



Cálculo de la potencia de una lente intraocular mediante técnicas de machine learning

David Carmona González
Estadística y Bioinformática
Machine Learning

Director/Consultor
Edwin Santiago Alférez Baquero

Fecha Entrega
05-06-2018

© (2018 - David Carmona González)

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Cálculo de la potencia de una lente intraocular mediante técnicas de Machine Learning</i>
Nombre del autor:	<i>David Carmona González</i>
Nombre del consultor/a:	<i>Edwin Santiago Alférez Baquero</i>
Nombre del PRA:	
Fecha de entrega (mm/aaaa):	06/2018
Titulación:	<i>Máster</i>
Área del Trabajo Final:	<i>Machine Learning</i>
Idioma del trabajo:	<i>Español</i>
Palabras clave	<i>Machine learning, lente intraocular, cirugía cataratas, biometría.</i>
<p>Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.</i></p>	
<p>OBJETIVOS: obtención de un algoritmo predictor para la potencia de la lente intraocular que implantaremos en un ojo.</p> <p>MÉTODOS: retrospectivamente se emplearon 260 ojos caracterizados biométricamente, con lentes intraoculares implantadas y resultados refractivos obtenidos. Implementamos diferentes algoritmos de regresión para predecir el valor de potencia dióptrica de la lente intraocular a implantar con la que vamos a obtener el mejor resultado refractivo. Se hizo validación externa del modelo y los resultados de predicción se compararon con Holladay II, Haigis, Barrett Universal II y Hill-RBF.</p> <p>RESULTADOS: se obtuvo un modelo de regresión combinado con un R^2 de 0.9863 y un RMSE de 0.6102. En las comparaciones con otras fórmulas de cálculo biométrico se obtiene un EAM de 0.34 ± 0.26 D (0.00/1.07) y un 75% de ojos incluidos en ± 0.50 D.</p> <p>CONCLUSIONES: se implementa un modelo regresivo no paramétrico con técnicas de machine learning, capaz de competir con las mejores fórmulas de cálculo biométrico.</p>	

Abstract (in English, 250 words or less):

PURPOSE: To develop a predictive algorithm to obtain the power of an intraocular lens to be implanted in an eye.

METHODS: Retrospectively, a total of 260 eyes biometric characterized with intraocular lens implanted and their corresponding refractive outcomes. Were used to get the best refractive outcomes, several regression algorithms were implemented to predict the dioptric power of the lens that will be implanted in a surgery.

RESULTS: The best result was a stacking regression model with a R^2 equal to 0.9863 and a RMSE equal to 0.6102. Comparing with other intraocular lens power calculation formulas, the outcomes were MAE of 0.34 ± 0.26 D (0.00/1.07) and the percentage of eyes within ± 0.50 D were 75%.

CONCLUSIONS: A non-parametric regressive model with machine learning techniques is implemented, capable of competing with the best intraocular lens power calculation formulas.

Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo.....	1
1.2 Objetivos del Trabajo.....	2
1.3 Enfoque y método seguido.....	2
1.4 Planificación del Trabajo.....	3
1.4.1 Tareas.....	3
1.4.2 Calendario y temporización.....	4
1.4.3 Hitos.....	5
1.4.4 Análisis de riesgos.....	5
2. Materiales y métodos.....	6
2.1 Exploración de los datos.....	6
2.2 Análisis descriptivo de los datos.....	11
2.3 Transformación de los datos.....	16
2.4 Implementación de algoritmos de machine learning.....	16
3. Resultados.....	22
3.1 Comparando modelos.....	22
3.2 Validación externa de los modelos.....	24
3.3 Ensamble de modelos.....	26
3.4 Selección del modelo definitivo.....	28
3.5 Comparaciones del modelo con las fórmulas biométricas reales.....	29
3.6 Comparaciones de los modelos con otros estudios publicados.....	31
4. Conclusiones.....	32
5. Glosario.....	33
6. Bibliografía.....	34

Lista de figuras

- Figura 1. Variables anatómicas corneales e intraoculares.
Figura 2. Descripción del equivalente esférico.
Figura 3. Estadística descriptiva e histogramas de la muestra total ($n = 260$).
Figura 4. Estadística descriptiva del nuevo dataset reducido a 12 variables.
Figura 5. Boxplot de todas las variables biométricas por grupos de longitud axial.
Figura 6. Diagramas de correlación entre variables predictoras.
Figura 8. Variables predictoras frente a residuos en modelo lineal.
Figura 9. QQplot y Test de Shapiro-Wilk para contrastar normalidad.
Figura 10a y 10b. Valores R^2 , RMSE y MAE de SVM Kernel Radial en bruto y con hiperparámetros.
Figura 11. Plots de normalidad, selección del modelo y distribución acumulada.
Figura 12. Comparación de modelos.
Figura 13. Plots predicción/observación entre train y test en los mejores modelos.
Figura 14. Importancia de las variables predictoras en los mejores modelos.
Figura 15. Diagrama y matriz de correlaciones entre modelos.
Figura 16. Boxplot de R^2 y contraste de hipótesis dos a dos en los tres modelos.
Figura 17. Boxplots de EAM y MedEA del modelo ENSEMBLE y el resto de fórmulas biométricas.
Figura 18. Porcentaje de ojos dentro del rango postoperatorio ± 0.50 D ($n = 52$).

Lista de Tablas

- Tabla 1. Tabla resumen de todas las variables biométricas sin discriminar por longitud axial.
Tabla 2. Valores. R^2 , RMSE y MAE de todos los modelos en bruto (muestra entrenamiento).
Tabla 3. Contraste dos a dos con Bonferroni.
Tabla 4. Comparación de métricas RMSE, R^2 y MAE en train y test en los mejores modelos.
Tabla 5. Resultados del modelo stacking.
Tabla 6. EP, EAM y MedAE de todas las fórmulas y el modelo ENSEMBLE.
Tabla 7. Porcentaje de ojos dentro del rango postoperatorio ± 0.75 D y ± 1.00 D.

1. Introducción

1.1 Contexto y justificación del Trabajo

Actualmente, la cirugía de cristalino se ha convertido en una sinergia multidisciplinar en la que podemos diferenciar claramente tres fases. La primera es prequirúrgica e interviene un optometrista clínico para la toma de medidas y de decisiones biométricas sobre el ojo del paciente. La segunda fase es la intraoperatoria en la que el médico oftalmólogo ejecuta la técnica quirúrgica y por último la etapa postquirúrgica en la que ambos especialistas siguen al paciente para explorar de forma paralela variables puramente médicas y otras solo ópticas.

El procedimiento clínico necesario para obtener la potencia dióptrica de una lente intraocular que será implantada en el ojo de un paciente se describe a continuación. En primera instancia medimos en el ojo del paciente mediante técnicas biométricas una serie de variables importantes que caracterizan matemáticamente la anatomía ocular. Dichas variables se introducen en unas fórmulas de cálculo que arrojan un valor de potencia para la lente intraocular, además de la predicción de un valor de graduación residual postoperatorio. El valor dióptrico de la lente que lleve asociado el menor valor refractivo residual es el que se implanta.

La forma en la que estas fórmulas biométricas son capaces de predecir la potencia y la graduación residual es muy diferente, de hecho, no todas operan con las mismas variables biométricas. Hay multitud de estudios [1,2,3] en la literatura oftalmológica que analizan y comparan la precisión y predictibilidad de unas y otras.

Desde el año 1967 en el que S. Fiodorov implementó la primera fórmula de cálculo biométrico hasta la actualidad, hemos pasado por 5 generaciones de evoluciones y mejoras en estos algoritmos matemáticos. No obstante, no hay una fórmula perfecta para todos los casos, dependiendo de las características anatómicas del ojo escogeremos una u otra fórmula. Hoy en día es aceptado por la comunidad científica [4] que un buen resultado es aquel que se encuentra en el rango ± 0.50 dioptrías postoperatorias.

Ninguna de las 9 fórmulas más utilizadas [5] en la actualidad ofrece la garantía de predecir perfectamente la potencia de la lente en todos los rangos anatómicos [6], ya que unas funcionan mejor que otras dependiendo de la queratometría, de la longitud axial o de la profundidad de cámara anterior.

Esta es precisamente la brecha dónde tiene cabida este trabajo. Explorar una nueva forma de predecir la potencia de la lente intraocular mediante técnicas de machine learning. Confirmar si esta manera de estudiar el problema nos permitirá mayor precisión en la predicción.

Hasta la fecha, solo hay una fórmula de reciente aparición promovida por el Dr. W. Hill llamada Hill-RBF, basada precisamente en un modelo no lineal con función gaussiana Radial Basis Function (RBF) y que actualmente está siendo sometida a multitud de estudios comparativos con las fórmulas tradicionales [7].

Clarke [8] en 1997 fue el pionero con su trabajo, en el que comparó la predicción de la potencia de la lente intraocular mediante una red neuronal con una fórmula absolutamente instaurada y fiable en la década de los 90, la Holladay I. Concluyó que los resultados obtenidos con la red neuronal disminuían significativamente la varianza del error. Desde entonces hasta la publicación de Hill-RBF, probablemente favorecido por el auge del aprendizaje automático, no ha habido este tipo de abordaje en el cálculo de lentes intraoculares.

1.2 Objetivos del Trabajo

Objetivo general:

El objetivo primario en este trabajo de investigación es proponer mediante técnicas de aprendizaje automático, un modelo robusto para predecir la potencia de la lente intraocular que se implantará a un paciente en una cirugía de cristalino a través de *inputs* biométricos obtenidos en clínica.

Objetivos específicos:

- Filtrar, ordenar y limpiar los datos.
- Implementar diferentes métodos de preprocesamiento y selección de características que permitan tomar las más apropiadas para la regresión.
- Implementar diferentes algoritmos de regresión, para obtener diferentes modelos que se comparan con el fin de seleccionar el de mejor desempeño.
- Comparar con las fórmulas de cálculo biométrico que ya existen. No obstante, se seleccionarán las más utilizadas de tercera, cuarta y quinta generación; Haigis, Holladay II, Barrett Universal Formula y Hill-RBF respectivamente.

1.3 Enfoque y método seguido

Se cuenta con una base de datos original en la que todos los ojos ya están operados y con una lente intraocular implantada. De este modo la graduación residual posoperatoria ya está calculada y por tanto sabemos si el resultado obtenido ha sido bueno o no. Contamos por tanto con un dataset con el registro de la potencia de la lente intraocular implantada que arroja los menores valores de graduación tras la cirugía en función de las particulares características anatómicas biométricas oculares.

Se implementarán por tanto algoritmos de clasificación supervisada para la obtención de la potencia de lente intraocular en base a una caracterización ocular numérica concreta.

1.4 Planificación del Trabajo

1.4.1 Tareas

- *Definición de los contenidos del Trabajo*

Estudio y definición inicial de todos los contenidos del trabajo. Se propone una línea de tiempo con los hitos y tareas que se llevarán a cabo en el Trabajo.

- *Dataset del trabajo*

El conjunto de datos es original y no disponible en ningún repositorio.

- *Ordenación de datos*

Se filtrará, ordenará y limpiará la base de datos para un tratamiento posterior y así obtener más y mejor información.

- *Exploración inicial del conjunto de datos a tratar*

Se hará un estudio inicial del tipo de variables con las que contamos en el dataset, así como una exploración descriptiva analítica y gráfica de dichas variables.

- *Preprocesamiento*

En esta fase analizaremos los datos mediante técnicas de transformación y reducción de dimensiones para abordar la siguiente etapa del trabajo con las mayores garantías.

- *Selección de métodos clasificatorios*

En esta fase analizaremos cuál de los diferentes enfoques clasificatorios es el más eficiente. Probaremos las siguientes metodologías; k-Nearest Neighbors, Naive Bayes, Artificial Neural Network, Support Vector Machine con diferentes Kernel, Decision tree, Random Forest, etc.

- *Soporte informático del proyecto*

Respecto al software para todo el desarrollo del trabajo se utilizará el entorno y lenguaje de programación R (versión 3.4.3), a través del conocido IDE RStudio (versión 1.1.423) para plataformas Windows.

Se trabajará con la librería CARET versión 6.0-78 (Classification and Regression Training de Max Kuhn).

En cuanto al hardware se ha empleado un Laptop Lenovo con Windows 10 y un procesador Intel core i7-6700 HQ con 16 GB de RAM.

- *Comparación y evaluación de los distintos métodos*

Se comparará el resultado de todos los modelos utilizados y determinaremos cual es el mejor mediante métricas de R^2 , RMSE y MAE en la predicción.

- *Elaboración del informe final*

Se procederá a realizar un informe científico con los resultados obtenidos, analizándolos y mostrando las conclusiones del trabajo.

- *Difusión*

Si los resultados obtenidos mejoran las predicciones realizadas con las herramientas que existen a nuestra disposición actualmente en la clínica diaria, divulgaremos el trabajo a través de los circuitos y canales científicos convencionales, publicación en revista JCR y presentación en congresos como la American Society of Cataract and Refractive Surgery (ASCRS) y/o la European Society of Cataract and Refractive Surgery (ESCRS).

1.4.2 Calendario y temporización

El trabajo fin de máster conlleva 300 horas aproximadamente que se desglosarán a continuación en una planificación temporal por horas y tareas acompañada de un cronograma (Figura.1) con todos los plazos del trabajo.

