

UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR



CALIBRACIÓN DE PUNTUACIONES PROCEDENTES DE SISTEMAS BIOMÉTRICOS

-PROYECTO FIN DE CARRERA-

Sandra Uceda Ontivero

Julio 2014

CALIBRACIÓN DE PUNTUACIONES PROCEDENTES DE SISTEMAS BIOMÉTRICOS

AUTOR: Sandra Uceda Ontivero

TUTOR: Daniel Ramos Castro



ATVS Grupo de Reconocimiento Biométrico
(<http://atvs.ii.uam.es>)
Dpto. de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid

Agradecimientos

Y aunque sea de lo primero que aparece, en realidad los agradecimientos suponen la guinda final del pastel. Tras seis años de duro trabajo y esfuerzo la carrera llega a su fin.

Para comenzar, gracias a Javier Ortega por ser quien me introdujo en el mundo de la biometría, confiar en mí, para darme la oportunidad de participar en numerosas actividades que enriquecieron mi formación en la universidad. Por consiguiente, haber podido formar parte del grupo de investigación ATVS, y en consecuencia realizar el proyecto con ellos, por lo que gracias a mi tutor, Daniel Ramos, que ha podido enseñarme particularmente el mundo de la huella dactilar, ha sido un gran descubrimiento para mí. Desde otro punto de vista, gracias a todos los miembros del departamento de lofoscopia de la Guardia Civil, en especial al teniente Herrero, que me enseñó la parte más técnica de la dactiloscopia forense.

Durante estos tres años gracias a todos los compañeros con los que he compartido clases, prácticas, trabajos, exámenes, academia, y finalmente buenos ratos. De todos he podido aprender experiencias, algunas buenas y otras no tan buenas. Pero siempre me acordaré de los 10 ‘toligos’ que acabamos en Berlín (+Eva), Berta, María, Eva, Rodri, Richi, Fátima, Sara, Jaime y Helia.

Por supuesto, muchas gracias a mi familia, a mi madre y a mi hermana, que son las que más me conocen y me apoyan incondicionalmente, aportándome el equilibrio para levantarme y para bajarme de la nube también, ¡¡GRACIAS chicas!!

A mis abuelos que tan orgullosos se sienten de mí y yo me siento de ellos. GRACIAS

Y a mi compañero de experiencias, Adrián, no sé qué hubiera pasado si el primer año de carrera no hubieras sabido estar diciéndome que luce y espere, que lo bueno llegará, y tenías razón, ya ha llegado, está llegando, y en verdad he podido con todo y más!! Así que gracias por hacérmelo ver, por haber sufrido conmigo y por ayudar a que esto sea posible.

Dedicado a todos.

Sandra

Julio 2014

Resumen

El proyecto está basado en el cálculo de relaciones de verosimilitud para poder dar un peso o un grado de apoyo en la decisión final del juez. Se persigue evitar afirmaciones categóricas para no caer en errores de declarar culpable a un inocente o viceversa. Para lograrlo se estudiará el rendimiento obtenido de los LR calculados mediante tres modelos distintos: Gaussiano, regresión logística y KDF. De esta manera, el perito aportará un grado de apoyo basado en las evidencias científicas en vez de tomar una decisión entre dos hipótesis posibles, culpable o inocente.

Para la realización de este proyecto se han utilizado muestras procedentes de huellas dactilares reales cedidas por el departamento de identificación de la Guardia Civil. Con ellas se ha realizado una base de datos, en sus dos formatos, la huella de individuo conocido junto a su huella coincidente extraída de una escena del crimen. Finalmente en la base de datos se tienen muestras de 248 individuos distintos. Para cada individuo se pueden tener de una a cuatro huellas distintas coincidiendo cada muestra con un dedo distinto de cada individuo. Por lo que en total se disponen de 258 pares de huellas para la realización de las pruebas posteriores. Como extensión a esta base de datos se continuó con huellas palmares, con unas características parecidas de las que se reunieron muestras para un total de 14 individuos distintos, con un total de 20 pares de huellas palmares.

Esta base de datos ha permitido realizar pruebas de entrenamiento con 10.000 combinaciones diferentes de pares de huellas y pruebas de test con 24.964 combinaciones diferentes de pares de huellas. Dividiendo, para ello, las muestras, en primer lugar entre entrenamiento y test y en segundo lugar realizando dos tipos de combinaciones, genuinas, cuando pertenecen a la misma fuente e impostoras, cuando pertenecen a fuentes distintas.

Con las puntuaciones (scores) obtenidas de todas las comparaciones realizadas se van a aplicar tres métodos de cálculo de relaciones de verosimilitud (LR, Likelihood Ratios). Estas nuevas puntuaciones serán interpretables por sí mismas y nos permitirán aportar un grado de fiabilidad al par de huellas que compara el sistema que estemos usando, en este caso supone una metodología aceptada por las necesidades de la ciencia forense. Con él se podrá dar un grado de apoyo, a partir de las evidencias en un caso judicial. En este proyecto no se pretende poder tomar una decisión final de entre las dos hipótesis que se barajan: culpable o inocente.

Así mismo, se mostrará un estudio y comparación del rendimiento obtenido por los métodos propuestos. Para evaluar los sistemas se estudia su rendimiento mediante curvas de entropía cruzada (ECE) y se proponen también previo al cálculo de LR tres modelos de transformación de los scores para una correcta normalización del modelo.

Palabras Clave

Biometría, reconocimiento de huella dactilar, sistemas biométricos, ciencias forenses, relaciones de verosimilitud.

Abstract

The project is based on the calculation of likelihood ratios to give a weight or degree of support in the final decision of the judge. It seeks to avoid categorical statements to avoid falling into errors of convicting an innocent or vice versa. To achieve it will study the performance of the LR calculated using three different models: Gaussian, logistic regression and KDF. Thus, the expert will provide a degree of support based on evidence rather than make a decision between two possible, guilty or innocent hypothesis.

For making this project we have used samples from actual fingerprints provided by the department identifying of Guardia Civil. With them was made a database in two formats, individual known footprint with its matching fingerprint taken from a crime scene. Finally in the database 248 will have different individual samples. For each individual can have one to four different fingerprints coincide each sample with a different finger to each individual. So that in total they have 258 pairs of traces for conducting further tests. As an extension to this database continued with palm prints, with similar characteristics of the samples for a total of 14 different individuals were combined, a total of 20 pairs of palm prints.

This database has enabled 10,000 training trials with different combinations of pairs of traces and test trials with 24,964 different combinations of pairs of fingerprints. Dividing to do, specimens, first between training and test and secondly performing two types of combinations, genuine, when they belong to the same source and impostors when they belong to different sources.

With scores (scores) obtained from all comparisons are to be applied three methods of calculating likelihood ratios (LR, Likelihood Ratios). These new scores will be interpretable by themselves and allow us to provide a degree of assurance to the pair of tracks that compares the system you are using, in this case is a methodology accepted by the needs of forensic science. With it you can take a degree of support from the evidence in a court case. This project is not intended to make a final decision between the two hypotheses being considered: guilty or innocent.

Likewise, a study and comparison of the performance achieved by the proposed methods is shown. To evaluate the system performance it is studied using cross-entropy curves (ECE) and also propose LR prior to the calculation of three models of transformation of the scores for correct normalization model.

Keywords

Biometrics, Fingerprint recognition, biometrics systems, forensics Science, likelihood ratios.

Glosario de Acrónimos

AFIS: Automated Fingerprint Recognition system.

ECE: Empirical Cross Entropy.

FA: Falsa Aceptación.

FR: Falso Rechazo.

KDF: Kernel Density Function.

LR: Likelihood Ratio.

PAV: Pool Adjacent Violators.

DGGC: Dirección General de la Guardia Civil.

INDICE DE CONTENIDOS

1. Introducción	1
1.1 Motivación	4
1.2 Objetivos	5
1.3 Metodología y plan de trabajo.	6
2. Estado del arte	9
2.1 Biometría	9
2.1.1 Características de los rasgos biométricos	9
2.1.2 Clasificación de los rasgos biométricos.....	10
2.1.3 Biometría en ámbito forense	11
2.2 Huella Dactilar	13
2.2.1 Formación de la huella dactilar	13
2.2.2 Reconocimiento dactilar en ámbito forense.....	15
2.2.3 Identificación dactilar en ámbito forense	17
2.3 Sistemas Biométricos	19
2.3.1 Tipos de sistemas biométricos	19
2.3.2 Sistemas forenses de identificación dactilar.....	21
2.3.3 Extracción de características de los sistemas forenses dactilares	22
3. Creación de la Base de Datos: Proyecto de huella con la Dirección General de la Guardia Civil	24
3.1 Base de Datos Lofoscópicos del ATVS	25
3.1.1 Pre-procesamiento de imagen.....	26
3.1.2 Marcado de minucias.....	28
3.1.3 Guardado de puntos	29
3.1.4 Resultado	30
3.1.5 Base de datos de huellas palmares	31
3.2 Herramientas gráficas de adquisición	32
3.2.1 MINUCIA.....	32
3.2.2 LR-LOFOSCOPIA.....	34
4. Teoría de Decisión Bayesiana	37
4.1 Errores contemplados	40
4.2 Evaluación mediante Likelihood Ratio, LR	40
4.2.1 SCORE VS LR	41
4.2.2 Probabilidad basada en Posibilidades.....	42
4.2.3 Precisión del LR: Calibración y Poder de discriminación.....	42
4.3 Normalización de los scores	44
5. Métodos de Cálculo de LR	47
5.1 Modelo Gaussiano	48
5.2 KDF	49
5.3 Regresión Logística	50
5.4 Evaluación del rendimiento de LR	54

5.4.1 ECE PLOT. Empirical Cross-Entropy	54
5.4.2 NCE. Normalized Cross-Entropy.....	58
5.4.3 Regla del 30 de doddington	58
5.5 Criterios de decisión.....	59
6. Resultados Experimentales	62
6.1 Primera parte de los experimentos:.....	64
Implementación de los modelos estudiados.....	64
6.1.1 Modelo Gaussiano	64
6.1.2 KDF	66
6.1.3 Regresión Logística.....	68
6.1.4 Conclusión	69
6.2 Segunda parte de los experimentos:.....	70
Normalización de los scores	70
6.2.1 Modelo Gaussiano con transformación previa	75
6.2.2 Regresión logística con transformación previa	77
.....	77
6.2.3 KDF con transformación previa.....	78
6.3 Tercera parte de los experimentos:	84
Realización de los experimentos a partir de bases de datos más pequeñas	84
6.3.1 Primer Criterio: ROBUSTEZ.....	89
6.3.2 Segundo Criterio: CALIBRACIÓN	90
6.3.3 Tercer Criterio: CLLR.....	91
6.4 Cuarta parte de los experimentos:.....	93
Resolución del análisis del estudio de los modelos	93
6.4.1 NCE: Normalized Cross-Entropy.....	94
6.4.2 Regla del 30 de Doddington	97
7. Conclusiones y trabajo futuro	100
7.1 Conclusiones.....	100
7.2 Trabajo Futuro	103
8. Bibliografía.....	104
9. PRESUPUESTO	106
10. PLIEGO DE CONDICIONES	107

INDICE DE TABLAS

TABLA 3-1 ESQUEMA DEL FUNCIONAMIENTO DE MINUCIA	33
TABLA 3-2 ESQUEMA DEL FUNCIONAMIENTO DE MINUCIA	34
TABLA 3-3 ESQUEMA DEL FUNCIONAMIENTO DE LR-LOFOSCOPIA EN MODO MARCADO DE MINUCIAS	35
TABLA 5-1 OBJETIVO A RELLENAR: CRITERIO PARA LA SELECCIÓN DE LOS MODELOS	61
TABLA 6-1 DATOS DE ROBUSTEZ.....	89
TABLA 6-2 VALORES DE CLLR SEGÚN LA CALIBRACIÓN	90
TABLA 6-3 VALORES DE CLLR	91
TABLA 6-4 MODELOS ESCOGIDOS PARA CADA CRITERIO	93

ÍNDICE DE FIGURAS

FIGURA 1-1 HUELLAS EXTRAÍDAS DEL CASO MAYFIELD, DE IZQUIERDA A DERECHA, MAYFIELD, ENCONTRADA EN LA BOLSA, DAOUD. [3]	2
FIGURA 1-2 MUESTRAS DE LOS PARES DE HUELLAS COINCIDENTES. [15]	3
FIGURA 1-3 PROCESO LOGÍSTICO QUE SIGUE EL PROYECTO FIN DE CARRERA.	5
FIGURA 2-1 EJEMPLOS DE RASGOS BIOMÉTRICOS FISIOLÓGICOS	10
FIGURA 2-2 EJEMPLO DE RASGOS BIOMÉTRICOS CONDUCTUALES	11
FIGURA 2-3 HUELLA NATURAL, LATENTE E IMPRESIÓN DIGITAL RESPECTIVAMENTE. [16] [17] [18]	14
FIGURA 2-4 TIPOS DE HUELLA SEGÚN EL PRIMER NIVEL. [19]	15
FIGURA 2-5 TIPOS DE MINUCIAS	16
FIGURA 2-6 EXTRACCIÓN DE CARACTERÍSTICAS DACTILAR. [20]	18
FIGURA 2-7 ESQUEMA DEL PROCESO DE REGISTRO	20
FIGURA 2-8 ESQUEMA DEL PROCESO DE IDENTIFICACIÓN	20
FIGURA 2-9 ESQUEMA DEL PROCESO DE VERIFICACIÓN	21
FIGURA 2-10 EXTRACCIÓN DE CARACTERÍSTICAS	22
FIGURA 3-1 COTEJO LOFOSCÓPICO	26
FIGURA 3-2 PLANTILLA DE DATOS	26
FIGURA 3-3 HUELLA CON TESTIGO MÉTRICO. [17]	27
FIGURA 3-4 ESQUEMA DEL CÓDIGO DE IDENTIFICACIÓN DE MINUCIAS	29
FIGURA 3-5 EJEMPLO DE FICHERO TXT. PLANTILLA DE LA HUELLA. CUANDO LA MINUCIA NECESITA MÁS DE UN PUNTO PARA IDENTIFICARLA SE ESCRIBE UNA NUEVA LÍNEA PARA CADA PUNTO, COMO AQUÍ OCURRE CON LA NÚMERO 11.	30
FIGURA 3-6 HUELLA PALMAR. [21]	31
FIGURA 3-7 INTERFAZ GRÁFICA DE MINUCIA	33
FIGURA 3-8 INTERFAZ GRÁFICA DE LR-LOFOSCOPIA	35
FIGURA 3-9 FUNCIÓN PARA CALCULAR LR	36
FIGURA 3-10 FUNCIÓN 'NUEVA HUELLA'	36
FIGURA 4-1 EJEMPLO DE CURVA DET	43
FIGURA 4-2 EJEMPLO DEL RENDIMIENTO DE UN SISTEMA MEDIANTE UNA PROBETA	44
FIGURA 4-3 SCORES NON-TARGET SIN NORMALIZAR	45
FIGURA 4-4 SCORES TARGET SIN NORMALIZAR	45
FIGURA 4-5 DISTRIBUCIÓN DE LOS SCORE TRAS APLICAR LAS TRANSFORMACIONES 1, 2 Y 3	46
FIGURA 5-1 PODER DE DISCRIMINACIÓN, MODELO GAUSSIANO	48
FIGURA 5-2 REPRESENTACIÓN SCORE-LOG (LR)	48
FIGURA 5-3 MODELO BASADO EN KDF [22]	49
FIGURA 5-4 REPRESENTACIÓN LOG LR-SCORE	50
FIGURA 5-5 FUNCIÓN SIGMOIDEA	51
FIGURA 5-6 REPRESENTACIÓN LOG-LR-SCORE	53
FIGURA 5-7 ECE NEUTRAL	55
FIGURA 5-8 REPRESENTACIÓN DE LAS CURVAS DE ENTROPÍA ECE	57
FIGURA 5-9 REPRESENTACIÓN DEL INTERVALO DE CONFIANZA BASADO EN LA REGLA DEL 30 DE DODDINGTON. SOBRE LA DISTRIBUCIÓN DE LOS LR OBTENIDOS SE CALCULA EL INTERVALO DE CONFIANZA.	59
FIGURA 5-10 OBJETIVOS PARA EL CRITERIO DE DECISIÓN	60
FIGURA 6-1 DISTRIBUCIÓN DE LOS SCORE TARGET (IZQUIERDA) Y NON-TARGET (DERECHA) RESPECTIVAMENTE.	64
FIGURA 6-2 ECE PLOT EN MODELO GAUSSIANO SIN TRANSFORMACIÓN	65
FIGURA 6-3 DISTRIBUCIÓN DE LOS SCORES CON EL MODELO KDF	66

FIGURA 6-4 ECE PLOT SCORE SIN TRANSFORMACIÓN MODELO KDF	66
FIGURA 6-5 TRANSFORMACIÓN SCORE-LOG (LR)	66
FIGURA 6-6 DISTRIBUCIÓN DE LOS SCORE SIN TRANSFORMACIÓN MODELO REGRESIÓN LOGÍSTICA	68
FIGURA 6-7 TRANSFORMACIÓN SCORE-LOG (LR) CON REGRESIÓN LOGÍSTICA	68
FIGURA 6-8 ECE PLOT SCORE SIN TRANSFORMACIÓN CON REGRESIÓN LOGÍSTICA	69
FIGURA 6-9 DISTRIBUCIÓN DE LOS SCORE NON-TARGET CON TRANSFORMACIÓN 1	70
FIGURA 6-10 DISTRIBUCIÓN SCORE TARGET CON TRANSFORMACIÓN 1	70
FIGURA 6-11 PODER DE DISCRIMINACIÓN DE LOS SCORE CON TRANSFORMACIÓN 1	71
FIGURA 6-12 TRANSFORMACIÓN SCORE-LOG (LR) GAUSSIANO	71
FIGURA 6-13 ECE PLOT SCORE CON TRANSFORMACIÓN 1 GAUSSIANO	72
FIGURA 6-14 DISTRIBUCIÓN DE LOS SCORE CON LAS TRES TRANSFORMACIONES	73
FIGURA 6-15 PODER DE DISCRIMINACIÓN Y ECE PLOT CON LA T1 EN PRIMERA FILA, T2 EN LA SEGUNDA Y T3 EN LA TERCERA FILA RESPECTIVAMENTE EN GAUSSIANO	75
FIGURA 6-16 PODER DE DISCRIMINACIÓN Y ECE PLOT CON T1, T2 Y T3 RESPECTIVAMENTE CON REGRESIÓN LOGÍSTICA	77
FIGURA 6-17 DISTRIBUCIÓN DE LOS SCORE TARGET Y NON-TARGET	79
FIGURA 6-18 TRANSFORMACIÓN DE SCORE A LR CON TRANSFORMACIONES. ECE PLOT CON TRANSFORMACIONES EN KDF	80
FIGURA 6-19 COMPORTAMIENTO DE LA TRANSFORMACIÓN 1	81
FIGURA 6-20 COMPORTAMIENTO DE LA TRANSFORMACIÓN 2	82
FIGURA 6-21 COMPORTAMIENTO DE LA TRANSFORMACIÓN 3	82
FIGURA 6-22 ECE PLOT PARA LOS 9 MODELOS CON UN 10% MENOS DE EVIDENCIAS	85
FIGURA 6-23 ECE PLOT PARA LOS 9 MODELOS CON UN 20 % MENOS DE EVIDENCIAS	86
FIGURA 6-24 ECE PLOT PARA LOS 9 MODELOS CON UN 50% MENOS DE EVIDENCIAS	87
FIGURA 6-25 REPRESENTACIÓN DE LAS	88
FIGURA 6-26 GENERACIÓN DE 100 MUESTRAS PARA CADA BASE DE DATOS CON UN 50 % MENOS DE EVIDENCIAS	92
FIGURA 6-27 NCE PARA KDF T2 SIN 10%	94
FIGURA 6-28 NCE PARA KDF T1 SIN 10%	94
FIGURA 6-29 NCE PARA LOS MEJORES MODELOS CON UN 50% MENOS DE EVIDENCIAS	95
FIGURA 6-30 NCE PARA LOS MEJORES MODELOS CON UN 50% MENOS DE EVIDENCIAS	96
FIGURA 6-31 ECE PLOT KDF T2	97
FIGURA 6-32 ECE PLOT KDF T1	97
FIGURA 6-33 ECE PLOT REG LOG T2	98
FIGURA 6-34 ECE PLOT GAUSS T3	98
FIGURA 6-35 ECE PLOT KDF T2	98
FIGURA 7-1 REPRESENTACIÓN DEL RENDIMIENTO DE LOS 9 MODELOS PARA UNA BASE DE DATOS PEQUEÑA Y 100 IMPLEMENTACIONES DISTINTAS	102

1. Introducción

Parece una necesidad básica entre los seres humanos la tarea de identificación. En todo momento de nuestro día desde que amanecemos hasta que anochece el cerebro humano realiza procesos de identificación de lo más simples e intuitivos, la mayoría de ellos se realizan de forma inconsciente.

Sin embargo, cuando a la identificación se le suma la seguridad, el proceso no resulta tan sencillo. Un ser humano puede identificarse bien porque sepa algo (una contraseña), porque contenga algo (una llave) y por último porque sea algo (biometría). De todos existe la forma de falsificarlos, trucarlos u obtenerlos de algún modo, sin embargo, el único del que uno no se puede desprender es el tercero.

En este punto surge la biometría, características innatas al ser humano, inherentes a él o incluso aprendidas, como la firma. Concretamente la huella dactilar presenta una estructura que hasta el momento es altamente discriminante [1]. Además, con el paso de los años esta estructura no cambia, tiende a regenerarse adoptando la misma forma, excepto en el caso de lesiones graves, como una cicatriz. Para mayor seguridad en las tareas de identificación existen técnicas que permiten comprobar si una huella proviene de un individuo con vida.

Con el paso del tiempo se ha avanzado mucho en el ámbito biométrico y son muchos los sistemas que han incorporado estas técnicas. A pesar de ello la huella dactilar se sitúa entre las primeras posiciones de porcentaje de uso debido a su gran tasa de fiabilidad y de aceptación por parte del usuario [2].

Los sistemas de adquisición de huella dactilar obtienen imágenes muy nítidas de la estructura de las crestas y los surcos, por lo que la comparación de huellas para la identificación del usuario resulta muy discriminativa. Sin embargo, el problema surge cuando la adquisición de las huellas se produce en entornos no tan favorables, como ocurre en los casos de ámbito forense; situaciones que surgen de una investigación criminalística obtenidas de escenas forenses reales.

El problema más aparente es la falta de calidad de las imágenes de las huellas que se obtienen. Puede que se tenga solo un fragmento de la huella, que el material donde está genere mucho ruido en la imagen o que los materiales químicos que ayudan a resaltar las crestas interrumpen la claridad visual de éstas.

A pesar de que este rasgo sea muy discriminativo también se cometen errores, un error muy mediático tuvo lugar durante los atentados terroristas de Madrid, el 11 de Marzo de 2004. Cuando los expertos del FBI acusaron al abogado de Oregón, Brandon Mayfield, alegando haber encontrado su huella dactilar en una bolsa relacionada con los atentados. Tras unas semanas, las autoridades españolas verificaron que la misma huella pertenecía a Daoud Ouhnane, de origen argelino. Tres agentes del FBI habían examinado la huella cometiendo el mismo error: el análisis dactilar que había llevado al arresto de Mayfield era erróneo [3]. En la imagen, a la izquierda la huella de Mayfield, a la derecha la de Daoud y en medio la encontrada en la bolsa.



Figura 1-1 Huellas extraídas del caso Mayfield, de izquierda a derecha, Mayfield, encontrada en la bolsa, Daoud. [3]

Tras este tipo de acontecimientos surgió la necesidad [14] de realizar exhaustivos análisis con pruebas empíricas sobre la verificación de comparaciones de huellas. Particularizando a este caso, resultó que la probabilidad de que la huella perteneciera al abogado era alta, pero no se tuvo suficientemente en cuenta que no existiese otra huella que pudiera parecerse a la encontrada en el caso.

Por ello, en este proyecto se persigue poner solución a esta suposición de afirmar que no existan dos huellas lo suficientemente parecidas como para dar lugar a error. Teniendo en cuenta que el sistema de identificación español actual declara que dos huellas son coincidentes si se localizan las mismas 12 minucias, se pueden cometer errores al basarse en un umbral subjetivo. ¿Por qué doce minucias? Si se hubieran establecido 15 minucias, o 7, ¿esas huellas también serían coincidentes? Es importante evitar tomar este tipo de decisiones basadas en la elección de un umbral que no está justificado empíricamente.

Por este motivo se pretende reducir la gravedad de un error humano, no realizando únicamente procesos de aceptación o exclusión sino de dar un peso que refleje el grado de apoyo que presentan las pruebas. Si además de la información recogida en un caso se aporta un grado de apoyo que esté basado en las evidencias científicas, la decisión final, que debe tomarla el juez, estará basada de forma conjunta a todas las pruebas recogidas.

Para llevar a cabo este proceso, en primer lugar se adquirió una base de datos de huellas reales obtenidas de escenas del crimen. Para ello se colaboró desde un equipo del ATVS con los peritos lofoscópicos de la Dirección General de la Guardia Civil y se buscó reproducir el mismo procedimiento que realizan para extraer los datos necesarios de las huellas, así los resultados que se obtengan en este proyecto final de carrera no estarán basados en casos hipotéticos sino que la fuente de donde se obtienen es real. Esto es positivo en dos aspectos:

- Poder colaborar con las autoridades para intentar mejorar sus tareas siguiendo los protocolos que habitualmente realizan.
- Extraer conclusiones de unos resultados que provienen de fuentes reales, y por tanto simulan casos forenses reales.

Gracias a esta base de datos se almacena un par de huellas por cada individuo, por un lado la huella conocida, que pertenece a un individuo identificado, y por otro lado la huella latente, aquella que se ha extraído de una escena del crimen y que ha sido previamente identificada por los peritos. Es decir, se conoce de antemano a quien pertenece cada huella extraída de una escena de un crimen. Esta clasificación se realiza para establecer dos tipos de comparaciones:



Figura 1-2 Muestras de los pares de huellas coincidentes. [15]

-Entre huellas que pertenecen al mismo individuo (la identificada con la extraída del crimen), dan lugar a comparaciones entre **huellas genuinas**.

-Entre huellas que pertenecen a individuos distintos (la identificada con la extraída del crimen), dando lugar a comparaciones entre **huellas impostoras**.

Conociendo previamente de qué tipo es cada comparación, nos permite extraer puntuaciones de cada una de las dos hipótesis que se barajan; que pertenezcan al mismo individuo o que pertenezcan a individuos distintos. Estas puntuaciones (scores) serán sometidas a estudios probabilísticos [4] con el fin de obtener una puntuación útil por sí misma (LR, Likelihood Ratios) con la que poder valorar el peso de la prueba en el caso.

Para realizar este cálculo se estudiarán y probarán tres métodos, modelado gaussiano, regresión logística y KDF (Kernel Density Function). Además se proponen 3 modos de normalización de las puntuaciones extraídas de las huellas. El proyecto también incluye todos los detalles que se han seguido para la implementación de estas técnicas, así como los resultados obtenidos tras ellas.

1.1 Motivación

El ámbito en que se va a mover este proyecto es el forense, las ciencias forenses se entienden como la aplicación de las técnicas científicas a investigaciones relacionadas con actividades criminales para bien estimar la existencia de un crimen o bien identificar al autor del mismo. Es mediante la biometría con lo que se pretende mejorar las técnicas de identificación [5].

Es el papel del perito forense evaluar las pruebas y exponerlas ante el juez, sin embargo, el perito no es quien debe tomar una decisión categórica, sino que debe poder evaluar de forma cuantitativa la validez de las pruebas. Es el juez quien, teniendo en cuenta todas las pruebas del caso, debe tomar decisiones basándose en la información disponible. El principal objetivo de obtener puntuaciones probabilísticas de las evidencias es de proporcionar una medida que permita justificar la validez de las pruebas y con qué seguridad éstas son fiables. De manera que se justifique qué grado de apoyo se presenta ante el juez tras analizar las evidencias.

Teniendo en cuenta el ámbito del que se trata, se debe tener en cuenta que la importancia de cometer errores es muy elevada, ya que solo están implicadas dos categorías (inocente, culpable) [6]. Como errores entendemos dos tipos, afirmar que dos huellas pertenecen a la misma fuente cuando en realidad no (declarar culpable a un inocente, falsa aceptación), y por el contrario afirmar que dos huellas no pertenecen a la misma fuente cuando en realidad si (declarar inocente a un culpable, falso rechazo). En ambos casos la responsabilidad de este proceso es muy alta y por ello hay que tratarlo con el cuidado que se merece.

Para no caer en la tentación de cometer dichos errores la alternativa que se propone es proporcionar un valor probabilístico de apoyo a las hipótesis planteadas, con el fin de evaluar la importancia que toma la evidencia en el caso judicial.

Como para todo estudio se necesitan unos datos contrastados, en este proyecto se recurre a la base de datos creada de huellas dactilares en un proyecto con el Departamento de Lofoscopia de la Guardia Civil. Las huellas que se tienen son las obtenidas por el sistema AFIS (Sistemas de Identificación de Huella Dactilar) de la Guardia Civil y las obtenidas de casos forenses reales. Gracias a estas últimas y a las dificultades que presentan en su tratamiento se puede afirmar que los datos con los que se van a trabajar presentan una estrecha relación con la realidad forense.

1.2 Objetivos

Este proyecto está enfocado en el ámbito forense y se va a realizar en colaboración con el departamento de Identificación del Servicio de Criminalística de la Guardia Civil. Se describe mediante un procedimiento global de principio a fin que trate de proporcionar a los sistemas forenses actuales, en primer lugar la agilización en el proceso de adquisición de las evidencias, y en segundo lugar, una mejora en la toma de decisiones sobre la hipótesis correcta.

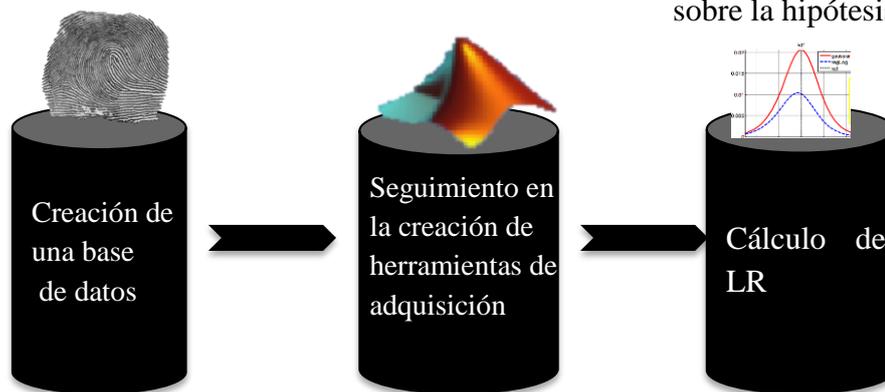


Figura 1-3 Proceso logístico que sigue el proyecto fin de carrera.

Así se detalla a continuación cuales van a ser los pasos a seguir para la realización del proyecto:

1. Participación en la creación de una base de datos de huellas procedentes de casos forenses reales y sus huellas coincidentes del sistema AFIS. Se va a realizar con la colaboración de la Dirección General de la Guardia Civil quienes proporcionan las evidencias a utilizar.
2. Colaboración con la creación de dos herramientas visuales diseñadas una para imitar el proceso de adquisición de los peritos y la otra para realizar el cálculo de LR. De forma paralela se realizará la introducción de mejoras a dicha herramienta mediante continuas pruebas al sistema.
3. A partir de las puntuaciones obtenidas de los pares de huellas genuinos e impostores se realizará la transformación a LR mediante 3 modelos distintos:
 - a. Modelo Gaussiano
 - b. Regresión Logística
 - c. KDF

Previamente será necesario realizar un estudio de las mejores transformaciones a las puntuaciones para la normalización de los datos. Así, se buscará la que mejor defina un comportamiento adecuado basado en la teoría de la decisión bayesiana.

1.3 Metodología y plan de trabajo.

1. **Documentación previa:** Durante los primeros meses se va a realizar un estudio de la biometría principalmente, y de la huella dactilar particularmente. Posteriormente se seguirá con los métodos probabilísticos de cálculo de LR. Además, se completará el estudio con el análisis de distintos métodos para evaluar el rendimiento de un sistema y sobre la medida de rendimiento de dichos métodos.
2. **Desarrollo:** Cuando los conceptos estudiados estuvieran claros se comenzará la implementación de los distintos modelos de cálculo de LR y la realización de pruebas que consoliden las interpretaciones.
 - a. Construcción de una base de datos de huellas.
 - b. Evaluar y testear el funcionamiento de las herramientas de adquisición.
 - c. Estudiando el formato en que se presentan los datos será necesario la realización de pruebas para escoger el mejor método de normalización de éstos.

- d. Se implementarán tres métodos de cálculo de LR. El objetivo final de los experimentos es evaluar según distintos criterios qué métodos se comportan mejor para los distintos sistemas propuestos.
3. **Extracción de conclusiones:** Evaluar los resultados obtenidos estudiando la validez de éstos en los entornos forenses. Para ello, se hará uso de la interpretación de las gráficas y sus distintas formas de presentación, además se propone un análisis de decisión para estudiar cuál es el modelo que mejor funciona según el escenario en el que se encuentre. Según los criterios elegidos para la decisión se valorarán cada una de las alternativas algorítmicas que se hayan seleccionado para el estudio, en este caso los diferentes escenarios propuestos.
4. **Memoria:** La creación de la memoria se irá completando durante el proceso de desarrollo, en los meses finales se terminará elaborando al completo, en ella se incluyen los siguientes capítulos:

CAPÍTULO 1: Introducción a la temática del proyecto, la biometría, enfocando los objetivos y la motivación que ha llevado a la realización de este proyecto.

CAPÍTULO 2: Estudio del arte de la biometría, concretamente de la huella dactilar. Estudio de sus particularidades, las maneras de representación y reconocimiento y su enfoque en el mundo forense. Se explican los sistemas utilizados, tanto en biometría como en ciencia forense.

CAPÍTULO 3: Desarrollo de la primera parte del proyecto, en el que se colaboró con la Guardia Civil en la creación de una base de datos de huellas dactilares y la obtención de características en formato electrónico. Tutorial del funcionamiento de las herramientas empleadas para su obtención.

CAPÍTULO 4: Análisis de la teoría de decisión bayesiana empleada para la interpretación de los resultados. Explicación de todos los conceptos empleados en el proceso, necesarios para el entendimiento del cálculo de relaciones de verosimilitud.

CAPÍTULO 5: Exposición de los métodos de cálculo de LR empleados en el proyecto. Técnicas utilizadas para evaluar el rendimiento de los sistemas, basadas en entropía y criterios que se han seguido para el análisis de decisión.

CAPÍTULO 6: Desarrollo experimental de la aplicación de los modelos estudiados a las puntuaciones obtenidas de la base de datos creada.

CAPÍTULO 7: extracción de conclusiones tras finalizar el proyecto, porqué es importante y que se puede deducir de los experimentos obtenidos. Para finalizar se proponen los trabajos futuros que podrían surgir tras este proyecto.

En este esquema se detalla el procedimiento que se va a seguir para la elaboración del proyecto. Durante el primer año se realizará la adquisición de la base de datos mediante el uso de las herramientas de adquisición creadas por el ATVS. Posteriormente y con la base de datos completada se realizaran los siguientes pasos que se han explicado anteriormente.

