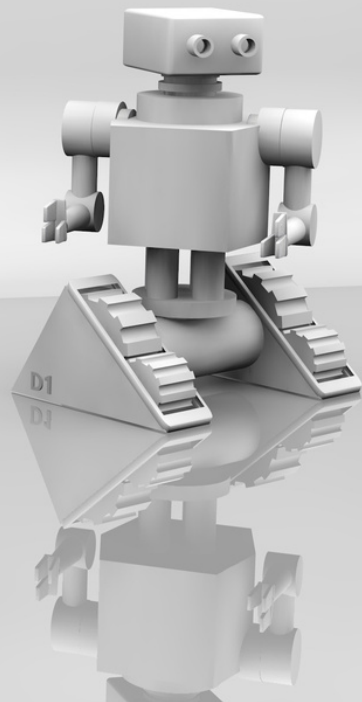


Avances en Informática y Automática

Decimocuarto workshop



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL



Avances en Informática y Automática

Decimocuarto Workshop

Avances en Informática y Automática

Decimocuarto Workshop

Editores

Jorge Vicente Gabriel
Iñaki Arberas Vicente

Publicado en España por:

Departamento de Informática y Automática Facultad de Ciencias
Universidad de Salamanca
Plaza de los Caídos s/n
37008, Salamanca, España
Tel.: + 34 923294653
Fax: + 34 923294514
Web: <http://mastersi.usal.es>
Email: mastersi@usal.es

ISBN 978-84-09-25507-8

Editores:

Jorge Vicente Gabriel
Iñaki Arberas Vicente

Prólogo

El Máster Oficial en Sistemas Inteligentes de la Universidad de Salamanca tiene como principal objetivo promover la iniciación de los estudiantes en el ámbito de la investigación. El congreso organizado por el Departamento de Informática y Automática que se celebra dentro del Máster en Sistemas Inteligentes de la Universidad de Salamanca proporciona la oportunidad ideal para que sus estudiantes presenten los principales resultados de sus Trabajos de Fin de Máster y obtengan una realimentación del interés de los mismos.

La decimocuarta edición del *workshop* “Avances en Informática y Automática”, correspondiente al curso 2019 - 2020, ha sido un encuentro interdisciplinar donde se han presentado trabajos pertenecientes a un amplio abanico de líneas de investigación, desde los sistemas multiagente y la visualización de la información hasta la minería de datos pasando por otros campos relacionados. Todos los trabajos han sido supervisados por investigadores de reconocido prestigio pertenecientes a la Universidad de Salamanca, proporcionando el marco idóneo para sentar las bases de una futura tesis doctoral. Entre los principales objetivos del congreso se encuentran:

- Ofrecer a los estudiantes un marco donde exponer sus primeros trabajos de investigación.
- Proporcionar a los participantes un foro donde discutir ideas y encontrar nuevas sugerencias de compañeros, investigadores y otros asistentes a la reunión.
- Permitir a cada estudiante una realimentación de los participantes sobre su trabajo y una orientación sobre las futuras direcciones de investigación.
- Contribuir al desarrollo del espíritu de colaboración en la investigación.

Organización

El *workshop* “Avances en Informática y Automática” está organizado por el Departamento de Informática y Automática de la Universidad de Salamanca.

Comité Organizativo

Belén Perez Lancho
Belén Curto Diego
Roberto Therón Sánchez
Jose Luís Alonso Berrocal
Pastora Isabel Vega Cruz
Angel Luis Sánchez Lázaro

Índice general

Inteligencia artificial en construcción de perfiles de asesinos en serie	2
<i>Alberto Botana López, Angélica González Arrieta, Ana Gil González, Ana de Luis Reboredo</i>	
Sistemas de reconocimiento y detección de poses	28
<i>Francisco García Encinas, María N. Moreno García, and Juan Francisco de Paz Santana</i>	
Estudio sobre viabilidad de interacción gestual en el tratamiento de mapas en entornos móviles	44
<i>David Peral García, José Rafael García-Bermejo Giner</i>	
Sistema de Clasificación Emocional basado en Estímulos Externos	66
<i>Nuria Mateos, Belén Pérez, Ana Belén Gil, Ana de Luis</i>	
IIoT, Big Data y Sistemas inteligentes bajo un modelo Edge Computing	92
<i>V. Gabriel, Juan M. Corchado, Ana Belén Gil</i>	
Machine Learning y Puntuación de Riesgo Poligénico para la Enfermedad de Alzheimer	109
<i>Nibeth Mena Mamani</i>	
Visualización de incertidumbre en grafos sociales para la exploración de bibliotecas digitales	126
<i>Iñaki Arberas, Alejandro Benito, Roberto Therón</i>	
Sistema inteligente entorno al mundo del grafiti	145
<i>Adilson Isaac González, Angélica González Arrieta, Daniel López Sánchez</i>	
Design and implementation of a Pipeline for advanced DNA analysis (NSG) by using Machine Learning Techniques	165
<i>Zakieh Alizadehsani, Juan M. Corchado</i>	
Autores	188

Inteligencia artificial en construcción de perfiles de asesinos en serie

Alberto Botana López,
Angélica González Arrieta, Ana Gil González y Ana de Luis Reboredo

Facultad de Ciencias, Universidad de Salamanca
Plaza Caídos, 1, 37008 Salamanca
{alberto.botana, angelica, abg, adeluis}@usal.es

Resumen Desde finales del siglo XIX, las técnicas de perfilado criminal han probado ser de gran utilidad a la hora de las investigaciones en casos de asesinos en serie. En este trabajo se lleva a cabo el desarrollo de un sistema de obtención de reglas de asociación con el fin de obtener información relevante a la obtención de perfiles de asesinos en serie. Se desarrolla además una aplicación de redes neuronales para predecir el arma homicida, el grupo racial del criminal y la relación entre víctima y agresor en nuevos casos, así como completando casos actuales con faltas de información. Se efectúa una descripción de los problemas a tratar, tras lo cual se lleva a cabo un desglose del código y discusión de los resultados.