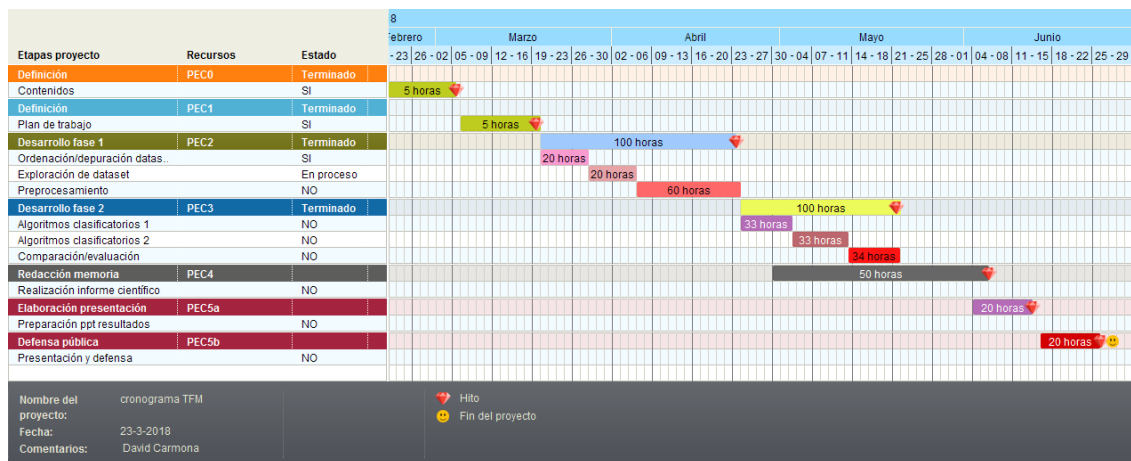


Figura 1. Diagrama de Gantt con definiciones temporales del proyecto completo

Las dos primeras PECs conforman 10 horas en conjunto. La PEC 2 que corresponde a la primera fase del desarrollo del trabajo abarcará 100 horas de trabajo distribuidas en tres etapas tal y como se puede observar en el diagrama anterior. La primera serán 20 horas para la ordenación y depuración del dataset, una segunda etapa de 20 horas dedicadas a la exploración en la que se hará un estudio del tipo de variables con las que contamos en el dataset, así como una evaluación descriptiva analítica y gráfica de dichas variables. Y las últimas 20 horas de este bloque se dedicarán al preprocesamiento de los datos.

La PEC 3 o la segunda fase del desarrollo del trabajo también corresponderá a 100 horas distribuidas igualmente en tres fases diferenciadas. En la primera fase de 33 horas implementaremos un primer bloque de algoritmos clasificatorios. A continuación, en una segunda etapa de 33 horas implementaremos otro bloque de algoritmos nuevos y por último en una fase estimada en 34 horas se analizarán las comparaciones entre modelos.

A la redacción de la memoria se dedicarán 50 horas que se distribuirán en 27 días, de los cuales 13 se decide solapar con la anterior tarea (PEC 3) para optimizar los recursos ya que en esas fechas tengo concentración de congresos nacionales e internacionales.

Por último, las PEC 5a y 5b suman 40 horas para completar el total del proyecto final de máster, distribuidas en 20 horas cada actividad.

1.4.3 Hitos

Los hitos representados en el cronograma de la figura 1 por diamantes, se corresponden con las entregas programadas en la asignatura.

1.4.4 Análisis de riesgos

- Coste de tiempo:

Actualmente se cuenta con una base de datos útil de 200 ojos intervenidos con todos los cálculos realizados. Con motivo de la mejora en la obtención de resultados, se pretende recopilar más datos para ampliar el dataset. Esta tarea implica un coste de tiempo adicional que se asume sin que repercuta en las tareas establecidas.

Se da como tiempo límite el comienzo de la PEC 3, en cuyo caso si no se ha conseguido aumentar la muestra continuaremos con la fijada inicialmente.

- Costo económico:

El hardware y software son de uso personal y con licencia GNU. El procesador de texto Word (Microsoft Office 365) utilizado para desarrollar el trabajo, es bajo licencia privada permanente. No se necesita ninguna inversión adicional ni ninguna otra infraestructura para completar el trabajo.

- Coste personal:

En caso de enfermedad u otro imprevisto personal, se recuperará el tiempo perdido ampliando los horarios previstos.

- Implicaciones éticas y legales:

La base de datos utilizada es original y no se encuentra en ningún repositorio. Preserva absolutamente el anonimato y cumple con los preceptos de la declaración de Helsinki, además ha sido revisada y evaluada por el comité de bioética del hospital en el que trabajo.

2. Materiales y métodos

2.1 Exploración de los datos

A continuación, describimos las 23 variables que forman parte del dataset de trabajo ($n = 260$) y en la Figura 1 se muestra la ubicación y dimensión anatómica de todas las variables biométricas que se emplearon para realizar este trabajo.

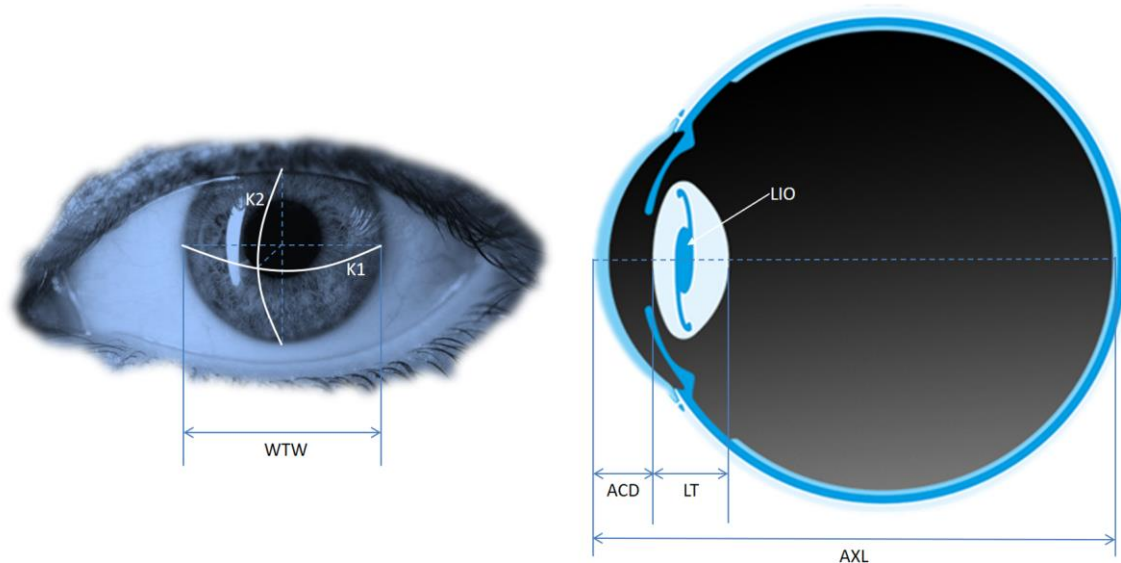


Figura 1. Variables anatómicas corneales e intraoculares.

AXL:

Axial Length o longitud axial. Es una variable numérica medida en milímetros y hace referencia a la longitud antero-posterior del globo ocular. Se determinó clínicamente con un biómetro IOL Master 700 (Carl Zeiss Meditec, Gena, Alemania) basado en la tecnología Optical Coherence Tomography SWEPT Source.

ACD:

Anterior Chamber Depth o profundidad de la cámara anterior. Es una variable numérica medida en milímetros y representa la distancia ortogonal entre el epitelio corneal y la superficie anterior del cristalino. Se midió con el mismo biómetro reseñado en el párrafo anterior.

LT:

Lens Thickness o espesor del cristalino. Es una variable numérica medida en milímetros que establece el grosor cristalino. Se determina con el mismo dispositivo comentado anteriormente.

WTW:

White To White o blanco-blanco. Es una variable numérica medida en milímetros y hace referencia al diámetro horizontal de la córnea. Se utiliza como variable biométrica que caracteriza anatómicamente el ojo, porque tiene una correlación con el saco capsular donde va alojada la lente intraocular y este no se puede medir directamente en clínica.

El diámetro horizontal corneal también se mide con el biómetro especificado anteriormente.

ASD:

Anterior Segment Depth o profundidad del segmento anterior. Es una variable numérica expresada en milímetros que representa la suma de la ACD y del LT.

$$ASD = ACD + LT$$

Se incluyó este parámetro por la posible relevancia anatómica ocular, ya que clásicamente los ojos se han dividido en tres categorías discriminando por longitud axial; cortos, normales y largos respectivamente. En consecuencia, podríamos encontrarnos ojos cortos con longitudes anteroposteriores inferiores a 21.00 mm y sin embargo poseer una ASD alta, o al contrario, presentarse un ojo axialmente largo con un segmento anterior pequeño.

K1_F, K2_F y KM_F:

Keratometry1_Front, Keratometry2_Front y KeraometryMean_Front. Estas variables hacen referencia a la queratometría de la superficie corneal anterior, es decir, los radios de curvatura medidos en milímetros de los meridianos principales, así como el valor medio de ambos. Se obtienen con el mismo biómetro.

La córnea es un dioptro de geometría astigmática, lo que implica la existencia de dos meridianos ortogonales de máxima y mínima potencia refractiva (Figura 2). La diferencia de potencia entre ambos ejes corresponde al astigmatismo de la córnea.

K1_B, K2_B y KM_B:

Keratometry1_Back, Keratometry2_Back y KeraometryMean_Back. Al igual que las anteriores hacen referencia a la queratometría, pero de la cara posterior de la córnea.

En este caso el dispositivo que se utilizó para determinar los radios de curvatura de la superficie posterior corneal fue el Pentacam HR (Oculus Optikgerate GmbH, Wetzlar, Alemania).

R_B/F:

Ratio_Back/Front. Esta variable es numérica adimensional. Se refiere al ratio entre el radio de curvatura de la cara posterior de la córnea y anterior respectivamente multiplicado por 100.

$$R_{B/F} = \frac{KM_B}{KM_F} \times 100$$

Se incluye como caracterizador biométrico porque puede ser un factor diferencial y concluyente en ojos anatómicamente diferentes. Según la literatura revisada en ojos normales que no han sufrido ningún procedimiento quirúrgico a nivel ocular, el promedio poblacional de este ratio es del 82.2% [9].

MOD_IOL:

Modelo_IntraOcular Lens. Es una variable categórica y hace referencia al modelo de lente intraocular que se ha implantado en el ojo. Esta variable se ha tenido en cuenta porque el abordaje actual con las fórmulas de cálculo biométrico necesita discriminar por tipo de lente, para predecir un valor de potencia y de residuo refractivo.

Cada modelo de lente intraocular tiene un diseño, un material, un tamaño, una geometría y todos estos parámetros están aglutinados en una constante llamada constante del fabricante o constante A, única para cada modelo de lente. Para realizar el cálculo con cualquier fórmula, se necesitan introducir todas las variables biométricas con las que trabaje dicha fórmula, además de identificar la lente intraocular que se quiere implantar mediante las constantes adecuadas, para que el cálculo se personalice para esa lente en concreto.

Para la implementación de modelos teóricos con machine learning se prescindirá de esta variable, y solo para la presentación de resultados comparativos con las fórmulas biométricas, transformaremos la media del error predictivo a cero. De este modo se eliminará el error sistemático derivado del uso de diferentes constantes no optimizadas.

IOL_I:

IntraOcular Lens_Implanted. Se refiere a la potencia dióptrica de la lente intraocular implantada. Es una variable numérica, se expresa en dioptrías y como se puede apreciar tiene una precisión de media dioptría, ya que no se manufacturan lentes de cuarto de dioptría, salvo excepciones. Este valor es el que predice la fórmula biométrica empleada para el cálculo.

IOL_P:

IntraOcular_Lens_Perfect. Se refiere a la potencia de la lente intraocular que tendríamos que haber implantado en vez de IOL_I. Es una variable numérica y expresada en dioptrías.

Se debe tener especial precaución con las equivalencias de potencia en los diferentes planos de trabajo. Los planos útiles son el corneal y el de la lente intraocular, separados uno de otro por la posición efectiva de la lente o effective lens position (ELP) y por el índice de refracción del humor acuoso que aproximadamente es 1.336.

El valor de IOL_P se obtiene de la diferencia entre la lente implantada (IOL_I) y el equivalente esférico postoperatorio (SE_post) con la distometría cornea-saco aplicada. Teniendo en cuenta que el trabajo es retrospectivo, se sabe que residuo refractivo le ha quedado al paciente con la lente implantada seleccionada, de modo que se puede obtener el valor de la lente que se tendría que haber seleccionado para dejarlo sin refracción residual postoperatoria.

T_HOL2, T_HAI, T_BAR y T_HILL:

Target_Holladay 2, target_Haigis, target_Barrett y target_Hill. Son los valores de refracción en formato equivalente esférico que predicen las fórmulas biométricas con las que se ha calculado la potencia de la lente intraocular. Son variables numéricas medidas en dioptrías.

Como se dijo anteriormente la anatomía corneal con sus meridianos de máxima y mínima potencia refractiva, producen dos líneas focales perpendiculares entre sí y separadas longitudinalmente en proporción a la diferencia de potencia entre ellas. Como se puede observar en la figura 2, el punto medio entre las dos focales imágenes se conoce como círculo de mínima confusión (CMC) o equivalente esférico (SE).

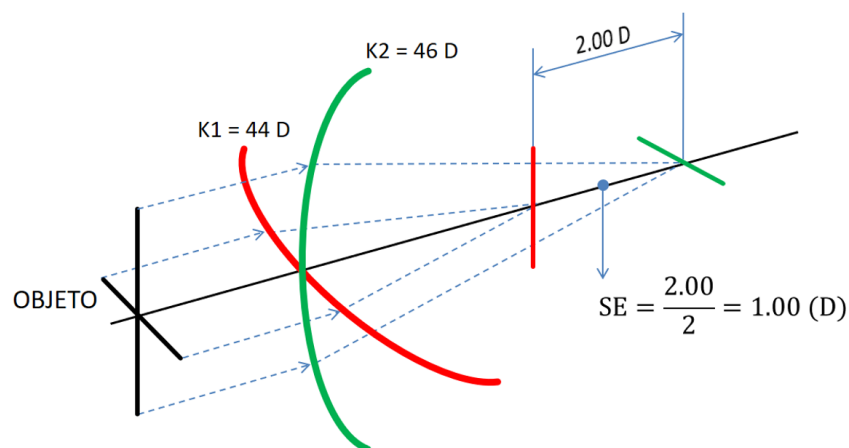


Figura 2. Descripción del equivalente esférico.

Las fórmulas biométricas utilizadas en el trabajo son Holladay 2, Haigis, Universal formula Barrett 2 y Hill-RBF 2.0. Todas ellas están incorporadas en el software del biómetro utilizado en este trabajo, exceptuando Hill-RBF 2.0 que

es de acceso libre a través de su web <https://rbfcalculator.com/online/index.html>.

Estas fórmulas biométricas son las más empleadas en la actualidad y en este trabajo se han calculado para comparar sus resultados refractivos con el modelo regresivo más eficiente obtenido en el presente estudio.

SE_POST:

Spherical Equivalent_post o Equivalente Esférico_postcirugía. Hace referencia al equivalente esférico residual postoperatorio. Es una variable numérica y expresada en dioptrías.

Las fórmulas biométricas predicen un valor dióptrico de lente intraocular y un valor asociado de equivalente esférico como residuo refractivo. Una vez que la lente se implanta y se da el alta médica al paciente, se le comprueba la graduación objetiva y subjetivamente para determinar el equivalente esférico residual final (SE_post).

