

Sistema de apoyo a la decisión para trasplante de páncreas en pacientes con diabetes tipo 1 basado en modelado de datos

Pilar Moreno Alfaro

Tutor: Vicente TRAVER SALCEDO
Cotutor: Antonio MARTINEZ-MILLANA

Trabajo Fin de Máster presentado en la Escuela
Técnica Superior de Ingenieros de Telecomunicación de
la Universitat Politècnica de València, para la obtención
del Título de Máster en Ingeniería de Telecomunicación

Curso 2016–17

Valencia, 11 de septiembre de 2017

Resumen

Este proyecto se centra en el desarrollo de un sistema de ayuda a profesionales sanitarios en la toma de decisión respecto a la realización de trasplantes de páncreas y riñón en pacientes con Diabetes Mellitus tipo 1.

Se ha realizado un análisis exploratorio del conjunto de datos de 82 pacientes sometidos a trasplante de páncreas y riñón entre los años 2002 y 2015 en el departamento de salud La Fe. Este análisis permite determinar la presencia de indicadores del éxito o fracaso de la intervención. Los indicadores más significativos son utilizados para el desarrollo de un modelo de clasificación que permite realizar predicciones en función de los datos de entrada de un nuevo paciente. El usuario accede al modelo a través de un interfaz web y el intercambio de información se realiza mediante un sistema de comunicaciones basado en la orquestación de servicios. Con todo esto, se ha conseguido implementar un sistema distribuido y autónomo que permite predecir antes de realizar la intervención, en base a un número reducido de variables, si un paciente trasplantado es susceptible de perder el páncreas a largo plazo con una precisión del 82,9 %.

Abstract

This master thesis is focused on the development of a clinical decision support system to help Clinicians on the decision about pancreas and kidney transplant in patients with type 1 Diabetes Mellitus.

To do so, an exploratory data analysis has been performed using a group of 82 patients, which have been transplanted in La Fe health department between 2002 and 2015. This analysis allows to describe the sample population and all the variables from the study. Among them, the most significant indicators are used for developing a classification model to predict the transplant success and failure. Afterwards, the model is integrated into a distributed system so the users can interact with the model through a web interface. The information exchange is performed as means as an orchestrated system, implemented as an distributed and autonomous system to predict the pancreas loss with a 82,9 % accuracy with a short group of variables.

Resum

Aquest projecte es centra en el desenvolupament d'un sistema d'ajuda a la decisió per a professionals sanitaris respecte a la realització de trasplantaments de pancrees i ronyó en pacients amb Diabetis Mellitus de tipus 1.

S'ha realitzat un anàlisi exploratori del conjunt de dades de 82 pacients sotmesos a trasplantament de pàncrees i ronyó entre els anys 2002 i 2015 en el departament de salut La Fe. Aquesta anàlisi ha permès determinar la presència d'indicadors de l'èxit o fracàs de la intervenció. Els indicadors més significatius són utilitzats per al desenvolupament d'un model de classificació que permet realitzar prediccions en funció de les dades d'entrada d'un nou pacient. L'usuari accedeix al model a través d'un interfície web i l'intercanvi d'informació es realitza mitjançant un sistema de comunicacions basat en l'orquestració de serveis. Amb tot això, s'ha aconseguit implementar un sistema distribuït i autònom que permet predir si un pacient trasplantat és susceptible de perdre el pàncrees a llarg termini amb una precisió del 82,9% abans de realitzar la intervenció y en base a un nombre reduït de variables.

Índice general

Resumen	II
1. Introducción	1
1.1. Motivación	1
1.2. La diabetes	2
1.2.1. Definición	2
1.2.2. Clasificación y diagnóstico	3
1.2.3. Complicaciones de la Diabetes Mellitus	4
1.2.4. Contexto social	5
1.3. Enfoque terapéutico	5
1.3.1. Nutrición y deporte	6
1.3.2. Control glucémico	6
1.3.3. Insulinoterapia	7
1.3.4. Trasplante de páncreas y riñón	8
1.3.4.1. Historia	8
1.3.4.2. Tipos	9
1.3.4.3. Selección de donante y receptor	9
1.3.4.4. Beneficios del trasplante	10
2. Antecedentes	11
3. Objetivos	15
3.1. Objetivo principal	15
3.1.1. Objetivos secundarios	15
4. Materiales y métodos	17
4.1. Introducción	17
4.2. El modelo de datos relacional	17
4.2.1. Sistema gestor de base de datos	20
4.2.1.1. MySQL	20
4.2.1.2. PostgreSQL	21
4.2.1.3. SQLite	21
4.2.1.4. Oracle	21
4.2.1.5. Microsoft Office Access	22
4.2.1.6. Microsoft SQL Server	22
4.2.1.7. IBM DB2	22
4.3. Datos disponibles para el proyecto	22

4.4.	Análisis exploratorio de datos	23
4.4.1.	Análisis univariante	24
4.4.2.	Análisis bivariante	24
4.5.	Modelos de clasificación	26
4.5.1.	Evaluación de los modelos de clasificación	28
4.6.	Software Matlab	31
4.6.1.	DataBase Toolbox	32
4.6.2.	Statistics and Machine Learning Toolbox	32
4.6.3.	App Classification Learner	33
4.6.4.	Técnicas de reducción de la dimensionalidad	37
4.6.5.	Metodología de validación	38
4.7.	Coreografía de servicios	39
4.8.	Propuesta de sistema de ayuda a la decisión clínica	41
5.	Resultados	43
5.1.	Estructuración de los datos	43
5.2.	Estudio descriptivo de los datos	45
5.3.	Análisis bivariante	55
5.3.1.	Evento final número 1: Supervivencia del paciente	57
5.3.2.	Evento final número 2: Explante de páncreas	59
5.3.3.	Evento final número 3: Pérdida de injerto	60
5.3.4.	Evento final número 4: Necesidad de insulina	63
5.4.	Análisis de los resultados obtenidos	65
5.5.	Preparación de los datos para el modelado	67
5.6.	Evaluación de los modelos de predicción	68
5.6.1.	Primera iteración	69
5.6.2.	Segunda iteración	70
5.6.3.	Tercera iteración	72
5.6.4.	Cuarta iteración	74
5.7.	Implementación del sistema de ayuda a la decisión clínica	78
6.	Discusión	84
7.	Conclusión	87
7.1.	Revisión de los objetivos	87
7.2.	Líneas futuras	88
A.	Análisis exploratorio de datos	89
B.	Variables del estudio	118
	Bibliografía	121

Índice de figuras

1.1. Relación entre niveles de HbA1c y glucemia	7
4.1. Pasos en el proceso de clasificación	26
4.2. Diferentes hiperplanos para un conjunto de muestras	28
4.3. Matriz de confusión	29
4.4. (a) ROC de un modelo de clasificación perfecto ; (b)ROC de un modelo de clasificación aceptable; (c) ROC de un modelo de clasificación que no aporta ninguna información	31
4.5. Database Toolbox (arriba) y código SQL generado automáticamente (abajo)	32
4.6. Descripción del método Cross-Validation	36
4.7. Esquema de un sistema orquestado	40
4.8. Ejemplo mensaje XMSG	41
4.9. Estructura funcional del sistema	41
5.1. Esquema de la base de datos relacional	44
5.2. Frecuencia de pacientes en función de la edad al trasplante	48
5.3. Distribución de la variable <i>HbA1c</i> previa al trasplante	48
5.4. Distribución de la variable <i>GLUCEMIA</i> previa al trasplante	49
5.5. Distribución de la variable <i>SEXO</i>	49
5.6. Número de trasplantes realizados por año	50
5.7. Número de fallecidos por año en que se realizó el trasplante	50
5.8. Evolución de los niveles de creatinina en los pacientes sin necesidad de insulina	52
5.9. Evolución de los niveles de urea en los pacientes sin necesidad de insulina	53
5.10. Evolución de los niveles de filtrado glomerular en los pacientes sin necesidad de insulina	53
5.11. Gráfico de tipo Boxplot para la variable IMC	54
5.12. (a) Variable <i>CREATININA</i> con outlier ; (b)Variable <i>CREATININA</i> sin outlier	55
5.13. (a) Variable <i>DosisCT</i> con outlier ; (b)Variable <i>DosisCT</i> sin outlier	55
5.14. Descripción de las muestras utilizadas para el estudio en función de la supervivencia del paciente	57
5.15. Descripción de las muestras utilizadas para el estudio en función del ex- plante del páncreas	59
5.16. Descripción de las muestras utilizadas para el estudio en función de la pérdida de funcionalidad del páncreas	61
5.17. Descripción de las muestras utilizadas para el estudio en función de la necesidad de insulina	63
5.18. Función PCA del toolbox Classification Learner	70

5.19. Resultados PCA	70
5.20. Pesos de cada variable a la clasificación	71
5.21. Resultados del modelo aplicando NCA	72
5.22. Variables con el influencia en el resultado final excluyendo las variables del donante	73
5.23. Ejemplo de uno de los modelos con una precisión del 80,0%	74
5.24. Resultados del modelo de clasificación para las variables UREA y TG	76
5.25. (a) Matriz de confusión con el número de observaciones ; (b) Matriz de confusión con ratios	76
5.26. Curva ROC y valor del Área Bajo la Curva (AUC)	77
5.27. Esquema del sistema orquestado	78
5.28. Interfaz del servicio Sniffer	79
5.29. Interfaz del servicio TestMatlab1	79
5.30. Interfaz web	80
5.31. (a) Configuración de la dirección del orquestador; (b) Configuración de los servicios	81
5.32. Configuración actual	81
5.33. Resultado positivo del modelo de predicción	81
5.34. Resultado negativo del modelo de predicción	81
5.35. Cabecera del mensaje XMSG	82
5.36. Contenido del mensaje XMSG del remitente	82
5.37. Contenido del mensaje XMSG del destinatario	83

Índice de cuadros

4.1. Comparativa entre los distintos sistemas gestores de bases de datos	20
4.2. Pruebas según la naturaleza de las variables y el tipo de análisis	25
4.3. Descripción de los test de contraste de hipótesis más usados	26
4.4. Velocidad y memoria de cada modelo	35
5.1. Resumen de resultados del análisis descriptivo de las variables cuantitativas	46
5.2. Resumen de resultados del análisis descriptivo de las variables cualitativas	47
5.3. Resumen de los resultados variables cuantitativas para el evento final 1 . .	58
5.4. Resumen de los resultados variables cualitativas para el evento final 1 . .	58
5.5. Resumen de los resultados variables cuantitativas para el evento final 2 . .	59
5.6. Resumen resultados variables cualitativas para el evento final 2	60
5.7. Resumen resultados variables cuantitativas para el evento final 3	61
5.8. Resumen resultados variables cualitativas para el evento final 3	62
5.9. Resumen resultados variables cuantitativas para el evento final 4	64
5.10. Resumen resultados variables cualitativas para el evento final 4	64
5.11. Variables significativas para cada evento final	66
5.12. Conjunto de variables inicial utilizadas para el modelado	69
5.13. Variables con influencia en el resultado final	71
5.14. Variables del donante excluidas del modelado	72
5.15. Variables influyentes con sus pesos	73
5.16. Variables utilizadas como predictoras en el modelo de clasificación	74
5.17. Resultados de precisión para cada modelo entrenado	75
5.18. Parámetros del contenido del mensaje XMSG	82
A.1. Resultados del análisis descriptivo de las variables cuantitativas	90
A.2. Resultados del análisis descriptivo de las variables cualitativas	96
A.3. Resultados variables cuantitativas para el evento final 1: Supervivencia del paciente	97
A.4. Resultados variables cualitativas para el evento final 1: Supervivencia del paciente	102
A.5. Resultados variables cuantitativas para el evento final 2: Explante de páncreas	103
A.6. Resultados variables cualitativas para Explante de páncreas	107
A.7. Resultados variables cuantitativas para el evento final 3: Pérdida de injerto	108
A.8. Resultados variables cualitativas para el evento final 3: Pérdida de injerto	112
A.9. Resultados variables lina	113
A.10. Resultados variables cualitativas para el evento final 4: Necesidad de insulina	117
B.1. Descripción de las variables disponibles en el estudio	120

Abbreviations

DM	Diabetes Mellitus
DMT1	Diabetes Mellitus Tipo 1
SADC	Sistemas de Apoyo a la Decisión Clínica
UI	Unidades Internacionales
IMC	Índice de Masa Corporal
HbA1c	Hemoglobina glicosilada
SPK	Simultaneous Pancreas-Kidney
PAK	Pancreas After Kidney
PTA	Pancreas Transplantation Alone
KNN	k Nearest Neighbors
IAM	Infarto de miocardio
ACV	Ataque cerebrovascular (ictus)
DosisCT	Dosis de corticoides que toma el paciente después del trasplante
NCA	Neighborhood Component Analysis
TG	Triglicéridos
AUC	Área bajo la curva ROC

Capítulo 1

Introducción

1.1. Motivación

Actualmente, la incidencia de la Diabetes Mellitus está aumentando de forma exponencial en todo el mundo. En las últimas décadas, han aumentado de forma alarmante el número de casos y la prevalencia de la enfermedad. Según las estimaciones, 422 millones de adultos en todo el mundo tenían diabetes en 2014, frente a los 108 millones de 1980 [1]. La prevalencia mundial de la diabetes casi se ha duplicado desde ese año, pues ha pasado del 4,7% al 8,5% en la población adulta. Además, se estima que en 2012 la diabetes fue causa directa de 1,5 millones de muertes. A estos datos podría añadirse un porcentaje desconocido de muertes por enfermedad cardiovascular que puede ser directamente atribuible a esta enfermedad [1].

En los últimos años, la medicina ha seguido una tendencia que consiste en detectar las enfermedades de forma precoz y usar el diagnóstico con el fin de seleccionar la terapia óptima en base a observaciones individuales. La posibilidad de almacenar acceder a grandes cantidades de información junto con los continuos avances en el campo de las Tecnologías de la Información y la Comunicación han permitido la aparición de Sistemas de Apoyo a la Decisión Clínica (SADC) para la gestión de pacientes.

Los SADC son sistemas informáticos que proporcionan apoyo para los profesionales clínicos en la toma de decisiones médicas relacionadas con el diagnóstico, pronóstico, tratamiento y gestión de pacientes. Estos sistemas son una valiosa herramienta para los

profesionales sanitarios, usándose siempre como un apoyo complementario al criterio de los expertos y nunca como un sustituto.

Los SADC utilizan modelos matemáticos para obtener la probabilidad de un paciente individual de desarrollar cierta condición clínica. Estos modelos pueden basarse en relaciones de probabilidad entre múltiples variables las cuales han sido recogidas en un contexto de estudio específico.

Este proyecto describe el proceso de diseño, desarrollo y evaluación de un sistema SADC y su adaptación a un entorno donde sea fácilmente accesible para su uso como herramienta de apoyo para profesionales sanitarios.

1.2. La diabetes

1.2.1. Definición

La Diabetes Mellitus (DM) es una enfermedad crónica que aparece cuando el páncreas no produce insulina suficiente o cuando el organismo no utiliza eficazmente la insulina que produce [2].

La insulina es una hormona segregada por el páncreas que regula el nivel de glucosa, una forma de azúcar en sangre. El cuerpo necesita glucosa para tener suficiente energía, abasteciéndose de la misma mediante la ingesta de alimentos. Después de comer, la glucosa se transporta por la sangre hasta los tejidos y la hormona producida por las células- β del páncreas hace posible el traspaso de la glucosa desde el torrente sanguíneo hasta las células de los tejidos.

La principal causa en la aparición de Diabetes Mellitus es la inexistencia de insulina, o bien la imposibilidad de la misma para hacer llegar la glucosa a los tejidos. No obstante, se pueden distinguir varios tipos de DM.

1.2.2. Clasificación y diagnóstico

Según su etiología, la diabetes se puede clasificar en los siguientes grupos [3]:

- Diabetes Mellitus tipo 1 (DMT1): También se conoce como insulino dependiente, juvenil o de inicio en la infancia. Se caracteriza por una producción deficiente de insulina, por lo que requiere una administración diaria de insulina exógena. Suele diagnosticarse durante la niñez y la juventud.
- Diabetes Mellitus tipo 2: También conocida como no insulino dependiente o de inicio en la edad adulta. Se caracteriza por una gestión ineficaz de la insulina. Aparece cuando el páncreas se encuentra sometido a este sobreesfuerzo debido a una mala alimentación, y los tejidos generan una resistencia a la acción de la hormona, anulando a la acción de la misma. Es la variante más extendida, se asocia normalmente al sedentarismo y el sobrepeso.
- Diabetes Mellitus Gestacional: Se caracteriza por hiperglucemia que aparece durante el embarazo y alcanza valores que, pese a ser superiores a los normales, son inferiores a los establecidos para diagnosticar una diabetes [3]. Tanto las mujeres con este tipo de Diabetes como sus hijos corren mayor riesgo de padecer diabetes de tipo 2 en el futuro.
- Otros tipos específicos de diferente etiología:
 - Diabetes Mellitus por defectos genéticos en la célula β : En este grupo se encuentran la Diabetes tipo MODY (Maturity-Onset Diabetes of the Young) que es un tipo poco frecuente asociado a defectos genéticos, y la Diabetes Neonatal que se detecta a las pocas semanas o meses de nacer.
 - Diabetes por defectos genéticos a la acción de la insulina.
 - Diabetes por enfermedades relacionadas con el páncreas. En general, cualquier enfermedad que afecte de forma difusa al páncreas puede provocar diabetes.
 - Endocrinopatías que incrementan el número de hormonas que inhiben la producción de insulina.
 - Diabetes inducida por fármacos que pueden producir hiperglucemia.
 - Diabetes inducida por infecciones.

1.2.3. Complicaciones de la Diabetes Mellitus

Todos los tipos de diabetes pueden provocar complicaciones en el organismo e incrementar el riesgo general de muerte prematura. Las complicaciones debidas a la DMT1 se pueden dividir en dos grupos: las agudas y las crónicas.

- **Complicaciones agudas:** Son las que se producen en un determinado momento en el tiempo.
 - **Hipoglucemia:** El nivel de azúcar en sangre disminuye peligrosamente. Esto puede causar síntomas como pérdida de consciencia e incluso la muerte.
 - **Cetoacidosis:** Es la condición en la que el cuerpo no puede usar la glucosa como fuente de energía debido a que no hay insulina o es insuficiente. En este caso, se usan las grasas como fuente de energía. Cuando las grasas se descomponen para proveer la energía provocan la absorción de unos químicos llamados cetonas que, en niveles altos, son tóxicos. Esto puede causar síntomas como la disminución del estado de conciencia, rigidez muscular y náuseas.
- **Complicaciones crónicas:** Son aquellas que aparecen debido a la persistencia de la enfermedad durante mucho tiempo. Esto se asocia, a largo plazo, a la aparición de complicaciones macro y microvasculares [4] que disminuyen la calidad y esperanza de vida del paciente, siendo la esperanza de vida de un paciente diabético diez años menor que la de un paciente que no sufre esta enfermedad.
 - **Complicaciones microvasculares:** El aumento persistente de azúcar en sangre hace que los vasos sanguíneos se vayan obstruyendo, de manera que cada vez es más difícil que la sangre circule hacia los órganos. Esto puede derivar en ceguera (Retinopatía diabética), insuficiencia renal terminal (Nefropatía diabética) y daños irreversibles en los nervios (Neuropatía diabética).
 - **Complicaciones macrovasculares:** Afectan a los vasos sanguíneos más grandes del cuerpo como la aorta, la carótida o los vasos sanguíneos de las extremidades. Pueden provocar mala circulación de la sangre, infarto agudo de miocardio o infarto cerebral.

1.2.4. Contexto social

En las últimas décadas, el número de caso y prevalencia de la DM ha aumentado alarmantemente. Según se estima, el número de adultos con diabetes en el año 2014 asciende a 422 millones, frente a 108 millones estimados en 1980 [1], representando la DMT1 entre un 5 % y 10 % del total de la población diabética [5].

La Organización Mundial de la Salud estima que la prevalencia mundial de esta enfermedad en adultos ha ascendido del 4,7 % registrado en 1980 al 8,5 % en 2014 [1].

En España se ha registrado una incidencia de DMT1 de 17 casos por cada 100.000 habitantes/año. Mientras que la prevalencia se estima en torno a un 0,2 % para población adulta [6].

La DM fue la causa del 3 % de las muertes en España, de las cuales un 42.8 % fueron hombres y un 57.2 % fueron mujeres [7]. De manera indirecta podría añadirse un porcentaje de muertes por enfermedad cardiovascular que puede ser directamente atribuible a esta enfermedad. La DM se asocia con mayor número de hospitalizaciones, mayor tasa de reingresos y aumento de la duración de la estancia hospitalaria con respecto a la población no diabética. Se estima que los costes directos de la DM en España ascienden a 2,40 - 2,67 billones de euros, lo que se traduce entorno a un 6,3 y 74 % del gasto sanitario total, siendo el coste directo por individuo entre 1.290 y 1.476 euros anuales, muy superior al de las personas sin esta enfermedad, que se estiman unos 865 euros anuales [8].

1.3. Enfoque terapéutico

La DMT1 se caracteriza por una deficiencia en la producción de insulina, por lo que el tratamiento sustitutivo con insulina es necesario para la supervivencia del paciente. Este tratamiento intenta mantener el nivel medio de glucosa en sangre dentro de valores considerados normales [3]. Lo más adecuado es la combinación de la ingesta de hidratos de carbono, el ejercicio y la administración de medicamentos (insulina o antidiabéticos).

Un control efectivo de los niveles de glucosa en sangre minimiza el progreso de la enfermedad, ya que la exposición continua a elevados niveles de glucosa se asocia, a largo

plazo, a la aparición de complicaciones macro y microvasculares que disminuyen la calidad y esperanza de vida del paciente. Tanto es así, que la esperanza de vida de un paciente diabético puede ser de diez años menor que la de un paciente que no sufre esta enfermedad.

1.3.1. Nutrición y deporte

La intervención nutricional es una parte fundamental en la estrategia de tratamiento de los pacientes con DMT1. Sus objetivos son promover y apoyar patrones alimenticios saludables, alcanzar y mantener el peso corporal y alcanzar un control glucémico adecuado para retrasar o prevenir las complicaciones de la diabetes citadas en la sección 1.2.3. En general, los requerimientos nutricionales de los pacientes con DMT1 no difieren de la población general ya que son aquellos necesarios para mantener un peso saludable y un crecimiento adecuado, evitando el sobrepeso. Se considera imprescindible que todas las personas con diabetes deben reciban atención nutricional individualizada, preferentemente proporcionada por una dietista/enfermera capacitada para ello [9].

Se estima que la vida sedentaria y ausencia de ejercicio físico en la vida cotidiana es la causa de entre el 10 % y 16 % de casos de DM. Las personas que están en baja o moderada forma física tienen seis veces más probabilidades de desarrollar esta enfermedad que las que están en buena forma física [10].

1.3.2. Control glucémico

Diversos estudios han puesto de manifiesto la asociación entre el pobre control glucémico y la aparición de complicaciones [11].

El control glucémico más fiable, según la Asociación Americana de Diabetes, consiste en la combinación de dos técnicas que son la monitorización de los niveles de glucemia y el análisis de la hemoglobina glicosilada (HbA1c).

La monitorización de la glucemia en el manejo de la diabetes es una herramienta esencial a la hora de planificar la pauta nutricional y la actividad física, y ajustar la pauta de insulina del paciente.

Por otro lado, la HbA1c es un indicador que refleja los niveles medios de glucosa en sangre durante los 2 o 3 últimos meses. Además, diversos estudios han demostrado que la HbA1c es el mejor predictor de la aparición de complicaciones de la diabetes. Es por ello, que diversas sociedades científicas recomiendan la monitorización del HbA1c de forma paralela al control glucémico.

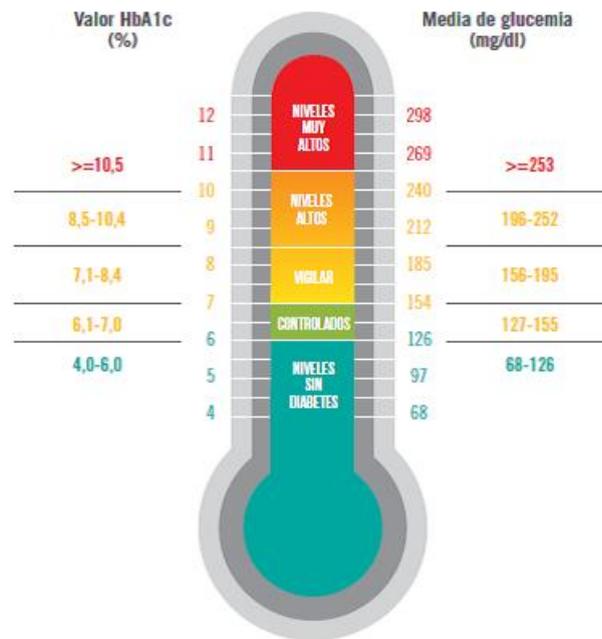


FIGURA 1.1: Relación entre niveles de HbA1c y glucemia

1.3.3. Insulinoterapia

En una persona sin DMT1 la secreción de insulina sigue dos procedimientos: la insulina basal, de forma continua a través de las células β del páncreas; y la insulina prandial, después de cada ingesta de alimentos se refuerza con una producción extra de insulina que ayude a catalizar el exceso de glucosa presente en la sangre.

Sin embargo, una persona que padece DMT1 no tiene la capacidad de producir insulina de forma efectiva, por lo que la terapia de insulina consiste en simular la tarea del páncreas mediante la administración de insulina, bien sea mediante múltiples dosis de insulina o bien en forma de infusión continua subcutánea.

La elección de una forma de administración de insulina u otra debe ser tomada de forma conjunta con el paciente, valorando no sólo riesgo-beneficio, sino también los costes económicos.

Para la administración de insulina mediante múltiples dosis de insulina se puede seguir diferentes pautas pero todas se caracterizan por el uso de un tipo de insulina basal (destinada a simular la liberación continuada de insulina por el páncreas) y otro tipo de insulina prandial (administrada antes de las comidas para controlar los niveles glucémicos tras la ingesta).

Las necesidades diarias de insulina oscilan entre 0,2-0,4 unidades internacionales (UI) de insulina por kg y día. Con ello obtenemos la dosis diaria total de insulina que precisa el paciente y se administra un 40-50 % como insulino terapia basal y el 50-60 % restante como dosis prandiales.

La infusión subcutánea continua de insulina se realiza a través de una bomba de insulina. La bomba de insulina es un dispositivo de pequeño tamaño, que permite la administración de insulina a un ritmo programable. De esta forma permite aportar insulina a lo largo de todo el día a un ritmo continuo en función de los requerimientos del paciente y también permite la administración de cantidades puntuales de insulina extra.

1.3.4. Trasplante de páncreas y riñón

1.3.4.1. Historia

El primer trasplante de páncreas en animales se realizó en 1892 cuando se implantó una porción de páncreas en un perro, previniendo el desarrollo de diabetes. En 1922, con el descubrimiento de la insulina, se frenó la investigación sobre el trasplante de páncreas, ya que el empleo de la insulina permitía aumentar la supervivencia de los pacientes diabéticos [12]. Sin embargo, no tardaron en empezar a aparecer las complicaciones crónicas derivadas de la diabetes. En la década de los 60 los avances en la inmunosupresión permitieron realizar el primer trasplante de páncreas en humanos.

Fue en 1983 cuando se realizó en España el primer trasplante de páncreas combinado y simultáneo con el trasplante renal (SPK), en el Hospital Clínic de Barcelona. Desde entonces y hasta 2015 se han realizado en España un total de 1632 trasplantes de páncreas. En la Comunidad Valenciana se inició el programa de trasplante de páncreas en 2002, siendo el Hospital Universitari i Politècnic La Fe, de Valencia, el único hospital de la Comunidad, donde se realiza.

1.3.4.2. Tipos

El trasplante de páncreas es el tratamiento electivo para pacientes con DM tipo 1, aunque también se ha indicado, pero de forma excepcional, en pacientes con DM tipo 2 o a los pacientes con diabetes tipo MODY y que cumplan unos criterios concretos. Existen tres tipos de trasplantes de páncreas:

1. **Trasplante combinado y simultáneo con trasplante renal** (Simultaneous Pancreas-Kidney, SPK): Este tipo es el más común y frecuente, suponiendo más del 85 % de los trasplantes incluidos en el *Registro Internacional de Trasplante de Páncreas* (IPTR) [13].
2. **Trasplante pancreático en pacientes con trasplante renal previo** (Pancreas After Kidney, PAK): Supone el 10 % de los trasplantes del IPTR.
3. **Trasplante aislado de páncreas** (Pancreas Transplantation Alone, PTA): Es el tipo de trasplante menos indicado, suponiendo el 5 % de los trasplantes en el IPTR.

1.3.4.3. Selección de donante y receptor

Los resultados del trasplante de páncreas, en cualquiera de sus modalidades, han mejorado significativamente en los últimos años. Como consecuencia la lista de espera y el número de trasplantes han aumentado en gran medida.

En nuestro país, el número de páncreas ofertados es mayor al número de demandados, aunque no ocurre lo mismo con el riñón. Por ello, se opta por una extracción multiorgánica en donantes cadáver para permitir la obtención del páncreas y el riñón del mismo donante. Actualmente, no se permite que los pacientes con insuficiencia renal terminal solo reciban un trasplante de riñón, ya que con el trasplante SPK la supervivencia de los pacientes y la calidad de vida es mayor a largo plazo.

En España, en 2015 se generaron 1682 donantes, de los cuales 252 fueron ofertados, se extrajeron 150 y finalmente se implantaron 97 [14].

Los pacientes candidatos de recibir a un trasplante SPK son pacientes diabéticos, generalmente de tipo 1, con insuficiencia renal terminal y que precisan de programas de

diálisis. Los criterios de selección varían ligeramente en función del tipo de trasplante al que se va a someter el paciente.

1.3.4.4. Beneficios del trasplante

Los principales beneficios que aporta el trasplante de páncreas son el descenso de la mortalidad y la mejoría de la calidad de vida del paciente. El descenso de la mortalidad es debido, fundamentalmente al trasplante renal, ya que los pacientes en diálisis presentan una mayor mortalidad frente a los que no se trasplantan [15]. La mejoría en la calidad de vida es derivada, no sólo del abandono de la diálisis, sino también del abandono de la insulina evitándose inyecciones diarias. La mejoría en los resultados del trasplante de páncreas ha ido aumentando de forma progresiva conforme han ido mejorando los criterios de selección de donante y receptor, las técnicas quirúrgicas, los inmunosupresores y la supervisión del injerto, de forma que la supervivencia media del paciente y de los injertos pancreático y renal ha aumentado de forma progresiva.

La supervivencia del paciente a los 5 años del trasplante fue del 89,8 %, siendo similar en los 3 tipos de trasplante de páncreas [16]. Por ello, el trasplante de páncreas en cualquiera de sus modalidades conlleva un aumento notable en la supervivencia del paciente, de alrededor de 10 años, en comparación con aquellos que permanecen en lista de espera.

Este trabajo se centra en el análisis de datos retrospectivos de pacientes que han sido sometidos a SPK con el fin de encontrar indicadores de salud que permitan predecir el éxito/fracaso en la reaparición de la DM tipo 1. En el capítulo 2 se realiza un análisis de los antecedentes en el uso de técnicas de modelado de datos clínicos y se particulariza el análisis para estudios previos relacionados con el trasplante SPK. En el capítulo 3 se definen y enumeran el objetivo principal y objetivos secundarios de este TFM. El capítulo 4 se detallan las técnicas y los procedimientos seguidos para la realización de este proyecto. En el capítulo 5 se exponen detalladamente los resultados obtenidos. En el capítulo 6 se ofrece una interpretación a los resultados que se han obtenido, además de exponer diferentes problemáticas a las que ha hecho frente. Finalmente, en el capítulo 7 se realiza una conclusión global del trabajo y se definen futuras líneas de investigación.

Capítulo 2

Antecedentes

El continuo desarrollo tecnológico en el sistema sanitario Valenciano, junto con el incremento computacional y de capacidad de almacenamiento de los ordenadores actuales propician la implementación de los Sistemas de Ayuda a la Decisión Clínica.

Los SADC están diseñados para colaborar con los médicos en el cuidado de los pacientes de muy diversas formas, ya sea definiendo las terapias oportunas para el paciente, determinando el suministro de medicamentos o estableciendo las probabilidades de éxito de una intervención. Para ello, utilizan modelos matemáticos basados en un amplio abanico de técnicas según la complejidad del problema. No obstante, los más comunes son los sistemas de clasificación y predicción, que determinan la probabilidad de un paciente de desarrollar cierta condición clínica. Estos modelos pueden basarse basan en relaciones de probabilidad entre múltiples variables. El modelo debe garantizar la generalidad del modelo frente a diferentes muestras, pero en la práctica a menudo los modelos pueden no ser igual de efectivos cuando se aplican a otros individuos. Esto puede deberse tanto a deficiencias en el modelado, como a diferencias impredecibles entre las muestras utilizadas para el desarrollo del modelo y las nuevas muestras.

Idealmente, un modelo de predicción debe ser desarrollado utilizando un extenso y definido grupo de individuos con características similares. Los individuos son clasificados en función del padecimiento o no padecimiento de cierto problema o condición clínica, así se comparan los datos de cada paciente.

Dependiendo de la temporalidad del estudio, se puede dividir entre estudios prospectivos y retrospectivos. Los estudios prospectivos son realizados desde el presente siguiendo a

los pacientes en el tiempo. Dentro de esta clasificación, los estudios de cohortes se caracterizan por el hecho de que inicialmente no se conoce el desenlace final de la condición de interés. Durante un periodo de tiempo suficiente, los pacientes deben ser monitorizados siguiendo un protocolo riguroso de actuación para evaluar la evolución de los factores de interés. Finalmente, el conjunto de pacientes se podrá clasificar en dos subconjuntos: uno formado por los pacientes que han desarrollado el problema o condición clínica y otro grupo formado por los pacientes que no lo han padecido. Así, se podrá comparar la incidencia de diferentes eventos entre ambos grupos de pacientes. A pesar de ser estudios muy costosos por requerir gran cantidad de tiempo y recursos económicos, son estudios de gran solidez ya que las observaciones a largo plazo permiten reconocer cualquier variable de confusión potencial.

