 0.1 – 0.9, 1 - 10]

El rango del parámetro cw fue restringido a 21 posibles valores cercanos a la proporción real de 1.9 de ambas clases (81 control, 42 Qc), limitando así el *trade-off* a favor de la sensibilidad y evitando un sobreajuste al conjunto de datos utilizado. El parámetro “C” tiene 3 valores, mientras que el parámetro gamma fue definido con 37 posibles valores. De acuerdo con esto, el proceso de búsqueda exhaustiva considera 2394 combinaciones posibles. Por otra parte, se realizó una búsqueda manual considerando de forma independiente el efecto de la afinación de cada parámetro en los resultados de la clasificación.



Posteriormente, se realiza la búsqueda exhaustiva de la combinación de valores que otorguen los mejores resultados en términos de las métricas seleccionadas. Finalmente, el modelo se construye utilizando los parámetros definidos en la búsqueda y esta configuración representa nuestro modelo propuesto de screening test para la identificación del queratocono.

### **3.6. Evaluación del modelo propuesto como Screening Test**

Para evaluar la funcionalidad como screening test del modelo de 7 características propuesto, se configura un segundo algoritmo SVM como referencia que utilice todas las características disponibles, con el propósito de comparar rendimientos.

#### **3.6.1 Configuración de modelo SVM de referencia**

Este segundo algoritmo implementado utiliza las 23 características disponibles por cada muestra en el conjunto de datos recuperado; las 7 características utilizadas por el modelo anterior más 16 características provenientes de exámenes realizados con un topógrafo corneal avanzado. Estas 16 características (definiciones en tabla 3.1) son: Kmax, CV, RPIavg, RPImin, RPImax, AEmax, PEmax, AEdif, PEdif, D, IVA, ISV, KI, CKI, IHD, IHA. Se utiliza la misma afinación de parámetros definida en la construcción del screening test propuesta. Este modelo representa la mejor herramienta de identificación del queratocono de acuerdo al conjunto de datos disponible para la presente investigación, y será considerada como modelo de referencia.

#### **3.6.2 Comparación de rendimientos**

Finalmente, se comparan los desempeños del modelo de 7 características propuesto como screening test versus el modelo de referencia de 23 características previamente mencionado. Se evalúan y comparan los rendimientos de acuerdo a la sensibilidad, especificidad y precisión en la clasificación. Adicionalmente, se analiza el impacto en los resultados al añadir de forma paulatina las características disponibles. De esta forma, se visualiza el impacto en los resultados al agregar las mediciones derivadas de los distintos equipos oftalmológicos utilizados, entrenando el sistema con 1, 3, 5, 7, 9, 16 y 23 características; siendo el sistema de 7 características el screening test propuesto y el sistema de 23 características el modelo de referencia.

## Capítulo 4. Resultados

En el presente capítulo se muestran los resultados obtenidos en el preprocesamiento de los datos, en el manejo del desbalance de clases, en la implementación del screening test y en la comparación de los resultados del modelo propuesto versus el modelo de referencia.

### 4.1. Resultados del test de significancia

Los resultados de los test de significancia de valor p y valor f para cada característica se muestran en la tabla 4.1

**Tabla 4.1 Test de significancia de características predictoras. Fuente: Elaboración propia**

Métrica	Control	Queratocono	P value	F value
Sph	-1.57 ± 2.2	-2.47 ± 2.2	0.041	4.3
Cyl	-2.23 ± 1.9	-2.93 ± 1.7	< 0.001	27.0
Axis	83.25 ± 72.5	84.71 ± 63.4	0.014	6.2
K1	41.48 ± 1.3	44.81 ± 3.5	< 0.001	57.3
K2	43.77 ± 1.6	47.96 ± 3.8	< 0.001	83.4
AP	534.17 ± 24.5	483.95 ± 37.1	< 0.001	80.9
TP	528.51 ± 24.2	475.88 ± 38.0	< 0.001	87.4
Kmax	44.16 ± 1.7	53.03 ± 6.4	< 0.001	138.1
CV	59.85 ± 3.0	56.79 ± 2.8	< 0.001	21.4
RPlavg	1.07 ± 0.2	1.78 ± 0.7	< 0.001	80.5
RPlmin	0.77 ± 0.2	1.35 ± 0.6	< 0.001	63.3
RPlmax	1.34 ± 0.2	2.54 ± 1.1	< 0.001	89.8
Aemax	8.14 ± 4.8	18.95 ± 8.8	< 0.001	77.0
Pemax	15.51 ± 6.5	36.60 ± 17.4	< 0.001	94.2
Aedif	10.41 ± 4.9	32.74 ± 17.8	< 0.001	112.5
Pedif	19.86 ± 7.0	65.17 ± 36.8	< 0.001	115.4
D	1.26 ± 0.7	6.86 ± 4.5	< 0.001	119.3
ISV	23.94 ± 10.7	72.36 ± 35.9	< 0.001	126.3
IVA	0.12 ± 0.01	0.75 ± 0.45	< 0.001	154.7
KI	1.02 ± 0.002	1.19 ± 0.12	< 0.001	159.7
CKI	0.99 ± 0.1	1.05 ± 0.05	0.001	10.9
IHA	5.28 ± 4.3	26.68 ± 21.7	< 0.001	73.8
IHD	0.01 ± 0.006	0.08 ± 0.06	< 0.001	98.4

Se incluyen las medidas de media y desviación estándar por cada categoría para representar las propiedades del set de datos adquirido.

Se puede observar que, de las características utilizadas para el screening test, la capacidad predictiva de esfera y axis es levemente significativa, mientras que las características corneales derivadas de topografía presentan un alto grado de significancia como características predictoras. Se desprende entonces que las medidas que más aportan a la predicción, en el screening test de 7 características implementado, son las derivadas de los exámenes queratometría y paquimetría.

Las características que fueron eliminadas por no ser estadísticamente significativas o por ser linealmente dependientes de otra característica presente en el set de datos son: Rmin, Km, distancia TP – AP, AChD y ChV.