Keywords: minería de datos, reglas de asociación, inteligencia artificial, redes neuronales, criminología, perfilado de asesinos en serie

1. Introducción y motivaciones.

La primera constancia de los asesinatos en serie se remontan al siglo IV a.C, donde se establece el Círculo del Veneno en la Antigua República Romana, en la que una coalición de 170 “matronas”, mujeres de clase alta y distinción dentro de la República, se dedicaron a envenenar entre todas a más de 90 hombres [12]. Pero la primera mención de un asesino en serie que actuó por su propia cuenta y de forma no organizada fue Liu Pengli, un príncipe chino en el siglo II a.C quien, durante un periodo entre los años 144 y 116 a.C, asesinó a más de 100 civiles del reino puramente por entretenimiento. Una vez fue descubierto, en lugar de ser ejecutado, como le recomendó la Corte Imperial a su tío, el Emperador Jing de Han, fue exiliado del reino y sus derechos reales le fueron revocados, devolviéndolo a un estatus de persona mundana [16]. Desde entonces, los asesinos en serie han tenido apariciones relativamente constantes a lo largo de la historia. Una de las técnicas más efectivas para combatir a estos criminales es el análisis de conducta criminal, que comenzó en el año 1888 de la mano de los doctores George Phillips y Thomas Bond, quienes analizaron al conducta de Jack el Destripador [3]. Esta técnica permite, basándose en el comportamiento de

criminales anteriores, extrapolar las próximas acciones del delincuente a estudiar. En este trabajo se busca una forma de, utilizando las técnicas de computación y minería de datos más relevantes de los últimos años, construir un programa que pueda facilitar el trabajo de la perfilación psicológica de estos asesinos.

Este trabajo se ha focalizado en la implementación de un sistema para obtener reglas de asociación en el conjunto de datos utilizado por la Unidad de Análisis de Conducta Criminal de la Universidad de Salamanca (UACC-USAL) así como en la aplicación de un sistema de predicción con redes neuronales a un conjunto de datos criminalísticos americanos con el fin de poder aplicarlo en bases de datos españolas una vez estén disponibles.

En la sección 1 se presenta una revisión del estado del arte del uso de las reglas de asociación así como de las redes neuronales, las dos técnicas utilizadas para el desarrollo del programa. A continuación, en la sección 1 se desgranar los objetivos de este trabajo, mientras que en la sección 1 se lleva a cabo una descripción de la base de datos a utilizar, tras lo cual, en la sección 1 se explica de forma esquemática los pasos que llevará el programa, el cual arrojará unos resultados discutidos en la sección 1. Por último, en la sección 1 se presentan las conclusiones así como las posibles futuras líneas de expansión del trabajo.

2. Revisión del estado del arte

Si bien el estudio de perfiles de asesinos en serie es un tema sobre el que se ha trabajado de forma más que exhaustiva, la aplicación de inteligencia artificial o de técnicas computacionales a este campo no está del todo extendida, encontrando un único artículo en toda la bibliografía consultada referente al tema, en el que Yasnitsky *et al.* [21] desarrollan un modelo de redes neuronales para predecir las tendencias violentas y poder predecir nuevos casos antes de que los asesinos empiecen a actuar. A continuación se efectúa una revisión del estado del arte de las técnicas de inteligencia artificial utilizadas en este trabajo.

2.1 Reglas de asociación

La primera mención de las reglas de asociación como se conocen hoy en día aparece en el año 1966 [7], pero la primera aplicación práctica de estas así como la definición que las puso en el panorama de la ciencia de datos viene en el año 1993 en el artículo de Agrawal *et al.* [1]. En él se muestra un algoritmo pensado para iterar sobre una gran cantidad de transacciones de un supermercado con el objetivo de mostrar qué productos suelen comprarse juntos, pudiendo así las personas responsables de la ubicación de los productos utilizar esta información para planificar la colocación de estos dentro de sus centros. En este artículo, por ejemplo, se identifica la regla de que el 90 % de las personas que compraron pan y mantequilla compraron también leche.

Las reglas obtenidas con este algoritmo presentan tres métricas que serán utilizadas extensivamente en este trabajo para expresar el valor de dicha regla:

- Confidence, o *confianza*: Definido por la ecuación $conf = \frac{A+B}{A}$, significa el número de veces que los elementos A y B aparecen juntos dividido entre el número de veces que aparece sólo A, esto es, *cómo de a menudo aparece B cuando aparece A*.
- Support, o *soporte*: Definido por la ecuación $supp = \frac{A+B}{total}$, significa el *número de veces que aparecen los elementos A y B juntos en total* de la base de datos.
- Lift, o *palanca*: Se define por $lift = \left(\frac{\left(\frac{A+B}{A} \right)}{\left(\frac{B}{total} \right)} \right)$, es el concepto más complejo de comprender, y su definición formal es que significa el aumento de confianza de que el elemento B vaya a aparecer sabiendo que A ha aparecido. Otra forma de definirlo es la correlación que hay entre A y B frente a la que habría si fueran independientes. De esta forma, *si A y B tienen un lift de 5, significa que es 5 veces más probable que aparezcan juntos que si fueran independientes*. Un valor de lift superior a 1 indica que están más relacionados de lo normal, y un valor menor a 1 indica que están inversamente relacionados, esto es, que es menos probable que B aparezca si A aparece.
- Rule Power Factor (RPF) o *factor de potencia de la regla*: Métrica reciente y no utilizada en este trabajo y propuesto por Kumar *et al.* [10], se obtiene multiplicando la confianza por el soporte tal que $RPF = conf(A \rightarrow B) \cdot supp(A)$.

En aplicaciones más recientes, Alwidian *et al.* aplican sistemas de minería de reglas de asociación para la detección del cáncer de mama [2] en Amman, Jordania obteniendo ratios de acierto en nuevos casos del 96 % partiendo de datos médicos de los pacientes; mientras que Yao *et al.* los aplican para la búsqueda de clientes en la industria automovilística [20].