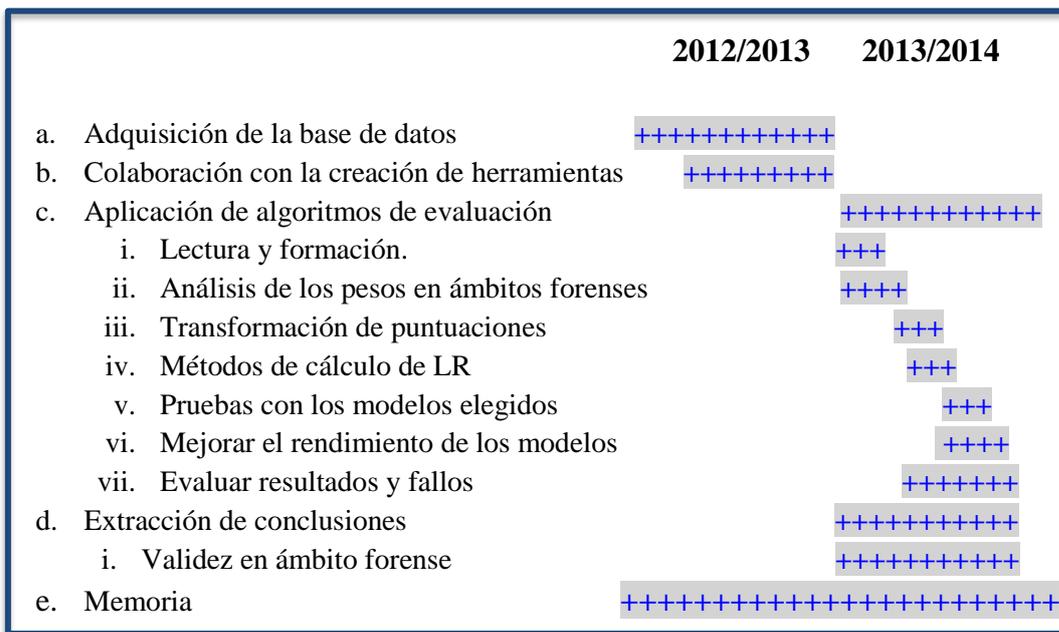


Tabla 1-1 Organización del proyecto fin de carrera.

2. Estado del arte

2.1 Biometría

La Biometría es la ciencia que estudia el reconocimiento de personas por sus rasgos fisiológicos o de comportamiento. Al igual que hacemos las personas cuando reconocemos por la cara o la voz del hablante, la biometría busca automatizar estos mismos procesos. Generar reconocimiento automático mediante los distintos rasgos biométricos que existan.

2.1.1 Características de los rasgos biométricos

Se entiende por rasgo biométrico a la característica propia de un ser humano en concreto, la cual le hace identificarse como individuo. Se pueden dividir en dos tipos, los rasgos anatómicos, aquellos que provienen de la biología del ser humano, genéticos y heredados, y los rasgos de comportamiento, adquiridos o aprendidos.

Para que cierta cualidad del ser humano pueda ser considerada rasgo biométrico, es necesario que ésta sea válida para identificar a la persona en cuestión a partir de dicho rasgo y para ello ha de ser coherente con las siguientes premisas de forma estricta:

1. Universalidad: Todo individuo debe poseer ese rasgo
2. Unicidad: La identificación de un rasgo debe ser único para un individuo, dos individuos distintos no pueden identificarse con el mismo rasgo.
3. Permanencia o estabilidad: Debe perdurar sin modificaciones con el tiempo. Perenne, al menos a corto plazo.
4. Mensurabilidad: Debe poder ser evaluado de forma cuantitativa

De forma opcional:

5. Aceptabilidad: El nivel de invasión al individuo debe ser muy bajo, para su evaluación lo ideal es que presente una buena colaboración del usuario.
6. Rendimiento: Precisión, fiabilidad, eficacia y velocidad de adquisición y evaluación de los rasgos.
7. Fraude o ataques: La capacidad para ser falseado debe ser mínima o existir un método de veracidad del individuo.

2.1.2 Clasificación de los rasgos biométricos

Como se ha citado antes dentro de los rasgos biométricos se diferencian dos subgrupos; los rasgos fisiológicos y los de comportamiento.

A continuación se muestra una pequeña descripción de algunos de estos rasgos resaltando únicamente la característica más relevante de su campo.

Rasgos fisiológicos:

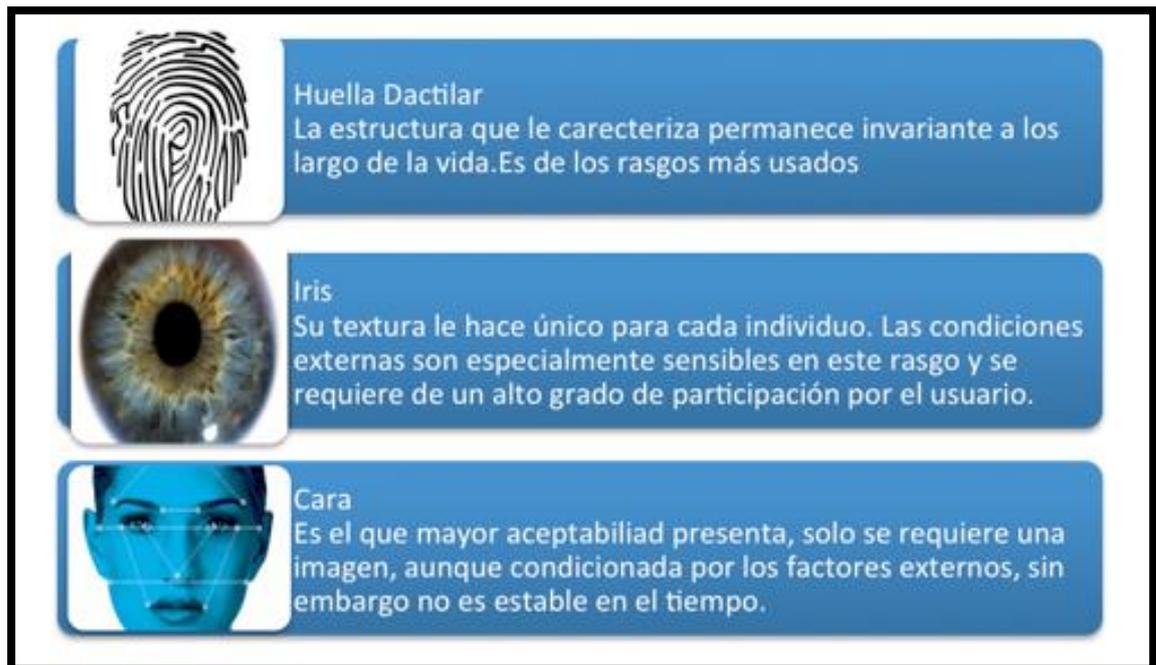


Figura 2-1 Ejemplos de rasgos biométricos fisiológicos

Rasgos conductuales:

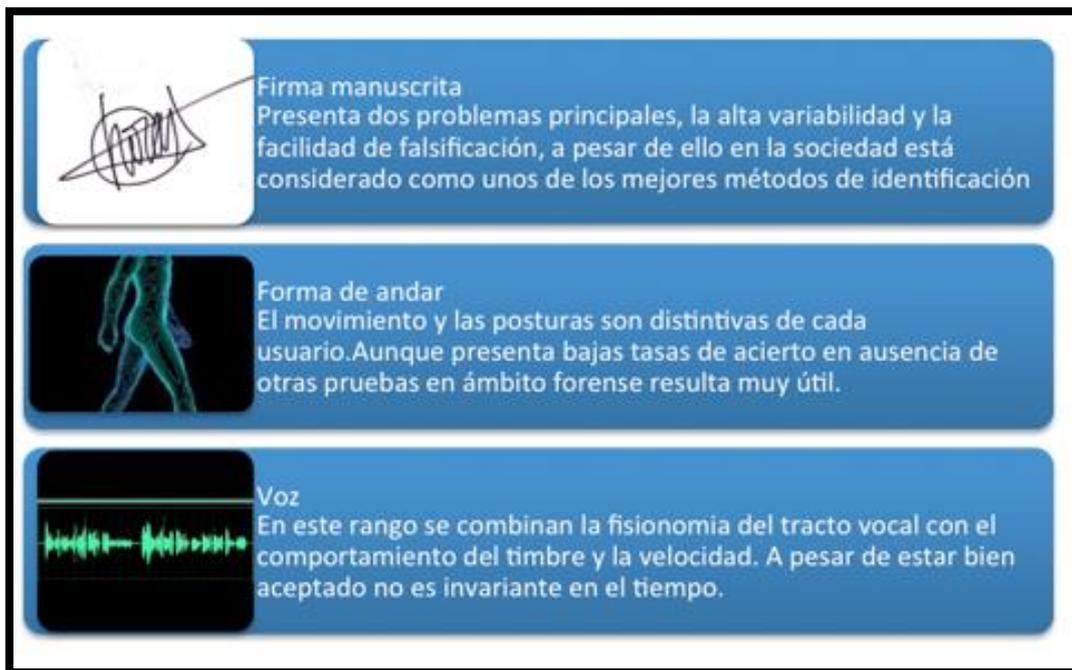


Figura 2-2 Ejemplo de rasgos biométricos conductuales

Además de éstos existen otros tipos de rasgos, algunos menos usados que no sirven de forma única para la identificación del individuo, sino que se usa para dar apoyo e información complementaria, como la altura, el peso, el color de ojos, etc. A estos se les denomina “soft biometrics” dado que únicamente proporcionan información añadida.

Por otro lado, surgen nuevos campos de investigación que actualmente no constituyen un sistema verificable o que su uso es mínimo, son las modalidades biométricas emergentes, tales como la forma de la oreja, el reconocimiento de olor o la termología facial, entre otras.

2.1.3 Biometría en ámbito forense

La biometría se puede aplicar en diferentes entornos de nuestra vida cotidiana, cada vez más son los formatos en que se pueden encontrar como por ejemplo, reconocimiento facial en los móviles, reconocimiento de huella en algunos ordenadores portátiles o reconocedores de lengua hablada en aplicaciones para móviles inteligentes. La ciencia forense es un entorno más en que la biometría juega un papel importante.

Se entiende ciencia forense como la aplicación de la ciencia a actividades de investigación criminal que buscan mejorar con la ayuda de los avances tecnológicos.

Los científicos forenses han demostrado que los rasgos físicos y de comportamiento pueden proporcionar información acerca de la identidad de personas, que en este caso, están implicadas en un caso de investigación criminal [3].

Sin embargo, no es una novedad, hoy en día, la utilización de la biometría en este entorno, por lo que ahora lo que se persigue es la automatización de los procesos de adquisición e identificación [2].

Referente a los procesos de adquisición se pretende mejorar la extracción de características, tratando de crear herramientas que agilicen el proceso además de intentar minimizar el error humano. Este es uno de los objetivos principales que marcan los peritos de la Guardia Civil, y por lo que este proyecto ha estado involucrado en tareas relacionadas con herramientas que los peritos usan para un mejor almacenamiento de las características de cada huella.

Con respecto a la identificación, resulta útil que un programa realice búsquedas frente a una gran base de datos y mediante comparaciones cuantitativas extraiga una lista con los candidatos que mayor similitud presentan según el modelo aplicado. Por un lado se reduce el tiempo de búsqueda ya que el perito solo tendría que verificar la identificación con un número bastante reducido de muestras. Por otro lado se aumenta la fiabilidad en los resultados ya que la biometría es capaz de aportar métodos de cálculo que han sido entrenados previamente para verificar un correcto funcionamiento.

La principal desventaja es la mala calidad de las pruebas. Hay que tener en cuenta que en este caso las muestras que se tienen han sido dejadas accidentalmente en la escena del crimen. La mayoría tan solo son fragmentos de huellas, han sido tratadas por sustancias químicas que resaltan las crestas, o se ven interferidas por objetos que no tienen que ver con la huella. A diferencia de las aplicaciones biométricas en entornos controlados la colaboración del individuo es escasa, esto provoca una variabilidad de las muestras muy elevada. En términos cuantitativos, provoca en los sistemas de identificación un poder de discriminación más limitado que en escenarios en los que las huellas son de mejor calidad.

En biometría forense se utilizan actualmente numerosos rasgos para la identificación:

- Reconocimiento facial en videos o imágenes.
- Cartas manuscritas.
- Huellas dactilares o palmares.
- Manchas de fluidos de los que extraer el ADN.
- Grabaciones de voz.

2.2 Huella Dactilar

En nuestros días se usan las huellas dactilares en ámbito policial para las investigaciones, o en identificación de individuos para controles de acceso. Se ha convertido en un método de identificación personal aceptado y formalizado por la sociedad.

Gracias a que el proceso de adquisición de huellas es poco invasivo para el usuario se pueden recopilar grandes bases de datos, las cuales sin la ayuda de la tecnología, los procesos de identificación resultarían realmente largos e ineficaces. En las bases de datos se dispone de las impresiones de huellas en las que se conoce la fuente. Cuando se necesita comprobar la identidad de una huella anónima, una huella que ha sido extraída de la escena del crimen, huella latente, se enfrenta contra esa base de datos para establecer comparaciones con todas, el sistema entonces propone una lista con los candidatos más probables.

Gracias a la tecnología actual se consigue automatizar los procesos, lo que ha provocado un rápido crecimiento de las aplicaciones comerciales en ordenadores o Smartphone.

A esto se le suma la importancia que tiene la huella dactilar en concreto, es el rasgo biométrico más utilizado debido a su fiabilidad, unicidad y a los buenos resultados que obtiene.

2.2.1 Formación de la huella dactilar

La forma que adquiere cada huella dactilar se obtiene a partir del séptimo mes de gestación y permanece invariable en el tiempo hasta el fallecimiento de la persona excepto en casos de accidentes como la aparición de cicatrices o en casos de enfermedad.

Lo más interesante de la figura que conforma la huella es que es muy improbable encontrar dos patrones de crestas de huellas iguales, desde el punto de vista de las características forenses. Este es el principal motivo que hace que la huella se sitúe en la cabeza de los rasgos más usados.

El dactilograma de la huella está formado por crestas y surcos; las crestas se refieren al relieve que forman y los surcos son las regiones que se encuentran entre las crestas. Los poros de la piel se encuentran sobre las crestas, éstos ayudan a la hora de diferenciar en una imagen lo que es cresta y lo que es surco. Juegan un papel muy importante en la seguridad, se utilizan para la comprobación de si existe vida en el individuo que está introduciendo la huella ya que cuando a la piel se le aporta calor los poros se dilatan.

El dactilograma de la huella puede ser analizado desde tres puntos de vista que serán los usados posteriormente para el reconocimiento y además se pueden encontrar en tres formatos distintos:

2.2.1.1 Natural

Es la observación directa de las huellas de los dedos.

2.2.1.2 Latente

Es la impresión producida por el contacto con una superficie. Generalmente no son visibles al ojo humano por ello requieren de una luz especial o productos químicos que reaccionen con las secreciones cutáneas producidas por el contacto.

Las huellas latentes son las que se obtienen de la escena del crimen directamente, por lo tanto la fuente es anónima, lo cual requiere de un estudio posterior de identificación. Además, suelen encontrarse en baja calidad, pueden presentar interrupciones o no estar completas, lo que dificulta su procesamiento. Debido a esto la identificación de huellas latentes es un proceso mucho más complicado que el de las impresiones digitales.

2.2.1.3 Artificial (impresión digital)

Es una reproducción gráfica de la huella natural, normalmente se produce en entornos controlados. Por ejemplo, cuando nos hacemos el DNI o cuando se detiene a un sospechoso se captan las huellas de ambas manos usando para ello tinta o un escáner directamente.

Esta captura puede ser por impresiones rodadas cuando se toma la huella por el rodamiento del dedo sobre el papel o por impresiones posadas, apoyando únicamente el dedo sobre el papel sin presentar movimiento.

Estas huellas son más controladas ya que se realiza en situaciones de calidad, la división entre crestas y surcos es muy clara y siempre permanece entera. Esto facilita los procesos de identificación resultando relativamente sencillo.



Figura 2-3 Huella natural, latente e impresión digital respectivamente. [16] [17] [18]

2.2.2 Reconocimiento dactilar en ámbito forense

El reconocimiento de la huella dactilar se basa en la forma que adquiere el relieve, crestas y surcos, y la manera en que crea un patrón de referencia. Para analizarla existen tres niveles distintos por los que se caracteriza una huella dactilar.

2.2.2.1 Primer Nivel

La forma que adquiere la huella teniendo una visión global, general. Se basa principalmente en dos zonas características, el núcleo y las deltas. El núcleo de la huella es aproximadamente el centro de ésta, el grado de curvatura de la alineación de las crestas. Cada huella adquiere una composición de crestas característica, además debido a las deltas, que son crestas que se dividen en varias crestas en direcciones distintas, las direcciones de las crestas se desvían. La información que extraemos de las deltas es el número de veces que aparece o incluso la ausencia de ello. Debido a ambas situaciones estas son algunas de las formas conocidas.

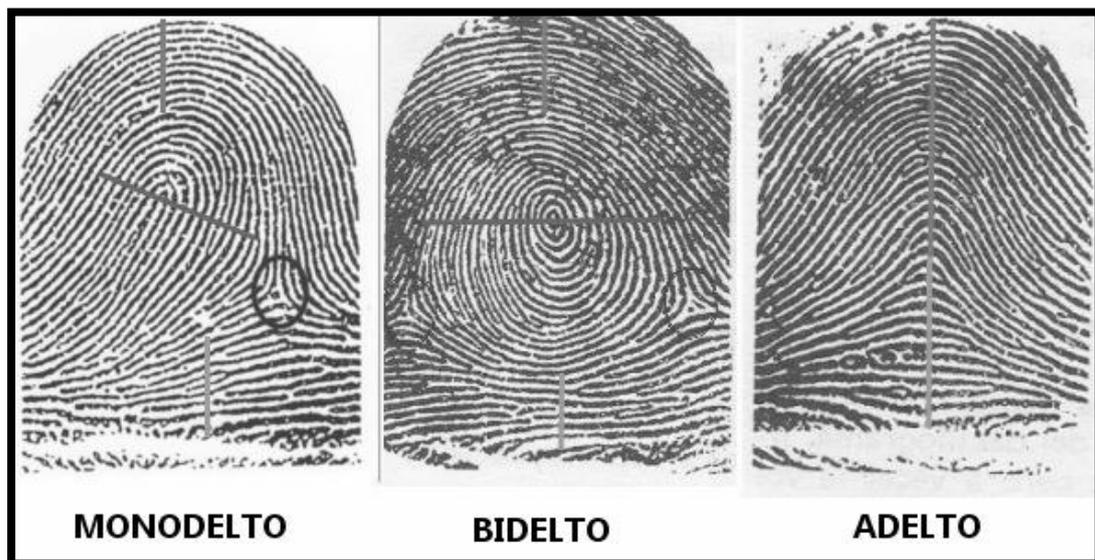


Figura 2-4 Tipos de huella según el primer nivel. [19]

A. ADELTO

Se caracteriza por la ausencia de delta, esto provoca un núcleo poco definido, en forma de arco suave, en el que las crestas de la base son muy uniformes con las del margen.

B. DESTRODELTO

Se aprecia porque está formado por un solo delta efectivo dirigido hacia la derecha, además el núcleo queda definido hacia el mismo lado.

C. SINISTRODELTO

Adquiere la misma composición que el caso anterior pero en sentido contrario, hacia la izquierda, tanto el delta principal como el núcleo.

D. BIDELTO

Presentan dos deltas, al menos uno de ellos es efectivo, provoca un núcleo muy definido debido a las orientaciones que provienen de ambos deltas. Estos deltas pueden estar en cualquier situación del dactilograma.

2.2.2.2 Segundo Nivel

Analizando más a fondo el dactilograma desde un punto de vista más concreto encontramos las minucias. Éstas se establecen porque la huella dactilar está formada por crestas y surcos, y la composición que crean son únicas para cada individuo. De ellas se puede extraer la ubicación, el tipo de minucia y la orientación que sigue la cresta.

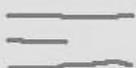
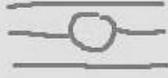
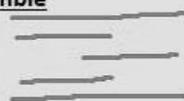
Terminación Abrupta 	Convergencia/Bifurcación 
Punto 	Desviación 
Empalme 	Fragmento 
Interrupción 	Ojal 
Secante 	Círculo 
Delta 	Vuelta 
Transversal 	Ensamble 
M 	

Figura 2-5 Tipos de minucias

2.2.2.3 Tercer Nivel

Esta última es la más detallista, centrándose únicamente en las crestas. Analiza el ancho de las crestas, cómo son sus contornos y los poros que se sitúan sobre ella. Se puede analizar el tamaño de los poros, la posición o forma.

Su uso es escaso ya que es más difícil, depende mucho de la calidad de las muestras que se tengan.

En este proyecto se utilizará un reconocimiento basado en minucias utilizando el mismo protocolo que los peritos de la Guardia civil y que más adelante se describirá.

2.2.3 Identificación dactilar en ámbito forense

Con el término identificación se hace referencia a la individualización de cada muestra mediante el análisis de las características extraídas. Por ello el proceso de identificación comienza tras la extracción de características de la muestra, en este caso, la huella dactilar.

Para proceder con la identificación es necesario establecer unos criterios previos en los que se defina un protocolo de actuación, un convenio común para identificar huellas anónimas con huellas identificadas. De esta manera se consigue que la identificación de los individuos dueños de las muestras sea siempre igual, siga un mismo convenio para que sea justo y válido.

Como este proyecto se ha realizado en colaboración con la Dirección General de la Guardia Civil el protocolo seguido para la extracción de características es el mismo que ellos siguen. Se aplica al método de reconocimiento basado en minucias, por ello, se establece que debe existir un número mínimo de minucias coincidentes entre huellas para que la identificación sea válida y puede llevarse a cabo. De esta manera se excluyen aquellas huellas que no se localice un número mínimo de minucias de la evidencia debido a la mala calidad de la imagen o porque el fragmento que se posee sea pequeño. Existiendo algunas excepciones que a continuación se detallan.

Criterio de identificación:

1. Es necesario un mínimo de doce minucias coincidentes.
2. Si el número de minucias que se localizan en una huella es menor que 12, esa evidencia no es válida para la identificación excepto en dos casos:

- a. Que tenga un mínimo de 8 minucias y además una de ellas sea una minucia de baja frecuencia de aparición, es decir, que no sean terminaciones abruptas, bifurcaciones o convergencias.
- b. Que sea una huella simultánea, esto es, que entre dos huellas de dedos distintos de un mismo individuo reúnan 12 minucias teniendo al menos una de las huellas un mínimo de 8 minucias.

Este es el sistema que se utiliza en España y en la mayoría de los países, sin embargo, la elección de situar el umbral en 12 minucias puede parecer arbitrario.

Tras la comparación de minucias de ambas huellas los peritos establecen una decisión:

- Identificada: Se ha cumplido con los requisitos y además coinciden.
- Excluida: No cumple con algún requisito y se descarta de la comparación.
- No concluyente: Cumpliendo con los requisitos no son huellas coincidentes.

El proceso de toma de decisión la realiza el perito analizando las dos huellas, observando las minucias extraídas y valorando las rarezas entre las huellas. Esta valoración siempre está supervisada por otro perito o un supervisor. Sin embargo, esta valoración se basa únicamente en la experiencia del experto y no en procedimientos empíricos repetibles.

Los errores que pueden cometer son dos, un falso positivo, declarar culpable a un inocente, o un falso negativo, declarar inocente a un culpable. Aunque ambos tipos de errores sean muy importantes, entre los expertos se considera que cometer un falso positivo es mucho más grave que cometer un falso negativo, ya que éste último se puede deber a una falta de pruebas suficientes que consigan demostrar la culpabilidad.

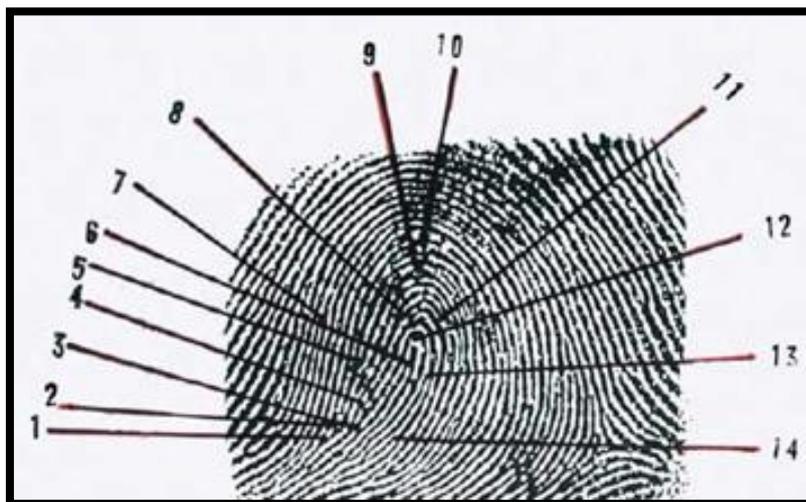


Figura 2-6 Extracción de características dactilar. [20]

2.3 Sistemas Biométricos

Los sistemas biométricos son procedimientos automáticos que buscan reconocer patrones de una determinada muestra y localizar la identidad del usuario al que pertenece. De la muestra se extraen características individuales que son las que se evalúan por el sistema de forma distinta según el tipo de rasgo a analizar [2].

Un sistema biométrico engloba generalmente los siguientes pasos:

- **Adquisición de datos:** Se define como el proceso de digitalización de las muestras a usar en los procesos posteriores. Se recoge y almacena el rasgo biométrico. En este paso tiene especial importancia la calidad con que se realice la captura ya que influirá en el rendimiento posterior del sistema o en la fiabilidad de los resultados. Es común el uso de lectores o escáner biométricos.
- **Pre-procesamiento:** De los rasgos anteriormente recogidos se preparan en el formato correcto o bien se acondiciona la imagen digital para recoger únicamente las regiones de interés o bien eliminar todo aquello que moleste, ruido o elementos ajenos.
- **Extracción de características:** Según el rasgo biométrico a usar se recogen los datos que vayan a caracterizar cada muestra para representar a un único individuo diferenciado del resto.
- **Cálculo de similitud:** Con las características anteriores se realizan comparaciones entre las muestras de la base de datos para cuantificar la similitud que existe entre ellas. De esta manera se obtiene un score para cada comparación realizada.
- **Toma de decisión:** Según la puntuación obtenida se valora si dos muestras son coincidentes o no basándose en umbral de decisión como ocurre en los sistemas de verificación. En este proyecto se pretende huir del establecimiento de un umbral para la toma de decisión ya que la elección de dicho umbral puede resultar subjetiva. La alternativa propuesta será presentar la decisión en términos de cantidad de apoyo que presenta una hipótesis para un rango de valores, que representan los distintos casos que se pueden dar. Valorándolo así sobre todos los umbrales posibles.

2.3.1 Tipos de sistemas biométricos

En el estado del arte existen distintos tipos de sistemas biométricos según la finalidad y el uso al que vayan a ser destinados. Encontramos dos tipos principalmente, sistemas de verificación y sistemas de identificación y una posible aplicación a ambos que establecemos como un subgrupo, los sistemas de registros.

- **Registro previo:** Es necesario cuando el sistema lleva incluido la creación de la base de datos por él mismo, es decir, un sistema en el que para su uso es necesario que los individuos que vayan a usar el sistema se registren previamente. De esta manera se almacenan las características del rasgo biométrico junto con la identidad de la persona para su uso posterior. Suele ser común, para mayor seguridad, solicitar al usuario que introduzca varias veces su rasgo biométrico para tener en cuenta la variabilidad de éste.



Figura 2-7 Esquema del proceso de registro

- **Identificación:** Trabajando en este modo se tiene una base de datos completa previamente de forma que para el usuario que introduce su rasgo biométrico, inicialmente anónimo, únicamente se realiza una búsqueda de si está o no entre los que forman la base de datos. Se dice que este modo trabaja comparando “uno a muchos” lo que conlleva un gran coste computacional. El resultado de esta aplicación puede ser únicamente categorizar si el usuario está o no en la base de datos o bien mostrar una lista con los candidatos que mayor similitud presenten. En estos sistemas normalmente puede que el usuario pretenda no ser encontrado en dicha base de datos, por lo que este modo es el que más se usa en entornos forenses, en el que el usuario pretende no ser identificado.

Para la obtención de la cantidad de similitud con las muestras los scores obtenidos se transforman a LR para una mejor interpretación.

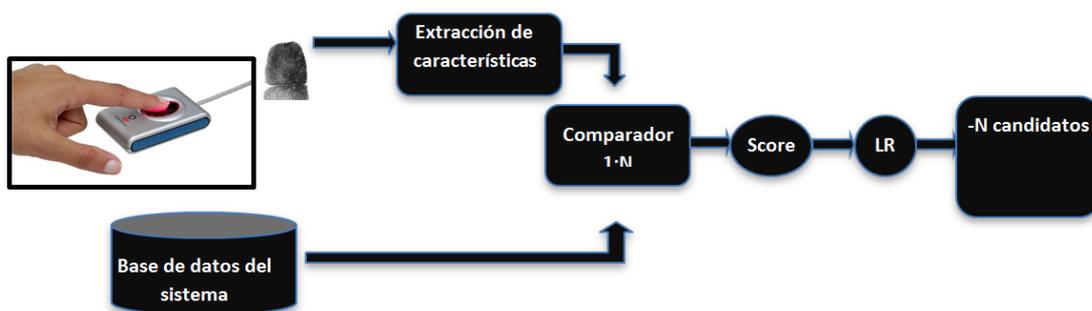


Figura 2-8 Esquema del proceso de identificación

- **Verificación:** A diferencia del modo anterior, en este caso la comparación realizada es de “uno a uno” en la que el usuario se identifica y se compara con la muestra que anteriormente se había registrado de él. El resultado del sistema es confirmar si el usuario es quien dice ser, por lo que a la salida habrá un sí o no. Esta decisión se obtiene mediante un umbral que categoriza los scores obtenidos

de la comparación y que opcionalmente han sido transformados a LR. Según la importancia del sistema se puede imponer un umbral más restrictivo o por el contrario bajarlo debido a que el sistema no requiere mucha seguridad.

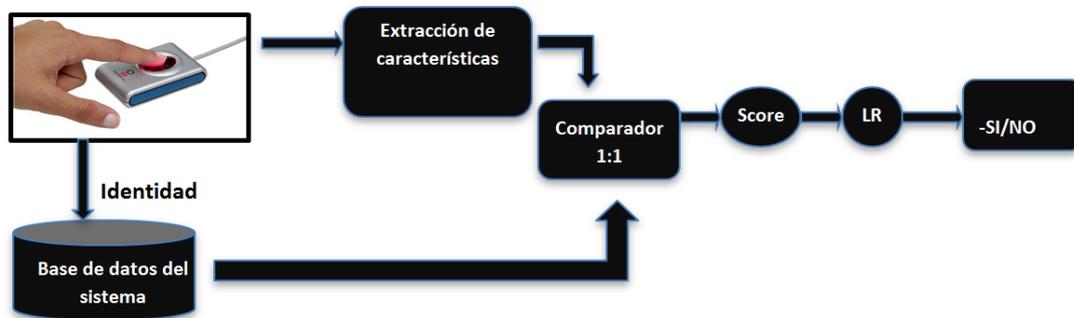


Figura 2-9 Esquema del proceso de verificación

2.3.2 Sistemas forenses de identificación dactilar

De los sistemas biométricos que anteriormente se han explicado en el campo de las ciencias forenses es el que se pone en práctica es el sistema de identificación, en el que el usuario no quiere que sea identificado por eso se desconoce la identidad de la muestra. Es decir, con la muestra extraída del caso se compararía con la base de datos para encontrar cuales son los candidatos más probables.

La principal diferencia con respecto a los sistemas de verificación está en el número de comparaciones que se realizan. En los sistemas de verificación se realiza una única comparación y como resultado una respuesta afirmativa o negativa. Por el contrario en los sistemas de identificación se compara una muestra con todas las existentes en la base de datos que se tenga, existiendo tantas comparaciones como muestras de características exista. Además el resultado de la búsqueda tampoco coincide con el de los sistemas de verificación ya que en este caso se devuelve una lista de las muestras con más probabilidad de acierto, ordenadas de mayor a menor.

Se concluye con que los sistemas forenses de reconocimiento dactilar, en este caso, utilizan sistemas biométricos de identificación, que, a partir de una huella anónima y la extracción de características se compara con toda la base de datos finalizando con una lista de las huellas más probables que serán finalmente evaluadas por un experto humano. Estos sistemas se conocen con el nombre de AFIS en el ámbito dactilar (Automated Fingerprint identification System). En español se conoce como SAID (Sistema Automático de Identificación Dactilar) y es de donde se han extraído las huellas identificadas que se utilizarán para la base de datos que se va a crear en este proyecto final de carrera.

2.3.3 Extracción de características de los sistemas forenses dactilares

El sistema español, SAID, concretamente, utiliza un método de reconocimiento basado en minucias en el que se realiza la extracción de características para obtener una plantilla de datos para cada huella.

Para que dos huellas se consideren coincidentes, pertenecen al mismo individuo al que llamaremos fuente de los datos, las minucias extraídas de ambas deben coincidir en la ubicación de cada una, el tipo y el número de minucias localizadas. Para ello los peritos extraen 3 datos de cada minucia identificada con el fin de obtener una plantilla que englobe todos los datos extraídos.



Figura 2-10 Extracción de características

- **Ubicación:** Para conocer la ubicación de donde se sitúa cada minucia es necesario conocer las coordenadas donde se localiza el punto característico. Las coordenadas se recogen todas manteniendo un sistema de referencias para posteriormente poder compararlas entre sí. Es importante que la ubicación dentro de la huella coincida en ambas muestras a comparar, pero también es importante la distancia y posición entre minucias de una misma huella. Así, daría información de cómo están distribuidas por la huella, incluso de poder reproducir esta plantilla.
- **Orientación:** Para cada minucia localizada se indica la orientación que sigue la cresta de ese punto con respecto al plano horizontal. Gracias a la orientación se puede estimar la composición que adquieren las crestas dentro de la huella.
- **Tipo:** Por convenio se numeran todos los tipos de minucia conocidos [Figura 2.5]. Facilitando así la tarea de reconocimiento mediante el uso de un mismo código identificativo.

Actualmente existen métodos y algoritmos que se encargan de la extracción de características automáticamente, no solo mediante el método de minucias sino extrayendo otras características propias de la huella. Sin embargo, este proyecto no se focaliza en el proceso de automatización de la extracción, ya que los peritos todavía no se muestran muy familiarizados y convencidos de utilizar estos sistemas y por ello se sigue necesitando de la participación de un humano para estas tareas.

3. Creación de la Base de Datos: Proyecto de huella con la Dirección General de la Guardia Civil

El proyecto final de carrera comenzó con una beca de colaboración en un proyecto de la Guardia Civil en el que participaron tanto peritos miembros de la DGGC como integrantes del grupo ATVS de la Universidad Autónoma de Madrid.

Entre las funciones a desempeñar en este proyecto se detallarán aquellas que atañen al proyecto final de carrera que aquí se expone. Está enfocado en el entorno de la huella dactilar, en colaboración con el departamento de lofoscopia de la Guardia Civil dentro de criminalística.

La motivación del proyecto es mejorar el proceso que siguen los peritos para la adquisición de minucias en una huella dactilar, se precisa que el procedimiento sea más rápido y eficiente. Para ello en primer lugar es necesario que los peritos indiquen los procedimientos que ellos usan, además de proporcionar el material suficiente para que el trabajo se pueda llevar a cabo.

Asimismo, desde el ATVS se persiguen dos objetivos:

-Creación de una herramienta visual que imite el proceso de extracción de características. Facilitando así las tareas diarias de los peritos, de forma que el proceso sea más rápido además de almacenar de forma electrónica todos los datos usados, las huellas implicadas y las características extraídas.

-A partir de la herramienta anterior, se diseñará una nueva que mediante las características extraídas realice cálculos de razones de verosimilitud, LR, con el fin de extraer una puntuación que estime el grado de apoyo de los resultados.

El fin es que los peritos puedan, de aquí en adelante, tomar estas herramientas como método de trabajo y poco a poco ir introduciendo técnicas de cálculo para la valoración de la prueba en un caso judicial.

La **contribución principal de este proyecto final de carrera** en el proyecto con la Guardia Civil es la creación de una base de datos de huellas dactilares y palmares mediante el uso de esta herramienta gráfica llamada “MINUCIA”. Mientras se realizó la base de datos, se fueron corrigiendo y arreglando problemas que iban surgiendo en el uso de la herramienta. Con la realización de la base de datos se consiguen dos objetivos:

- La utilización de esta herramienta permite ponerla a prueba antes de entregársela a la DGGC. Se prueba su funcionamiento de forma paralela a la creación de la base de datos, esto permite introducir mejoras y solucionar errores. Se consigue así un mejor funcionamiento y obtener un sistema adaptado a los procedimientos que realizan los peritos en su día a día.
- Disponer de una base de datos para el grupo ATVS procedente de huellas dactilares reales, obtenidas de casos forenses reales, lo cual será relevante a la hora de evaluar los resultados finales que se obtengan. Una vez terminada se quiso ampliar aplicando los métodos a las huellas palmares que presentan procedimientos muy parecidos.