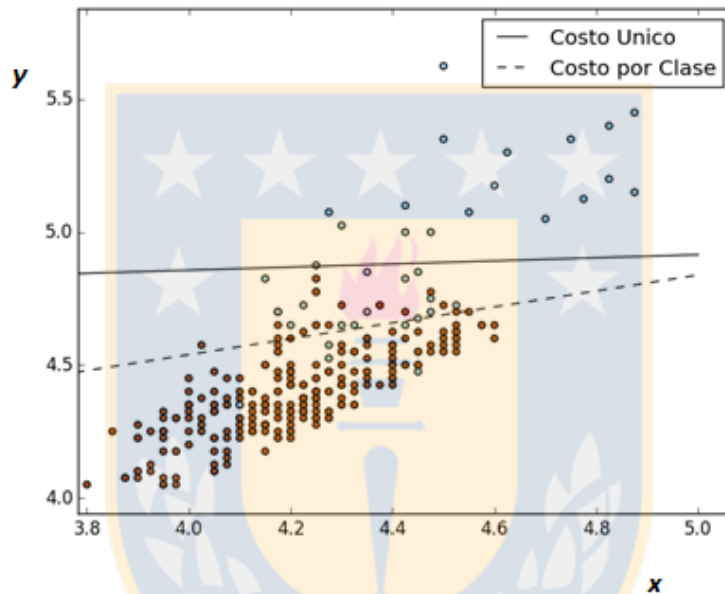
## **4.2. Representación del desbalance de clases**

Se expone a continuación los efectos del desbalance de clases en la predicción del modelo y la solución implementada en el presente trabajo que permite atenuar este efecto latente en el set de muestras del presente trabajo. A modo de ejemplo para el lector, se realizó un proceso de clasificación aparte del trabajo realizado, con el propósito de visualizar el efecto del desbalance de clases y también visualizar la solución implementada. Para esto, se realizó un proceso de clasificación que pudiese representar en 2 dimensiones el efecto de la utilización del costo diferenciado por clase “cw” en el algoritmo SVM, logrando así una comprensión visual de las modificaciones realizadas.

El clasificador implementado para este apartado es de carácter simple ya que ocupa una cantidad menor de características y no posee todas las afinaciones realizadas, solo las necesarias para explicar el fenómeno descrito.

Se configura el clasificador con un kernel lineal y se seleccionan 2 características presentes en el set de datos. Para este caso, se utilizaron mediciones de queratometrías, las cuales se representan gráficamente en el espacio de las características. La primera queratometría  $K1$  representa el eje  $x$  y la segunda queratometría  $K2$  representa el eje  $y$ .

Al utilizar sólo estos dos parámetros, cada registro se puede visualizar en el plano de características como un punto  $P$  de coordenadas  $(K1, K2)$ . Luego, al implementar el algoritmo se obtienen dos rectas, las cuales son los hiperplanos de decisión para casos de 2 dimensiones. La primera recta corresponde al algoritmo que presenta el mismo parámetro  $C$  de penalización de error en clasificación, mientras que la segunda recta en espunteado corresponde a la función de decisión que se obtiene al implementar una penalización diferenciada por clase para los errores de clasificación, tal como se ve en la Fig. 4.1.



**Fig. 4.1 Costo diferenciado por clase.** Fuente: Elaboración propia

En la Fig. 4.1, los registros de la clase queratocono están representados en el plano con el color azul y los datos de control se visualizan en el plano como puntos de color rojo. Aquí, se puede observar claramente el efecto que tiene la incorporación del concepto de costo diferenciado por clases en la formulación matemática del algoritmo, logrado a través del parámetro “class weight” mencionado. Esto influye en la función de decisión obtenida por el modelo haciendo un “*trade-off*” de los errores de clasificación, lo que atenúa el efecto del desbalance de clases en el rendimiento del clasificador, incrementando la cantidad de predicciones correctas en la clase de interés.

### 4.3. Afinación de parámetros SVM y modelo propuesto

Una vez que los datos han sido preprocesados y se han seleccionado las métricas definitivas resultantes de los test de significancia, se procede a construir el screening test propuesto con las 7 características no topográficas previamente definidas: esfera, astigmatismo, eje del astigmatismo, queratometría más plana, queratometría perpendicular, grosor de la córnea en el ápex y grosor mínimo. Mediante validación cruzada tipo Leave One Out, se entrena y evalúa el sistema recursivamente con la totalidad de los 123 datos, cada uno de ellos con las 7 características mostradas. Las métricas seleccionadas para evaluar el sistema son sensibilidad, especificidad y precisión, las cuales derivan de las componentes de la matriz de confusión generada en el proceso de predicción: verdaderos positivos (TP), falsos negativos (FN), falsos positivos (FP), verdaderos negativos (TN)

A los parámetros internos del algoritmo mencionados en metodología, se les asignó el rango numérico presentado en la tabla 4.2, el cual aplica para ambos kernel a implementar. Para el parámetro “C”, se definen 3 valores posibles. Para el parámetro “cw” se definen opciones numéricas cercanas a la proporción del desbalance de clases entre la clase control vs la clase queratocono. El parámetro “ $\gamma$ ” es solo pertinente en kernel rbf.

**Tabla 4.2: Rango numérico de parámetros SVM. Fuente: Elaboración propia**

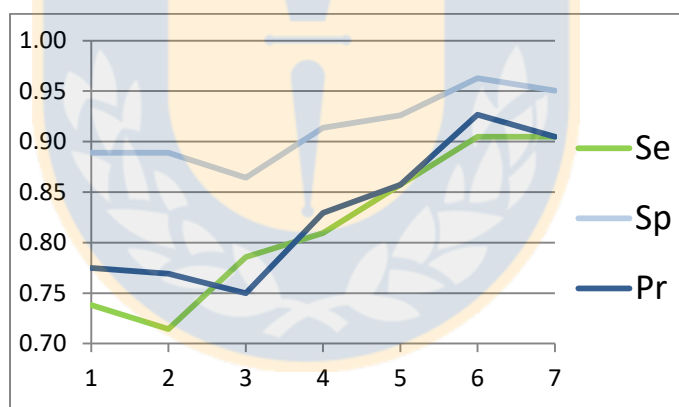
Parámetro	Rango
C	[1, 10, 100]
Cw	[1.0 – 3.0]
$\gamma$	[0.001 – 0.009, 0.01 – 0.09, 0.1 – 0.9, 1 - 10]

Posteriormente, se realiza la búsqueda exhaustiva de posibles combinaciones de valores de los parámetros de ajuste del algoritmo SVM que logren los mejores resultados en términos de las métricas seleccionadas. Durante esta búsqueda, se observó que el ajuste de cada parámetro incidía, generalmente, de forma independiente a los resultados de clasificación. Esto, en búsquedas manuales o en otras implementaciones, puede significar una disminución drástica de la cantidad de combinaciones posibles a considerar, desde 2394 posibilidades distintas a 85 combinaciones, de acuerdo al rango para cada parámetro

previamente definido. Curiosamente en este trabajo, al realizar el proceso de búsqueda exhaustiva computacional de estas 2394 combinaciones se llega a los mismos resultados que realizando la búsqueda de forma manual considerando independientes los efectos en los resultados de clasificación de cada parámetro. Debido a esto, se realiza el análisis del impacto de agregar los parámetros modificados en el desempeño en la clasificación, en el orden siguiente: C, cw y gamma.