Cuanto más cerca de cero esté, mejor es el resultado visual para el paciente y por tanto mejor la fórmula empleada.

Tal y como se aprecia en el dataset se puede encontrar valores negativos y positivos. Los negativos significan que la fórmula empleada sobreestimó la potencia de la lente a implantar y el ojo se quedó miope. Por el contrario, los casos positivos implican infravaloración de la potencia refractiva dejando el ojo hipermetrope.

MEP_HOL2, MEP_HAI, MEP_BAR y MEP_HILL:

Mean Error Predictive_Holladay2, Mean Error Predictive_Haigis, Mean Error Predictive_Barrett y Mean Error Predictive_Hill. Estas variables hacen referencia a los errores predictivos de las fórmulas. Son variables numéricas y medidas en dioptrías.

Se obtienen por la diferencia entre el equivalente esférico postoperatorio y el valor de predicción de la fórmula en cuestión. Por ejemplo, para la fórmula del doctor Jack Holladay sería:

$$MEP_{HOL2} = SE_{post} - T_{HOL2}$$

Si es cero la fórmula predijo con un 100% de exactitud el residuo de graduación postquirúrgica. En caso de observarse un valor negativo, estamos ante una sobreestimación de la potencia refractiva de la lente y por tanto una miopía residual, y si fuera positivo el Mean Error la fórmula habría una infracorrección del valor de la lente o lo que es lo mismo una hipermetropía postquirúrgica.

SIZE_EYE:

Se refiere al tamaño del ojo en términos de AXL. Es una variable categórica con tres niveles: S, N y L (Short, Normal y Long respectivamente).

En la siguiente tabla se pueden comprobar los puntos de corte para cada rango:

SIZE_EYE	S	N	L
AXL	≤ 22.00 mm	(22.00 - 25.00 mm)	≥ 25.00 mm

Es habitual hacer este tipo de divisiones, por el diferente comportamiento de las fórmulas biométricas en función de esta particularidad anatómica.

En este trabajo solo se utilizará esta variable para clasificar anatómicamente la muestra, pero por la limitación de esta y la heterogeneidad de las distribuciones de ojos, se prescindirá de ella en la modelización.

2.2 Análisis descriptivo de los datos

En la Figura 3 se puede observar la estadística descriptiva del conjunto de datos, posibles datos faltantes o NA's, el tipo de variables y su distribución.

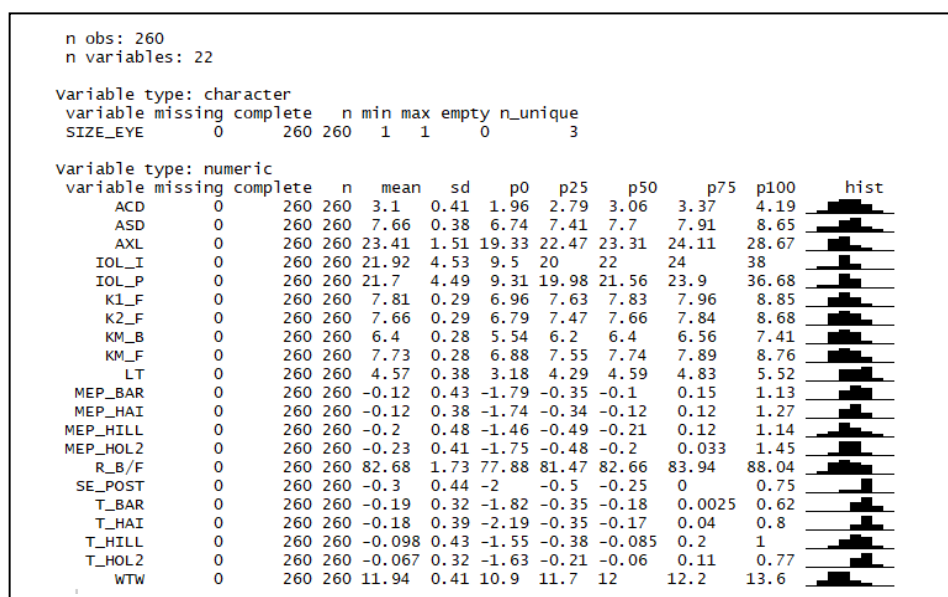


Figura 3. Estadística descriptiva e histogramas de la muestra total (n=260).

Como se explicó anteriormente, las lentes intraoculares (IOL) llevan asociada una constante A de fábrica dependiente de las características físico-ópticas de la lente. Por tanto, la variable categórica **MOD_IOL** que representa el modelo de IOL implantada, implícitamente tiene relación con dicha constante. Se elimina esta variable del conjunto de datos, porque buscamos un modelo predictivo que sea capaz de predecir sin dependencia de IOL ni de constantes. Se propondrá un modelo regresivo que proponga una potencia refractiva de IOL independientemente del tipo de lente intraocular que se quiera implantar.

Actualmente no existe ninguna fórmula de cálculo biométrico que funcione con dicha independencia.

Por otro lado, a efectos de la implementación del modelo eliminaremos del dataset las variables **T_BAR**, **T_HOL2**, **T_HILL**, **T_HAI**, **MEP_BAR**, **MEP_HOL2**, **MEP_HILL** y **MEP_HAI**. Todos estos parámetros tienen relación con los cálculos biométricos realizados con las fórmulas Barrett Universal II, Holladay 2, Hill-RBF y Haigis respectivamente. Estas variables se utilizarán posteriormente en el objetivo secundario de este trabajo, comparación de la precisión del modelo obtenido con las fórmulas existentes en la clínica actualmente.

La variable **IOL_I** o potencia refractiva de la lente intraocular implantada tiene un factor de corrección en su variable homóloga **IOL_P** como se explicó anteriormente en el apartado de exploración de datos. De modo que trabajaremos con los modelos de regresión solo con **IOL_P** para evitar duplicidades.

Por último, la variable **SE_post** o equivalente esférico postquirúrgico también quedará eliminada por ser el factor de compensación en **IOL_I**.

El resto de variables no eliminadas son caracterizadores anatómicos que particularizan el ojo que tengamos que operar e implantar, y por tanto serán variables predictoras para la modelización quedando como variable dependiente **IOL_P**.

Reconfigurando y reduciendo la dimensionalidad del dataset con todos estos cambios propuestos, el análisis descriptivo quedará como en la Figura 4.

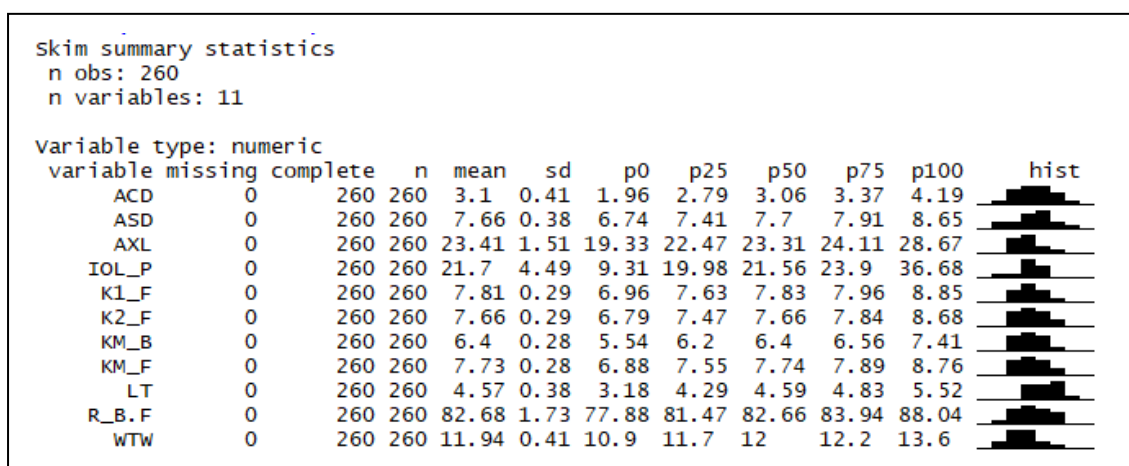


Figura 4. Estadística descriptiva del nuevo dataset reducido a 12 variables.

A continuación (Figura 5) se muestran todas las variables biométricas en un diagrama de boxplot discriminando por tamaños oculares S, N y L. Se observa en todas estas variables una correlación de proporcionalidad entre la

longitud axial y la categoría a la que corresponde dicha variable. Exceptuando LT (Lens Thickness) o espesor cristaliniano, dónde se puede comprobar que para los ojos que pertenecen al grupo Long (ojos ≥ 25 mm) la mediana está por debajo de 4.50 mm, cabiendo esperar a priori tamaños promedio mayores que en ojos normales o cortos.

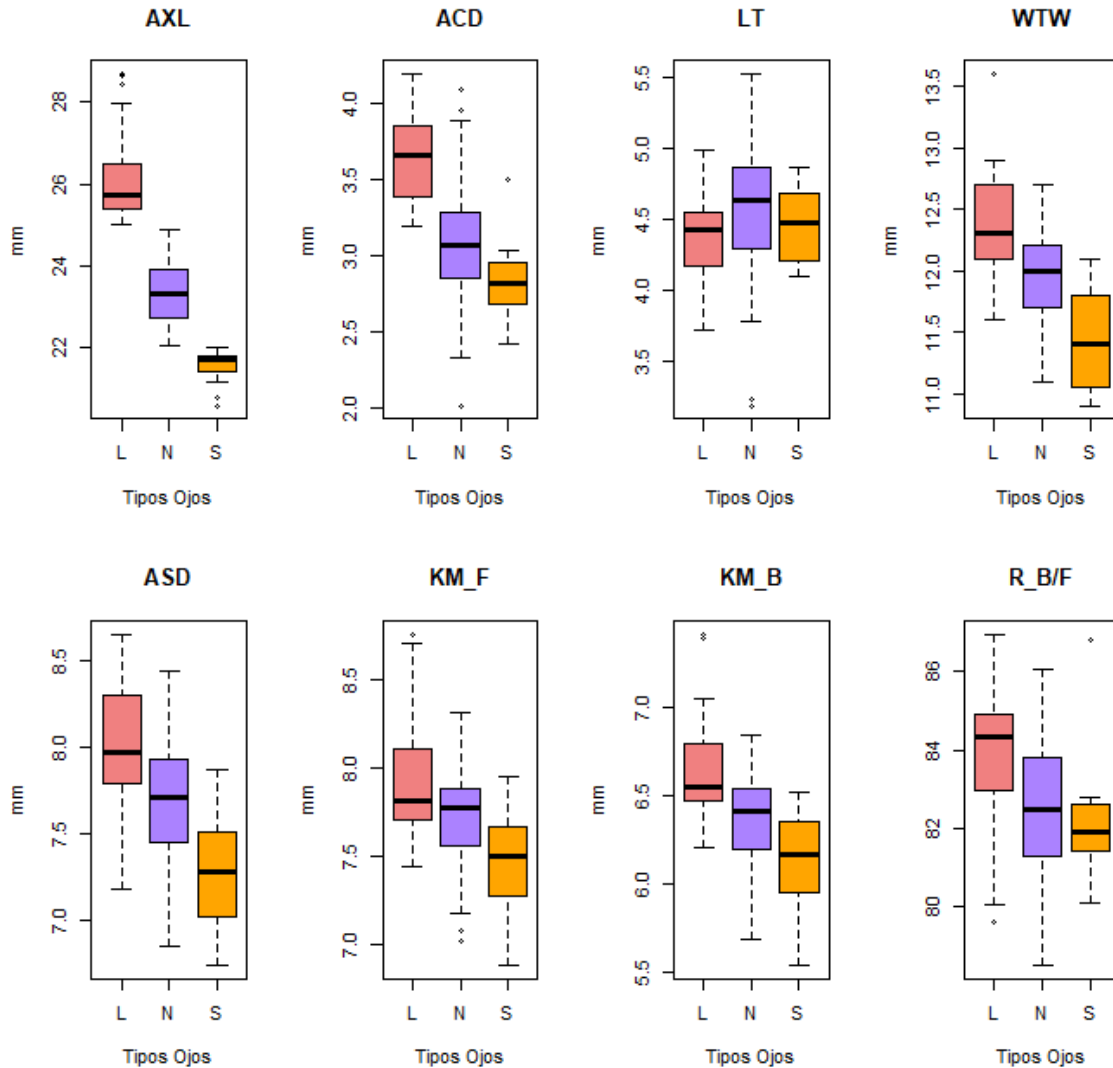


Figura 5. Boxplot de todas las variables biométricas por grupos de longitud axial.

Parámetros	Mean \pm SD	Rango
Ratio B/F	82.80 \pm 1.73	[77.88, 88.04]
AXL (mm)	23.41 \pm 1.51	[19.33, 28.67]
ACD (mm)	3.10 \pm 0.41	[1.96, 4.19]
LT (mm)	4.57 \pm 0.38	[3.18, 5.52]
WTW (mm)	11.94 \pm 0.41	[10.90, 13.60]
K media anterior (D)	7.73 \pm 0.28	[6.68, 8.76]
K media posterior (D)	6.40 \pm 0.28	[5.54, 7.41]
ASD (mm)	7.66 \pm 0.38	[6.96, 8.85]

Tabla 1. Tabla resumen de todas las variables biométricas sin discriminar por longitud axial.