2.2 Redes neuronales

Las redes neuronales, cuyo concepto fue introducido en 1943 por McCulloch y Warren [13], son modelos computacionales de aprendizaje automático inspirados en las redes neuronales biológicas presentes en los cerebros de los animales. Estos modelos llevan a cabo su aprendizaje analizando ejemplos en pares “datos-resultado” aprendiendo a asociar las entradas que recibe con la salida correspondiente. Uno de los usos que le dio su fama fue la aplicación de estos modelos por LeCun *et al.* [11] en el Sistema Postal Americano con la finalidad de reconocer automáticamente los números manuscritos.

Las redes neuronales están compuestas por tres conjuntos de capas: Una de entrada, con un número determinado de neuronas; uno de salida, con un número de neuronas de salida dependiente del problema al que se le asocia; y otro conjunto intermedio de capas ocultas, cada una con un número de neuronas establecido arbitrariamente. El diseño de esta estructura de capas ocultas puede afectar significativamente al rendimiento de la red, pero no hay ningún tipo de reglas conocidas sobre qué forma deben tomar.

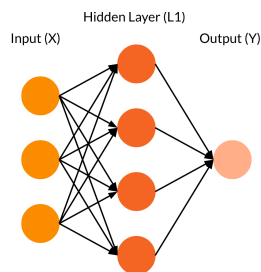


Figura 1: Estructura básica de una red neuronal con una capa oculta.

Cada capa de una red neuronal está conectada a la siguiente mediante una matriz de pesos de dimensiones $m \times n$, siendo m el número de neuronas de la primera capa y n el de la segunda. Estos pesos obtienen valores numéricos que oscilan entre -1 y 1, se inicializan de forma aleatoria, y sus valores se irán modificando a medida que la red se entrene sobre un conjunto de datos. Para ello se utiliza una “función de activación”, que dictará cuánto tienen que variar los pesos en función de cuánto dista el valor predicho del valor original que debería de tener. Las funciones de activación más comunes en este ámbito son:

- reLU: Definida por la función $f(x) = \max(0, x)$ y utilizada por primera vez por Hahnloser *et al.* en el año 2000 [6], es comúnmente aplicada a redes neuronales convolucionales y en sistemas de aprendizaje profundo.
- Identidad: Definida por la función $f(x) = x$, aplicada ampliamente en problemas linealmente separables.
- Sigmoide: Definida por la función $f(x) = \frac{1}{(1+e^{-x})}$, es posiblemente la curva más aplicada en este ámbito.

Otro concepto a explicar es el de matrices de confusión y métricas de rendimiento. Cuando se entrena una red neuronal sobre una base de datos, esta base se divide en dos subconjuntos: Uno de entreno y otro de prueba. La red a continuación se entrena sobre el conjunto de entreno y, una vez listo, se le pide a la red que prediga una serie de valores partiendo de los valores presentes en el conjunto de prueba. De esta forma, analizando si los valores emitidos por la red coinciden con los presentes en el conjunto de prueba se pueden establecer diversas métricas de rendimiento. En una matriz de confusión se representan gráficamente el número de aciertos que ha tenido el modelo y el número de fallos, clasificándolos según cómo haya fallado. En la matriz de confusión de ejemplo presente en la tabla 1 se muestra una matriz de ejemplo explicando los posibles valores que se pueden obtener:

- Verdadero Positivo (TP): Cuando la red predijo que el valor era positivo y, en efecto, era positivo.
- Falso Negativo (FN): La red predice como negativo, pero el resultado real era positivo.
- Falso Positivo (FP): Se predice el resultado como positivo, pero realmente era negativo.
- Verdadero Negativo (TN): La red predice un resultado negativo que, en efecto, era negativo.

		Valores predichos	
		Sí	No
Valores reales	Sí	TP	FN
	No	FP	TN

Tabla 1: Matriz de confusión de ejemplo.

Partiendo de los valores obtenidos en una matriz de confusión, se pueden calcular diversas métricas de rendimiento para un sistema, siguiendo las siguientes fórmulas:

- Accuracy o exactitud: Definida por la función $\frac{TP+TN}{TP+TN+FP+FN}$, esta métrica nos proporciona el porcentaje de casos totales que el modelo ha acertado.
- Recall o exhaustividad: Dada por la ecuación $\frac{TP}{TP+FN}$, nos proporciona el porcentaje de casos positivos que el modelo ha clasificado como tal.
- Precisión: Expresado como $\frac{TP}{TP+FP}$, nos informa de la proporción de los casos predichos como positivos que eran realmente positivos.
- Puntuación F1: Métrica dada por la ecuación $F1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$ que combina la información de precisión y exhaustividad.

La principal ventaja que aporta utilizar un modelo de redes neuronales es la gran versatilidad que ofrece, dado que un sistema relativamente simple puede cobrar una gran fuerza utilizando datos de cualquier tipo, siempre y cuando sean procesados de la manera correcta para su implementación. De esta forma, se puede utilizar una red neuronal tanto para reconocer gatos en fotos [5] como para generar canciones de jazz [18], aunque su uso que más importancia ha cobrado en los últimos años es su aplicación en la medicina para la detección y predicción de nuevos casos de cáncer [19] [17].

Recientemente, Crespo *et al* [4] investigaron la aplicación de reglas de asociación junto con redes neuronales en un entorno de monitorización energética y gestión de recursos. En su investigación proponen un "Proceso de Monitorización de Rendimiento de Recursos y Gestión de Fiabilidad", el cual propone recomendaciones para usos alternativos de recursos así como su reutilización y mantenimiento. Además, el proceso propuesto analiza las diferentes opciones de renovación y reinversión disponibles para recursos físicos. Para ello, idearon un sistema dividido en cuatro módulos como se presenta en el diagrama de modelado funcional presente en la figura 2.