**Tabla 4.3: Afinación de parámetros SVM. Fuente: Elaboración propia**

Config.	Kernel	C	Cw	$\gamma$	TP	FN	FP	TN	Se	Sp	Pr
1	Rbf	1	-	-	31	11	9	72	0.74	0.89	0.78
2	Lineal	1	-	-	30	12	9	72	0.71	0.89	0.77
3	Rbf	1	1.9	-	33	10	11	70	0.79	0.86	0.75
4	Lineal	10	-	-	34	8	7	74	0.81	0.91	0.83
5	Rbf	10	1.9	0.2	36	6	6	75	0.86	0.93	0.86
<b>6</b>	<b>Lineal</b>	<b>10</b>	<b>1.9</b>	<b>-</b>	<b>38</b>	<b>4</b>	<b>3</b>	<b>78</b>	<b>0.90</b>	<b>0.96</b>	<b>0.93</b>
7	Rbf	1	1.9	0.02	38	4	4	77	0.90	0.95	0.90



**Fig. 4.2: Evolución en la afinación de parámetros. Fuente: Elaboración propia**

Los mejores resultados para cada ajuste son mostrados en la tabla 4.3. En esta tabla, en primer lugar se muestran los resultados de clasificación sin ningún tipo de afinación, tanto para el kernel rbf como para el kernel lineal, destacando en sensibilidad el kernel rbf por sobre el lineal, con 0.74 versus 0.71 respectivamente. Luego, con un valor de penalización “C = 10” se incrementa la sensibilidad para ambas configuraciones, no obstante, rbf disminuye en especificidad y precisión. Como se puede observar también en la Fig. 4.2, el impacto de las afinaciones es más notorio en el kernel lineal.

Con respecto al factor “cw”, mencionado previamente en la sección 4.2 del presente capítulo, los mejores resultados se obtienen con un parámetro de penalización diferenciada  $cw = 1.9$ , de acuerdo a sensibilidad, especificidad y precisión. Curiosamente, la razón entre la clase de control y la clase de queratocono para nuestro set de datos es de 1.9, por lo que se desprende que el valor óptimo para cw para todas las métricas coincide con la razón entre clases. Un incremento en el valor cw por sobre 1.9, significaría un trade-off para incrementar sensibilidad en desmedro de la precisión o especificidad. Esta facultad puede ser importante a la hora de detectar enfermedades críticas sin importar los falsos positivos, quedando a criterio de los especialistas médicos si es necesario este trade-off en favor de la clase de interés para esta enfermedad en particular. No obstante, al realizar trade-off se corre el riesgo de un sobreajuste.

Finalmente, el mejor resultado (sin *trade-off*) lo logra el kernel lineal en la configuración 6, con un parámetro de penalización  $C = 10$  y  $cw = 1.9$ ; alcanzando una sensibilidad de 0.90, una especificidad de 0.96 y una precisión de 0.93. Cabe destacar que el kernel rbf en la configuración 7 logra los mismos resultados en sensibilidad, sin embargo en precisión solo logra 0.90. Esto corresponde, para nuestros datos, a solo un caso más de falsos positivos. Pese a que la diferencia es leve y es posible que con un set de datos más extenso el algoritmo con kernel rbf tome ventaja, la simplicidad de cálculo computacional y los resultados obtenidos decantan a la elección de la configuración 6 como resultado final de esta etapa de afinaciones de parámetros del algoritmo.

Entonces, el screening test para la identificación del queratocono propuesto por el presente trabajo consiste en la implementación del algoritmo SVM con 7 características predictoras, utilizando los valores afinados de parámetros de la configuración 6 mostrada, resultantes del proceso de búsqueda exhaustiva.

El progreso en los resultados de clasificación al modificar los parámetros C, cw y gamma se puede visualizar en la Fig. 4.2, la cual presenta un gráfico que muestra la evolución de los valores de sensibilidad, especificidad y precisión según las distintas configuraciones mencionadas en la tabla 4.3

#### 4.4. Evaluación del modelo propuesto como Screening Test

Para evaluar la funcionalidad como screening test de nuestro modelo de 7 características propuesto, se configura un segundo sistema. Este segundo modelo utiliza las 7 características utilizadas por el modelo anterior más 16 características provenientes de exámenes avanzados realizados con topografía corneal, sumando un total de 23 características. Las 16 características agregadas fueron mostradas en la tabla 4.1, y son las que prosiguen a las 7 primeras. Para la afinación de parámetros se utiliza la misma configuración de ajuste desarrollada para el modelo de 7 características; kernel lineal, parámetro de penalización “C” = 10 y peso por clase “cw” = 1.9. Este modelo de 23 características representa la mejor herramienta de identificación del queratocono disponible para nuestro set de datos y será considerado como el modelo de referencia.

El propósito de la construcción del modelo de referencia es la evaluación del desempeño en la clasificación de nuestro modelo de 7 características versus el desempeño del mejor modelo disponible. Para esto, se propone visualizar el impacto de agregar de a sub grupos –desde cero– las métricas de acuerdo al equipo médico de procedencia y de acuerdo a la simplicidad de obtención de estas; y entrenar y evaluar el sistema cada vez con las características que se vayan agregando. Con esto, se espera ver la evolución en los resultados de clasificación de acuerdo a una escala de métricas añadidas de 1, 3, 5, 7, 16 y 23 características, siendo 7 y 23 el modelo propuesto y el de referencia, respectivamente, tal como muestra posteriormente en la tabla 4.4.

Las primeras métricas que se agregan son las derivadas del autorefractómetro, agregando la medida de esfera, luego la medida de astigmatismo y el eje del astigmatismo. Seguidamente, se agregan las medidas de ambas queratometrías K1 y K2 para un total de 5 características y luego se agregan ambas medidas del grosor corneal –AP y TP– para completar el modelo propuesto de 7 características simples. Posteriormente, se agregan 2 y luego 8 características principales provenientes de un topógrafo corneal avanzado. Finalmente se agregan 8 índices especializados más, basados en desviaciones, asimetrías y aberraciones; también provenientes del mismo topógrafo. Con estas últimas mediciones de vanguardia se completa el sistema de referencia.