Los estudios retrospectivos comparan datos de población que tiene el problema o condición de interés con población que no la tiene. Es decir, en contraste con el estudio anterior, en este tipo de estudio inicialmente se elige la muestra de población con el problema o condición médica a evaluar. Este tipo de estudio es más sencillo y rápido que el anterior, aunque es susceptible fallo al enmascarar eventos poco frecuentes o raros.

Los modelos de predicción suelen ser evaluados por su poder estadístico [17], que se cuantifica en base a los siguientes indicadores:

- **Discriminación:** Evalúa la capacidad del modelo para clasificar las muestras según el evento final definido. En este caso, clasificar entre un grupo que desarrolla una cierta condición clínica y otro grupo que no. La medida de discriminación más frecuente es el área definida bajo la curva ROC (AUC).
- **Calibración:** Examinar la correspondencia entre los eventos observados y los eventos predichos por el modelo. Las medidas de calibración más utilizadas son la pendiente de calibración y el test de bondad de ajuste.
- **Generalización:** Capacidad para conseguir buena discriminación y calibración en muestras de población relativamente diferentes a las utilizadas para el desarrollo del modelo.

Estos parámetros se deben utilizar tanto para realizar la validación interna (con los mismos datos con los que se ha realizado el modelo), como para la validación externa

(con nuevos datos no utilizados para la creación del modelo). Desafortunadamente, la validación externa no siempre es posible debido a la falta o escasez de datos.

Hoy en día, los SADC están tomando relevancia y son cada vez más utilizados en el entorno médico. Por ello está en auge el número de investigaciones dedicadas a este ámbito.

Uno de los estudios más conocidos es el Estudio de corazón de Framingham, destinado a identificar los factores y características que se atribuyen a enfermedades cardiovasculares. El estudio comenzó incluyendo a 5.000 pacientes que no habían desarrollado ningún síntoma de enfermedad cardiovascular. La monitorización a lo largo de los años ha permitido la identificación de los principales factores de riesgo en enfermedades cardiovasculares así como los efectos de estos factores [18].

Otro estudio relevante en la comunidad médica es el Estudio prospectivo de la diabetes en el Reino Unido. Este estudio monitorizó durante 10 años a 5.000 pacientes del Reino Unido con diabetes tipo II y ha aportado gran cantidad de información sobre diferentes tratamientos y factores de riesgo de complicaciones microvasculares [19].

En concreto, el trasplante de páncreas ha sido objeto de estudio por numerosos autores ya que se asocia a un alto porcentaje de complicaciones quirúrgica que finalmente derivan en la pérdida funcionalidad del injerto, especialmente si se compara con trasplantes de otros de órganos [20]. A pesar de que, con el paso de los años, las técnicas quirúrgicas y los fármacos para este tipo de intervención han ido mejorando el ratio de pérdidas del páncreas trasplantado sigue sin descender. El estudio realizado por E.B.Finger [21] se centra en el análisis de factores de riesgo que pueden estar directamente relacionados con la pérdida del páncreas a corto plazo (periodo menor de 90 días desde el trasplante). Encuentran diferentes factores de riesgo significativamente estadísticos asociados al donante del injerto, entre los que se encuentran la edad del donante, los niveles de Creatinina, el IMC, la aparición de pancreatitis y el tiempo de preservación del órgano hasta el momento del trasplante. Con estos factores, han desarrollado un modelo de evaluación para posibles donantes de órganos que ha permitido determinar el impacto de la presencia de alguno de estos factores de riesgo presentes en el donante sobre la futura pérdida del injerto en el receptor. Con esto, han concluido que donantes que presentan solo uno

de estos factores de riesgo no supone una influencia relevante en el resultado final, mientras que la presencia de al menos 3 factores de riesgo puede disminuir la supervivencia del injerto hasta un 50 %.

En este proyecto, el objetivo es analizar variables retrospectivas para elaborar un modelo que permita identificar de manera predictiva los paciente que van a aceptar y rechazar el injerto pancreático.

Capítulo 3

Objetivos

3.1. Objetivo principal

Desarrollar e implementar de un sistema que Apoyo a la Decisión Clínica que permita predecir de manera fiable el éxito de una intervención quirúrgica de trasplante de páncreas y riñón en pacientes con Diabetes Mellitus tipo 1, entendiendo como éxito el mantenimiento íntegro de la funcionalidad del injerto pancreático.

3.1.1. Objetivos secundarios

Para la consecución del objetivo principal se han fijado los objetivos secundarios descritos a continuación.

- **Crear un modelo estructurado de datos:** Diseño e implementación de una base de datos relacional adecuada para albergar el amplio conjunto de datos necesarios para el desarrollo del modelo de apoyo a la decisión.
- **Realizar un análisis de las variables de estudio:** Estudio estadístico descriptivo de cada variable clínica para determinar las características de las mismas y conocer a fondo el conjunto de datos del que se dispone.
- **Realizar un estudio bivalente de las variables de estudio:** Elaboración de un análisis estadístico del conjunto de variables para determinar posibles relaciones

intrínsecas entre las muestras y la influencia de cada variable en el evento final de estudio.

- **Evaluar técnicas de modelado:** Evaluación de diferentes modelos de predicción que sean capaces de predecir, en base a unas variables de entrada determinadas, la pérdida o no pérdida de funcionalidad del injerto pancreático en pacientes con DMT1 a los que se les ha realizado un trasplante simultáneo de páncreas y riñón.
- **Implementar un sistema distribuido de comunicaciones:** Diseño e implementación de un sistema que permita a los profesionales médicos acceder al modelo de predicción en cualquier momento y desde cualquier lugar, sirviendo como apoyo en la toma de decisiones.

Capítulo 4

Materiales y métodos

4.1. Introducción

En este capítulo se detallan los materiales y métodos que han sido necesarios para la elaboración de este proyecto. En primer lugar, en la sección 4.2 se describen los sistemas gestores de bases de datos disponibles para la adecuada estructuración y almacenamiento de los datos. En la sección 4.3 se ofrece una descripción de los datos disponibles para el estudio. En la sección 4.4 se presentan las técnicas disponibles para el análisis estadístico de los datos disponibles. En la sección 4.5 se analizan diferentes métodos utilizados para el desarrollo de modelos de clasificación y predicción. Las herramientas disponibles para la aplicación y práctica de las técnicas y modelos de estas dos secciones anteriores se expone en la sección 4.6. Por último, en las secciones 4.7 y 4.8 se describe el sistema utilizado para la comunicación e intercambio de información entre el usuario final y el modelo de predicción.

4.2. El modelo de datos relacional

El correcto almacenamiento y estructuración de los datos se considera imprescindible no solo para ofrecer más comodidad al acceso a la información, si no que también aporta un valor añadido al conjunto de datos.

Un sistema gestor de bases de datos (SGBD) consiste en una colección de datos relacionados en un contexto y un conjunto de programas para acceder a dichos datos. La

colección de datos, normalmente denominada base de datos, contiene información relevante para una entidad. El objetivo principal de un SGBD es proporcionar una forma de almacenar y recuperar la información de una base de datos de manera que sea tanto práctica como eficiente.

La base de datos se construye respetando un modelo de datos. El modelo aporta una colección de herramientas conceptuales para describir los datos, las relaciones y la semántica de la base de datos.

En 1970 Edgar Frank Codd, de los laboratorios IBM, postuló las bases para el diseño e implementación del modelo relacional y la forma de acceder a los datos de las bases de datos. Su fuerte base matemática hace que el modelo sea predecible, fiable y seguro, lo que ha propiciado su consolidación en la actualidad como el modelo más utilizado. Todo modelo de datos debe tener en consideración tres aspectos:

- Estructura de los datos
- Integridad de los datos
- Manejo de los datos

En cuanto a la **estructura de los datos**, la idea fundamental es el uso de relaciones para el almacenamiento de la información. De manera simplificada, la fila de una tabla, denominada tupla, representa una relación entre un conjunto de valores, siendo una tabla el conjunto de dichas relaciones. A su vez, una tabla tiene un conjunto de columnas, denominadas atributos, cuyo nombre debe ser único. Entre las propiedades más importantes que deben respetar dichas relaciones se destacan: *i*) el conjunto de valores de los atributos para una relación deben considerarse unidades indivisibles; *ii*) En una tabla no puede haber ninguna tupla repetida.

Como consecuencia de estas propiedades, surgen las claves. Las están formadas por uno o más atributos cuyos valores identifican de forma única dos tuplas de una relación. El tipo de clave más importante es la clave primaria, es una única clave elegida por el diseñador de la base de datos como elemento principal para identificar de forma única cada tupla (debe ser elegida por el diseñador porque en una base de datos suele haber varias claves que podrían considerarse primarias), de manera que no pueden existir dos o más registros con el mismo valor del campo o campos clave. Este campo o conjunto

de campos se conoce como clave única y pueden existir varias claves únicas en una determinada tabla.

Por otro lado, la **integridad de los datos** de un modelo relacional se define mediante una serie de restricciones. La regla de integridad de entidades especifica que ninguna tupla de una relación puede tomar valores nulos en los atributos que forman parte de su clave primaria, aunque sí pueden ser nulos valores de otras claves. La regla de integridad referencial dictamina que si una una tabla incluye atributos de la clave primaria de otra tabla (se denomina clave externa), sus valores deben coincidir con valores de la clave primaria a la que hace referencia, o bien, deben ser todos nulos.

Por tanto, una base de datos, desde el punto de vista relacional, está formada por un conjunto de relaciones. El esquema lógico de una base de datos consistirá en la unión de los esquemas de todas las relaciones que componen la base de datos, conjuntamente con todas las restricciones de integridad que afectan a esas relaciones.

Por último, para el **manejo de datos** se utiliza el lenguaje relacional. Este lenguaje, a su vez, se divide en dos: el álgebra relacional y el cálculo relacional. El álgebra relacional consta de un conjunto de operaciones que toman como entrada una o dos relaciones y producen como resultado una nueva relación. Las operaciones fundamentales del álgebra relacional son selección, proyección, unión, diferencia de conjuntos, producto cartesiano y renombramiento. Además de las operaciones fundamentales hay otras operaciones, por ejemplo, intersección de conjuntos, reunión natural, división y asignación que se definirán en términos de las operaciones fundamentales. Por otro lado, el cálculo relacional es un lenguaje de consultas que describe la información deseada sin dar un procedimiento específico para obtenerla. Éstos son lenguajes formales para la creación de consultas. Sin embargo, los sistemas de bases de datos comerciales precisan de un lenguaje de consultas cómodo para el usuario. Uno de los lenguajes más utilizados es el Structured Query Language (SQL). SQL usa una combinación de álgebra relacional y construcciones del cálculo relacional.

4.2.1. Sistema gestor de base de datos

Dado que se va a trabajar con el software Matlab, se han analizado las opciones disponibles que son compatibles con el toolbox que ofrece para trabajar con bases de datos relacionales, llamado Database Explorer. Este Toolbox permite conexiones JDBC y ODBC con rápidas lecturas y escrituras mediante un interfaz ODBC. También ofrece funciones para ejecutar consultas usando archivos y declaraciones SQL, e importación directa de datos en una estructura de tablas para formar un conjuntos de datos. El toolbox soporta conexiones con los sistemas gestores de bases de datos detallados en la tabla 4.1, ofreciendo a continuación una visión detallada de cada una de ellas.

	MySQL	PostgreSQL	SQLite	Oracle	Microsoft Access	Microsoft SQL Server	IBM DB2
Software libre	✓	✓	✓	✗	✗	✗	✗
Fácil instalación	✓	✓	✓	✗	✓	✗	✗
Fácil uso	✓	✗	✓	✗	✓	✓	✓
Consume pocos recursos	✓	✗	✓	✗	✓	✗	✗
Mecanismos para garantizar la seguridad	✓	✓	✓	✓	✓	✓	✓
Manejo de grandes volúmenes de datos	✓	✓	✗	✓	✗	✓	✓

CUADRO 4.1: Comparativa entre los distintos sistemas gestores de bases de datos

4.2.1.1. MySQL

Es el DBMS open source más popular a nivel mundial. Es conocido por su escalabilidad, flexibilidad y fácil uso. Además, ofrece excelentes medidas de seguridad que garantiza la absoluta protección de los datos (como pueden ser las técnicas que permiten el acceso a los datos solo para usuarios autorizados). Otra de las razones de su popularidad es que proporciona un soporte integral dependiendo de las necesidades de cada usuario, como pueden ser conectores, drivers (ODBC, JDBC, etc.) y soporte para diferentes lenguajes de programación como pueden ser C, C++, Java, PHP, Perl, Python, Ruby, etc. Por otra parte, la comunidad que usa MySQL es muy activa, por lo se cuenta con gran bibliografía e información sobre el mismo.

4.2.1.2. PostgreSQL

Sistema de gestión de base de datos objeto-relacional cuyo código fuente disponible libremente. Utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema, es decir, un fallo en uno de los procesos no afectará al resto y el sistema continuará funcionando. El nombre hace referencia a la evolución del proyecto original (Post Ingres) desarrollado por la Universidad de California, Berkeley.

4.2.1.3. SQLite

Es un sistema de gestión de base de datos relacional de dominio público cuya principal característica es que permite almacenar información en dispositivos embebidos de una manera sencilla, potente y rápida. Debido a su pequeño tamaño, SQLite se ha utilizado en sistemas como Android o BlackBerry.

Además, gracias a que es libre es posible encontrar una gran cantidad de componentes, librerías y drivers para interactuar con SQLite desde una gran diversidad de lenguajes y plataformas de programación como Java, Perl, Python, PHP, Ruby, etc.

Matlab tiene una aplicación específica que te permite la conexión de manera muy sencilla con una base de datos SQLite.

Sin embargo, el interfaz de Matlab para SQLite tiene muchas limitaciones ya que solo admite los datos DOUBLE, INT64 y CHAR, no soporta valores NULL, y no permite el uso de la aplicación Database Explorer para facilitar la comunicación con la base de datos, para eso hay que utilizar la línea de comandos.

4.2.1.4. Oracle

Oracle database es un sistema de gestión de base de datos de tipo objeto-relacional desarrollado por Oracle Corporation. Es considerado uno de los sistemas de base de datos más completos debido a estabilidad, escalabilidad y soporte multiplataforma. Por ello es uno de los más usados mundialmente.

4.2.1.5. Microsoft Office Access

Es un sistema de gestión de bases de datos incluido en el paquete Office. Su uso es muy sencillo e intuitivo debido a su interfaz gráfica (similar a todos los programas de Office). Sin embargo, no está recomendada para el uso de grandes volúmenes de datos debido a su escasa robustez y sus limitaciones en cuanto a búsquedas.

4.2.1.6. Microsoft SQL Server

Sistema de gestión de base de datos relacionales desarrollado por Microsoft. Está basado en el lenguaje Transact-SQL (TSQL). Transact-SQL es fundamental para trabajar con SQL Server. Todas las aplicaciones que se comunican con SQL Server lo hacen enviando instrucciones Transact-SQL al servidor, independientemente de la interfaz de usuario de la aplicación.

Puede ser configurado para utilizar varias instancias en el mismo servidor físico. También aporta soporte para transacciones y procedimientos almacenados. Además incluye un entorno gráfico de administración, que permite el uso de comandos DDL y DML gráficamente.

4.2.1.7. IBM DB2

Es un sistema de gestión de bases de datos relacionales desarrollado por la empresa IBM. Entre sus características principales se pueden destacar su integridad ya que asegura la protección de datos aun en caso de colpaso, escalabilidad, facilidad de uso e instalación. DB2 tiene APIs para REXX, C++, C, Delphi, .NET CLI, Java, Python, Perl, PHP, Ruby y otros lenguaje de programación. También soporta la integración en entornos como Eclipse y Visual Studio.

4.3. Datos disponibles para el proyecto

Los datos del estudio fueron tomados en el Hospital La Fe (Valencia, España) a lo largo de 13 años, comenzando en septiembre del año 2002 y finalizando en agosto de 2015, a 82 pacientes a los que se les ha practicado un trasplante de páncreas y riñón, ya sea de

manera simultánea (SPK) o solo el páncreas después de un trasplante de riñón anterior (PAK). Es necesario señalar que la labor de selección y recopilación de dichas variables queda fuera del alcance de este trabajo.

Los datos de los pacientes se encuentran almacenados en una base de datos confeccionada con el programa SPSS, y cuenta con un total de 82 filas y 795 columnas. Las filas se corresponden con el total de pacientes que han participado en el estudio, mientras que en las columnas se almacenan los valores de las variables observadas en distintos periodos, teniendo la misma variable desdoblada en varias columnas según el periodo del estudio en que fue tomada.

4.4. Análisis exploratorio de datos

El objetivo del Análisis Exploratorio de Datos (EDA) o Estadística Descriptiva consiste en ordenar, describir y sintetizar la información disponible. La realización de gráficas y diagramas forma parte fundamental en este análisis dado que proporciona una manera visual directa de organizar la información, permitiendo además la fácil identificación de datos atípicos y outliers. De manera que la finalidad del análisis exploratorio de datos no es consiste en extraer conclusiones generales sobre el fenómeno de estudio, sino solamente su descripción.

Para describir un conjunto de datos se comienza con un análisis individual de cada variable, análisis estadístico univariante, y posteriormente se estudian las relaciones entre variables, análisis estadístico multivariante.

El primer paso para el desarrollo del análisis de datos consiste en determinar la naturaleza de las variables con las que se trabaja ya que determina el tipo de análisis a realizar.

- Variables cualitativas: son aquellas que expresan cualidades o atributos. Aparecen agrupadas en categorías pudiendo presentar un orden subyacente (ordinales) o no (nominales).
- Variables cuantitativas: son aquellas que pueden expresarse numéricamente. A su vez, se puede distinguir entre variables cuantitativas discretas cuando solo pueden tomar valores enteros, y continuas cuando contienen cifras decimales.

4.4.1. Análisis univariante

El análisis univariante estudia la distribución individual de cada variable. Este análisis se centra en dos aspectos: obtener la tendencia central de la distribución y su dispersión. Atendiendo al tipo de variable se realizan los siguientes análisis:

- Variables cualitativas: para este tipo de variables el análisis consiste en expresar, normalmente en forma de porcentaje, la frecuencia de cada categoría. Se suelen representar mediante diagramas de barras o sectores.
- Variables cuantitativas: para el análisis de la tendencia central se utiliza la media aritmética, mientras que para estudiar el grado de dispersión se utiliza la desviación típica o varianza. Una variable cuantitativa discretas con un número pequeño de valores puede representarse gráficamente de manera similar a una variable cualitativa (mediante diagrama de barras o sectores). Si la variable es continua se representa gráficamente mediante histogramas o diagramas de cajas.

4.4.2. Análisis bivariante

El análisis bivariante estudia el efecto de una variable sobre otra. Para ello utiliza dos tipos de técnicas estadísticas: las pruebas paramétricas y las no paramétricas. Las pruebas más comunes para determinar la relación entre dos variables según el tipo de técnica son las siguientes [22]:

- Pruebas paramétricas: prueba t, análisis de correlación de Pearson y prueba ANOVA.
- Pruebas no paramétricas: prueba Mann-Whitney-Wilcoxon, análisis de correlación de Spearman y de Kendall tau, prueba de Kruskal-Wallis, prueba Chi-cuadrado (χ^2) y prueba Fisher.

El uso de las pruebas paramétricas está restringido a las muestras que cumplen los siguientes requisitos:

- Distribución normal de las variables: la población estudiada debe seguir una distribución Normal.

- Uniformidad de la varianza: se requiere que la varianza de una variable no dependa del nivel de otra variable.
- Escala de medida: las variables deben poder medirse numéricamente.
- Independencia: las muestras de un sujeto no dependen de las de otro.

Las dos últimas restricciones se contrastan siguiendo el sentido común del investigador. En caso de no cumplir alguna restricción, se hará uso de las pruebas no paramétricas.

La siguiente tabla muestra un resumen de las pruebas que se pueden realizar según a naturaleza de la variable y el tipo de análisis.

TIPO VARIABLES	TIPO PRUEBAS		PRUEBA
Cualitativa-Cualitativa	Pruebas no paramétricas		Prueba Fisher (tablas 2x2) Prueba χ^2
Cuantitativa-Cualitativa	Distribución normal	Paramétricas	Prueba t
	Distribución no normal	No paramétricas	Análisis media (Mann-Whitney-Wilcoxon) Análisis correlación (Spearman, Kendall tau) Análisis varianza (Kruskal-Wallis)

CUADRO 4.2: Pruebas según la naturaleza de las variables y el tipo de análisis

Como se ha señalado anteriormente, las análisis bivariate tienen como objetivo determinar la relación de una variable con otra haciendo uso de un conjunto de pruebas atendiendo a la naturaleza de las variables observadas.

Estas pruebas se basa en el contraste de hipótesis. Esto es un procedimiento para juzgar si una propiedad que se supone en una población estadística es compatible con lo observado en una muestra de dicha población. En otras palabras, los test de contraste de hipótesis analizan si las diferencias entre muestras de una distribución son debidas a la aleatoriedad o, por el contrario, indican que pertenecen a dos distribuciones diferentes. Los conceptos asociados con una prueba de hipótesis son:

- Valor del estadístico: es el valor que se obtiene a partir de los datos de la muestra aplicando la fórmula correspondiente a la prueba elegida.
- Nivel de significación: nivel a partir del cual se rechaza la hipótesis nula.
- Hipótesis nula (H_0): hipótesis que se desea contrastar.
- Hipótesis alternativa (H_1): hipótesis que se acepta en el caso de rechazar la hipótesis nula.

Prueba	Estadísticos	Hipótesis Nula (H0)
Comparación de medias	t, Mann-Whitney, Wilcoxon	Las medias son iguales, por tanto no hay diferencias entre los dos grupos
Independencia entre dos variables nominales	χ^2 , Fisher	Las dos variables son independientes
ANOVA	F, Kruskal-Wallis	El factor no tiene efecto sobre la variable dependiente
Homogeneidad de la varian za	Barlett, Cochran, Hartley	Homogeneidad de la varian za entre grupos

CUADRO 4.3: Descripción de los test de contraste de hipótesis más usados

En la tabla 4.3 se resumen algunos de los contrastes de hipótesis más usuales en el análisis estadístico [22], incluyendo la formulación de la hipótesis nula y su significado.

4.5. Modelos de clasificación

Los modelos de clasificación tratan de analizar el efecto simultáneo de las relaciones de un conjunto de más de dos variables con el fin de extraer una conclusión. Dado un conjunto de datos que se pueden dividir en dos o más subconjuntos, tratan de clasificar los datos en el subconjunto que les corresponde. Para el desarrollo de los modelos es necesario disponer de datos conocidos, es decir, que incluyan tanto la muestra como su clasificación. Una vez se dispone del modelo, se puede utilizar para predecir resultados en función de nuevos datos de entrada.



FIGURA 4.1: Pasos en el proceso de clasificación

Este tipo de modelos se conocen como modelos de Aprendizaje Supervisado. Por el contrario, los modelos que se desarrollan a partir de datos de los que no se conocen su clasificación se denominan modelos de Aprendizaje No Supervisado.

Dentro de estas categorías existe una amplia variedad de modelos y técnicas, sin embargo, en este proyecto solo se van a describir los modelos de Aprendizaje Supervisado más interesantes para el presente estudio.

- **Regresión Logística:** Una regresión es un modelo que intenta aproximar matemáticamente las relación entre una variable dependientes y otras independientes. Cuando la variable dependiente se trata de una variable categórica (Si/No, Muere/Vive, etc), se conoce como Regresión Logística.
- **Árboles de decisión:** Los Árboles de Decisión tratan de abordar el problema de clasificación dividiéndolo en diferentes problemas de clasificación más pequeños. De este modo, se crean nodos consecutivos en los cuales los datos se van clasificando desde el conjunto de datos inicial (raíz) hasta el resultado (hojas).
- **k-Vecinos más cercanos (*k-nearest neighbors*, *KNN*):** Este algoritmo de clasificación se basa en que un dato se clasifica en función de la clase más frecuente de sus k datos más cercanos (con menor distancia). Una cuestión importante es la selección del valor k . Está demostrado que el porcentaje de aciertos en la clasificación depende de k , siendo valores recomendados entre 5 y 7 [23]. También se pueden utilizar diferentes técnicas para el cálculo de la distancia mínima entre los datos: distancias euclídeas, distancia Chebychev, distancia de Hamming, distancia del coseno, etc.
- **Support Vector Machines (SVM):** Las SVM son capaces de clasificar datos mediante la construcción de un hiperplano¹ que separa las muestras en función de la clase a la que pertenecen. Este hiperplano se construye observando las muestras y determinando el vector que separa las clases de la forma más eficiente. Para ello se busca la distancia máxima entre el hiperplano y las muestras de diferente clase más cercanas al hiperplano. En la figura 4.2 se aprecia que los hiperplanos H_2 y H_3 dividen las muestras según su clase, pero únicamente el hiperplano H_3 mantiene la distancia máxima entre las muestras de cada clase más cercanas. En el caso más sencillo de SVM, la separación entre clase se realiza mediante una linea recta o plano recta pero, a medida que aumenta la complejidad de la clasificación, es necesario utilizar otro tipo de funciones para la construcción del hiperplano. Las funciones de Kernel se utilizan en problemas mucha complejidad computacional. Son funciones que mapean las muestras a un nuevo espacio de dimensión superior en el que es más sencilla la clasificación.

¹En un espacio unidimensional (como una recta), un hiperplano es un punto: divide una línea en dos líneas. En un espacio bidimensional (como el plano xy), un hiperplano es una recta: divide el plano en dos mitades. En un espacio tridimensional, un hiperplano es un plano corriente: divide el espacio en dos mitades.

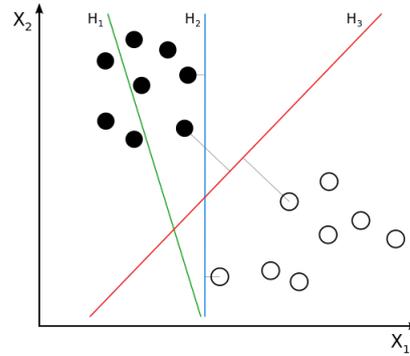


FIGURA 4.2: Diferentes hiperplanos para un conjunto de muestras

- Análisis Discriminante:** Trata de encontrar las relaciones entre las variables de entrada para obtener un modelo matemático que permita clasificar las muestras. Se puede considerar un análisis de regresión en el que la variable dependiente se divide en varias categorías y, además, las variables independientes son continuas. Hay diferentes tipos de Análisis Discriminante: el Análisis Discriminante Paramétrico (Lineal y Cuadrático) y el No Paramétrico. El primer tipo está sujeto a más restricciones que el segundo. Este análisis asume que se puede obtener una función de distribución Normal con las muestras de cada una de las clases. De modo que trata de buscar un método para establecer una frontera entre ambas distribuciones que permita clasificar adecuadamente las muestras.
- Conjunto de clasificadores (*Ensemble Classifiers*):** Es un método muy útil cuando la dificultad del problema de clasificación es muy elevada. Este método divide el dataset completo en diferentes dataset y cada uno de ellos se utiliza para el entrenamiento y prueba de un modelo de clasificación, de manera que se obtienen un conjunto de tantos modelos como subconjuntos se han realizado del dataset original (clasificadores). Finalmente, la clasificación de datos se hace teniendo en cuenta el resultado obtenido para cada uno de los modelos del conjunto.

Se pueden encontrar una amplia variedad de tipos de este método en función del modelo de clasificación a utilizar y el método de división del dataset original en los nuevos datasets.

4.5.1. Evaluación de los modelos de clasificación

Para medir cuantitativamente la capacidad de clasificación de un modelo se utiliza la Matriz de Confusión. Es una matriz en la que se visualizan el número de muestras que se

han clasificado correctamente respecto a la que se han sido clasificadas incorrectamente. En concreto, se representan en las filas las categorías reales en las que se dividen los datos, mientras que en las columnas se representan los resultados obtenidos de la predicción. De este modo se pueden comparar los resultados reales frente a los resultados de la predicción. En la imagen 4.3 se puede observar una matriz de confusión para un caso de clasificación binario con las clases Positivo y Negativo.

Resultado de la clasificación			
Resultado real	POSITIVO	Verdadero Positivo (VP)	Falso Negativo (FN)
	NEGATIVO	Falso Positivo (FP)	Verdadero Negativo (VN)

FIGURA 4.3: Matriz de confusión

En las celdas de la primera diagonal se sitúan los datos correctamente clasificados en sus correspondientes clases, mientras que en otra diagonal se encuentran los erróneos. La celda 12 representa el número de datos que se han clasificado como Negativo siendo realmente de la clase Positivo (falso Negativo), mientras que la celda 21 muestra el número de datos que se han clasificado de la clase Positivo perteneciendo realmente a la clase Negativo (falso Positivo).

A partir de esta matriz, se obtienen varias métricas que reflejan numéricamente el nivel de exactitud del modelo.

$$Exactitud = \frac{NumPrediccionesCorrectas}{NumPrediccionesTotales} = \frac{VP + VN}{VP + FP + VN + FN} \tag{4.1}$$

$$RatioError = \frac{NumPrediccionesIncorrectas}{NumPrediccionesTotales} = \frac{FP + FN}{VP + FP + VN + FN} \tag{4.2}$$

En el ámbito clínico, el grado de eficacia de una prueba diagnóstica se evalúa mediante la sensibilidad y la especificidad. Si suponemos que las clases anteriores, Positivo/Negativo,

se refieren a pacientes que tienen una determinada enfermedad o no la tienen, respectivamente, la **sensibilidad** es un parámetro que mide la probabilidad de detectar un paciente enfermo dentro del grupo de pacientes que verdaderamente están enfermos. La sensibilidad es un parámetro especialmente importante en los casos en que una enfermedad no debe pasar desapercibida o cuando la vida del paciente depende de la detección precoz de la enfermedad. Por el contrario, la **especificidad** mide la probabilidad de que se clasifiquen como negativo a pacientes que realmente no tienen la enfermedad. La especificidad es realmente importante cuando cuando la enfermedad tiene cierto estigma o el tratamiento supone un riesgo innecesario para la vida del paciente.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (4.3)$$

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (4.4)$$

A partir de la matriz de confusión, también se puede obtener otra medida de evaluación del modelo de clasificación llamada Área bajo la curva ROC. La curva ROC muestra la relación entre falsos positivos (1 - especificidad) y verdaderos positivos (sensibilidad). Se representa gráficamente con el valor de sensibilidad en el eje y , y el valor de (1 - especificidad) en el eje x , de manera que se obtiene un punto de la recta cada vez que se evalúa el modelo de clasificación según se va variando el umbral de discriminación del modelo. Para los modelos que proporciona como resultados valores continuos de probabilidad de pertenecer a una clase u otra, la curva se dibuja median la unión de cada uno de los puntos que se obtienen como resultado de cada modificación del umbral entre clases.

A partir de la curva ROC, se pueden obtener diferentes estadísticos para determinar el rendimiento del modelo. El más utilizado es el Área Bajo la Curva (AUC). El AUC determina, para estos ejemplos, la capacidad del modelo de distinguir entre pacientes enfermos y no enfermos. Si el AUC toma el valor 1, el modelo estaría clasificando el 100 % de los pacientes enfermos como enfermos y el 100 % de los pacientes no enfermos como sanos. En cambio, si toma el valor 0.5, el modelo clasifica con enfermo con la misma probabilidad a un paciente que tiene la enfermedad como a un paciente que no la

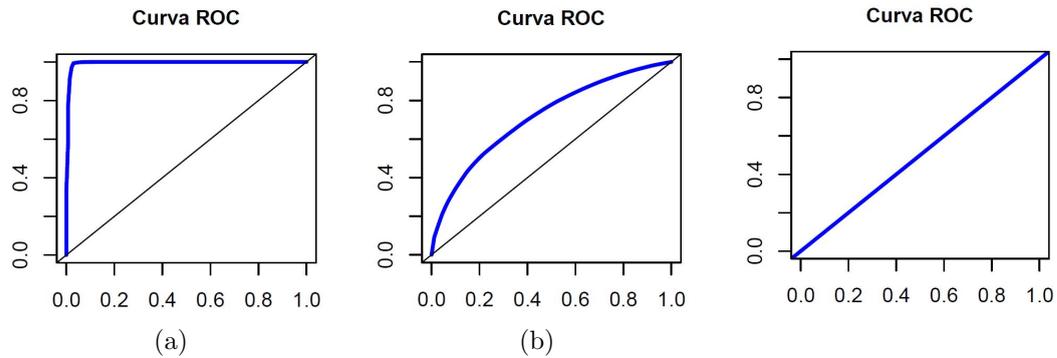


FIGURA 4.4: (a) ROC de un modelo de clasificación perfecto ; (b) ROC de un modelo de clasificación aceptable; (c) ROC de un modelo de clasificación que no aporta ninguna información

tiene, de manera que el modelo no aportaría ninguna información clasificatoria. Cuanto mayor sea el valor del AUC mayor capacidad de discriminación tendrá el modelo.

4.6. Software Matlab

Para el desarrollo del proyecto se ha trabajado con un gran conjunto de variables que serán analizadas mediante métodos matemáticos y probabilísticos. Además, es necesario interactuar con distintos ficheros y bases de datos para la extracción de la información deseada.

En el mercado podemos encontrar una extensa variedad de entornos de desarrollo matemáticos como Octave o Python, entre otros. Se ha optado por el uso de Matlab² 2017a con la licencia Académica de la Universitat Politècnica de València ya que es uno de los softwares más extendidos a nivel académico.

Matlab es una herramienta de software matemático que ofrece un entorno de desarrollo integrado con su propio lenguaje de programación. Es utilizado para procesamiento de señales, procesamiento de imágenes, visión artificial, comunicaciones, finanzas computacionales, diseño de control, robótica y muchos otros campos. Una gran ventaja es que Matlab puede ampliar sus capacidades mediante una extensa variedad de toolbox disponibles para la comunidad de usuarios.

Además de las funciones básicas de Matlab, en este proyecto se ha hecho uso de los toolbox descritos en las siguientes secciones.