3.1 Base de Datos Lofoscópicos del ATVS

Durante el primer año de este proyecto final de carrera se realizó la tarea de adquisición de una base de datos de huellas dactilares.

El objetivo es conseguir tener una base de datos de huellas procedente de casos reales, como las usadas en los casos de lofoscopia forense, en formato electrónico. Además, conseguir unificar el formato que presentan las huellas, sus impresiones y sus características.

Para ello el proceso de adquisición se ha realizado con la herramienta “MINUCIA” del ATVS, diseñada también para el proyecto de lofoscopia. Durante la utilización de esta herramienta se han ido incorporando soluciones a los problemas que a medida iban surgiendo.

Sin embargo, en primer lugar, el proceso se realizó “a mano” utilizando otra herramienta gratuita, GIMP. Es un editor de imagen que permitió marcar las minucias sobre la imagen de la huella. De esta manera se comenzó el proceso de formación para comprender cómo trabajaban los peritos. Posteriormente se pasó a utilizar MINUCIA para toda la base de datos notándose una disminución considerable en el tiempo de ejecución. En comparación, sin la herramienta podíamos realizar en una hora 1,25 pares

de huellas¹ de media a la hora, tras la utilización de la herramienta se aumentaba a 2,75 pares de huellas a la hora.

El proceso de adquisición no incluye únicamente el marcado de las minucias sino que previamente es necesario el procesamiento de las imágenes para que las coordenadas indicadas se realicen bajo el mismo criterio, por ello el trabajo realizado se divide en tres etapas: Pre-procesamiento de la imagen, marcado de minucias y guardado de puntos.

Para comprender de donde se parte y hasta donde queremos llegar es necesario definir los dos tipos de huellas que se van a utilizar.

Por un lado tenemos las huellas procedentes del sistema SAID, de las que se conoce su origen, se llaman huellas indubitadas. Por otro lado, para cada huella indubitada existe una huella latente procedente de la escena del crimen denominada huella dubitada. De la huella dubitada el perito realiza su identificación, gracias a la cual podemos realizar nuestros experimentos.

Conjunto de información inicial

- Cotejos en papel, con las minucias identificadas.
- Imagen huella dubitada
- Imagen huella indubitada

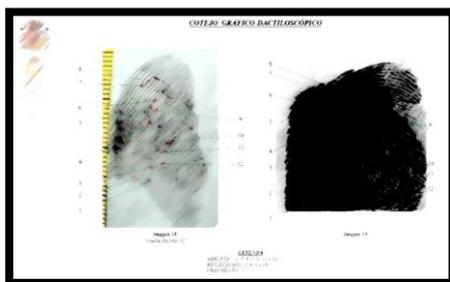


Figura 3-1 Cotejo lofoscópico

Formato de la Base de Datos final.

- Archivo txt. Plantilla de características
- Imagen huella dubitada en PNG a 500 ppp
- Imagen huella indubitada en PNG a 500 ppp

```
1| 138 285 196 2
2| 115 298 10 1
3| 121 276 16 1
4| 97 231 221 1
5| 110 215 56 1
6| 182 176 251 1
7| 206 118 231 1
8| 197 182 266 1
9| 149 226 239 1
10| 165 238 262 1
11| 155 270 334 5
11| 163 273 162 5
12| 195 284 154 1
```

Figura 3-2 Plantilla de datos

3.1.1 Pre-procesamiento de imagen

Para poder extraer posteriormente las minucias es necesario que las imágenes a utilizar posean un formato común. En este caso existen dos tipos de huellas cada una en un formato distinto.

¹ Se entiende como pares de huella a la huella latente de origen desconocido con su huella identificada coincidente.

- Huella indubitada, proviene del SAID, presenta una resolución más estándar de 500 o 1000 ppp.
- Huella dubitada, proviene directamente de la escena del crimen tomada con una cámara por lo que cada una se encuentra a una resolución y tamaños distintos. Esto supone un problema de homogeneidad de escala a la hora de trabajar con las imágenes. Ya que no permite tratarlas del mismo modo.

Para la base de datos se toma la decisión de presentar todas las imágenes a una resolución de 500 píxeles por pulgada. Esta elección se debe a dos motivos, en el caso de las huellas indubitadas siempre están a 500 o a 1000 ppp, y como no disponen de testigo métrico es necesario escoger una de esas dos resoluciones. Entre ellas es una transformación sencilla de dividir a la mitad la resolución en caso de ser de 1000 ppp y de descartar aquellas que difieran de 500 o 1000 ppp ya que éstas no presentan testigo métrico para poder hacer las transformaciones adecuadas. En el caso de las huellas dubitadas no existe ningún problema en realizar una transformación a escala real a 500ppp. Siempre y cuando presenten testigo métrico se podrían transformar a cualquier resolución. Aunque es conveniente no aumentar la resolución con respecto a la original ya que se estarían introduciendo píxeles de relleno. La más adecuada es la de 500 ppp, coincidente con la resolución de las impresiones procedentes de la mayoría de los sistemas estándar de identificación.

Para realizar la transformación es necesario tener claros algunos conceptos.

- Tamaño: El tamaño de la imagen nos lo va a indicar el número de píxeles que posea la imagen en las dos dimensiones, horizontal y vertical.
- Resolución: La resolución viene definida como el número de píxeles por pulgada. Entendiendo la pulgada a tamaño real. (1 pulgada=2.54 cm)
- Testigo métrico: Es requisito obligatorio que todas las imágenes de las huellas dubitadas traigan en ella un testigo métrico. Es la única forma de conocer el tamaño real de la huella y poder re escalarla para conseguir la resolución adecuada.

En definitiva, se necesita modificar el tamaño de la imagen en píxeles para conseguir una resolución de 500 píxeles en una pulgada, obteniendo la equivalencia de los píxeles a pulgadas real mediante la referencia del testigo métrico.



$$\text{Tamaño}_i(\text{pulgadas}) = \text{Tamaño}_i(\text{píxeles}) / \text{Resolución}_i \quad (3-1)$$

$$\text{Tamaño}_f(\text{píxeles}) = 500\text{ppp} * \text{Tamaño}_i(\text{pulgadas}) \quad (3-2)$$

Tras esto ya quedarían todas las imágenes en la resolución requerida y preparadas para la extracción de características.

Figura 3-3 Huella con testigo métrico. [17]

3.1.2 Marcado de minucias

Como define el estado del arte, las huellas dactilares están formadas por crestas y surcos, y la formación que constituye el dactilograma es muy improbable que se repita para otro individuo. Para el caso de la extracción de características basadas en minucias se adquieren tres datos para cada minucia, la coordenada, orientación y tipo de minucia.

Se establecen 15 tipos distintos de minucias según el dibujo que describa la cresta. Para indexar los tipos de minucias se implementará un código de numeración en el que cada tipo de minucia quede representado por un número.

Además, se extrae las coordenadas de la minucia en el plano de la huella estableciendo unos ejes y por último se marca también la orientación que sigue la cresta, manteniendo el trazado de ésta, así, se aporta más información de la situación en la que se encuentra la minucia en la huella.

Dentro de los 15 tipos de minucias que barajamos podemos encontrar unos que se clasifican como minucias de alta frecuencia de aparición: terminación abrupta, convergencia y bifurcación. El resto de tipos de minucia son los de baja frecuencia de aparición, son aquellos que son más difíciles de identificar y que en menos ocasiones están presentes, por ejemplo, transversal o ensamble.

Además, debido a la complejidad de algunos tipos de minucia, no es suficiente un solo punto de coordenadas para definir su situación en la huella sino que se propone establecer un criterio de marcado en el que una minucia se pueda identificar con más de un punto, los suficientes para que la estructura del tipo de minucia quede perfectamente representada. Se presenta a continuación un esquema de cómo queda marcada cada tipo de minucia, el número de coordenadas que necesita, la orientación que sigue y el número identificativo.

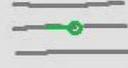
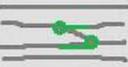
Terminación Abrupta 	1	Convergencia/Bifurcación 	2
Punto 	8	Desviación 	3
Empalme 	4	Fragmento 	5
Interrupción 	6	Ojal 	7
Secante 	9	Círculo 	11
Delta 	12	Vuelta 	15
Transversal 	10	Ensamble 	13
M 	14		

Figura 3-4 Esquema del código de identificación de minucias

En el proceso de adquisición actual de los peritos forenses de lofosocopia definen que son 12 las minucias mínimas necesarias para la identificación de la huella por lo que la base de datos ha sido diseñada para que cada huella esté definida únicamente con 12 minucias. Es necesario que todas sigan un mismo esquema por ello no se contemplan las excepciones y aquellas que no cumplan los requisitos quedarían excluidas de esta base de datos.

3.1.3 Guardado de puntos

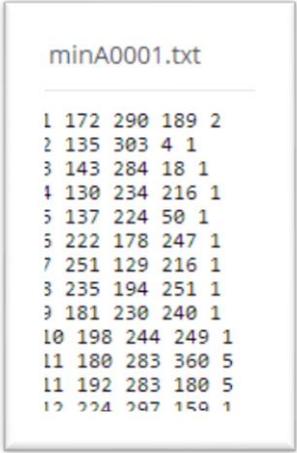
En último lugar es necesario guardar en formato electrónico los datos que se han extraído de las huellas. En esta etapa es donde se crea una plantilla para cada huella donde se encontrarán todos los datos de la extracción de características.

Esta plantilla es un archivo plano ".txt". Es una ventaja la utilización de este formato ya que ocupa muy poco espacio. Además, es muy cómodo para su uso posterior en programación y algoritmos porque sigue siempre una misma estructura para todas las huellas y permite trabajar con los datos sin necesidad de volver a necesitar las imágenes, que son lo que más espacio ocupan.

Gracias a esta plantilla y a partir de ella se podría volver a dibujar las minucias para representar cómo se sitúan gráficamente. Si se quisiera comparar con otra solo habría que pintar una plantilla encima de la otra para poder estimar si las doce minucias son coincidentes en el espacio.

Para la estructura que sigue esta plantilla se necesita, manteniendo el mismo orden en el que se definen:

- **Numero de minucia (orden):** Indica la numeración del 1 al 12 de cada minucia en la huella, es necesario para la comparación que las minucias coincidan en numeración, de las de la huella dubitada con sus homólogas en la indubitada.
- **Coordenadas del punto marcado:** Se guardan las coordenadas sobre el punto principal de la minucia que queda indicado con el círculo en el esquema propuesto (Figura 3-4).
- **Orientación:** Se indica los grados del ángulo que forma.
- **Tipo de minucia:** El tipo viene definido previamente por el esquema propuesto (Figura 3-4).



```
minA0001.txt
1 172 290 189 2
2 135 303 4 1
3 143 284 18 1
4 130 234 216 1
5 137 224 50 1
5 222 178 247 1
7 251 129 216 1
3 235 194 251 1
3 181 230 240 1
10 198 244 249 1
11 180 283 360 5
11 192 283 180 5
12 224 207 150 1
```

Figura 3-5 Ejemplo de fichero txt. Plantilla de la huella. Cuando la minucia necesita más de un punto para identificarla se escribe una nueva línea para cada punto, como aquí ocurre con la número 11.

Las herramientas creadas para el proyecto de la Guardia Civil imitan el proceso que se acaba de detallar. Disponen de todas las opciones necesarias para la realización de la base de datos y para su futuro uso por parte de los especialistas. Gracias a las herramientas el proceso se realizará más rápido, seguro y útil para futuras comprobaciones, modificaciones o tratamiento de los datos a través de las plantillas.

3.1.4 Resultado

Base de datos del ATVS:

- 258 pares de huellas (Huella dubitada con su huella indubitada coincidente)
- 248 individuos distintos implicados, ya que para algunos individuos se tenían muestras de varios dedos. Entre 1 y 4 pares de huellas por individuo.
- Formato electrónico: un archivo txt para cada huella.
- Formato unificado para todas las huellas.

3.1.5 Base de datos de huellas palmares

Finalmente, una vez terminado con las muestras de huellas dactilares que los peritos tenían disponibles, se comenzó a introducir la posibilidad de ampliar la base de datos. Esta vez mediante la adquisición de huellas palmares. Del mismo modo que ocurría en huella dactilar, en este caso para cada muestra de un caso judicial que proviene de una zona de la palma se identifica cuál es su huella palmar coincidente.

Para la realización de esta base de datos también se usa la herramienta MINUCIA ya que el proceso de adquisición, extracción de características y guardado de los puntos son de la misma forma que en el caso de la huella dactilar.

Sin embargo, existen algunas particularidades asociadas a la huella palmar:

- La huella palmar consta de dos regiones, la palma y los hipotenares, el derecho y el izquierdo para ambos casos.
- Debido a que este rasgo presenta un tamaño mayor que la huella dactilar se permite que aparezcan más de 12 minucias, por lo que no hay un número fijo establecido, el objetivo es encontrar las máximas minucias en común que sea posible.
- Pre-procesamiento de la imagen de la huella indubitada. Para la huella palmar dubitada el procedimiento es el mismo, siendo necesario siempre la presencia de un testigo métrico, sin embargo, para la indubitada también es necesario una transformación previa. Se parte de un cotejo en PDF en el que se presenta la huella de la palma derecha, izquierda y los hipotenares derecho e izquierdo respectivamente. Dentro del PDF se muestra también un testigo métrico. Éste tiene un papel imprescindible, ya que va a ser el que nos permita recortar la región de interés que vaya a estar involucrada en la identificación y mediante el uso del testigo realizar un cambio en la resolución de la imagen para conseguir 500 píxeles por pulgada real, igual que se procede en el caso de las dubitadas.



Figura 3-6 Huella palmar. [21]

3.1.5.1 Resultado:

Base de datos palmares del ATVS:

- 20 pares de huellas palmares. Cada huella latente con la huella indubitada.
- Procedentes de 14 individuos distintos, para alguno de ellos se tienen de 1 a 3 regiones palmares distintas.
- Formato electrónico, plantillas en archivo txt.
- Formato unificado y similar al de las huellas dactilares.

3.2 Herramientas gráficas de adquisición

Para el desarrollo de este proyecto se colaboró en la creación de dos herramientas, Minucia y Lr-Lofoscopia. La tarea principal que se aporta desde este proyecto no fue la de crear la herramienta en sí, sino la de testear, mediante prácticas continuas, sus funciones para velar por el buen funcionamiento de la herramienta. La mejor manera de realizar las pruebas fue creando la base de datos lofoscópicas mediante la utilización de este programa.

Las líneas que se han seguido para el funcionamiento del programa son las que los peritos forenses de la Guardia Civil necesitaban, ya que ellos son los usuarios finales de este programa. Se comentará el funcionamiento y el proceso que siguen.

3.2.1 MINUCIA

Como se podrá comprobar, la forma de actuar de la herramienta **MINUCIA** será idéntica a la explicación de los pasos a seguir para la creación de la base de datos en el apartado 3.1, la principal diferencia y ventaja a la vez es que la transformación a la resolución de 500 ppp y la creación de la plantilla en el fichero txt se realizan de forma automática durante la intervención del usuario. Éste solo tiene que intervenir para indicar dónde se encuentran las minucias. Gracias a esto el proceso de creación de la base de datos se agilizó considerablemente.



Figura 3-7 Interfaz Gráfica de MINUCIA

Gráfico	Función	Descripción
	ABRIR IMAGEN	Al abrir la imagen, si desea reescalarla debe indicar la unidad de medida que indicará el usuario sobre el testigo métrico.
	DIBUJAR MINUCIA	El usuario debe marcar el punto primero y seguido hacía donde se dirige la dirección de la cresta, posteriormente indicará el orden de la minucia y el tipo de ésta. Así para las doce minucias.
	EXPORTAR FICHERO	Cuando se dibujan las doce minucias se podrá generar el archivo donde irán guardados los datos y se terminará el proceso.

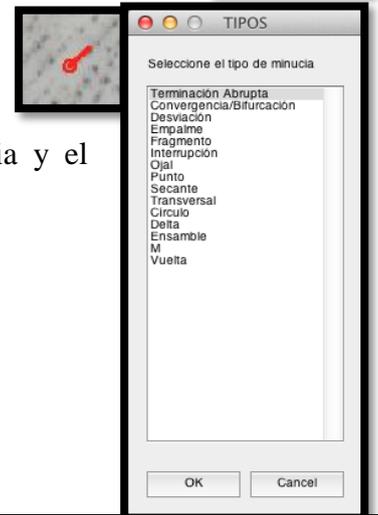


Tabla 3-1 Esquema del funcionamiento de MINUCIA

Gráfico	Función	Descripción
	IMPORTAR FICHERO DE MINUCIAS	Tras abrir una imagen se puede importar el fichero de minucias que se haya realizado antes para volver a dibujar las minucias sobre ésta.
	EDITAR MINUCIA	Se puede modificar tanto el punto marcado como la orientación haciendo uso de las flechas. También se puede eliminar directamente.

Tabla 3-2 Esquema del funcionamiento de MINUCIA

3.2.2 LR-LOFOSCOPIA

La herramienta LR-LOFOSCOPIA surge como complementación al proceso que realiza MINUCIA. Cuando los datos se han extraído de la huella y se ha generado el fichero de minucias este programa utiliza un algoritmo mediante el cual se calcula la relación de verosimilitud entre dos huellas, a través de sus ficheros de minucias.

Además, esta herramienta también incorpora la posibilidad de realizar el marcado de las minucias desde la misma interfaz que se calcula el LR. Sin embargo, sus funciones son más limitadas ya que las imágenes deben estar ya en la misma resolución.

En la interfaz gráfica se ve que está dividida en dos partes idénticas, una reservada para la huella dubitada, la huella anónima, y la otra destinada para la huella indubitada, la impresión digital. Cada una de las partes tiene las mismas funcionalidades, cuyo uso es idéntico al que se explicó para 'MINUCIA'. Además, cada huella tiene su tabla de minucias donde se representan los datos del fichero.

Esta herramienta puede trabajar en dos modos:

- Modo marcado de minucias
- Modo cálculo de LR

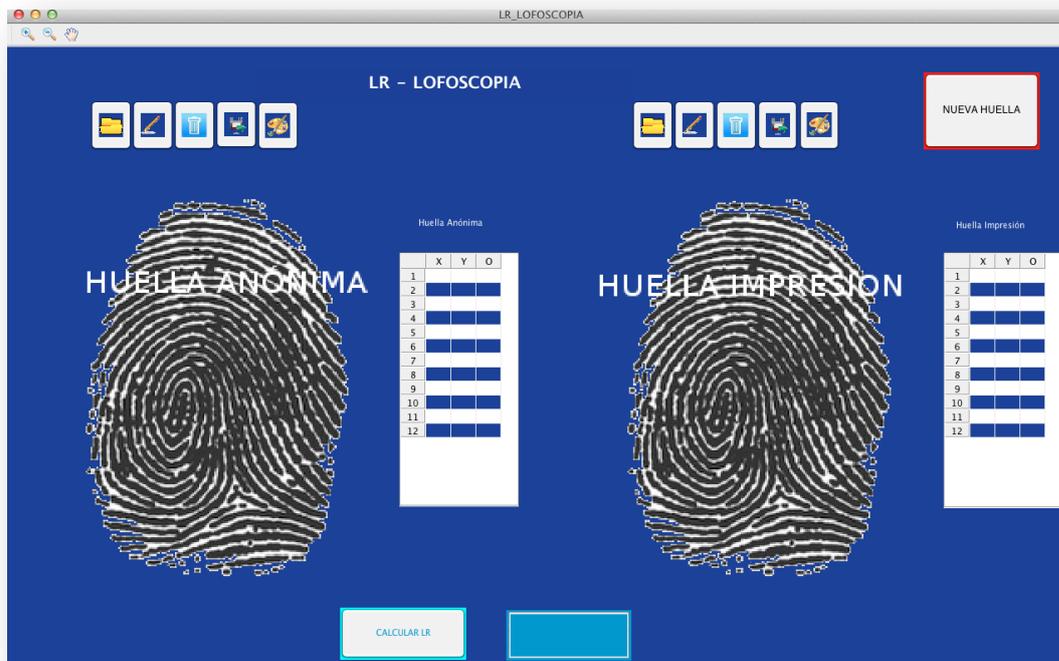


Figura 3-8 Interfaz Gráfica de LR-LOFOSCOPIA

MODO MARCADO DE MINUCIAS:

Gráfico	Función
	ABRIR IMAGEN Y REESCALADA
	DIBUJAR MINUCIA
	ELIMINAR MINUCIA
	EXPORTAR FICHERO DE MINUCIAS
	IMPORTAR FICHERO DE MINUCIAS

Tabla 3-3 Esquema del funcionamiento de LR-LOFOSCOPIA en modo marcado de minucias

MODO CÁLCULO DE LR

Cuando las dos imágenes de las huellas están marcadas y se han importado para ambas el fichero de minucias se puede utilizar la función Cálculo de LR. El sistema devolverá una puntuación como resultado del cálculo interno que realiza.



Figura 3-9 Función para calcular LR

Tras las pruebas realizadas por los peritos forenses de la Guardia Civil, se recogieron las impresiones y valoraciones, en consecuencia, este proyecto final de carrera ha estado involucrado en tareas de mejoras y solucionar problemas en la herramienta 'LR-LOFOSCOPIA' tales como:

- A pesar de que la base de datos del ATVS está formada por huellas que se caracterizan con 12 minucias obligatoriamente (no se permite ni más ni menos), la Guardia Civil necesita que se permita que el número de minucias sea menor o mayor, por ello se ajusta el programa quitando las restricciones del número de minucias permitidas para generar el archivo de datos.
- Evitar que los peritos puedan marcar las minucias fuera de la imagen habilitada para ello, daría lugar a incoherencia en los cálculos de LR posteriores, así como una consecución de errores en el proceso de marcado.
- Creación de la función 'Nueva huella' para poder arrancar el programa de nuevo bien porque se haya producido un problema o bien porque se desee continuar con las huellas siguientes. De esta manera no habría que cerrar y abrir de nuevo el proceso.



Figura 3-10 Función 'Nueva huella'

4. Teoría de Decisión Bayesiana

La teoría de decisión bayesiana permite valorar las hipótesis disponibles a partir de las evidencias de un caso. Proporciona las herramientas matemáticas necesarias para ello. Permite al encargado de la toma de decisión establecer la probabilidad de si el individuo es culpable o inocente [4] en conjunto con toda la información del caso. Los resultados de este proyecto se basan en esta teoría adaptándose a las dos categorías que se tienen, mediante la comparación de dos huellas procedentes de un caso judicial.

En la teoría bayesiana existen dos categorías posibles para un vector de puntuaciones que provienen del sistema biométrico a estudiar. Las categorías definidas son: pertenecer a la misma fuente y pertenecer a fuentes distintas, entendiendo fuente como el individuo, la huella, o la evidencia a tratar. Esta teoría se aplica cuando se tienen las probabilidades de ambas categorías, P y Q.

$$P = \text{prob}(H1|E) \quad 4-1$$

$$Q = \text{prob}(H2|E)$$

Siendo 'E' la evidencia científica, y 'H1' y 'H2' las dos hipótesis disponibles, pertenecer a la misma fuente y pertenecer a fuentes distintas respectivamente.

Por lo tanto, si se quiere minimizar la probabilidad de error, si $P > Q$ se escoge la hipótesis H1, y si $P < Q$, la hipótesis contraria. Sin embargo, en un caso judicial es muy complejo, en general, conocer previamente las probabilidades que tienen las evidencias de pertenecer o no a la misma fuente. Por ello, la teoría bayesiana nos permite entender esta probabilidad desde el otro sentido, y es analizar las probabilidades de que esa evidencia pertenezca o no a la misma fuente bajo suposiciones de una u otra categoría.

Se basa en la definición de dos nuevas probabilidades, por un lado la probabilidad de que una puntuación obtenida de las evidencias pertenezca a la misma fuente suponiendo que son de la misma fuente y la probabilidad de que esa misma puntuación pertenezca a la misma fuente suponiendo que viene de fuentes distintas [10].

$$P' = \text{prob}(x|H1) \quad 4-2$$

$$Q' = \text{prob}(x|H2)$$

Teniendo esto en cuenta si $P' > Q'$ entonces es más probable que x apoya a la hipótesis $H1$. Sin embargo, si se da el caso en que $P' = Q'$ no aporta ninguna información relevante para tomar la decisión.

En definitiva, la teoría bayesiana relaciona las probabilidades de cada hipótesis antes y después de analizar las evidencias [4].

Para aplicar este modelo a un caso forense real definimos los siguientes conceptos:

- T: Toda la información disponible del caso
- E: Evidencias, las pruebas, pueden ser imágenes, características o materiales, depende del sistema biométrico utilizado.
- I: Se corresponde con toda la información del caso exceptuando las evidencias, se puede definir como $I = T - E$.

Disponemos de dos categorías, dos hipótesis que a priori desconocemos el valor correcto y desconocemos cuales son las probabilidades de ocurrencia de cada una de ellas:

- H_p : Pertenecer a la misma fuente (target).
- H_d : Pertenecer a fuentes distintas (non-target).

El objetivo es determinar la probabilidad de que la evidencia pertenezca o no al sospechoso y poder determinar una decisión con un manejo adecuado de la incertidumbre. Sin embargo, esta no es tarea de los peritos técnicos, sino más bien de los jueces instruidos en el caso. Por ello la labor de este proyecto es generar una solución que permita analizar las evidencias procedentes de la biometría para que éstas puedan ser interpretadas como pruebas evaluables y se pueda definir así una decisión científica que sea transparente.

Desarrollando las probabilidades de interés usando las leyes de la probabilidad.

$$P(H = H_p|T) = P(H = H_p|I, E) = P(H_p|I, E) \quad 4-3$$

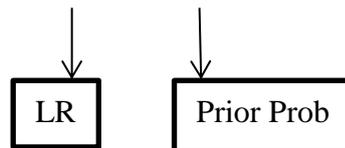
$$P(Hp|I, E) = \frac{P(E|Hp, I) * P(Hp|I)}{P(E|I)} \quad 4-4$$

$$= \frac{P(E|Hp, I) * P(Hp|I)}{P(E|Hp, I) * P(Hp|I) + P(E|Hd, I) * P(Hd|I)} \quad 4-5$$

Se considera que $P(Hp|E, I) = 1 - P(Hd|E, I)$

Dividiendo una probabilidad por su contraria:

$$\frac{P(Hp|E, I)}{P(Hd|E, I)} = \frac{P(E|Hp, I) P(Hp|I)}{P(E|Hd, I) P(Hd|I)} \quad 4-6$$



Para analizar la tasa anteriormente descrita que relaciona las probabilidades de las dos hipótesis teniendo en cuenta toda la información se divide la decisión en dos partes:

- LR: Likelihood ratio, se corresponde con las relaciones de verosimilitud, analizar las evidencias considerando las hipótesis complementarias. En un caso forense ésta sería la tarea del **perito**.
- Prior Prob: Referidas a las probabilidades de cada hipótesis a priori, sin tener en cuenta las evidencias. En el caso forense esta parte es la que le corresponde al **juez**.

En definitiva, resulta útil el uso de relaciones de verosimilitud debido a que no podemos estimar la probabilidad de cada hipótesis conociendo toda la información del caso, y para ello dividimos el problema en dos valoraciones diferentes, invirtiendo el orden de los parámetros. De esta manera se puede estimar, analizando las evidencias y teniendo en cuenta la opinión a priori.

4.1 Errores contemplados

Cuando se toman decisiones categóricas se pueden cometer errores, que deben ser contemplados. En este caso, al ser 2 las categorías disponibles, los errores son:

- Decidir que las evidencias son de la misma fuente cuando en realidad no lo son, se trata de una falsa aceptación (FA). En el caso forense, declarar culpable a un inocente.
- Decidir que no son de la misma fuente cuando en realidad si lo son, falso rechazo, (FR). En el caso forense, declarar inocente a un culpable.

Teniendo en cuenta que se trata de casos forenses en el que se intenta localizar el culpable, la importancia de cometer estos errores no conlleva las mismas consecuencias. Es por ello que se pueden establecer costes o penalizaciones, en este caso resulta más grave cometer una FA que un FR ya que resulta más grave que un inocente acabe en la cárcel.

Pero el seguimiento de este proyecto se va a centrar en el cálculo de los LR calibrados y no en la selección de los costes asociados a cada decisión.

4.2 Evaluación mediante Likelihood Ratio, LR

Calcular LR como relaciones de verosimilitud proporciona información que representa los grados de peso o de apoyo que tiene una hipótesis frente a la otra. En ningún momento se toma una decisión basada únicamente en LR sino que complementa la decisión del juez, puede ser [4]:

- Apoyando su misma hipótesis, lo cual dará más seguridad en el momento de la toma de decisión.
- Apoyando lo misma hipótesis pero con un grado muy pequeño. Esto quiere decir que cuanto mayor sea el LR más apoya la hipótesis, y por el contrario cuanto menor sea el LR menos apoya la hipótesis de manera que no se considere muy relevante la opinión extraída del perito.
- Apoye la hipótesis contraria. Puede dar lugar a contradicciones ya que la información recogida del caso lleva a decisiones muy distintas de las que se llegan con las pruebas científicas .Puede resultar igualmente útil ya que se puede interpretar de dos maneras; planteándose si las pruebas son de suficiente relevancia como para replantearse el caso, o por el lado contrario que las pruebas no puedan ser consideradas en el caso por su falta de veracidad.

En resumen, gracias al LR se puede definir una decisión científica que permita evaluar las pruebas en un juicio. Aunque no sea una condición suficiente, se pretende que pueda resultar necesaria o complementaria.

4.2.1 SCORE VS LR

No se debe confundir las puntuaciones score con las LR. Ambas son puntuaciones que extraemos del sistema biométrico que se está estudiando, sin embargo lo que de ellas se puede interpretar es muy distinto [4].

El score se obtiene de la extracción de características de las evidencias, la manera de obtenerlo puede ser muy variada, no sigue un mismo modelo para todos los sistemas biométricos por lo que no presenta una uniformidad a la hora de manejarlos después.

Esto supone un problema en cuanto a que no permite tratar estas puntuaciones de manera general o directa. Es decir, a partir de la obtención de un score aislado no se pueden establecer decisiones (Score=20, no se puede interpretar cuanta similitud presentan las evidencias, pero si se tienen dos: Score1=20 y Score2=100, se interpreta que el 2 presenta mayor similitud que el 1). Se necesitan entonces dos puntuaciones para poder establecer comparaciones entre ellas o tener un umbral definido previamente sobre el cual realizar las comparaciones con el resto de puntuaciones (Si el umbral es 200, el score=20 presenta poca similitud, sin embargo, con el umbral en 5 el score=20 presenta más similitud que antes). Este umbral solo sería válido para ese sistema, por lo que los scores solo podrían ser comparables entre puntuaciones obtenidas del mismo.

Para solucionarlo se propone el cálculo mediante relaciones de verosimilitud, LR.

- Interpretación de un score: Cuanto mayor sea el score más refuerza la hipótesis de pertenecer a la misma fuente, y viceversa, cuanto menor sea el score más refuerza la hipótesis de pertenecer a fuentes distintas. Sin embargo, no se puede determinar un peso probabilístico a partir únicamente de un score, sino que necesitamos de un grupo de scores para establecer comparaciones ya que no disponemos de una referencia generalizada como se mostrará posteriormente. Si se usa un score en el mecanismo de evaluación de los LR, las probabilidades que se obtienen tras las evidencias pueden ser engañosas.
- Interpretación de un LR: Se usa del mismo modo que el score para medir la similitud. Cuanto mayor sea más apoya la hipótesis de misma fuente y cuanto menor sea más apoya la hipótesis de fuentes distintas. Pero además tiene interpretación probabilística por sí mismo, diferenciador entre hipótesis, usando un proceso de decisión bayesiana. Lo que lo diferencia del score es la función que desempeña esta puntuación. Ya que es válida por sí misma, no necesita

compararla con otros LR para establecer una valoración (LR=3 significa “apoyo 3 a 1 la hipótesis de pertenecer a la misma fuente”).

4.2.2 Probabilidad basada en Posibilidades

Para comprender mejor el funcionamiento de LR recurrimos al concepto de “posibilidad” [4]. Definimos a priori la posibilidad de que pertenezca a la misma fuente como la probabilidad de que pertenezca frente a la probabilidad de que no pertenezcan a la misma fuente. De la misma manera se está definiendo la posibilidad de que pertenezca a la misma fuente tras analizar las evidencias.

$$O(Hp) = \frac{P(Hp)}{P(Hd)} \qquad O(Hp|E) = \frac{P(Hp|E)}{P(Hd|E)} \qquad 4-6$$

Teniendo en cuenta que $O(Hp) = 1/O(Hd)$ y $O(Hp|E) = 1/O(Hd|E)$ se relaciona las posibilidades antes y después de las evidencias recurriendo al LR como:

$$O(Hp|E) = LR * O(Hp) \qquad 4-7$$

De esta manera queda reflejado que en dos casos distintos con mismo LR, presentando distintas posibilidades a priori pueden dar lugar a decisiones distintas tras analizar las evidencias. Al mismo tiempo que dos casos distintos pueden dar lugar a la misma decisión tras analizar las evidencias con distintos pesos (LR), incluso aunque las posibilidades a priori sean las mismas.

De esta manera parece más intuitivo comprender la relevancia que puede presentar el análisis de las evidencias para cambiar la decisión de un caso judicial. Así se tienen en cuenta ambos factores, antes y después de analizar las evidencias, evaluando al final la posibilidad que queda al combinar ambas decisiones.

4.2.3 Precisión del LR: Calibración y Poder de discriminación

Se busca conseguir una buena precisión centrándose en estos dos conceptos, calibración y poder de discriminación. Para ello se busca la máxima optimización manteniendo el poder de discriminación, de esta manera se consigue mejorar la calibración.

4.2.3.1 Poder de discriminación

Se refiere a poder de discriminación como la capacidad que tiene el sistema para diferenciar una hipótesis de la otra. En este caso diferenciar de cuando las evidencias provienen de la misma fuente, son genuinas, o cuando vienen de fuentes distintas, impostoras.

Para poder estimar el poder de discriminación representamos en una gráfica la probabilidad de falsa aceptación frente a la de falso rechazo. Esta representación se conoce como Curva DET (Detection Error Tradeoff o curva de compensación por error de detección).

Se persigue que la curva esté lo más cercana posible al origen de coordenadas en el cual la probabilidad de error de ambos errores es cero.

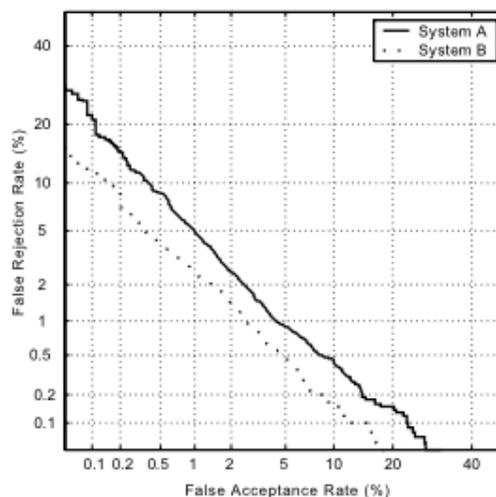


Figura 4-1 Ejemplo de curva DET

En el ejemplo de la figura 4-2, el poder de discriminación proviene del sistema biométrico empleado, es decir, no se puede modificar mediante el modelo estadístico. Hay que trabajar con ello, mejorar la precisión sin modificar el poder de discriminación ya que no depende del método de cálculo aplicado sino del sistema del que provienen los scores.

Para poder realizar la evaluación acudimos al algoritmo PAV (Pool Adjacent Violators), el cual nos dará una estimación de cuáles son los valores de LR óptimos a los que el sistema puede llegar, es decir, cuál sería el mejor valor que se puede alcanzar. El punto que se alcanza después de aplicar PAV ya no se puede mejorar porque el sistema biométrico no lo permite, para mejorarlo habría que estudiar la manera de mejorar el proceso de adquisición de las evidencias, o el proceso de cálculo del score, lo cual no es tarea de este proyecto.

4.2.3.2 Calibración

Con el fin de mejorar la precisión del método de cálculo aplicado sin modificar el poder de discriminación del sistema surge la necesidad de calibrarlo. Para ello, se intenta que los datos utilizados sigan una distribución parecida al modelo con que se vaya a calcular posteriormente las puntuaciones definitivas, el más común es una distribución gaussiana.