En este diagrama se muestran los cuatro subprocesos de este método. En el primero, denominado de Procesado de Datos, se ejecutan las tareas de preparación de los datos para su utilización en los siguientes módulos así como de su normalización. El segundo módulo, denominado de Modelado de Red Neuronal, es el encargado de dar un algoritmo para identificar desviaciones de los recursos de sus comportamientos habituales bajo condiciones de operación regular. Estas desviaciones vienen dadas en forma de errores de predicción. El tercer módulo, de Minado de Datos para Reglas de Asociación, será el encargado de la medición

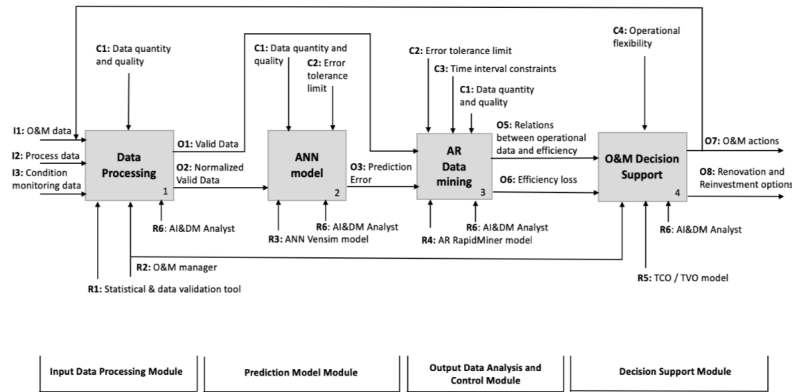


Figura 2: Sistema de gestión de recursos de Crespo et al.

de dicha desviación y obtención de la relación entre datos de operación y eficiencia, tras lo cual calculará la pérdida de eficiencia en un tiempo determinado. Finalmente, el último módulo toma esta información para tomar decisiones de gestión de recursos, y reevaluar posibles futuras acciones.

3. Objetivos

Los objetivos de este proyecto son los siguientes:

1. Importado de la base de datos y preprocesamiento para su aplicación práctica.
2. Implementación del algoritmo Apriori para la obtención de reglas de asociación útiles en investigaciones policiales.
3. Desarrollo de un sistema de filtrado de dichas reglas para deshechar aquellas triviales o inservibles.
4. Desarrollo de un sistema complementario de filtrado interactivo para que el usuario pueda obtener las reglas que necesite al momento.
5. Entrenamiento e implementación de un sistema de redes neuronales que prediga los valores de ciertas categorías determinadas de importancia prioritaria a la hora de la investigación.
6. Analizar y estudiar el rendimiento de dichas redes, optimizándolas para obtener los mejores resultados posibles.

Debido a que en mitad del desarrollo del proyecto aparecieron los efectos de la pandemia mundial del COVID-19, fue necesario hacer un cambio en el planteamiento del proyecto, dado que la base de datos que se estaba utilizando se encontraba en proceso de digitalización, que se vio detenido debido a los sucesos ya mencionados. Cuando se paralizó el proceso, la base de datos constaba únicamente de 28 entradas, por lo que fue necesario buscar una nueva base de datos de índole similar, pero de un tamaño más adecuado para la aplicación

de algoritmos de inteligencia artificial y más ajustado a estos objetivos. En el momento de este cambio de rumbo el desarrollo del sistema referente a las reglas de asociación estaba casi terminado, por lo que se decidió tomar esta primera base de datos como una prueba de concepto sobre la que acabar de desarrollar el sistema de reglas de asociación y estudiar su resultado, a fin de poder reutilizar el código sobre la segunda base de datos. De esta forma, cuando la situación acompañe, se podrá continuar con el desarrollo que inicialmente se tomó.

A fin de profundizar en el proceso, se explicará este primer desarrollo para la primera base de datos en la sección 1, en la sección 1 en qué consistieron los cambios realizados y, en la sección 1, cómo éstos proporcionaron resultados estadísticamente más relevantes tras su aplicación a la segunda base de datos.

4. Base de datos

4.1 Proyecto de Investigación de Asesinos en Serie de la UACC-USAL

En este trabajo se han utilizado dos bases de datos distintas. Para el primer apartado, se ha utilizado una base ya existente en papel que contiene datos pertinentes al comportamiento psicológico y criminal de diferentes asesinos en serie. En el momento en el que se detuvo la digitalización de la base de datos, había 28 asesinos diferentes registrados, y recogía cada uno un total de 298 características distintas, registradas en el libro publicado por de Santiago y Sánchez-Gil [8]. Las variables registradas, las cuales se pueden ver en su totalidad en el trabajo original, se ordenan en las siguientes categorías:

- Variables Sociodemográficas y Físicas: Datos como nombre, apellidos, alias, descripción física (Estatura, complexión) y contenidos sociales como estado civil, nivel económico y educativo o ideología política.
- Variables Psicológicas: Variables referentes a la presencia de anomalías psicológicas en el criminal como rasgos psicopáticos, distintas patologías mentales, conductas parafilicas o de abuso de sustancias.
- Antecedentes en Infancia y Adolescencia: Datos relativos a eventos traumáticos ocurridos durante la infancia y adolescencia, así como consecuencias que puedan tener en la actualidad: Pérdida de alguien importante, maltrato físico o psicológico, negligencia o abandono, presencia actual de tratamiento psicopatológico al respecto...
- Historial en edad adulta: Similar a la sección anterior, pero en la actualidad.
- Antecedentes familiares: Variables que califican la presencia de patologías mentales en la familia, funcionalidad de la misma, así como su nivel socio-económico.
- Historial delictivo y de violencia: Presencia de episodios previos, antecedentes, condenas o estancias en prisión.
- Escena: Datos pertinentes a la actividad criminal (Número de crímenes, máximo y mínimo de escenas por crimen, ubicación...) y tipificación de la escena así como su contenido.

- **Modus Operandi:** Categoría más extensa de la base de datos, incluye todo tipo de detalles en la ejecución de los crímenes: Desde nivel de planificación y métodos de ataque y control que utiliza el asesino hasta franja horaria en la que se comete el crimen.
- **Ritual:** Definido como actos perpetrados en la escena del crimen que eran innecesarios para la realización del asesinato, y que se encuentra presente en más de dos víctimas del mismo asesino. Estas variables incluyen comportamientos como la toma de trofeos o interacciones sexuales con las víctimas.
- **Comportamiento Geográfico:** Datos referentes a la zona o zonas de actuación del criminal, así como el método en el que llega a ella, y cuánto se desplaza como máximo.
- **Victimología:** Características de las víctimas del asesino: Relación con ellas, grupo racial, orientación sexual, sexo...
- **Tipologías:** Seis categorías distintas de asesino tipificadas por distintos autores de lo largo de los años por distintos autores como el Holmes y DeBurguer [9], Canter y Salfati [15], Fox y Levin [14] o por organizaciones como el FBI.