²<https://es.mathworks.com/products/matlab.html>

4.6.1. DataBase Toolbox

El Database Toolbox de Matlab ofrece un conjunto de funciones y una app para trabajar con bases de datos relacionales. Permite la conexión con distintos gestores de bases de datos sin software ni drivers adicionales. Facilita tanto el acceso a la información de la base de datos como la modificación de la misma.

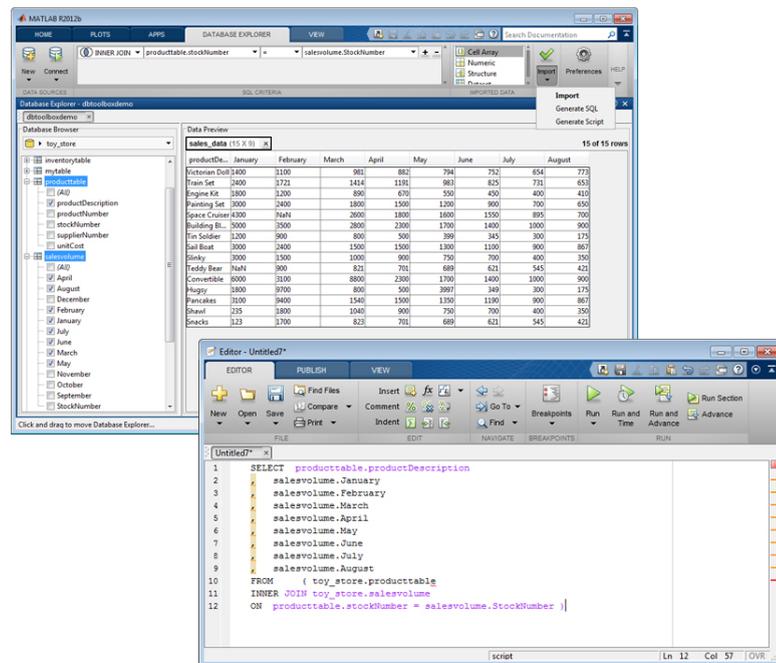


FIGURA 4.5: Database Toolbox (arriba) y código SQL generado automáticamente (abajo)

4.6.2. Statistics and Machine Learning Toolbox

Statistics and Machine Learning Toolbox proporciona funciones y apps para describir, analizar y modelizar datos.

En la sección 4.4 se ha descrito los métodos y necesarios para la realización de un análisis exploratorio de datos. Para ello, Matlab dispone de un extenso abanico de funciones que pueden ser utilizadas según las necesidades del usuario.

- Estadística descriptiva:
 - Medidas de tendencia central como la media aritmética (función *mean*).

- Medidas de dispersión, incluida la varianza, la desviación estándar y el rango intercuartílico (funciones *var*, *std* y *iqr*)
 - Construcción de tablas de contingencia (función *tabulate*)
 - Representación tanto de variables cuantitativas mediante histogramas y diagramas de cajas, como de variables cualitativas mediante diagramas de barras y de sectores (funciones *histogram*, *boxplot*, *bar* y *pie* respectivamente).
- **Contraste de hipótesis:**
- Pruebas para determinar la dependencia entre una variable cuantitativa y otra cualitativa. Como la prueba t en el caso de distribuciones paramétricas, y la prueba Mann-Whitney-Wilcoxon o prueba de suma de rangos Wilcoxon en el caso de distribuciones no paramétricas (funciones *ttest2* y *ranksum* respectivamente)
 - Pruebas para determinar la dependencia entre dos variables cualitativas a partir de las tablas de contingencia como la prueba χ^2 o la prueba de Fisher (funciones *crosstab* y *fishertest* respectivamente).
 - Pruebas para determinar la independencia entre dos variables (función *chi2gof*).

4.6.3. App Classification Learner

Esta app pertenece al paquete Statistics and Machine Learning Toolbox. Permite ejecutar varios algoritmos para entrenar y validar diferentes tipos de modelos de clasificación. Así, se podrán comparar diferentes modelos y elegir el más adecuado. Entre los tipos de modelos disponibles se encuentran los Árboles de decisión, Análisis discriminante, Support Vector Machines, Regresión logística, k-Vecinos más cercanos y Conjunto de clasificadores (Ensemble Classifiers). A su vez, cada tipo dispone de algoritmos con diferentes variantes.

- **Árboles de decisión:** Dispone de tres algoritmos diferentes en función del número máximo de divisiones que puede tener el árbol, desde la raíz hasta las ramas, para la toma de decisiones. Reciben el nombre de **Simple Tree**, **Medium Tree** y **Complex Tree**.

- Análisis discriminante: Este método asume que los datos que pertenecen a cada clase se basan en una distribución Normal diferente. Es decir, con un conjunto de datos dividido en dos clases se pueden obtener dos distribuciones Normales diferentes. Este método busca la frontera entre ambas distribuciones que permita clasificar el dato correctamente en una clase u otra. Este modelo cuenta con dos algoritmos diferentes según el tipo de frontera utilizada para la clasificación. **Linear Discriminant**, usa fronteras lineales, y **Quadratic Discriminant** usa elipses, parábolas o hipérbolas.
- Regresión Logística: El modelo disponible de regresión logística recibe el nombre de **Logistic Regression**.
- Support Vector Machines: Este modelo dispone de diferentes algoritmos en función del tipo de hiperplano utilizado para la clasificación de las muestras. Los algoritmos **Linear SVM**, **Quadratic SVM** y **Cubic SVM** generan hiperplanos mediante funciones lineales, cuadráticas y cúbicas, respectivamente. Para los problemas de clasificación con más dificultad son aconsejables los algoritmos **Fine Gaussian SVM**, **Medium Gaussian SVM** y **Coarse Gaussian SVM** que utilizan el método Kernel para la construcción del hiperplano.
- k-Vecinos más cercanos: Este modelo posee diferentes algoritmos en función de la métrica utilizada para calcular la distancia a los vecinos, así como del número de vecinos más cercanos que se tienen en cuenta para la clasificación. Los algoritmos **Fine KNN**, **Medium KNN** y **Coarse KNN** utilizan 1, 10 y 100 vecinos, respectivamente, y distancias euclídeas para la clasificación. Mientras que los algoritmos **Cosine KNN**, **Cubic KNN** y **Weighted KNN** utilizan distancia de coseno, distancia Minkowski y distancia ponderada, respectivamente, y un total de 10 vecinos para la clasificación de los datos.
- Conjunto de clasificadores (Ensemble Classifiers): Este método se basa en la división del conjunto de datos disponibles en diferentes subconjuntos de datos para crear un modelo (llamado clasificador) a partir de cada subconjunto. Los algoritmos **Boosted Trees** y **Bagged Trees** generan clasificadores a partir de modelos de Árbol de Decisión. En el primer caso, se generan secuencialmente nuevos dataset de entrenamiento mediante la asignación de pesos a las muestras que previamente han sido clasificadas erróneamente ya que esto permite que el nuevo modelo se

ajuste para clasificarlas de forma correcta. El resultado final de la clasificación será la ponderación, en función del error, de los resultados obtenidos para cada modelo. En el segundo caso, los nuevos dataset de entrenamiento se obtienen mediante el muestreo aleatorio con reemplazo del conjunto original, de manera que estos nuevos dataset para la creación de los clasificadores siempre serán más pequeños que el original. El resultado final será el promedio de los resultados para cada clasificador. El algoritmo **RUSBoost Trees** tienen el mismo fundamento que el Boosted Tree, se diferencian en que el RUSBoost Tree se utiliza para dataset con variables que tienen diferente número de muestras porque toma como referencia el número de muestras de la variable que menos datos tenga y escoge aleatoriamente ese número de muestras para las demás variables. Por otro lado, los algoritmos **Subspace Discriminant** y **Subspace KNN** utilizan los modelos de Análisis discriminante y k-vecinos mas cercanos, respectivamente. En estos casos se crea un conjunto de clasificadores entrenados usando un número concreto de variables para cada clasificador. El número de variables utilizadas para cada clasificador se puede escoger por el usuario, y son escogidas aleatoriamente del conjunto de datos cada vez que se entrena un nuevo clasificador.

La precisión de los modelos varía en gran medida dependiendo de las propiedades del conjunto de datos disponible, ya sea el número de muestras, naturaleza de las variables, etc. En la tabla 4.4 se detallan las características típicas de cada modelo en cuanto a velocidad³ y memoria⁴.

Modelo	Velocidad de predicción	Uso de memoria
Árbol de decisión	Rápido	Pequeño
Análisis discriminante	Rápido	Pequeño
Regresión logística	Rápido	Medio
Support Vector Machine	Medio para tipo lineal y bajo para el resto	Medio para lineal y alto para el resto
k-vecinos	Bajo para cúbico y medio para el resto	Medio
Conjunto de clasificadores	Entre rápido y medio	Entre bajo y alto

CUADRO 4.4: Velocidad y memoria de cada modelo

En esta App, es necesario elegir el esquema de validación que se quiere utilizar para determinar la precisión de los modelos. La validación sirve para estimar la precisión del

³Rápido: tiempo medio de ejecución 0,01 segundo; Medio: 1 segundo; Lento: 100 segundos

⁴Bajo: 1MB; Medio: 4MB; Alto: 100MB

modelo con nuevos datos frente a datos de entrenamiento utilizados para la creación del modelo. Matlab permite elegir entre tres esquemas de validación para ajustarse a las características del dataset utilizado.

- **Cross-Validation:** Divide el dataset en un número entero de particiones. Para cada partición, se crea un dataset de entrenamiento con todas las muestras disponibles excluyendo las muestras de dicha partición y se evalúa con las muestras de esa partición. Se realizan tantas interacciones como particiones de la muestra. Finalmente la precisión se corresponde con la media de error de validación en cada una de las interacciones.

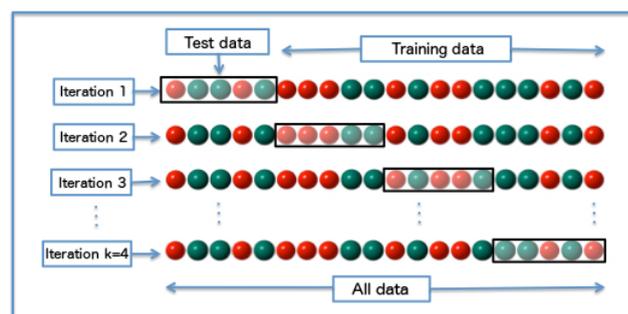


FIGURA 4.6: Descripción del método Cross-Validation

Este método es una manera eficiente de utilizar los datos, con lo que se recomienda cuando se dispone de datasets pequeños.

- **Holdout Validation:** Permite seleccionar un porcentaje de las muestras que se reservan para realizar la validación del modelo. Este tipo de validación se recomienda únicamente cuando se dispone de un dataset con una gran cantidad de datos.
- **No Validation:** No utiliza datos de validación. Calcula la precisión del modelo en base a los datos de entrenamiento. No es recomendable porque se obtienen estimaciones que no se corresponden con la realidad.

Una vez que se han obtenido los modelos, la app ofrece varios métodos para analizar los resultados obtenidos. Se puede observar un gráfico de dispersión que muestra las predicciones obtenidas haciendo uso del modelo frente a los datos reales, además de la matriz de confusión y el Área Bajo la Curva ROC del modelo correspondiente.

Es posible visualizar la curva ROC que representa el ratio de verdaderos positivos frente al ratio de falsos positivos, así como el ratio de verdaderos negativos frente al ratio de

falsos negativos. En esta curva aparece un marker para el valor del ratio de verdaderos positivos y el ratio de falsos positivos obtenidos para el modelo que se ha desarrollado. Una perfecta clasificación del modelo se representaría como una línea ROC recta en la parte superior de la gráfica, mientras que una pobre clasificación se correspondería con una línea de 45 grados.

4.6.4. Técnicas de reducción de la dimensionalidad

El hecho de disponer de gran cantidad de variables muy dispares conlleva la necesidad de recurrir a métodos de reducción de dimensionalidad con lo el fin de eliminar del modelado los datos que no aporten ningún tipo de información y conseguir el mínimo número de predictores posible.

El método de componentes principales identifica componentes principales, características o direcciones en las cuales agrupaciones de variables se mueven en sentidos similares. Esto puede deberse a factores conocidos o desconocidos, por ejemplo, que una variable sea consecuencia de otra variable, o que un grupo de variables se mueva en la misma dirección debido a la respuesta del paciente. Si esto sucede, se puede transformar el conjunto de variables original en un nuevo conjunto de denominado componentes principales. Este nuevo conjunto se construye mediante una transformación lineal derivada de la matriz de covarianza de las muestras. Cada componente principal está incorrelado entre sí y se ordena en base a la información que aporta, identificada por la varianza. Así, a primera componente principal tendrá mayor varianza, mientras que la última componente será la de menor varianza.

El número de componentes principales es igual al número de variables original, siendo igual la suma de las varianzas de las variables en ambos casos. La diferencia reside en que, si las variables originales están muy correlacionadas, con pocos componentes principales se puede obtener casi la misma información de variabilidad que con el conjunto completo de variables. Si las variables originales estuvieran completamente incorrelacionadas entre sí, el análisis de componentes principales no tendría interés, ya que las componentes principales coincidirían con las variables originales.

Matlab permite el uso de esta técnica de reducción de dimensionalidad en la app Classification Learner. De manera que el conjunto de datos se transforma en un conjunto de

componentes principales que serán utilizados para el entrenamiento del modelo seleccionado. Por defecto, solo se utilizarán los componentes que expliquen un total del 95 % de la varianza de los datos originales. Sin embargo, esta herramienta tienen la limitación de que no se aplica a las variables categóricas.

Otro método de reducción de dimensionalidad muy utilizado es el método de extracción de características empleando distancias euclidianas y componentes de vecindarios (NCA). Este método se utiliza con el fin de maximizar la precisión en algoritmos de clasificación. La función *fscnca* calcula, a partir de las variables de estudio y el resultado final, la probabilidad de una variable de influir en el resultado final. Con esto, asigna a cada variable un *peso* numérico según la relevancia de la variable, dando peso 0 a las variables que no tienen ninguna relevancia sobre la clasificación. De esta manera se pueden excluir del modelo de clasificación las variables que no aporten ninguna información para la clasificación.

4.6.5. Metodología de validación

La metodología de validación se basa en una serie de iteraciones donde se evalúan los modelos obtenidos y se intenta mejorar la precisión de los mismos. Para ello, se utiliza la siguiente técnicas:

- Aplicación de métodos de reducción de dimensionalidad en el caso de que el conjunto de variables predictoras sea muy elevado (> 10 variables).
- Modificación del conjunto de variables de entrada utilizadas, según el interés que tenga su aportación al modelo.
- Análisis de las variables significativas y exclusión en los casos que se requiera.
- Modificación del esquema de validación de los modelos, cambiando la proporción de muestras utilizadas para entrenamiento y prueba.

Aplicando esta metodología, se ha fijado un máximo de 5 iteraciones para encontrar el modelo que más se ajusta a los requerimientos del proyecto.

4.7. Coreografía de servicios

El objetivo de este trabajo no solo consiste en la obtención de un modelo predictivo aislado para ser usado por personal experto, sino que se pretende aportar un sistema práctico y real con un modelo de predicción que pueda ser utilizado por profesionales médicos en un entorno clínico. Las TIC juegan un papel crucial para la implementación de un sistema con estas características, aportando herramientas y soluciones para el intercambio de información. El objetivo que se persigue es la implementación e integración de un sistema cuyas principales tareas se han dividido en las siguientes: *i)* Proveer a usuarios no expertos en el sistema una plataforma de acceso al modelo de predicción. *ii)* Ejecutar el modelo para la obtención de los resultados de la predicción.

Una de las herramientas más conocida y utilizadas para este tipo de problemática son los Servicios Web (Web Services). Es un herramienta que aporta soluciones para el intercambio de información entre aplicaciones software sin necesidad de que utilicen el mismo lenguaje. Esto se consigue gracias al uso de estándares ya definidos. Aunque su uso es muy extendido, los Servicios Web tienen ciertas limitaciones. En cuanto a escalabilidad y dinamismo del diseñador, hay herramientas que ofrecen soluciones más acordes.

Una Arquitectura Orientada a Servicios (SOA) es un tipo de arquitectura destinada al intercambio de información que se basa en el uso de servicios. Una SOA asume que las todas funcionalidades del sistema se pueden englobar en unidades independientes llamadas servicios. Los servicios se comunican e interactúan entre sí mediante un protocolo de comunicación establecido. Dentro de esta arquitectura surgen dos paradigmas: coreografía y orquestación. La coreografía se basa en que los servicios que componen el sistema saben en qué momento se deben ejecutar y con qué servicio deben interactuar en cada caso, sin que surja la necesidad de un coordinador. Por el contrario, la orquestación se basa en la existencia de un componente central que sirve como coordinador de los servicios del sistema.

Este proyecto se ha basado en un sistema de orquestación ya existente [24], desarrollado por el grupo de Innovaciones Tecnológicas para la Salud y el Bienestar⁵ (SABIEN) el cuál forma parte del Instituto de Aplicaciones de las Tecnologías de la Información

⁵<http://www.sabien.upv.es/>

y de las Comunicaciones Avanzadas⁶ (ITACA) adscrito a la Universidad Politécnica de Valencia⁷(UPV). Este sistema orquestado, dispone de un coordinador que define exactamente el orden e interacción de los servicios disponibles. El coordinador central (orquestador) controla todos los servicio y la interacción entre ellos, y además contiene la declaración del total de servicios disponibles y su flujo de ejecución. Los servicios pueden conectarse al núcleo localmente o remotamente, usando un servicio encapsulado en un protocolo TCP.

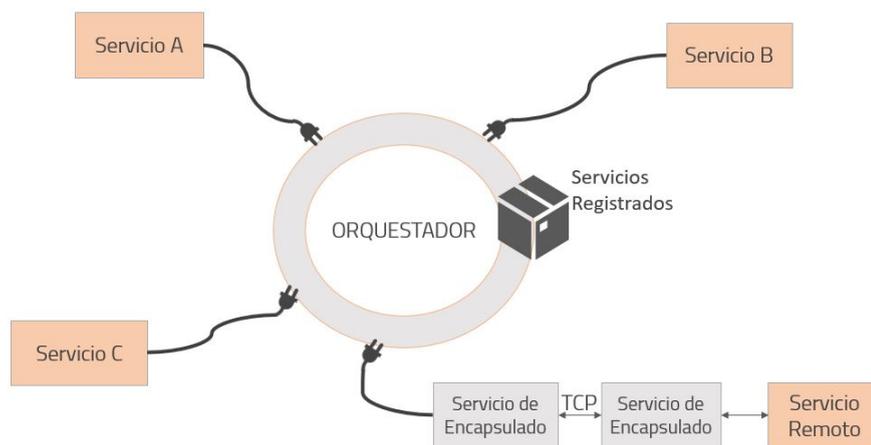


FIGURA 4.7: Esquema de un sistema orquestado

La comunicación entre servicios se realiza mediante el protocolo XMSG de envío de mensajes XML. La tecnología XML trata de expresar la información de forma estructurada y fácilmente interpretable.

En la figura 4.8 se puede observar un ejemplo de mensaje XMSG. En la primera parte del mismo se define el servicio remitente, el destinatario y el tipo de mensaje (Request, Inform, Event, Failure, Admin y NoSpecified). En la siguiente parte del mensaje se especifica el contenido, donde se define la llamada al método del servicio requerido. En el ejemplo, el mensaje ha sido enviado desde el Servicio A, cuya dirección es `java.ServicioA`, hasta el Servicio B cuya dirección es `java.orchestrator.ServicioB`. En el cuerpo del mensaje se invoca al método `LigthProcess`, el cual necesita las direcciones de las luces como parámetro de entrada.

⁶<http://www.itaca.upv.es>

⁷<http://www.upv.es/>

```

<xmsg>
  <message-id>18f4a193e342-4a80-93ca-9f67a60adb1</message-id>
  <sender> java.ServicioA</sender>
  <receiver>java.orchestrator.ServicioB</receiver>
  <protocol>request</protocol>
  <language>XMSG</language>
  <content>
    <Envelope xmlns:soap="http://www.w3-org/2003/05/soap-envelope">
      <Body>
        <LigthProcess xmlns="http://tempuri.org">
          <ligth1 type="xs:string">
            java.orchestrator.ligth.1
          </ligth1>
          <ligth1 type="xs:string">
            java.orchestrator.ligth.2
          </ligth1>
          <ligth1 type="xs:string">
            java.orchestrator.ligth.3
          </ligth1>
        </LigthProcess>
      </Body>
    </Envelope>
  </content>
</xmsg>

```

FIGURA 4.8: Ejemplo mensaje XMSG

4.8. Propuesta de sistema de ayuda a la decisión clínica

En base a la arquitectura de un sistema orquestado se ha propuesto un conjunto de tres módulos que definen funcionalmente el sistema de ayuda a la decisión clínica.

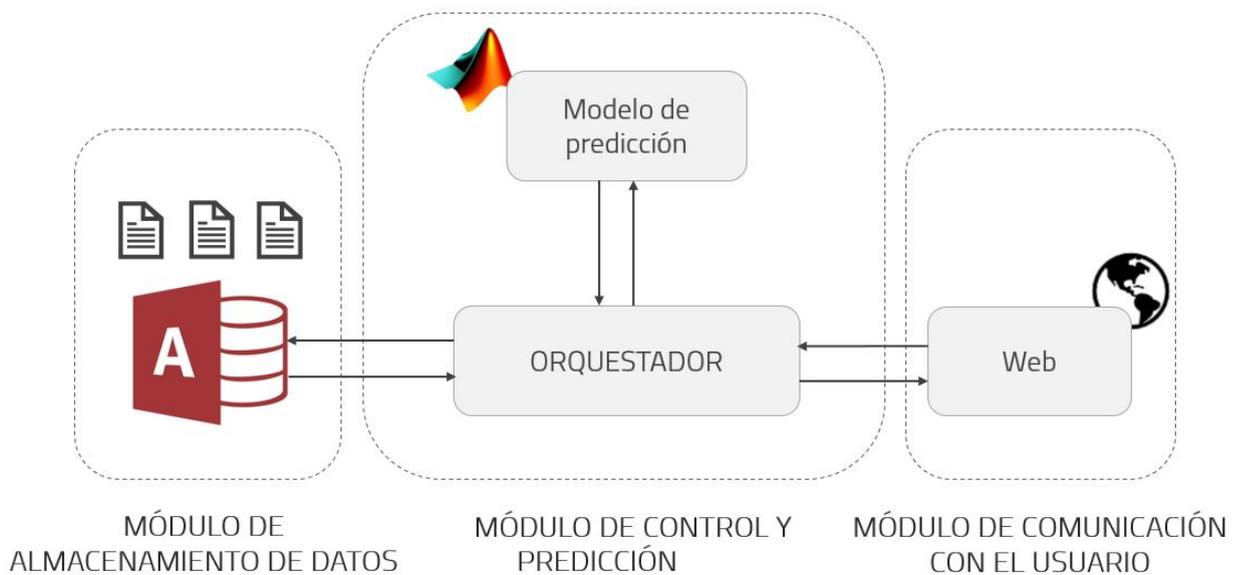


FIGURA 4.9: Estructura funcional del sistema

1. Módulo de Almacenamiento de Datos: contiene una base de datos donde se puede almacenar un histórico de los resultados de las predicciones obtenidos por el usuario mediante la ejecución del modelo de clasificación.

2. Módulo de Control y Predicción: está formado por dos componentes. El componente orquestador, definido en el apartado anterior, el cuál es el director funcional que controla el flujo de servicios del sistema. Y el componente predictivo, que contienen el motor para ejecutar el modelo y obtener los resultados de predicción en base a unos datos de entrada.
3. Módulo de Comunicación con el Usuario: diseñado para proveer a los usuarios finales una aplicación web donde introducir los datos deseados sobre un determinado paciente y recibir los resultados de la ejecución del modelo de predicción.

Capítulo 5

Resultados

5.1. Estructuración de los datos

La tarea inicial ha consistido en crear un documento descriptivo de cada una de esas variables donde se ha detallado una breve descripción de las mismas, el tipo de dato almacenado, los posibles valores que pueden tomar o el rango, así como los caracteres especiales. Este análisis resulta imprescindible para conocer el alcance del conjunto de datos y poder llevar a cabo una correcta manipulación de los datos de los que se dispone.

Los datos de los pacientes han sido recogidos desde el momento en el que fueron seleccionados para el trasplante hasta el final del estudio, tomándose los primeros datos antes de la intervención y prolongándose en ventanas temporales de días, semanas o meses hasta que se alcanzan los 2 años posteriores al trasplante, a partir de ese momento los datos se han tomado anualmente. En cada columna se han almacenado los datos correspondientes a una variable observada en un determinada fecha relativa al momento del trasplante. Esto supone que en la base de datos inicial se encuentran las variables duplicadas tantas veces como número de observaciones se ha realizado al paciente a lo largo del tiempo. Por ejemplo, las cifras de colesterol de un paciente medido a los 3 meses y a los 6 meses del trasplante se han tomado como dos variables distintas. Además, no todos los datos han sido recogidos sistemáticamente en periodos de tiempo iguales. Atendiendo a toda la información de la que se dispone, se ha concluido que todos los datos se puede clasificar atendiendo a tres categorías: nombre de la variable observada, número del paciente del que se ha obtenido la muestra y periodo de tiempo en que se extrajo la muestra.

Ante este panorama, se ha optado por la creación de una base de datos relacional. Para ello se ha utilizado el SGBD Microsoft Access 2016 debido a que su uso es muy intuitivo porque cuenta con un entorno similar a todos los productos de Microsoft, es muy utilizada en el ámbito académico y es adecuada para el volumen de datos de este estudio.

La base de datos se ha estructurado en 4 tablas:

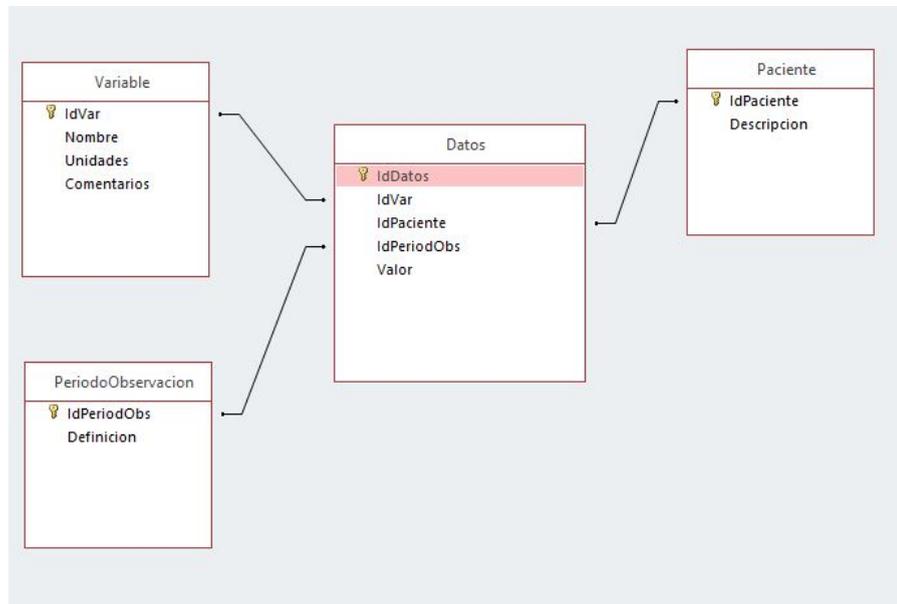


FIGURA 5.1: Esquema de la base de datos relacional

- Tabla Variable: Contiene el número identificador de cada variable junto con el nombre de la misma, las unidades en las que se mide y una breve descripción.
- Tabla Paciente: Contiene el número identificador de cada paciente junto con un nombre asociado.
- Tabla Periodo Observación: Se han establecido un total de 23 periodos de observación atendiendo a la fecha de trasplante.
- Tabla Datos: Se considera la tabla principal, ya que está enlazada con el resto de tablas y contiene el total de datos recopilados en el estudio. Éstos están clasificados según el identificador de la variable que se trata, el identificador del paciente al que pertenecen y el momento temporal en el que se tomaron esos datos respecto.

Para la inserción de los datos en la base de datos Access diseñada, ha sido imprescindible la realización de varios pasos intermedios para estructurar los datos del origen según el

esquema deseado. Para ello se ha realizado una función en Matlab cuya misión es cargar los datos desde el origen, dividirlos en función de la variable a la que pertenecen y dentro de la misma, clasificarlos de forma estructurada según el identificador del paciente y el periodo de observación. Todas las variables se han concatenado en una matriz siguiendo esta estructura y añadiendo un identificador en cada fila para disponer de una clave única en dicha matriz. Con el fin de obtener la mayor cantidad de información posible en la base de datos, en este proceso también se han transformado los datos de las variables categóricas, sustituyendo el código que identifica cada categoría por el significado de la categoría en sí misma. Es decir, en el caso de una variable categórica como puede ser el sexo del paciente, se ha sustituido el código 1 y 2 por el valor que representa Hombre y Mujer respectivamente. Según esta estructura, la matriz obtenida consta un total de 252.724 filas (82 pacientes x 23 periodos de observación x 134 variables). Sin embargo, teniendo en cuenta que no todas las variables han sido observadas a lo largo de todos los periodos de observación ni tienen datos para la totalidad de pacientes, se han eliminado los registros cuyo valor estaba vacío o se consideraba nulo, resultando una matriz con un total de 22.385 filas.

Esta matriz de datos se han cargado en la base de datos Access directamente desde el workspace de Matlab creando una conexión por medio del toolbox Database Explorer desde donde se ejecuta la sentencia SQL para la inserción de datos en la misma. Esta matriz se corresponde con la tabla llama Datos. El resto de tablas se han introducido los datos manualmente desde Access ya que la extensión de las mismas es mucho menor.

5.2. Estudio descriptivo de los datos

Según su naturaleza, las variables se pueden clasificar como cuantitativas o numéricas y cualitativas o categóricas. El conjunto de variables de estudio cuenta con un total de 134 variables, las cuales se han dividido según esta clasificación para la realización del análisis más adecuando en cada caso atendiendo a esta clasificación. Hay un gran número de variables que se han recogido en más de un periodo de observación, en estos casos se ha realizado en el análisis para el primer periodo de observación registrado ya que suele corresponder al periodo con mayor cantidad de muestras, lo que aportará una descripción más significativa de la variable.

El conjunto de variables consta de un total de 62 variables cuantitativas. Como se ha detallado en el apartado 4.4, el análisis descriptivo más acorde para este tipo de variables consiste en la obtención de la media aritmética y la desviación típica de las muestras. En la tabla 5.1 se muestra un resumen de los resultados obtenidos para un resumido conjunto de variables, los resultados del conjunto total de variables se muestran en el anexo A. De estos resultados se han excluido las variables *FechaTXPR* y *AÑO_NACIMIENTO* debido a que presentan un formato de fecha *DD/MM/AAAA*. Estas dos variables no se han utilizado en ningún análisis ya que ofrecen información redundante sobre el paciente. En su lugar, se dispone de las variables *AÑO_DEL TRASPLANTE* y *AÑOS_DEL PACIENTE AL TRASPLANTE* que aportan toda la información necesaria para el estudio.

Variable	Media	Desviación	N
PESO	66,599	12,735	82
IMC	24,045	3,391	82
TAS	141,173	23,693	75
TAD	77,224	11,851	76
GLUCEMIA	186,901	83,026	81
HBA1C	8,191	1,596	79
UREA	110,382	41,331	76
FG	16,557	13,183	14
CT	159,205	41,272	73
HDL	47,957	15,949	70
LDL	84,943	30,235	70
TG	134,919	64,424	74
PEPTIDO_C	0,232	0,204	82
AÑO_DIAG_DM	1983,573	6,821	82
AÑOS_DE_EVOLUCION_DE_LA_DM	25,524	6,508	82
AÑOS_EN_DIALISIS	1,914	1,075	81
UIDIA	37,622	19,619	82
AÑO_DEL TRASPLANTE	2009,354	3,557	82
AÑOS_DEL PACIENTE AL TRASPLANTE	37,366	5,717	82
MESES_EXITUS_TRAS_TX	6,5	13,005	8
MESES_EXPLANTEPANC_TRAS_TX	2,1	4,254	10
MESES_EXPLANTERENAL_TRAS_TX	2,5	4,359	4
PAK_TIEMPO_TRAS_TX_RENAL	59,5	62,089	4
Edaddonante	31,077	9,379	78
INSULINA	18,68	9,073	5

CUADRO 5.1: Resumen de resultados del análisis descriptivo de las variables cuantitativas

El total de 72 variables restantes, se han clasificado como variables cualitativas. En el caso de las variables cualitativas, el análisis más adecuado consiste en la obtención del número de muestras que pertenecen a cada categoría y el porcentaje sobre el total. En la tabla 5.2 se muestran los resultados obtenidos para un número reducido de variables, los resultados para el conjunto de variables completo se encuentran en el anexo A. En este análisis no se han incluido la variable *TIPOANTIHTA* debido al tipo de codificación que presenta, además de que se han detectado una gran cantidad de muestras erróneas.

Variable	%	N
Sexo		
Hombre	58,537	48
Mujer	41,463	34
EXITUS		
No	90,244	74
Si	9,756	8
PERD_INJERTO		
Si	31,707	26
No	68,293	56
TIPO_TRASPL		
SPR	98,78	81
PAK	1,22	1
EXPLANTE_DE_PANCREAS		
Si	13,415	11
No	86,585	71
NEC_INSULINA_TRAS_TX		
No	65,672	44
Si	34,328	23
Trasplantepreviorenal		
No	97,468	77
Si	2,532	2
Causafallecimientodonante		
CCV	44,156	34
TCE	42,857	33
Otras	12,987	10
Sexodonante		
Hombre	52,055	38
Mujer	47,945	35
RECIDIVA_DM1		
No	96,341	79
Si	3,659	3

CUADRO 5.2: Resumen de resultados del análisis descriptivo de las variables cualitativas

Los métodos gráficos tienen mucha presencia en el análisis descriptivo de datos ya que permite extraer información a simple vista de forma muy sencilla.