Hacemos un análisis de correlación entre variables (Figura 6), crítico para identificar los mejores predictores para los modelos, así como observar

qué variables tienen relación entre sí y por último identificar colinealidad entre ellas. Para ello se calculan los coeficientes de correlación para cada par de variables, así como las gráficas de dispersión e histogramas para la distribución:

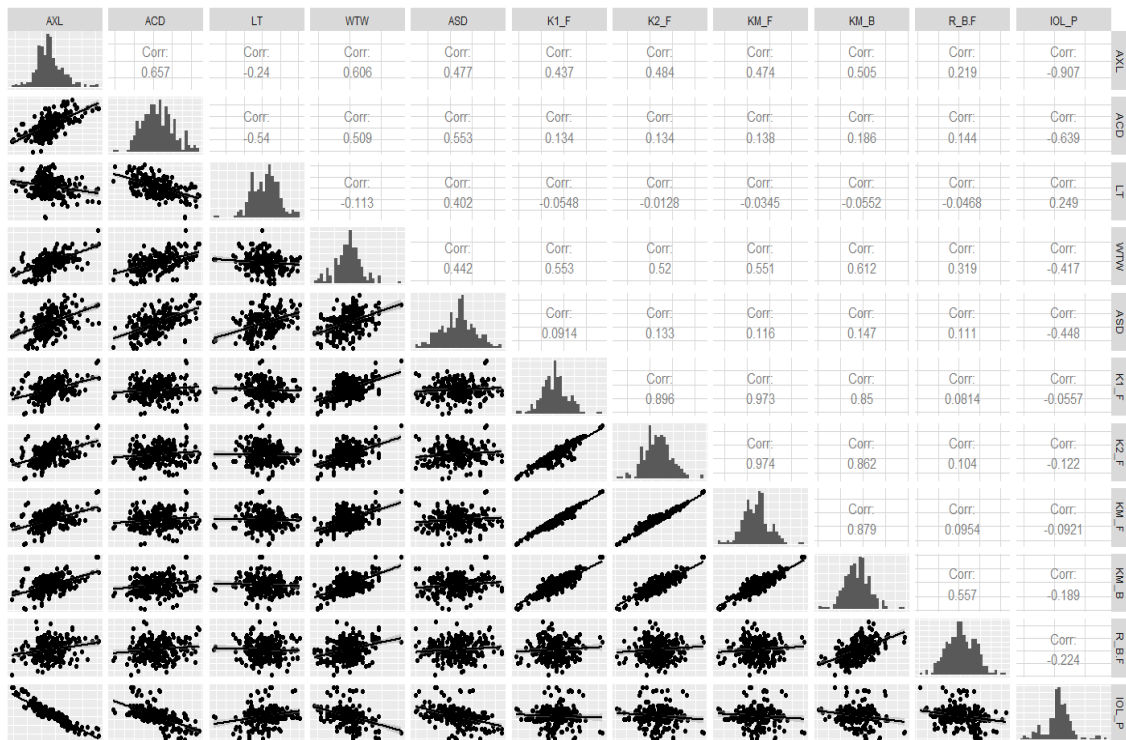


Figura 6. Diagramas de correlación entre variables predictoras.

Vemos que las variables que más relación tienen con la variable respuesta IOL_P son AXL (-0.907) y ACD (-0.639). A su vez ACD y AXL tienen una correlación de 0.657, lo que probablemente implique que podamos prescindir de ACD en el modelo predictivo final. La profundidad de la cámara anterior es una variable que tiene mucho peso específico en la caracterización anatómica del ojo para cálculo de lentes intraoculares, de modo que eliminarla de la ecuación predictiva podría ser arriesgado. Se generarán modelos diferentes, uno que incluya ACD y LT como variables predictoras y otro que incluya ASD que aportará información indirecta de ACD y LT, ya que es una combinación lineal de ambas.

Por otro lado, como KM_F es el promedio de las variables queratométricas K1_F y K2_F, están muy correlacionadas (0.97), lo que implica su eliminación por el aporte escaso de información al modelo. Al mismo tiempo, KM_F y KM_B también son variables correlacionadas (0.88), es por lo que plantearemos distintas estrategias para minimizar el impacto de la colinealidad sobre el modelo. La primera será eliminar KM_B y poder valorar el peso específico de esta en el modelo predictivo. La variable K_B/F indirectamente aporta información de KM_B al modelo, ya que se trata del ratio de curvaturas corneales posterior/anterior (R_Back/Front).

2.3 Transformación de datos

En muchas ocasiones, se hace necesario la transformación de los datos con el fin de obtener los mejores y más eficientes modelos de regresión. Para las tareas de transformación y posterior análisis de regresión, se utilizará la librería caret (Classification And REgression Training) de Max Kuhn [10] en R Studio versión 1.1.423.

A priori, es difícil saber que metodología de preprocesamiento [11] de datos aplicar, aunque existen determinadas *normas* que suelen funcionar bien. Los modelos basados en distancias, como los vecinos cercanos (kNN) por ejemplo, funcionan mejor si las variables de entrada tienen la misma escala. Generalmente, las transformaciones tienen más probabilidades de ser útiles para algoritmos de regresión, k-Nearest Neighbour (kNN), Artificial Neural Network (ANN) o Support Vector Machine (SVM). Sin embargo, para métodos basados en árboles y reglas como por ejemplo Random Forest (RF) y Tree Decision (TD), las transformaciones suelen resultar menos útiles.

No obstante, incluso empleando dichas transformaciones de datos en los modelos predictores, puede que no se mejore la capacidad predictiva de dichos modelos, produciéndose mejores resultados con los datos en crudo. De modo que lo mejor, es chequear la eficiencia del modelo implementado con y sin transformación.

No es necesario imputar variables ausentes, no tenemos variables con varianza próxima a cero y hemos reducido la dimensionalidad del dataset de trabajo. Algoritmos como SVM o ANN son sensibles a la escala de medida de las variables y la magnitud de la varianza. Pero en este caso las unidades de medida son las mismas en todas las variables predictoras.

2.4 Implementación de algoritmos de machine learning

Para la implementación, ajuste, validación y comparación de los modelos se utilizó la librería caret de Max Khun. Este paquete aglutina bajo un único marco, cientos de funciones de diferentes librerías facilitando en gran medida el análisis y evaluación de los modelos propuestos.

Desde una perspectiva regresiva los algoritmos empleados fueron; regresión lineal multivariante, vecinos próximos KNN, modelo general lineal, SVM con kernel radial y lineal, redes neuronales, random forest, árboles de decisión, regresión LASSO, regresión adaptativa multivariante spline y stochastic gradient boosting.

Se compararon los 11 modelos en bruto (Tabla 2) con diferentes combinaciones de variables predictoras, tal y como se apuntó en el apartado 2.2. Un dataset incluía LT y ACD, sin ASD y otro ASD sin LT y ACD. El resultado fue que ninguno de los modelos se resentía al eliminar LT y ACD como variables explicativas. Posteriormente, comparamos de nuevo todos los modelos con otra combinación diferente, en este caso se excluyó KM_B comprobando de nuevo la no afectación en el desempeño de los modelos.

Por tanto, el dataset definitivo con el que se decidió trabajar excluía ACD, LT y KM_B.

Se aleatoriza la muestra total y se divide en dos partes, la de entrenamiento (train) y la de validación (test) siguiendo una proporción 80/20. En los entrenamientos de los modelos propuestos, se utilizan las funciones *method*, *preProcess* y *tuneGrid* de caret para el ajuste, modificando los métodos de remuestreo y el preprocesamiento, dejando los hiperparámetros para el ajuste fino posterior.

Construimos los 11 modelos en bruto con la muestra *train*, obteniendo los siguientes valores de R^2 , RMSE y MAE mostrados en la Tabla 2.

	Rsquared	RMSE	MAE
kNN	0.8039487	2.0768438	1.5441374
LM	0.9721034	0.7215392	0.5501115
GLM	0.9714642	0.7197618	0.5491877
SVMRadial	0.9150324	1.2193712	0.7781057
SVMlineal	0.9736337	0.7173611	0.5379122
ANN	NaN	21.3895085	20.9710529
RF	0.9386487	1.0544459	0.7339256
RPART	0.5906365	2.7326017	2.1024889
GLMNET	0.9726885	0.7206895	0.5483842
SPLINE	0.9786902	0.6233555	0.4856596
GBM	0.9216737	1.2133778	0.8918284

Tabla 2. Valores. R^2 , RMSE y MAE de todos los modelos en bruto (muestra entrenamiento).

Desechamos los siguientes modelos: KNN, ANN, RPART y GBM, por ser los peores según las métricas seleccionadas. Sin embargo, se desarrollaron estos otros: GLM, SVMRadial y lineal, RF, GLMNET y SPLINE para posteriormente compararlos y seleccionar el modelo de mejor desempeño.

En el caso de SVMRadial, a pesar de su R^2 bajo, se quiso hiperparametrizarlo para mejorarlo, porque la fórmula de cálculo biométrico del Dr. W. Hill se basa en este gaussiano.

Entre los mejores modelos se encuentra LM o modelo de regresión lineal multivariante, con un 97.21% de explicación de varianza. El problema de este modelo es el incumplimiento de alguna hipótesis.

En la validación de condiciones para la regresión múltiple lineal, se comprobó la relación lineal entre los predictores numéricos y la variable respuesta IOL_P. Si la relación es lineal, los residuos (Figura 8) deben distribuirse aleatoriamente en torno a 0 con una variabilidad constante a lo largo del eje X:

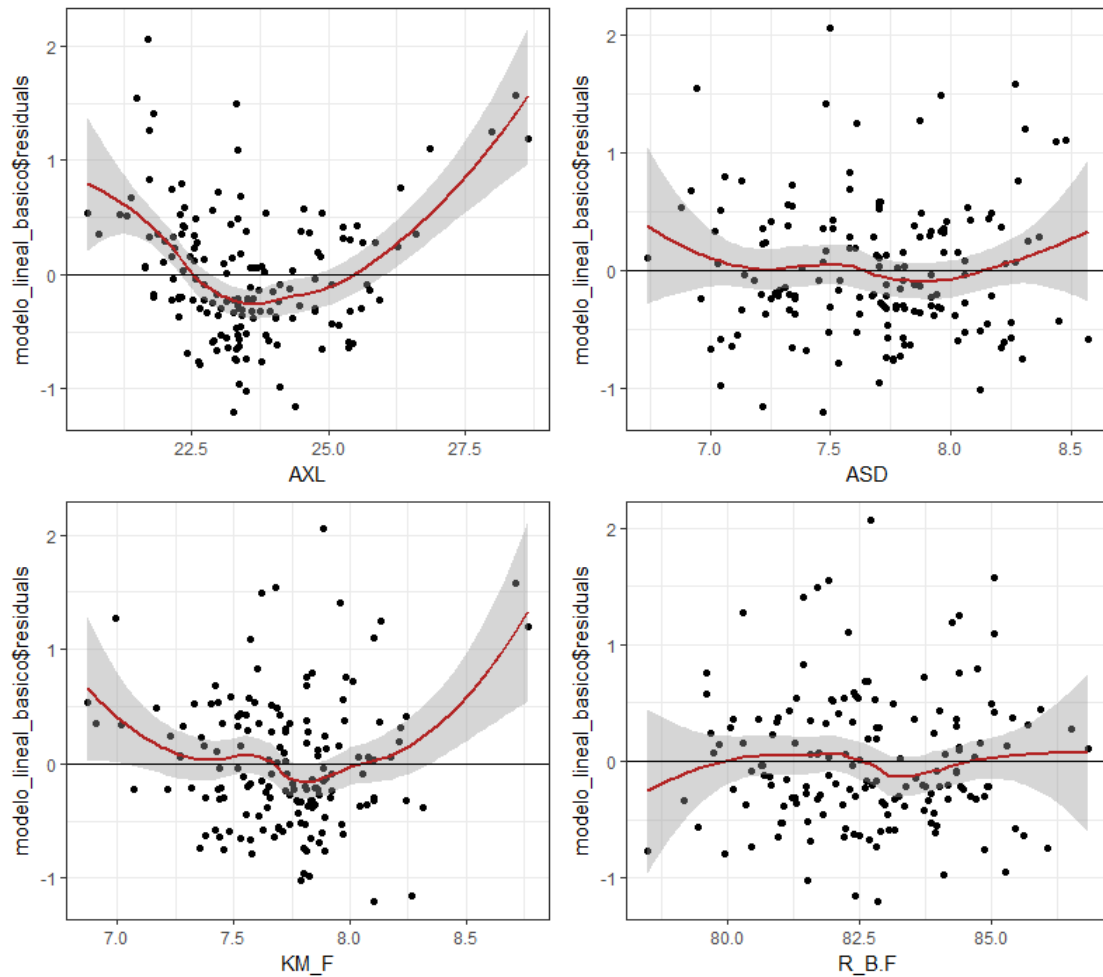


Figura 8. Variables predictoras frente a residuos en modelo lineal.

En la variable AXL y KM_F se observa que la distribución residual no cumple el criterio. En el QQ-plot y el test Shapiro-Wilk (Figura 9) se confirma la no normalidad de la distribución de los residuos:

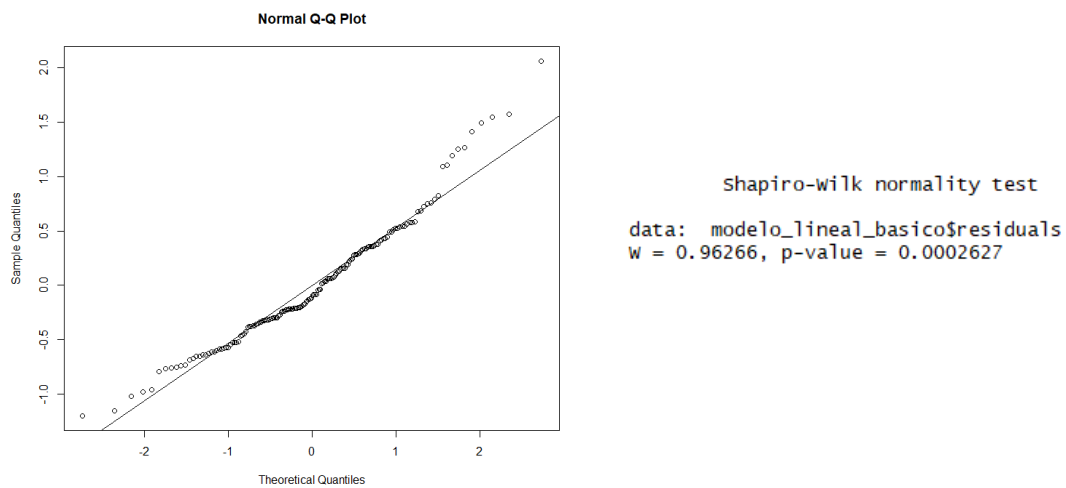


Figura 9. QQplot y Test de Shapiro-Wilk para contrastar normalidad.

Por lo tanto, desechamos la posibilidad del modelo lineal básico multivariante por incumplimiento de la hipótesis de normalidad, siendo los estimadores mínimos-cuadráticos ineficientes. No obstante, se podría utilizar alguna transformación para conseguir normalidad, pero optamos por el análisis de otros modelos que son laxos en este particular.

Otro modelo bueno con un 97.15% de explicación de varianza (Tabla 2), es el modelo generalizado lineal o *generalized linear model* (GLM). Los GLM son una extensión de los modelos lineales que permiten utilizar distribuciones no normales de los errores (binomiales, Poisson, gamma, etc.) y varianzas no constantes.