Por este motivo se busca que los score obtenidos de las comparaciones realizadas con la base de datos sigan una distribución gaussiana. Como a priori la gaussiana no se ajusta a la distribución de los datos se recurre a transformaciones previas de los scores. Siguiendo esta línea se probarán distintas transformaciones para analizar con cuales el sistema se consigue mejorar, y por tanto, calibrar.

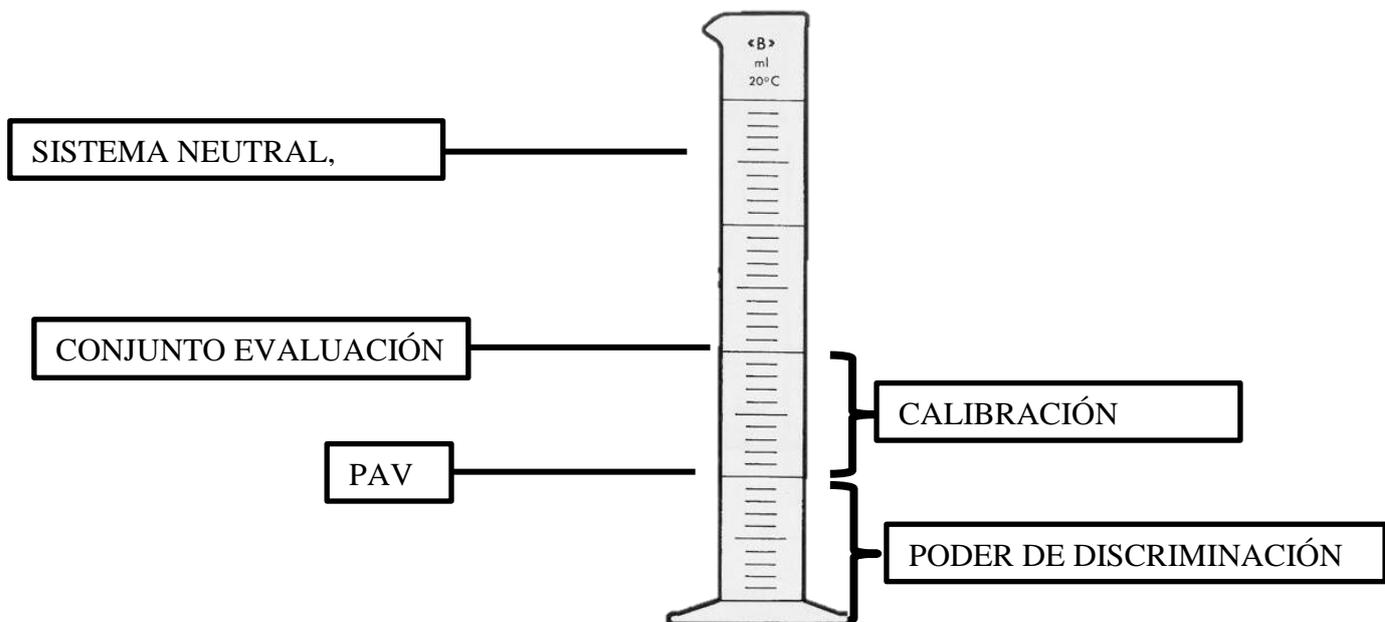


Figura 4-2 Ejemplo del rendimiento de un sistema mediante una probeta

4.3 Normalización de los scores

Se normalizan según el modelo usado y la teoría de decisión bayesiana para que todas las puntuaciones puedan ser tratadas con el mismo criterio.

Estas transformaciones tienen que ser lineales, crecientes o decrecientes, se debe aplicar la misma transformación a todos los scores, tanto a los obtenidos en comparaciones

genuinas como en comparaciones impostoras para que su tratado posterior sea ecuaníme.

De la base de datos obtenemos dos grupos de scores, los que se usan para entrenamiento y los que se usarán para realizar pruebas sobre el sistema entrenado.

Con los datos de entrenamiento se ajustan las transformaciones, en este proyecto se proponen tres. A continuación se muestra cómo queda la distribución antes y después de las transformaciones. En el capítulo 6 de experimentos se adjunta el proceso por el que se han escogido las transformaciones y el criterio seguido.

1. $-\log(\text{score})$
2. $-\log(\log(\log(\text{score})))$
3. $1/(\text{score}-0.5)$

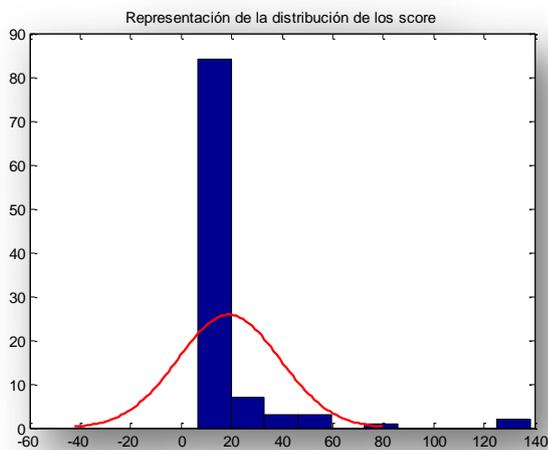


Figura 4-4 Scores target sin normalizar

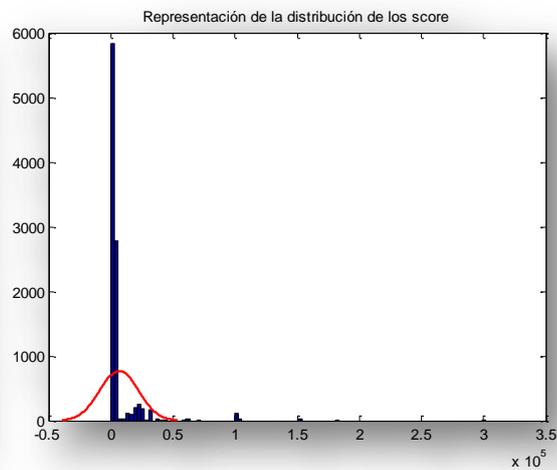


Figura 4-3 Scores non-target sin normalizar

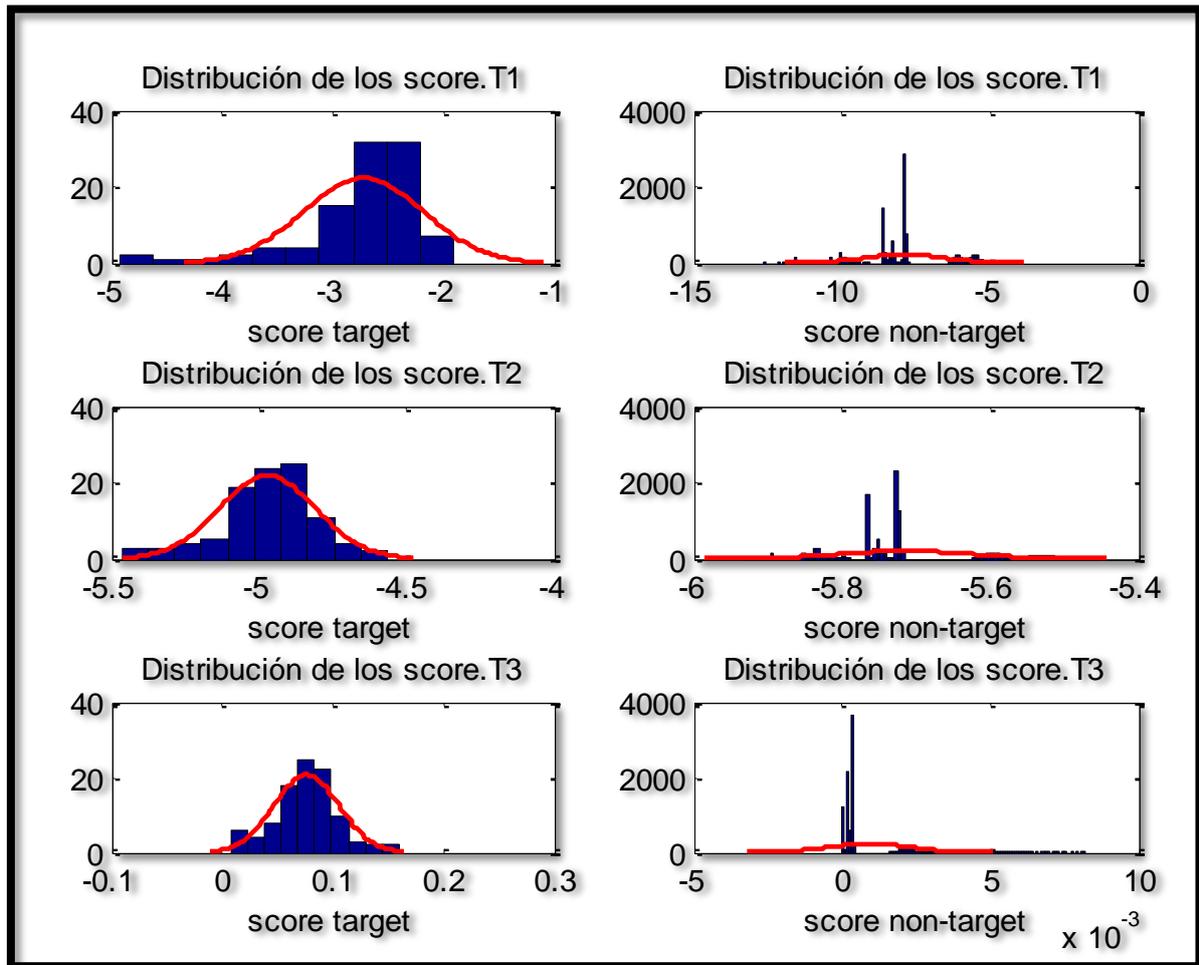


Figura 4-5 Distribución de los score tras aplicar las transformaciones 1, 2 y 3

5. Métodos de Cálculo de LR

A partir de las puntuaciones obtenidas de los scores normalizados se calcula el LR para poder valorar el peso de las evidencias forenses.

Para este proyecto se han usado tres métodos de cálculo de LR a partir de los scores obtenidos tras las transformaciones elegidas para la calibración.

El objetivo es encontrar un modelo que maximice la similitud de la distribución de los datos con la distribución que sigue el modelo, generalizando bien para nuevos datos.

Para proceder al cálculo de LR dividimos la base de datos a usar en dos grupos: Un primer grupo de entrenamiento y un grupo posterior de test, de prueba. Dentro de cada grupo se realizan dos tipos de comparaciones de pares de evidencias, entre evidencias pertenecientes a la misma fuente, y entre evidencias no pertenecientes a la misma fuente, enfrentando una contra todas las huellas que son de fuentes diferentes.

Entrenamiento	ScoreTargetTrain	100 comparaciones
	ScoreNonTargetTrain	9900 comparaciones
Test	ScoreTargetTest	158 comparaciones
	ScoreNonTargetTest	24806 comparaciones

El procedimiento será el mismo para los tres modelos, en primer lugar se entrena el sistema con los datos de entrenamiento obteniendo los valores necesarios que definan el modelo usado. Éstos serán los que se usen en un segundo paso. Este paso es el de probar con los valores obtenidos del entrenamiento los datos de test y evaluar el rendimiento del sistema en último lugar.

1. **Modelo Gaussiano**
2. **KDF**
3. **Regresión logística**

5.1 Modelo Gaussiano

Una de las distribuciones más comunes es la distribución normal. Esta distribución permite modelar el conjunto de datos definidos como una probabilidad.

$$P(E|\mu_{target}\sigma_{target}) = \frac{1}{\sqrt{2\pi}\sigma_{target}} e^{-\frac{(x-\mu_{target})^2}{2\sigma_{target}^2}} \quad 5-1$$

$$P(E|\mu_{nontarget}\sigma_{nontarget}) = \frac{1}{\sqrt{2\pi}\sigma_{nontarget}} e^{-\frac{(x-\mu_{nontarget})^2}{2\sigma_{nontarget}^2}} \quad 5-2$$

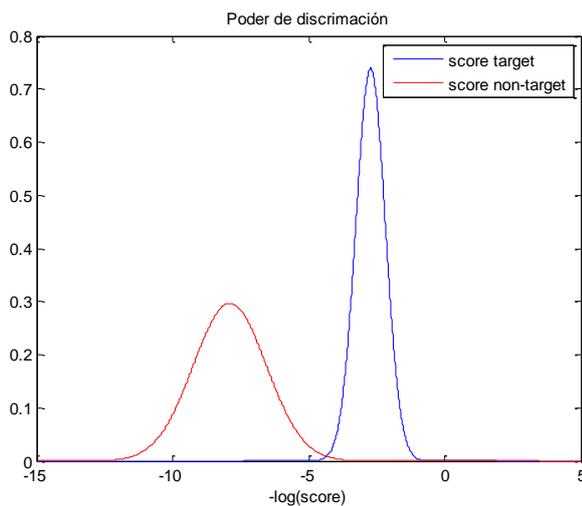


Figura 5-1 Poder de discriminación, modelo gaussiano

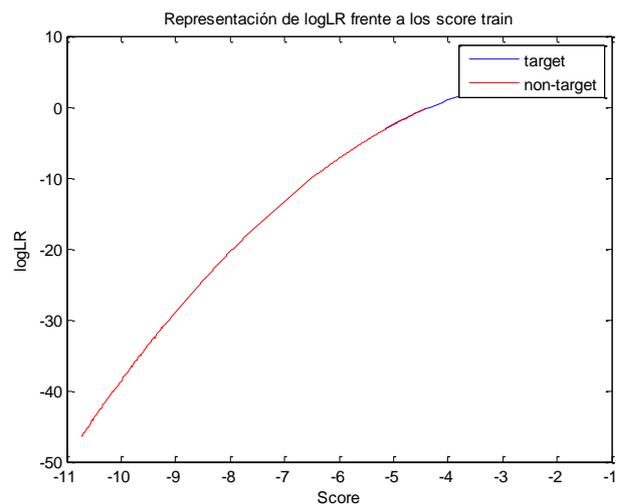


Figura 5-2 Representación Score-log (LR)

Ésta sería la representación de cómo se distribuyen los datos para cada hipótesis siguiendo distribuciones gaussianas y la transformación de obtener log-LR a partir de score para un par de densidades fijas.

Una vez obtenidos los parámetros que definen la distribución, en este caso dos, la media y la varianza, para cada hipótesis se calcula el LR de la siguiente forma para un score E:

$$LR = \frac{P(E|\mu_{target}\sigma_{target})}{P(E|\mu_{nontarget}\sigma_{nontarget})} \quad 5-3$$

Este método se ajusta bien a los datos si las distribuciones son gaussianas, además es sencillo computacionalmente hablando ya que solo son necesarios dos parámetros para caracterizar el conjunto de datos. Además, para un mejor manejo se trabaja en escalas logarítmicas que permite tratar todo el conjuntos de datos de forma unificada y visualizar todos los datos a pesar de sus grandes diferencias.

5.2 KDF

Este método trata de ajustarse más a los datos que se tienen, en este caso, a los score. Sigue un modelado gaussiano, sin embargo, se usa cuando los datos presentan una forma gaussiana repartida formando varias campanas. Por ello, trata de crear varias ventanas, concretamente una ventana gaussiana para cada score de entrenamiento, de manera que defina distribuciones gaussianas entorno a todos los datos. Por lo tanto, el objetivo es definir el kernel, es decir, la ventana gaussiana que represente los datos, fija para todos los datos de entrada.

A continuación se muestra un ejemplo de cómo funciona, a la izquierda se encuentra la distribución de los datos y a la derecha se muestra en rojo los kernel, las ventanas aplicadas a cada dato y en azul la totalidad de ellas. Además el área que abarca la función de densidad debe ser 1.

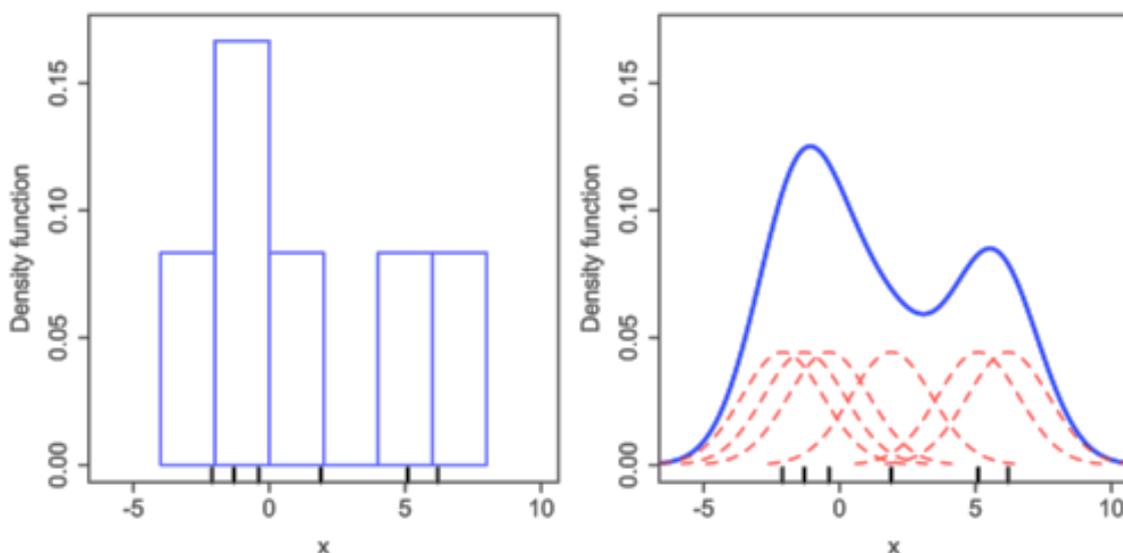


Figura 5-3 Modelo basado en KDF [22]

Para la base de datos usada en este proyecto, la distribución de los datos de entrenamiento non-target no pueden modelarse fácilmente de forma gaussiana, y por lo tanto se motiva el uso de KDF.

A pesar de que el ajuste es más exacto tiene el inconveniente de que presenta una alta complejidad computacional. Los parámetros que se necesitan calcular son los que definen a una distribución gaussiana, la media y la covarianza, para cada uno de los scores a evaluar.

Además se puede cometer un sobreajuste, y es que si la precisión es muy exacta se corre el riesgo de que algún score observado fuera del conjunto de datos de entrenamiento se trate de forma errónea.

Un ejemplo de transformación de score a log-LR, en este caso quedaría de la siguiente forma:

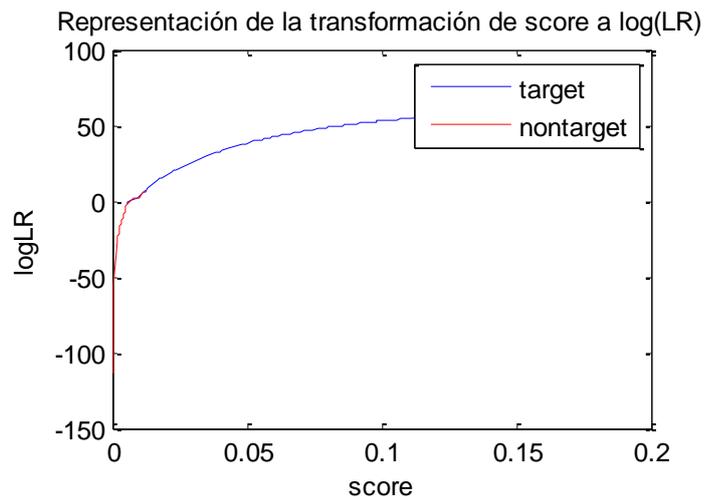


Figura 5-4 Representación log LR-Score

5.3 Regresión Logística

En los métodos discriminativos se busca la máxima similitud de los datos condicionada a la hipótesis que definan.

El propósito de este método es buscar una función de coste con el objetivo de encontrar los valores que minimicen dicha función. De esta manera nos acercaremos a los resultados deseados.

En este caso, hay dos tipos de regresiones, regresión lineal cuando la variable a tratar es continua, y regresión logística cuando la variable a tratar toma dos posibles valores. En este caso se trata de un caso de regresión logística ya que disponemos de dos valores, dos hipótesis, que pertenezcan a la misma fuente (target) y que pertenezcan a fuentes distintas (non-target).

Nuestros dos casos son:

- muestras que pertenecen a la misma fuente: Target $y = 1$.
- muestras que pertenecen a fuentes distintas: Non-Target $y = 0$.

Función de coste:

$$J(w) = \frac{1}{m} \sum_{i=1}^m (\text{cost}(h_w(x), y)) \quad 5-4$$

Siendo m el número de ejemplos y $hw(x)$ la hipótesis que quedaría definida de forma polinómica $hw(x) = w_0 + w_1x_1 + w_2x_2 + \dots + w_Nx_N$.

De manera que x son las características que definen nuestra hipótesis (score), N el número de características y w los parámetros que las definen. Sin embargo, para la hipótesis en los casos de regresión logística se emplea la función sigmoidea definida entre 0 y 1.

$$g(z) = \frac{1}{1 + e^{-z}} \quad Hw(x) = g(hw(x)) = \frac{1}{1 + e^{-(w_0 + w_1x_1 + \dots + w_nx_n)}} \quad 5-5$$

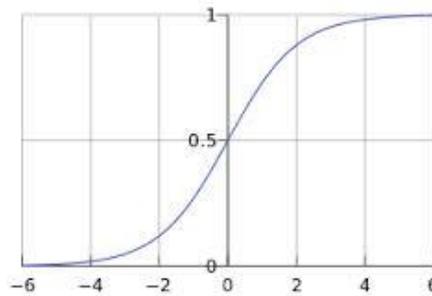


Figura 5-5 Función sigmoidea

Por lo tanto la hipótesis se puede interpretar como la probabilidad de que pertenezcan a la misma fuente conociendo los datos de las características y sus valores.

$$Hw(x) = P(Hp|X, W) = g(hw(x)) \quad 5-6$$

Al igual que hacíamos antes para el cálculo del LR aplicando relaciones de verosimilitud lo aplicamos de nuevo para la regresión logística.

$$\frac{P(Hp|E, I)}{P(Hd|E, I)} = \frac{P(E|Hp, I) P(Hp|I)}{P(E|Hd, I) P(Hd|I)} \quad 5-7$$

$$\frac{P(Hp|E, I)}{P(Hd|E, I)} = LR * Odd \quad 5-8$$

$$\frac{P(Hp|E, I)}{1 - P(Hp|E, I)} = LR * Odd \quad 5-9$$

$$P(Hp|E, I) = (1 - P(Hp|E, I)) * LR * Odd \quad 5-10$$

$$P(Hp|E, I) * (1 + LR * Odd) = LR * Odd \quad 5-11$$

$$P(Hp|E, I) = \frac{LR * Odd}{1 + LR * Odd} = \frac{1}{1 + (LR * Odd)^{-1}} = \frac{1}{1 + e^{-\log(LR) - \log(Odd)}} \quad 5-12$$

Si lo relacionamos con la expresión anterior con que definíamos la hipótesis:

$$w_0 + \sum_1^m w_i * x_i = \log(LR) + \log(Odd) \quad 5-13$$

Por lo tanto la función de coste define dos posibles hipótesis, en este caso, que pertenezcan a la misma fuente y que no pertenezcan a la misma fuente.

Cuando Hp es cierta, $y=1$, por el contrario cuando Hd es cierto el menor valor de la función de coste se da en 0. Si cuando Hp es cierta se deduce que no pertenecen a la misma fuente, la función de coste es elevada, y por lo tanto el error cometido mayor.

De esta manera la función de coste queda expresada:

$$J(w) = -\frac{1}{m} \left[\sum_{i \in target} \log(P(Hp|x_i, w)) + \sum_{j \in nontarget} \log(P(Hd|x_j, w)) \right] \quad 5-14$$

$$J(w) = -\frac{1}{m} \sum_{i=1}^m y^{(i)} \log(h_w(x^{(i)})) + (1-y^{(i)}) \log(1 - h_w(x^{(i)})) \quad 5-15$$

OBJETIVO: Encontrar $\{w_0, w_1, \dots, w_n\}$ que hagan que $P(Hp|E)$ esté lo más cercana a 1 cuando se trata de un caso de misma fuente, y lo más cercana a 0 cuando se trate de un caso de fuentes distintas.

Para encontrar los parámetros adecuados se pueden utilizar distintos algoritmos como por ejemplo, el de gradiente descendiente o el de gradiente conjugado. Para el de gradiente descendiente por cada iteración se desciende por la función de coste hasta lograr el mínimo. Este algoritmo es adecuado usarlo cuando el número de ejemplos de entrenamiento es elevado. El de gradiente conjugado también es un método iterativo útil con sistemas grandes y para métodos de optimización. Como en este caso necesitamos

encontrar los parámetros óptimos para la función de coste se va a realizar mediante el algoritmo de gradiente conjugado ya implementado [12].

Una vez tenemos los parámetros óptimos que definen el comportamiento, realizamos la transformación de score a LR con los datos de test. Para ello usamos los parámetros obtenidos con los datos de entrenamiento.

Recordando la relación existente entre los LR y la expresión que define una regresión logística extraemos los valores de los LR.

$$\text{Hipótesis} = P(Hp|X, W) = g(w_0 + w_1x_1 + \dots + w_mx_m) \quad 5-16$$

$$= g(\log LR + \log(Odd))$$

$$\log(LR) = w_0 + w_1x - \log(Odd) \quad 5-17$$

Siendo w_0 y w_1 los parámetros obtenidos a partir de los scores de entrenamiento, siendo 'x' los scores y $\log(Odd)$ las probabilidades a priori. Como estas últimas no podemos saberlas realizamos el cálculo de LR para un valor concreto, por ejemplo $Odd=1$.

Finalmente la transformación de score a LR quedaría así:

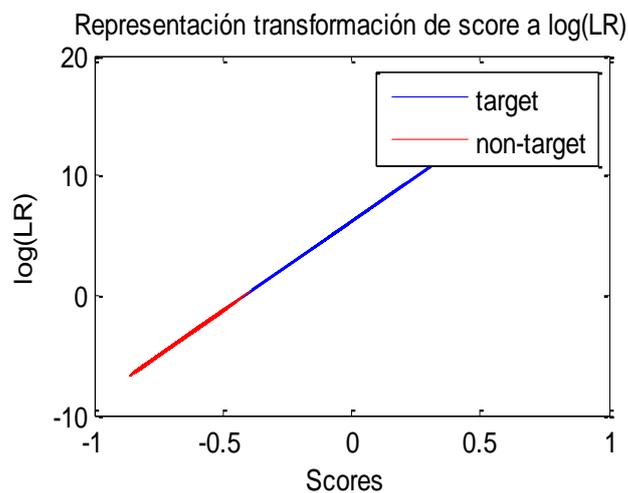


Figura 5-6 Representación log-LR-Score

5.4 Evaluación del rendimiento de LR

5.4.1 ECE PLOT. Empirical Cross-Entropy

Para determinar el rendimiento de los LR, realizamos una **representación con curvas ECE** (Empirical Cross Entropy). Esta curva nos da información del poder de discriminación del sistema biométrico utilizado y de la calibración obtenida para los LR calculados. Se encarga de calcular la entropía, la incertidumbre que presentan los datos con respecto al resto de información presente en un caso. Cuanto mayor sea el valor de la entropía, mayor incertidumbre presenta y por lo tanto se necesitaría mas información añadida para poder estimar una decisión. Por el contrario, si presenta valores pequeños de ECE, hay una incertidumbre baja, es decir, se necesita poca información añadida para determinar la hipótesis.

La expresión que lo define es la siguiente:

$$ECE = -\frac{P(Hp|I)}{Np} \sum_{i=Np} \log_2 P(Hp|E_i, I) - \frac{P(Hd|I)}{Nd} \sum_{i=Nd} \log_2 P(Hd|E_i, I) \quad 5-18$$

Para ello lo definimos en función de los valores de LR y de las probabilidades a priori.

$$ECE = \frac{P(Hp)}{Np} \sum_{i=1}^{Np} \log_2 \left(1 + \frac{1}{LRi * Odd} \right) + \frac{P(Hd)}{Nd} \sum_{j=1}^{Nd} \log_2 (1 + LRj * Odd) \quad 5-19$$

Como desconocemos las probabilidades a priori que provienen del juez, hemos de probar para una variación de éstas. De esta manera necesitamos definir las probabilidades de que las evidencias pertenezcan o no a la misma fuente y hacer el cálculo para cada valor.

$$Odds(Hp) = P(Hp)/P(Hd) \quad 5-20$$

Para la representación se tienen 3 curvas;

- En primer lugar representamos la curva ECE para el sistema en el que LR=1 siempre. Este sistema no aporta ninguna información útil ya que no tiene poder de discriminación. Es un **sistema neutral** por lo que lo tomaremos de referencia como el caso peor, es decir, cualquier valor que se sitúe por debajo de esta

curva estará obteniendo resultados mejores y cuanto más alejado esté menos información se necesitará y mejor será el resultado.

Desarrollo de la expresión para LR=1:

$$ECE_{peor} = \frac{P(Hp)}{Np} Np * \log_2 \left(1 + \frac{1}{Odd} \right) + \frac{P(Hd)}{Nd} Nd * \log_2(1 + Odd) = \tag{5-21}$$

$$= P(Hp) * \log_2 \left(\frac{1 + Odd}{Odd} \right) + (1 - P(Hp)) * \log_2(1 + Odd) =$$

$$= P(Hp) * [\log_2(1 + Odd) - \log_2(Odd - \log_2(1 + Odd))] + \log_2(1 + Odd) =$$

$$ECE_{neutral} = -P(Hp) * \log_2(Odd) + \log_2(1 + Odd) \tag{5-22}$$

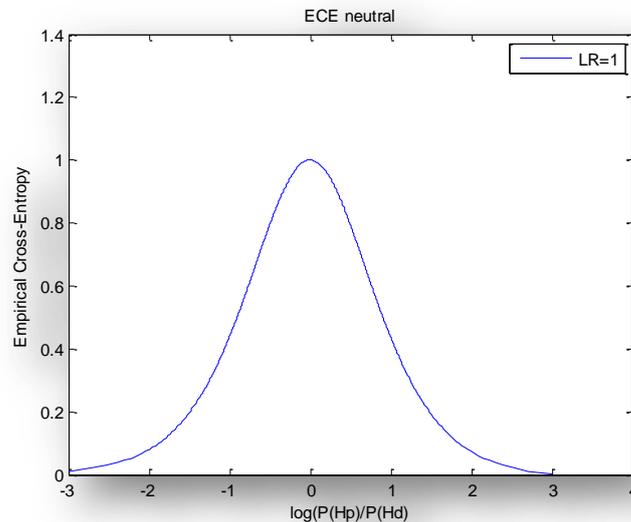


Figura 5-7 ECE neutral

Necesitamos probarlo para distintos valores posibles de probabilidades a priori, y a partir de estas calcular las probabilidades de cada hipótesis.

$$P(Hp) = \frac{LR * Odds}{1 + LR * Odds} \tag{5-23}$$

Es importante tener en cuenta que manejamos escalas logarítmicas para simplificar los cálculos.

- En segundo lugar representaremos la curva ECE con el conjunto de LR reales, los obtenidos de las pruebas de test. Nos indica cual es la entropía del conjunto

de evaluación. Cuanto más alta sea esta curva peor es el sistema en cuanto a que necesita mayor información para poder tomar una decisión.

Desarrollo de la expresión para el caso real:

$$ECEp = \frac{P(Hp)}{Np} * \sum_{i=1}^{Np} \log_2 \left(1 + \frac{1}{LRpi * Odd} \right) =$$

5-24

$$P(Hp) * \text{mean}(\log_2 \left(1 + \frac{1}{LRp * Odd} \right))$$

$$ECEd = \frac{P(Hd)}{Nd} * \sum_{j=1}^{Nd} \log_2 \left(1 + \frac{1}{LRdj * Odd} \right) =$$

5-25

$$P(Hd) * \text{mean}(\log_2(1 + LRd * Odd))$$

$$ECE = ECEp + ECEd$$

5-26

- Por último se representa el valor más bajo que se puede alcanzar. Se corresponde con los valores óptimos del sistema, es el límite inferior al que el sistema puede llegar para mejorar su calibración sin modificar el poder de discriminación. Estos valores se consiguen aplicando el algoritmo **PAV (Pool Adjacent Violator)**. La diferencia entre la curva del conjunto de evaluación y la curva obtenida tras PAV nos indica la calibración del sistema.

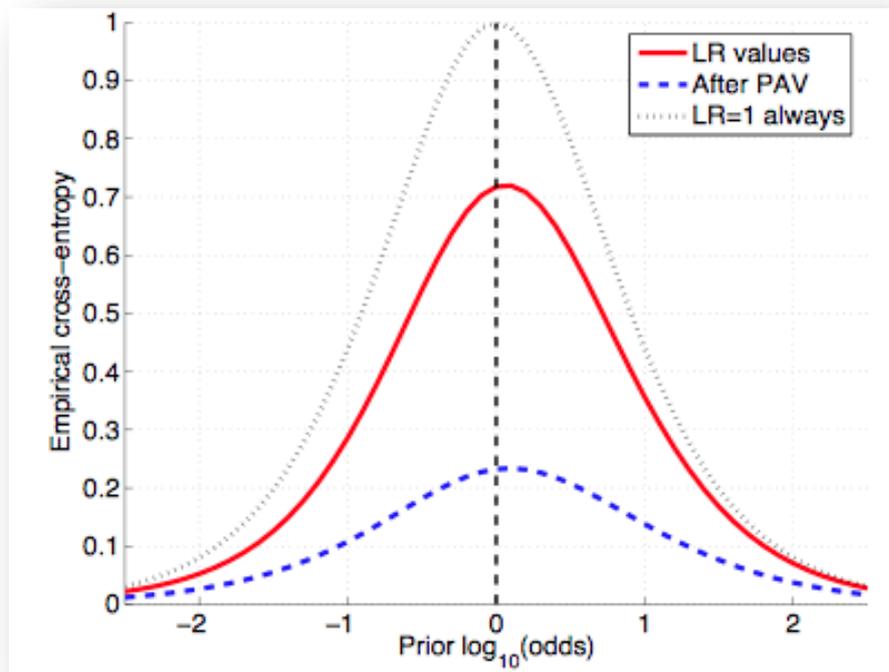


Figura 5-8 Representación de las curvas de entropía ECE

El inconveniente que presenta esta representación es que desconocemos las probabilidades a priori, por ello, se realiza la representación sobre los posibles valores de éstas. Sin embargo, esto es adecuado en un caso forense, en el que precisamente las probabilidades a priori no son conocidas. Esta es la gran ventaja de este tipo de curvas frente a otros métodos de medida del rendimiento.

A partir de las ECE plot se puede interpretar que un sistema es bueno cuando presenta pequeños valores de ECE, ya que indica que se necesita poca información para que la decisión tomada sea fiable. Por el contrario, es importante también para determinar si el método utilizado es inadecuado cuando alguna región del conjunto de evaluación presenta una curva ECE mayor que la del sistema neutral. Quiere decir que el resultado es peor que si no se analizaran las evidencias. Por tanto, el sistema utilizado no es bueno o no vale para ese rango de probabilidades. En ese caso sería conveniente intentar mejorar la calibración del sistema o descartar la evaluación de las evidencias. Esto último puede ocurrir cuando las condiciones de las evidencias son malas porque estén en mala calidad.

En definitiva, la representación de la curva ECE resulta intuitiva ya que por lógica, cuando tienes mucha información de algo, la opinión que das tiene mucho más valor, ya sea para bien o para mal. Pero cuando no tienes muchos datos en los que basarte, no estás seguro de ello, además de que necesitas más información para llegar a una conclusión, la opinión que das no apoya demasiado ni a una ni a otra hipótesis.

5.4.2 NCE. Normalized Cross-Entropy

La principal desventaja de las curvas ECE es que la proporción visual del centro de la gráfica con respecto a los extremos no es realista para su interpretación. Por eso se recurre a las curvas NCE.

Mediante representaciones NCE se consigue ver lo que ocurre en los extremos del eje x de una curva ECE, sobre todo cuando el rango de representación es grande. Como referencia todas las comprobaciones y comparaciones que se realizan son tomando el punto del centro, en el que ambas probabilidades son las mismas, sin embargo, esas conclusiones no son del todo certeras. De esta manera no solo se puede evaluar en cero sino que se ve el comportamiento en todo el rango.