4.2 Crímenes en EE.UU 1980-2014

La segunda base de datos, disponible en la página de Kaggle a través de este enlace¹, contiene datos del Informe Suplementario de Homicidios del FBI desde 1980 hasta el 2014, así como 22.000 casos que fueron desclasificados gracias al Acta de Libre Información. Esta base de datos, de más de 638.000 casos, contiene las siguientes características.

1. ID: Número de identificación del caso.
2. Código y nombre de comisaría: Características burocráticas para identificación de la comisaría en la que se archivó el caso.
3. Tipo de agencia: Indica a qué cuerpo de seguridad pertenecía la agencia en la que se clasificó el caso: Sheriff, Policía municipal, estatal, regional, especial...
4. Ciudad y estado en la que se cometió el crimen.
5. Mes y año del crimen.
6. Incidente: Esta columna tienen un significado desconocido, dado que en la base de datos no venía explicada y, en la investigación para realizar este trabajo no se consiguió encontrar los documentos oficiales del FBI referentes a la base de datos original, por lo cual no se puede atribuir un significado claro. Por esto mismo se ignorará este parámetro.
7. Tipo de crimen: Esta base de datos clasifica todas las entradas en dos tipos de crimen: Homicidios por negligencia y Asesinato/Homicidio. Para la realización de este trabajo, se descartarán aquellos producidos por negligencia.
8. Caso resuelto: Variable binaria que indica si el caso ha sido resuelto a fecha de publicación de la base de datos o no.
9. Datos de la víctima: Sexo, edad, etnia y grupo racial de la víctima.
10. Datos del criminal: Sexo, edad, etnia y grupo racial de el criminal.
11. Relación entre víctima y criminal.

¹<https://www.kaggle.com/murderaccountability/homicide-reports>

12. Arma utilizada en el crimen
13. Víctimas adicionales
14. Criminales adicionales
15. Fuente de información: FBI o Acta de Libre Información (FOIA) por sus siglas en inglés.

5. Modelo propuesto

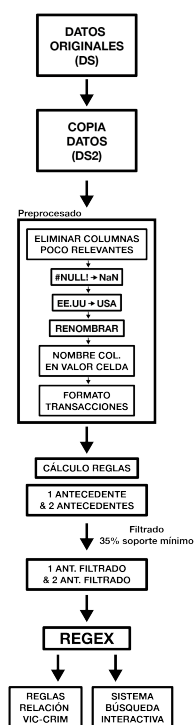


Figura 3: Diagrama del primer programa, que actúa sobre la base de datos UACC-USAL.

El modelo propuesto, cuyo código se encuentra disponible públicamente en este enlace² y en este enlace³, y se esquematiza en las figuras 3 y 4, consta de dos programas: Uno por cada base de datos utilizada.

5.1 Proyecto de Investigación de asesinatos en serie de la UACC-USAL

En el primer programa se ejecutará el código pertinente a la base de datos de la UACC-USAL. En él, el primer paso a tomar será la importación y preprocesado de datos, eliminando columnas innecesarias o que aporten poca o nada de información. Además, se aplica un algoritmo diseñado expresamente para este trabajo que añade a cada celda el nombre de su columna correspondiente. Esto se hace porque, al obtener las reglas, sólo se mencionan los valores, sin proporcionar información de qué columna provienen, proporcionando resultados de "Sí⇒No", sin ningún contexto de a qué variables se refiere. Tras ello, se crea la base de datos de transacciones que se le dará al algoritmo Apriori que calculará las reglas de asociación. Con las reglas una vez obtenidas, se lleva a cabo un filtrado de las reglas y se define un algoritmo para representarlas de forma más fácilmente inteligible.

5.2 Crímenes en EE.UU 1980-2014

En el segundo programa, que trabaja sobre la base de datos americanos, se realiza primeramente un preprocesado en el que se borrarán las columnas que no aportan datos utilizables para las predicciones o establecimiento de reglas, como el número de identificación del caso o la fuente de la cual proviene la información del caso. En este preprocesado cabe destacar la eliminación de las entradas cuyo valor del parámetro "Perpetrator Age" se encuentre entre 0 y 10 años, bien porque se considera como errata o porque este trabajo busca la aplicación de métodos de inteligencia artificial en casos de asesinato, y se toma como que un niño de

²<https://colab.research.google.com/drive/1lgEsY5CieON3Cs-yevGzcFsulcrnukZQ?usp=sharing>

³https://colab.research.google.com/drive/1SHvPnbQ9VCSS_SNQqzUBpjMBN83ea5xA?usp=sharing

menos de 10 años no podría llevar a cabo un asesinato con premeditación, siendo en tal caso un homicidio y quedando fuera del marco de objetivos del proyecto. Además, para intentar mejorar los resultados en una de las redes neuronales, se hace un estudio de la densidad de aparición de ciertas relaciones entre víctima y agresor, para reducir dimensionalidad descartando los valores cuya aparición sea menor a un 1 % de los casos totales. Una vez realizado el preprocesado, la base de datos se divide en tres grupos, divididos cada uno en subconjuntos de entreno y prueba partiendo de una misma semilla para garantizar iguales divisiones entre grupos. Con los grupos divididos e inicializados, se entrenarán tres redes neuronales diferentes. Para llevar a cabo el entrenamiento, se da la opción al usuario de realizar un estudio de rendimiento de las redes neuronales, el cual probará varias estructuras de capas ocultas y devolverá la que obtenga mejor resultado. En caso de no llevar a cabo el estudio completo, se analiza el rendimiento de la estructura seleccionada. Una vez las redes han sido entrenadas, se realiza un estudio detallado del mejor modelo obtenido. Para ello se calculan sus matrices de confusión y, con ellas, se calculan las métricas de rendimiento escogidas, lo cual se explica más en detalle en la sección 1. Tras obtener las redes neuronales, se aplican los algoritmos desarrollados en el primer programa para obtener reglas de asociación mediante el algoritmo Apriori y se lleva a cabo un filtrado de éstas, además de presentar una herramienta de búsqueda y filtrado interactivo. Finalmente, por petición de la UACC-USAL, se procede a buscar reglas referentes a la resolución del crimen, con el fin de obtener información respecto a qué factores pueden llevar a la paralización de una investigación.