En este caso, el método de remuestreo que mejores resultados obtuvo fue validación cruzada con repetición “repeatedcv” con 10 particiones o folds y 5 repeticiones. Se probaron diferentes metodologías de preprocesamiento “center”, “scale”, “range”, “BoxCox”, “pca”, no siendo necesario en este dataset tratar las observaciones de ningún modo. Caret no ofrece la posibilidad de utilizar ajuste de parámetros (*tuneGrid*) en este tipo de modelos.

En cuanto a los modelos brutos propuestos basados en máquinas de soporte vectorial (SVM), se puede comprobar (Tabla 2) que si utilizamos un kernel radial obtenemos un R^2 de 0.9150. Sin embargo, si el kernel es lineal conseguimos un modelo que explique el 97.36% de la varianza.

Con SVM cuando los datos no se pueden separar linealmente, se hace una transformación no lineal a un espacio de dimensión mayor donde sí se puedan separar linealmente. Para esto, se usan funciones núcleo (kernel), que calculan el producto escalar en el espacio de características. Caret nos ofrece la posibilidad de ajustar los hiperparámetros sigma y C (cost) de suavizado y penalización respectivamente en el modelo con kernel radial y solo C en el lineal, mediante la función *tuneGrid*.

Mediante el parámetro *tuneLength* de la función *train()* podemos establecer el número de niveles para cada parámetro de tuning, si no se han especificado valores concretos con *tuneGrid*. Para obtener unos valores aproximados de sigma y C de manera automática por validación cruzada es la mejor opción, y posteriormente a través de *tuneGrid* construiremos un dataframe en el entorno de sigma y C detectados en el paso anterior como propone Hsu et al [14].

Se procede con esta técnica en SVM con kernel RBF y posteriormente con núcleo lineal. Tal y como se puede comprobar en la figura 10b, el modelo SVM radial bruto arroja un valor *cost* de 4 y un valor sigma de 0.2058. En la figura 10a se puede comprobar que usando los hiperparámetros sigma y C con la función *tuneGrid*, hemos mejorado el modelo sustancialmente, obteniendo un R^2 de 0.9788.

MODELO_SVMRadial hiperparámetros					MODELO_SVMRadial sin hiperparámetros			
sigma	C	RMSE	Rsquared	MAE	C	RMSE	Rsquared	MAE
0.0161	16.525	0.6240448	0.9788389	0.4830690	0.25	2.214559	0.7977909	1.2763230
0.0161	16.530	0.6239304	0.9788432	0.4829397	0.50	1.875576	0.8414264	1.0929521
0.0161	16.535	0.6239829	0.9788408	0.4830058	1.00	1.564478	0.8794489	0.9407611
0.0162	16.525	0.6241835	0.9788319	0.4831847	2.00	1.268969	0.9110292	0.8092560
0.0162	16.530	0.6241467	0.9788320	0.4831428	4.00	1.219371	0.9150324	0.7781057
0.0162	16.535	0.6241297	0.9788358	0.4831343	8.00	1.225436	0.9147531	0.7867885
0.0163	16.525	0.6245243	0.9788109	0.4835389	16.00	1.255223	0.9104290	0.8086468
0.0163	16.530	0.6244371	0.9788186	0.4834719	32.00	1.292385	0.9051586	0.8419537
0.0163	16.535	0.6244535	0.9788171	0.4834989	64.00	1.343664	0.8986648	0.8910119

Figura 10. Valores R^2 , RMSE y MAE de SVM Kernel Radial en bruto y con hiperparámetros.

Sin embargo, cuando se empleó el kernel lineal con hiperparámetros, no mejoró en términos de RMSE y R^2 .

Otro de los modelos implementados que consiguió resultados aceptables fue random forest (RF), obtuvo un R^2 de 0.9386.

Random Forest [15] es una técnica de machine learning que consiste en la construcción de un gran número de árboles con las siguientes características:

- Está basado en muestreo bootstrap. Cada árbol está basado en una muestra aleatoria con reemplazo de las observaciones.
- Cada división del árbol está basada en una muestra aleatoria de los predictores.
- Los árboles no se cortan, son tan largos como sea posible.

Caret tiene integrado el paquete *randomForest* con su método "rf" que solo permite optimizar un hiperparámetro, el número de predictores (*.mtry*) que son evaluados en cada división. Por otro lado, caret a través del método "ranger" incluido en la librería *ranger*, nos da la posibilidad de cambiar tres hiperparámetros:

- *mtry*: número de predictores seleccionados aleatoriamente en cada selección.
- *min.node.size*: tamaño mínimo que tiene que tener un nodo para poder ser dividido.
- *splitrule*: criterio de división.

Ajustando los 3 hiperparámetros y estableciendo el número de árboles en 700, para no consumir mucho tiempo de computación, se mejoró el modelo bruto precursor ligeramente, consiguiendo un R^2 de 0.9537.

GLMNET es una librería que implementa el modelo regresivo conocido como LASSO perteneciente a los modelos de regresión penalizada (Penalized Regression). Esta metodología permite reducir la varianza además de hacer selección de predictores a diferencia del método Ridge.

Los hiperparámetros que podemos modificar con caret en estos modelos son, Alpha y lambda. Lambda determina el grado de penalización del modelo y es encontrado automáticamente por validación cruzada. Optimizando estos

parámetros a través de *tuneGrid* obtenemos un R^2 de 0.9726, lo que significa que no mejoramos respecto al modelo bruto GLMNET.

Por último, de los modelos sin depurar seleccionados para optimización tenemos SPLINE, que pertenece a Multivariate Adaptive Regression Spline (MARS) [16] [17] implementado por Friedman en los años 90. Es un método de modelación no paramétrico que extiende el modelo lineal incorporando no linealidades e interacciones de variables. Este fue el mejor de todos en términos de precisión con un R^2 0.9787 en la muestra de entrenamiento.

Los modelos MARS son una combinación de regresión polinómica y *step functions*. Mediante esto último se divide el rango del predictor en k subintervalos de forma que en cada uno se emplean únicamente las observaciones que pertenecen a la región para ajustar el modelo. Gracias a los polinomios podemos introducir curvatura al modelo, debido a la generación de nuevos predictores, que se obtienen al elevar todos o algunos de los predictores originales a distintas potencias.

Los hiperparámetros que *caret* nos permite modificar en estos modelos son dos, *nprune* y *degree*. *Nprune* correspondería a los subintervalos k de división y *degree* al grado polinómico que se quiere en el modelo. *Caret* como con todos los modelos con los que trabaja, obtiene por validación cruzada los mejores valores para ambos hiperparámetros, aunque se pueden ajustar y modificar posteriormente.

No se consigue mejorar el modelo inicial SPLINE con un R^2 de 0.9787. Se obtuvo dicho valor mediante un spline polinómico de primer grado con 8 subdivisiones. Se probaron polinomios cuadráticos y cúbicos, no mejorando la precisión y consumiendo mucho tiempo computacional. El método de remuestreo que mejores resultados obtuvo fue validación cruzada sin repetición.

La Figura 11 muestra gráficos de normalidad, los dos de abajo, como se distribuyen los residuos por cada respuesta del modelo y un QQplot. En los de arriba, empezando por la derecha se muestra el modelo seleccionado en función de los *nprunes* y del Generalized R Squared (GRSq), así como la distribución acumulada de los valores de los residuos en función de los cortes (*nprunes*) en el de la izquierda. En este último plot, en el eje vertical por cada $1/n$ (n = casos entrenados) en el eje horizontal nos movemos al siguiente residuo absoluto más grande.

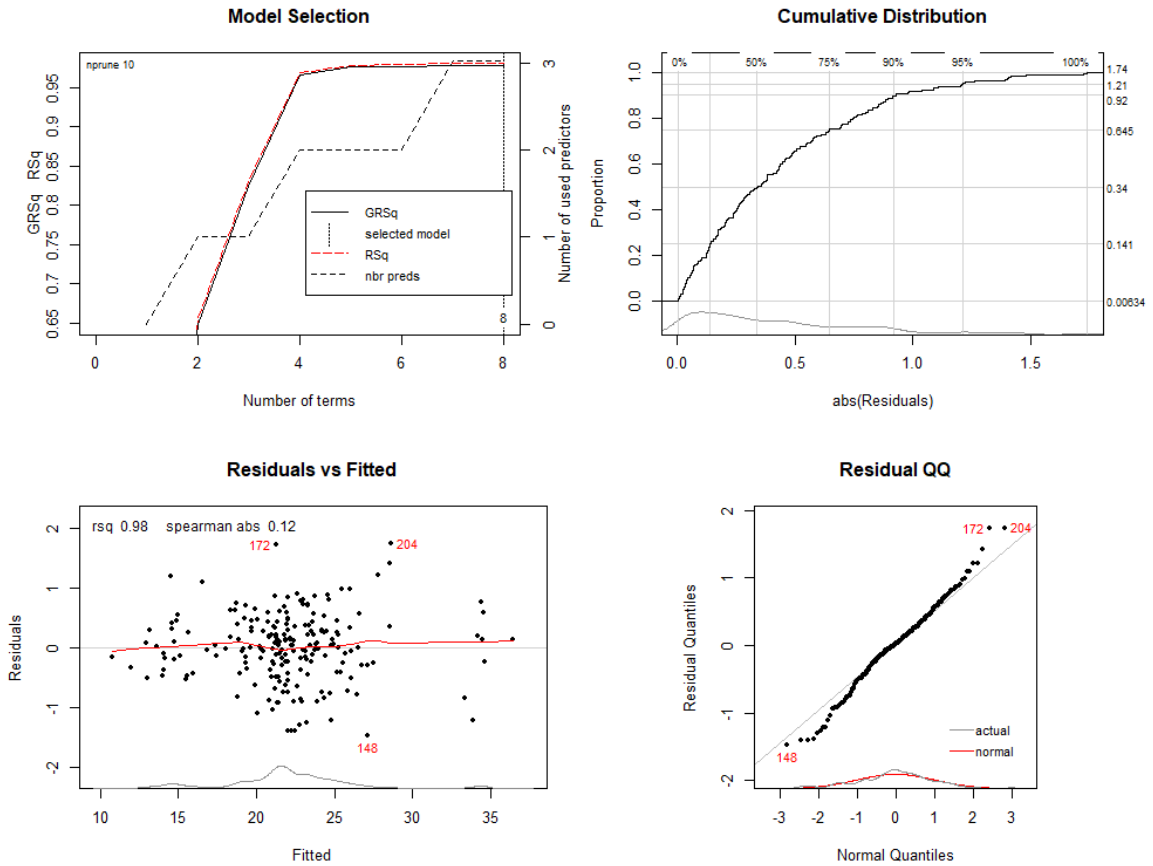


Figura 11. Plots de normalidad, selección del modelo y distribución acumulada.

3. RESULTADOS

3.1 Comparando modelos

A continuación, se comparan los 5 modelos (Figura 12) seleccionados para el desarrollo con hiperparámetros y entrenados con la muestra *train*. Las métricas utilizadas son R^2 , RMSE y MAE.

Hay dos modelos claramente superiores al resto, SPLINE y SVMRadial, con un R^2 de 0.9788 y 0.9774 respectivamente, obtenidos mediante MARS y SVM con gaussiano RBF. El modelo implementado mediante Random Forest es notablemente el peor de todos, con un coeficiente de determinación de 0.9542. El tercer modelo obtenido con SVM y kernel lineal, explica un 97.32% de la varianza de la muestra de entrenamiento.

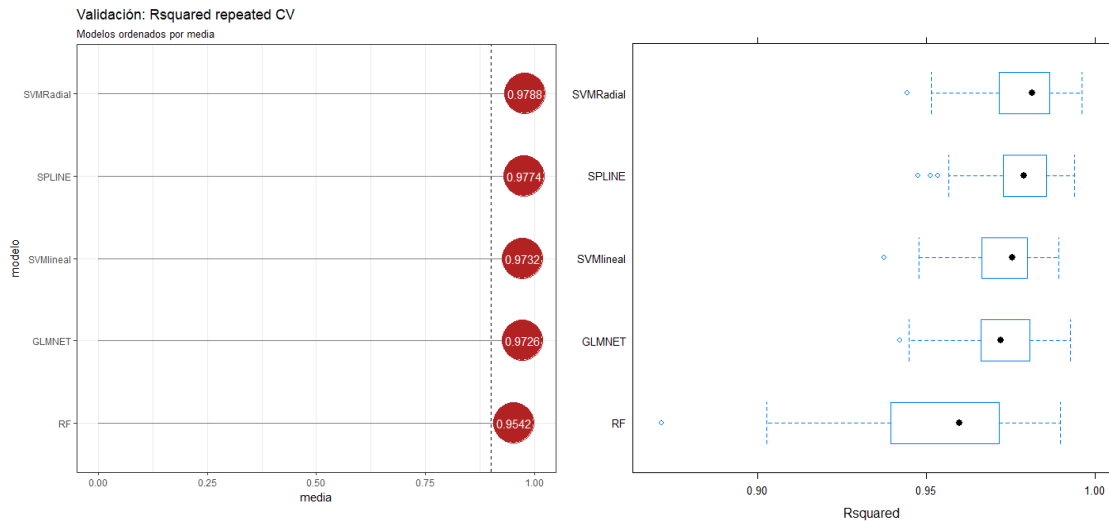


Figura 12. Comparación de modelos.

Las métricas obtenidas por validación cruzada de forma automática con caret, son estimaciones de la capacidad que tiene el modelo para predecir nuevas observaciones. Pero como cualquier estimación tiene asociada una varianza. Por tanto, para aseverar si un modelo es mejor que otro, además de observar los máximos o mínimos (dependiendo de la métrica), es necesario analizar la varianza.