Para calcularla se realiza la división de los datos del rendimiento para el caso real entre el rendimiento para el sistema neutral, $LR=1$. De esta manera se consigue que las proporciones estén normalizadas a la comparación con el caso neutral de cada punto calculado.

Sin embargo, aun expandiendo la representación de las probabilidades a priori y a pesar de que tengamos más información de lo que pasa en los extremos, observamos que a veces el comportamiento no es el adecuado en algunas regiones, debido a la falta de datos conforme nos vamos acercando a los extremos de las probabilidades a priori. Para tener un conocimiento mayor de la fiabilidad de los datos de entrenamiento se podría restringir un intervalo para el cual el comportamiento es válido, y fuera de él el comportamiento no es muy estable.

5.4.3 Regla del 30 de doddington

Esta regla surge como solución al problema que se ha citado antes, establecer únicamente unas regiones para las que el comportamiento de los datos si es fiable, y fuera de dicha región no lo es.

Establece que para tener un 90% de confianza de que la tasa de error verdadera está dentro del $\pm 30\%$ de la tasa de error observada, debe haber al menos 30 errores.

Para que esta regla se cumple se tiene que definir el umbral adecuado para que el número de errores sea de al menos 30. En este caso tenemos dos posibles, que siendo target se considere non-target y viceversa, por lo tanto disponemos de dos umbrales

adecuados, es decir, queda definido un intervalo de confianza. Intervalo para el que los resultados obtenidos se pueden considerar válidos con una confianza del 90%.

- Con Θ_{up} definimos que sea máximo 30 el número de errores de considerar non target cuando lo es. $N(\text{falso rechazo}) \leq 30$
- Con Θ_{down} definimos que sea mínimo 30 el número de errores de considerar target cuando no lo es. $N(\text{falsa aceptación}) \geq 30$

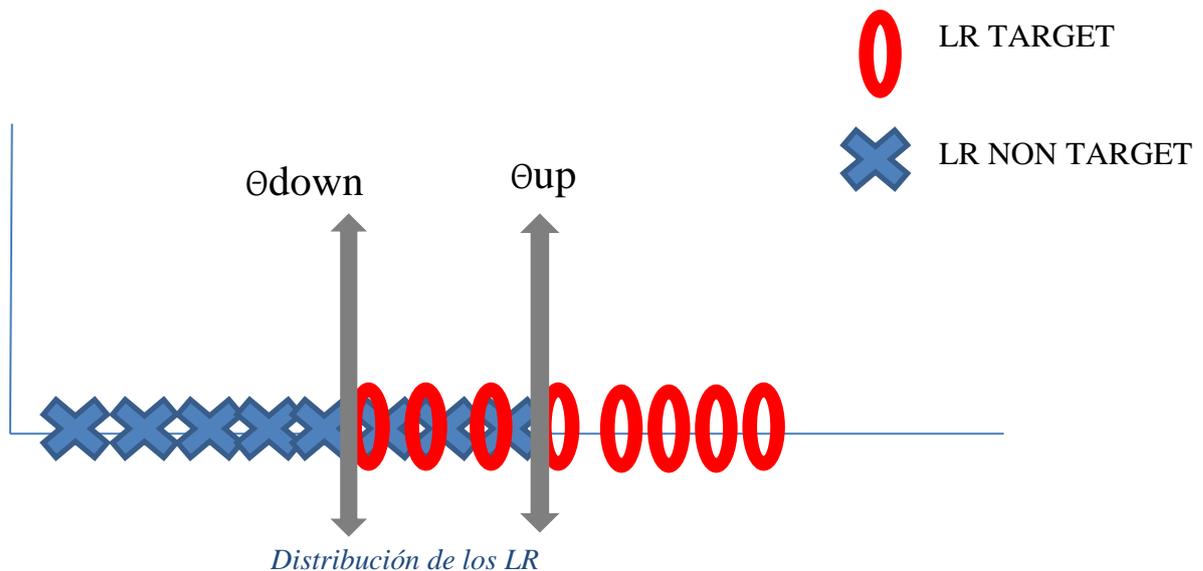


Figura 5-9 Representación del intervalo de confianza basado en la regla del 30 de dodginton. Sobre la distribución de los LR obtenidos se calcula el intervalo de confianza.

5.5 Criterios de decisión

Cuando se realizan las pruebas con la base de datos y a partir de los distintos escenarios propuestos lo que se propone es un estudio del rendimiento que se obtiene con dichos datos según los distintos escenarios.

Lo ideal sería establecer una única conclusión, determinar cuál de los métodos propuestos es el que mejor se comporta, el que mejores resultados obtiene además de ser el más fiable de todos. Sin embargo, esto no es posible, ya que son muchas las variables implicadas que determinan si el sistema es bueno. Además, en unos escenarios puede resultar óptimo y sin embargo, en otros pésimo. Por ello, como en toda decisión, existe un análisis y según este, depende de la variable que se estudie la conclusión cambia.

En definitiva en este análisis de decisión que se propone se realiza una matriz de objetivos por alternativas, en el que las alternativas son los distintos escenarios o modelos propuestos, y los objetivos son los criterios que se persiguen en la evaluación de los resultados.

Los objetivos en los que se basan los criterios de selección son los siguientes.

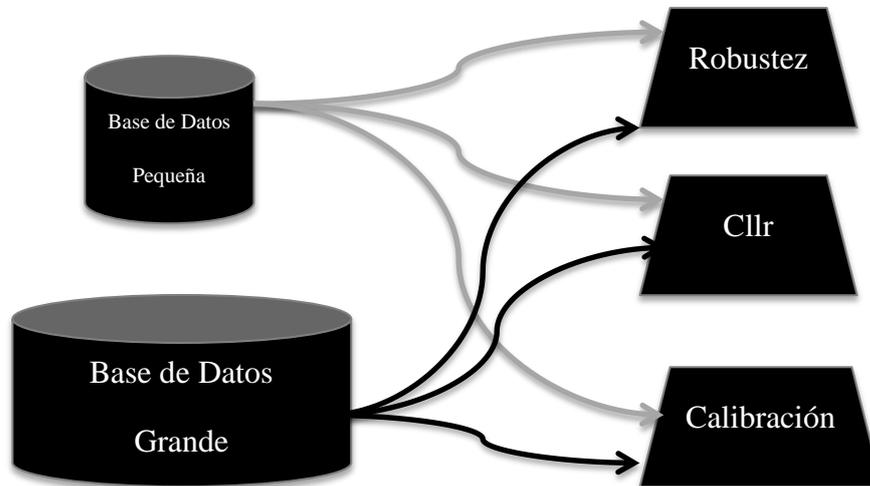


Figura 5-10 Objetivos para el criterio de decisión

- **Cllr:** El valor del Cllr es el punto medio en el eje x de las representaciones obtenidas en las ECE plot. Se define como el cálculo de entropía ECE cuando las probabilidades de ambas hipótesis son iguales. Cuando más bajo este valor mejor ya que estará mejorando el caso peor independientemente de lo que se obtenga tras el algoritmo PAV, ya que éste límite no se puede mejorar con el modelo aplicado.
- **Calibración:** Se entiende la calibración del sistema como el valor obtenido de realizar la diferencia entre el valor del Cllr y el de la representación obtenida con el algoritmo PAV. Cuanto menor sea este valor, más están los datos ajustados al valor óptimo para ese sistema.
- **Robustez:** Se define como el rango de variación de los valores de Cllr al probar con distintas implementaciones aleatorias de bases de datos del mismo tamaño. Cuanto menor sea, más robusto será el sistema, más fiable serán los valores cuando el sistema se ponga en funcionamiento.

La diferencia entre evaluar el Cllr únicamente y la calibración, está en que aunque PAV obtenga el mejor resultado alcanzable con la calibración, esto no quiere decir que el resultado que se vaya a obtener sea bueno, ya que si el sistema biométrico utilizado para la adquisición de datos es malo, los resultados con PAV serán igualmente malos. Por lo que prima más que el Cllr sea bueno, y para ello presente poca incertidumbre, a que la diferencia de éste con PAV sea pequeña.

Calibración de puntuaciones procedentes de sistemas biométricos

Además, la valoración de estos tres criterios cambia mucho según si se trata de una base de datos pequeña o grande, por lo que según sea el caso se establecerán distinciones.

- **Base de Datos Pequeña:** Cuando disponemos de pocos o insuficientes datos cómo para representar un comportamiento general. A falta de datos se elegirá cuál de las alternativas disponibles funciona mejor pese a esta mala situación. Se valora cual sería más estable y razonablemente buena.
- **Base de Datos Grande:** Cuando la falta de datos no es tanto problema, escogemos el modelo que mejor funcione en todos los sentidos, el óptimo, el más arriesgado ya que la consistencia de los datos va a ser buena y podemos arriesgarnos más.

En definitiva se persigue que tras los diferentes experimentos realizados se pueda rellenar la siguiente tabla, en lugar de dar un único modelo válido.

	Robustez	Calibración	Clr
BD Grande			
BD Pequeña			

Tabla 5-1 Objetivo a rellenar: Criterio para la selección de los modelos

6. Resultados Experimentales

En este capítulo se van a desarrollar los experimentos realizados durante el proyecto final de carrera. Los experimentos se han realizado sobre la base de datos creada para este proyecto al inicio de él. Esta base de datos cuenta con 258 pares de huellas dactilares. Estas huellas se han dividido en dos grupos, uno de ellos se utilizará para un entrenamiento previo de los datos, y el otro grupo se usará como prueba de los parámetros obtenidos con las huellas del primer grupo. De esta manera, servirá como simulación de un escenario real en el que pueda llegar cualquier huella desconocida para el sistema entrenado.

Para cada grupo se han realizado dos tipos de comparaciones, entre huellas que pertenecen a la misma fuente, las que se han recopilado en esta base de datos, y la segunda comparación se ha enfrentado a una huella dubitada contra el resto de las indubitados generando comparaciones entre huellas que pertenecen a fuentes distintas.

Entrenamiento	ScoreTargetTrain	100 comparaciones
	ScoreNonTargetTrain	9900 comparaciones
Test	ScoreTargetTest	158 comparaciones
	ScoreNonTargetTest	24806 comparaciones

Para cada una de estas comparaciones se ha obtenido una puntuación, un score, este proyecto final de carrera trabajará directamente desde estos scores. El objetivo es aplicar distintos métodos de cálculo de LR para realizar un estudio sobre cuál es el que mejor se adapta según las condiciones y los criterios aplicados.

En primer lugar se implementa y se aplica sobre los scores el desarrollo de los tres sistemas a utilizar, modelo gaussiano, kdf y regresión logística. Posteriormente y con el

fin de que los datos a utilizar, en este caso, scores, se ajusten mejor al modelo y a las hipótesis propuestas, se realizan pruebas para la normalización de los datos. Estas pruebas se consiguen mediante transformaciones para que una vez aplicadas las transformaciones se vuelvan a aplicar los tres modelos y comprobar si ha habido mejoras en el rendimiento.

Para asegurarnos de que los datos son estables a condiciones adversas y con el fin de solucionar algunos problemas que persisten se someterán los mismos estudios realizados pero esta vez con una base de datos más pequeña. De esta manera, se analizará el comportamiento de los modelos en situaciones críticas y se valorará cuál de ellos es más estable. Por último y para una mejor interpretación de los datos, sin dar lugar a confusiones, se proponen las curvas NCE únicamente para los modelos que mejores resultados han obtenido, así como analizar la robustez de cada uno.

6.1 Primera parte de los experimentos:

Implementación de los modelos estudiados.

6.1.1 Modelo Gaussiano

En primer lugar con los scores obtenidos de la base de datos se representan comparándolos con una distribución gaussiana típica.

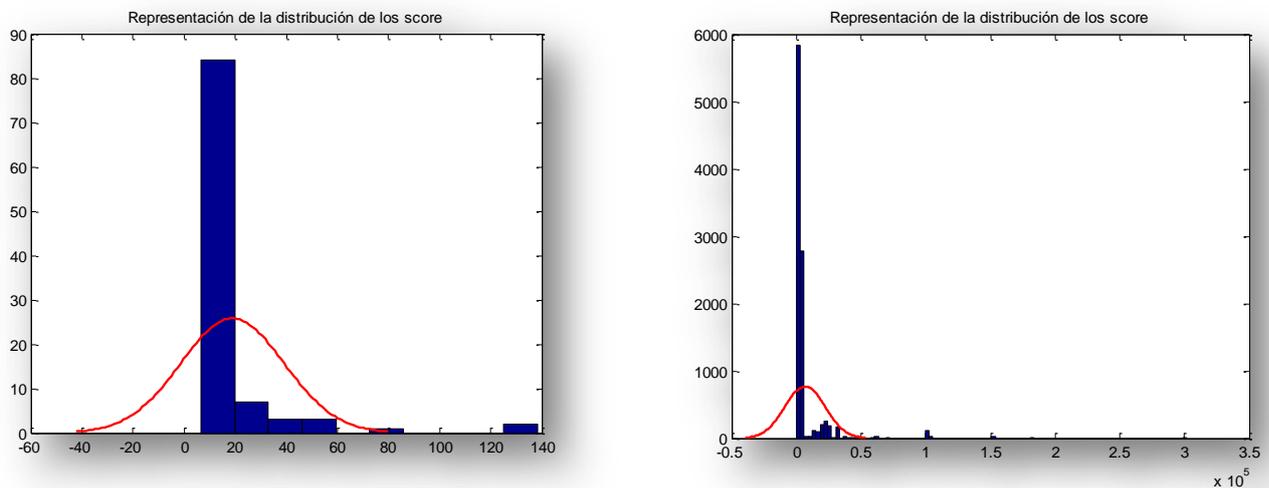


Figura 6-1 Distribución de los score target (izquierda) y non-target (derecha) respectivamente.

Como se puede ver a simple vista el comportamiento que presenta no guarda una estrecha relación con la gaussiana de referencia. En el caso de los target puede ser debido a que no se tienen suficientes datos para representarlo y en el caso de los non-target puede deberse a que los datos están muy concentrados en un punto en concreto.

Para poder comparar el rendimiento de este sistema acudimos a las curvas ECE. Como se puede observar no se obtiene ninguna mejora con respecto al caso peor (LR=1), por lo que este sistema no aporta ninguna información relevante. Puede ser debido a que se estén tratando los datos con un sistema gaussiano cuando su comportamiento no es gaussiano.

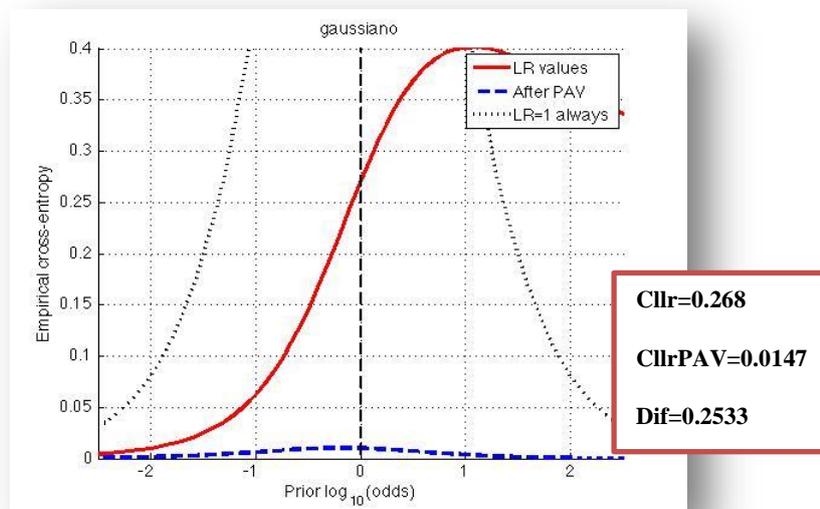


Figura 6-2 ECE plot en modelo gaussiano sin transformación

Aplicando el modelo sobre los scores directamente no se obtienen buenos resultados, ya que para los LR obtenidos la curva es peor que el caso neutro en un rango de probabilidades. Además observando la distribuciones de la figura 6.1 no se cumple la interpretación correcta de LR: cuanto mayor sea el LR más apoya la hipótesis de pertenecer a la misma fuente y cuanto menor sea, menos la apoya. Lo que está ocurriendo es al revés las puntuaciones de las non-target son mayores que las puntuaciones target, las puntuaciones que definen las hipótesis están invertidas con respecto a la interpretación adecuada.

6.1.2 KDF

Si no aplicamos ninguna transformación a los scores de entrada, ésta sería su representación siguiendo un modelo KDF.

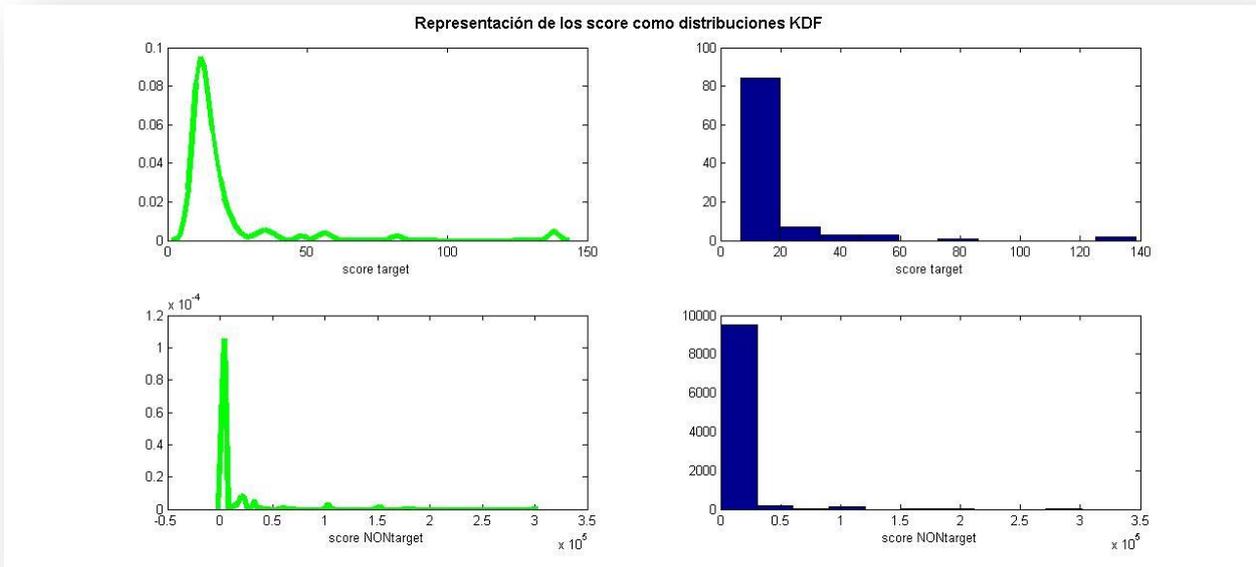


Figura 6-3 Distribución de los scores con el modelo KDF

Si observamos la distribución se puede apreciar que el modelo KDF se ajusta bastante a priori al comportamiento que siguen los scores. La distribución de los scores principalmente de los non-target define muy bien una distribución KDF, en la que se aprecian diferentes subcampanas y no una sola como ocurre en el modelo gaussiano.

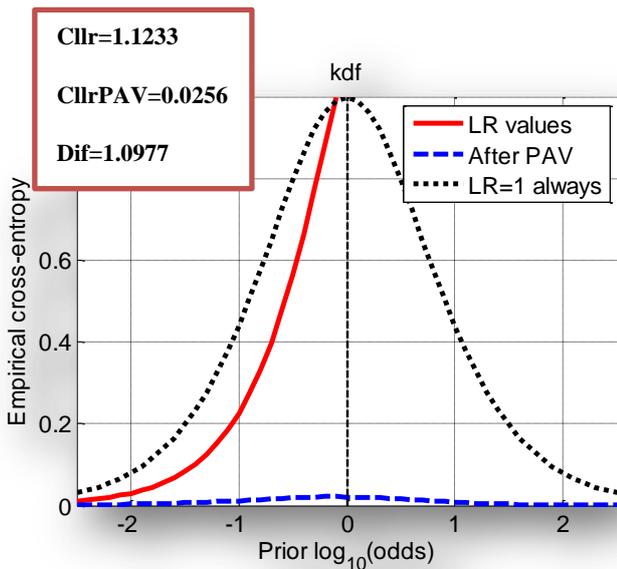


Figura 6-4 ECE plot score sin transformación modelo KDF

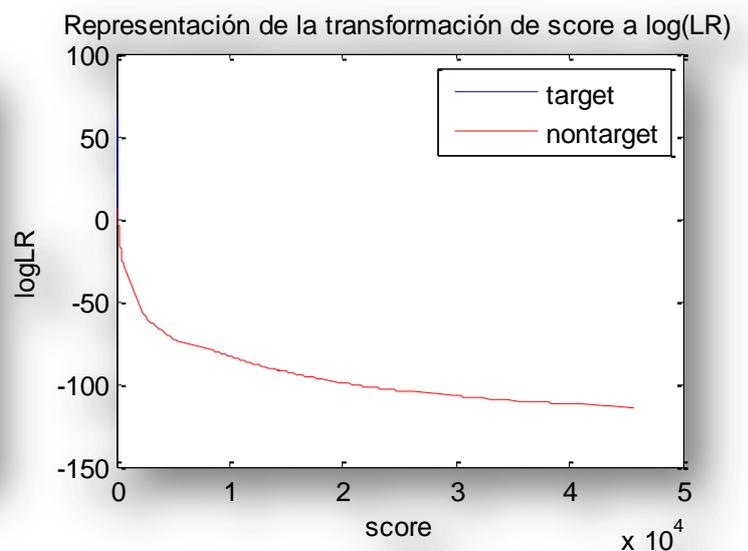


Figura 6-5 Transformación score-log (LR)

Los resultados son muy malos, superando el caso peor en la región de probabilidades positivas que se corresponde cuando la probabilidad de pertenecer a la misma fuente es mayor que la de pertenecer a fuentes distintas. Es decir, los resultados malos e deben más al caso de las puntuaciones target, puede deberse a que los datos target no definen una distribución acorde al modelo KDF.

6.1.3 Regresión Logística

Cuando utilizamos modelos discriminativos para calcular LR, como en el caso de la regresión logística, al proceso de transformar un score a LR se le suele llamar “calibración”. Mediante el uso de regresión logística se calculan los parámetros ‘w’ que minimicen la función de coste.

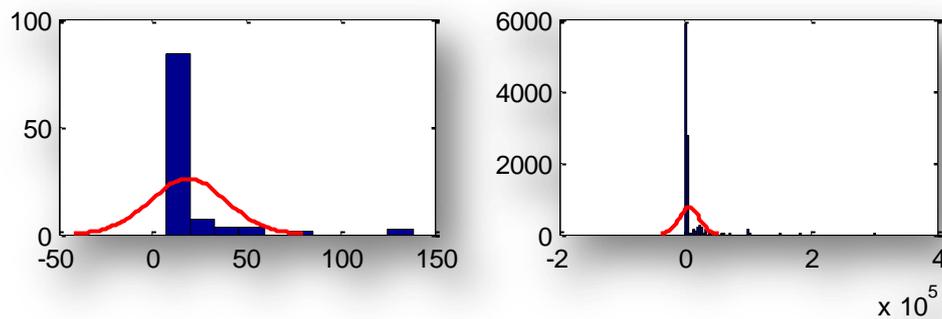


Figura 6-6 Distribución de los score sin transformación modelo Regresión Logística

- Mediante la función *train_llr_fusion* [12] generamos los parámetros óptimos para los score de entrenamiento. (Θ_0, Θ_1)
- Una vez definidos los parámetros, mediante la función *lin_fusion* [12] se realiza la transformación de los score de test a LR para cada una de las hipótesis. Ésta es la relación de score frente a log (LR).

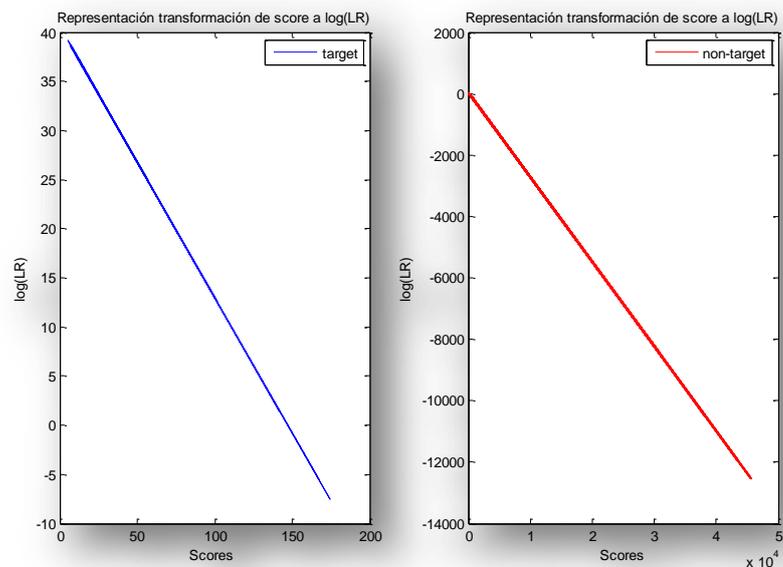


Figura 6-7 Transformación score-log (LR) con Regresión Logística

- Por último recurrimos a las ECE plot para comprobar el rendimiento de este método.

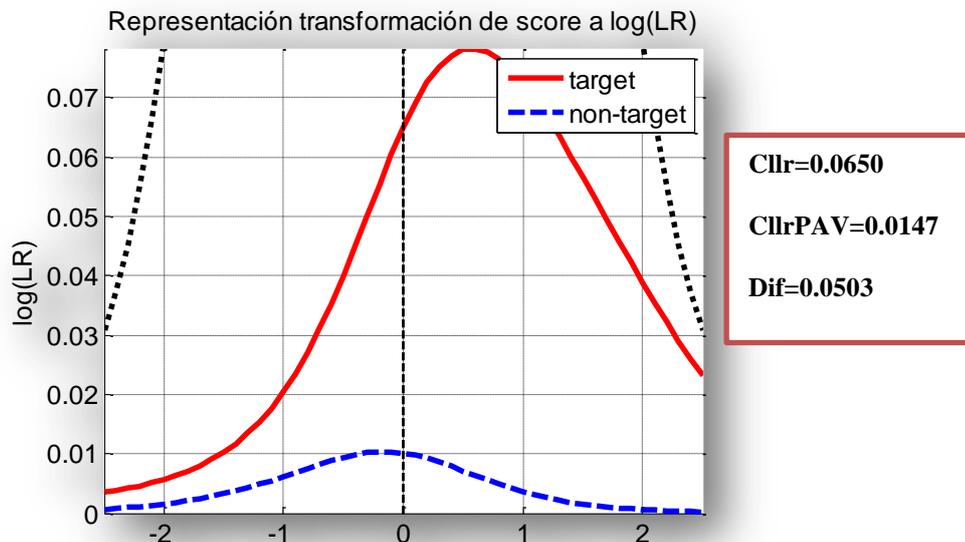


Figura 6-8 ECE plot score sin transformación con regresión logística

La curva de los LR es mucho más reducida y por tanto mejor que la del sistema neutral por lo que los resultados son buenos, sin embargo, hay una región de probabilidades a priori para los que los resultados no serían válidos. En el tramo de la derecha la curva para los LR reales es superior a la representación para el caso peor. Además, hay una separación con respecto a los valores más óptimos, lo cual indica que la calibración podría mejorarse notablemente.

6.1.4 Conclusión

Parece que la distribución KDF se ajusta bastante a los scores directamente sin necesidad de transformaciones, sin embargo, la entropía está muy cercana a 1, esto quiere decir que con este modelo no se aporta mucha información por parte de las evidencias ya que no mejora mucho el caso neutral, en el que no se analizan las evidencias. Además, en el modelo gaussiano y de regresión logística los resultados son buenos exceptuando la región de la derecha, en donde las probabilidades a priori son altas, que corresponde cuando la probabilidad de los target es mayor que los non-target. En ella se aprecian unas colas que sobrepasan los valores del sistema neutral, lo cual resulta muy negativo. Puede ser debido a que alguno de los scores no sean representados correctamente por la gaussiana utilizada ya que como se ve en la representación de su histograma en el caso de los target algunos de ellos se encuentran fuera de la campana de gauss desarrollada.

Por este motivo se propone buscar una transformación previa de los scores para lograr que adquieran de la forma más precisa posible el comportamiento de una gaussiana. Se proponen tres transformaciones. Además se probará si estas transformaciones mejoran

todavía más los resultados obtenidos o si por el contrario una distribución más gaussiana empeora los resultados de los otros modelos.

6.2 Segunda parte de los experimentos:

Normalización de los scores

Se propone la normalización de los scores para conseguir que éstos sigan una distribución gaussiana, para ello, mediante la ayuda del modelo gaussiano implementado se han probado transformaciones hasta que los scores de entrenamiento han mejorado su ajuste. Para las transformaciones propuestas es necesario que sean transformaciones monótonas, es decir, sin máximos ni mínimos, ya que éstas no cambian el poder de discriminación del sistema, únicamente afectan a la mejora de la calibración. Recordemos que el poder de discriminación venía limitado por el sistema biométrico utilizado en la adquisición por lo que no se debe modificar.

Para obtener la primera transformación se detalla el proceso seguido mediante el uso del modelo gaussiano:

La transformación propuesta es $T = -\log(\text{Score})$, de esta manera y con los datos de entrenamiento, la distribución quedaría más parecida a una gaussiana que en el caso anterior cuando no se utilizaba transformación.

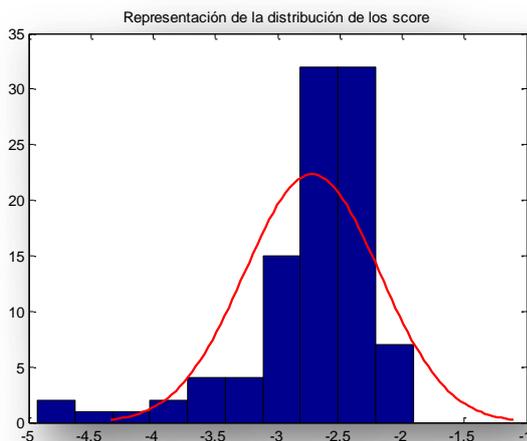


Figura 6-10 Distribución score target con transformación 1

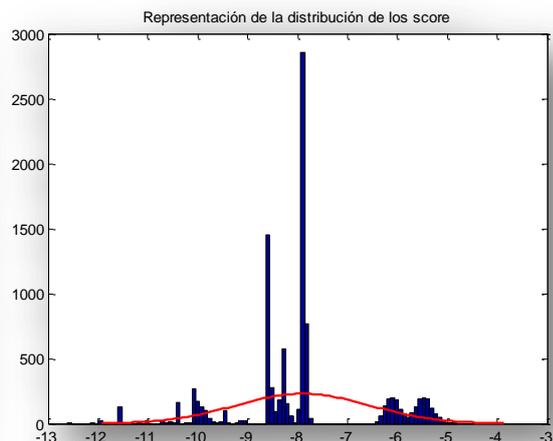


Figura 6-9 Distribución de los score non-target con transformación 1

Además, también cumple la hipótesis con la que tratamos al LR, cuanto mayor sea el LR más apoya la hipótesis de pertenecer a la misma fuente, y viceversa, cuanto menor sea menos apoya la hipótesis de pertenecer a la misma fuente.

Se puede ver en la siguiente figura la representación de los scores de ambas hipótesis con la transformación ya aplicada. Se observa a priori que presenta un buen poder de discriminación de las puntuaciones target con respecto a las non-target debido a la capacidad de diferenciar entre ambas.

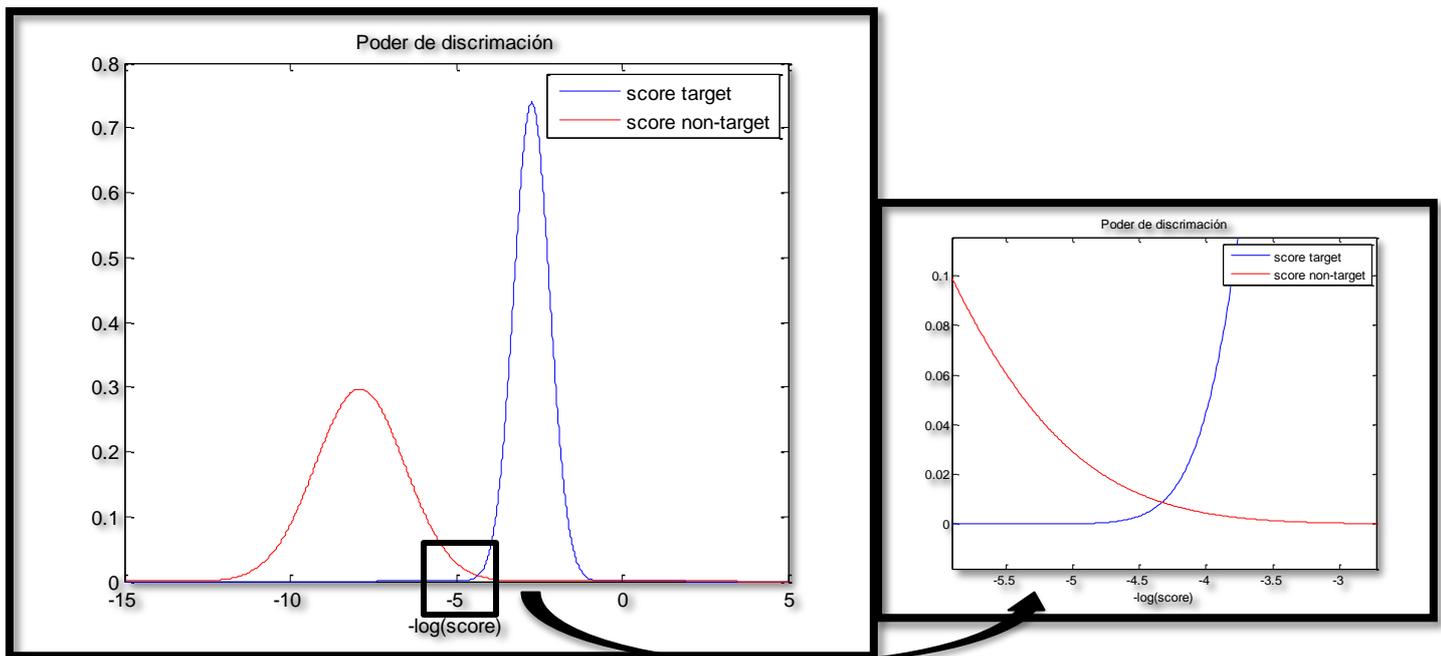


Figura 6-11 Poder de discriminación de los score con transformación 1

Representación de la transformación de los scores (target y non-target) a \log (LR) mediante cálculos bayesianos.

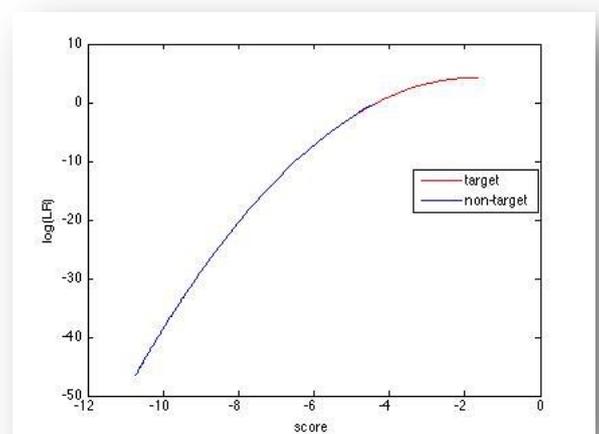


Figura 6-12 Transformación score-log (LR) gaussiano

Para comprobar el rendimiento del sistema representamos las curvas ECE para los LR obtenidos de los scores de test.