La figura 5.2 muestra gráficamente la información proporcionada por la variable *AÑOS_DEL_PACIENTE_AL_TRASPLANTE*. Como se ha visto en la sección 1.2, la edad está íntimamente ligada a la incidencia de diabetes. La edad media de los pacientes en el momento del trasplante es de 37,4 años, con una desviación típica de 57,5 años. La franja de edad en la que más trasplantes se han producido es entre los 35 y 39, disminuyendo significativamente el número de trasplantes a partir de los 45 años.

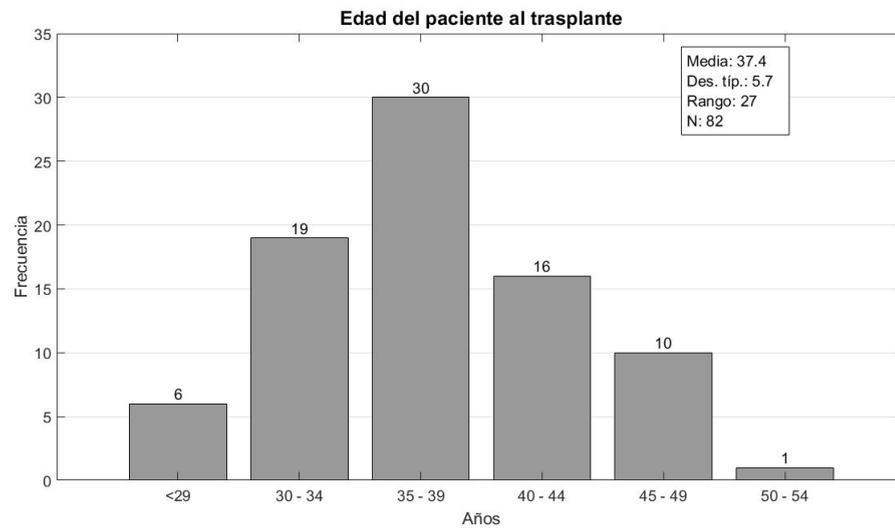


FIGURA 5.2: Frecuencia de pacientes en función de la edad al trasplante

También se ha mencionado anteriormente que el control glucémico más fiable se realiza mediante la monitorización de los niveles en sangre de glucosa y HbA1c. Se corresponde con las variables *GLUCEMIA* y *HBA1C* respectivamente. El valor medio de la distribución de HbA1c es de 8,2%, con una desviación típica de 1,6% (figura 5.3).

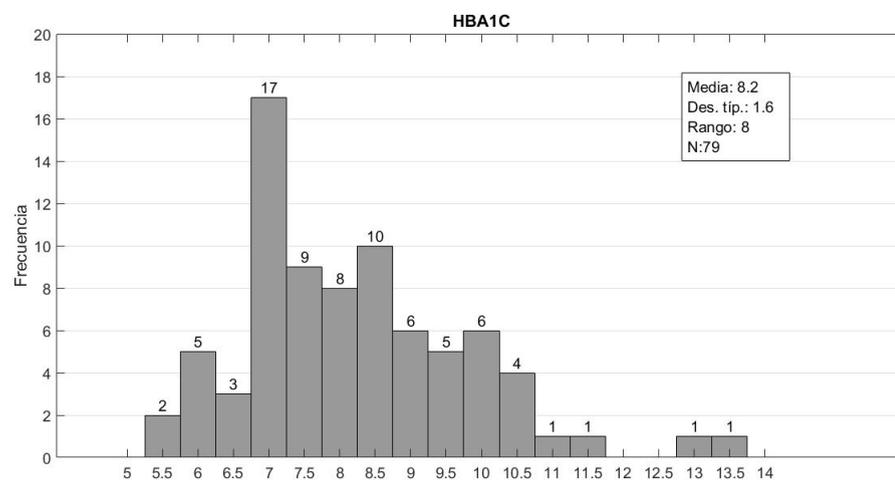


FIGURA 5.3: Distribución de la variable *HbA1c* previa al trasplante

El valor medio de glucemia, reflejado en la figura 5.4, es de 186,901 mg/dl y la desviación típica es 83,026 mg/dl.

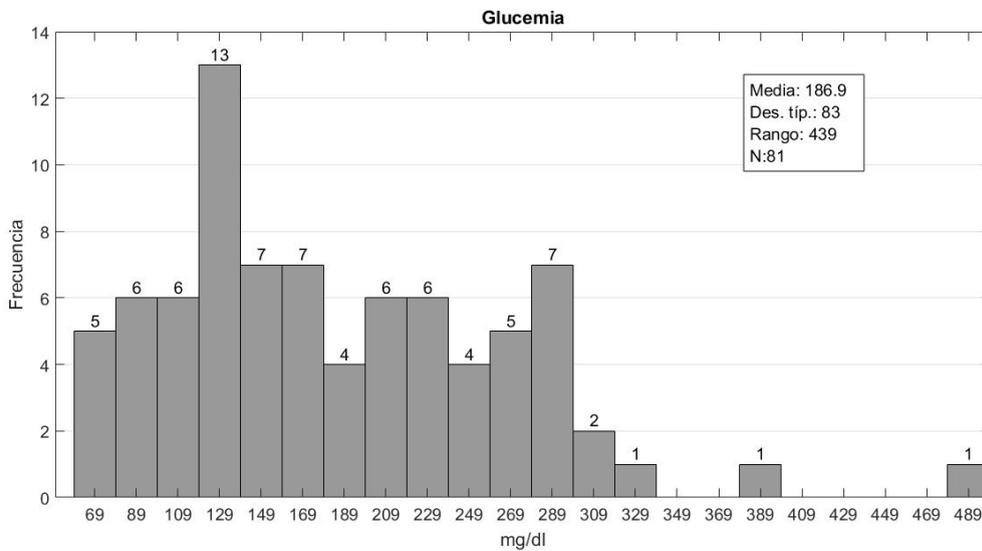


FIGURA 5.4: Distribución de la variable *GLUCEMIA* previa al trasplante

En cuanto a características generales de los pacientes del estudio, de los 82 trasplantes que se realizaron 48 casos eran hombres (58,8%) y 34 (41,5%) eran mujeres (figura 5.5).

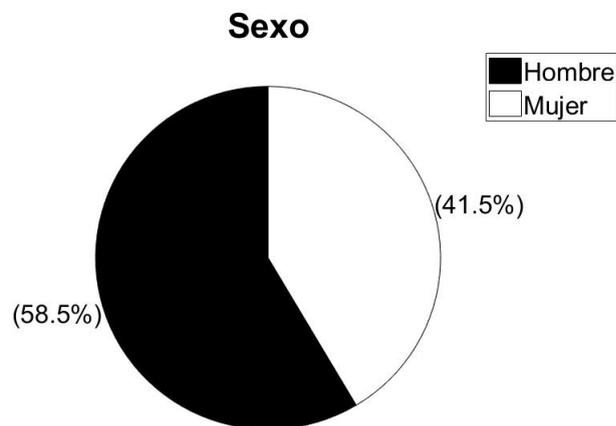


FIGURA 5.5: Distribución de la variable *SEXO*

En la gráfica 5.6 se detalla el número de trasplantes realizados cada año desde 2002 hasta 2015.

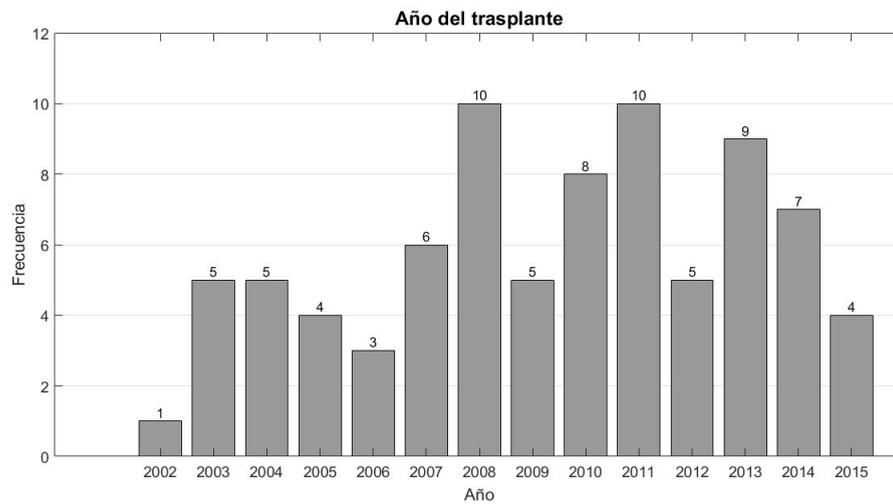


FIGURA 5.6: Número de trasplantes realizados por año

Cuando se finalizó el estudio, el número total de fallecidos ascendió a 8 (9,8%). Sin embargo, el gráfico 5.7 refleja una mayor cantidad de fallecimientos entre los pacientes trasplantados en la primera mitad del periodo de estudio. Siendo la supervivencia para el primer periodo (año 2002 al año 2008) del 85,3% y para el segundo periodo (año 2009 al año 2015) del 93,8%.

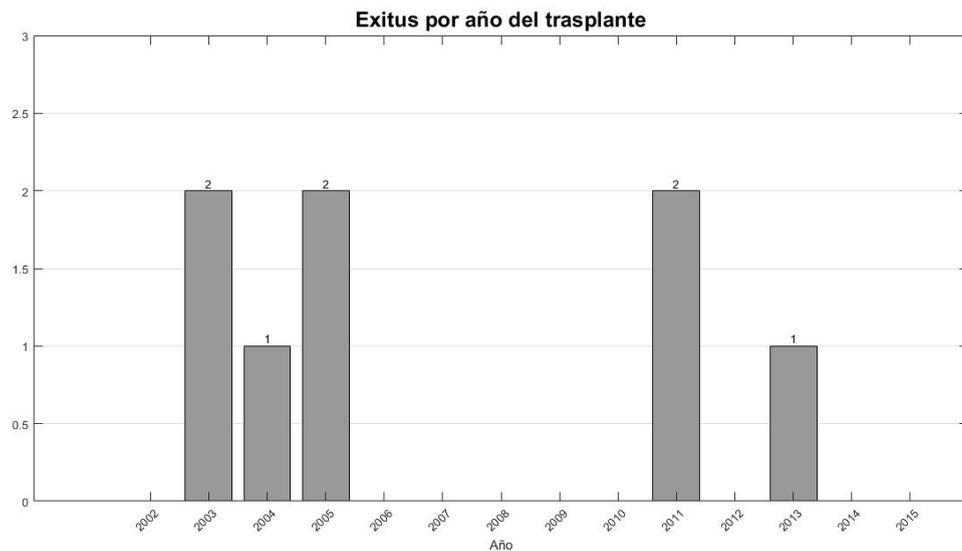


FIGURA 5.7: Número de fallecidos por año en que se realizó el trasplante

Según se ha detallado en secciones anteriores, el valor de cada variable ha sido tomado

en uno o varios periodo de observación diferentes. De modo que se puede diferenciar claramente entre variables que únicamente se han registrado en un periodo de observación y variables de las que se han tomado muestras sistemáticamente a lo largo del tiempo, a partir del momento en el que se produjo el trasplante. Ya sea porque estas variables únicamente tienen interés clínico observadas en cierto momento temporal o bien porque no tienen significado temporal. Por ejemplo, las variables *SEXO* y *CAUSA_EXITUS* no tienen sentido temporal por lo que únicamente se han registrado una vez, mientras que variables como *IAM* o *ACV* sí tienen sentido temporal pero solo ha interesado, según el criterio médico del profesional encargado de la recogida de las variables de estudio, registrarlas antes del trasplante.

Con esta premisa, las variables también se pueden clasificar en temporales o no temporales atendiendo a su significado. Por ello, se ha realizado otro tipo de análisis descriptivo para las variables temporales, ya que lo que se pretende es observar su evolución a lo largo del tiempo.

En este caso, únicamente se han analizado las variables que tienen más interés a nivel clínico. Por ejemplo, los niveles de creatinina, urea en sangre y la tasa de filtración glomerular (variables *CREATININA*, *UREA* y *FG* respectivamente) son especialmente interesantes porque hacen referencia a la función renal del paciente.

La creatinina es un compuesto que se forma en los músculos y células nerviosas. Este compuesto se elimina a través de los riñones, de modo que sirve como parámetro para determinar la función de filtrado renal. Un aumento de los niveles de creatinina en sangre es indicador de fallo renal. Un rango normal de creatinina en sangre se encuentra entre 0,6 y 1,3 mg/dl, aunque también dependerá de la masa muscular del paciente. En el siguiente gráfico (figura 5.9) se ha representado, mediante un conjunto de gráficos boxplot, las variaciones del nivel de creatinina en los pacientes a medida que aumenta el tiempo post-trasplante. Se puede observar que en el periodo Basal, es decir antes del trasplante, los brazos o *bigotes* de la caja abarcan un largo rango de valores, además de que el tamaño de la propia caja es bastante grande comparado con los demás periodos. Esto refleja unos niveles muy altos de creatinina, ya que la mediana se encuentra muy alejada de los valores normales, además de un conjunto de valores muy dispares en estos pacientes. En el resto de periodos, los tamaños de las cajas y los bigotes son mucho menores, además de que las medianas se encuentran prácticamente alineadas, lo

que indica que los valores de creatinina en la mayoría de pacientes se mantienen estables dentro de la normalidad a lo largo del tiempo. A partir de los tres 10 años post-trasplante apenas se puede diferenciar la caja debido a la escasez de datos disponibles para esos periodos.

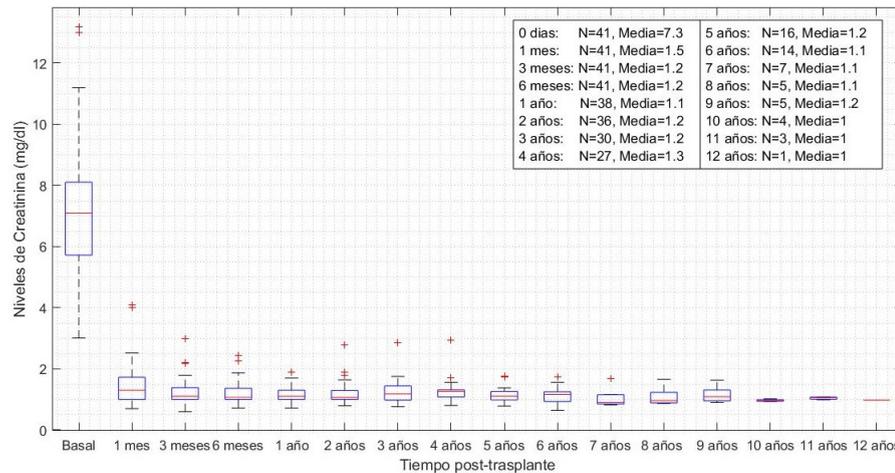


FIGURA 5.8: Evolución de los niveles de creatinina en los pacientes sin necesidad de insulina

La urea es el resultado final del metabolismo de las proteínas. Normalmente, los riñones filtran la urea de la sangre, pero cuando los riñones no funcionan bien, la cantidad de urea filtrada es menor y aumenta en la sangre. Los valores normales de urea en sangre son entre 7 y 20 mg/dl. Los valores más altos de 100 mg/dl se deben a un fallo renal importante. De nuevo, en el gráfico 5.9 se observa una caja y bigotes para el periodo Basal. En el resto de periodos, la media mantiene una tendencia dentro de unos límites aunque con cierta variación. En esta ocasión el tamaño de las cajas es mucho mayor frente al gráfico 5.8 lo que puede deberse a que, en la creatinina, el margen entre los niveles considerados normales es mucho menor y más crítica que en el caso de la urea. Para esta variable también se observa un número de casos menor para los últimos periodos de observación.

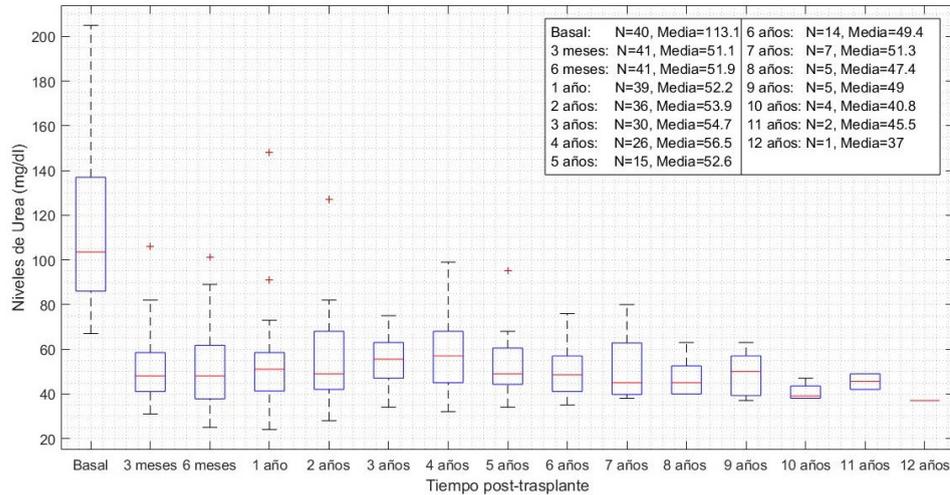


FIGURA 5.9: Evolución de los niveles de urea en los pacientes sin necesidad de insulina

El examen de la tasa de filtración glomerular mide la capacidad de los riñones para filtrar la sangre. Se consideran cifras normales entorno a 90-120 ml/min. Un resultado por debajo de 15 ml/min indica una insuficiencia renal total. En el siguiente gráfico (figura 5.10) podemos observar que, a diferencia de los gráficos anteriores, los niveles Basales se encuentran muy concentrados debido a que únicamente se dispone de datos de estudio para dos pacientes, por lo que no aporta ninguna información relevante. En los periodos restantes, se observa un tamaño de cajas y una mediana muy desigual. El desbalanceo en cuanto al número de muestras recogidas en cada periodo es muy notorio, llegando en algunos caso a representarse únicamente la caja sin los bigotes (a los 8 años, 10 años y 12 años).

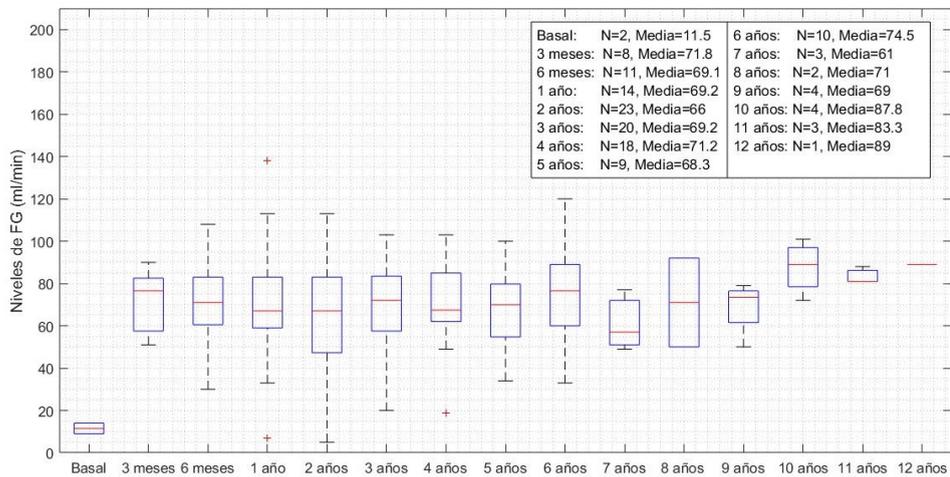


FIGURA 5.10: Evolución de los niveles de filtrado glomerular en los pacientes sin necesidad de insulina

En análisis descriptivo gráfico también ha contribuido a la detección de valores erróneos en las muestras mediante el uso de gráficos Boxplot. Teóricamente, estos gráficos consideran como erróneos los datos que se encuentran alejados del límite superior e inferior (límite de los bigotes). Sin embargo, en el ámbito médico esta restricción no es del todo exacta porque se pueden dar situaciones en las que se tomen valores anormalmente altos o bajos una muestra que no se consideren erróneos, sino que sean indicadores de enfermedad grave. Por lo tanto, estos valores no se pueden considerar erróneos. Por ejemplo, en la imagen 5.11 se ha representado un boxplot de la variable *IMC* de los pacientes que no han sobrevivido al trasplante.

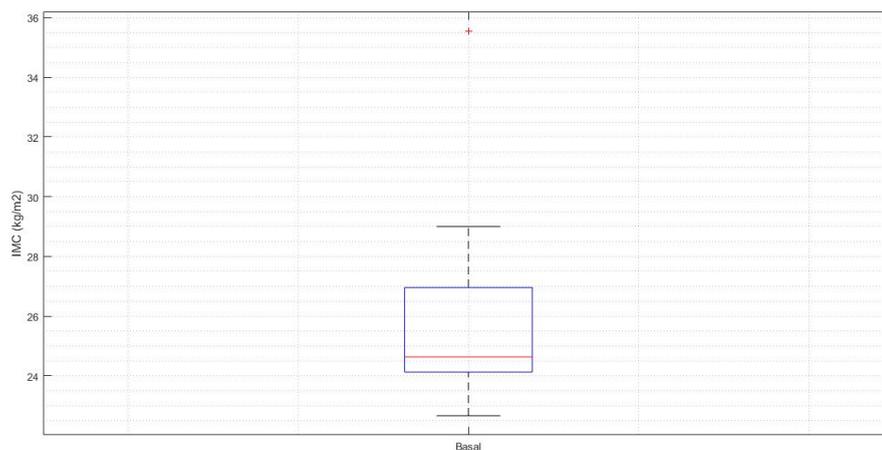


FIGURA 5.11: Gráfico de tipo Boxplot para la variable *IMC*

En este caso se puede observar un dato que excede en gran medida el umbral superior con lo que, en principio, sería un dato susceptible de ser erróneo. Sin embargo, se ha comprobado que este valor es tan alto debido a que el paciente padecía una obesidad importante. Por tanto, este tipo de situaciones se han analizado cuidadosamente. Finalmente, se han encontrado una serie de datos erróneos para ciertas variables en las que las muestras han sido tomadas erróneamente. La variable *CREATININA*, cuyos datos han sido tomados a los tres años del trasplante en pacientes que finalmente no han sido insulino dependientes, tiene una muestra errónea en uno de los pacientes ya que indica un valor excesivamente alto. Se ha corroborado con la médica directora del estudio que ese dato de 92 mg/dl se trata realmente de un valor de 0,92 mg/dl.

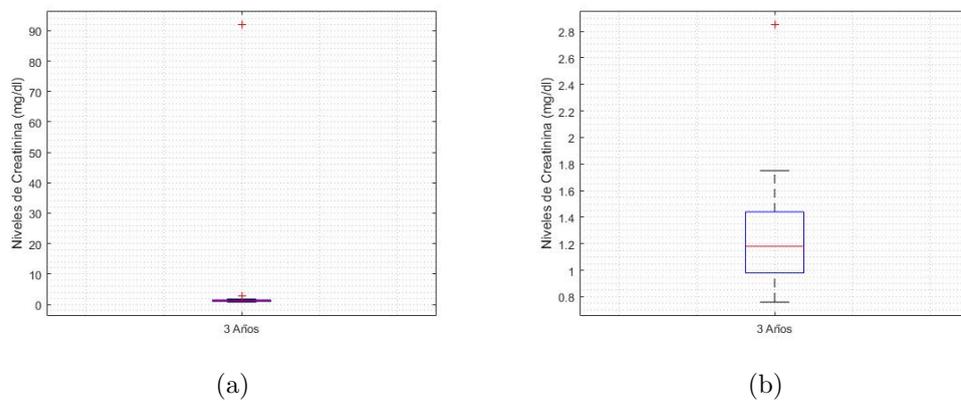


FIGURA 5.12: (a) Variable *CREATININA* con outlier ;
(b) Variable *CREATININA* sin outlier

De la misma manera, la variable *DosisCT* para estos mismos pacientes pero tomada a los 5 años tiene un valor erróneo, 75 mg, que realmente se corresponde con 7,5 mg.

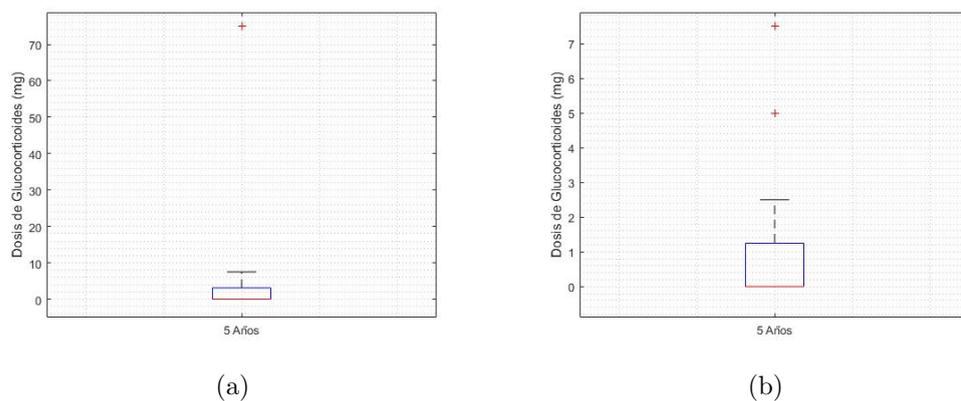


FIGURA 5.13: (a) Variable *DosisCT* con outlier ; (b) Variable *DosisCT* sin outlier

5.3. Análisis bivalente

El análisis multivariante de las variables de estudio está destinado a la identificación de variables que puedan contribuir directamente en un cierto evento final. En concreto, el análisis bivalente se centra en el efecto de una variable sobre otra. La naturaleza de cada variable dictamina el tipo de análisis que se puede realizar en cada caso, ver tabla 4.2.

En este estudio, el evento final siempre será una variable cualitativa que tomará dos valores y se comparará con otra variable que podrá ser cuantitativa o cualitativa.

En el caso de las variables cuantitativas, se realiza la prueba t-Student si siguen una distribución Normal. Si no se puede suponer la Normalidad de las muestras se utiliza la prueba no paramétrica Mann-Whitney-Wilcoxon. Para determinar si las muestras siguen una distribución Normal se utilizará el test de bondad de ajuste. El objetivo de estas pruebas es corroborar si los datos estudiados proceden de distribuciones con iguales medias, siendo ésta la hipótesis nula (H_0), quedando como hipótesis alternativa (H_1) que las medias son diferentes. Los elementos de interés son aquellos cuya $H_0=1$, indicando que la hipótesis nula puede ser rechazada con un nivel de significación del 5 %.

En el caso de variables cualitativas, se calcula la tabla de contingencia para cada variable. A partir de esta tabla, se puede analizar la significancia entre las clases de una variable mediante el test estadístico χ^2 y el p valor asociado. Se plantea como hipótesis nula (H_0) la independencia de las muestras de la variable analizada (ver tabla 4.3). Valores de p inferiores a 0.05 rechazan la H_0 con un 95 % de probabilidad. Por tanto, si el valor de p es inferior a 0.05 se rechaza H_0 , es decir que las muestras de la variable analizada tienen dependencia entre sí. Esto conlleva que pertenecen a distribuciones diferentes ya que un conjunto de variables aleatorias se consideran independientes e idénticamente distribuidas si cada variable aleatoria tiene la misma distribución de probabilidad y todas son mutuamente independientes.

El objetivo final del análisis bivalente es determinar si puede existir alguna variable susceptible de afectar directamente al resultado de un evento. En otras palabras, si el estado inicial de cierta variable observada en un paciente antes del trasplante puede condicionar directamente en el resultado final de dicho paciente después del trasplante. Por esta razón, en el análisis bivalente solamente se han analizado las variables registradas antes del trasplante o basales.

Debido a la gran cantidad de variables dispares disponibles en el estudio, se ha realizado un análisis exhaustivo para determinar el conjunto de muestras que pueden ser propensas a mostrar algún tipo de relación entre las variables. Esto ha llevado a realizar un total de 4 estudios multivariantes donde se comparan las variables de diferentes muestra poblacional atendiendo a diferentes resultados finales.

5.3.1. Evento final número 1: Supervivencia del paciente

Para este estudio multivariante se han tomado todas las muestras disponibles de las variables observadas antes del trasplante pertenecientes a los pacientes que han sido sometidos a una intervención SPK (Simultaneous Kidney Pancreas). El análisis se ha realizado en función de la supervivencia o no supervivencia del paciente después del trasplante.

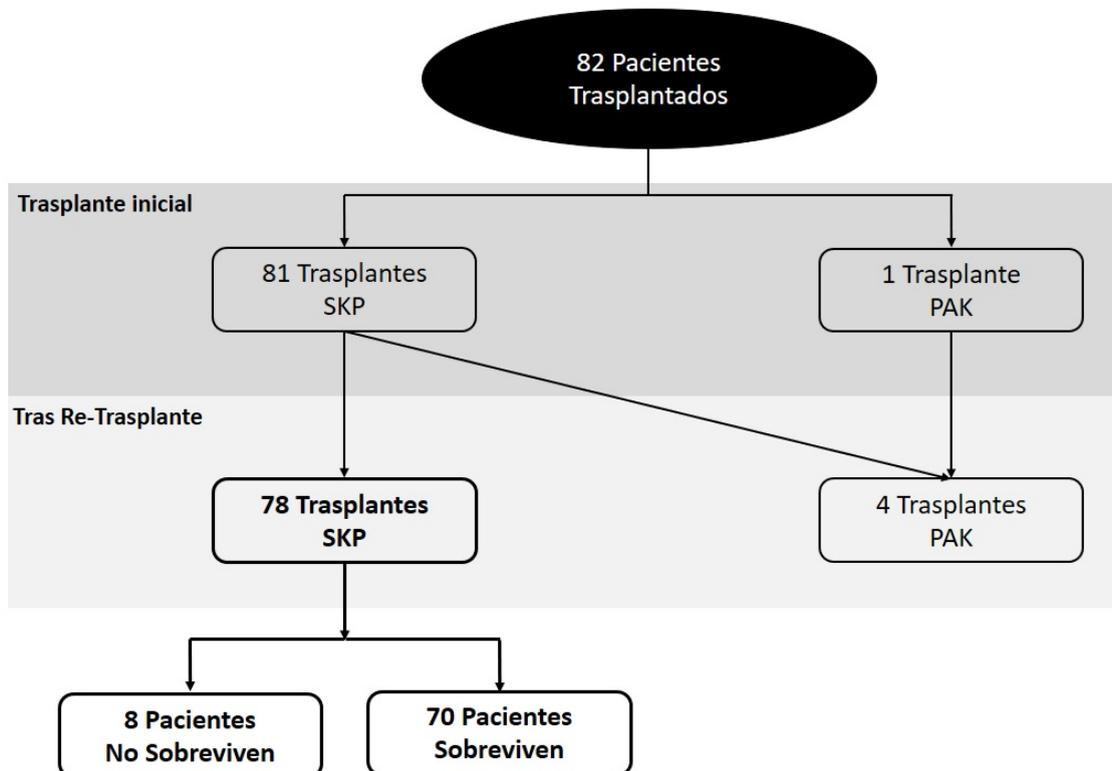


FIGURA 5.14: Descripción de las muestras utilizadas para el estudio en función de la supervivencia del paciente

Para las variables numéricas se ha elaborado una tabla que contiene el nombre de la variable, el número de muestras de pacientes que han sobrevivido (No exitus), el número de muestras de pacientes que no han sobrevivido (Exitus), el p valor y si se rechaza o se cumple la hipótesis nula (H_0), siendo $H_0=0$ si se acepta y $H_0=1$ si se rechaza.

En el caso de las variables categóricas, se ha elaborado una tabla de contingencia donde se refleja el porcentaje y número de muestras para No Exitus y Exitus, así como el valor χ^2 y el p.

Para este conjunto muestral se han obtenido los mostrados en las tablas A.3 y 5.4. Estas tablas contienen un resumen, los resultados del total de variables analizadas se encuentran en el anexo A.

Variable	No Exitus (N)	Exitus (N)	p Valor	H0
IMC	70	8	0,085	0
TAD	65	7	0,947	0
GLUCEMIA	70	7	0,127	0
HDL	61	6	0,662	0
LDL	61	6	0,394	0
TG	64	6	0,066	0
PEPTIDO_C	70	8	0,573	0
UIDIA	70	8	0,07	0
UI_KG_DIA	70	8	0,179	0
DOSIS_ESTATINA	56	5	0,318	0
ORTOSTATISMO	54	2	0,216	0
MANIOBRA_DE_VALSALVA	54	2	1	0
NEUROTESIOMETRO_PIE_DERECHO	53	6	0,008	1
ITB_DERECHO	36	1	0,8	0
ITB_IZQUIERDO	34	1	0,44	0
TALLA	70	8	0,055	0
AÑO_DEL_TRASPLANTE	70	8	0,031	1
AÑO_DE_INICIO_DE_LA_DIALISIS	69	8	0,017	1

CUADRO 5.3: Resumen de los resultados variables cuantitativas para el evento final 1

Variable	No Exitus		Exitus		χ^2	p Valor
	%	N	%	N		
Sexo						
Hombre	61,429	43	62,5	5	0,003	0,953
Mujer	38,571	27	37,5	3		
Ex-fumador	22,857	16	-	-		
DISLIPEMIA						
No	27,143	19	25	2	0,017	0,897
Si	72,857	51	75	6		
SISTEMA_NERVIOSO_PARASIMPATICO						
Afectación leve	3,636	2	-	-	14,541	0,002
Afectación severa	67,273	37	-	-		
Afectación moderada	27,273	15	50	1		
Normal	1,818	1	50	1		
RETINOPATIA_DIABETICA_EN_OJO_DERECHO						
Ceguera	2,857	2			11,683	0,069
Fotocoagulada	82,857	58	75	6		
Edema macular diabético	7,143	5	-	-		
Proliferativa	2,857	2	-	-		
Leve	1,429	1	-	-		
Moderada	2,857	2	12,5	1		
Ausente	-	-	12,5	1		
CATETERISMO						
Normal	22,727	15	-	-	2,724	0,256
No realizado	71,212	47	100	7		
Estenosis arteria coronaria	6,061	4	-	-		

CUADRO 5.4: Resumen de los resultados variables cualitativas para el evento final 1

5.3.2. Evento final número 2: Explante de páncreas

Para este estudio multivariante se han tomado las muestras de las variables de los pacientes que han sido sometidos a una intervención SPK y han sobrevivido al estudio. El análisis se ha realizado en función del explante o no explante del páncreas trasplantado debido a complicaciones sufridas por la intervención quirúrgica.