Como la comparación dos a dos la haremos entre los 5 modelos, necesitamos un test paramétrico t-test con ajuste de Bonferroni. Caret implementa una función para este cometido llamada *diff()*. Los resultados para las tres métricas de esta comparación se muestran en la siguiente tabla:

Upper diagonal: estimates of the difference Lower diagonal: p-value for H0: difference = 0					
MAE					
	SVMRadial	SVMlineal	RF	GLMNET	SPLINE
SVMRadial		-0.05461	-0.20122	-0.06562	-0.01257
SVMlineal	0.00174		-0.14662	-0.01102	0.04203
RF	2.977e-09	2.694e-05		0.13560	0.18865
GLMNET	0.01320	1.00000	2.731e-05		0.05305
SPLINE	1.00000	0.19842	1.617e-08	0.04328	
RMSE					
	SVMRadial	SVMlineal	RF	GLMNET	SPLINE
SVMRadial		-0.096432	-0.347427	-0.099504	-0.017776
SVMlineal	7.550e-05		-0.250996	-0.003072	0.078655
RF	4.242e-09	9.242e-06		0.247923	0.329651
GLMNET	0.001523	1.000000	6.071e-06		0.081727
SPLINE	1.000000	0.020031	1.063e-08	0.010180	
Rsquared					
	SVMRadial	SVMlineal	RF	GLMNET	SPLINE
SVMRadial		0.0056095	0.0245311	0.0061600	0.0013760
SVMlineal	0.0598647		0.0189216	0.0005505	-0.0042335
RF	2.138e-07	0.0001132		-0.0183711	-0.0231551
GLMNET	0.0368417	1.000000	6.276e-05		-0.0047840
SPLINE	1.000000	0.5875694	3.723e-06	0.4403622	

Tabla 3. Contraste dos a dos con Bonferroni.

A la vista de las comparaciones múltiples, se puede asegurar que todos los modelos son mejores en todas las métricas de precisión (MAE, RMSE y R^2) que el construido con Random Forest. Sin embargo, las diferencias entre los demás no son estadísticamente significativas (Tabla 3). Por tanto, se puede asegurar que no existen evidencias suficientes para considerar que la capacidad predictiva de los modelos SVMRadial, SVMlineal, GLMNET y SPLINE es distinta.

3.2 Validación externa de los modelos

Aunque está demostrado que las metodologías de validación de un modelo mediante técnicas de validación cruzada o bootstrapping, consiguen estimaciones muy buenas del error que comete el modelo, es conveniente hacer una predicción final con nuevas observaciones para asegurar que durante la optimización no se haya generado *overfitting* o sobreajuste.

Caret tiene una función llamada *extractPrediction ()* para este cometido cuando se trabaja con varios modelos en paralelo. La Tabla 4 muestra la comparación de las tres métricas en los modelos validados con la muestra de entrenamiento y las predicciones hechas con la muestra de test.

	RMSE-train	Rsq-train	MAE-train	RMSE-test	Rsq-test	MAE-test
SVMRadial	0.6212077	0.9787730	0.4832524	0.5600179	0.9846905	0.4374304
SPLINE	0.6233555	0.9786902	0.4856596	0.5781456	0.9833987	0.4488461
GLMNET	0.7207115	0.9727174	0.5485444	0.7570984	0.9716138	0.5566454
SVMlineal	0.7176392	0.9731600	0.5378587	0.7611950	0.9714026	0.5519271

Tabla 4. Comparación de métricas RMSE, R^2 y MAE en train y test en los mejores modelos.

Todos los modelos predicen mejor en la validación externa con la muestra test, que en el proceso de entrenamiento. No obstante, los mejores son SVMRadial y SPLINE (MARS).

En la Figura 13 se pueden comprobar las observaciones (eje y) contra las predicciones (eje x) en todos los modelos tanto en la muestra de entrenamiento *train* como en la de validación *test*. Todos los modelos, excepto SPLINE y SVMRadial, predicen en la validación externa con la muestra *test* desviándose respecto al entrenamiento, siendo más evidente en el modelo RF.

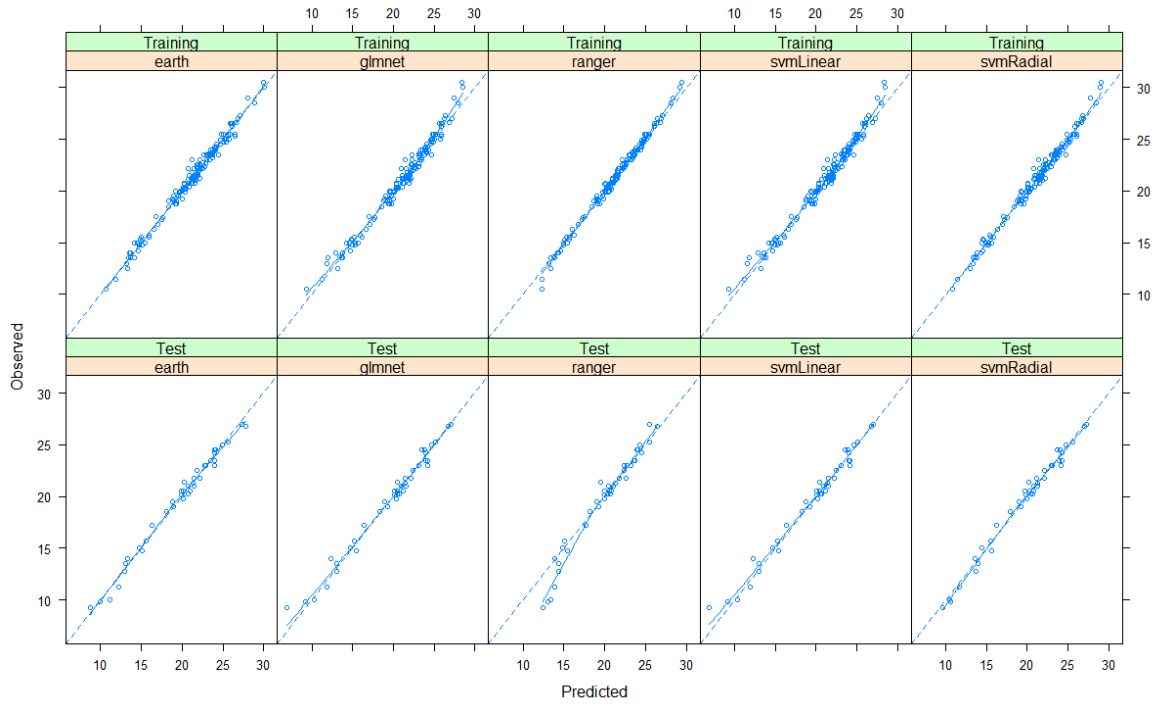


Figura 13. Plots predicción/observación entre train y test en los mejores modelos.

En la Figura 14, se observa la importancia o peso específico de cada variable dentro de los modelos mejores mediante la función `varImp()` de `caret`.

	SVMRadial	SPLINE
AXL	0.84311	100.000
ASD	0.20547	3.715
WTW	0.13221	0.000
R_B.F	0.11328	0.000
KM_F	0.05797	39.806

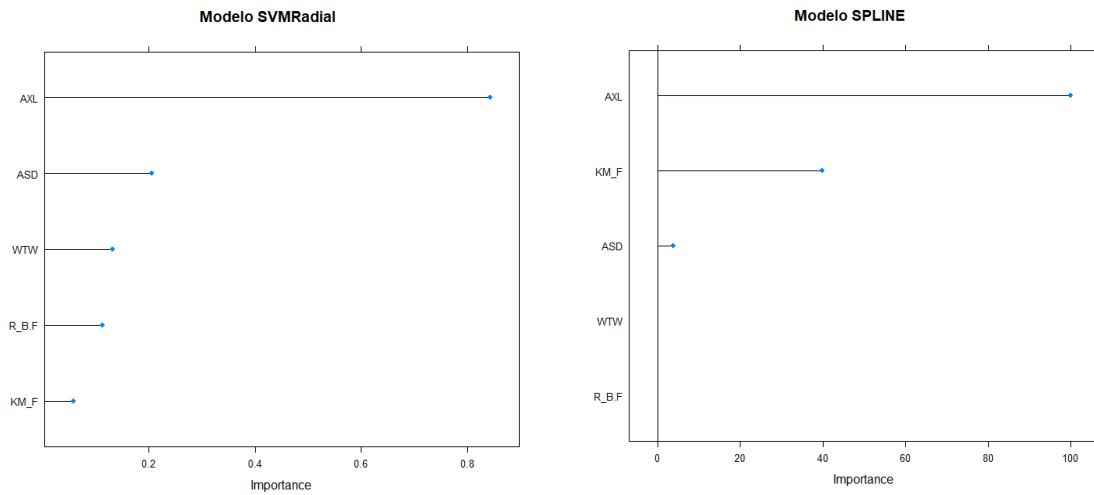


Figura 14. Importancia de las variables predictoras en los mejores modelos.

Como se puede comprobar, ambos modelos tienen a la variable predictora AXL como la más importante. En segundo lugar, SPLINE otorga más peso específico dentro del modelo a la queratometría de la cara anterior corneal (KM_F). Sin embargo, la profundidad del segmento anterior (ASD) es la segunda más relevante para el modelo regresivo SVM con gaussiano RBF.

Para el modelo SPLINE el diámetro corneal horizontal (WTW) y el ratio de cara corneal posterior/anterior (R_B/F) no se tienen en cuenta como predictores.

En el modelo SVMRadial, es importante destacar la inclusión del ratio corneal por delante de la queratometría media simulada, ya que este puede ser un factor diferencial respecto a las fórmulas de cálculo biométrico actuales que no lo incorporan como input.

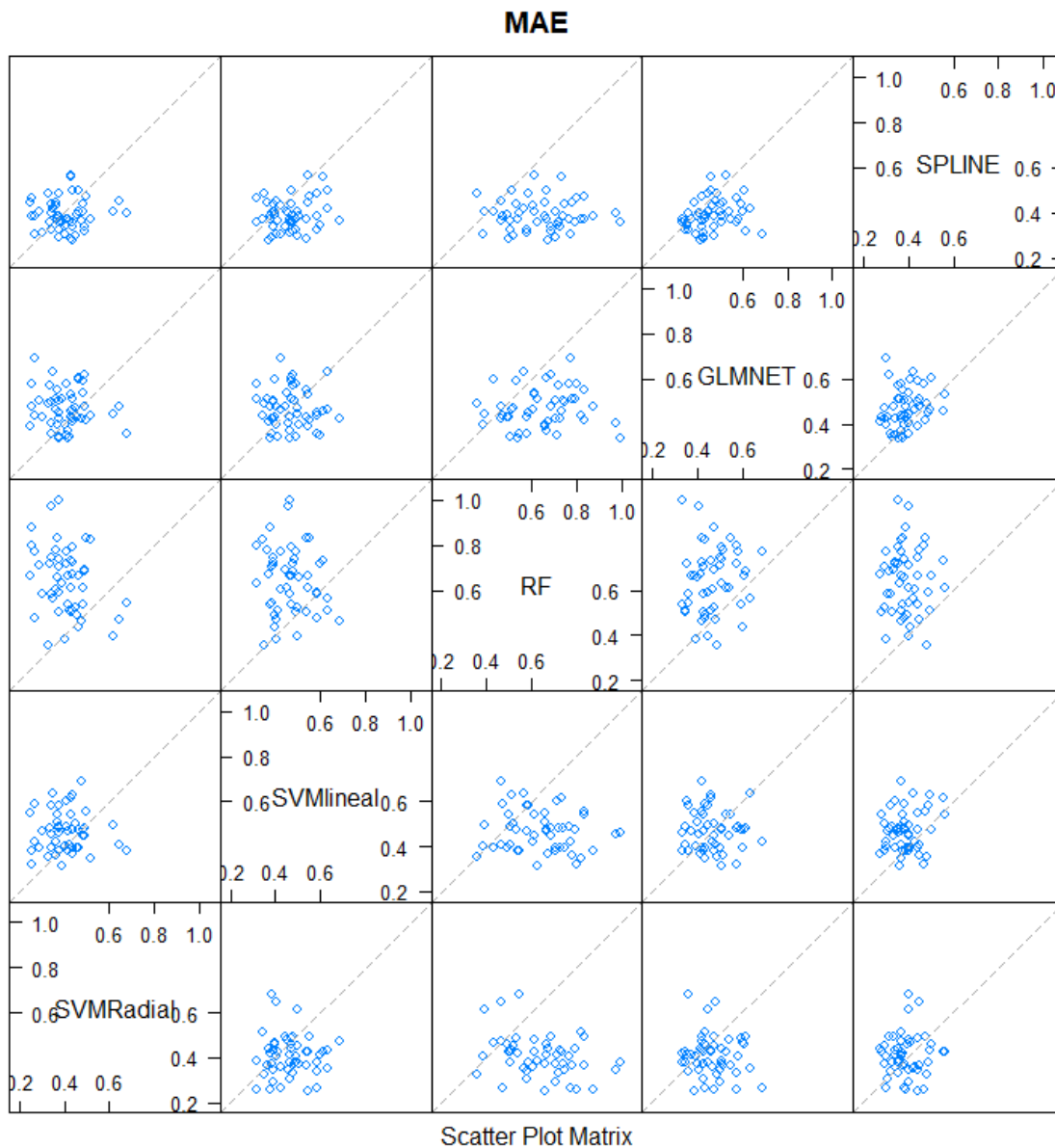
3.3 Ensamblaje de modelos

Si distintos modelos entrenados independientemente han obtenido muy buenos resultados con estrategias predictoras diferentes, combinando dos o más modelos muy buenos, se obtendrá una descripción más detallada de la verdadera estructura subyacente en las observaciones, a esto se le conoce como *model ensembling*.

En primer lugar, se comprueba la correlación entre los mejores modelos del apartado anterior; SPLINE, SVMRadial, GLMNET, RF y SVMLineal, tanto gráfica como analíticamente. No existe ninguna correlación entre ninguno de los modelos, de modo que a priori podríamos asociarlos para potenciar la predicción.

Para el ensamblaje de modelos se utilizará la técnica *stacking*, en la que combinaremos predicciones de dos modelos primarios para construir un modelo supervisor que aprenda la mejor manera de recombinar esas predicciones.

Los modelos seleccionados para el *stacking* serán SVMRadial ($R^2 = 0.9788$) y SPLINE-MARS ($R^2 = 0.9787$). En la Figura 15 se observa la incorrelación entre ambos, condición fundamental para aplicar esta técnica de ensamblaje, así como son predictores por regresión con estrategias algorítmicas muy dispares, otra de las condiciones para implementar *stacking*.



	SVMRadial	SVmlineal	RF	GLMNET	SPLINE
SVMRadial	1.00000000	-0.008337091	-0.308600865	-0.10838420	0.054253418
SVmlineal	-0.008337091	1.00000000	-0.121577542	-0.06166902	0.180795211
RF	-0.308600865	-0.121577542	1.00000000	0.100006025	-0.005970942
GLMNET	-0.108384202	-0.061669024	0.100006025	1.00000000	0.212340125
SPLINE	0.054253418	0.180795211	-0.005970942	0.21234012	1.00000000

Figura 15. Diagrama y matriz de correlaciones entre modelos.