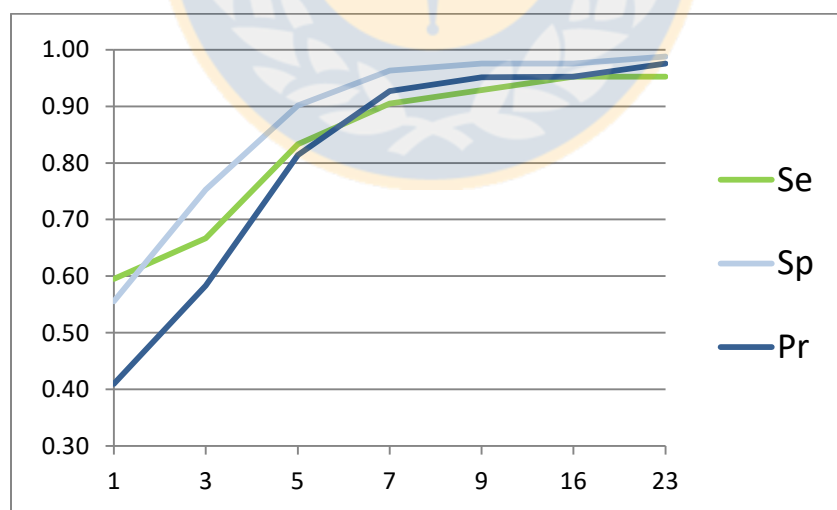


Como se puede observar en la tabla 4.4 y Fig. 4.3, las métricas de errores refractivos por si solas obtienen un pobre desempeño en sensibilidad, especificidad y precisión. El salto cualitativo en el desempeño se logra al agregar las medidas de queratometría, pasando de 0.67 Se, 0.75 Sp y 0.58 Pr; a 0.83 Se, 0.90 Sp y 0.81 Pr respectivamente. Finalmente, al agregar las métricas de grosor corneal se logran los resultados destacados en verde en la tabla de abajo, los cuales son los resultados obtenidos por el modelo de screening test propuesto.

**Tabla 4.4: Evaluación del modelo como Screening Test**

Fuente: Elaboración propia

Features	TP	FN	FP	TN	Se	Sp	Pr
1	25	17	36	45	0.60	0.56	0.41
3	28	14	20	61	0.67	0.75	0.58
5	35	7	8	73	0.83	0.90	0.81
<b>7</b>	<b>38</b>	<b>4</b>	<b>3</b>	<b>78</b>	<b>0.90</b>	<b>0.96</b>	<b>0.93</b>
9	39	3	2	79	0.93	0.98	0.95
16	40	2	2	79	0.95	0.98	0.95
<b>23</b>	<b>40</b>	<b>2</b>	<b>1</b>	<b>80</b>	<b>0.95</b>	<b>0.99</b>	<b>0.98</b>

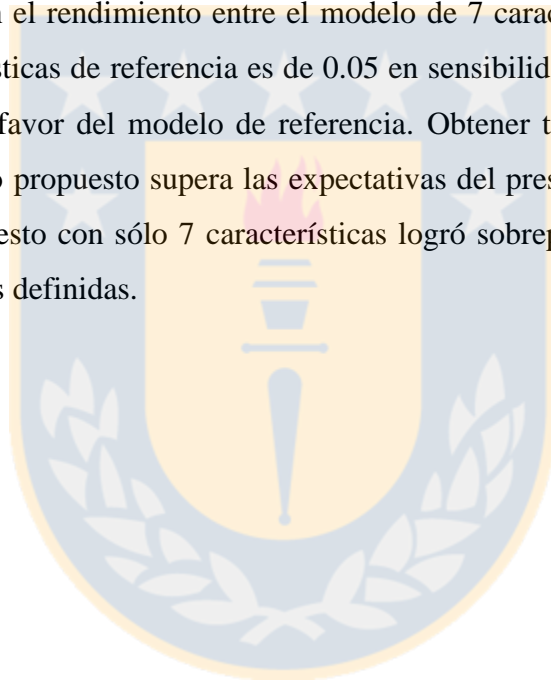


**Fig. 4.3: Rendimiento al agregar características predictorás**

Fuente: Elaboración propia

Es evidente que al agregar el resto de las 16 características disponibles existe un impacto positivo en el desempeño de clasificación del modelo, lo cual se muestra destacado en azul en la tabla anterior. No obstante, el incremento en el desempeño es menor comparado al incremento observado al agregar las primeras mediciones, tal como muestra la Fig. 4.3. Con esto, se puede inferir que lo robusto del algoritmo necesita de solo un par de características para lograr buenos resultados, mientras que todas las características adicionales sirven para un proceso de refinación o depuración de estos resultados; y si se partiera el proceso de evaluación con datos topográficos en primer lugar, se pueden esperar mejores resultados aún.

La diferencia en el rendimiento entre el modelo de 7 características propuesto y el modelo de 23 características de referencia es de 0.05 en sensibilidad, 0.03 en especificidad y 0.05 en precisión, a favor del modelo de referencia. Obtener tan poca diferencia en el rendimiento del modelo propuesto supera las expectativas del presente trabajo, destacando que el algoritmo propuesto con sólo 7 características logró sobrepasar un rendimiento del 90% en las tres métricas definidas.



## Capítulo 5. Conclusiones

---

En el presente capítulo, se expone un breve análisis y discusión sobre las distintas etapas del desarrollo del modelo de screening test propuesto. Además, se mencionan los posibles aspectos que se pueden profundizar y mejorar en investigaciones futuras relacionadas a la presente investigación

### 5.1 Discusión

En la presente investigación se propuso un método de identificación del queratocono en base a un algoritmo SVM entrenado y evaluado con las mediciones oftalmológicas de mayor accesibilidad, de tal forma que el modelo pueda ser utilizado como screening test en los lugares donde otros exámenes más avanzados de córnea no estén disponibles, y así, servir como una herramienta de apoyo al diagnóstico que pueda identificar correctamente los pacientes con la patología y derivarlos a los centros especializados donde si existan herramientas de diagnóstico más potentes.

Para completar este modelo, se propuso una metodología de trabajo en donde primeramente se adquieren los datos desde los centros oftalmológicos avanzados para luego ser pre procesados e ingresados al ambiente de desarrollo. Para construir el modelo se seleccionan las características de mayor disponibilidad y se definen las métricas de evaluación. Luego, se busca la forma de afrontar el desbalance de clases y se selecciona un modelo de validación cruzada acorde al set de datos adquirido. Posteriormente, se procede a la afinación de parámetros del modelo y a evaluar los resultados obtenidos con estos parámetros, con el fin de seleccionar el conjunto de valores de parámetros con el que el algoritmo de clasificación obtiene el mejor desempeño, de acuerdo a las métricas seleccionadas. Finalmente, se evalúa el funcionamiento del modelo como screening test, al ser comparado con un modelo de referencia que utiliza la totalidad de mediciones de vanguardia disponibles en el set de datos.