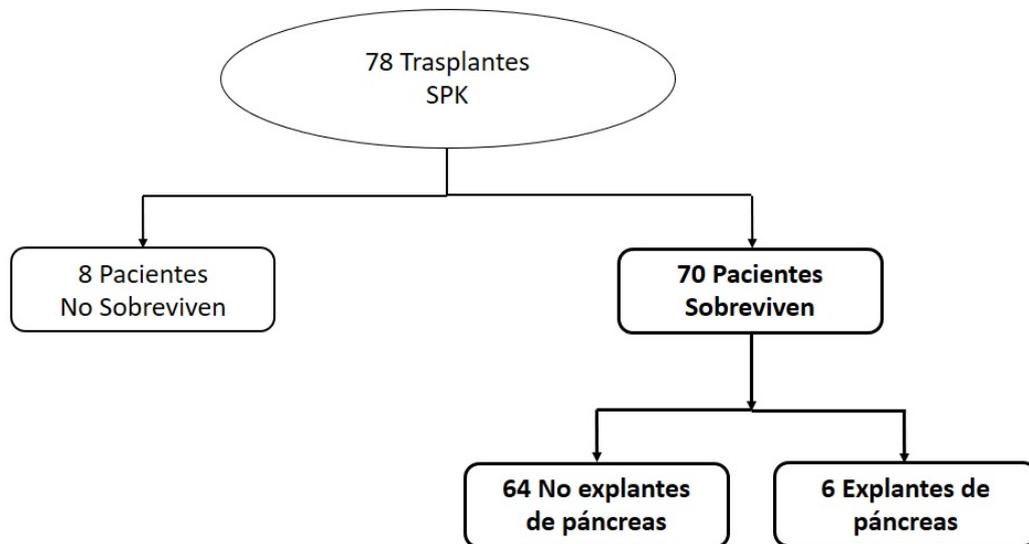


FIGURA 5.15: Descripción de las muestras utilizadas para el estudio en función del explante del páncreas

Para este conjunto muestral se han obtenido los siguientes resultados de forma resumida.

En el anexo A se encuentran los resultados para el conjunto total de variables.

Variable	No explante (N)	Explante (N)	p Valor	H0
AÑO_DIAG_DM	64	6	0,08	0
AÑO_DE_INICIO_DE_LA_DIALISIS	63	6	0,003	1
NºPINCHAZO	63	6	0,078	0
ANTIHTA	62	5	0,036	0
AÑOS_DE_EVOLUCION_DE_LA_DM	64	6	0,579	0
PEPTIDO_C	64	6	0,059	0
UREA	60	6	0,995	0
MANIOBRA_DE_VALSALVA	51	3	0,925	0
INDICE_3015	51	3	0,297	0
DEEPBREATHING	46	1	0,772	0
AÑO_DEL_TRASPLANTE	64	6	0,004	1
AÑOS_DEL_PACIENTE_AL_TRASPLANTE	64	6	0,343	0
ANTIHTA	62	5	0,036	0
DOSIS_ESTATINA	54	2	0,89	0
PESO	64	6	0,6	0
IMC	64	6	0,774	0
TAS	60	5	0,883	0
Edaddonante	61	5	0,585	0

CUADRO 5.5: Resumen de los resultados variables cuantitativas para el evento final 2

Variable	No explante		Explante		χ^2	p Valor
	%	N	%	N		
RMN_VASCULAR						
AE leve	17,188	11	-	-	9,979	0,019
Normal	78,125	50	66,667	4		
Severa	1,563	1	-	-		
AE moderada	3,125	2	33,333	2		
RETINOPATIA_DIABETICA_EN OJO_DERECHO						
Fotocoagulada	87,5	56	66,667	4	22,361	0,001
Leve	1,563	1	-	-		
Proliferativa	1,563	1	-	-		
Edema macular diabético	6,25	4	-	-		
Moderada	3,125	2	-	-		
Grave o pre-proliferativa	-	-	16,667	1		
Ceguera	-	-	16,667	1		
MANIOBRA_DE_VALSALVA2						
Anómalo	82,353	42	33,333	1	6,796	0,033
Normal	5,882	3	-	-		
Límite	11,765	6	66,667	2		
Causafallecimientodonante						
TCE	42,623	26	60	3	1,138	0,566
CCV	40,984	25	40	2	-	
Otras	16,393	10	-	1	-	
Sexodonante						
Hombre	47,368	27	80	4	1,958	0,162
Mujer	52,632	30	20	1		
DEEPBREATHING2						
Anómalo	89,362	42	100	1	0,119	0,942
Normal	4,255	2	-	-		
Límite	6,383	3	-	-		

CUADRO 5.6: Resumen resultados variables cualitativas para el evento final 2

5.3.3. Evento final número 3: Pérdida de injerto

Para este estudio multivariante se han tomado las muestras de las variables de los pacientes que han sido sometidos a una intervención SPK y han sobrevivido al estudio. El análisis se ha realizado en función de la pérdida o no pérdida de funcionalidad del injerto pancreático trasplantado.

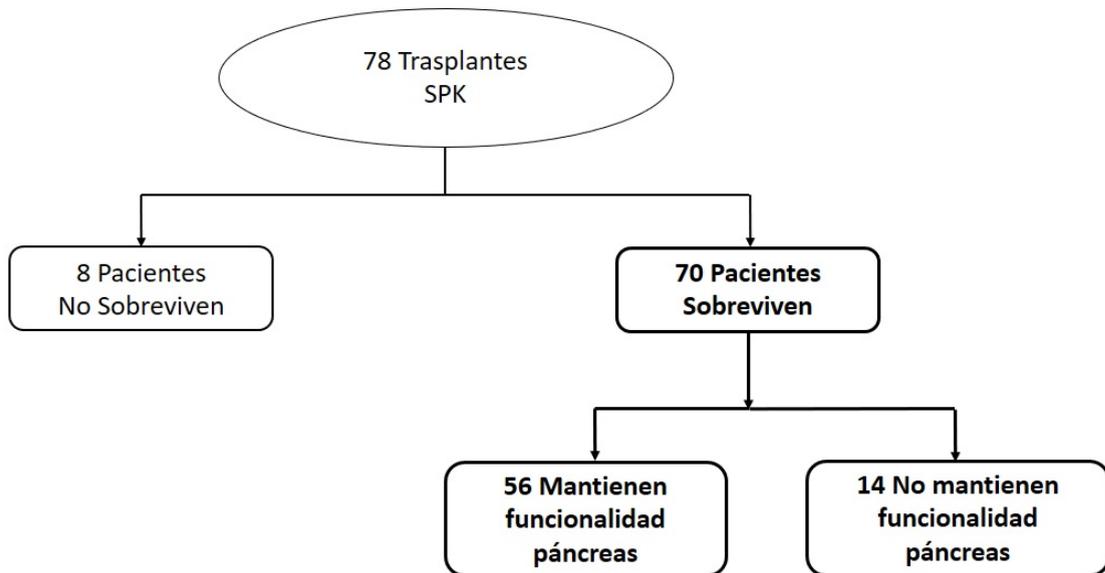


FIGURA 5.16: Descripción de las muestras utilizadas para el estudio en función de la pérdida de funcionalidad del páncreas

Para este conjunto muestral se han obtenido los siguientes resultados de forma resumida. En el anexo A se encuentran los resultados para el conjunto total de variables.

Variable	No Pérdida injerto (N)	Sí Pérdida injerto (N)	P	H0
PESO	56	14	0,708	0
IMC	56	14	0,859	0
UREA	53	13	0,117	0
TG	50	14	0,094	0
DOSIS_ESTATINA	49	7	0,877	0
ORTOSTATISMO	45	9	0,95	0
INDICE_3015	44	10	0,012	1
DEEPBREATHING	41	6	0,911	0
ITB_DERECHO	28	8	0,057	0
TALLA	56	14	0,783	0
N°CIG	56	13	0,041	1
CVmáximaCMV	55	14	0,083	0
AÑO_DEL_TRASPLANTE	56	14	0,006	1
AÑOS_DEL_PACIENTE_AL_TRASPLANTE	56	14	0,413	0
N°PINCHAZO	55	14	0,534	0
NEUROTESIOMETRO_PIE_IZQUIERDO	42	11	0,367	0
AÑO_DE_INICIO_DE_LA_DIALISIS	55	14	0,027	1
INCOMPATIBILIDADESHLA	40	12	0,074	0

CUADRO 5.7: Resumen resultados variables cuantitativas para el evento final 3

Variable	No Pérdida injerto		Si Pérdida injerto		χ^2	p Valor
	%	N	%	N		
Sexo						
Hombre	55,357	31	85,714	12	4,356	0,037
Mujer	44,643	25	14,286	2		
CATETERISMO						
No realizado	75,926	41	50	6	6,616	0,037
Normal	16,667	9	50	6		
Estenosis arteria coronaria	7,407	4	-	-		
RMN_VASCULAR						
AE leve	17,857	10	7,143	1	8,705	0,033
Normal	78,571	44	71,429	10		
Severa	1,786	1	-	-		
AE moderada	1,786	1	21,429	3		
RETINOPATIA_DIABETICA_EN OJO_DERECHO						
Fotocoagulada	87,5	49	78,571	11	13,854	0,031
Proliferativa	1,786	1	-	-		
Edema macular diabético	7,143	4	-	-		
Moderada	3,571	2	-	-		
Grave o pre-proliferativa	-	-	7,143	1		
Ceguera	-	-	7,143	1		
Leve	-	-	7,143	1		
RETINOPATIA_DIABETICA_EN OJO_IZQUIERDO						
Fotocoagulada	89,286	50	57,143	8	16,272	0,006
Proliferativa	1,786	1	7,143	1		
Edema macular diabético	5,357	3	14,286	2		
Moderada	3,571	2	-	-		
Ceguera	-	-	14,286	2		
Leve	-	-	7,143	1		
MANIOBRA_DE_VALSALVA2						
Anómalo	86,364	38	50	5	7,044	0,03
Normal	4,545	2	10	1		
Límite	9,091	4	40	4		
INDICE_3015_2						
Anómalo	75	33	40	4	12,282	0,002
Normal	11,364	5	60	6		
Límite	13,636	6	-	-		
DEEPBREATHING2						
VEBDonante						
Positivo	100	12	60	3	5,44	0,02
Negativo						

CUADRO 5.8: Resumen resultados variables cualitativas para el evento final 3

5.3.4. Evento final número 4: Necesidad de insulina

Para este estudio multivariante se han tomado las muestras de las variables de los pacientes que han sido sometidos a una intervención SPK, han sobrevivido al estudio y no han perdido la funcionalidad del injerto pancreático trasplantado, habiéndose realizado el trasplante antes del último año de estudio. El análisis se ha realizado en función de la necesidad o no de insulina del paciente después del trasplante.

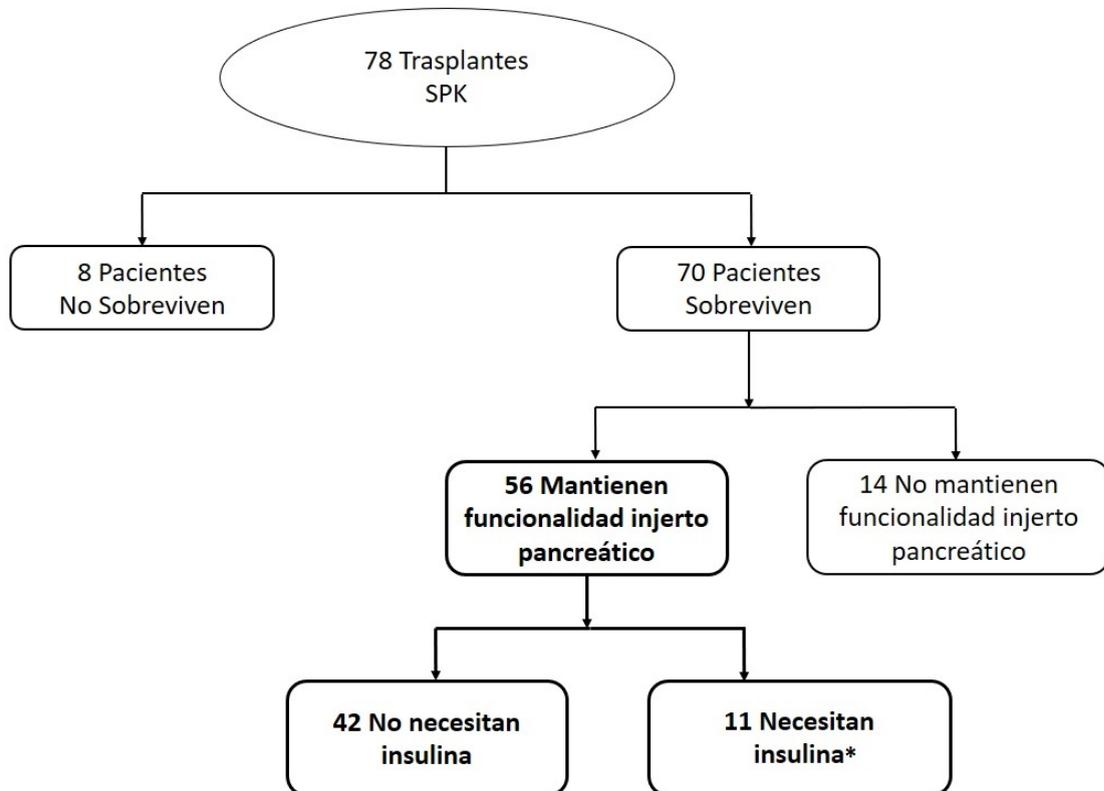


FIGURA 5.17: Descripción de las muestras utilizadas para el estudio en función de la necesidad de insulina

En la figura 5.17 se puede observar que de los 56 pacientes que conservan la funcionalidad del injerto pancreático, 42 pacientes no han necesitado insulina a largo plazo mientras que otros 11 pacientes si han resultado ser insulino dependientes. Los 3 pacientes restantes que aquí no se reflejan para el estudio bivariante, han sido trasplantado en el último año de estudio. Esto supone que, como medida para evitar el estrés del nuevo injerto, se inyectan dosis de insulina a todos los pacientes durante un cierto periodo post-trasplante. Esta medida es de carácter preventivo, con lo que no indica que los pacientes vayan a ser insulino dependientes a posteriori. Dado que los 3 pacientes que se encuentran en esta situación no aportan ninguna información se han excluido del análisis bivariante para

esta clasificación. Para este conjunto muestral se han obtenido los siguientes resultados de forma resumida. En el anexo A se encuentran los resultados para todas las variables.

Variable	No insulina (N)	Sí insulina (N)	p Valor	H0
PESO	42	11	0,863	0
IMC	42	11	0,674	0
TAS	39	10	0,954	0
FG	2	8	0,654	0
CT	36	11	0,096	0
LDL	33	11	0,263	0
TG	36	11	0,633	0
PEPTIDO_C	42	11	0,189	0
UIDIA	42	11	0,947	0
ORTOSTATISMO	31	11	0,366	0
MANIOBRA_DE_VALSALVA	30	11	0,414	0
INDICE_3015	30	11	0,039	1
CVmáximaCMV	41	11	0,509	0
AÑO_DEL_TRASPLANTE	42	11	0,003	1
AÑOS_DEL_PACIENTE_AL_TRASPLANTE	42	11	0,727	0
FCODL	41	11	0,659	0
AÑO_DE_INICIO_DE_LA_DIALISIS	42	11	0,004	1
Edaddonante	40	11	0,221	0

CUADRO 5.9: Resumen resultados variables cuantitativas para el evento final 4

Variable	No Insulina		Insulina		χ^2	p Valor
	%	N	%	N		
DIAPASON_PIE_IZQUIERDO						
Patológico grave	34,375	11	54,545	6	8,364	0,039
Normal	34,375	11	18,182	2		
Prepatológico	6,25	2	27,273	3		
Patológico leve	25	8	-	-		
Profilaxistrombosistipo						
Antiagregación	91,429	32	87,5	7	5,072	0,079
Anticoagulación	8,571	3	-	-		
Doble	-	-	12,5	1		
Sexodonante						
Hombre	55,263	21	20	2	4,028	0,045
Mujer	44,737	17	80	8		
Negativo	38,462	15	20	2	-	-
ITB_DERECHO2						
Normal	54,167	13	37,5	3	0,097	0,953
Calcificación arterial	33,333	8	50	4		
Enfermedad arterial moderada	12,5	3	12,5	1		
RMN_VASCULAR						
AE leve	16,667	7	18,182	2	0,8	0,849
Normal	78,571	33	81,818	9		
Severa	2,381	1	-	-		
AE moderada	2,381	1	-	-		

CUADRO 5.10: Resumen resultados variables cualitativas para el evento final 4

Como se ha comentado anteriormente, hay dos variables de estudio para las que no se ha realizado el análisis multivariante para ninguno de los escenarios finales debido al formato de las mismas. Estas variables son FechaTXPR y AÑO_NACIMIENTO.

5.4. Análisis de los resultados obtenidos

Con el análisis multivariante se puede extraer información acerca de la relación de cada variable de estudio con el evento final observado en cada caso. No hay que olvidar que este análisis es puramente estadístico y que es tarea del usuario analizar e interpretar de forma exhaustiva los resultados obtenidos con el objetivo de determinar la validez de los mismos.

Como se ha descrito anteriormente, el análisis bivariante se ha centrado únicamente en el análisis de pacientes que han sido sometidos a trasplante SPK, dejando de lado a los pacientes sometidos a PAK. La razón de esta exclusión se debe a la escasa cantidad de pacientes de estudio para este tipo de trasplante. Por esta misma razón, en las tablas que reflejan los resultados del análisis multivariante para cada uno de los eventos finales (anexo A) a menudo aparecen ciertas variables con celdas vacías. Ya sea indicando el valor erróneo con un guión – o con un NaN. En primer caso se debe a que no hay datos recogidos de una cierta variable para el subconjunto poblacional del análisis, mientras que el segundo caso aparece en las tablas de las variables cualitativas y se debe a que la variable únicamente tienen datos para una categoría haciendo imposible la obtención de los parámetros estadísticos.

En el cuadro 5.11 se muestran las variables de interés obtenidas en el análisis multivariante para cada resultado final. Como se ha detallado anteriormente, una variable será considerada de interés cuando el valor del estadístico p sea menor a 0.05, lo cual indica que las diferencias entre las muestras de cada distribución (resultado A / resultado B) son significativas y no se deben al azar.

En cualquiera de los subgrupos del análisis aparecen significativas algunas variables como AÑO_DEL_TRASPLANTE o AÑO_DE_INICIO_DE_LA_DIALISIS. Sin embargo, estas variables no tienen mayor relevancia en el estudio que el mero significado informativo, ya que este tipo de variables no pueden ser utilizadas para la realización de un

Exitus/No Exitus	p valor
NEUROTESIOMETRO_PIE_DERECHO	0,008
AÑO_DEL_TRASPLANTE	0,031
AÑO_DE_INICIO_DE_LA_DIALISIS	0,017
Trasplantepreviorenal	0,004
SISTEMA_NERVIOSO_PARASIMPATICO	0,002
Explante/No explante Páncreas	
AÑO_DEL_TRASPLANTE	0,004
AÑO_DE_INICIO_DE_LA_DIALISIS	0,003
ALCOHOL	0,034
RMN_VASCULAR	0,019
SISTEMA_NERVIOSO_PARASIMPATICO	0,039
Profilaxistrombosis	0,00003
Profilaxistrombosistipo	0,000009
RETINOPATIA_DIABETICA_EN_OJO_DERECHO	0,001
ANTIHTA	0,036
MANIOBRA_DE_VALSALVA2	0,033
Pérdida/No Perdida Injerto	
INDICE_3015	0,012
NºCIG	0,041
AÑO_DEL_TRASPLANTE	0,006
AÑO_DE_INICIO_DE_LA_DIALISIS	0,027
SEXO	0,037
CATETERISMO	0,037
RMN_VASCULAR	0,033
RETINOPATIA_DIABETICA_EN_OJO_DERECHO	0,031
RETINOPATIA_DIABETICA_EN_OJO_IZQUIERDO	0,006
MANIOBRA_DE_VALSALVA2	0,03
INDICE_3015_2	0,002
Necesidad/No necesidad de insulina	
INDICE_3015	0,039
AÑO_DEL_TRASPLANTE	0,003
AÑO_DE_INICIO_DE_LA_DIALISIS	0,004
CATETERISMO	0,05
DIAPASON_EN_PIE_IZQUIERDO2	0,039

CUADRO 5.11: Variables significativas para cada evento final

modelo predictivo porque se centran en los datos de los pacientes del estudio y nunca se podrían extrapolar a un modelo de predicción general. En cuanto a la variable AÑO_DE_INICIO_DE_LA_DIALISIS, tendría más sentido la significancia de la variable AÑOS_EN_DIALISIS aunque en ningún caso se haya alcanzado el valor de p necesario para ello.

También hay ciertas variables que, a priori, no ofrecen ninguna justificación clínica para los resultados obtenidos. Variables como DIAPASON y RETINOPATIA_DIABETICA

son analizadas para ambos pies y ojos respectivamente como variables separadas. Podrían considerarse variables susceptibles de estudio si resultaran significativas ambas medidas, sin embargo no tienen explicación aparente que aparezcan únicamente RETINOPATIA_DIABETICA_EN_OJO_DERECHO y DIAPASON_EN_PIE_IZQUIERDO2.

Hay ciertas variables que, aunque no han alcanzado los valores necesarios de p para considerarse significativas, podrían ser interesantes. Como la variable AÑOS_DE_EVOLUCION_DE_LA_DM, que en el análisis multivariante cuyo evento final es la necesidad o no necesidad de insulina el estadístico p toma un valor cercano a 0,05 (cuadro A).

Por otro lado, se ha optado por el desarrollo del modelo de predicción en base a los datos del estudio del evento final número 3, que determina la clasificación de los pacientes según la pérdida o no de la funcionalidad del injerto pancreático trasplantado. Se ha tomado esta determinación ya que es el mayor conjunto de pacientes de los que se tienen datos además de que es el conjunto con las dos clases más balanceadas.

5.5. Preparación de los datos para el modelado

El estudio bivariante confirma niveles relativamente bajos de clasificación para las pocas variables que han demostrado tener una significación estadística y significancia clínica con respecto de la clase observada. Por tanto, el análisis bivariante no permite extraer conclusiones certeras sobre las variables a incluir para el modelo de clasificación, con lo que se ha abordado el problema con todas las variables disponibles.

Se dispone de muchas variables que no tienen el suficiente número de muestras para poder considerarlas aptas para el modelado. Por ello, se ha fijado un umbral mínimo de muestras que debe tener una variable para ser incluida en el modelado. En el caso de las variables cualitativas este umbral es del 50 % de muestras sobre el total de muestras posible y en el caso de las cuantitativas un 60 %. Esta distinción únicamente se debe a que se disponen de una cantidad de datos considerablemente menor de variables categóricas. Tomando como resultado clases la pérdida o no pérdida de funcionalidad del injerto pancreático, se dispone de un total de 56 pacientes que mantienen la funcionalidad y 14 que la pierden. Según el umbral mínimo establecido, se excluirán del modelado las variables cualitativas que no cuenten, al menos, con muestras para el 50 % de los

pacientes, que serán un total de 28 para la clase que mantiene la funcionalidad y un total de 7 para los que no la mantienen. En el caso de variables cuantitativas el umbral se sitúa en el 60%, que se corresponde con mínimo de 34 muestras para la clase de mantenimiento de la funcionalidad y un mínimo de 9 muestras para la clase de pérdida de la funcionalidad. Por tanto, se excluyen del modelado las variables que no disponen como mínimos de este número de muestras.

Para el entrenamiento del modelo se ha creado una matriz con todas las variables correspondientes al periodo basal (antes de la intervención quirúrgica) situadas en las columnas y las muestras disponibles situadas en las filas, concatenando para cada variable las 14 muestras de la clase correspondiente a mantenimiento de la funcionalidad con las 56 muestras de la clase pérdida de funcionalidad. En la primera columna se ha indicado la clase a la que pertenecen las variables, correspondiendo el 1 a la clase Pérdida de la funcionalidad del injerto pancreático y el 0 a Mantenimiento de la funcionalidad del injerto pancreático.

El toolbox Classification Learner obliga que todas las variables de entrada que se utilicen para la generación del modelo deben tener el mismo número de muestras, sin embargo, esto no sucede. De modo que, para las variables que no han sido excluidas del modelado por no tener un número de muestras mínimo pero no tienen el total de 56 muestras de la clase 0 y 14 muestras de la clase 1, se ha realizado una estimación del valor medio de las muestras de cada variable para sustituir los valores ausentes.

Por otro lado, el hecho de disponer de gran cantidad de variables conlleva la necesidad de recurrir a métodos de reducción de dimensionalidad, descritos en la sección 4.6.4, para reducir en la medida de lo posible el número de variables de entrada del modelo (predictores).

5.6. Evaluación de los modelos de predicción

Con ayuda del toolbox Classification Learner se han obtenido diferentes modelos de predicción con la intención de encontrar el modelo más adecuado a las necesidades del estudio, intentando mejorar en cada caso las prestaciones de los mismos mediante las técnicas descritas anteriormente.

Aprovechando las prestaciones de la herramienta utilizada, la forma de proceder ha sido cargar en el toolbox la matriz correspondiente, en cada caso, a las variables con las que se desea desarrollar el modelo y entrenar todos los modelos disponibles en dicha herramienta para determinar cuál o cuáles de ellos ofrecen una mayor exactitud en la clasificación.

Después de realizar un total de 4 iteraciones reduciendo la dimensionalidad de los datos de utilizados para el modelo, se ha obtenido un modelo basado en KNN con una precisión del 82,9%. En las siguientes secciones se detallan los resultados obtenidos en cada iteración.

5.6.1. Primera iteración

El número de variables basales disponibles, excluyendo las que no cumplen los requisitos de muestras mínimas, asciende a un total de 85. Del conjunto de 43 variables basales cuantitativas se han excluido 7, conservando 36, mientras que del conjunto de 52 variables basales cualitativas se han descartado 3, manteniendo 49. Estas variables se muestran en la siguiente tabla:

PERD.INJERTO	TIPOFCODLBASAL	HDL
SEXO	CMVReceptor	LDL
HTA	VEBReceptor	TG
DISLIPEMIA	CMVDonante	PEPTIDO_C
TABACO	VHCReceptor	UIDIA
ALCOHOL	VHBReceptor	UL.KG.DIA
TIPO.DE.DIALISIS	VIHReceptor	ORTOSTATISMO
RETINOPATIA.DIABETICA.EN.OJO.DERECHO	Trasplantepreviorenal	MANIOBRA.DE.VALSALVA
RETINOPATIA.DIABETICA.EN.OJO.IZQUIERDO	SeroconversionCMV	INDICE.3015
TEST.DE.ESFUERZO	TipodeafectacionCMV	NEUROTESIOMETRO.PIE.DERECHO
ECOCARDIOGRAFIA	ProfilaxisATB	TALLA
EKG	Profilaxiscotrimoxazol	N.CIG
CATETERISMO	ProfilaxisCMV	ANO.DIAG.DM
IAM	Profilaxisantifungica	ANOS.DE.EVOLUCION.DE.LA.DM
ACV	Profilaxistrombosis	ANOS.EN.DIALISIS
CLAUDICACION.INTERMITENTE	Profilaxistrombositipo	CVmaximaCMV
URETEROGRAFIA	Causafallemcimentodonante	ANO.DEL.TRASPLANTE
ORTOSTATISMO.ANTES.DEL.TRASPLANTE2	Sexodonante	ANOS.DEL.PACIENTE.AL.TRASPLANTE
SISTEMA.NERVIOSO.PARASIMPATICO	VHCDonante	N.PINCHAZO
MANIOBRA.DE.VALSALVA.ANTES.DEL.TRASPLANTE2	VHBDonante	NEUROTESIOMETRO.PIE.IZQUIERDO
INDICE.3015.ANTES.DEL.TRASPLANTE2	VIHDonante	FCODL
NEUROTESIOMETRO.PIE.DERECHO.ANTES2	PESO	ANO.DE.INICIO.DE.LA.DIALISIS
NEUROTESIOMETRO.PIE.IZQUIERDO.ANTES2	IMC	INCOMPATIBILIDADESHLA
ITB.DERECHO.ANTES2	TAS	DIAPASON.EN.PIE.DERECHO
ITB.IZQUIERDO.ANTES2	TAD	DIAPASON.EN.PIE.IZQUIERDO
DIAPASON.EN.PIE.DERECHO.ANTES2	GLUCEMIA	ANTIHTA
DIAPASON.EN.PIE.IZQUIERDO.ANTES2	HBA1C	CREATININA
RMN.VASCULAR	UREA	Edaddonante
EMG	CT	

CUADRO 5.12: Conjunto de variables inicial utilizadas para el modelado

En primer lugar, se ha utilizado este conjunto completo de variables para el entrenamiento de los modelos, obteniéndose un nivel máximo de precisión del 80%.

5.6.2. Segunda iteración

Dado que es insostenible y nada práctico que un modelo de predicción tenga un número de variables de entrada tan elevado se han aplicado dos técnicas de reducción de dimensionalidad.

La técnica PCA se incluye de forma intrínseca en el toolbox de Matlab, de manera que es suficiente con habilitar dicha opción una vez que se ha cargado la matriz completa.

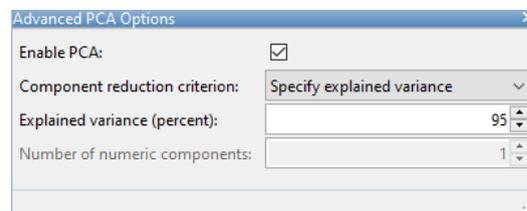


FIGURA 5.18: Función PCA del toolbox Classification Learner

Con esta opción habilitada, se han entrenado nuevamente los modelos disponibles. En la información de los modelos detallada en la figura 5.19, se puede observar que con la herramienta PCA se ha obtenido una matriz de componentes, cuya primera componente principal es suficiente para determinar más del 95 % de la varianza del conjunto de datos. Con este entrenamiento se ha obtenido nuevamente una precisión máxima del 80 %.

PCA

PCA is keeping enough components to explain 95% variance.

After training, 1 components were kept.

Explained variance per component (in order): 99.9%, 0.1%, 0.0%, 0.0%, 0.0%, 0.0%, 0.0%, 0.0%, 0.0%, 0.0% (variances of least important components hidden)

FIGURA 5.19: Resultados PCA

Esta técnica evita las variables categóricas y teniendo en cuenta que el conjunto de datos del estudio tienen más de la mitad de variables de ese tipo, se ha procedido a la utilización de otra técnica de reducción de dimensionalidad que puede obtener mejores resultados.

Esta técnica de reducción de dimensionalidad es el método de extracción de características empleando distancias euclídeas y componentes de vecindarios (*Neighborhood Component Analysis*, NCA). Se ha representado gráficamente en la figura 5.20 la influencia de las muestras de cada variable al resultado de la clasificación. Los números del eje x corresponden a cada variable, mientras que en el eje y se representa numéricamente la influencia o *peso*.

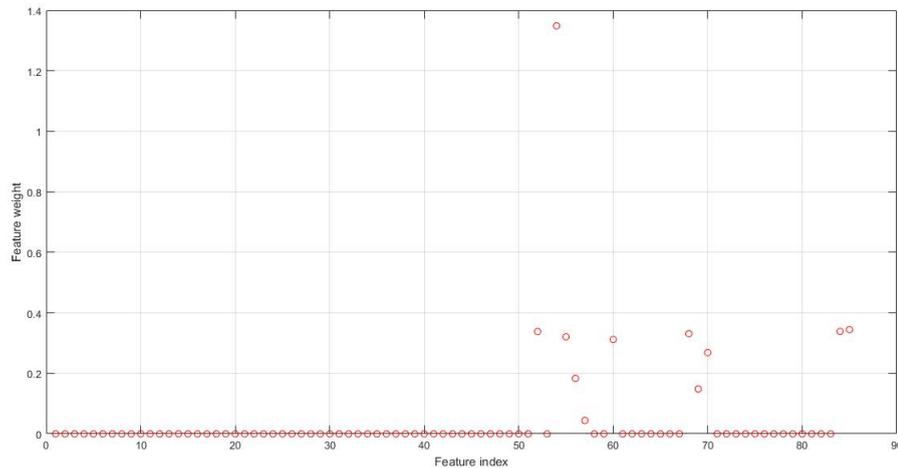


FIGURA 5.20: Pesos de cada variable a la clasificación

Según esta gráfica, las variables que mayor influencia tienen sobre el resultado final son las siguientes:

Variable	Peso
TAS	0.338
GLUCEMIA	1.349
HBA1C	0.321
UREA	0.183
CT	0.044
TG	0.312
TALLA	0.331
N_CIG	0.148
AÑO_DIAG_DM	0.268
CREATININA	0.338
Edaddonante	0.344

CUADRO 5.13: Variables con influencia en el resultado final

Este resultado muestra que la variable *AÑO_DIAG_DM* tiene cierta influencia sobre la clasificación, sin embargo, como se ha explicado anteriormente no tiene sentido incluirla en un modelo destinado a la predicción. Por tanto, se han entrenado los modelos de clasificación únicamente con las 10 variables que han mostrado un peso mayor que 0 en el resultado final (excluyendo la variable *AÑO_DIAG_DM*). La mejor precisión se ha obtenido para el modelo Ensemble Classifier del tipo Subespacio KNN, alcanzando un 81.4%.

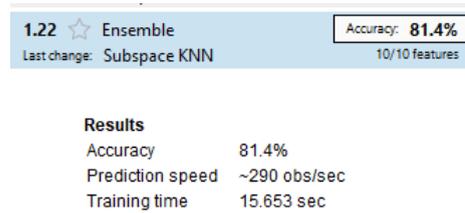


FIGURA 5.21: Resultados del modelo aplicando NCA

5.6.3. Tercera iteración

Para el desarrollo de del modelo anterior, se ha aplicado la técnica de reducción de dimensionalidad sobre el total de las variables que se muestran en la tabla 5.12.