6. Resultados experimentales

En esta sección se discutirán los resultados obtenidos en el programa. En el desarrollo del mismo se han calculado reglas para uno y dos antecedentes, pero en este informe, salvo que las reglas con dos antecedentes aporten información relevante, sólo se mostrarán las primeras

6.1 Proyecto de Investigación de Asesinos en Serie de la UACC-USAL

Las primeras reglas obtenidas en este caso de estudio se representan en la tabla 2, donde se recogen todas las reglas calculadas para uno y dos antecedentes en la base de datos del Proyecto de Investigación de Asesinos en Serie de la UACC-USAL. Como se puede observar, las reglas obtenidas son de carácter bastante trivial y redundante, ya que la información extraída de estas reglas es que “Si un asesino actúa en norteamérica/es de norteamérica, entonces es de Estados Unidos”, lo cual no siempre es cierto (El criminal podría ser canadiense) pero en este caso aplica siempre. Esto es debido a que la base de datos, en el momento de redacción de este trabajo, vio interrumpido su progreso de digitalización y, a causa de esto, no existe una gran variedad en los datos, causando que estas reglas tengan un soporte más alto del esperado.

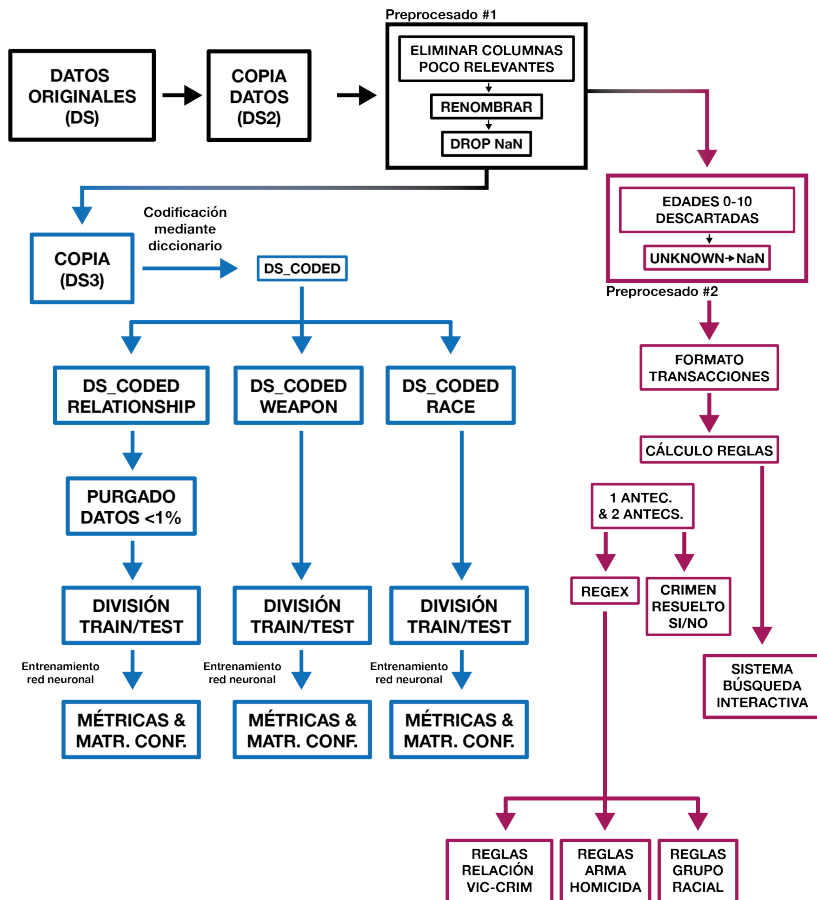


Figura 4: Diagrama del segundo programa. La separación en colores es para identificar los procesos de cada subprograma (Rojo: Reglas de asociación. Azul: Redes neuronales)

Antecedent	Consequent	Confidence	Support	Lift
ZONA.ACTUACION AMERICA NORTE	PAIS.ACTUACION USA	0,9000	0,3333	2,7000
PAIS.ACTUACION USA	ZONA.ACTUACION AMERICA NORTE	1,0000	0,3333	2,7000
PAIS.PROCEDENCIA USA	PAIS.ACTUACION USA	0,8182	0,3333	2,4545
PAIS.ACTUACION USA	PAIS.PROCEDENCIA USA	1,0000	0,3333	2,4545
FORMAS.CAUSAR.MUERTE.ARMA.BLANCA SI	ARMAS.BLANCAS SI	1,0000	0,3333	2,4545

Tabla 2: Reglas de asociación de la UACC-USAL con un antecedente.

Para solucionar estos problemas de trivialidad, se le aplica un filtrado al soporte mínimo necesario, obteniendo unas reglas como las que se muestran en la tabla 3. Se observa que ahora las reglas obtenidas presentan mayor utilidad, si bien se presenta alguna que otra regla bastante redundante (“La presencia de indicios de armas blancas en la escena del crimen implica que dicho crimen se cometió con un arma blanca”).