Durante el desarrollo de la primera sección del presente trabajo, se puede destacar que, en el contexto nacional, la adquisición de registros de pacientes anónimos relacionados con la presente investigación fue de difícil ejecución y se vio fuertemente mermada debido a burocracias internas en cada centro oftalmológico avanzado u hospitales de interés contactados. Aparte de la ya mencionada baja incidencia de la patología y del estado de subdiagnóstico que se postula, los factores principales que aumentaron la complejidad de la adquisición de muestras para la presente investigación son los siguientes:

- **Procesos de acreditación:** De acuerdo a la cartera del Ministerio de Salud de Chile, los procesos constantes de acreditación a los cuales son sometidos los centros de salud han aumentado el estándar de calidad en Chile. Sin embargo, algunas instituciones se han vuelto reacias a participar de cualquier tipo de investigación externa con el propósito de evitar cualquier tipo de problema o situación con el comité de fiscalización de la acreditación. Esto también denota cierta falta de voluntad y cultura investigativa que frena los intentos de cualquier persona externa a la institución que quiera hacer investigación en salud en nuestro país
- **Confidencialidad del paciente:** Nuevamente, para evitar problemas con cualquier entidad fiscalizadora (las cuales tampoco tienen una cultura de la investigación), la falta de claridad legal y la falta de diferenciación entre los distintos tipos de estudios por parte de las personas encargadas resulta en una negativa al acceso de datos, sin importar las diferencias entre distintos tipos de estudio o el anonimato especificado en este estudio en particular. La desinformación de todas las partes involucradas sobre cuándo aplicar consentimiento informado y cuándo no es necesario, resulta también en que los estudios que sí sean autorizados por el comité de ética de cada organismo, tengan que realizarse de forma prospectiva, sin poder utilizar las bases de datos recopiladas hasta ese entonces debido a una discutible aplicación de los conceptos de confidencialidad del paciente. Esto prolonga la duración del estudio más allá de los objetivos de este mismo.
- **Retraso tecnológico:** Finalmente, los centros oftalmológicos del país presentan un distinto estado tecnológico respecto a ficha electrónica, ficha unificada, o protocolos

de anotación (que los especialistas anoten en la ficha todo lo que corresponde). Muchas de las consultas y centros oftalmológicos con buena disposición a colaborar no tienen digitalizados sus registros o protocolizado las anotaciones en fichas, mientras que los centros privados que si están a la vanguardia presentan los problemas expuestos en los puntos anteriores. Todo esto hace que la adquisición de registros para el presente estudio, y para cualquier tipo de investigación en salud, sea un desafío engorroso.

Siguiendo con la metodología, en la etapa de preprocesamiento, aparte de los análisis realizados en la sección 4.1 del capítulo de resultados, se destaca que los datos recopilados para la categoría queratocono, incluían tanto queratoconos asintomáticos como queratoconos subclínicos. Estas diferencias significativas en el estado de la enfermedad hacen que la mayoría de las características predictoras de la categoría queratocono tengan en general una media considerablemente elevada con respecto a su par en la categoría de control. Similarmente, la dispersión de los datos en la categoría queratocono es mayor en todas las características predictoras respecto de la categoría de control, y en muchas de estas mediciones se puede apreciar, por la desviación estándar, que iba a existir un traslape entre las clases. Todo esto ratifica y aprueba la selección del algoritmo SVM del presente estudio, debido que al considerar mayormente para el cálculo de la función de decisión sólo los vectores de soporte, se logra un algoritmo robusto cuyo rendimiento no es significativamente afectado por la dispersión de los datos o por las medias muestrales elevadas debido a datos extremos (queratoconos graves).

También se puede apreciar en el test de significancia, que las características simples utilizadas por el modelo propuesto tienen un menor peso como predictoras en comparación a las características derivadas de topografía corneal. Aun con estas limitaciones, se pudo llegar a resultados que cumplen el objetivo de screening test. No obstante, otro sería el escenario si se pusiera un énfasis en la detección del queratocono subclínico, en donde claramente el modelo propuesto no obtendría resultados favorables utilizando sólo las características mencionadas.

El modelo propuesto obtuvo una sensibilidad de 0.9, una especificidad de 0.96 y una precisión de 0.93, sobrepasando así en todas las métricas un rendimiento del 90%. Estos resultados confirman que el modelo propuesto puede detectar sin problemas a los pacientes de queratocono definido vs pacientes con otras afecciones. Sin embargo, no se asegura un buen rendimiento al intentar detectar pacientes con queratocono subclínico. Para este último caso, es necesario ampliar el número de características predictoras agregando mediciones obtenidas de topógrafos corneales avanzados, lo cual va en contra del objetivo de funcionalidad como screening test del presente estudio. La única forma en que se puede ampliar el número de características predictoras sin perder de vista el objetivo es incorporando mediciones no complejas que no hayan sido utilizadas en el presente estudio, o esperar que nuevos equipos de vanguardia portátiles sean utilizados para estos fines, dependiendo así del tipo de mediciones que estos nuevos equipos realicen.

Cabe destacar también que los resultados presentes se obtuvieron sin realizar ningún *trade-off* para la clase de queratocono. Es decir, con las afinaciones propuestas y el manejo del desbalance de clases, aún se pueden mejorar los resultados obtenidos al sacrificar el rendimiento en la detección de la clase de control, aumentando la sensibilidad en la clasificación del queratocono, pero disminuyendo la precisión y especificidad. Usualmente, los screening test prefieren aumentar la tasa de falsos positivos por sobre la tasa de falsos negativos. Sin embargo, factores económicos pueden ser relevantes en este escenario. Cuánto *trade-off* se puede implementar en desmedro de un aumento de falsos positivos va a depender de la opinión de los especialistas y del énfasis que se le quiera dar a la detección de esta patología considerando todos los costos asociados.

Por otra parte, destacar que el modelo de 23 características predictoras – incluyendo mediciones topográficas– utilizado como modelo de referencia en el presente trabajo, obtiene resultados acorde a los rendimientos obtenidos por los modelos investigados en los antecedentes bibliográficos, logrando una sensibilidad de 0.95, una especificidad de 0.99 y una precisión de 0.98. Esto es equiparable a los resultados obtenidos por los últimos estudios revisados [13, 14, 15, 16, 17, 18]. De esta forma, se rectifica la utilización de este modelo de 23-c como modelo de referencia para evaluar el rendimiento como screening test del modelo de 7-c propuesto en el presente trabajo.