Sin embargo, se ha decidido orientar el modelo a la predicción en base a variables procedentes únicamente del propio paciente trasplantado para que el modelo pueda ser utilizado sin demandar información externa al paciente.

Por ello, se han excluido del modelado las variables relativas a los donantes de los injertos pancreático que han sido trasplantados en los pacientes. En concreto estas variables son:

Variable
Edaddonante
Sexodonante
Causafallecimientodonante
VHCDonante
VHBDonante
VIHDonante
CMVDonante

CUADRO 5.14: Variables del donante excluidas del modelado

Eliminando estas variables del conjunto, se ha obtenido el siguiente gráfico de pesos (figura 5.22):

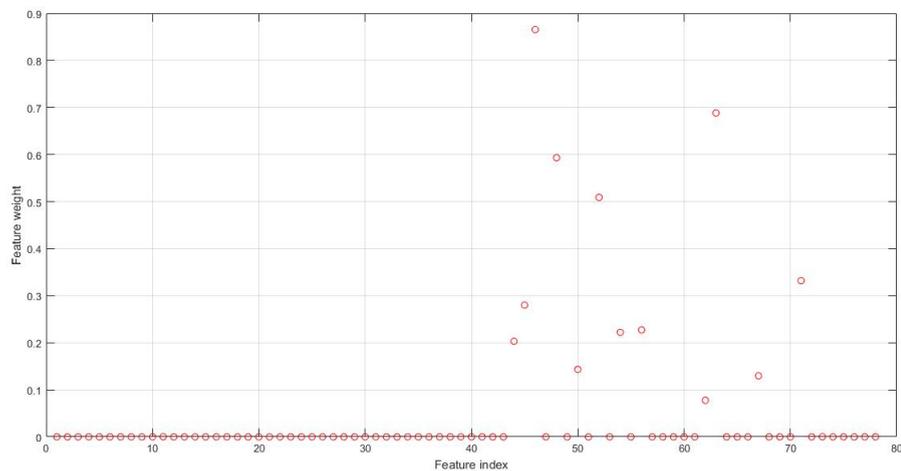


FIGURA 5.22: Variables con el influencia en el resultado final excluyendo las variables del donante

Estas variables significativas, junto con sus pesos se detallan en la tabla 5.15 :

Variable	Peso
PESO	0.203
IMC	0.280
TAS	0.866
GLUCEMIA	0.593
UREA	0.143
HDL	0.508
TG	0.222
UIDIA	0.227
TALLA	0.078
N_CIG	0.688
CVmáximaCMV	0.13
NEUROTENSIOMETRO_PIE_IZQUIERDO	0.332

CUADRO 5.15: Variables influyentes con sus pesos

Analizando críticamente este nuevo conjunto de variables, se ha decidido prescindir de algunas de ellas en el entrenamiento de los modelos. En concreto, se ha eliminado la variable *TALLA* porque es la que tienen un peso menor en el resultado de la clasificación además de que clínicamente no hay ningún fundamento para poder asociar la altura de un paciente con el resultado de un trasplante; la variable *NEUROTENSIOMETRO_PIE_IZQUIERDO* porque, como ha explicado en la sección 5.4, no tiene demasiado sentido que el pie izquierdo sea un predictor pero el derecho no lo sea; la variable *CVMaximaCMV* debido a que tiene un peso muy pequeño sobre el resultado además de que tiene muy pocos valores distintos de 0 y éstos son muy elevados y dispares, lo que

puede introducir mucho ruido en la clasificación; y la variable *PESO*, ya que la prueba realizada a las variables *PESO* e *IMC* para comprobar si existe una relación lineal entre ellas ha determinado un coeficiente de correlación de 0.836, muy próximo a 1 (un valor de 1 indica dependencia total entre variables). En este último caso se ha prescindido de la variable *PESO* porque, de las dos, es la que menor influencia tiene sobre el resultado final. Finalmente, se han entrado los modelos de clasificación con las siguientes variables:

Variable
IMC
TAS
GLUCEMIA
UREA
HDL
TG
UIDIA
N.CIG

CUADRO 5.16: Variables utilizadas como predictoras en el modelo de clasificación

Se ha obtenido una precisión máxima del 80,0% en varios modelos, entre los que se encuentran casi todos los algoritmos disponibles para el modelo KNN, el modelo SVM en las variantes que utiliza el método Kernel y varias variantes del modelo que se basa en el Conjunto de clasificadores.

```

Model 1.17: Trained

Results
Accuracy          80.0%
Prediction speed  ~880 obs/sec
Training time     11.731 sec

Model Type
Preset: Cubic KNN
Number of neighbors: 10
Distance metric: Minkowski (cubic)
Distance weight: Equal
Standardize data: true

```

FIGURA 5.23: Ejemplo de uno de los modelos con una precisión del 80,0%

5.6.4. Cuarta iteración

Utilizado el mismo conjunto de variables citadas anteriormente para el desarrollo del modelo, se ha modificado el método entrenamiento y prueba del modelo. Hasta el momento, se ha utilizado el método de validación Cross-Validation de 5 particiones. Sin embargo, en este caso se ha modificado el número de particiones a 4. De todos los modelos entrenados, se ha alcanzado un 82,9% de precisión de clasificación para el modelo

Ensemble Classifier del tipo Subespacio KNN. Este modelo utiliza un número de 5 variables elegidas aleatoriamente para la construcción de cada clasificador, de un total de 30. Cada clasificado es un modelo KNN que utiliza distancias euclídeas para determinar la distancia de los vecinos más cercanos.

Modelo	Algoritmo	Precisión
Árbol de decisión	Complex Tree	71.4 %
	Medium Tree	71.4 %
	Simple Tree	77.1 %
Análisis discriminante	Linear discriminant	72.9 %
	Quadratic Discriminant	72.9 %
Regresión logística	Logistic Regression	71.4 %
SVM	Linear SVM	80.0 %
	Quadratic SVM	68.6 %
	Cubic SVM	71.4 %
	Fine Gaussian SVM	80.0 %
	Medium Gaussian SVM	80.0 %
	Coarse Gaussian SVM	80.0 %
KNN	Fine KNN	74.3 %
	Medium KNN	80.0 %
	Coarse KNN	80.0 %
	Cosine KNN	80.0 %
	Cubic KNN	80.0 %
	Weighted KNN	80.0 %
Ensemble	Boosted Trees	80.0 %
	Bagged Trees	80.0 %
	Subspace Discriminant	81.4 %
	Subspace KNN	82.9 %
	RUSBoosted Trees	54.3 %

CUADRO 5.17: Resultados de precisión para cada modelo entrenado

En la imagen 5.24 se muestran los resultados obtenidos con este modelo de clasificación respecto del eje x, que se corresponde con la variable UREA, y el eje y, que se corresponde con la variable TG. Los puntos rojos son los datos que se han clasificado como Pérdida de la funcionalidad del páncreas, mientras que los puntos en azul son los datos que se han clasificado como Mantenimiento de la funcionalidad del páncreas. Los datos se representan como puntos o cruces en función de si el modelo ha realizado una clasificación correcta respecto a la clasificación real del dato o no, respectivamente.

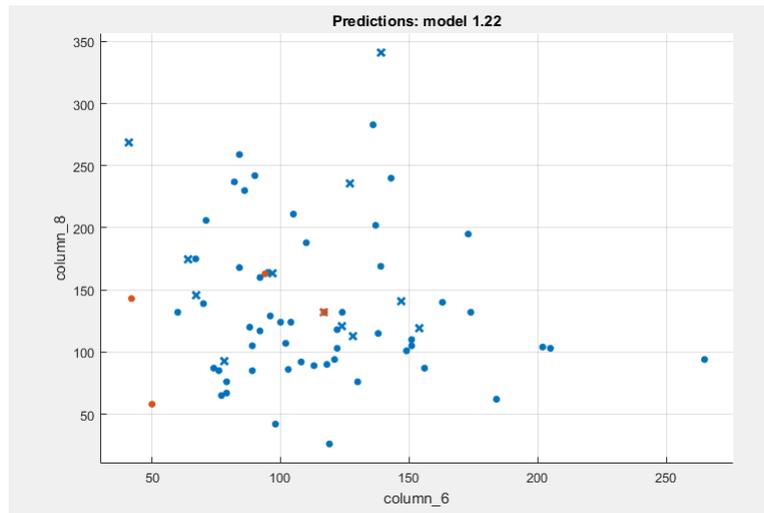


FIGURA 5.24: Resultados del modelo de clasificación para las variables UREA y TG

Con la matriz de confusión se puede ver de forma clara los resultados de clasificación obtenidos por el modelo.



FIGURA 5.25: (a) Matriz de confusión con el número de observaciones ;
(b) Matriz de confusión con ratios

En la imagen 5.25 (a) se puede observar el resultado obtenido con el modelo de clasificación (columnas) en función de la clasificación real de las muestras (filas). En la imagen 5.25 (b) se tienen estos mismos resultados pero en función del porcentaje de muestras reales que pertenecen a cada clasificación. En la columna de la parte derecha que aparece nombrada como *True Positive Rate* aparecen los porcentajes de muestras que se han clasificado correctamente tanto de una clase como de la otra respecto al número de muestras que pertenecen realmente a esa clase. Mientras que en la columna nombrada como *False Negative Rate* se observan el porcentaje de muestras que se han clasificado erróneamente. Este modelo consigue clasificar correctamente el 96 % de las muestras de

la clase 0 (Mantenimiento de la funcionalidad del páncreas) y el 21 % de las muestras de la clase 1 (Pérdida de funcionalidad del páncreas). En contraste, el 79 % de las muestras de la clase 1 y el 2 % de la clase 0 han sido clasificadas erróneamente.

Es decir, el 96 % se corresponde con la métrica de sensibilidad, que se identifica como el porcentaje de pacientes que el modelo clasifica como Mantenimiento de funcionalidad del páncreas sobre el número de pacientes que realmente les funciona correctamente el injerto. Sin embargo, la métrica de especificidad indica que solo el 21 % de pacientes que realmente han perdido la funcionalidad del injerto han sido clasificados por el modelo en esta clase.

Con el valor de falsos positivos frente al valor de verdaderos positivos que se obtienen a partir de la matriz de confusión se representa la gráfica de la curva ROC (figura 5.26).

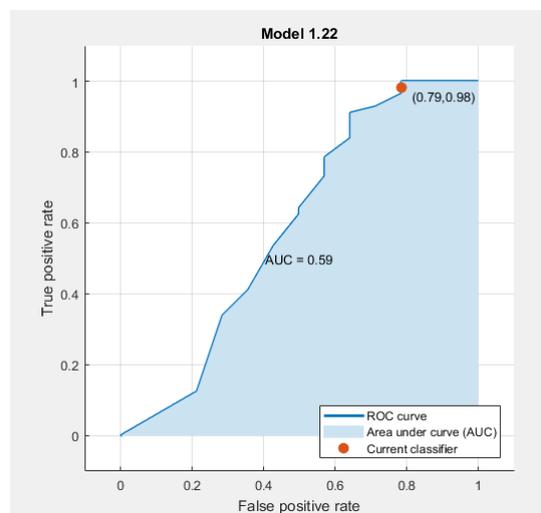


FIGURA 5.26: Curva ROC y valor del Área Bajo la Curva (AUC)

En esta figura se observa un punto naranja que indica el porcentaje de falsos positivos (79 %) y verdaderos positivos (98 %) para el modelo resultante, siendo los mismos valores que se indican en la matriz de confusión. Mientras que la métrica central indica un área bajo la curva de 0.59.

5.7. Implementación del sistema de ayuda a la decisión clínica

Según el esquema funcional del sistema de ayuda a la decisión clínica descrito en la sección 4.8, se ha adaptado el sistema de orquestación para satisfacer las necesidades del proyecto. En la imagen 5.27 se muestra la estructura de la solución adoptadas para el sistema orquestado.

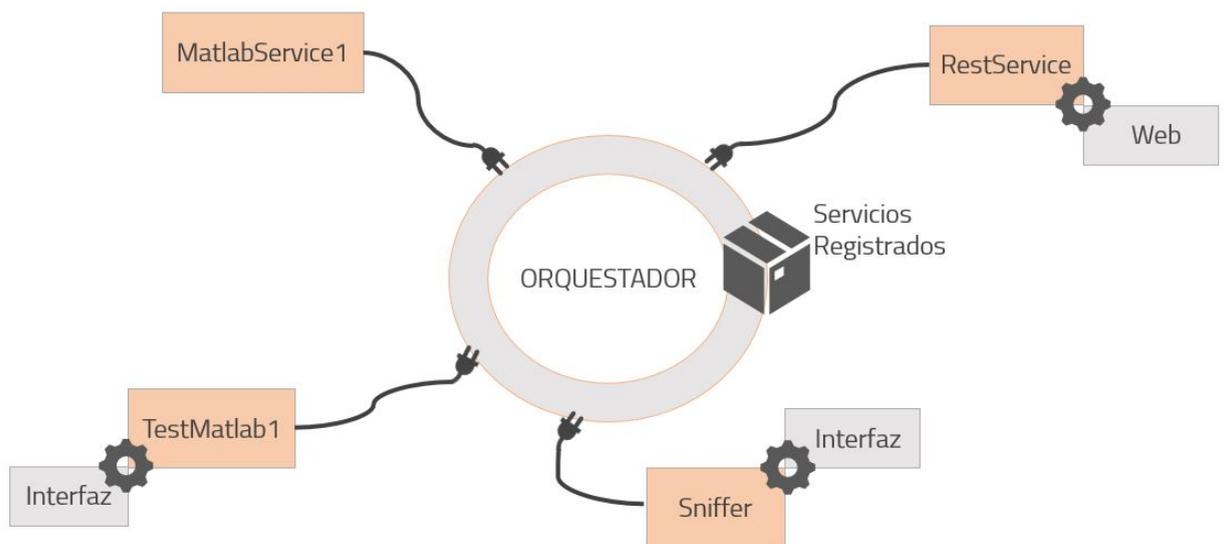
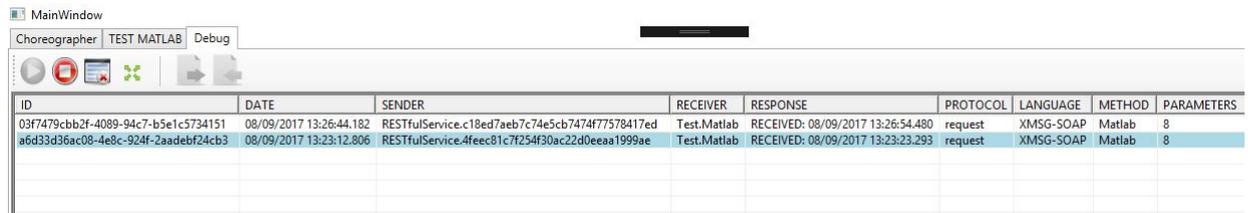


FIGURA 5.27: Esquema del sistema orquestado

Módulo de control y predicción: Está formado por el núcleo orquestador, que contiene un registro con todos los servicios disponibles en el sistema, y diferentes servicios que se comunican entre sí. Cada uno de ellos se detalla a continuación:

- Servicio MatlabService1: Este servicio ejecuta el modelo de predicción elegido anteriormente (Ensemble Classifier del tipo Subespacio KNN) mediante una llamada al motor de Matlab para obtener el resultado de mantenimiento o No mantenimiento de la funcionalidad del páncreas. Este servicio demanda una variables de entrada que se corresponde con los datos del paciente del que se desea conocer la predicción. Estos datos son introducidos por el usuario a través del interfaz web.
- Servicio Sniffer: Este servicio se utiliza por defecto para visualizar, mediante un interfaz, el intercambio de mensajes que realiza el orquestador.



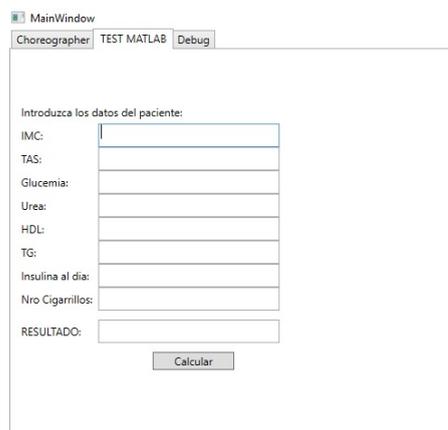
ID	DATE	SENDER	RECEIVER	RESPONSE	PROTOCOL	LANGUAGE	METHOD	PARAMETERS
03f7479-cbb2f-4089-94c7-b5e1c5734151	08/09/2017 13:26:44.182	RESTfulService.c18ed7aeb7c74e5cb7474f7578417ed	Test.Matlab	RECEIVED: 08/09/2017 13:26:54.480	request	XMSG-SOAP	Matlab	8
a6d53d36ac08-4e8c-924f-2aadebf24cb3	08/09/2017 13:23:12.806	RESTfulService.4feec81c77254f30ac22d0eeaa1999ae	Test.Matlab	RECEIVED: 08/09/2017 13:23:23.293	request	XMSG-SOAP	Matlab	8

FIGURA 5.28: Interfaz del servicio Sniffer

En la imagen 5.28 se puede la información que el servicio Sniffer proporciona sobre el intercambio de información que se produce en el sistema. En este caso, se han capturado dos comunicaciones que se corresponden con dos veces en los que el usuario ha demandado los resultados de predicción para un paciente. El servicio proporciona información del identificador de la comunicación, la fecha y hora, el servicio que envía el mensaje, el servicio que lo recibe, la fecha y hora en la que se recibe la respuesta, el protocolo, el lenguaje del mensaje, el método al que llama dentro del servicio receptor y en número de parámetros que se envía.

En este caso, el servicio que envía ambos mensajes es el servicio web (*RestService*) cuyo identificador es *RESTfulService* que está enviando información al servicio que ejecuta el modelo de predicción (*MatlabService1*) cuyo identificador es *Test.Matlab*. Además, el servicio Sniffer también permite visualizar los mensajes que se envían en cada comunicación.

- Servicio TestMatlab1: Este servicio se ha añadido para la realización de pruebas de comunicación con modelo de forma local sin necesidad del uso del interfaz web. El intercambio de información se produce directamente con el servicio *MatlabService1*.



MainWindow
Choreographer TEST MATLAB Debug

Introduzca los datos del paciente:

IMC:

TAS:

Glucemia:

Urea:

HDL:

TG:

Insulina al día:

Nro Cigarrillos:

RESULTADO:

Calcular

FIGURA 5.29: Interfaz del servicio TestMatlab1

Módulo de comunicación con el usuario: Sirve de nexo de unión del sistema con el usuario final. Actualmente, este módulo se encuentra en la misma máquina que el resto del sistema orquestador, sin embargo, está preparado para ser instalado en cualquier otra máquina y comunicarse con el servidor donde se encuentra el orquestador. Consta de un interfaz web donde el usuario introduce los datos del paciente y recibe el resultado de la predicción, figura 5.30. El servicio RestService es el encargado de encapsular en un mensaje XMSG estos datos y enviarlos al servicio que ejecuta el modelo de predicción para posteriormente, recibir el resultado y mostrarlo en la web.

Pilar Moreno - TFM

En esta web se ejecuta un modelo de predicción que determina la viabilidad de un trasplante de páncreas en base a la pérdida de funcionalidad del injerto, con un 82,9% de precisión.

Introduzca los datos del paciente:

Tensión arterial sistólica (mmHg)	10
Niveles de Glucosa (mg/dl)	10
Nivel de colesterol HDL (mg/dl)	10
Triglicéridos (mg/dl)	10
Urea en sangre (mg/dl)	20
Número de cigarrillos al día	0
Unidades de insulina al día	0
IMC (kg/m ²)	10.00

LIMPIAR EJECUTAR MODELO

FIGURA 5.30: Interfaz web

La web consta de un encabezado que conduce a las opciones de configuración. Estas opciones permiten modificar la dirección de la máquina donde se encuentra el orquestador, figura 5.31 (a), que en este caso indican la dirección de localhost porque todo el sistema se encuentra en la misma máquina. También se puede modificar el servicio que se utiliza para enviar los datos así como el servicio que los va a recibir y el método dentro de este servicio, figura 5.31 (b).



FIGURA 5.31: (a) Configuración de la dirección del orquestador;
(b) Configuración de los servicios



FIGURA 5.32: Configuración actual

Más abajo, la web permite introducir los datos de los pacientes mediante unos sliders. Una vez que se ha completado los datos del paciente y se ha ejecutado el modelo, la web muestra una caja con el resultado recibido:

El modelo de predicción determina que el paciente **SI** va a mantener la funcionalidad del páncreas.
Sí se recomienda la realización del trasplante.

FIGURA 5.33: Resultado positivo del modelo de predicción

El modelo de predicción determina que el paciente **NO** va a mantener la funcionalidad del páncreas.
No se recomienda la realización del trasplante.

FIGURA 5.34: Resultado negativo del modelo de predicción

Las siguientes imágenes se corresponden con un mensaje de las comunicaciones capturadas por el servicio Sniffer, ver imagen 5.28, en la que el servicio web (RestService) cuyo

identificador es *RESTfulService* envía un mensaje al servicio que ejecuta el modelo de predicción que contienen los datos que ha insertado el usuario en la web.

```
REQUEST: 08/09/2017 13:23:12.806
{
  "IdMensaje": "a6d33d36ac08-4e8c-924f2aadebf24cb3",
  "sender": "RESTfulService.4feec81c7f254f30ac22d0eeaa1999ae",
  "receiver": "Test.Matlab",
  "protocolo": 0,
  "MetaData": null,
  "Credential": null,
  "lenguaje": "XMSG-SOAP",
  "enrespuesta": ""
}
```

FIGURA 5.35: Cabecera del mensaje XMSG

```
"Contenido": {
  "metodo": "Matlab",
  "parametros": {
    "input02": {
      "type": "xs:double",
      "value": "1662"
    },
    "input03": {
      "type": "xs:integer",
      "value": "16"
    },
    "input04": {
      "type": "xs:integer",
      "value": "26"
    },
    "input05": {
      "type": "xs:integer",
      "value": "101"
    },
    "input06": {
      "type": "xs:integer",
      "value": "59"
    },
    "input07": {
      "type": "xs:integer",
      "value": "168"
    },
    "input08": {
      "type": "xs:integer",
      "value": "69"
    },
    "input09": {
      "type": "xs:integer",
      "value": "14"
    }
  }
}
```

FIGURA 5.36: Contenido del mensaje XMSG del remitente

Cada uno de los inputs se corresponde con las variables predictores de entrada en el modelo:

Variable	Parámetro
IMC	input02
TAS	input03
GLUCEMIA	input04
UREA	input05
HDL	input06
TG	input07
UIDIA	input08
N_CIG	input09

CUADRO 5.18: Parámetros del contenido del mensaje XMSG

A continuación, el servicio MatlabService1 responde al mensaje anterior con el resultado de la ejecución del modelo, que será un "value":1 en el caso de que el modelo determine que el paciente no va a mantener la funcionalidad del páncreas trasplantado, o "value":0 en el caso de que determine que sí la mantendrá.

```
RESPONSE: 08/09/2017 13:23:23.293
{
  "IdMensaje": "8ac3f369ff7b42e3ae93bedc6f211ece",
  "sender": "Test.Matlab",
  "receiver": "RESTfulService.4feec81c7f254f30ac22d0eaa1999ae",
  "protocolo": 1,
  "MetaData": null,
  "Credential": null,
  "lenguaje": "XMSG-SOAP",
  "enrespuesta": "a6d33d36ac08-4e8c-924f-2aadebf24cb3",
  "Contenido": {
    "metodo": "MatlabResponse",
    "parametros": {
      "MatlabResult": {
        "type": "xs:integer",
        "value": "1"
      }
    }
  }
}
```

FIGURA 5.37: Contenido del mensaje XMSG del destinatario

Capítulo 6

Discusión

La Diabetes Mellitus es un gran problema en la actualidad, ya que el número de personas que se ven afectadas por esta enfermedad aumenta sin control año tras año. A pesar de que el porcentaje de incidencias de la diabetes tipo 1 es mucho menor que la de tipo 2, la diabetes tipo 1 repercute gravemente en el estilo de vida de los pacientes, además de las repercusiones en la salud, ya que va a depender de por vida de una administración diaria de insulina. Por eso, no es de extrañar que se baraje como solución alternativa para el paciente la sustitución del órgano que está produciendo de forma deficiente la hormona insulina: el páncreas. Este tipo de trasplantes de páncreas se realiza de forma conjunta con el de riñón, resultando en dos modalidades diferentes, el trasplante simultáneo de páncreas y riñón (SPK), y el trasplante de páncreas a pacientes que ya se les ha trasplantado un riñón (PAK). Sin embargo, el éxito de este tipo de trasplantes no es demasiado elevado y no se conocen las causas que influyen en este hecho.

En este proyecto, se ha realizado un análisis estadístico bivariante con los datos disponibles de los pacientes a los que se les ha practicado un SPK, para determinar la influencia de las variables sobre diferentes resultados finales del paciente entre los que se encuentran la supervivencia del paciente a largo plazo, el explante del páncreas trasplantado, la pérdida de funcionalidad del páncreas y la necesidad de ingesta de insulina después del trasplante. Sin embargo, no se han obtenido resultados concluyentes sobre ninguno de estos análisis. No ha sido posible incluir en el análisis los datos de los pacientes con trasplante PAK debido a que únicamente se han realizado 4 intervenciones de este tipo. La escasez de variables influyentes en los resultados finales puede ser debido a que el

número de muestras disponibles para cada variable, 70 pacientes SPK, es muy pequeño para reflejar resultados concluyentes. Además, el número de muestras de cada clase es muy desigual para todos los resultados finales.

No se ha encontrado un conjunto de variables significativas, por lo que ha sido necesario abordar los modelos de clasificación con el total de variables disponibles, realizado otro tipo de técnicas para reducción de la dimensionalidad. Debido a que se ha encontrado una cantidad muy elevada de variables con muestras perdidas ha sido necesario descartar las variables que no disponían de un mínimo del 50% de las muestras, mientras que se han estimado los valores perdidos de las variables mediante asignándoles la moda/media de las muestras en cada variable. Esto puede suponer una influencia muy grande en el modelo de predicción porque el número de muestras que se han tenido que estimar porque estaban perdidas es muy elevado.

Después de varias iteraciones, se ha encontrado un modelo con una exactitud aceptable del 82,9%. Este modelo tiene una sensibilidad muy alta (98,0%), esto refleja que clasifica casi a la perfección a los pacientes que verdaderamente mantienen el páncreas, es decir, no determina casi en ningún caso que un paciente va a perder el páncreas siendo realmente que no lo perderá. En otras palabras, no excluye erróneamente a los posibles pacientes de un trasplante.

En contraste, este modelo tiene una especificidad muy baja (21,0%). Del total de pacientes que verdaderamente han perdido el páncreas ha clasificado como pérdida a muy pocos. Esto se puede traducir en que el modelo determina como aptos para el trasplante a gran cantidad de pacientes a menos que sea muy evidente que el paciente va a perder el páncreas.

Sin embargo, pese al valor considerablemente alto de precisión alcanzada por el modelo, los valores de área bajo la curva ROC (AUC) obtenidos (0.59) quedan muy lejos del valor que se atribuyen a un buen modelo de predicción (entorno a 0.80). Esto se debe, a que en la precisión influye el valor de los datos positivos clasificados correctamente (verdaderos positivos), donde el modelo consigue una clasificación casi perfecta debido a que la muestra es, fundamentalmente, de datos que pertenecen a esta clase. Sin embargo, el poder de clasificación para los datos de la clase negativa es más deficiente porque, debido al reducido número de datos para esta clase, no se ha podido establecer unos predictores fiables. Por consiguiente, el valor de AUC obtenido, donde se refleja la relación entre

los valores de la clase positiva que se clasifican correctamente y los valores de la clase negativa que se han confundido con valores de la clase positiva, es un resultado muy pobre.

Todo esto refleja que quizá el problema es demasiado complejo para estos modelos y surge la necesidad de utilizar otro tipo de modelos más complejos como las redes neuronales o el perceptrón multicapa, resultando necesaria la recolección de más datos de pacientes a los que se les ha realizado esta intervención en otros hospitales.

En cuanto al sistema de orquestación desarrollado, hay que resaltar que

Por otro lado, hay que resaltar que el sistema distribuido basado en orquestación de servicios que se ha desarrollado permite modificar o sustituir de forma muy sencilla el modelo predictivo, además de permitir el acceso descentralizado a profesionales de la salud. Lo que facilita la escalabilidad y la ampliación del sistema.

Capítulo 7

Conclusión

7.1. Revisión de los objetivos

En este capítulo se ofrece una revisión de los objetivos planteados al inicio de este proyecto.

Respecto al objetivo principal, se ha conseguido implementar un sistema distribuido destinado a la predicción del éxito de un trasplante de páncreas y riñón en pacientes con DM tipo 1 basado en un modelo de clasificación con una precisión del 82,9% al que se accede desde una página web.

En cuanto a los objetivos secundarios:

- Se ha creado un modelo estructurado de datos que permite a un usuario comprender fácilmente el conjunto de datos disponibles en el estudio, así como acceder a ellos fácilmente. Sin embargo, para el análisis de esos datos con el software Matlab, en ciertas ocasiones se ha comprobado que es más útil trabajar directamente con el origen inicial de datos, ya que dispone de la codificación numérica de las variables cualitativas.
- Se ha analizado de forma descriptiva el conjunto de variables. Esto ha permitido conocer más a fondo los datos del estudio así como detectar y corregir errores en las muestras.

- El análisis bivalente de los datos se ha realizado para diferentes escenarios, sin embargo, no ha arrojado conclusiones relevantes sobre las variables que influyen directamente sobre la clasificación de los pacientes en estos escenarios.
- Se ha evaluado diferentes modelos de clasificación, buscando mediante diferentes técnicas el modelo con mejores prestaciones. Finalmente, se ha conseguido un modelo que permite predecir el éxito del trasplante con un 82,9% de precisión de acierto. Sin embargo, la limitación en cuanto a datos realmente predictivos, ha dificultado la generalización del modelo, siendo necesaria una validación exhaustiva con datos externos.
- Se ha implementado un sistema de comunicación distribuido donde el usuario puede acceder a los resultados del modelo predictivo mediante un interfaz sencillo. No obstante, no ha incluido en este sistema el Módulo de almacenamiento de datos propuesto inicialmente para almacenar los datos de las predicciones. Hasta el momento, la base de datos se ha creado únicamente para almacenar el conjunto de datos de estudio, siendo independiente del sistema de orquestación.

7.2. Líneas futuras

En este proyecto se ha hecho latente la necesidad de disponer de un conjunto de datos adecuados para poder obtener un modelo de predicción que aporte precisión y generalización. Con lo que, como trabajo futuro, se plantea la revisión de los datos disponibles, así como la recolección de más muestras de pacientes con las mismas características para aportar más generalización y poder predictivo al modelo.

También se plantea el análisis de modelos de clasificación más acordes a problemas de alta complejidad, como redes neuronales y perceptrón multicapa.

Por otro lado, se propone la modificación de la estructura de la base de datos para insertar nuevas tablas donde se almacenen los datos de las predicciones realizadas y así, añadirla en el sistema orquestado. Para que este sistema resulte práctico en un contexto real, será necesario alojarlo en un servidor para permitir que el usuario pueda acceder a él sin necesidad de instalar el sistema en su propia máquina.

Apéndice A

Análisis exploratorio de datos

En este apéndice se muestran los resultados del análisis exploratorio de datos para el conjunto total de variables del estudio.

Variable	Media	Desviación	N
PESO	66,599	12,735	82
IMC	24,045	3,391	82
TAS	141,173	23,693	75
TAD	77,224	11,851	76
GLUCEMIA	186,901	83,026	81
HBA1C	8,191	1,596	79
UREA	110,382	41,331	76
FG	16,557	13,183	14
CT	159,205	41,272	73
HDL	47,957	15,949	70
LDL	84,943	30,235	70
TG	134,919	64,424	74
PEPTIDO_C	0,232	0,204	82
UIDIA	37,622	19,619	82
UI_KG_DIA	0,565	0,225	82
DOSIS_ESTATINA	14,968	19,465	63
ORTOSTATISMO	-21,897	16,427	58
MANIOBRA_DE_VALSALVA	1,104	0,097	58
INDICE_3015	1,049	0,086	58
DEEPBREATHING	1,099	0,138	48
NEUROTESIOMETRO_PIE_DERECHO	20,247	9,63	60
ITB_DERECHO	1,113	0,248	37
ITB_IZQUIERDO	1,11	0,245	35
TALLA	165,999	8,937	82

Sigue en la página siguiente.

Variable	Media	Desviación	N
Nº CIG	6,704	10,128	81
AÑO DIAG DM	1983,573	6,821	82
AÑOS DE EVOLUCION DE LA DM	25,524	6,508	82
AÑOS EN DIALISIS	1,914	1,075	81
CV máxima CMV	813,877	3337,863	81
AÑO DEL TRASPLANTE	2009,354	3,557	82
AÑOS DEL PACIENTE AL TRASPLANTE	37,366	5,717	82
Nº PINCHAZO	3,494	0,937	81
NEUROTESIOMETRO PIE IZQUIERDO	21,197	10,35	59
FCODL	0,679	0,496	81
AÑO DE INICIO DE LA DIALISIS	2007,325	3,463	80
INCOMPATIBILIDADESHLA	4,73	1,096	63
DIAPASON EN PIE DERECHO	3,891	3,009	64
DIAPASON EN PIE IZQUIERDO	3,969	2,889	64
ANTIHTA	2,423	1,363	78
CREATININA	7,064	2,759	81
AC_GAD	161,02	495,317	21
AC_IA2	26,701	59,773	20
PAK TIEMPO TRAS TX RENAL	59,5	62,089	4
Edad donante	31,077	9,379	78
SOG_GLUC	83	0	1
Albúmina	3,721	0,602	66
Proteinuria	0,283	0,441	59
INSULINA	18,68	9,073	5
Nº ADO	0	0	68
Dosistacrolimus	6,446	1,522	56
Nivelestacrolimus	10,665	6,324	51
DosisMMF	1991,525	65,094	59
DosisCT	500	0	77
TIEMPO EVOL DESDE EL TRASPLANTE	5,646	3,557	82
MESES EXITUS TRAS TX	6,5	13,005	8
MESES FISTULA TRAS TX	0	0	3
MESES EXPLANTE PANCRAS TRAS TX	2,1	4,254	10
MESES EXPLANTE RENAL TRAS TX	2,5	4,359	4
Cambiorégimen Tacrolimus M	22,393	20,619	28
Retirada CTM	25,88	22,48	50

CUADRO A.1: Resultados del análisis descriptivo de las variables cuantitativas

Variable	%	N
Sexo		
Hombre	58,537	48
Mujer	41,463	34
EXITUS		
No	90,244	74
Si	9,756	8
PERD_INJERTO		
Si	31,707	26
No Ex-fumador	19,512	16
TIPO_TRASPL		
SPR	98,78	81
PAK	1,22	1
NEC_INSULINA_TRAS_TX		
No	65,672	44
Si	34,328	23
Trasplantepreviorenal		
No	97,468	77
Si	2,532	2
Causafallecimientodonante		
CCV	44,156	34
TCE	42,857	33
Otras	12,987	10
Sexodonante		
Hombre	52,055	38
Mujer	47,945	35
RECIDIVA_DM1		
No	96,341	79
Si	3,659	3
VHCReceptor		
No	100	79
VHBReceptor		
No	98,734	78
Si	1,266	1
VIHReceptor		
No	100	79
CMVReceptor		
Si	72,152	57
No	27,848	22
VEBReceptor		
Si	95,313	61
No	4,688	3
URETEROGRAFIA		
Normal	91,139	72
Vejiga disfuncional o residuo postmiccional	8,861	7

Sigue en la página siguiente.