Antecedent	Consequent	Confidence	Support	Lift
ABUSO.SUSTANCIAS NO	FORMAS.CAUSAR.MUERTE.ARMA.BLANCA SI	1.0	0,3704	2,700
FORMAS.CAUSAR.MUERTE.ARMA.BLANCA SI	ABUSO.SUSTANCIAS NO	1.0	0,3704	2,700
ARMAS.BLANCAS SI	FORMAS.CAUSAR.MUERTE.ARMA.BLANCA SI	1.0	0,3704	2,700
FORMAS.CAUSAR.MUERTE.ARMA.BLANCA SI	ARMAS.BLANCAS SI	1.0	0,3704	2,700
ABUSO.SEX.ENTORNO.ESCOLAR NO	ARMAS.BLANCAS NO	1.0	0,3704	2,455

Tabla 3: Reglas de asociación de la UACC-USAL con soporte filtrado para un antecedente

Siguiendo las recomendaciones de la UACC-USAL, se desarrolló también un sistema de búsqueda dentro de las reglas obtenidas y, más concretamente, se denotó el interés de la organización por encontrar información relevante a la relación entre criminal y víctima, cuyas reglas obtenidas se muestran en las reglas de las tablas 4 y 5.

Antecedent	Consequent	Confidence	Support	Lift
MET.AT. FZA. CON O SIN ARMA	REL.VICTIMAS.NINGUNA SI	0,8333	0,3704	2,2500
ARMAS.BLANCAS NO	REL.VICTIMAS.DESCONOCE NO	0,9091	0,3704	2,2314
ARMAS.BLANCAS NO	REL.VICTIMAS.FAMILIAR.LEJANO NO	0,9091	0,3704	2,2314
ARMAS.BLANCAS NO	REL.VICTIMAS.OTRA NO	0,9091	0,3704	2,2314
ARMAS.BLANCAS NO	REL.VICTIMAS.RELACION.SENTIMENTAL NO	0,9091	0,3704	2,2314

Tabla 4: RA de la UACC-USAL con un antecedente y cuyo consecuente sea la relación entre asesino y víctima.

Se observa que las reglas obtenidas en este caso no son de gran utilidad, dado que la única información que aportan es que se desconoce la relación entre víctima y criminal o bien niega algunas opciones, pero sin aportar información esclarecedora. Esto, de nuevo, es debido a que la base de datos presenta ahora

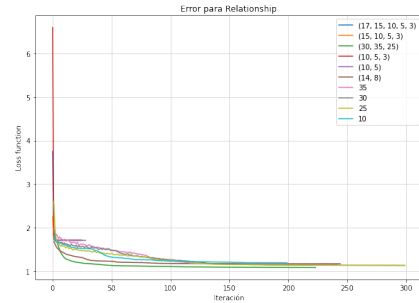
Antecedent 1	Antecedent 2	Consequent	Confidence	Support	Lift
ANTECED.POLICIALES SI	EDAD.VICTIMAS.DESCONOCE NO	REL.VICTIMAS.NINGUNA SI	1,0000	0,3704	1,9286
CANIBALISMO NO	TIEMPO.MIN.ENTRE.AGRESIONES 1-60 MIN	REL.VICTIMAS.NINGUNA SI	1,0000	0,3704	1,9286
CREENCIA.RELIG. CRISTIANA	REL.VICTIMAS.FAMILIAR.CERCANO NO	REL.VICTIMAS.NINGUNA SI	1,0000	0,3704	1,9286
EPIS.PREVIOS.VIOLENCIA SI	SEL.VICTIMAS.POSIBILIDAD SI	REL.VICTIMAS.NINGUNA SI	1,0000	0,3704	1,9286
EVID.INDICIOS.BIOLOGICOS SI	REL.VICTIMAS.FAMILIAR.CERCANO NO	REL.VICTIMAS.NINGUNA SI	1,0000	0,3704	1,9286

Tabla 5: RA de la UACC-USAL con dos antecedentes y cuyo consecuente sea la relación entre asesino y víctima

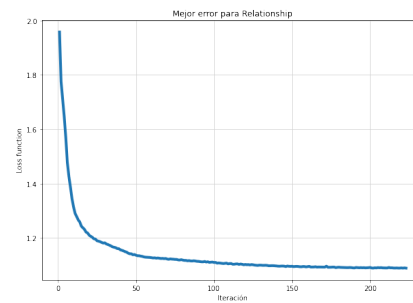
mismo un tamaño reducido pero, a medida que se expanda, las reglas proporcionadas tendrán mayor nivel de información útil para aportar. Por ello, los resultados obtenidos en esta sección deberán interpretarse como provisionales y pertenecientes a una prueba de concepto, que se presentan como una vía de trabajo sobre la que continuar una vez la situación acompañe.

6.2 Crímenes en EE.UU 1980-2014

6.2.1 Aplicación de redes neuronales Se ha llevado a cabo un estudio del rendimiento para distintas estructuras internas en las tres redes neuronales. Para cada una, se obtiene una representación gráfica de la evolución del rendimiento para todas las estructuras, y otra representación separada para la evolución de la estructura que obtuvo mejor puntuación. Esto se lleva a cabo para las funciones de activación “reLU” e “Identity”, dos de las más utilizadas en el estado del arte. A continuación, en las gráficas 5a a 7b, se muestran las evoluciones de las redes neuronales entrenadas con la primera función de activación, mientras que las evoluciones de los modelos entrenados con la función “Identity” se pueden ver en las figuras 8a a 10b.



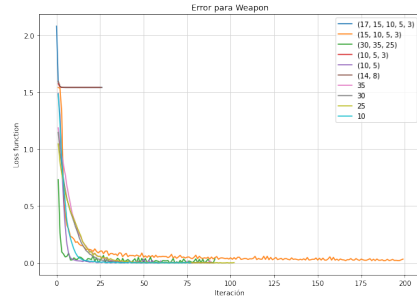
(a) Evolución de todas las estructuras para relación víctima-criminal (reLU).



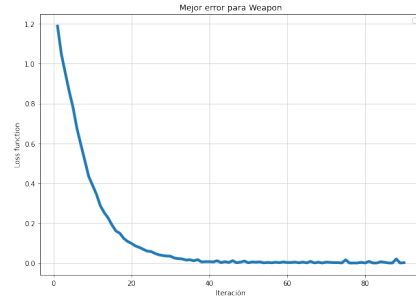
(b) Evolución de la mejor estructura para relación víctima-criminal (reLU).

Figura 5: Resultados obtenidos para relación víctima-criminal (reLU).