Finalmente, se destaca que el rendimiento del modelo propuesto, al ser comparado con el modelo de referencia, logra acortar la brecha a un 5% de diferencia en el rendimiento con respecto a la sensibilidad en la clasificación, un 3% de diferencia en el rendimiento con respecto a la especificidad y un 5% de diferencia en el rendimiento con respecto a la precisión. Todos estos resultados validan la funcionalidad como screening test del modelo propuesto, al lograr detectar la patología a un nivel muy cercano a como lo haría un modelo de clasificación que utiliza mediciones topográficas de vanguardia. Por estos motivos, se recomienda la utilización del modelo propuesto en distintos lugares donde, por variables de desarrollo, sociales, económicas o geográficas; no existe un acceso sencillo a exámenes de imagenología corneal.

## 5.2 Trabajo Futuro

Es necesario validar el modelo propuesto utilizando un conjunto de datos que mantenga las características utilizadas, pero con un número de muestras mayor. De esta forma se validan los resultados obtenidos y se evita cualquier indicio de sobreajuste al conjunto de datos utilizados en la investigación. Sería interesante también poder adquirir una base de datos que permita realizar clasificación multiclase, para de esta forma poder diferenciar la identificación del queratocono subclínico, leve, moderado o avanzado. De esta forma, se podría evaluar el rendimiento del modelo en estos segmentos específicos, y se podrían proponer soluciones para mejorar rendimiento en cada una de estas secciones, en especial, para incrementar la identificación del queratocono subclínico. Además, podría utilizarse para detectar algunas de las ametropías más comunes en los datos de control.

Por otra parte, se puede complementar el modelo propuesto al combinar los resultados con algunas características numéricas y categóricas que pueden resultar relevantes como: edad, predisposición genética, etnia, comorbilidades, alergias, frotamiento de ojos y seguimiento de la variación de la agudeza visual en el tiempo.

Finalmente, se pueden agregar a las características predictoras utilizadas, mediciones provenientes de tonómetro o de algún equipamiento oftalmológico portátil de vanguardia, con tal de incrementar la precisión en la detección de las distintas categorías de queratocono y funcionar como una herramienta de apoyo al diagnóstico.

## Bibliografía

- [1] Rabinowitz, Y. S. (1998). Keratoconus. *Survey of ophthalmology*, 42(4), 297-319.
- [2] Díaz Comínches, G., Cañas Ronda, A., Jiménez Cepeda, R., & Neira Peláez, R. (1999). Características epidemiológicas en pacientes portadores de queratocono. *Revista Cubana de Oftalmología*, 12(1), 0-0.
- [3] Nielsen, K., Hjortdal, J., Aagaard Nohr, E., & Ehlers, N. (2007). Incidence and prevalence of keratoconus in Denmark. *Acta ophthalmologica Scandinavica*, 85(8), 890-892.
- [4] Gokhale, N. S. (2013). Epidemiology of keratoconus. *Indian journal of ophthalmology*, 61(8), 382.
- [5] Millodot, M., Shneor, E., Albou, S., Atlani, E., & Gordon-Shaag, A. (2011). Prevalence and associated factors of keratoconus in Jerusalem: a cross-sectional study. *Ophthalmic Epidemiology*, 18(2), 91-97.
- [6] Gordon-Shaag, A., Millodot, M., & Shneor, E. (2012). The epidemiology and etiology of keratoconus. *Epidemiology*, 70(1).
- [7] Sherwin, T., & Brookes, N. H. (2004). Morphological changes in keratoconus: pathology or pathogenesis. *Clinical & experimental ophthalmology*, 32(2), 211-217.
- [8] Krachmer, J. H., Feder, R. S., & Belin, M. W. (1984). Keratoconus and related noninflammatory corneal thinning disorders. *Survey of ophthalmology*, 28(4), 293-322..
- [9] Maeda, N., Klyce, S. D., Smolek, M. K., & Thompson, H. W. (1994). Automated keratoconus screening with corneal topography analysis. *Investigative ophthalmology & visual science*, 35(6), 2749-2757.
- [10] Smolek, M. K., & Klyce, S. D. (1997). Current keratoconus detection methods compared with a neural network approach. *Investigative ophthalmology & visual science*, 38(11), 2290-2299.



- [11] Chastang, P. J., Borderie, V. M., Carvajal-Gonzalez, S., Rostène, W., & Laroche, L. (2000). Automated keratoconus detection using the EyeSys videokeratoscope. *Journal of Cataract & Refractive Surgery*, 26(5), 675-683.
- [12] Accardo, P. A., & Pensiero, S. (2002). Neural network-based system for early keratoconus detection from corneal topography. *Journal of biomedical informatics*, 35(3), 151-159.
- [13] Twa, M. D., Parthasarathy, S., Raasch, T. W., & Bullimore, M. (2003, May). Automated classification of keratoconus: A case study in analyzing clinical data. In *SIAM Intl. Conference on Data Mining*, San Francisco, CA.
- [14] Twa, M. D., Parthasarathy, S., Roberts, C., Mahmoud, A. M., Raasch, T. W., & Bullimore, M. A. (2005). Automated decision tree classification of corneal shape. *Optometry and vision science: official publication of the American Academy of Optometry*, 82(12), 1038
- [15] Marsolo, K., Twa, M., Bullimore, M. A., & Parthasarathy, S. (2007). Spatial modeling and classification of corneal shape. *Information Technology in Biomedicine, IEEE Transactions on*, 11(2), 203-212.
- [16] Souza, M. B., Medeiros, F. W., Souza, D. B., Garcia, R., & Alves, M. R. (2010). Evaluation of machine learning classifiers in keratoconus detection from orbscan II examinations. *Clinics*, 65(12), 1223-1228.
- [17] Arbelaez, M. C., Versaci, F., Vestri, G., Barboni, P., & Savini, G. (2012). Use of a support vector machine for keratoconus and subclinical keratoconus detection by topographic and tomographic data. *Ophthalmology*, 119(11), 2231-2238.
- [18] Arbelaez, M. C., & Sekito, M. B. (2013). Screening for subclinical keratoconus. *Oman journal of ophthalmology*, 6(1), 1.
- [19] Smadja, D., Touboul, D., Cohen, A., Doveh, E., Santhiago, M. R., Mello, G. R., ... & Colin, J. (2013). Detection of subclinical keratoconus using an automated decision tree classification. *American journal of ophthalmology*, 156(2), 237-246.