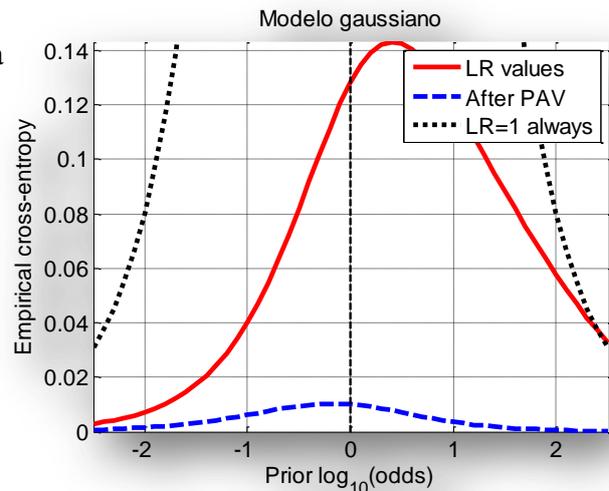


Figura 6-13 ECE plot score con transformación 1 gaussiano

Los resultados son bastante buenos y válidos, ya que se ha mejorado con respecto al caso peor en todo el rango de probabilidades a priori, por lo que nos aporta menor entropía, es decir, requerimos de menos información, casi la mitad de información menos. Sin embargo, se puede percibir un pequeño rango de probabilidades a priori en el que la curva roja es superior a la curva punteada negra, por lo que para esos valores el sistema no es adecuado. Esto puede deberse a que los datos utilizados no siguen de forma exacta una distribución gaussiana, como se aprecia en los histogramas.

Por un lado se ha conseguido mejorar el sistema gracias a la transformación realizada. Por otro lado notamos que el sistema no se acerca a la curva obtenida tras PAV, esto refleja una mala calibración. Por ello y teniendo en cuenta el rango de probabilidades a priori en el que se sale del caso neutro se propone una mejora en la calibración mediante una transformación que ajuste mejor los datos a una gaussiana.

Éstas son las tres transformaciones propuestas:

4. $-\log(\text{score})$
5. $-\log(\log(\log(\text{score})))$
6. $1/(\text{score}-0.5)$

Para cada transformación se ha representado cómo se distribuyen los scores target y non-target comparándolos con una distribución gaussiana típica y viendo cómo se acercan más a ella. Además gracias a estas tres transformaciones se consigue representar el planteamiento de la hipótesis de los scores y los LR, cuanto mayor sea esta puntuación más probabilidad hay de que pertenezca a la misma fuente, y viceversa.

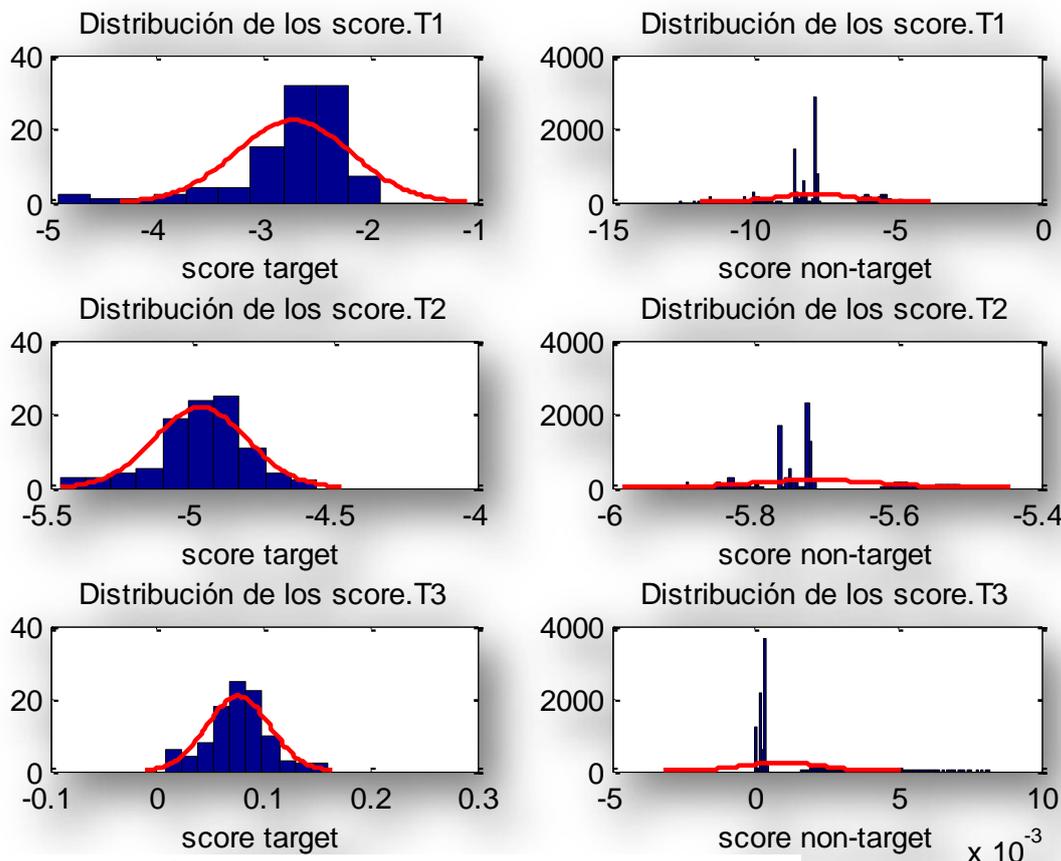


Figura 6-14 Distribución de los score con las tres transformaciones

El siguiente paso sería probar estas transformaciones en los tres modelos implementados, comprobar si funcionan correctamente en los tres modelos, o poder estimar cuál de las transformaciones se ajusta mejor a qué modelo.

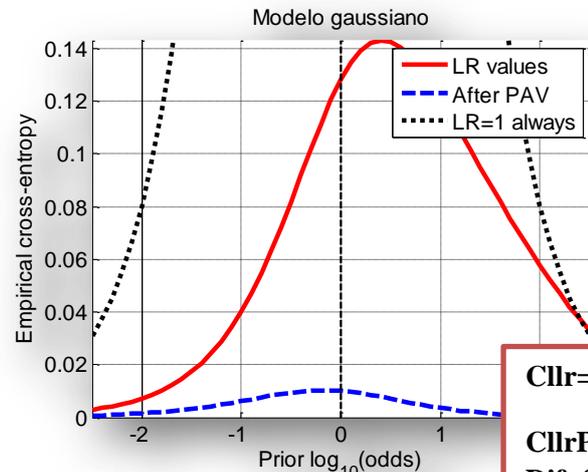
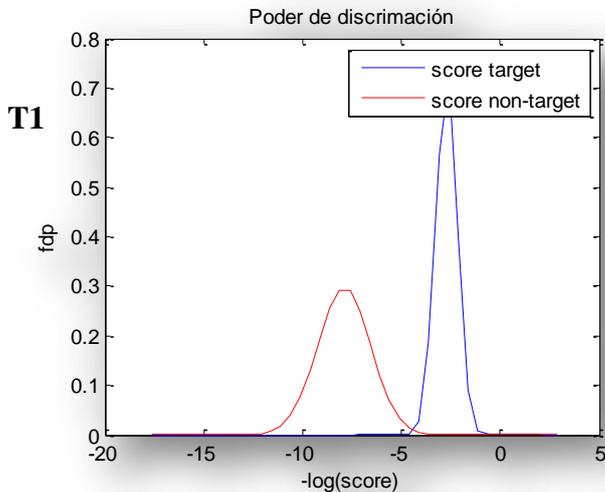
Se realizará el cálculo de LR para cada modelo partiendo de las tres transformaciones. Para poder apreciar las diferencias obtenidas al aplicar cada transformación sobre el mismo modelo se representará:

- El poder de discriminación de los scores normalizados mediante cada una de las transformaciones. El poder de discriminación ayudará a interpretar cómo se distribuyen las dos categorías y el carácter diferenciador que existe entre ellas.

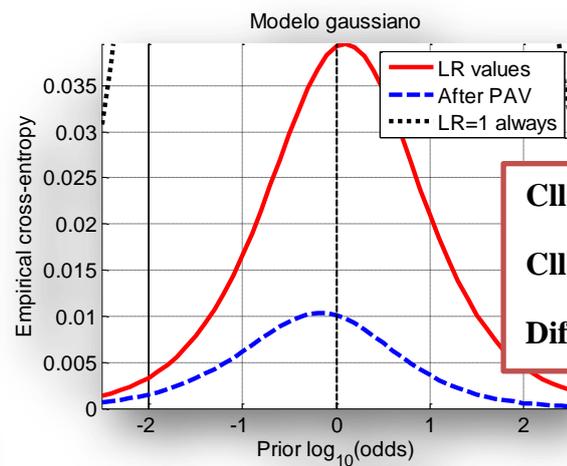
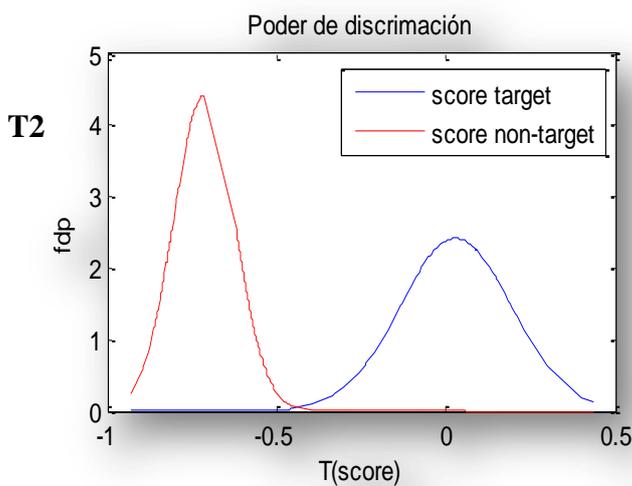
- Mediante curvas ECE se estudiará el rendimiento pudiendo obtener de ellas:
 - El valor Cllr.
 - El valor Cllr de los resultados que se obtienen tras PAV.
 - La diferencia del valor Cllr obtenido real con respecto al óptimo, nos da entender cuánto más se puede mejorar el sistema con respecto al mejor valor alcanzable. Para todas estas puntuaciones, cuanto menor, mejor.

Tras estas consideraciones se podrá realizar una comparación valorando cómo se comportan las distintas transformaciones al aplicar un mismo modelo, y poder establecer cuál de ellas es más adecuada para según qué modelo.

6.2.1 Modelo Gaussiano con transformación previa

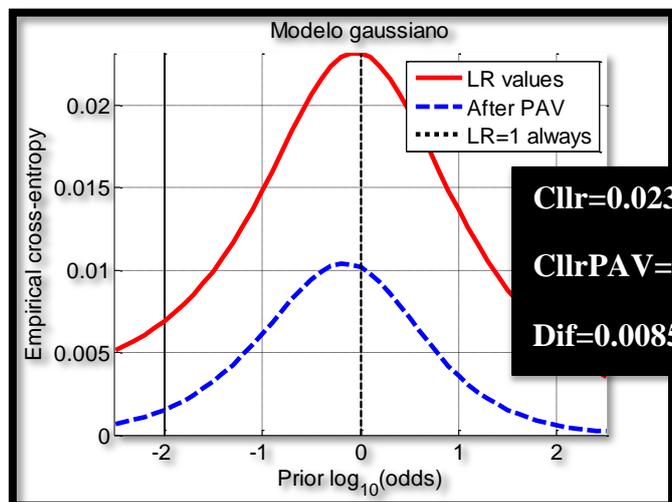
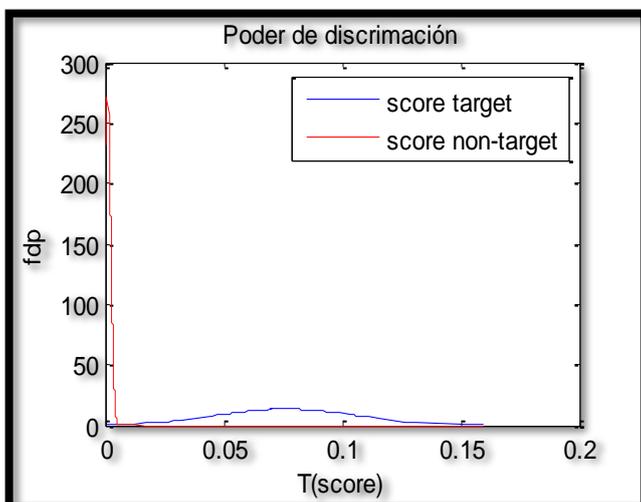


Cllr=0.1283
CllrPAV=0.0147
Dif=0.1136



Cllr=0.0393
CllrPAV=0.0147
Dif=0.0246

T3



Cllr=0.0232
CllrPAV=0.0147
Dif=0.0085

Figura 6-15 Poder de discriminación y ECE plot con la T1 en primera fila, T2 en la segunda y T3 en la tercera fila respectivamente en gaussiano

Comparación:

En este sistema el poder de discriminación se encuentra con un $Cllr=0.0147$, por lo que el objetivo es calibrar el sistema lo mejor posible para lograr estar lo más cerca posible de este valor, ya que representa los valores óptimos posibles que el sistema permite.

Si nos fijamos en la curva que obtiene la entropía para los LR obtenidos en el test, el mejor sistema funciona con la transformación 3, además de tener un valor de $Cllr$ más bajo, sirviéndonos de la representación de PAV, la diferencia entre la calibración y el poder de discriminación del sistema es menor en la transformación 3, lo cual indica una buena calibración de los datos.

Con respecto a la primera transformación aplicada, se ha mejorado en todo el rango de probabilidades a priori.

MODELO GAUSSIANO-> TRANSFORMACIÓN 3

6.2.2 Regresión logística con transformación previa

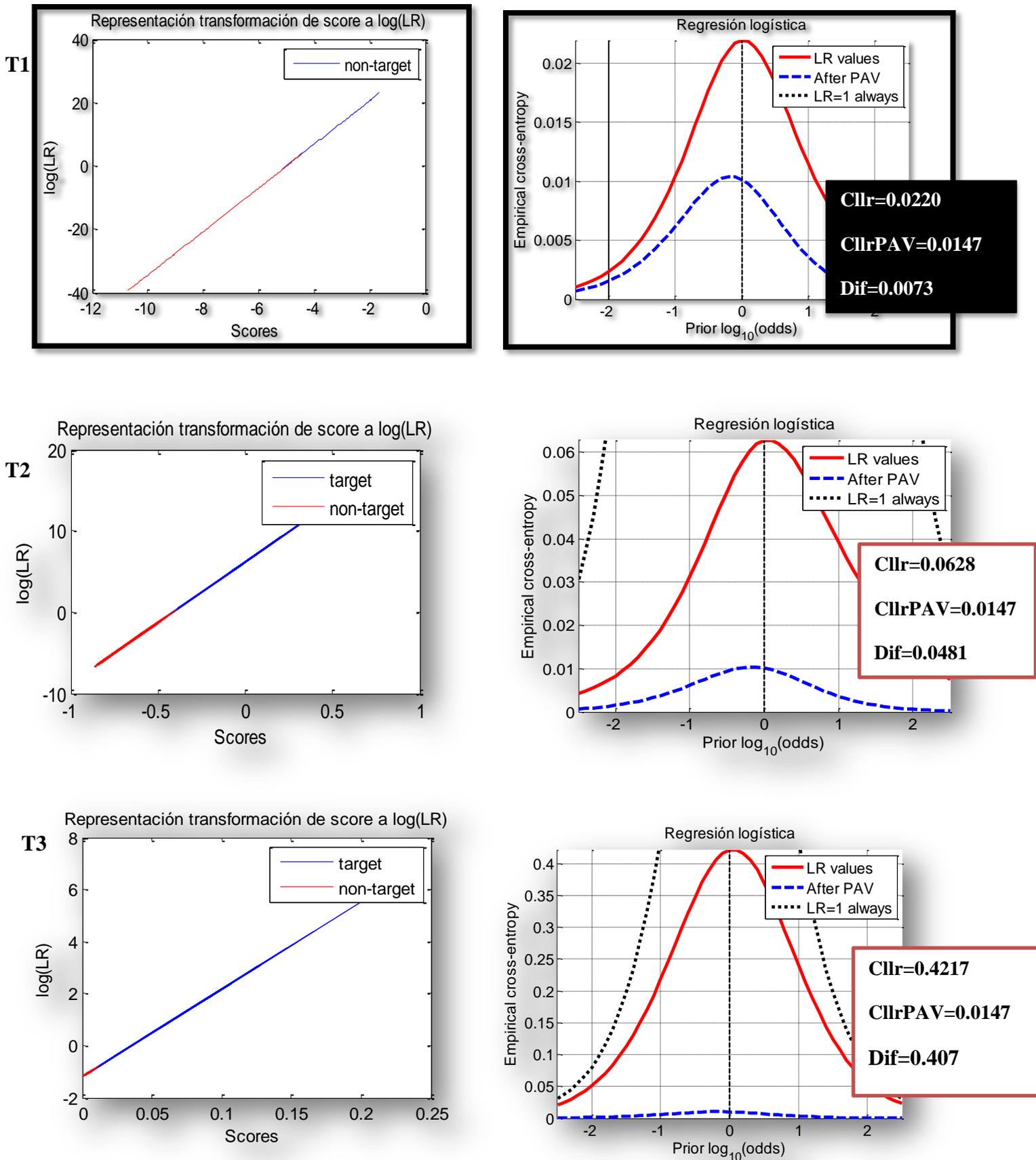


Figura 6-16 Poder de discriminación y ECE plot con T1, T2 y T3 respectivamente con regresión logística

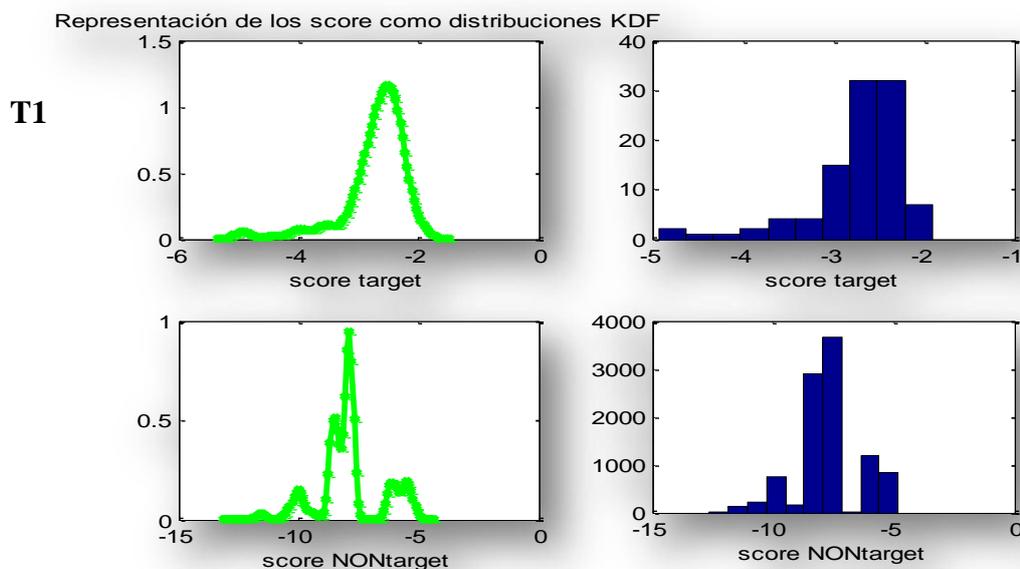
Comparación:

En este caso ocurre lo contrario que en el modelo gaussiano, cuanto mayor se aproximan los datos a una gaussiana, peor comportamiento presentan al tratarlos como un modelo de regresión logística. Por ello, de las tres transformaciones aplicadas, la primera es la que mejores resultados obtiene. Además de presentar un Cllr muy bajo, la diferencia con respecto a los valores óptimos es mínima, por lo que se deduce que con esta transformación la calibración es adecuada.

Si además lo comparamos con lo que obtuvimos sin aplicar ninguna transformación a los datos la mejora también es considerable, de 0.0650 a 0.022, consiguiendo además que todos los resultados se obtengan por debajo del peor caso.

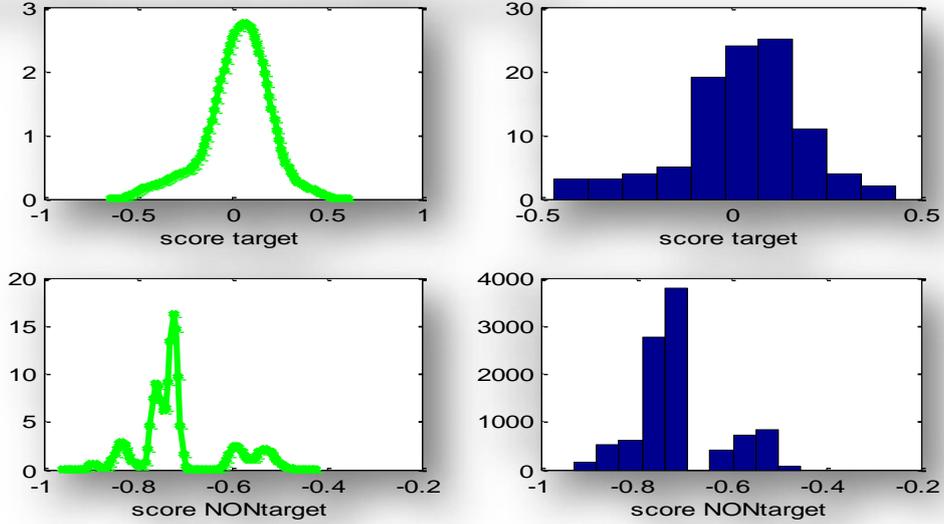
MODELO REGRESIÓN LOGÍSTICA-> TRANSFORMACIÓN 1

6.2.3 KDF con transformación previa



T2

Representación de los score como distribuciones KDF



T3

Representación de los score como distribuciones KDF

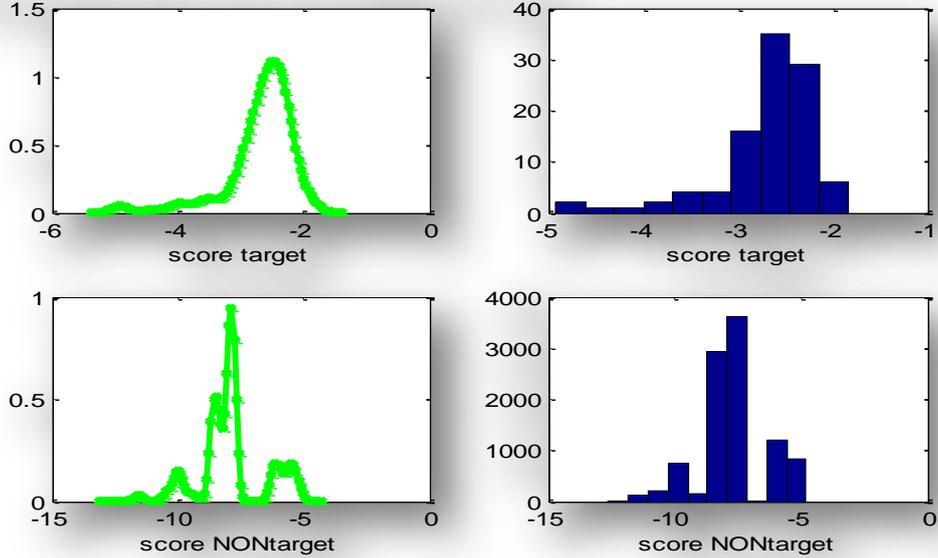
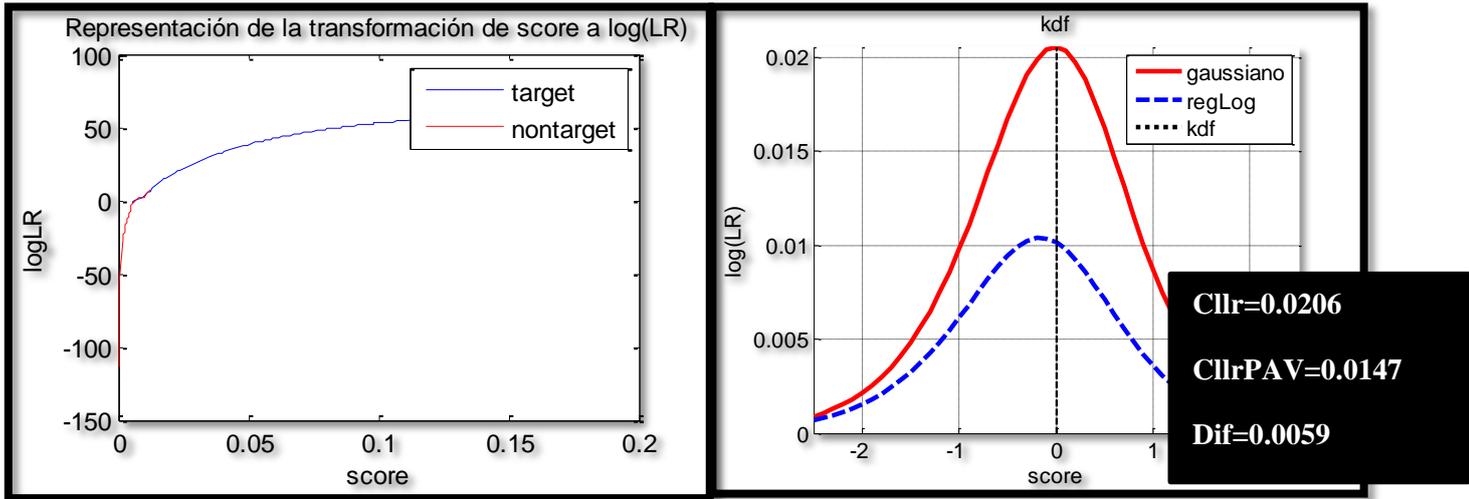
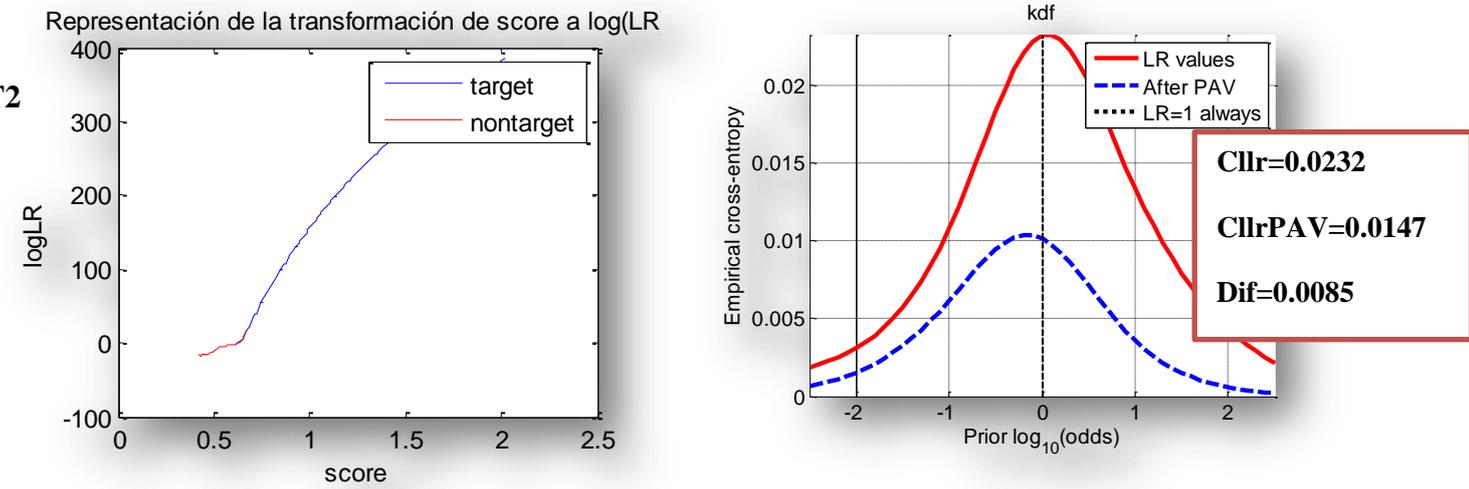


Figura 6-17 Distribución de los score target y non-target

T1



T2



T3

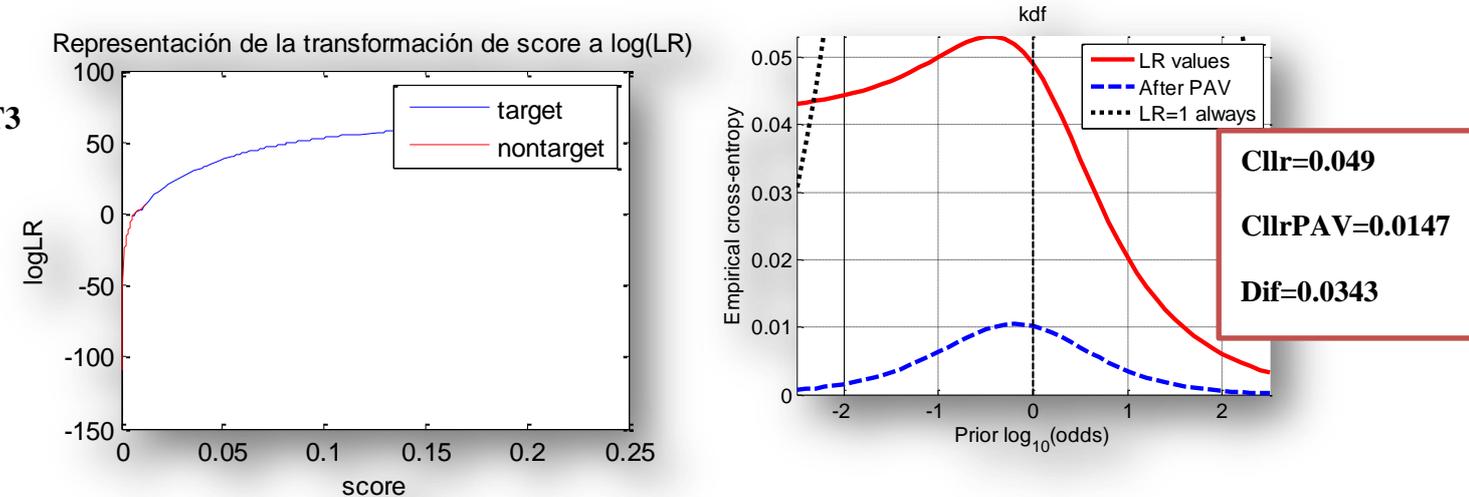


Figura 6-18 Transformación de score a LR con transformaciones. ECE plot con transformaciones en KDF

Comparación:

En este caso, para la transformación 1 se obtiene un valor más bajo de incertidumbre, buenos resultados, si únicamente nos fijamos en el punto de referencia del centro, el Cllr. Además, se observan otros problemas con la transformación 3, para ciertos valores los resultados obtenidos sobrepasan notablemente el caso peor. Esto da lugar a sistemas muy inestables.

MODELO KDF -> TRANSFORMACIÓN 1

Está claro que para poder valorar los resultados es necesario definir según qué factor se está analizando ya que existen muchos factores y según el caso se debe escoger el sistema adecuado.

Para finalizar el análisis de las transformaciones propuestas, resulta útil la representación de cómo se distribuye la relación entre la transformación de las puntuaciones score a las transformaciones LR para cada modelo propuesto. De esta manera se estudia cómo se comporta cada transformación según el modelo usado.

TRANSFORMACIÓN 1

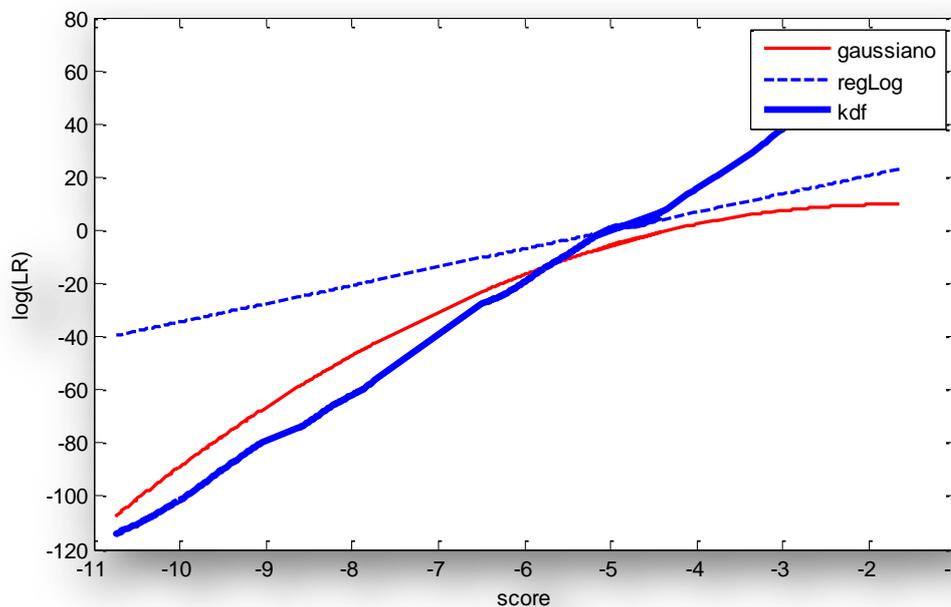


Figura 6-19 Comportamiento de la transformación 1

TRANSFORMACIÓN 2

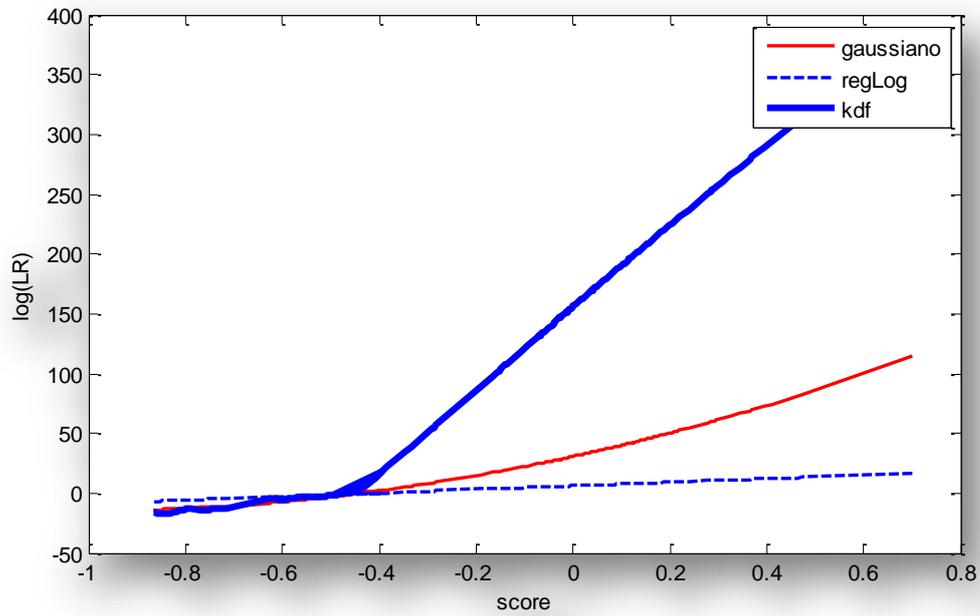


Figura 6-20 Comportamiento de la transformación 2

TRANSFORMACIÓN 3

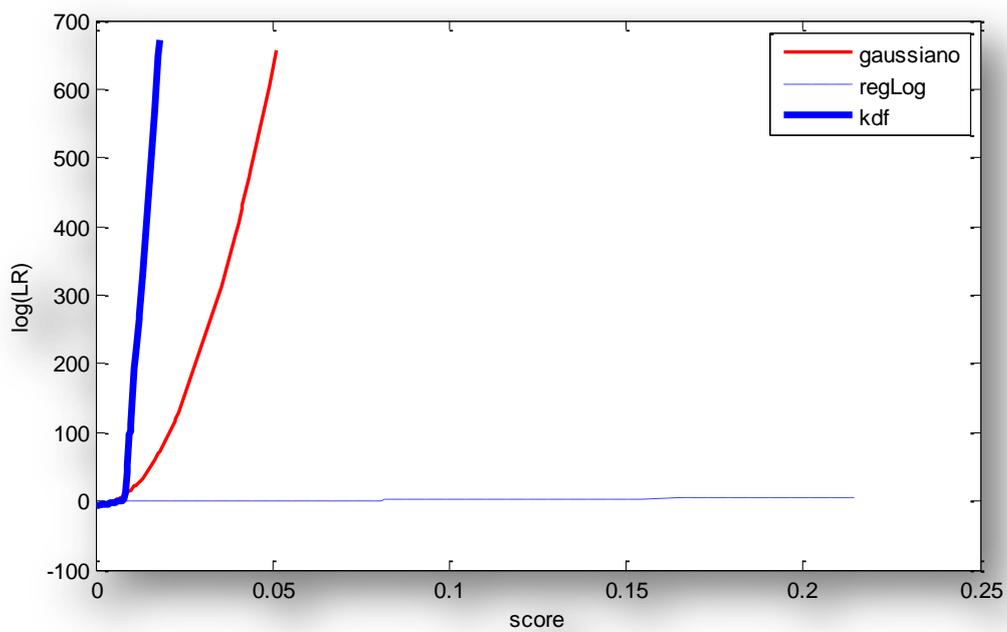


Figura 6-21 Comportamiento de la Transformación 3

Se han logrado encontrar unas transformaciones que representan mejor cada modelo en términos de unos resultados bajos de entropía. Sin embargo, la información que se deduce de la transformación score-log (LR) no es nada buena, sobre todo en la transformación 3.