Variable	%	N
Trasplantepreviorenal		
No	97,468	77
Si	2,532	2
SeroconversiónCMV		
Negativo	96,296	78
Positivo	3,704	3
ProfilaxisATB		
Si	100	73
Profilaxiscotrimoxazol		
Si	100	73
ProfilaxisCMV		
Si	100	73
Profilaxisantifúngica		
Si	100	73
EMG		
PNSM severa	45,833	33
PNSM moderada	36,111	26
Normal	6,944	5
PNSM leve	11,111	8
RMN_VASCULAR		
Normal	78,049	64
AE leve	14,634	12
AE moderada	6,098	5
Severa	1,22	1
TIPOFCODL		
Ninguno	35	28
Estatina	56,25	45
Fibrato	2,5	2
Ambos	5	4
Ezetimibe	1,25	1
SISTEMA_NERVIOSO_PARASIMPATICO		
Afectación leve	3,39	2
Afectación severa	62,712	37
Normal	3,39	2
Afectación moderada	30,508	18
TIPO_DE_DIALISIS		
Hemodiálisis	98,765	80
Prediálisis	1,235	1
RETINOPATIA_DIABETICA_EN_OJO_DERECHO		
Fotocoagulada	86,585	71
Ausente	1,22	1
Grave o pre-proliferativa	1,22	1
Ceguera	1,22	1
Leve	1,22	1
Proliferativa	1,22	1
Edema macular diabético	4,878	4
Moderada	2,439	2

Sigue en la página siguiente.

Variable	%	N
RETINOPATIA DIABETICA EN OJO IZQUIERDO		
Ceguera	2,439	2
Fotocoagulada	82,927	68
Ausente	1,22	1
Edema macular diabético	6,098	5
Moderada	3,659	3
Proliferativa	2,439	2
Leve	1,22	1
TEST DE ESFUERZO		
Normal	74,684	59
No submáximo	11,392	9
Isquemia leve	11,392	9
No concluyente	2,532	2
ECOCARDIOGRAFIA		
HVI	22,785	18
Normal	73,418	58
Isquemia leve en cara inferior	2,532	2
Derrame pleural leve	1,266	1
ECG		
Normal	82,192	60
BHH	4,11	3
Isquemia leve	5,479	4
HVI	8,219	6
TipodeafectaciónCMV		
Sin enfermedad	97,531	79
Enterocolitis	1,235	1
Hepatitis	1,235	1
ORTOSTATISMO2		
Anómalo	43,103	25
Normal	32,759	19
Límite	24,138	14
MANIOBRA DE VALSALVA2		
Límite	15,517	9
Anómalo	79,31	46
Normal	5,172	3
INDICE_3015_2		
Normal	22,414	13
Anómalo	67,241	39
Límite	10,345	6
DEEPBREATHING2		
Anómalo	89,796	44
Normal	4,082	2
Límite	6,122	3

Sigue en la página siguiente.

Variable	%	N
DIAPASON_EN_PIE_DERECHO2		
Normal	33,333	21
Patológico grave	28,571	18
Patológico leve	26,984	17
Prepatológico	11,111	7
DIAPASON_EN_PIE_IZQUIERDO2		
Normal	28,571	18
Patológico grave	34,921	22
Patológico leve	20,635	13
Prepatológico	15,873	10
NEUROTESIOMETRO_PIE_DERECHO2		
Normal	49,18	30
Patológico	36,066	22
Prepatológico	14,754	9
NEUROTESIOMETRO_PIE_IZQUIERDO2		
Normal	48,333	29
Patológico	33,333	20
Prepatológico	18,333	11
ITB_DERECHO2		
Normal	57,778	26
Calcificación arterial	33,333	15
Enfermedad arterial moderada	8,889	4
ITB_IZQUIERDO2		
Normal	55,556	25
Calcificación arterial	35,556	16
Enfermedad arterial severa	2,222	1
Enfermedad arterial moderada	6,667	3
Profilaxistrombosis		
No	6,78	4
Si	93,22	55
Profilaxistrombosistipo		
Antiagregación	91,071	51
Doble	1,786	1
Ninguna	1,786	1
Anticoagulación	5,357	3
Sexodonante		
Hombre	52,055	38
Mujer	47,945	35
VHCDonante		
Negativo	100	81
VHBDonante		
Negativo	100	81
VIHDonante		
Negativo	100	81
CMVDonante		
Negativo	32,895	25
Positivo	67,105	51
VEBDonante		
Positivo	84,211	16
Negativo	15,789	3

Sigue en la página siguiente.

Variable	%	N
TIPO_TRASPL		
SPR	98,78	81
PAK	1,22	1
PAK_TRAS_SPK		
No	0,951	78
Si	0,049	4
DOSIS_INS		
Nada	80	44
<20 UI	14,545	8
>20 UI	5,455	3
TIPOADO		
Ninguno	100	68
EXITUS		
No	90,244	74
Si	9,756	8
CAUSA_EXITUS		
Shock septico de origen urinario	0,286	2
Trombosis injerto pancreatico + absceso pancreas	0,143	1
Fracaso multiorgánico por fallo renal agudo	0,143	1
Fistula ileo-mesenterica	0,143	1
Infeccion de herida quirurgica	0,286	2
FISTULA		
Si	6,098	5
No	93,902	77
EXPLANTE_DE_PANCREAS		
Si	13,415	11
No	86,585	71
EXPLANTE_RENAL		
No	95,122	78
Si	4,878	4
TROMBOSIS_RENAL		
No	96,341	79
Si	3,659	3
TROMBOSIS_PANCREAS		
No	91,463	75
Si	8,537	7
COMPLICACIONES_QX		
Fístula	3,774	2
Ninguna	50,943	27
Suboclusion intestinal	9,434	5
Hematoma o sangrado	16,981	9
Colección intraabdominal	3,774	2
Infeccion herida quirurgica	11,321	6
Dehiscencia de la herida	1,887	1
Pseudoaneurisma art injerto pancreatico	1,887	1
RECIDIVA_DM1		
No	96,341	79
Si	3,659	3

Sigue en la página siguiente.

Variable	%	N
RECHAZO_PANCREAS		
No	91,463	75
Si	8,537	7
PERD_INJERTO		
Si	31,707	26
No	68,293	56
TIPO_TTO_ACTUAL		
Nada	62,687	42
Insulina >20	13,433	9
ADOS	2,985	2
Insulina <20	20,896	14
NEC_INSULINA_TRAS_TX		
No	65,672	44
Si	34,328	23
Protocolotacrólimus		
PROGRAF	20,313	13
CONVERSION PROGRAF-ADVAGRAF	39,063	25
CONVERSION PROGRAF-MTOR	3,125	2
ADVAGRAF	37,5	24
ContinuaCT		
NO	0,509	27
SI	0,491	26

CUADRO A.2: Resultados del análisis descriptivo de las variables cualitativas

Variable	No Exitus (N)	Exitus (N)	p Valor	H0
PESO	70	8	0,711	0
IMC	70	8	0,085	0
TAS	65	7	0,28	0
TAD	65	7	0,947	0
GLUCEMIA	70	7	0,127	0
HBA1C	68	7	0,438	0
UREA	66	7	0,257	0
FG	13	1	0,11	0
CT	63	6	0,887	0
HDL	61	6	0,662	0
LDL	61	6	0,394	0
TG	64	6	0,066	0
PEPTIDO_C	70	8	0,573	0
UIDIA	70	8	0,07	0
UL_KG_DIA	70	8	0,179	0
DOSIS_ESTATINA	56	5	0,318	0
ORTOSTATISMO	54	2	0,216	0
MANIOBRA_DE_VALSALVA	54	2	1	0
INDICE_3015	54	2	0,364	0
DEEPBREATHING	47	1	0,669	0
NEUROTESIOMETRO_PIE_DERECHO	53	6	0,008	1
ITB_DERECHO	36	1	0,8	0
ITB_IZQUIERDO	34	1	0,44	0
TALLA	70	8	0,055	0
N°CIG	69	8	0,483	0
AÑO_DIAG_DM	70	8	0,611	0
AÑOS_DE_EVOLUCION_DE_LA_DM	70	8	0,878	0
AÑOS_EN_DIALISIS	70	8	0,464	0
CVmáximaCMV	69	8	0,458	0
AÑO_DEL TRASPLANTE	70	8	0,031	1
AÑOS_DEL_PACIENTE_AL TRASPLANTE	70	8	0,372	0
N°PINCHAZO	69	8	0,885	0
NEUROTESIOMETRO_PIE_IZQUIERDO	53	5	0,255	0
FCODL	69	8	0,636	0
AC_GAD	—	—	—	—
AC_IA2	—	—	—	—
AÑO_DE_INICIO_DE_LA_DIALISIS	69	8	0,017	1
INCOMPATIBILIDADESHLA	52	8	0,404	0
DIAPASON_EN_PIE_DERECHO	59	4	0,312	0
DIAPASON_EN_PIE_IZQUIERDO	59	4	0,531	0
ANTIHTA	67	7	0,864	0
Edaddonante	66	8	0,875	0

CUADRO A.3: Resultados variables cuantitativas para el evento final 1: Supervivencia del paciente

Variable	No Exitus		Exitus		χ^2	p Valor
	%	N	%	N		
Sexo						
Hombre	61,429	43	62,5	5	0,003	0,953
Mujer	38,571	27	37,5	3		
HTA						
Si	95,714	67	87,5	7	0,996	0,318
No	4,286	3	12,5	1		
SISTEMA_NERVIOSO_SIMPATICO						
Afectado	65,714	23	—	—	1,82	0,177
Normal	34,286	12	100	1		
ALCOHOL						
No	97,143	68	100	8	0,235	0,628
Si	2,857	2	—	—		
TABACO						
No	55,714	39	75	6	2,333	0,311
Si	21,429	15	25	2		
Ex-fumador	22,857	16	—	—		
DISLIPEMIA						
No	27,143	19	25	2	0,017	0,897
Si	72,857	51	75	6		
CATETERISMO						
Normal	22,727	15	—	—	2,724	0,256
No realizado	71,212	47	100	7		
Estenosis arteria coronaria	6,061	4	—	—		
IAM						
No	97,143	68	100	8	0,235	0,628
Si	2,857	2	—	—		
ACV						
No	97,143	68	100	8	0,235	0,628
Si	2,857	2	—	—		
CLAUDICACION_INTERMITENTE						
No	97,143	68	100	8	0,235	0,628
Si	2,857	2	—	—		
VHCreceptor						
No	100	67	100	8	NaN	NaN
VHBReceptor						
No	98,507	66	100	8	0,121	0,728
Si	1,493	1	—	—		
VIHReceptor						
No	100	67	100	8	NaN	NaN
CMVReceptor						
Si	68,657	46	87,5	7	1,224	0,269
No	31,343	21	12,5	1		
VEBReceptor						
Si	94,915	56	100	3	0,16	0,689
No	5,085	3	—	—		

Sigue en la página siguiente.

Variable	No Exitus %	N	Exitus %	N	χ^2	p Valor
URETEROGRAFIA						
Normal	89,706	61	100	8	0,907	0,341
Vejiga disfuncional o residuo postmiccional	10,294	7	–	–		
Trasplantepreviorenal						
No	100	67	87,5	7	8,488	0,004
Si			12,5	1		
SeroconversiónCMV						
Negativo	95,652	66	100	8	0,362	0,547
Positivo	4,348	3	–	–		
ProfilaxisATB						
Si	100	62	100	8	NaN	NaN
Profilaxiscotrimoxazol						
Si	100	62	100	8	NaN	NaN
ProfilaxisCMV						
Si	100	62	100	8	NaN	NaN
Profilaxisantifúngica						
Si	100	62	100	8	NaN	NaN
EMG						
PNSM severa	44,444	28	71,429	5	3,306	0,347
PNSM moderada	38,095	24	14,286	1		
Normal	6,349	4	14,286	1		
PNSM leve	11,111	7	–	–		
RMN_VASCULAR						
Normal	77,143	54	75	6	0,683	0,877
AE leve	15,714	11	12,5	1		
AE moderada	5,714	4	12,5	1		
Severa	1,429	1	–	–		
TIPOFCODL						
Ninguno	36,765	25	25	2	1,521	0,823
Estatina	54,412	37	75	6		
Ambos	5,882	4	–	–		
Fibrato	1,471	1	–	–		
Ezetimibe	1,471	1	–	–		
SISTEMA_NERVIOSO PARASIMPATICO						
Afectación leve	3,636	2	–	–	14,541	0,002
Afectación severa	67,273	37	–	–		
Afectación moderada	27,273	15	50	1		
Normal	1,818	1	50	1		
TIPO_DE_DIALISIS						
Hemodiálisis	98,571	69	100	8	0,116	0,734
Prediálisis	1,429	1	–	–		

Sigue en la página siguiente.

Variable	No Exitus		Exitus		χ^2	p Valor
	%	N	%	N		
RETINOPATIA DIABETICA EN OJO DERECHO						
Fotocoagulada	85,714	60	87,5	7	9,896	0,195
Grave o pre-proliferativa	1,429	1	-	-		
Ceguera	1,429	1	-	-		
Leve	1,429	1	-	-		
Proliferativa	1,429	1	-	-		
Edema macular diabético	5,714	4	-	-		
Moderada	2,857	2	-	-		
Ausente			12,5	1		
RETINOPATIA DIABETICA EN OJO IZQUIERDO						
Ceguera	2,857	2	-	-	11,683	0,069
Fotocoagulada	82,857	58	75	6		
Edema macular diabético	7,143	5	-	-		
Proliferativa	2,857	2	-	-		
Leve	1,429	1	-	-		
Moderada	2,857	2	12,5	1		
Ausente			12,5	1		
TEST DE ESFUERZO						
Normal	71,642	48	100	8	3,038	0,386
No submáximo	11,94	8	-	-		
Isquemia leve	13,433	9	-	-		
No concluyente	2,985	2	-	-		
ECOCARDIOGRAFIA						
HVI	26,471	18	-	-	3,414	0,332
Normal	69,118	47	100	8		
Isquemia leve en cara inferior	2,941	2	-	-		
Derrame pleural leve	1,471	1	-	-		
ECG						
Normal	79,365	50	100	7	1,774	0,621
BHH	4,762	3	-	-		
Isquemia leve	6,349	4	-	-		
HVI	9,524	6	-	-		
TipodeafectaciónCMV						
Sin enfermedad	97,101	67	100	8	0,238	0,888
Enterocolitis	1,449	1	-	-		
Hepatitis	1,449	1	-	-		
ORTOSTATISMO2						
Anómalo	44,444	24	-	-	1,613	0,446
Límite	24,074	13	50	1		
Normal	31,481	17	50	1		
MANIOBRA DE VALSALVA2						
Límite	14,815	8	50	1	1,812	0,404
Anómalo	79,63	43	50	1		
Normal	5,556	3	-	-		
INDICE 3015.2						
Normal	20,37	11	50	1	1,11	0,574
Anómalo	68,519	37	50	1		
Límite	11,111	6				

Sigue en la página siguiente.

Variable	No Exitus		Exitus		χ^2	p Valor
	%	N	%	N		
DEEPBREATHING2						
Anómalo	89,583	43	100	1	0,116	0,944
Normal	4,167	2	-	-		
Límite	6,25	3	-	-		
DIAPASON_EN PIE_DERECHO2						
Normal	31,034	18	75	3	3,746	0,29
Patológico grave	29,31	17	25	1		
Patológico leve	27,586	16	-	-		
Prepatológico	12,069	7	-	-		
DIAPASON_EN PIE_IZQUIERDO2						
Normal	25,862	15	75	3	4,762	0,19
Patológico grave	36,207	21	25	1		
Patológico leve	20,69	12	-	-		
Prepatológico	17,241	10	-	-		
NEUROTENSIOMETRO PIE_DERECHO2						
Normal	46,296	25	66,667	4	1,203	0,548
Patológico	38,889	21	16,667	1		
Prepatológico	14,815	8	16,667	1		
NEUROTENSIOMETRO PIE_IZQUIERDO2						
Normal	44,444	24	80	4	2,549	0,28
Patológico	35,185	19	20	1		
Prepatológico	20,37	11	-	-		
ITB_DERECHO2						
Normal	56,818	25	100	1	0,747	0,688
Calcificación arterial	34,091	15	-	-		
Enfermedad arterial moderada	9,091	4	-	-		
ITB_IZQUIERDO2						
Normal	54,545	24	100	1	0,818	0,845
Calcificación arterial	36,364	16	-	-		
Enfermedad arterial severa	2,273	1	-	-		
Enfermedad arterial moderada	6,818	3	-	-		
Causafallecimientodonante						
CCV	40,909	27	57,143	4	1,455	0,483
TCE	43,939	29	42,857	3		
Otras	15,152	10	-	-		
Sexodonante						
Hombre	50	31	62,5	5	0,443	0,506
Mujer	50	31	37,5	3		
VHCDonante						
Negativo	100	69	100	8	NaN	NaN

Sigue en la página siguiente.

Variable	No Exitus		Exitus		χ^2	p Valor
	%	N	%	N		
VHBDonante						
Negativo	100	69	100	8	NaN	NaN
VIHDonante						
Negativo	100	69	100	8	NaN	NaN
CMVDonante						
Negativo	30,769	20	28,571	2	0,014	0,905
Positivo	69,231	45	71,429	5		
VEBDonante						
Positivo	88,235	15	50	1	1,968	0,161
Negativo	11,765	2	50	1		
Profilaxistrombosis	-	-	-	-	-	-
Profilaxistrombosistipo	-	-	-	-	-	-

CUADRO A.4: Resultados variables cualitativas para el evento final 1: Supervivencia del paciente

Variable	No explante (N)	Explante (N)	p Valor	H0
PESO	64	6	0,6	0
IMC	64	6	0,774	0
TAS	60	5	0,883	0
TAD	60	5	0,595	0
GLUCEMIA	64	6	0,374	0
HBA1C	62	6	0,273	0
UREA	60	6	0,995	0
FG	-	-	-	-
CT	57	6	0,406	0
HDL	55	6	0,293	0
LDL	55	6	0,303	0
TG	58	6	0,112	0
PEPTIDO_C	64	6	0,059	0
UIDIA	64	6	0,355	0
ULKG_DIA	64	6	0,522	0
DOSIS_ESTATINA	54	2	0,89	0
ORTOSTATISMO	51	3	0,105	0
MANIOBRA_DE_VALSALVA	51	3	0,925	0
INDICE_3015	51	3	0,297	0
DEEPBREATHING	46	1	0,772	0
NEUROTESIOMETRO_PIE_DERECHO	48	5	0,744	0
ITB_DERECHO	34	2	0,196	0
ITB_IZQUIERDO	33	1	0,246	0
TALLA	64	6	0,381	0
N°CIG	64	5	0,399	0
AÑO_DIAG_DM	64	6	0,08	0
AÑOS_DE_EVOLUCION_DE_LA_DM	64	6	0,579	0
AÑOS_EN_DIALISIS	64	6	0,456	0
CVmáximaCMV	63	6	0,501	0
AÑO_DEL_TRASPLANTE	64	6	0,004	1
AÑOS_DEL_PACIENTE_AL_TRASPLANTE	64	6	0,343	0
N°PINCHAZO	63	6	0,078	0
NEUROTESIOMETRO_PIE_IZQUIERDO	48	5	0,315	0
FCODL	63	6	0,99	0
AÑO_DE_INICIO_DE_LA_DIALISIS	63	6	0,003	1
INCOMPATIBILIDADESHLA	47	5	0,91	0
DIAPASON_EN_PIE_DERECHO	54	5	0,934	0
DIAPASON_EN_PIE_IZQUIERDO	54	5	0,621	0
AC_GAD	-	-	-	-
AC_IA2	-	-	-	-
ANTIHTA	62	5	0,036	0
Edaddonante	61	5	0,585	0

CUADRO A.5: Resultados variables cuantitativas para el evento final 2: Explante de páncreas

Variable	No Explante		Explante		χ^2	p Valor
	%	N	%	N		
Sexo						
Hombre	59,375	38	83,333	5	1,329	0,249
Mujer	40,625	26	16,667	1		
HTA						
No	3,125	2	16,667	1	2,452	0,117
Si	96,875	62	83,333	5		
SISTEMA_NERVIOSO_SIMPATICO						
ALCOHOL						
No	98,438	63	83,333	5	4,509	0,034
Si	1,563	1	16,667	1		
TABACO						
Si	21,875	14	16,667	1	2,775	0,25
No	57,813	37	33,333	2		
Ex-fumador	20,313	13	50	3		
DISLIPEMIA						
No	26,563	17	33,333	2	0,127	0,721
Si	73,438	47	66,667	4		
CATETERISMO						
No realizado	72,131	44	60	3	1,134	0,567
Normal	21,311	13	40	2		
Estenosis arteria coronaria	6,557	4	-	-		
IAM						
No	96,875	62	100	6	0,193	0,66
Si	3,125	2	-	-		
ACV						
No	96,875	62	100	6	0,193	0,66
Si	3,125	2	-	-		
CLAUDICACION_INTERMITENTE						
No	96,875	62	100	6	0,193	0,66
Si	3,125	2	-	-		
VHCreceptor						
No	100	61	100	6	NaN	NaN
VHBReceptor						
No	98,361	60	100	6	0,1	0,752
Si	1,639	1	-	-		
VIHReceptor						
No	100	61	100	6	NaN	NaN
CMVReceptor						
Si	67,213	41	83,333	5	0,66	0,417
No	32,787	20	16,667	1		
VEBReceptor						
Si	94,34	50	100	6	0,358	0,55
No	5,66	3	-	-		
URETEROGRAFIA						
Normal	90,323	56	83,333	5	0,289	0,591
Vejiga disfuncional o residuo postmiccional	9,677	6	16,667	1		
Trasplantepreviorenal						
No	100	61	100	6	NaN	NaN

Sigue en la página siguiente.

Variable	No Explante %	Explante N	Explante %	Explante N	χ^2	p Valor
SeroconversiónCMV						
Negativo	95,238	60	100	6	0,299	0,585
Positivo	4,762	3	-	-		
ProfilaxisATB						
Si	100	56	100	6	NaN	NaN
Profilaxiscotrimoxazol						
Si	100	56	100	6	NaN	NaN
ProfilaxisCMV						
Si	100	56	100	6	NaN	NaN
Profilaxisantifúngica						
Si	100	56	100	6	NaN	NaN
EMG						
PNSM moderada	39,655	23	20	1	2,96	0,398
Normal	6,897	4	-	-		
PNSM leve	12,069	7	-	-		
PNSM severa	41,379	24	80	4		
RMN_VASCULAR						
AE leve	17,188	11	-	-	9,979	0,019
Normal	78,125	50	66,667	4		
Severa	1,563	1	-	-		
AE moderada	3,125	2	33,333	2		
TIPOFCODL						
Ninguno	35,484	22	50	3	0,918	0,922
Estatina	54,839	34	50	3		
Ambos	6,452	4	-	-		
Fibrato	1,613	1	-	-		
Ezetimibe	1,613	1	-	-		
SISTEMA_NERVIOSO PARASIMPATICO						
Afectación severa	69,231	36	33,333	1	8,339	0,039
Afectación leve	1,923	1	33,333	1		
Afectación moderada	26,923	14	33,333	1		
Normal	1,923	1	-	-		
TIPO_DE_DIALISIS						
Hemodiálisis	98,438	63	100	6	0,095	0,758
Prediálisis	1,563	1	-	-		
RETINOPATIA_DIABETICA_EN OJO_DERECHO						
Fotocoagulada	87,5	56	66,667	4	22,361	0,001
Leve	1,563	1	-	-		
Proliferativa	1,563	1	-	-		
Edema macular diabético	6,25	4	-	-		
Moderada	3,125	2	-	-		
Grave o pre-proliferativa			16,667	1		
Ceguera			16,667	1		
VIHDonante	-	-	-	-	-	-

Sigue en la página siguiente.

Variable	No Explante %	N	Explante %	N	χ^2	p Valor
RETINOPATIA DIABETICA EN OJO_IZQUIERDO						
Fotocoagulada	85,938	55	50	3	23,49	0
Proliferativa	3,125	2	-	-		
Edema macular diabético	6,25	4	16,667	1		
Leve	1,563	1	-	-		
Moderada	3,125	2	-	-		
Ceguera	-	-	33,333	2		
TEST DE ESFUERZO						
No submáximo	13,115	8	-	-	1,161	0,762
Normal	70,492	43	83,333	5		
Isquemia leve	13,115	8	16,667	1		
No concluyente	3,279	2	-	-		
ECOCARDIOGRAFIA						
Normal	67,742	42	83,333	5	0,722	0,868
HVI	27,419	17	16,667	1		
Isquemia leve en cara inferior	3,226	2	-	-		
Derrame pleural leve	1,613	1	-	-		
ECG						
Normal	77,966	46	100	4	1,111	0,775
BHH	5,085	3	-	-		
Isquemia leve	6,78	4	-	-		
HVI	10,169	6	-	-		
TipodeafectaciónCMV						
Sin enfermedad	96,825	61	100	6	0,196	0,907
Enterocolitis	1,587	1	-	-		
Hepatitis	1,587	1	-	-		
ORTOSTATISMO2						
Anómalo	43,137	22	66,667	2	1,121	0,571
Límite	25,49	13				
Normal	31,373	16	33,333	1		
MANIOBRA DE VALSALVA2						
Anómalo	82,353	42	33,333	1	6,796	0,033
Normal	5,882	3	-	-		
Límite	11,765	6	66,667	2		
INDICE_3015_2						
Anómalo	70,588	36	33,333	1	4,269	0,118
Normal	17,647	9	66,667	2		
Límite	11,765	6	-	-		
DEEPBREATHING2						
Anómalo	89,362	42	100	1	0,119	0,942
Normal	4,255	2	-	-		
Límite	6,383	3	-	-		
DIAPASON EN PIE_DERECHO2						
Patológico grave	28,302	15	40	2	1,13	0,77
Normal	30,189	16	40	2		
Prepatológico	13,208	7	-	-		
Patológico leve	28,302	15	20	1		

Sigue en la página siguiente.

Variable	No Explante %	Explante N	Explante %	Explante N	χ^2	p Valor
DIAPASON_EN						
PIE_IZQUIERDO2						
Patológico grave	35,849	19	40	2	0,12	0,989
Patológico leve	20,755	11	20	1		
Normal	26,415	14	20	1		
Prepatológico	16,981	9	20	1		
NEUROTENSIOMETRO						
PIE_DERECHO2						
Patológico	38,776	19	40	2	0,149	0,928
Normal	46,939	23	40	2		
Prepatológico	14,286	7	20	1		
NEUROTENSIOMETRO						
PIE_IZQUIERDO2						
Patológico	32,653	16	60	3	1,706	0,426
Prepatológico	20,408	10	20	1		
Normal	46,939	23	20	1		
ITB_DERECHO2						
Normal	57,143	24	50	1	0,363	0,834
Calcificación arterial	33,333	14	50	1		
Enfermedad arterial moderada	9,524	4	-	-		
ITB_IZQUIERDO2						
Normal	55,814	24	-	-	1,791	0,617
Calcificación arterial	34,884	15	100	1		
Enfermedad arterial severa	2,326	1	-	-		
Enfermedad arterial moderada	6,977	3	-	-		
Profilaxistrombosis						
No	3,774	2	66,667	2	16,933	0,00003
Si	96,226	51	33,333	1		
Profilaxistrombosis tipo						
Antiagregación	92,157	47	50	1	26,035	0,0000093
Doble	1,961	1	-	-		
Anticoagulación	5,882	3	-	-		
Ninguna	-	-	50	1		
Causafallecimientodonante						
TCE	42,623	26	60	3	1,138	0,566
CCV	40,984	25	40	2	-	-
Otras	16,393	10	-	-	-	-
Sexodonante						
Hombre	47,368	27	80	4	1,958	0,162
Mujer	52,632	30	20	1	-	-
VHCDonante						
Negativo	100	63	100	6	NaN	NaN
VHBDonante						
Negativo	100	63	100	6	NaN	NaN
Negativo	100	63	100	6	NaN	NaN
CMVDDonante						
Positivo	70	42	60	3	0,217	0,641
Negativo	30	18	40	2	-	-
VEBDDonante						
Positivo	93,75	15	-	-	7,969	0,004
Negativo	6,25	1	100	1	-	-

CUADRO A.6: Resultados variables cualitativas para Explante de páncreas

Variable	No Pérdida injerto (N)	Sí Pérdida injerto (N)	P	H0
PESO	56	14	0,708	0
IMC	56	14	0,859	0
TAS	52	13	0,362	0
TAD	52	13	0,564	0
GLUCEMIA	56	14	0,397	0
HBA1C	55	13	0,534	0
UREA	53	13	0,117	0
FG	-	-	-	-
CT	50	13	0,459	0
HDL	47	14	0,288	0
LDL	47	14	0,661	0
TG	50	14	0,094	0
PEPTIDO_C	56	14	0,147	0
UIDIA	56	14	0,453	0
UI_KG_DIA	56	14	0,268	0
DOSIS_ESTATINA	49	7	0,877	0
ORTOSTATISMO	45	9	0,95	0
MANIOBRA_DE_VALSALVA	44	10	0,876	0
INDICE_3015	44	10	0,012	1
DEEPBREATHING	41	6	0,911	0
NEUROTESIOMETRO_PIE_DERECHO	42	11	0,568	0
ITB_DERECHO	28	8	0,057	0
ITB_IZQUIERDO	27	7	0,384	0
TALLA	56	14	0,783	0
N°CIG	56	13	0,041	1
AÑO_DIAG_DM	56	14	0,73	0
AÑOS_DE_EVOLUCION_DE_LA_DM	56	14	0,343	0
AÑOS_EN_DIALISIS	56	14	0,773	0
CVmáximaCMV	55	14	0,083	0
AÑO_DEL_TRASPLANTE	56	14	0,006	1
AÑOS_DEL_PACIENTE_AL_TRASPLANTE	56	14	0,413	0
N°PINCHAZO	55	14	0,534	0
NEUROTESIOMETRO_PIE_IZQUIERDO	42	11	0,367	0
FCODL	55	14	0,879	0
AÑO_DE_INICIO_DE_LA_DIALISIS	55	14	0,027	1
INCOMPATIBILIDADESHLA	40	12	0,074	0
DIAPASON_EN_PIE_DERECHO	46	13	0,869	0
DIAPASON_EN_PIE_IZQUIERDO	46	13	0,972	0
Edaddonante	54	12	0,656	0
AC_GAD	-	-	-	-
AC_AI2	-	-	-	-
ANTIHTA	54	13	0,109	0

CUADRO A.7: Resultados variables cuantitativas para el evento final 3: Pérdida de injerto

Variable	No Pérdida injerto		Si Pérdida injerto		χ^2	p Valor
	%	N	%	N		
Sexo						
Hombre	55,357	31	85,714	12	4,356	0,037
Mujer	44,643	25	14,286	2		
HTA						
No	3,571	2	7,143	1	0,348	0,555
Si	96,429	54	92,857	13		
SISTEMA_NERVIOSO_SIMPATICO						
Afectado	66,667	20	60	3	0,085	0,771
Normal	33,333	10	40	2		
ALCOHOL						
No	98,214	55	92,857	13	1,158	0,282
Si	1,786	1	7,143	1		
TABACO						
Si	23,214	13	14,286	2	0,663	0,718
Ex-fumador	23,214	13	21,429	3		
No	53,571	30	64,286	9		
DISLIPEMIA						
No	25	14	35,714	5	0,65	0,42
Si	75	42	64,286	9		
CATETERISMO						
No realizado	75,926	41	50	6	6,616	0,037
Normal	16,667	9	50	6		
Estenosis arteria coronaria	7,407	4	-	-		
IAM						
No	98,214	55	92,857	13	1,158	0,282
Si	1,786	1	7,143	1		
ACV						
No	96,429	54	100	14	0,515	0,473
Si	3,571	2	-	-		
CLAUDICACION_INTERMITENTE						
No	96,429	54	100	14	0,515	0,473
Si	3,571	2	-	-		
VHCReceptor						
No	100	53	100	14	NaN	NaN
VHBReceptor						
No	98,113	52	100	14	0,268	0,605
Si	1,887	1	-	-		
VIHReceptor						
No	100	53	100	14	NaN	NaN
CMVReceptor						
Si	66,038	35	78,571	11	0,808	0,369
No	33,962	18	21,429	3		
VEBReceptor						
Si	93,75	45	100	11	0,724	0,395
No	6,25	3				
URETEROGRAFIA						
Normal	89,091	49	92,308	12	0,118	0,731
Vejiga disfuncional o residuo postmiccional	10,909	6	7,692	1		
Trasplantepreviorenal						
No	100	53	100	14	NaN	NaN

Sigue en la página siguiente.