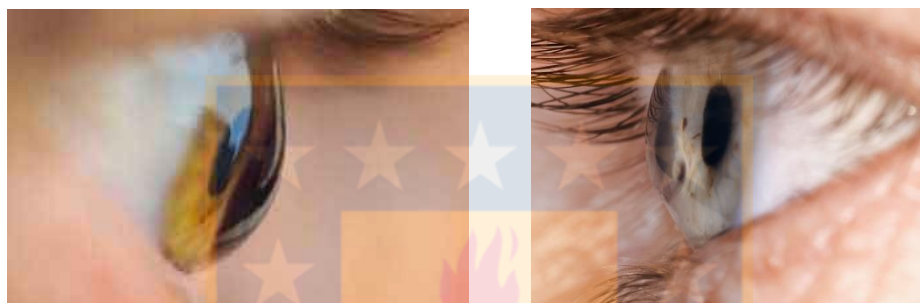
- [20] Hidalgo, I. R., Perez, P. R., Rozema, J. J., & Tassignon, M. J. B. (2014). Comparison of Machine Learning Methods to Automatically Classify Keratoconus. *Investigative Ophthalmology & Visual Science*, 55(13), 4206-4206.
- [21] Saad, A., & Gatinel, D. (2012). Evaluation of total and corneal wavefront high order aberrations for the detection of forme fruste keratoconus. *Investigative ophthalmology & visual science*, 53(6), 2978-2992.
- [22] Smadja, D., Touboul, D., & Colin, J. (2012). Comparative Evaluation of Elevation, Keratometric, Pachymetric and Wavefront Parameters in Normal Eyes, Subclinical Keratoconus and Keratoconus with a Dual Scheimpflug Analyzer. *International Journal of Keratoconus and Ectatic Corneal Diseases*, 1(3), 158-166.
- [23] Soto, D. (2014). *Identificación automática del hábito de tabaquismo utilizando información no estructurada de fichas clínicas electrónicas*, Memoria de Título, Ingeniero Civil Biomédico, Departamento de Ingeniería Eléctrica, Universidad de Concepción, Concepción.
- [24] Cortes, C., & Vapnik, V. (1995). Support vector machine. *Machine learning*, 20(3), 273.
- [25] Mikhail, B. (2003). *Support Vector Machines*, University of Texas at Austin, recuperado el 25-10-2015 desde <https://www.cs.utexas.edu/~mooney/cs391L/slides/svm.ppt>
- [26] Kim, E. (2013). *Kernel Trick*, recuperado el 30-05-2017 desde [www.eric-kim.net/eric-kim-net/posts/1/kernel\\_trick.html](http://www.eric-kim.net/eric-kim-net/posts/1/kernel_trick.html)
- [27] Joanneum, F. H. (2005). Cross-validation explained. *Graz, Austria: Institute for Genomics and Bioinformatics, Graz University of Technology*.
- [28] Domenech, J. (2011). *Validación Cruzada*, recuperado el 30-05-2016, desde [https://es.wikipedia.org/wiki/Validación\\_cruzada](https://es.wikipedia.org/wiki/Validación_cruzada)
- [29] Blake, C. R., Lai, W. W., & Edward, D. P. (2003). Racial and ethnic differences in ocular anatomy. *International ophthalmology clinics*, 43(4), 9-25

- [30] Oftalvist, *Queratocono*, recuperado el 08-10-2018, desde <https://www.oftalvist.es/es/especialidades/queratocono>
- [31] Rapuano, C. (2015, Diciembre). *Does Crosslinking Reduce the Need for Corneal Transplants for Keratoconus? It's Too Soon to Tell*, recuperado el 05-10-2018 desde <https://www.medscape.com/viewarticle/856181>
- [32] Stahl, *Keratoconus*, recuperado el 05-10-2018, desde <https://www.stahlny.com/eyecare-long-island/eye-conditions/keratoconus/>
- [33] Provision, *Queratocono*, recuperado el 05-10-2018, desde [http://www.provision.com.mx/shw\\_services/queratocono/index.html](http://www.provision.com.mx/shw_services/queratocono/index.html)
- [34] Nidek co., *HandyRef-K*, recuperado el 05-10-2018, desde [https://www.nidek-intl.com/product/ophthalptom/refraction/ref\\_auto/](https://www.nidek-intl.com/product/ophthalptom/refraction/ref_auto/)
- [35] DGH Technology, *The new pachmate 2 handheld pachymeter*, recuperado el 05-10-2018 desde <http://dgstechnology.com/product/pachmate-2/>
- [36] Optikon, *Keratron Scout the sharpest visión*, recuperado el 05-10-2018 desde <http://www.optikon.com/index/products-details/1/en/p/corneal-keratron-scout>
- [37] Ophthalmic Instruments, *Pentacam HR*, recuperado el 05-10-2018 desde <https://www.ophthalmic.com.sg/product/pentacam-hr/>

## Anexo

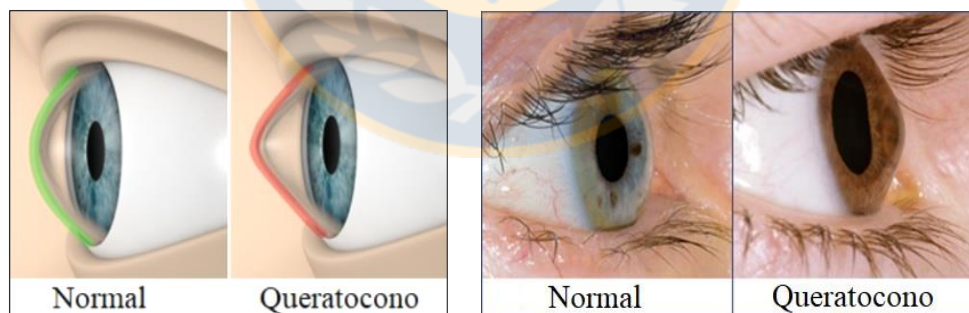
### A. Imágenes oftalmológicas

En el presente anexo, se muestran algunas imágenes del área oftalmológica con el propósito de ilustrar lo expuesto en los antecedentes del capítulo de introducción. Primeramente, en la Fig. A.1 se muestran imágenes reales de ojos con queratocono, donde se puede apreciar la deformación cónica que adopta la córnea.