Se construyó un *data.frame* con los resultados de predicción en la muestra test (52 samples) de los dos modelos primarios, SVMRadial y SPLINE respectivamente y los valores reales de IOL_P. El método de regresión seleccionado tras probar varios, fue *glmnet* con regresión penalizada LASSO.

Como método de resampling finalmente se utilizó LOOCV (leave-one-out cross validation) también conocido como “jackknife”, tras haber probado *bootstrapping*, estimación out-of-bag (oob) y *cross-validation* con y sin repetición.

LOOCV implica separar los datos de forma que para cada iteración tengamos una sola muestra para los datos de prueba y el resto conformando los datos de entrenamiento. En la Tabla 5 se pueden ver los resultados:

lambda	RMSE	Rsquared	MAE
0.003	0.6144099	0.9861090	0.4879840
0.013	0.6102468	0.9863002	0.4827170
0.023	0.6105309	0.9862951	0.4819089
0.033	0.6113995	0.9862658	0.4820905
0.043	0.6121798	0.9862428	0.4825643
0.053	0.6129177	0.9862241	0.4831950
0.063	0.6136223	0.9862091	0.4839217
0.073	0.6143332	0.9861962	0.4847084
0.083	0.6150598	0.9861849	0.4855373

Tabla 5. Resultados del modelo stacking.

Con el ensamblaje de dos modelos, SVMRadial y SPLINE (MARS), se consigue mejorar la potencia predictora, ya que para una lambda de 0.013 se obtiene un R^2 de **0.9863**, un RMSE de **0.6102** y un MAE de **0.4827**.

3.4 Selección del modelo definitivo

La elección está entre tres modelos; SVM con gaussiano RBF, SPLINE basado en MARS con polinomio de primer grado y el ensamblado por *stacking* de estos dos anteriores. En todos los casos el desempeño y las métricas conseguidas son similares.

En la Figura 16 se puede comprobar que no existen diferencias estadísticamente significativas, salvo entre ENSEMBLE y SVMRadial ($p < 0.05$).

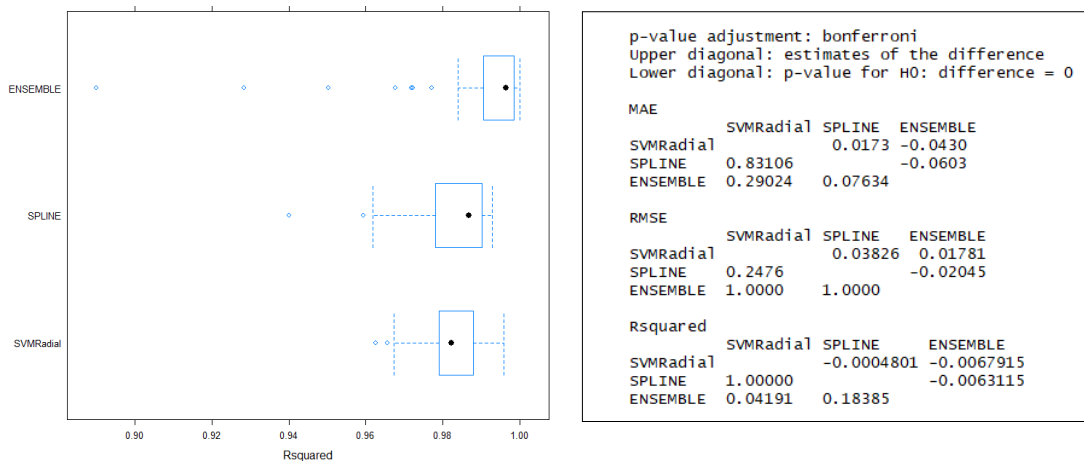


Figura 16. Boxplot de R^2 y contraste de hipótesis dos a dos en los tres modelos.

Teniendo en cuenta que en la comparación dos a dos con Bonferroni, existe diferencia estadísticamente significativa entre el modelo SVMRadial y el modelo combinado por *stacking*, además de mejorar la métrica R² con este último. Y que por otro lado, la información biométrica que aportan los modelos implementados al ensamblado es completa, se selecciona como modelo definitivo de predicción el combinado que hemos llamado ENSEMBLE.

3.5 Comparaciones del modelo con las fórmulas biométricas reales.

Se compararon los resultados de predicción del modelo definitivo obtenido en este trabajo en la muestra test, con los resultados obtenidos por cuatro fórmulas biométricas; Holladay II, Haigis, Barrett Universal II y Hill-RBF, utilizadas habitualmente en la práctica clínica diaria.

Para llevar a cabo esta comparativa tal y como se establece en el protocolo para los estudios de precisión en fórmulas biométricas [18], se calcularon 3 parámetros para cada fórmula y para el modelo combinado: el Error Predictivo (EP), el Error Absoluto Medio (EAM) y la Mediana del Error Absoluto (MedEA), según las siguientes expresiones:

$$EP = \frac{\sum(SE_{POST} - TARGET_{FORMULA})}{n}$$

$$EAM = \frac{\sum|SE_{POST} - TARGET_{FORMULA}|}{n}$$

$$MedAM = \text{mediana de EA}$$

Para calcular el EP del modelo definitivo se procedió de la siguiente manera. Al valor de IOL_P que se corresponde con el valor de lente intraocular que debería implantar sin cometer errores refractivos residuales postoperatorios, se le resta el valor de IOL_P obtenido por ENSEMBLE. De este modo, se obtiene el error refractivo cometido por el modelo en el plano capsular. Para obtener el error predictivo final del modelo combinado o EP, se aplica distometría saco capsular-córnea para conseguir la equivalencia de dicho error a nivel corneal para poder realizar la comparación con todas las fórmulas biométricas.

La distometría o equivalencia de potencia dióptrica entre planos se hizo a través de la siguiente fórmula:

$$Potencia_{cornea} = \frac{1}{ELP * 10^{-3} + \frac{n_{humor\ acuoso}}{Potencia_{saco}}}$$

Siendo ELP la posición efectiva de la lente, es decir, la posición final respecto al epitelio corneal en la que se ubica el implante intraocular en el saco del cristalino. Obsérvese que ELP medida en milímetros, se pasa a metros en la fórmula puesto que las potencias están expresadas en dioptrías. $n_{\text{humor acuoso}}$ es el índice de refracción del humor acuoso, aproximadamente 1.336.

La ELP depende de muchos factores no medidos en este trabajo, pero como se necesita para llevar a cabo la distometría, se aproxima de la siguiente manera, no influyendo en el resultado de equivalencia como se verá más adelante. La lente quedará posicionada en el saco capsular a una distancia menor que ASD (ACD + LT). Además, teniendo en cuenta la anatomía del cristalino, la lente quedará siempre por delante de la mitad del espesor del cristalino, de modo que podemos deducir la ELP con la siguiente fórmula:

$$ELP = ACD + \frac{LT}{2.5}$$

Esta aproximación no influye en la distometría, ya que las distancias entre planos no son superiores a 9 mm y además las potencias refractivas residuales son muy pequeñas, del orden de 0.50 D.

El EP para el resto de fórmulas biométricas ya está calculado en el dataset inicial, las variables correspondientes para cada fórmula son: MEP_HOL2, MEP_HAI, MEP_BAR y MEP_HILL respectivamente. El EAM es el valor absoluto de EP.

Como se comentó en el primer apartado de este trabajo, procederemos a la conversión de los promedios de EP de todas las fórmulas a 0.00 D [18] [19], para eliminar el error sistemático derivado de las diferentes constantes utilizadas en el cálculo y sus posibles no optimizaciones.

Además, se calcularon los porcentajes de ojos con lente intraocular implantada con resultados refractivos dentro de los siguientes intervalos: ± 0.50 D; ± 0.75 D y ± 1.00 D.

Por último, para contrastar los EAM y los MedAM de cada fórmula se utilizó un test no paramétrico de Friedman debido a la distribución no normal de la muestra en estas categorías, con un análisis post-hoc mediante una prueba de Nemenyi para las comparaciones múltiples.

En la Tabla 6 se observan todos los valores de EP, EAM y MedAE con sus desviaciones estándar, rangos máximos y mínimos de todas las fórmulas biométricas y el modelo combinado ENSEMBLE.

Fórmula	EP ± DE (D)	Rango	EAM ± DE (D)	Rango	MedAM ± DE(D)	Rango
ENSEMBLE	0.00±0.43	-1.03/1.07	0.34±0.26	0.00/1.07	0.34±0.26	0.00/1.07
Holladay II	0.00±0.38	-1.55/0.68	0.28±0.26	0.00/1.09	0.24±0.23	0.00/1.09
Haigis	0.00±0.36	-1.20/1.04	0.28±0.23	0.01/1.20	0.23±0.23	0.01/1.20
Barrett U.	0.00±0.38	-1.09/0.77	0.29±0.25	0.00/1.94	0.22±0.25	0.00/1.94
Hill-RBF	0.00±0.40	-0.73/0.97	0.30±0.25	0.00/0.97	0.26±0.25	0.00/0.97

Tabla 6. EP, EAM y MedAE de todas las fórmulas y el modelo ENSEMBLE.

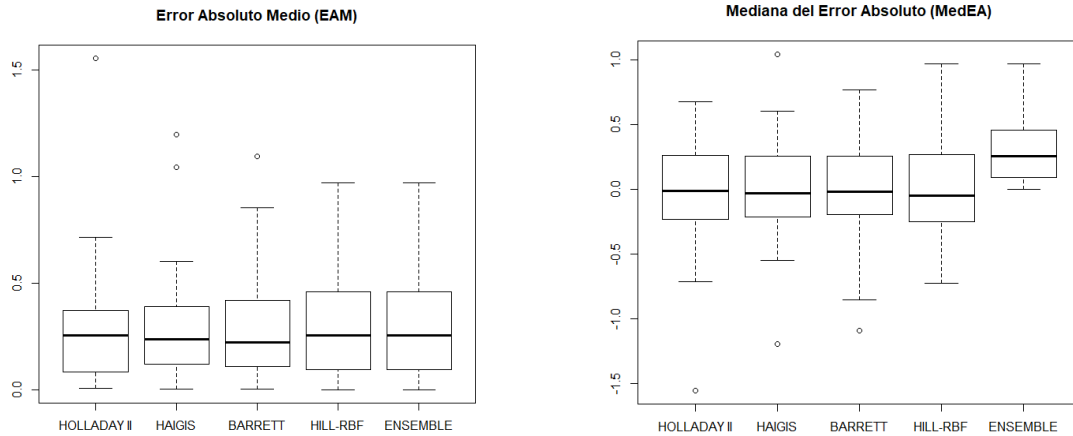


Figura 17. Boxplots de EAM y MedEA del modelo ENSEMBLE y el resto de fórmulas biométricas.

El error predictivo (EP) del modelo propuesto sin tener en cuenta el ajuste por constantes de lentes, es el mejor ya que está en cero, lo que implica que las predicciones en término medio no tienen una tendencia ametrópica, como sí la tienen suavemente hacia la miopía las 4 fórmulas de cálculo que se han usado para la comparación.

Otro factor a tener en cuenta es la desviación estándar, es la mayor en todas las categorías analizadas, aunque sin grandes diferencias (0.06 D respecto a la menor), lo que significa una mayor dispersión en las predicciones. Comprobamos este hecho en el análisis de los rangos max/min del modelo propuesto, donde se puede constatar que es el que mayor error comete en rango hipermetrópico (1.07 D), siendo Haigis la segunda con un error máximo de 1.04 D (Tabla 6). Respecto al mayor error miópico cometido por ENSEMBLE -1.03 D cuando Holladay II tiene un -1.55 D o Haigis un -1.20 D.

En las otras categorías, EAM y MedEA, el modelo final también es ligeramente peor, arrojando los valores más altos respecto a las fórmulas de comparación, 0.34 ± 0.26 D. Esto implica la obtención de mayores residuos refractivos postoperatorios en términos absolutos. No obstante, la diferencia en términos absolutos respecto a la fórmula Hill-RBF que se basa en técnicas similares de aprendizaje automático de máquinas es solo de 0.04 D. Respecto a Haigis que es la mejor en términos de EAM, la diferencia con el modelo propuesto es solo de 0.06 D.

Por último, en la Figura 18 se puede comprobar el porcentaje de ojos que están dentro del rango ± 0.50 D:

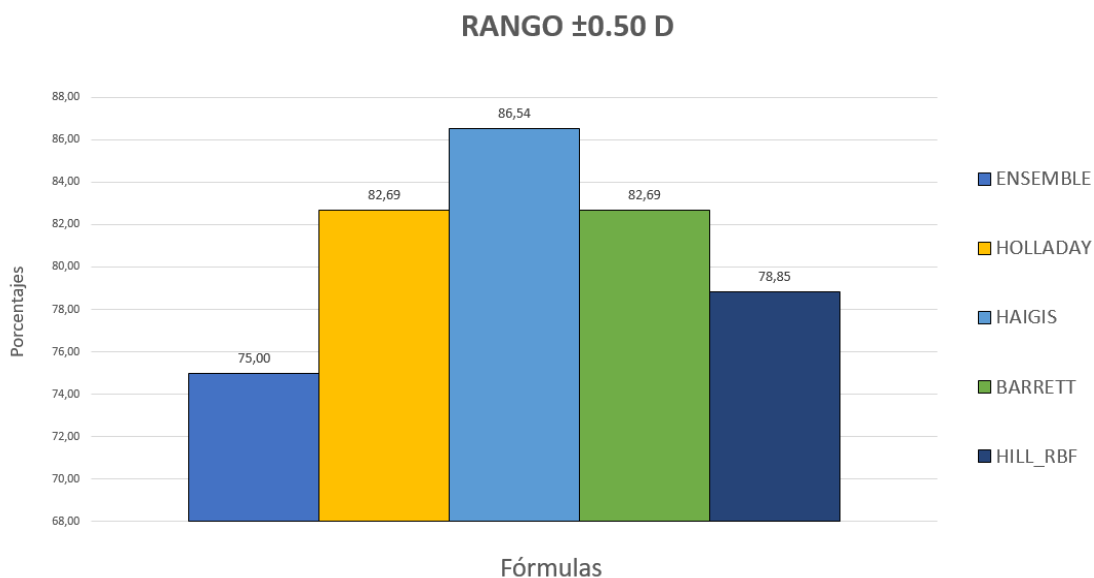


Figura 18. Porcentaje de ojos dentro del rango postoperatorio ± 0.50 D ($n = 52$).