Variable	No Pérdida injerto		Si Pérdida injerto		χ^2	p Valor
	%	N	%	N		
SeroconversiónCMV						
Negativo	96,364	53	92,857	13	0,33	0,566
Positivo	3,636	2	7,143	1		
ProfilaxisATB						
Si	100	48	100	14	NaN	NaN
Profilaxiscotrimoxazol						
Si	100	48	100	14	NaN	NaN
ProfilaxisCMV						
Si	100	48	100	14	NaN	NaN
Profilaxisantifúngica						
Si	100	48	100	14	NaN	NaN
EMG						
PNSM moderada	42	21	23,077	3	3,574	0,311
PNSM leve	12	6	7,692	1		
PNSM severa	42	21	53,846	7		
Normal	4	2	15,385	2		
RMN_VASCULAR						
AE leve	17,857	10	7,143	1	8,705	0,033
Normal	78,571	44	71,429	10		
Severa	1,786	1	-	-		
AE moderada	1,786	1	21,429	3		
TIPOFCODL						
Ninguno	35,185	19	42,857	6	0,807	0,937
Estatina	55,556	30	50	7		
Ambos	5,556	3	7,143	1		
Fibrato	1,852	1	-	-		
Ezetimibe	1,852	1	-	-		
SISTEMA_NERVIOSO PARASIMPATICO						
Afectación severa	71,111	32	50	5	6,436	0,092
Afectación leve	2,222	1	10	1		
Afectación moderada	26,667	12	30	3		
Normal	-	-	10	1		
TIPO_DE_DIALISIS						
Hemodiálisis	98,214	55	100	14	0,254	0,615
Prediálisis	1,786	1				
RETINOPATIA_DIABETICA_EN OJO_DERECHO						
Fotocoagulada	87,5	49	78,571	11	13,854	0,031
Proliferativa	1,786	1	-	-		
Edema macular diabético	7,143	4	-	-		
Moderada	3,571	2	-	-		
Grave o pre-proliferativa			7,143	1		
Ceguera			7,143	1		
Leve			7,143	1		
RETINOPATIA_DIABETICA_EN OJO_IZQUIERDO						
Fotocoagulada	89,286	50	57,143	8	16,272	0,006
Proliferativa	1,786	1	7,143	1		
Edema macular diabético	5,357	3	14,286	2		
Moderada	3,571	2	-	-		
Ceguera	-	-	14,286	2		
Leve	-	-	7,143	1		

Sigue en la página siguiente.

Variable	No Pérdida injerto		Si Pérdida injerto		χ^2	p Valor
	%	N	%	N		
TEST DE ESFUERZO						
No submáximo	11,321	6	14,286	2	0,62	0,892
Normal	71,698	38	71,429	10		
Isquemia leve	13,208	7	14,286	2		
No concluyente	3,774	2				
ECOCARDIOGRAFIA						
Normal	69,091	38	69,231	9	0,821	0,844
HVI	25,455	14	30,769	4		
Isquemia leve en cara inferior	3,636	2	-	-		
Derrame pleural leve	1,818	1	-	-		
ECG						
Normal	80,769	42	72,727	8	5,963	0,113
Isquemia leve	7,692	4	-	-		
BHH	1,923	1	18,182	2		
HVI	9,615	5	9,091	1		
TipodeafectaciónCMV						
Sin enfermedad	96,364	53	100	14	0,524	0,769
Enterocolitis	1,818	1	-	-		
Hepatitis	1,818	1	-	-		
ORTOSTATISMO2						
Anómalo	44,444	20	44,444	4	1,33	0,514
Límite	26,667	12	11,111	1		
Normal	28,889	13	44,444	4		
MANIOBRA DE VALSALVA2						
Anómalo	86,364	38	50	5	7,044	0,03
Normal	4,545	2	10	1		
Límite	9,091	4	40	4		
INDICE_3015_2						
Anómalo	75	33	40	4	12,282	0,002
Normal	11,364	5	60	6		
Límite	13,636	6				
DEEPBREATHING2						
Anómalo	90,476	38	83,333	5	3,03	0,22
Normal	2,381	1	16,667	1		
Límite	7,143	3				
DIAPASON_EN PIE_DERECHO2						
Patológico grave	28,261	13	33,333	4	1,536	0,674
Normal	34,783	16	16,667	2		
Prepatológico	10,87	5	16,667	2		
Patológico leve	26,087	12	33,333	4		
DIAPASON_EN PIE_IZQUIERDO2						
Patológico grave	39,13	18	25	3	5,764	0,124
Normal	30,435	14	8,333	1		
Prepatológico	13,043	6	33,333	4		
Patológico leve	17,391	8	33,333	4		
NEUROTENSIOMETRO PIE_DERECHO2						
Patológico	38,095	16	41,667	5	0,138	0,933
Normal	47,619	20	41,667	5		
Prepatológico	14,286	6	16,667	2		

Sigue en la página siguiente.

Variable	No Pérdida injerto		Si Pérdida injerto		χ^2	p Valor
	%	N	%	N		
NEUROTENSIOMETRO						
PIE_IZQUIERDO2						
Patológico	33,333	14	41,667	5	0,775	0,679
Normal	47,619	20	33,333	4		
Prepatológico	19,048	8	25	3		
ITB_DERECHO2						
Normal	54,286	19	66,667	6	1,224	0,542
Calcificación arterial	34,286	12	33,333	3		
Enfermedad arterial moderada	11,429	4	-	-		
ITB_IZQUIERDO2						
Normal	52,778	19	62,5	5	1,006	0,8
Calcificación arterial	36,111	13	37,5	3		
Enfermedad arterial severa	2,778	1	-	-		
Enfermedad arterial moderada	8,333	3	-	-		
Profilaxistrombosis						
No	4,444	2	18,182	2	2,515	0,113
Si	95,556	43	81,818	9		
Profilaxistrombosistipo						
Antiagregación	90,698	39	90	9	5,231	0,156
Doble	2,326	1	-	-		
Anticoagulación	6,977	3	-	-		
Ninguna	-	-	10	1		
Causafallecimientodonante						
TCE	42,593	23	50	6	1,969	0,374
CCV	44,444	24	25	3	-	-
Otras	12,963	7	25	3	-	-
Sexodonante						
Hombre	47,059	24	63,636	7	0,995	0,319
Mujer	52,941	27	36,364	4	-	-
VHCDonante						
Negativo	100	55	100	14	NaN	NaN
VHBDonante						
Negativo	100	55	100	14	NaN	NaN
VIHDonante						
Negativo	100	55	100	14	NaN	NaN
CMVDonante						
Positivo	67,308	35	76,923	10	0,451	0,502
Negativo	32,692	17	23,077	3	-	-
VEBDonante						
Positivo	100	12	60	3	5,44	0,02
Negativo						

CUADRO A.8: Resultados variables cualitativas para el evento final 3: Pérdida de injerto

Variable	No insulina (N)	Sí insulina (N)	p Valor	H0
PESO	42	11	0,863	0
IMC	42	11	0,674	0
TAS	39	10	0,954	0
TAD	39	10	0,172	0
GLUCEMIA	42	11	0,877	0
HBA1C	41	11	0,901	0
UREA	40	10	0,233	0
FG	2	8	0,654	0
CT	36	11	0,096	0
HDL	33	11	0,428	0
LDL	33	11	0,263	0
TG	36	11	0,633	0
PEPTIDO_C	42	11	0,189	0
UIDIA	42	11	0,947	0
UI_KG_DIA	42	11	0,861	0
DOSIS_ESTATINA	35	11	0,485	0
ORTOSTATISMO	31	11	0,366	0
MANIOBRA_DE_VALSALVA	30	11	0,414	0
INDICE_3015	30	11	0,039	1
DEEPBREATHING	27	11	0,859	0
NEUROTESIOMETRO_PIE_DERECHO	29	10	0,733	0
ITB_DERECHO	19	6	0,829	0
ITB_IZQUIERDO	19	5	0,357	0
TALLA	42	11	0,834	0
N°CIG	42	11	0,542	0
AÑO_DIAG_DM	42	11	0,706	0
AÑOS_DE_EVOLUCION_DE_LA_DM	42	11	0,095	0
AÑOS_EN_DIALISIS	42	11	0,382	0
CVmáximaCMV	41	11	0,509	0
AÑO_DEL_TRASPLANTE	42	11	0,003	1
AÑOS_DEL_PACIENTE_AL_TRASPLANTE	42	11	0,727	0
N°PINHAZO	41	11	0,454	0
NEUROTESIOMETRO_PIE_IZQUIERDO	29	10	0,837	0
FCODL	41	11	0,659	0
AÑO_DE_INICIO_DE_LA_DIALISIS	42	11	0,004	1
INCOMPATIBILIDADESHLA	33	7	0,207	0
DIAPASON_EN_PIE_DERECHO	32	11	0,606	0
DIAPASON_EN_PIE_IZQUIERDO	32	11	0,559	0
ANTIHTA	41	10	0,319	0
AC_GAD	9	7	0,482	0
AC_IA2	9	6	0,563	0
Edaddonante	40	11	0,221	0

CUADRO A.9: Resultados variables línea

Variable	No insulina		Insulina		χ^2	p Valor
	%	N	%	N		
Sexo						
Hombre	59,524	25	36,364	4	1,18	0,277
Mujer	40,476	17	63,636	7		
HTA						
No	4,762	2	-	-	0,691	0,406
Si	95,238	40	100	11		
SISTEMA_NERVIOSO_SIMPATICO						
Afectado	68,421	13	75	6	0,072	0,789
Normal	31,579	6	25	2		
ALCOHOL						
No	97,619	41	100	11	0,339	0,56
Si	2,381	1	-	-		
TABACO						
Si	19,048	8	27,273	3	1,682	0,431
Ex-fumador	23,81	10	27,273	3		
No	57,143	24	45,455	5		
DISLIPEMIA						
No	28,571	12	9,091	1	1,143	0,285
Si	71,429	30	90,909	10		
CATETERISMO						
No realizado	67,5	27	100	11	5,993	0,05
Normal	22,5	9	-	-		
Estenosis arteria coronaria	10	4	-	-		
IAM						
No	97,619	41	100	11	0,339	0,56
Si	2,381	1	-	-		
ACV						
No	97,619	41	90,909	10	0,691	0,406
Si	2,381	1	9,091	1		
CLAUDICACION_INTERMITENTE						
No	97,619	41	100	11	0,691	0,406
Si	2,381	1	-	-		
VHCReceptor						
No	100	42	100	11	NaN	NaN
VHBReceptor						
No	97,619	41	100	11	0,267	0,605
Si	2,381	1	-	-		
VIHReceptor						
No	100	42	100	11	NaN	NaN
CMVReceptor						
Si	64,286	27	72,727	8	0,277	0,599
No	35,714	15	27,273	3		
VEBReceptor						
Si	92,308	36	100	9	0,738	0,39
No	7,692	3	-	-		
URETEROGRAFIA						
Normal	90,244	37	90,909	10	0,22	0,639
Vejiga disfuncional o residuo postmiccional	9,756	4	9,091	1		
Trasplantepreviorenal						
No	100	42	100	11	NaN	NaN

Sigue en la página siguiente.

Variable	No insulina		Insulina		χ^2	p Valor
	%	N	%	N		
SeroconversiónCMV						
Negativo	97,561	40	90,909	10	0,659	0,417
Positivo	2,439	1	9,091	1		
ProfilaxisATB						
Si	100	40	100	8	NaN	NaN
Profilaxiscotrimoxazol						
Si	100	40	100	8	NaN	NaN
ProfilaxisCMV						
Si	100	40	100	8	NaN	NaN
Profilaxisantifúngica						
Si	100	40	100	8	NaN	NaN
EMG						
PNSM moderada	37,838	14	60	6	1,985	0,576
PNSM leve	10,811	4	20	2		
PNSM severa	45,946	17	20	2		
Normal	5,405	2	-	-		
RMN_VASCULAR						
AE leve	16,667	7	18,182	2	0,8	0,849
Normal	78,571	33	81,818	9		
Severa	2,381	1	-	-		
AE moderada	2,381	1	-	-		
TIPOFCODL						
Ninguno	37,5	15	27,273	3	1,28	0,865
Estatina	52,5	21	72,727	8		
Ambos	5	2	-	-		
Fibrato	2,5	1	-	-		
Ezetimibe	2,5	1	-	-		
SISTEMA_NERVIOSO						
PARASIMPATICO						
Afectación severa	74,194	23	72,727	8	1,209	0,546
Afectación leve	3,226	1	-	-		
Afectación moderada	22,581	7	27,273	3		
TIPO_DE_DIALISIS						
Hemodiálisis	100	42	100	11	3,055	0,081
RETINOPATIA_DIABETICA_EN						
OJO_DERECHO						
Fotocoagulada	85,714	36	100	11	1,061	0,786
Proliferativa	2,381	1	-	-		
Edema macular diabético	7,143	3	-	-		
Moderada	4,762	2	-	-		
RETINOPATIA_DIABETICA_EN						
OJO_IZQUIERDO						
Fotocoagulada	88,095	37	100	11	1,138	0,768
Proliferativa	2,381	1	-	-		
Edema macular diabético	4,762	2	-	-		
Moderada	4,762	2	-	-		
TEST_DE_ESFUERZO						
No submáximo	12,5	5	10	1	1,364	0,714
Normal	70	28	80	8		
Isquemia leve	15	6	10	1		
No concluyente	2,5	1	-	-		
ECOCARDIOGRAFIA						

Sigue en la página siguiente.

Variable	No insulina		Insulina		χ^2	p Valor
	%	N	%	N		
Normal	70,732	29	63,636	7	1,111	0,774
HVI	24,39	10	27,273	3		
Isquemia leve en cara inferior	2,439	1	9,091	1		
Derrame pleural leve	2,439	1	-	-		
EKG						
Normal	86,842	33	54,545	6	4,876	0,181
Isquemia leve	5,263	2	18,182	2		
HVI	7,895	3	18,182	2		
BHH	-	-	9,091	1		
TipodeafectaciónCMV						
Sin enfermedad	97,561	40	90,909	10	3,294	0,193
Enterocolitis	2,439	1	-	-		
Hepatitis	-	-	9,091	1		
ORTOSTATISMO2						
Anómalo	51,613	16	27,273	3	2,104	0,349
Límite	22,581	7	45,455	5		
Normal	25,806	8	27,273	3		
MANIOBRA_DE_VALSALVA2						
Anómalo	83,333	25	90,909	10	2,272	0,321
Normal	3,333	1	9,091	1		
Límite	13,333	4	-	-		
INDICE_3015_2						
Anómalo	70	21	81,818	9	1,271	0,53
Normal	13,333	4	9,091	1		
Límite	16,667	5	9,091	1		
DEEPBREATHING2						
Anómalo	85,714	24	100	11	2,211	0,331
Normal	3,571	1	-	-		
Límite	10,714	3	-	-		
DIAPASON_EN						
PIE_DERECHO2						
Patológico grave	25	8	36,364	4	1,002	0,801
Normal	37,5	12	27,273	3		
Prepatológico	9,375	3	9,091	1		
Patológico leve	28,125	9	27,273	3		
DIAPASON_EN						
PIE_IZQUIERDO2						
Patológico grave	34,375	11	54,545	6	8,364	0,039
Normal	34,375	11	18,182	2		
Prepatológico	6,25	2	27,273	3		
Patológico leve	25	8	-	-		
NEUROTENSIOMETRO						
PIE_DERECHO2						
Patológico	37,931	11	40	4	0,025	0,988
Normal	48,276	14	40	4		
Prepatológico	13,793	4	20	2		

Sigue en la página siguiente.

Variable	No insulina		Insulina		χ^2	p Valor
	%	N	%	N		
ITB_DERECHO2						
Normal	54,167	13	37,5	3	0,097	0,953
Calcificación arterial	33,333	8	50	4		
Enfermedad arterial moderada	12,5	3	12,5	1		
ITB_IZQUIERDO2						
Normal	54,167	13	33,333	3	2,065	0,559
Calcificación arterial	37,5	9	44,444	4		
Enfermedad arterial severa	4,167	1				
Enfermedad arterial moderada	4,167	1	22,222	2		
Profilaxistrombosis						
No	5,405	2			0,453	0,501
Si	94,595	35	100	8		
Profilaxistrombosistipo						
Antiagregación	91,429	32	87,5	7	5,072	0,079
Anticoagulación	8,571	3				
Doble	-	-	12,5	1		
Causafallecimientodonante						
TCE	45	18	36,364	4	1,39	0,499
CCV	40	16	54,545	6	-	-
Otras	15	6	9,091	1	-	-
Sexodonante						
Hombre	55,263	21	20	2	4,028	0,045
Mujer	44,737	17	80	8	-	-
VHCDonante						
Negativo	100	41	100	11	NaN	NaN
VHBDonante						
Negativo	100	41	100	11	NaN	NaN
VIHDonante						
Negativo	100	41	100	11	NaN	NaN
CMVDonante						
Positivo	61,538	24	80	8	2,36	0,125
Negativo	38,462	15	20	2	-	-
VEBDonante						
Positivo	100	10	100	2	NaN	NaN
NEUROTENSIOMETRO						
PIE_IZQUIERDO2						
Patológico	27,586	8	50	5	1,393	0,498
Normal	51,724	15	30	3		
Prepatológico	20,69	6	20	2		

CUADRO A.10: Resultados variables cualitativas para el evento final 4: Necesidad de insulina

Apéndice B

Variables del estudio

En este apéndice se ofrece una descripción del conjunto total de variables del estudio.

Nº	Nombre	Unidad	Descripción
1	PESO	kg	Peso del paciente
2	IMC	kg/m ²	Índice de masa corporal
3	TAS	mmHg	Tensión arterial sistólica
4	TAD	mmHg	Tensión arterial diastólica
5	GLUCEMIA	mg/dl	Nivel de glucosa en sangre
6	SOG.GLUC	mg/dl	Resultado de la prueba sobrecarga oral de glucosa
7	HBA1C	%	Nivel de Hemoglobina glicosilada
8	UREA	mg/dl	Niveles de Urea en sangre
9	CREATININA	mg/dl	Niveles de Creatinina en sangre
10	FG	ml/min	Niveles de filtración glomerular
11	CT	mg/dl	Cifras de colesterol Total
12	HDL	mg/dl	Nivel de colesterol (HDL)
13	LDL	mg/dl	Nivel de colesterol (LDL)
14	TG	mg/dl	Valores de triglicéridos en sangre
15	Albúmina	g/dl	Nivel de Albúmina en sangre
16	Proteinuria	g/24horas	Nivel de proteínas en orina de 24 horas
17	INSULINA	mcU/mL	Nivel de insulina en sangre
18	PEPTIDO_C	ng/mL	Nivel de péptido C en sangre
19	AC_GAD	U/ml	Niveles de anticuerpos Anti glutamato descarboxilasa
20	AC_IA2	U/ml	Niveles de Autoanticuerpos anti- tirosin fosfatasa 2
21	NºPINCHAZO	-	Nro pinchazos insulina al día
22	DOSIS_INS	-	Dosis insulina en cada pinchazo
23	UIDIA	UI/día	UI de insulina al día
24	ULKG_DIA	UI/kg/día	UI por kg del paciente al día
25	NºADO	-	nro de fármacos antidiabéticos orales que toma el paciente al día
26	TIPOADO	-	tipo de fármacos antidiabéticos orales que toma al día
27	ANTIHTA	-	Nro de fármacos para la hipertensión arterial diarios
28	TPOANTIHTA	-	Tipo de fármacos para la hipertensión arterial
29	FCODL	-	Nro de fármacos para el colesterol que toma el paciente
30	TIPOFCODL	-	Tipo de fármacos para el colesterol que toma el paciente
31	DOSIS.ESTATINA	mg	Dosis de estatina
32	ORTOSTATISMO	mmHg	Hipotensión ortóstica (caída de 20 mmHg o más en la TAS)
33	ORTOSTATISMO2	-	Clasificación Ortostatismo en el paciente

Sigue en la página siguiente.

Nº	Nombre	Unidad	Descripción
34	SISTEMA_NERVIOSO_SIMPATICO	-	Estado sistema nerviosos simpatico del paciente
35	SISTEMA_NERVIOSO_PARASIMPATICO	-	Estado sistema nerviosos parasimpatico del paciente
36	MANIOBRA_DE_VALSALVA	mmHg	Tipo de maniobra que se realiza para analizar si hay neuropatía después del trasplante
37	MANIOBRA_DE_VALSALVA2	-	Clasificación según los resultados de la maniobra de Valsalva
38	INDICE_3015	mmHg	Tipo de maniobra que se realiza para analizar si hay neuropatía después del trasplante
39	INDICE_3015_2	-	Clasificación según los resultados de la maniobra de Índice 30:15
40	DEEPBREATHING	mmHg	Tipo de maniobra que se realiza para analizar si hay neuropatía después del trasplante
41	DEEPBREATHING2	-	Clasificación según los resultados de Deepbreathing
42	NEUROTESIOMETRO_PIE_DERECHO	mV	Resultados neurotensiómetro pie derecho
43	NEUROTESIOMETRO_PIE_DERECHO2	-	Clasificación resultados neurotensiómetro pie derecho
44	NEUROTESIOMETRO_PIE_IZQUIERDO	mV	Resultados neurotensiómetro pie izquierdo
45	NEUROTESIOMETRO_PIE_IZQUIERDO2	-	Clasificación resultados neurotensiómetro pie izquierdo
46	ITB_DERECHO	mmHg	Índice de tobillo-brazo derecho
47	ITB_DERECHO2	-	Clasificación según el índice de tobillo-brazo derecho
48	ITB_IZQUIERDO	mmHg	índice de tobillo-brazo izquierdo
49	ITB_IZQUIERDO2	-	Clasificación según el índice de tobillo-brazo izquierdo
50	DIAPASON_EN_PIE_DERECHO	-	Resultado diapason en pie derecho
51	DIAPASON_EN_PIE_DERECHO2	-	Clasificación según los resultados diapason en pie derecho
52	DIAPASON_EN_PIE_IZQUIERDO	-	Resultado diapason en pie en pie izquierdo
53	DIAPASON_EN_PIE_IZQUIERDO2	-	Clasificación según los resultados diapason en pie izquierdo
54	Dosistacrólimus	-	Dosis suministrada de Tacrolimus
55	Nivelestacrólimus	ng/mL	Niveles de tacrólimus en sangre después de la operación
56	DosisMMF	mg	Dosis suministrada de micofenolato de mofetilo
57	DosisCT	mg	Dosis de corticoides que toma el paciente
58	Sexo	-	Sexo del paciente
59	TALLA	cm	Talla del paciente
60	FechaTXPR	-	Fecha del trasplante
61	AÑO_DEL_TRASPLANTE	-	Año del trasplante
62	TIEMPO_EVOL_DESDE_EL_TRASPLANTE	años	Años desde el trasplante hasta el último año de seguimiento
63	AÑO_NACIMIENTO	año	Año de nacimiento del paciente
64	AÑOS_DEL_PACIENTE_AL_TRASPLANTE	años	Edad del paciente en el momento del trasplante
65	HTA	-	Hipertensión arterial
66	DISLIPEMIA	-	Dislipemia
67	TABACO	-	Paciente fumador
68	NºCIG	cig/dia	Nro de cigarrillos al día
69	ALCOHOL	-	Paciente alcoholico
70	AÑO_DIAG_DM	año	Año en el que se le diagnosticó la diabetes al paciente
71	AÑOS_DE_EVOLUCION_DE_LA_DM	años	Años desde que se le diagnostica la DMT1 hasta el trasplante
72	AÑO_DE_INICIO_DE_LA_DIALISIS	año	Año en el que se inició el tratamiento de diálisis
73	AÑOS_EN_DIALISIS	año	Años desde que se inició la diálisis hasta que se realizó el trasplante
74	TIPO_DE_DIALISIS	-	Tipo de diálisis a la que fue sometida el paciente
75	RETINOPATIA_DIABETICA_EN_OJO_DERECHO	-	Tipo retinopatía en el ojo derecho
76	RETINOPATIA_DIABETICA_EN_OJO_IZQUIERDO	-	Tipo retinopatía en el ojo izquierdo
77	TEST_DE_ESFUERZO	-	Resultado del test de esfuerzo
78	ECOCARDIOGRAFIA	-	Resultado de la ecocardiografía
79	ECG	-	Resultados del electrocardiograma
80	CATETERISMO	-	Técnica para ver las arterias coronarias del corazón
81	IAM	-	Infarto de miocardio
82	ACV	-	Ataque cerebrovascular (ictus)
83	CLAUDICACION_INTERMITENTE	-	Aparece cuando las arterias de las piernas están ocluidas.
84	RMN_VASCULAR	-	Resultados AngioResonancia
85	EMG	-	Resultados Electromiografía
86	URETEROGRAFIA	-	Resultados Ureterografía
87	EXITUS	-	Muerte del paciente
88	MESES_EXITUS_TRAS_TX	-	Meses hasta la muerte del paciente desde el trasplante
89	CAUSA_EXITUS	-	Causa de la muerte
90	TIPO_TRASPL	-	Tipo de trasplante realizado al paciente
91	PAK_TRAS_SPK	-	Pacientes que primero han tenido un trasplante de riñón y luego se les hace solo un trasplante de páncreas.
92	PAK_TIEMPO_TRAS_TX_RENAL	meses	Meses desde el trasplante de riñón hasta el trasplante de páncreas
93	FISTULA	-	Aparición de fístula
94	MESES_FISTULA_TRAS_TX	meses	Nro de meses entre el trasplante y la aparición de la fístula
95	EXPLANTE_DE_PANCREAS	-	Extracción del páncreas trasplantado por una complicación quirúrgica
96	MESES_EXPLANTEPANCRAS_TRAS_TX	meses	Meses transcurridos entre el trasplante y el explante de páncreas
97	EXPLANTE_RENAL	-	Extracción del riñón trasplantado
98	MESES_EXPLANTERENAL_TRAS_TX	meses	Meses transcurridos entre el trasplante y el explante de riñón
99	TROMBOSIS_PANCREAS	-	Se ha producido una tomobosis en el páncreas del paciente
100	TROMBOSIS_RENAL	-	Se ha producido una tomobosis en el riñón del paciente
101	COMPLICACIONES_QX	-	Complicaciones derivadas de la técnica quirúrgica
102	RECIDIVA_DM1	-	Reaparición de la diabetes tras un periodo de tiempo
103	RECHAZO_PANCREAS	-	El rechazo del injerto por el sistema inmune del receptor

Sigue en la página siguiente.

N°	Nombre	Unidad	Descripción
104	PERD.INJERTO	-	érdida del injerto del páncreas. Engloba pérdidas por complicaciones quirúrgicas, por rechazo del paciente o por reaparición de la diabetes
105	VHCReceptor	-	Aparición de Hepatitis C en el receptor
106	VHBReceptor	-	Aparición de Hepatitis B en el receptor
107	VIHReceptor	-	Aparición de VIH en el receptor
108	CMVReceptor	-	Aparición de citomegalovirus en el receptor
109	VEBReceptor	-	Virus de Epstein-Barr
110	Trasplantepreviorenal	-	Pacientes que antes han tenido un trasplante renal
111	Edad donante	años	Edad del donante
112	Causa fallecimiento donante	-	Causa de fallecimiento del donante
113	Sexo donante	-	Sexo del donante
114	VHCDonante	-	Hepatitis C
115	VHBDonante	-	Hepatitis B
116	VIHDonante	-	Virus de inmunodeficiencia humana
117	CMVDonante	-	Citomegalovirus
118	VEBDonante	-	Virus de Epstein-Barr
119	CV máxima CMV	-	Carga máxima del citomegalovirus
120	Seroconversión CMV	-	Seroconversión
121	Tipo de afectación CMV	-	Tipo de afectación citomegalovirus
122	INCOMPATIBILIDAD HLA	-	Número de HLA que tienen en común el donante del órgano con el receptor
123	Profilaxis ATB	-	Si han precisado administrar algún antibiótico antes del trasplante
124	Profilaxis cotrimoxazol	-	Al paciente se le administra Cotrimoxazol
125	Profilaxis CMV	-	El paciente toma medidas contra el citomegalovirus
126	Profilaxis antifúngica	-	El paciente toma medidas antifúngicas
127	Profilaxis trombosis	-	El paciente toma medidas contra la trombosis
128	Profilaxis trombosis tipo	-	Tipo de tratamiento para evitar la trombosis
129	TIPO_TTO_ACTUAL	-	Tipo de tratamiento actual contra la diabetes
130	NEC_INSULINA_TRAS_TX	-	El paciente necesita tratamiento de insulina tras el trasplante
131	Protocolo tacrolimus	-	Inmunodepresor
132	Cambio régimen Tacrolimus M	meses	Meses postrasplante a los que se hizo el cambio de medicación.
133	Retirada CTM	meses	Meses a los que se retiraron los corticoides
134	Continúa CT	-	El paciente continúa con suministro de corticoides

CUADRO B.1: Descripción de las variables disponibles en el estudio

Bibliografía

- [1] World Health Organization 2016. Informe mundial sobre la diabetes. Technical report, www.who.int/diabetes/global-report, .
- [2] American Diabetes Association et al. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes care*, 26(suppl 1):s5–s20, 2003.
- [3] American Diabetes Association et al. 2. classification and diagnosis of diabetes. *Diabetes Care*, 40(Supplement 1):S11–S24, 2017.
- [4] Ralph A DeFronzo, Eleuterio Ferrannini, Kurt George Matthew Mayer Alberti, Paul Zimmet, and George Alberti. *International Textbook of Diabetes Mellitus, 2 Volume Set*, volume 1. John Wiley & Sons, 2015.
- [5] David M Maahs, Nancy A West, Jean M Lawrence, and Elizabeth J Mayer-Davis. Epidemiology of type 1 diabetes. *Endocrinology and metabolism clinics of North America*, 39(3):481–497, 2010.
- [6] Miguel Ruiz-Ramos, Antonio Escolar-Pujolar, Eduardo Mayoral-Sánchez, Florentino Corral-San Laureano, and Isabel Fernández-Fernández. La diabetes mellitus en españa: mortalidad, prevalencia, incidencia, costes económicos y desigualdades. *Gaceta Sanitaria*, 20:15–24, 2006.
- [7] World Health Organization 2016. Perfiles de los países para la diabetes. Technical report, http://www.who.int/diabetes/country-profiles/esp_es.pdf?ua=1, .
- [8] Juan Oliva, Félix Lobo, Begoña Molina, and Susana Monereo. Direct health care costs of diabetic patients in spain. *Diabetes care*, 27(11):2616–2621, 2004.
- [9] American Diabetes Association et al. Nutrition recommendations and principles for people with diabetes mellitus. *Diabetes care*, 17:522, 1994.

- [10] Marion J Franz, John P Bantle, Christine A Beebe, John D Brunzell, Jean-Louis Chiasson, Abhimanyu Garg, Lea Ann Holzmeister, Byron Hoogwerf, Elizabeth Mayer-Davis, Arshag D Mooradian, et al. Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications. *Diabetes care*, 25(1):148–198, 2002.
- [11] Mark E Molitch, Michael W Steffes, Patricia A Cleary, and David M Nathan. Baseline analysis of renal function in the diabetes control and complications trial. *Kidney international*, 43(3):668–674, 1993.
- [12] FG Banting and CH Best. The internal secretion of the pancreas. *Translational Research*, 80(4):465–480, 1972.
- [13] Angelika C Gruessner and David ER Sutherland. Pancreas transplant outcomes for united states (us) and non-us cases as reported to the united network for organ sharing (unos) and the international pancreas transplant registry (iptr) as of june 2004. *Clinical transplantation*, 19(4):433–455, 2005.
- [14] Organización Nacional de Trasplantes 2015. Memoria anual del trasplante de páncreas. Technical report, http://www.who.int/diabetes/country-profiles/esp_es.pdf?ua=1.
- [15] Jeffrey M Venstrom, Maureen A McBride, Kristina I Rother, Boaz Hirshberg, Trevor J Orchard, and David M Harlan. Survival after pancreas transplantation in patients with diabetes and preserved kidney function. *Jama*, 290(21):2817–2823, 2003.
- [16] Rainer WG Gruessner and AC Gruessner. Pancreas after islet transplantation: A first report of the international pancreas transplant registry. *American Journal of Transplantation*, 16(2):688–693, 2016.
- [17] Gary S Collins, Johannes B Reitsma, Douglas G Altman, and Karel GM Moons. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): the tripod statement. *BMC medicine*, 13(1):55, 2015.
- [18] Ralph B D’Agostino, Ramachandran S Vasan, Michael J Pencina, Philip A Wolf, Mark Cobain, Joseph M Massaro, and William B Kannel. General cardiovascular risk profile for use in primary care: The framingham heart study. *Circulation*, 117(6):743–753, 2008.

-
- [19] Paromita King, Ian Peacock, and Richard Donnelly. The uk prospective diabetes study (ukpds): clinical and therapeutic implications for type 2 diabetes. *British journal of clinical pharmacology*, 48(5):643–648, 1999.
- [20] David M Nathan, Howard Fogel, Dennis Norman, Paul S Russell, Nina Tolloff-Rubin, Francis L Delmonico, Hugh Auchincloss Jr, Janice Camuso, and A Benedict Cosimi. Long-term metabolic and quality of life results with pancreatic/renal transplantation in insulin-dependent diabetes mellitus. *Transplantation*, 52(1):85–90, 1991.
- [21] Erik B Finger, David M Radosевич, Ty B Dunn, Srinath Chinnakotla, David ER Sutherland, Arthur J Matas, Timothy L Pruett, and Raja Kandaswamy. A composite risk model for predicting technical failure in pancreas transplantation. *American Journal of Transplantation*, 13(7):1840–1849, 2013.
- [22] Manuel Arrianza Balmón. Guía práctica de análisis de datos. 2006.
- [23] Peter Hall, Byeong U Park, and Richard J Samworth. Choice of neighbor order in nearest-neighbor classification. *The Annals of Statistics*, pages 2135–2152, 2008.
- [24] Antonio Martinez-Millana, Carlos Fernández-Llatas, Lucia Sacchi, Daniele Segagni, Sergio Guillén, Riccardo Bellazzi, and Vicente Traver. From data to the decision: A software architecture to integrate predictive modelling in clinical settings. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 8161–8164. IEEE, 2015.