**Fig. A.1 Ojos con queratocono [30,31]**

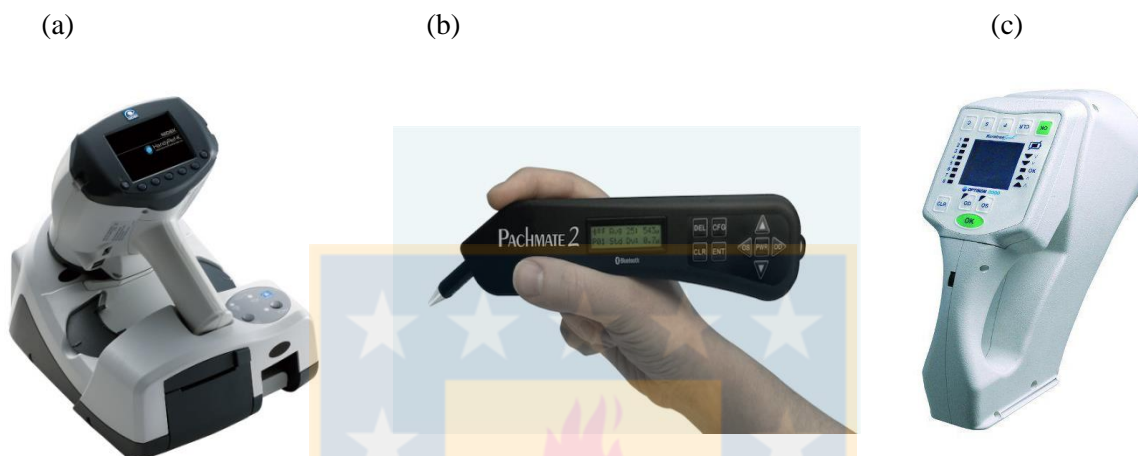
Para visualizar de mejor forma las diferencias entre una córnea normal y una córnea con queratocono, la Fig. A.2 muestra una ilustración y una imagen real en donde se comparan ambos casos.



**Fig. A.2 Córnea normal vs córnea con queratocono [32,33]**

El objetivo del presente trabajo contempla la identificación de la patología utilizando sólo los instrumentos oftalmológicos que pueden ser accesibles para la población general en consultas oftalmológicas comunes o eventualmente en operativos médicos sociales. Estos equipos oftalmológicos son de características más simples, pero esto no significa que no puedan ser equipos modernos. Autorefractómetro, queratómetro y

paquímetro son equipos simples ampliamente utilizados en las consultas oftalmológicas comunes. Por lo mismo, existen versiones portátiles de estos que cumplen con el propósito de comodidad y movilidad del instrumento hacia el paciente. Es más, existen topógrafos corneales avanzados de características portátiles que en un futuro podrían ser utilizados para el tamizaje de distintas enfermedades tal como se muestra en la Fig A.3



**Fig. A.3 Versiones portátiles de equipos oftalmológicos [34-36]**

- (a) Autorefractómetro-queratómetro Nidek HandyRef-K
- (b) Paquímetro de ultrasonido Pachmate-2
- (c) Topógrafo Optikon Keratron Scout

Tal como se mencionó en la introducción, la topografía corneal avanzada Pentacam –o equivalentes– se ha considerado como el *gold standar* para lograr el diagnóstico completo de esta patología, especialmente por sus propiedades de obtener mapas de elevación de la parte posterior de la córnea, así como también integrar la función de paquímetro y queratómetro.

Las características utilizadas para la construcción del modelo de referencia del presente trabajo contempló la utilización de 16 características provenientes de los mapas: elevación/Belin, refractivo y Topométrico; presentes en exámenes de topografía Pentacam realizados por el centro oftalmológico especializado. En la Fig. A.4 se muestra un topógrafo Pentacam en su instalación usual, mientras que en la Fig. A.5 se muestran los mapas derivados del examen.



Fig. A.4 Oculus Pentacam [37]

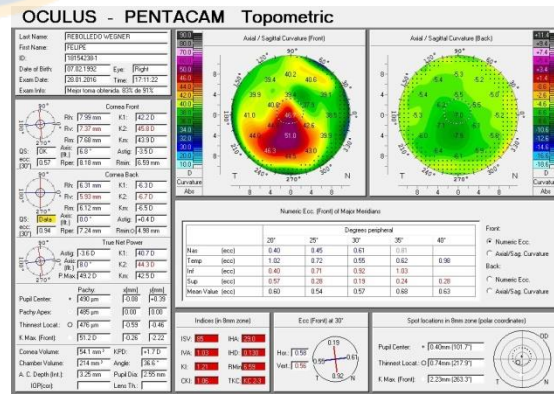
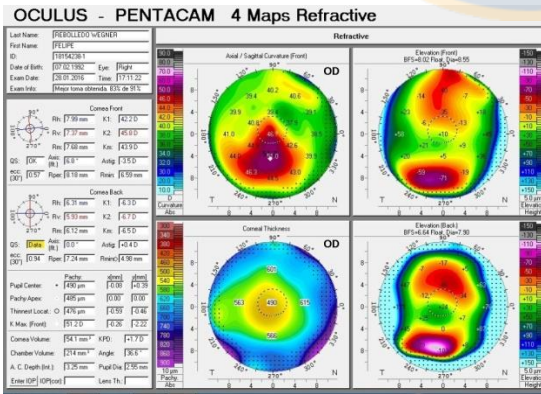
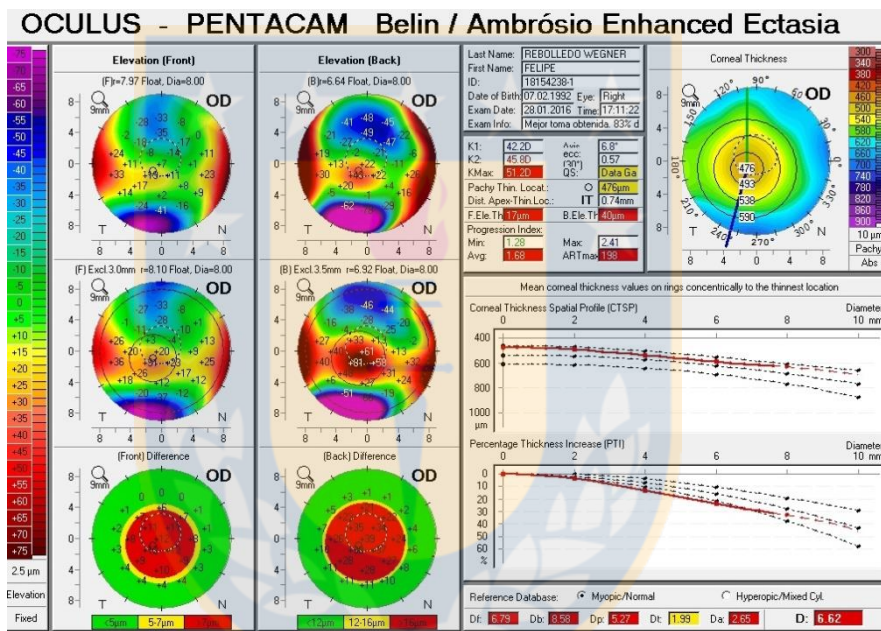


Fig. A.5 Mapa de elevación/Belin, refractivo y topométrico  
Fuente: Elaboración propia