Es muy inestable que para pequeñas variaciones de los score se produzcan grandes cambios en los LR, ya que si existe alguna evidencia engañosa, es decir una evidencia que se aleje un poco de su distribución y que siendo target se considere non-target, se va a convertir en un log (LR) muy elevado, quiere decir que estará apoyando la hipótesis errónea con mucha intensidad, lo cual provoca unos resultados muy engañosos.

Además la base de datos que estamos usando es pequeña, y no se puede tener la certeza de que estos resultados sean buenos. Se podría interpretar que al ser una base de datos tan ajustada, los datos que manejamos sean muy parecidos entre sí (los de test con los de train), por lo que si se tuviera una base de datos mayor en la que exista mayor variabilidad, el sistema podría comportarse de una forma más inestable al manejar distintas magnitudes de datos.

Por ello, como este base de datos no se puede incrementar ya que en la DGGC tampoco se dispone de unas grandes bases de datos de huellas pertenecientes a casos forenses reales, se va a proceder a realizar las mismas pruebas con bases de datos más pequeñas.

En la tercera parte de este capítulo se centrará en cómo se obtienen las bases de datos más pequeñas y la forma de realizar los experimentos para conseguir extraer resultados fiables.

6.3 Tercera parte de los experimentos:

Realización de los experimentos a partir de bases de datos más pequeñas

Las bases de datos más pequeñas se extraen de la que se ha usado antes pero eliminando evidencias de forma aleatoria. De esta manera comprobaremos si los resultados empeoran, o si por el contrario algún modelo funciona correctamente a pesar de la escasez de datos, además estableceremos unos límites para los cuales el estudio dejaría de ser interesante para su evaluación.

En primer lugar vamos a diseñar tres bases de datos:

- Con un 10% menos de evidencias $B1=\{B1_1 \dots B1_k\}$
- Con un 20% menos de evidencias $B2=\{B2_1 \dots B2_k\}$
- Con un 50% menos de evidencias $B3=\{B3_1 \dots B3_k\}$

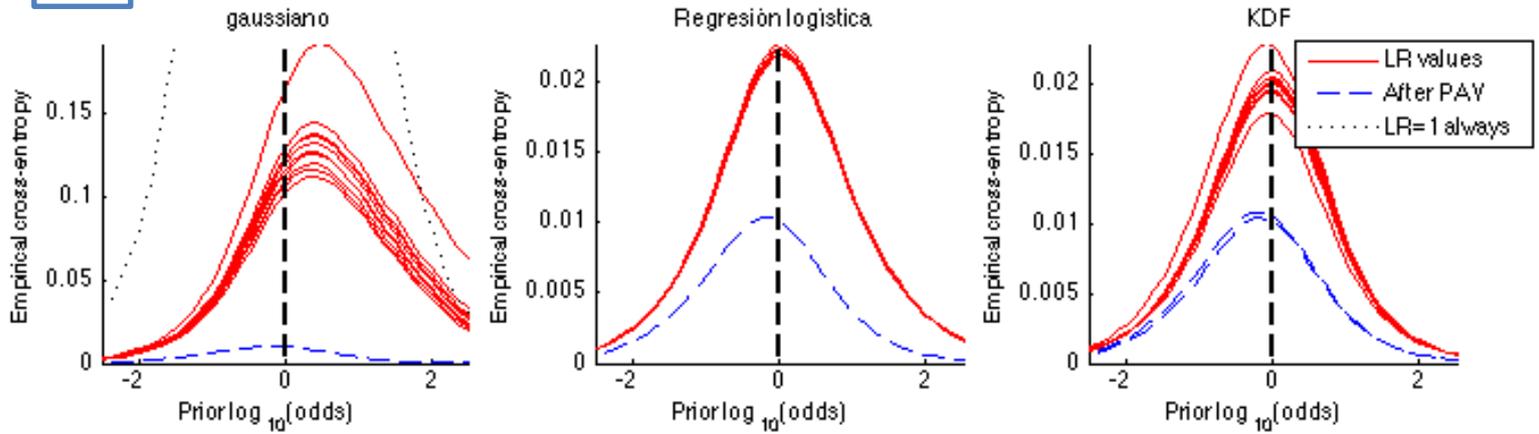
En cada base de datos se van a generar $k=10$ casos aleatorios en los que se eliminarán un porcentaje de evidencias (10%, 20% o 50%). Para cada base de datos se repetirán los tres modelos con las tres transformaciones propuestas representando la curva ECE de los 9 casos generados aleatoriamente.

Para realizar la comparación entre todos los casos obtenidos, nos fijaremos únicamente en el punto Cllr. Este punto lo tomaremos de referencia para ver cómo se comporta tanto la curva de entropía de los LR obtenidos como la curva de los LR óptimos obtenidos tras PAV.

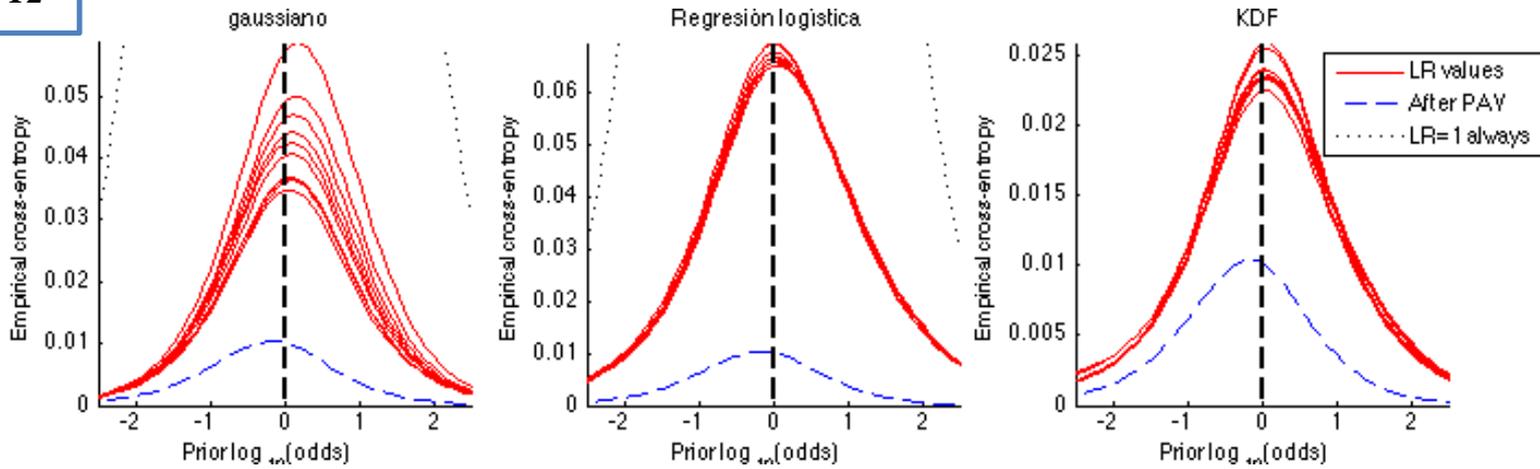
Valorar únicamente los experimentos mediante el punto de referencia establecido parece poco aprovechable, ya que de esta manera se escapan muchos datos del estudio. Es en este momento donde aplicaremos los tres criterios que se definieron anteriormente en el capítulo 5, robustez, calibración y Cllr. Ya que no se puede definir únicamente un modelo ganador, por así decirlo, se realizará el análisis del estudio para decidir según qué escenario tengamos escoger un modelo adecuado.

10% menos

T1



T2



T3

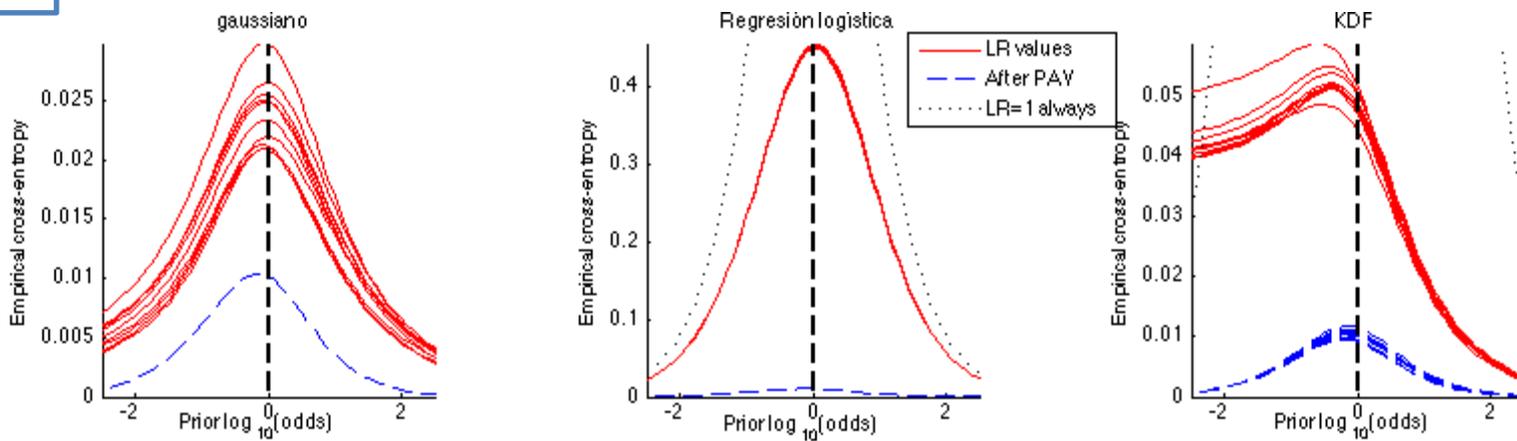
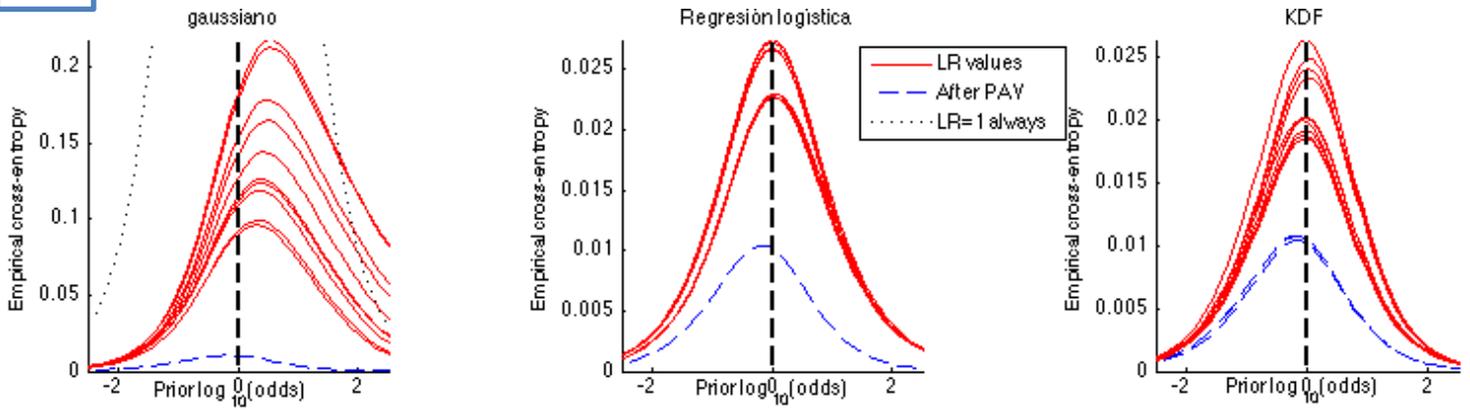


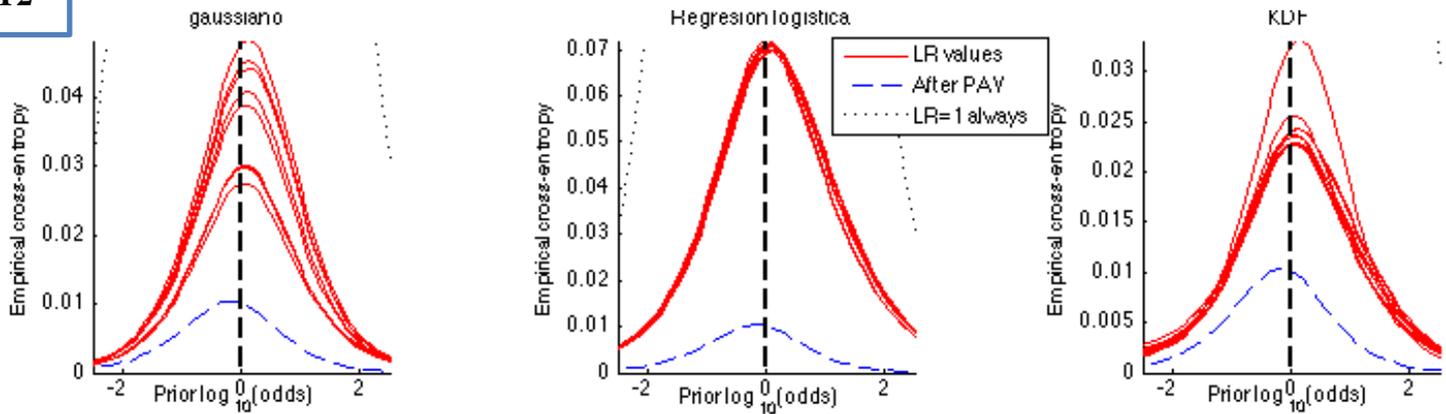
Figura 6-22 ECE plot para los 9 modelos con un 10% menos de evidencias

20% menos

T1



T2



T3

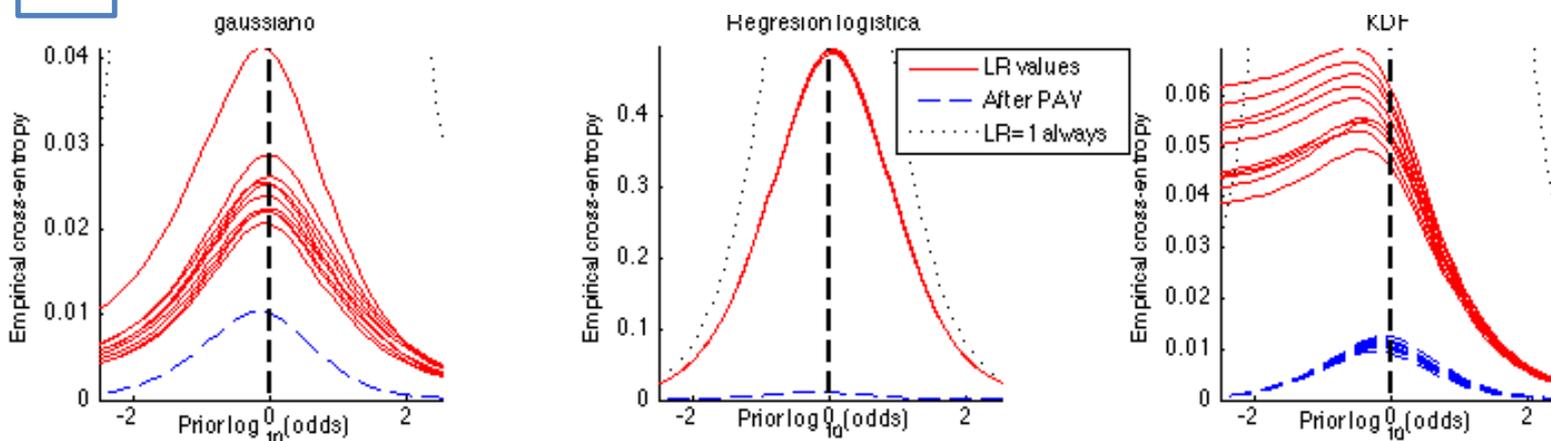
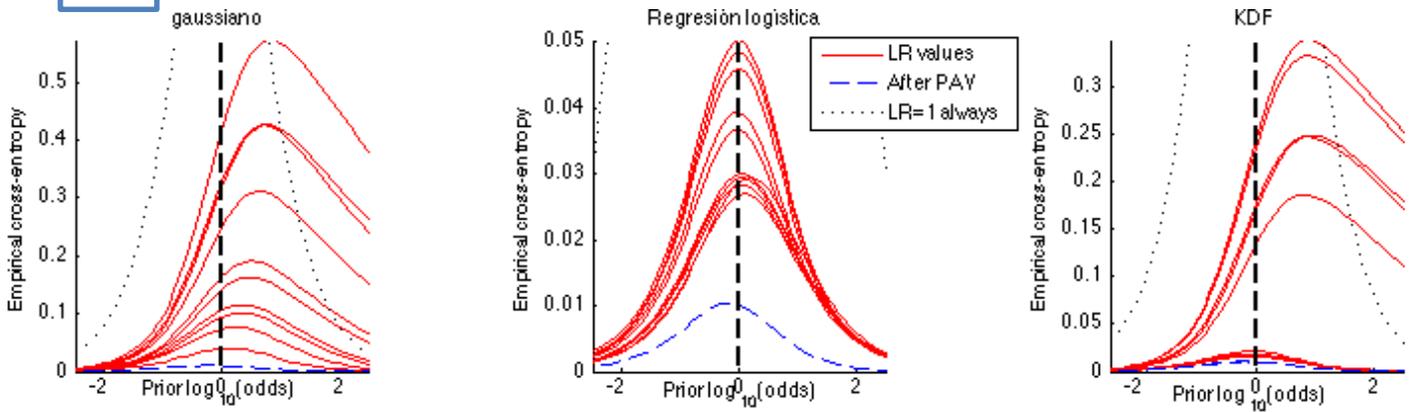


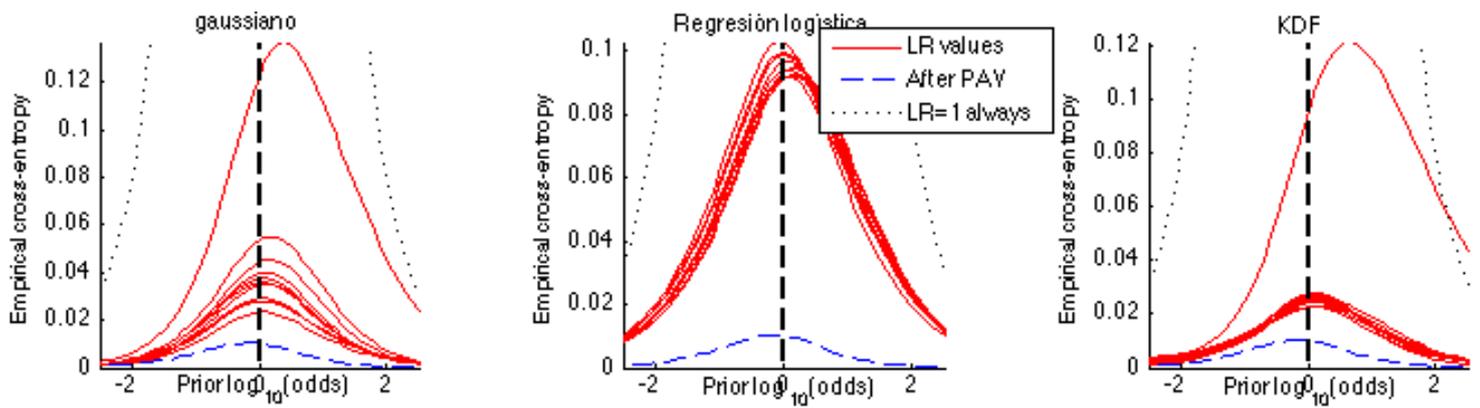
Figura 6-23 ECE plot para los 9 modelos con un 20 % menos de evidencias

50% menos

T1



T2



T3

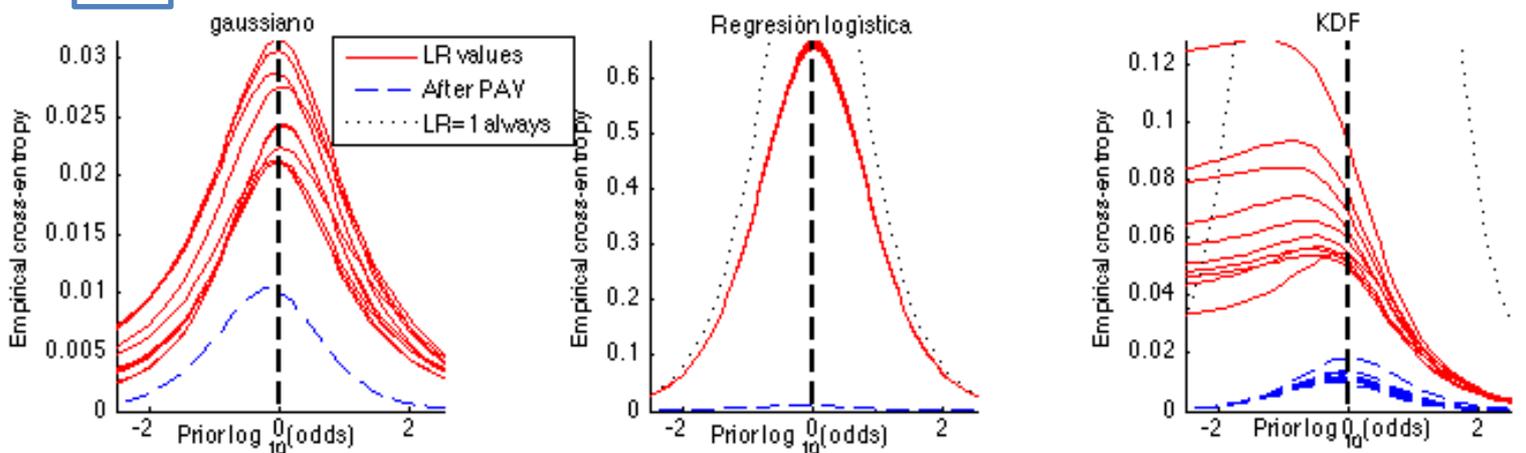
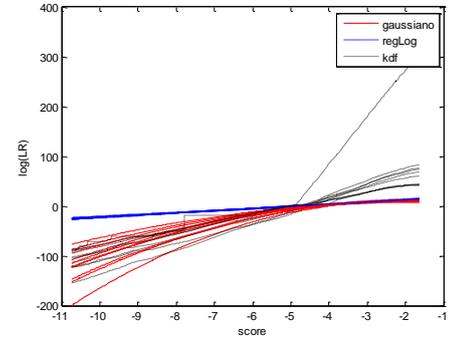
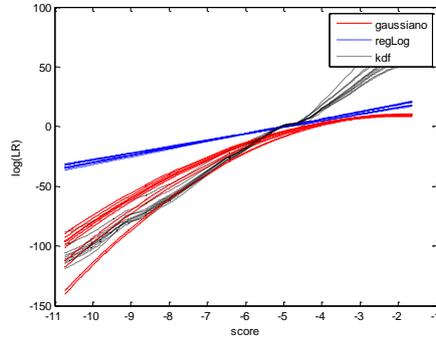
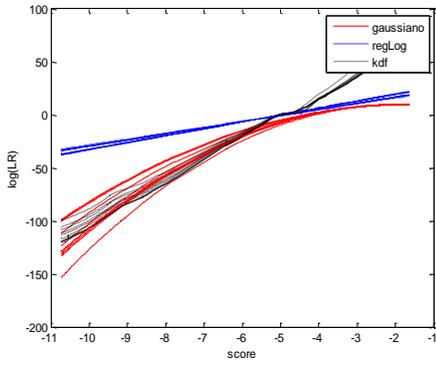


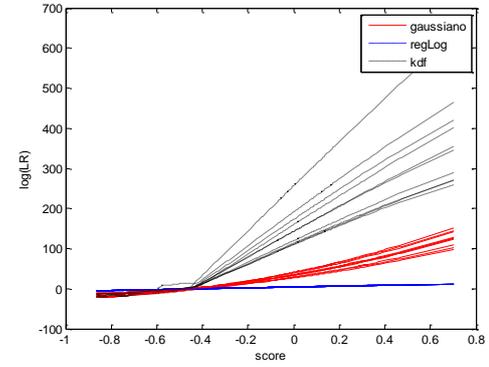
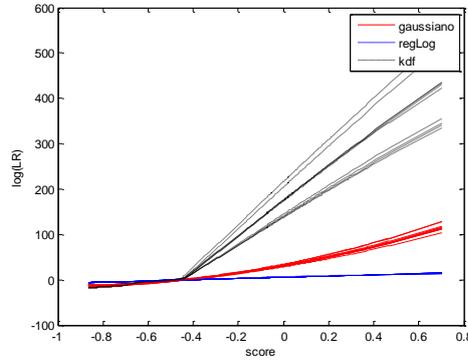
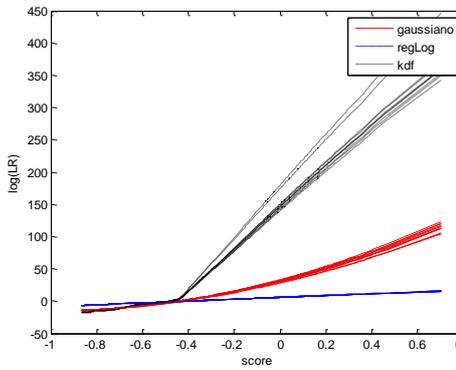
Figura 6-24 ECE plot para los 9 modelos con un 50% menos de evidencias

10%
20%
50%

T 1



T 2



T 3

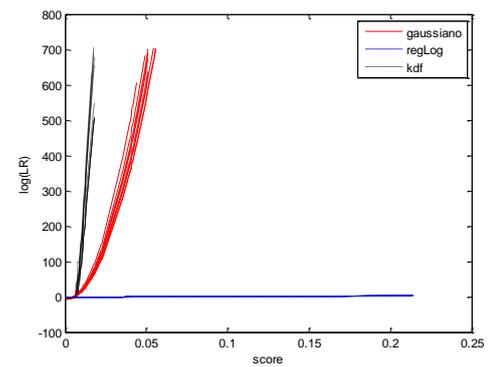
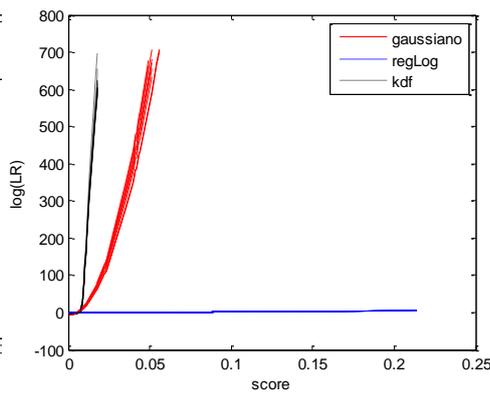
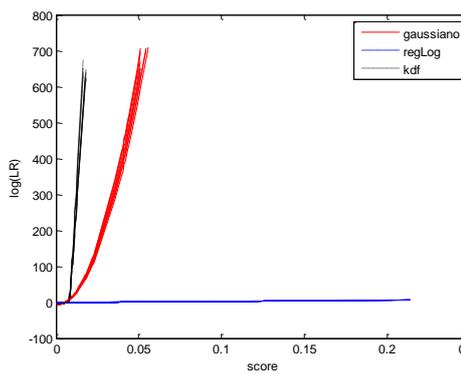


Figura 6-25 Representación de las Transformaciones en cada una de las bases de datos

6.3.1 Primer Criterio: ROBUSTEZ

Se define como el rango de variación de los valores de Cllr al probar con distintas implementaciones aleatorias de bases de datos del mismo tamaño. Cuanto menor sea, más robusto será el sistema, más fiable serán los valores cuando el sistema se ponga en funcionamiento

	T1			T2			T3		
	10%	20%	50%	10%	20%	50%	10%	20%	50%
Gaussiano	0.1008	0.1426	0.2749	0.0181	0.0287	0.0666	0.0078	0.0064	0.0262
Reg.Log	0.0047	0.0069	0.0248	0.0034	0.0054	0.0095	0.0060	0.0125	0.0165
KDF	0.0085	0.0158	0.2625	0.0031	0.0025	0.9763	0.0074	0.0195	0.0374

Tabla 6-1 Datos de robustez

Comentarios:

- El modelo regresión logística es el que mayor robustez presentan los datos si reducimos aleatoriamente la base de datos al 50% y además usamos la transformación 2.
- El modelo de Kdf presenta unos rangos de variación parecidos al de regresión logística excepto cuando se reducen los datos un 50% que aumenta considerablemente. Además anteriormente se indicó que la mejor transformación para KDF era la 1, sin embargo, en cuanto robustez de los datos la opción más acertada es la 2, incluso reduciendo la base de datos hasta en un 20%.

Teniendo en cuenta estas dos consideraciones podemos deducir que si disponemos de datos insuficientes es conveniente usar regresión logística, en concreto la transformación2, sin embargo si la cantidad de datos no es muy escasa es conveniente usar una opción más arriesgada con KDF ya que los datos son más estables. En cualquier caso la transformación 2 es la que mejor está funcionando en cuanto a estabilidad de los datos.

6.3.2 Segundo Criterio: CALIBRACIÓN

Se mide la calibración de cada modelo a la diferencia del punto medio de los LR reales con respecto a los obtenidos tras PAV

	T1			T2			T3		
	10%	20%	50%	10%	20%	50%	10%	20%	50%
Gaussiano	0.1058	0.1161	0.1785	0.0277	0.0215	0.0303	0.0093	0.0111	0.0106
Reg.Log	0.0076	0.0098	0.0217	0.0521	0.0557	0.0806	0.4391	0.4769	0.6470
KDF	0.0054	0.0068	0.0907	0.0094	0.0098	0.0174	0.0337	0.0378	0.0452

Tabla 6-2 Valores de Cllr según la Calibración

Conclusiones adicionales:

- Para el modelo gaussiano al reducir la base de datos un 50% menos la empeora no es considerable. Combinado con la robustez de los datos, se podría reducir un 20%, ya que con la mitad el aumento de variación resulta inestable.
- Con regresión logística el mejor calibrado se obtiene en la transformación 1, al igual que ocurría con la base de datos completa, en la que reduciendo un 10% la base de datos el calibrado no varía mucho. Manteniendo una robustez en los datos buena.
- En KDF con la T1 se puede reducir un 10% los datos obteniendo el mismo calibrado, incluso reduciendo un 20% menos KDF presenta el mejor calibrado con las transformación 1

El sistema mejor calibrado es KDF en la primera transformación con un 10% menos, incluso se podría reducir un 20% que el calibrado es mejor que los otros modelos. Cuando se disponga de datos suficientes conviene usar kdf con T1.

Si se reduce la base de datos en un 50% el que mejor calibración presenta es el modelo gaussiano con la transformación 3. Si los datos son insuficientes el mejor resultado se obtiene con el modelo gaussiano en la T3.

6.3.3 Tercer Criterio: CLLR

Establecido como el punto de referencia, en el que las probabilidades de ambas hipótesis son iguales. Cuando menor, mejor. Debido a que se tiene un número de experimento lanzados $k=10$, se estimará el punto medio de estas aportaciones.

	T1			T2			T3		
	10%	20%	50%	10%	20%	50%	10%	20%	50%
Gaussiano	0.1473	0.1289	0.1131	0.0430	0.0373	0.0349	0.0228	0.0268	0.0290
Reg.Log	0.0253	0.0257	0.0297	0.0669	0.0702	0.0942	0.453	0.4934	0.6613
KDF	0.0343	0.0331	0.0394	0.0253	0.0245	0.0272	0.0494	0.0491	0.0581

Tabla 6-3 Valores de Cllr

Comentarios:

- El Cllr más bajo de todos los casos analizados, y por tanto, el mejor, es de 0.0228 en el modelo gaussiano usando la transformación 3 y reduciendo tan solo un 10% la base de datos. Con la base de datos original, el Cllr se localizaba en 0.0232, prácticamente son iguales.
- Si analizamos como afecta la reducción de la base de datos en cada uno de los 9 casos por separado, se percibe que aunque se reduzca la base de datos los valores medios de Cllr se mueven en torno a unos valores parecidos. Se podría enunciar que **“La reducción del tamaño de la base de datos no afecta de forma significativa al valor del Cllr”**.
- En cuanto a elegir el mejor modelo cuando la base de datos es pequeña, (la reducimos al 50% que supondría el peor caso que estamos analizando.) habría que escoger el modelo KDF con la transformación 2, seguido muy de cerca del modelo gaussiano con la transformación 3.
- Por el lado contrario, si disponemos de una base de datos con un tamaño considerable el modelo que más bajo Cllr presenta es el modelo gaussiano con las transformación 3.
 - Este valor de Cllr puede resultar confuso ya que se ha estimado el valor medio para el número de repeticiones probadas lo cual depende en gran medida de la robustez de los datos en el modelo, analizada anteriormente.

Caso extremo: Generación aleatoria de k=100 bases de datos distintas con un 50 % menos de evidencias

Exponiendo un caso extremo, con tantas generaciones de los experimentos como lo es una cantidad de 100, permite ver a simple vista cómo se comportan las puntuaciones obtenidas, y permite decidir de forma rápida y general cuál se comporta mejor o cuál es más estable. A partir de aquí se tendría que decidir que se valora más, obtener entropías bajas aunque no siempre, u obtener resultados estables, fiables, que con cada realización del experimento el resultado no varíe mucho.

T1

Se puede determinar que el más estable es el modelo de regresión logística,

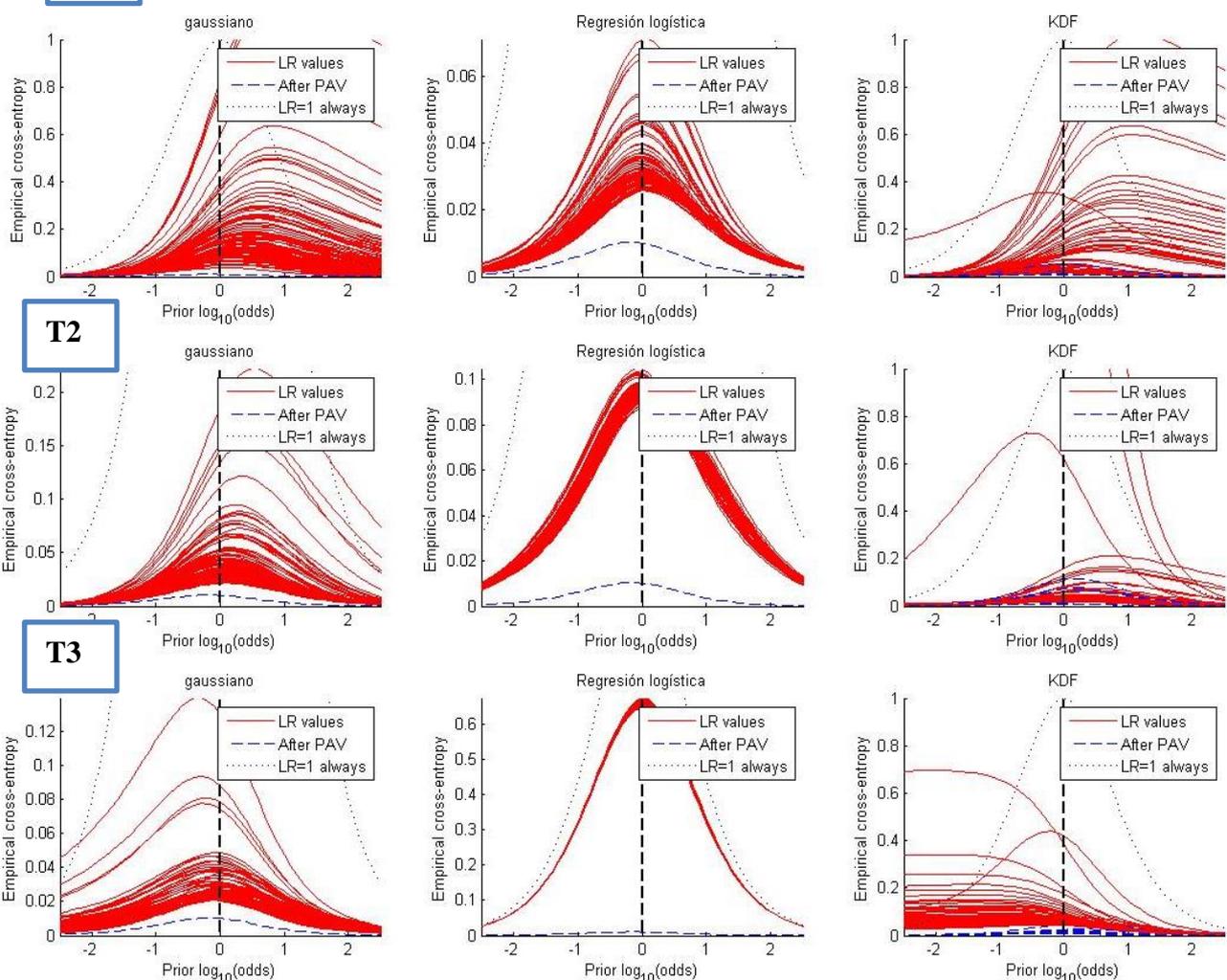


Figura 6-26 Generación de 100 muestras para cada base de datos con un 50 % menos de evidencias

ofreciendo buenos valores de entropía salvo para la transformación 3.