Con el modelo ENSEMBLE se consigue introducir un 75.00% de los ojos dentro del rango ± 0.50 D, no siendo un mal resultado pero muy inferior al 86,54% que alcanza Haigis, o el 82,69% de Holladay II y Barrett Universal II. En la Tabla 7 se puede observar los porcentajes en los rangos ± 0.75 D y ± 1.00 D:

Fórmula	± 0.75 D (%)	± 1.00 D (%)
ENSEMBLE	90.38	96.15
Holladay II	96.15	98.08
Haigis	96.15	98.08
Barrett U.	94.23	98.08
Hill-RBF	90.38	100

Tabla 7. Porcentaje de ojos dentro del rango postoperatorio ± 0.75 D y ± 1.00 D.

Con el modelo combinado se consigue introducir el 90.38% y 96.15% de los resultados refractivos dentro de los intervalos ± 0.75 D y ± 1.00 D respectivamente, valores similares a los conseguidos con Hill-RBF o con el resto de fórmulas.

3.6 Comparaciones de los modelos con otros estudios publicados

En el estudio de Kane et al. [20] para una muestra de 3122 ojos, los valores de EP, EAM y MedEA para fórmulas como Haigis, Holladay II y Hill-RBF son de 0.41 D, superiores a los obtenidos con el modelo ENSEMBLE (Tabla 6). Incluso el porcentaje de ojos incluidos en el rango ± 0.50 D en estas

fórmulas es inferior al 70%, bastante menor al conseguido con el modelo combinado.

En el trabajo de Melles et al. [21] con una muestra de 18501 ojos, la mejor fórmula de las comparadas es Barrett Universal II con un EAM de 0.311 ± 0.40 D, seguida de Haigis 0.338 ± 0.44 D y de Holladay II con 0.350 ± 0.45 D. Hill-RBF no está incluida en el estudio. Por tanto, valores peores en el caso de Haigis y Holladay II respecto a los obtenidos en este trabajo con el modelo ENSEMBLE: 0.34 ± 0.26 D (Tabla 6).

Sin embargo, el modelo ENSEMBLE obtiene un porcentaje de ojos dentro del rango ± 0.50 D igual que Holladay II (75%) pero inferior que Haigis y Barrett Universal II, 77.1 % y 80.8% respectivamente.

Roberts et al. [22] en su artículo constatan la inclusión del 80% de la muestra ($n = 400$) dentro del rango ± 0.50 D en Holladay II, Barrett Universal II y Hill-RBF. El modelo ENSEMBLE incluye 75% en este rango, muy cerca de estas fórmulas. Por otro lado, los valores de EAM con Barrett Universal II son 0.30 ± 0.24 D, con Hill-RBF 0.32 ± 0.24 D y 0.32 ± 0.27 D con Holladay II, no estando incluida Haigis en este estudio. Lo que significa que el modelo propuesto solo difiere 0.04 D en EAM respecto a Barrett Universal II que es la mejor fórmula de este trabajo.

4. CONCLUSIONES

En este trabajo se han implementado 11 modelos de regresión no lineal basados en técnicas de machine learning. De todos ellos, se seleccionaron 2 por ser los que mejor rendimiento presentaban, el primero mediante el algoritmo SVM con kernel gaussiano RBF, y el segundo con *models adaptive regression splines* con un polinomio de segundo orden. Se hiperparametrizaron para optimizar su desempeño, mejorando sus métricas respecto a los modelos en bruto. Por último, se implementó un modelo combinado o ensamblado de los anteriores con técnica *stacking* empleando regresión penalizada LASSO.

El abordaje para el cálculo biométrico desde la perspectiva del aprendizaje automático de máquinas no es inédito, ya que el Dr. Clarke [23] fue pionero en los años 90 con una red neuronal para cálculos biométricos, comparando sus resultados por aquel entonces con la primera versión de Holladay. El Dr. Clarke es el fundador de FullMonteIOL, primer *website* en la que mediante el método de cadenas de Monte Carlo-Markov y redes neuronales, realiza predicciones de potencia de lentes intraoculares. Estas metodologías fueron rechazadas en este trabajo por el escaso rendimiento predictivo que mostraban frente a otros modelos.

El Dr. W. Hill es el precursor de Hill-RBF [24], la otra fórmula de cálculo biométrico basada en técnicas de machine learning presentada en 2016. En su segunda versión, el dataset con el que se obtuvo el modelo no lineal con

gaussiano RBF, fue de 12.419 ojos, muy lejos de la muestra de este estudio ($n = 260$). Esto implica un gran avance en precisión respecto a la primera versión que contaba con 3445 ojos, colocándose actualmente entre las mejores fórmulas de cálculo biométrico.

En términos de EAM, la mejor métrica para comparar rendimiento entre fórmulas biométricas, Hill-RBF obtiene 0.30 ± 0.25 D (0.00/0.97) contra 0.34 ± 0.26 D (0.00/1.07) del modelo ENSEMBLE. Téngase en cuenta la diferencia entre muestras, 12.419 ojos contra 260 ojos de este trabajo.

Es interesante recalcar, que el modelo predictivo de la fórmula del Dr. Hill no tiene en cuenta el diámetro horizontal corneal (WTW). Aunque sea un parámetro solicitado en el formulario de entrada de datos de su web, no tiene peso en el cálculo. En el modelo aquí presentado, el aporte sobre todo de SVM-RBF al modelo ENSEMBLE de esta variable biométrica, es un factor diferenciador a tener en cuenta.

Otros factores diferenciales de este trabajo han sido la inclusión de la cara posterior de la córnea, que ninguna fórmula de cálculo actual valora. Así como, la implementación de varios modelos regresivos no paramétricos con acoplamiento para potenciar la predictibilidad. Por otro lado, el manejo de una nueva variable predictora llamada ASD (Anterior Segment Depth), que sustituye a LT y ACD, adquiriendo más protagonismo en el modelo SVM con kernel radial que la queratometría media anterior.

Respecto a la literatura oftalmológica revisada y comparada, ENSEMBLE obtiene muy buenos resultados, incluso por encima de alguna fórmula de uso habitual en clínica actualmente.

Los resultados del modelo presentado en este trabajo son muy prometedores con la limitación impuesta por una muestra no demasiado grande ($n = 260$), y con prevalencia de ojos con longitudes axiales normales. En futuros estudios se incrementará la muestra para mejorar resultados, intentando homogeneizar la distribución con la inclusión de más ojos cortos por debajo de 22 mm y más largos por encima de 25 mm.

Otro factor que se tendrá en cuenta en trabajos venideros será la variabilidad producida por el cirujano ante ojos con las mismas características biométricas, y la variabilidad debida por medidas realizadas con diferentes tecnologías. Se incrementará la muestra teniendo en cuenta estos hechos.

En conclusión, se ha obtenido una fórmula de cálculo biométrico que tiene cabida en el ámbito de la cirugía de cataratas actual y que alcanzará la precisión predictiva de las mejores cuando amentemos la muestra.

5. GLOSARIO

ELP: del inglés Effective Lens Position. Corresponde a la distancia que queda la lente intraocular implantada de la córnea. Esta variable es la que predicen las fórmulas de cálculo biométrico clásicas.

AXL: Axial Length o longitud axial. Corresponde a la longitud expresada en milímetros determinada desde la córnea hasta el epitelio pigmentario de la retina en caso de medirse con un biómetro óptico.

ACD: Anterior Chamber Depth o profundidad de la cámara anterior. Distancia expresada en milímetros que hay desde el epitelio corneal hasta la cara anterior del cristalino.

LT: Lens Thickness o espesor cristaliniano. Se corresponde con el grosor central del cristalino.

WTW: White to White o blanco-blanco. Es el diámetro horizontal corneal.

ASD: Anterior Segment Depth o profundidad del segmento anterior. Es la distancia en milímetros que hay desde el endotelio corneal a la cristaloides posterior, es decir, la suma de LT y ACD.

IOL: Intra Ocular Lens o lente intraocular. Implante de polímero plástico con un diseño óptico y una potencia refractiva concreta, que va alojado en el interior del ojo en diferentes posiciones anatómicas, dependiendo del tipo de cirugía que se practique.

k-NN: del inglés K- Nearest Neighbour o k-vecinos cercanos. Es un método de clasificación supervisada, que clasifica nuevas instancias como la clase mayoritaria de entre los k vecinos más cercanos entre los datos de entrenamiento.

GLM: del inglés Generalized Linear Model o modelo lineal generalizado. Son una extensión de los modelos lineales que permiten utilizar distribuciones no normales de los errores (binomiales, Poisson, gamma, etc.) y varianzas no constantes.

SVM: del inglés support vector machine o máquinas de vector soporte. Son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik. Se pueden emplear en problemas de clasificación y de regresión.

ANN: del inglés Artificial neural network o redes neuronales. Una red neuronal artificial es un grupo interconectado de nodos similar a la vasta red de neuronas en un cerebro biológico.

RF: del inglés Random Forest, es un algoritmo desarrollado por Leo Breiman y Adele Cutler. Es una combinación de árboles predictores tal que cada árbol

depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos.

LASSO: del inglés Least Absolute Shrinkage and Selection Operator. Es un método de análisis de regresión que realiza selección de variables y regularización para mejorar la exactitud del modelo estadístico.

RMSE: del inglés Root Mean Square Error o también conocido como error cuadrático medio, mide la diferencia entre el estimador y lo que se estima.

6. Bibliografía

[1] Wang Q, Jiang W, Lin T, Zhu Y, Chen C, Lin H, Chen W. Accuracy of intraocular lens power calculation formulas in long eyes: a systematic review and meta-analysis. *Clin Exp Ophthalmol*. 2018 Mar 2.

[2] Faramarzi A, Aghajani A, Ghiasian L. Accuracy of Various Intraocular Lens Power Calculation Formulas in Steep Corneas. *J Ophthalmic Vis Res*. 2017 Oct-Dec;12(4):385-389.

[3] Jeong J, Song H, Lee JK, Chuck RS, Kwon JW. The effect of ocular biometric factors on the accuracy of various IOL power calculation formulas. *BMC Ophthalmol*. 2017 May 2;17(1):62.

[4] Abulafia A, Barrett GD, Koch DD, Wang L, Assia EI. Protocols for Studies of Intraocular Lens Formula Accuracy. *Am J Ophthalmol*. 2016 Apr; 164:149-50.

[5] Cooke DL, Cooke TL. Comparison of 9 intraocular lens power calculation formulas. *J Cataract Refract Surg*. 2016 Aug;42(8):1157-64.

[6] Gökce SE, Zeiter JH, Weikert MP, Koch DD, Hill W, Wang L. Intraocular lens power calculations in short eyes using 7 formulas. *J Cataract Refract Surg*. 2017 Jul;43(7):892-897.

[7] Reitblat O, Levy A, Kleinmann G, Lerman TT, Assia EI. Intraocular lens power calculation for eyes with high and low average keratometry readings: Comparison between various formulas. *J Cataract Refract Surg*. 2017 Sep;43(9):1149-1156.

[8] Clarke GP, Burmeister J. Comparison of intraocular lens computations using a neural network versus the Holladay formula. *J Cataract Refract Surg*. 1997 Dec;23(10):1585-9.

[9] Ho JD, Tsai CY, Tsai RJ, Kuo LL, Tsai IL, Liou SW. Validity of the keratometric index: evaluation by the Pentacam rotating Scheimpflug camera. *J Cataract Refract Surg.* 2008;34:137-145.

[10] <http://topepo.github.io/caret/index.html>

[11] Kotsiantis, D. Kanellopoulos, D. Pintelas, P. Data Preprocessing for Supervised Learning. *International Journal of Computer Science*,1 (2), 111-117. 2006.

[12] D. W. Aha, D. Kibler, and M. K. Albert. Instance-based learning algorithms. *Machine Learning*, (1):37–66, 1991.

[13] I. Rish. An empirical study of the naive Bayes classifier. in *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, (22):41–46, 2001.

[14] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin. *A Practical Guide to Support Vector Classification*. Department of Computer Science National Taiwan University.2016.

[15] Breiman L. Random Forests. *Machine Learning*. 45 (1): 5–32, 2001.

[16] Jerome H. Friedman. Multivariate Adaptive Regression Splines (with discussion). *Annals of Statistics* 19/1, 1991. <https://statistics.stanford.edu/research/multivariate-adaptive-regression-splines>. Cited on pages 4, 5, and 54.

[17] Jerome H. Friedman. Fast MARS. Stanford University Department of Statistics, Technical Report 110, 1993. <https://statistics.stanford.edu/research/fast-mars>. Cited on pages 4, 5, and 11.

[18] Hoffer KJ, Aramberri J, Haigis W, Olsen T, Savini G, Shammas HJ, Bentow S. Protocols for studies of intraocular lens formula accuracy. *Am J Ophthalmol.* 2015 Sep;160(3):403-405.

[19] Koch DD, Hill W, Abulafia A, Wang L. Pursuing perfection in intraocular lens calculations: I. Logical approach for classifying IOL calculation formulas. *J Cataract Refract Surg.* 2017 Jun;43(6):717-718.

[20] Wang L, Koch DD, Hill W, Abulafia A. Pursuing perfection in intraocular lens calculations: III. Criteria for analyzing outcomes. *J Cataract Refract Surg.* 2017 Aug;43(8): 999–1002.

[21] Kane JX, Van Heerden A, Atik A, Petsoglou C. Accuracy of 3 new methods for intraocular lens power selection. *J Cataract Refract Surg.* 2017 Mar;43(3):333-339.

[22] Melles RB, Holladay JT, Chang WJ. Accuracy of Intraocular Lens Calculation Formulas. *Ophthalmology.* 2018 Feb;125(2):169-178.

[23] Roberts TV, Hodge C, Sutton G, Lawless M. Comparison of Hill-radial basis function, Barrett Universal and current third generation formulas for the calculation of intraocular lens power during cataract surgery. Clin Exp Ophthalmol. 2018 Apr;46(3):240-246.

[24] Clarke GP. FullMonte IOL. Disponible en: <https://fullmonteiol.com/>.

[25] Hill W. Hill-RBF Calculator. Disponible en: <https://rbfcalculator.com/>.