6.4 Cuarta parte de los experimentos:

Resolución del análisis del estudio de los modelos

Como se propuso en el capítulo 5, el objetivo del análisis de decisión de entre los modelos propuestos no es el de llegar a un único modelo ganador sino a poder establecer según la situación que tengamos, el modelo que mejor se adapte.

	Robustez	Calibración	Cllr
BD Grande (10%)	KDF T2	KDF T1	KDF T1
BD Pequeña (50%)	Reg.Log T2	Gauss T3	KDF T2

Tabla 6-4 Modelos escogidos para cada criterio

Sin embargo, estas valoraciones presentan algunos inconvenientes, y es que en esta elección solo ha influido el punto medio, en el que las probabilidades son iguales.

Además no se tiene muy en cuenta lo que ocurre en los extremos debido a dos motivos:

- Las proporciones que se aprecian en las representaciones de las curvas ECE pueden resultar engañosas. La diferencia de la entropía obtenida para los LR obtenidos con respecto al valor del caso neutral en el que los LR son iguales a 1 no se aprecian del mismo modo en el centro de la gráfica que en los extremos. Por lo que se puede caer en conclusiones engañosas ya que las proporciones en los extremos son más pequeñas, y se puede malinterpretar que se esté comportando mejor que en el centro. Con proporciones mayores parece que la variación es mayor, y por tanto peor en cuanto a robustez. De hecho se puede deducir de antemano que probablemente en los extremos exista más sensibilidad a una mala robustez, ya que en los extremos un valor grande de LR tiene una influencia mucho mayor en la degradación de entropía, esto ocurre en los extremos donde la probabilidad de una hipótesis es mucho más alta que la otra.
- El rango de probabilidad para los que se han probado los experimento se ha elegido de forma subjetiva, se podría expandir este intervalo y observar qué ocurre más allá.

Para tratar de solucionar el primer punto, se recurre a las curvas de entropía normalizadas NCE, comparándose así con las curvas ECE utilizadas, además de probar a ampliar el rango de probabilidades a priori.

6.4.1 NCE: Normalized Cross-Entropy

Se aplica las curvas NCE únicamente para los 6 mejores modelos establecidos según los criterios anteriores.

BD sin10%

Robustez

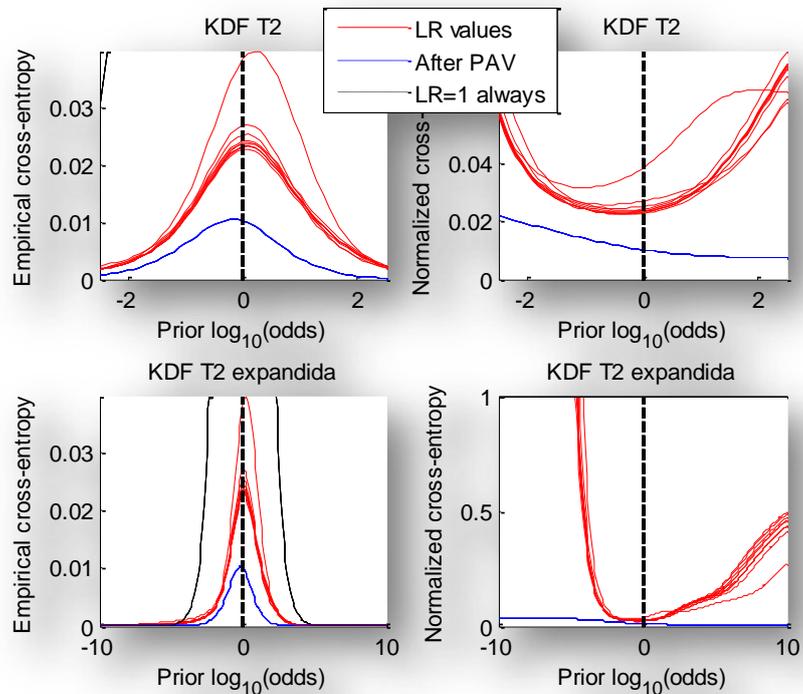


Figura 6-27 NCE para KDF T2 sin 10%

Calibración y Cllr

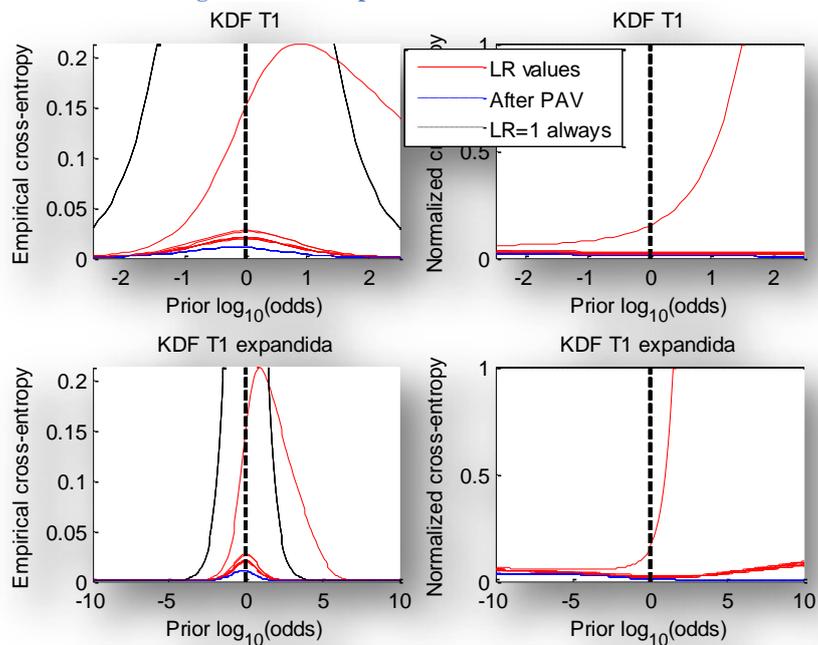
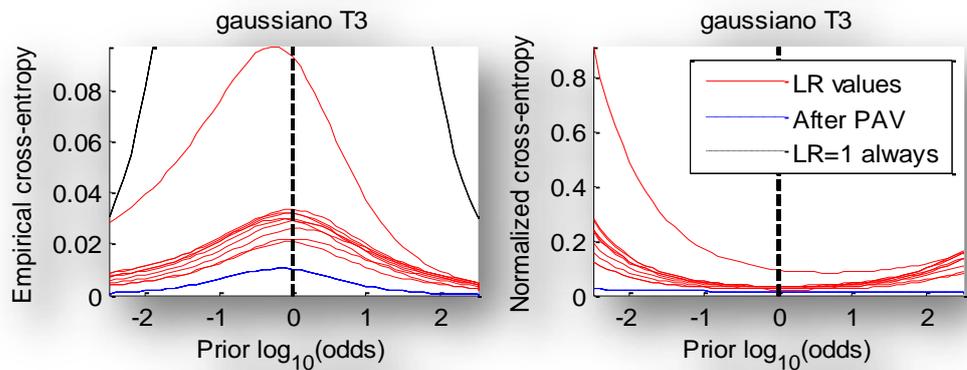
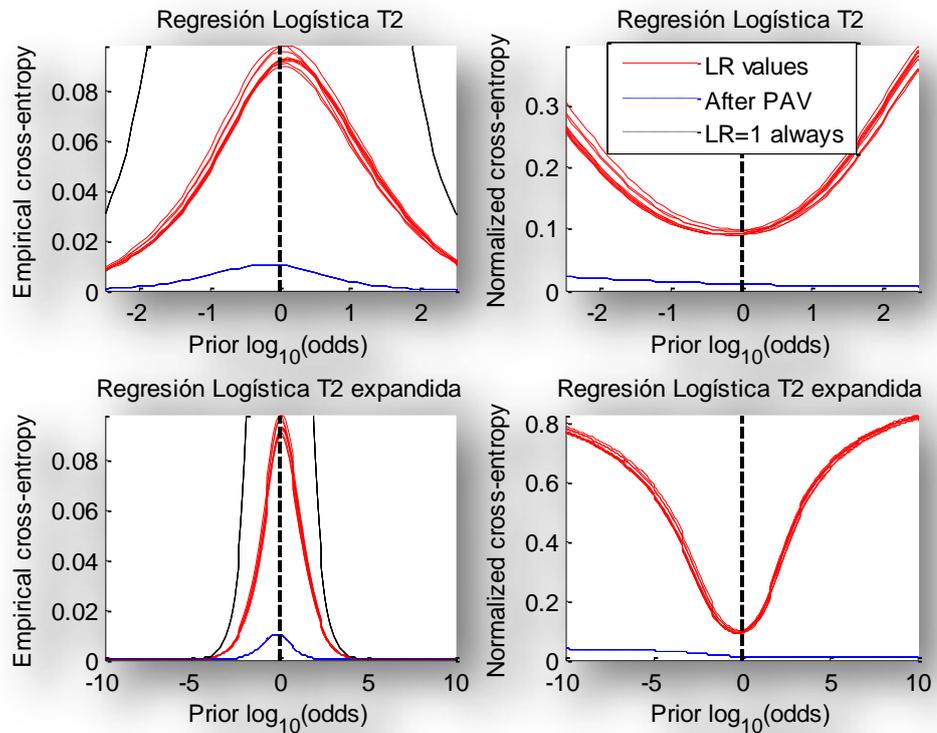


Figura 6-28 NCE para KDF T1 sin 10%

BD sin50%

Robustez



Calibración

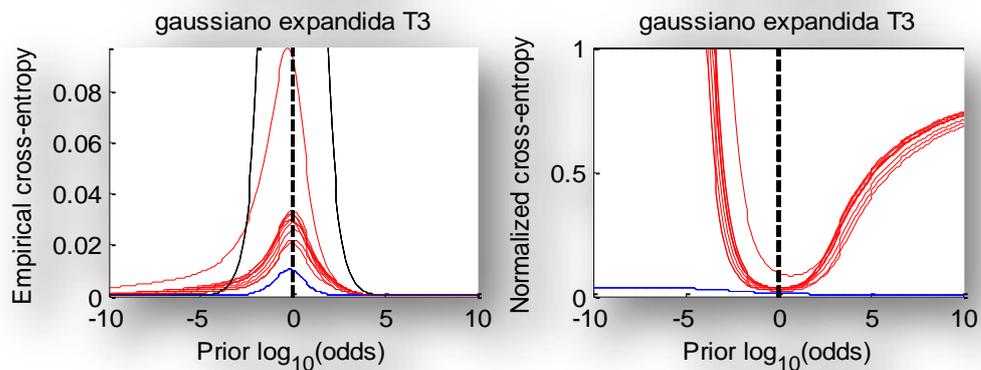


Figura 6-29 NCE para los mejores modelos con un 50% menos de evidencias

Cllr

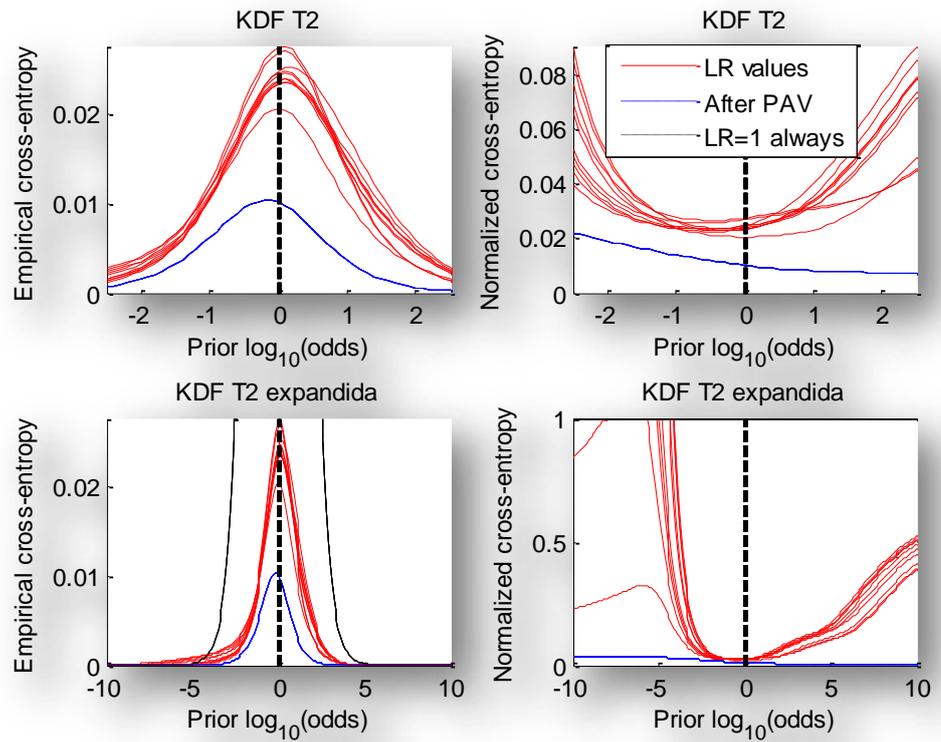


Figura 6-30 NCE para los mejores modelos con un 50% menos de evidencias

Si al alejarnos del cero hay más dispersión en los lados es un indicativo de falta de robustez, lo cual era de esperar desde un principio. Esto supone que no son valores fiables para el entrenamiento del sistema.

Sin embargo, expandiendo la representación de las probabilidades a priori a pesar de que tengamos más información de lo que pasa en los extremos observamos que a veces el comportamiento no es estable. Para tener un conocimiento mayor de la fiabilidad de los datos de entrenamiento se podría restringir un intervalo para el cual el comportamiento es válido, y fuera de él el comportamiento no es muy estable.

6.4.2 Regla del 30 de Doddington

Aplicando esta regla se realiza la representación gráfica indicando intervalos de confianza, dentro de los cuales la regla del 30 aplica un intervalo para cada grupo de LR obtenidos según el sistema utilizado.

- Base de Datos grande

KDF T2. [-0,02: 55.8]

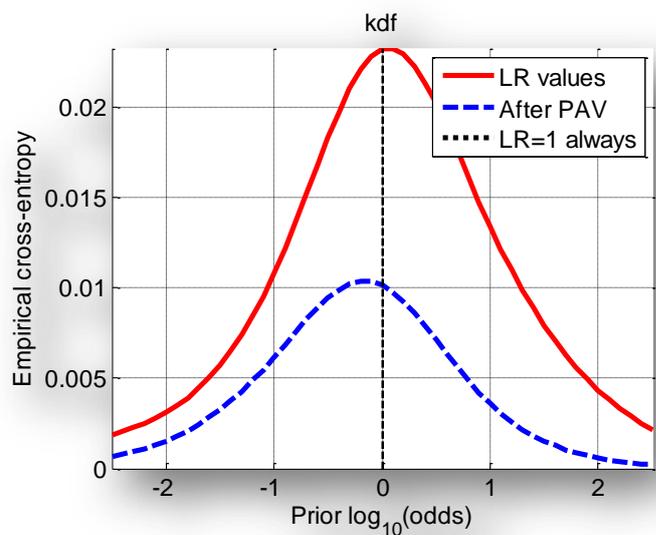


Figura 6-31 ECE plot KDF T2

KDF T1 [0.18:18.7]

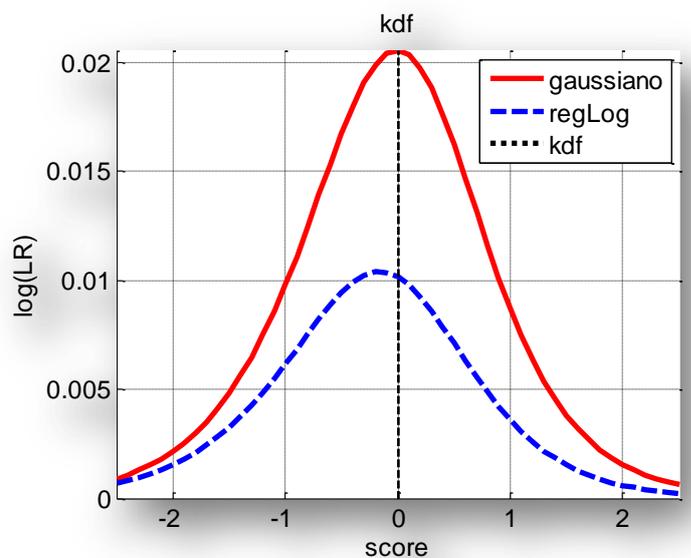


Figura 6-32 ECE plot KDF T1

- Base de Datos pequeña

RegLog T2 [-0.4:1.5]

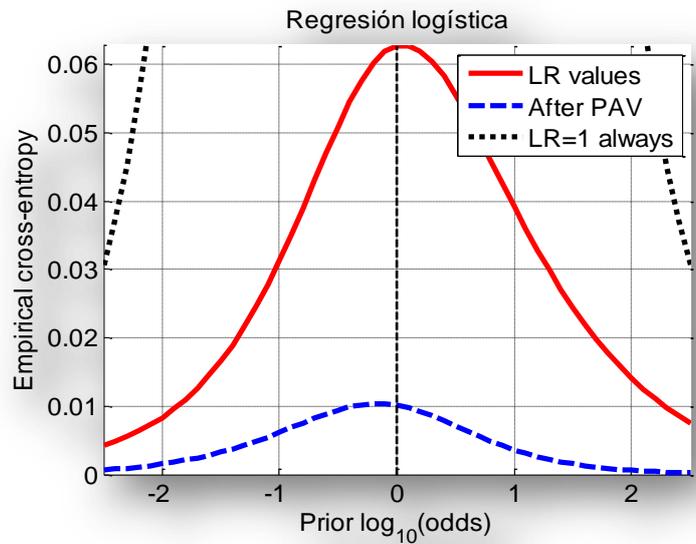


Figura 6-33 ECE plot Reg Log T2
Modelo gaussiano

Gauss T3 [1.6:∞]

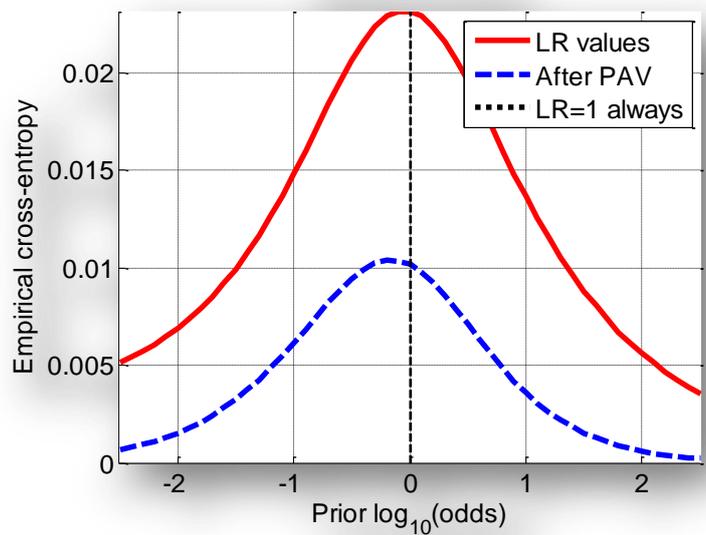


Figura 6-34 ECE plot Gauss T3

KDF T2 [0.06:49.8]

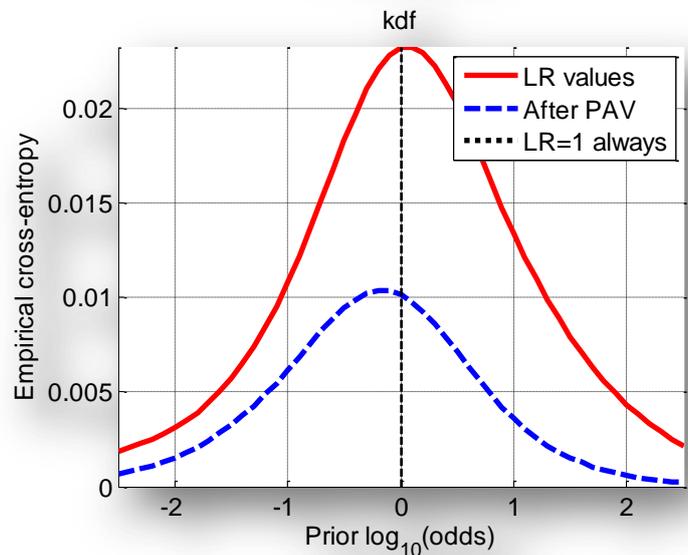


Figura 6-35 ECE plot KDF T2

Gracias al análisis de los intervalos de confianza se puede comprobar cómo la reducción de la base de datos para algunos modelos, como el caso del modelo gaussiano, degrada bastante el resultado, obteniendo LR de valores muy grandes, infinito. Es lo que ocurre en el modelo gaussiano con T3 para cuando se reduce un 50% la base de datos, cuando se analiza el intervalo completo se observa que con falta de datos el modelo gaussiano no es un sistema que represente bien los datos.

En general, cuando el intervalo alcanza valores muy elevados no representa a un sistema muy estable, en el caso de KDF también se están alcanzando unos valores elevados por el límite superior. De esta manera el modelo de regresión logística es el más estable se está comportando.

7. Conclusiones y trabajo futuro

7.1 Conclusiones

Con la realización de este proyecto se puede comprender la necesidad de una evidencia científica en un caso judicial y como ésta puede influir en la toma de decisión entre las dos únicas hipótesis que se barajan en un juicio, culpable o inocente.

Recientemente se pueden ver muchos casos en la prensa de cómo las pruebas extraídas y analizadas científicamente pueden influir o por el contrario, cómo una falta de ellas provocan que no se tenga suficiente información como para declarar una relación de culpabilidad. Existen otros casos, como por ejemplo el de los niños Ruth y José, en que una antropóloga forense declaró de forma precisa y categórica que los huesos hallados en la finca del acusado no pertenecían a los hijos de éste. Tras este acontecimiento y una sucesión de solicitud de segunda revisión, se declaró, por parte de otro antropólogo forense, que sí que había indicios para afirmar que esos huesos pertenecían a los niños.

Analizando este asunto no se puede entender cómo si se ha afirmado que los huesos no pertenecían a la fuente, es más, pertenecían a roedores, ahora se pueda plantear que sí pertenecen y que el estudio previo era erróneo. Esto fue lo mismo que se preguntó el abogado defensor. *“El abogado defensor, José María Sánchez de Puerta, se ha interesado por cómo es posible que en su informe diga que ninguno de los huesos tienen origen animal y, en cambio, la perito Josefina Lamas diga todo lo contrario, que ninguno tiene naturaleza humana “[13].* Con qué seguridad habría que basarse en el estudio anterior y con qué seguridad habría que basarse en el estudio último.

Éste es el principal objetivo en el que estaba basado este proyecto. Si en vez de categorizar con un culpable o inocente en relación a las evidencias se hubiera podido establecer un porcentaje o un peso por un lado de la validez de las pruebas y por otro de la seguridad con que se afirma la hipótesis, resultaría más fácil comprender el cambio de opinión basado en las evidencias. Ya que quizás la primera hipótesis que se lanzó

poseía un peso muy bajo, es decir, apoyaba poco la decisión o las pruebas eran insuficientes. Mientras que la segunda hipótesis disponía de más argumentos y pruebas que provocaban un peso más elevado, dando mayor seguridad a la decisión.

Gracias al cálculo de LR mediante teorías bayesianas que permiten analizar las evidencias sin conocer el resto de información del caso se consigue que la ciencia tenga un carácter más transparente, lógico y racional para la toma de decisión. Sin olvidar que la tarea del perito no es la de establecer una decisión final categórica, sino la de aportar la máxima información posible de la mejor forma para que el juez disponga de más información añadida, bien para reforzar su decisión o bien para rebatir la decisión basándose en hechos válidos.

Para poder llevar a cabo esta metodología es de vital importancia la colaboración con los cuerpos de seguridad y la concienciación de éstos mediante formación continua. Por ello el uso de herramientas como las usadas en este proyecto para la creación de la base de datos resultan bastante útiles para poder aplicar los cálculos de LR.

De los diferentes modelos propuestos para el cálculo de LR, nueve, se ha podido ver que no se pretende escoger un único modelo. Tras el estudio de los distintos casos realizado se extraen conclusiones, pero dependiendo del criterio a perseguir y del escenario que se tenga.

En cuanto a la robustez de los resultados, si disponemos de datos insuficientes es conveniente usar regresión logística, en concreto la transformación 2, sin embargo, si la cantidad de datos no es escasa es conveniente usar una opción más arriesgada con KDF ya que los datos son más estables, o incluso con regresión logística también. En cualquier caso la transformación 2 es la que mejor está funcionando en cuanto a estabilidad de los datos.

En cuanto a la calibración de los resultados con respecto a PAV, el sistema mejor calibrado cuando se disponga de datos suficientes es KDF con T1, incluso reduciendo un 20% la base de datos el calibrado es mejor que en los otros modelos.

Si los datos son insuficientes el mejor resultado se obtiene con el modelo gaussiano en la T3. Si se reduce la base de datos en un 50% el que mejor calibración presenta es el modelo gaussiano con la T3.

En cuanto al valor medio de Cllr, cuando la base de datos es pequeña, (la reducimos al 50% que supondría el peor caso que estamos analizando) habría que escoger el modelo KDF con la transformación 2, seguido muy de cerca del modelo gaussiano con la transformación 3.

Por el lado contrario, si disponemos de una base de datos con un tamaño considerable el modelo que más bajo Cllr presenta es el modelo gaussiano con las transformación 3. La reducción del tamaño de la base de datos no afecta de forma significativa al valor del Cllr.

Analizando los intervalos de confianza, nos permite cuantificar el comportamiento en un rango más amplio que si se analiza únicamente el punto Cllr, con esto se ha demostrado que el modelo gaussiano con la T3 no es recomendable en el caso de una base de datos pequeña. Por el contrario, regresión logística sigue manteniendo su estabilidad, se confirma cuando se ha calculado el intervalo de confianza.

En general, se puede analizar de una forma conjunta, analizar cada criterio observando los otros dos, ya que quizás compensa obtener una peor calibración a costa de una mejora considerable en cuanto a robustez.

Gracias a la figura, se observa que regresión logística funciona bastante bien, por dos motivos, se observa muy bien que es el que mejor robustez presenta, provocando poca variabilidad. Además, la normalización de los score se realizó basándose en modelos gaussianos, lo que puede provocar que el modelo de regresión logística puede mejorar todavía más si se buscan otros tipos de normalizaciones más acorde a su modelo.

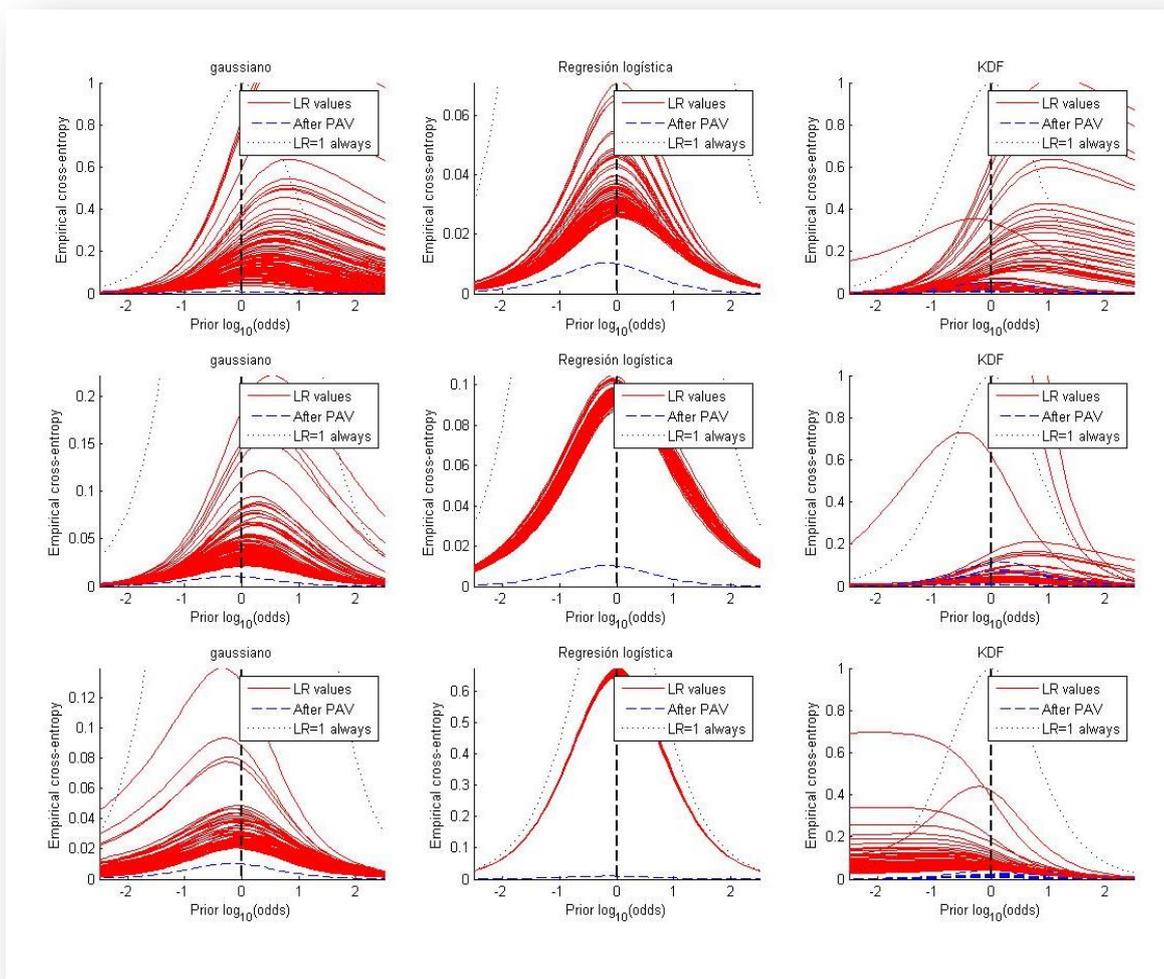


Figura 7-1 Representación del rendimiento de los 9 modelos para una base de datos pequeña y 100 implementaciones distintas

Debido a esta normalización, el modelo gaussiano, aunque es muy inestable, con la última transformación se consiguen valores de entropía muy bajos.

El modelo KDF es altamente inestable para una de las regiones, lo que a pesar de obtener buenos resultados numéricos y llegar a alcanzar valores muy bajos de entropía, no se puede dejar de lado que no siempre se comportará igual y que los datos de calibración y Cllr se han estimado como un valor medio de todas las representaciones de las distintas bases de datos.

7.2 Trabajo Futuro

A partir de este proyecto surgen nuevas líneas de investigación:

Creación de bases de datos más grandes. Gracias a la herramienta ya implementada se propone seguir aumentando la base de datos con los nuevos cotejos que se vayan obteniendo. Sobre todo, ampliar la base de datos de huellas palmares y ampliar el estudio en ellas.

Repetición de los experimentos a nuevas bases de datos. Se pueden comparar con los resultados obtenidos de otras bases de datos, de esta manera se verá la diferencia de tratar con huellas latentes reales o con huellas extraídas de entornos controlados.

Aplicar los modelos a otros rasgos biométricos. Comprobar si los mismos modelos y normalizaciones propuestas funcionan igual cuando la entrada de datos es otra, los scores provienen de comparaciones de otros rasgos biométricos, tales como locuciones o muestras faciales. Estableciéndose así un criterio de normalización útil para cualquier sistema biométrico.

Búsqueda de nuevos métodos de normalización mediante transformaciones monótonas que se adapten mejor a los modelos de regresión logística.

8. Bibliografía

1. Champod, C. and Evett, I.W. "A probabilistic approach to fingerprint evidence". *Journal of Forensic Identification*, pages 101-122. Vol. 51(2), 2001.
2. Ratha, N. and Bolle, R., "Automatic Fingerprint Recognition Systems". Springer, 2003.
3. Fine, G.E. A Review of the FBI's Handling of the Brandon Mayfield Case. U.S. Department of Justice, Office of the Inspector General, 2006.
4. Ramos, D. Forensic evaluation of the evidence using automatic speaker recognition systems. PhD. Thesis. UAM, Madrid, 2007.
5. Champod. Forensic applications, overview. In S. Li, editor, *Encyclopedia of Biometrics*. Springer, New York, 2009. 3, 26
6. Evett. Expressing evaluative opinions: a position statement. *Science & Justice*, 51(1):1-2, 2011. Several signatories
7. C. Neumann, I. Evett, and J. Skerrett. Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*
8. Anil K. Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Trans. Circuits Syst. Video Techn.*, 14(1):4-20, 2004.
9. Jain, A.K., Flynn, P. and Ross, A.A. editors. *Handbook of biometrics*. Springer, 2007.

10. D. Meuwly and R. Veldhuis. Forensic biometrics: From two communities to one discipline. In Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the, pages 1{12. IEEE,2012
11. Richard O.Duda, Peter E.Hart, David G.Stark. "Pattern Classification". Second Edition.
12. "FoCal", <https://sites.google.com/site/nikobrummer/focal>.
13. <http://www.elmundo.es/elmundo/2013/07/03/andalucia/1372841743.html>
14. The coming paradigm shift in forensic identification science
15. <http://lectordehuellasdigitales.org/cotejo-de-huellas-digitales/>
- 16.http://www.aplicacionestecnologicas.com/Biometria_Aplicada/Huella_Digital/index.html
17. <http://insigniass.blogspot.com.es/2011/09/cursos-perito-judicial.html>
18. <http://www.biotronicfs.com/captahuellas.htm>
19. <http://www.mianamnesia.com/2011/12/dime-como-es-tu-huella-dactilar-y-te-dire-quien-eres/>
20. http://scielo.isciii.es/scielo.php?pid=S1135-76062005000200006&script=sci_arttext
21. <http://www.advancedsourcecode.com/palmid.asp>
22. http://en.wikipedia.org/wiki/Kernel_density_estimation

9. PRESUPUESTO

1) Ejecución Material

- Compra de ordenador personal (Software incluido)..... 2.000 €
- Alquiler de impresora láser durante 6 meses 50 €
- Material de oficina 150 €
- Total de ejecución material 2.200 €

2) Gastos generales

- 16 % sobre Ejecución Material 352 €

3) Beneficio Industrial

- 6 % sobre Ejecución Material 132 €

4) Honorarios Proyecto

- 1280 horas a 15 € / hora 19200 €

5) Material fungible

- Gastos de impresión 200 €
- Encuadernación 20 €

6) Subtotal del presupuesto

- Subtotal Presupuesto 22104 €

7) I.V.A. aplicable

- 21% Subtotal Presupuesto 10.525,6 €

8) Total presupuesto

- Total Presupuesto 22.114,5 €

Madrid, Julio de 2014

La Ingeniera Jefa de Proyecto

Fdo.: Sandra Uceda Ontivero

Ingeniera de Telecomunicación

10. PLIEGO DE CONDICIONES

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de la Calibración de puntuaciones procedentes de sistemas biométricos. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego. Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.
2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.
3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.
4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.
5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata,

según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad “Presupuesto de Ejecución de Contrata” y anteriormente llamado “Presupuesto de Ejecución Material” que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.