Los resultados de entrenamiento de todas las estructuras iteradas para todos los modelos probados se muestran en la tabla 6, y las redes neuronales una vez

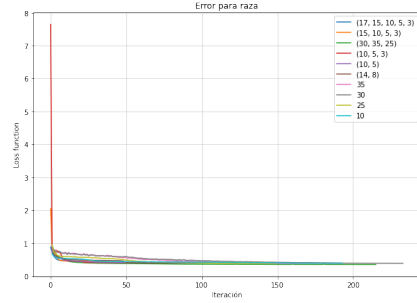


(a) Evolución de todas las estructuras para arma homicida (reLU).

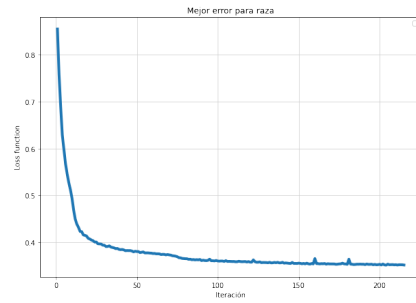


(b) Evolución de la mejor estructura para arma homicida (reLU).

Figura 6: Resultados obtenidos para arma homicida (reLU).



(a) Evolución de todas las estructuras para Race (reLU).



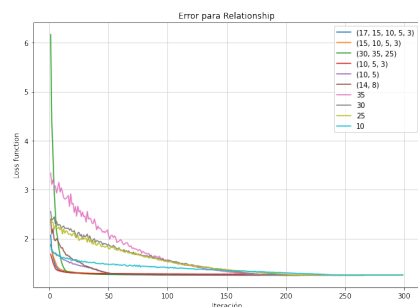
(b) Evolución de la mejor estructura para Race (reLU).

Figura 7: Resultados obtenidos para Race (reLU).

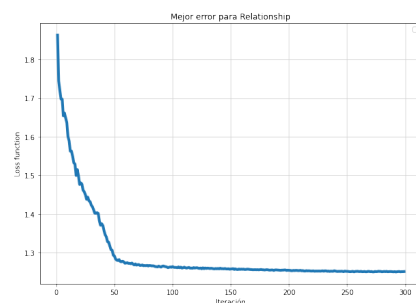
optimizadas se recogen en la tabla 7, junto con la estructura y función de activación, además del ratio de aciertos obtenido. La notación de estas estructuras se interpreta de la siguiente forma: Una estructura tal que (15, 10, 5, 3) se deberá tomar como una estructura de 4 capas ocultas, cada una con 15, 10, 5 y 3 neuronas, respectivamente. Se observa en la tabla 7 que la red neuronal pertinente al arma homicida tiene menos resultados. Esto es debido a que, como en varias pruebas se obtuvo un rendimiento máximo, se optó por no llevar a cabo más pruebas con ese modelo. Se observa también que, de forma sistemática, se obtienen mejores resultados en aquellas redes neuronales entrenadas con la función de activación “reLU”, presentando un aumento del rendimiento de entorno al 5 %.

Además de las puntuaciones preliminares obtenidas al entrenar, se ha calculado también las métricas de rendimiento para las tres redes, mostradas en la tabla 8.

Se observa que para las redes respectivas a la predicción del arma homicida así como del grupo racial del criminal se obtienen resultados bastante buenos,

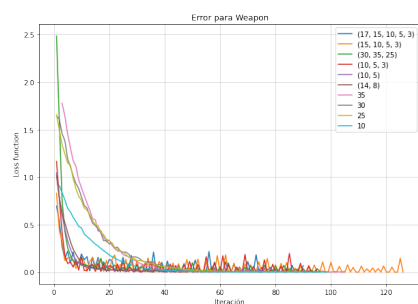


(a) Evolución de todas las estructuras para relación víctima-criminal (Identity).



(b) Evolución de la mejor estructura para relación víctima-criminal (Identity).

Figura 8: Resultados obtenidos para relación víctima-criminal (Identity).

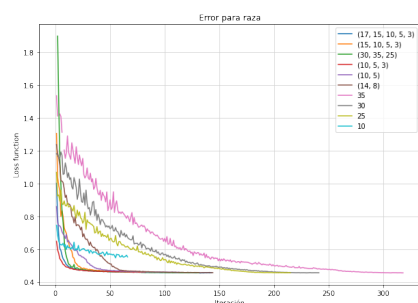


(a) Evolución de todas las estructuras para arma homicida (Identity).

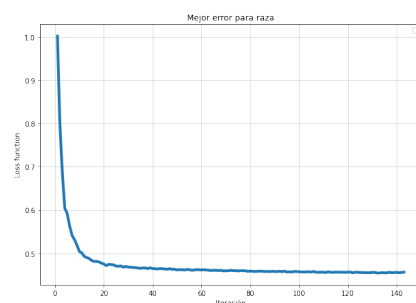


(b) Evolución de la mejor estructura para arma homicida (Identity).

Figura 9: Resultados obtenidos para arma homicida (Identity).



(a) Evolución de todas las estructuras para Race (Identity).



(b) Evolución de la mejor estructura para Race (Identity).

Figura 10: Resultados obtenidos para Race (Identity).

Estructura	RELACIÓN Activación		ARMA Activación		RAZA Activación	
	Relu	Identity	Relu	Identity	Relu	Identity
(17,15,10,5,3)	0,3870	0,4815			0,8668	0,8718
(15, 10, 5, 3)	0,3870	0,4759			0,8645	0,8714
(30, 35, 25)	0,5564	0,4981			0,8792	0,8713
(10, 5, 3)	0,3870	0,4840	0,9862	1,0000	0,8726	0,8686
(10, 5)	0,3870	0,4995	0,9941	1,0000	0,8729	0,8701
(14, 8)	0,5072	0,4949			0,8689	0,8709
35	0,5270	0,4882			0,8738	0,8716
30	0,5390	0,4978			0,8756	0,8683
25	0,5388	0,4987			0,8735	0,8715
10	0,5169	0,4985	1,0000	1,0000	0,8725	0,8689

Tabla 6: Resultados del entreno de estructuras para las redes neuronales.

	Estructura capas ocultas	Función act.	Aciertos
Relación	(30, 35, 25)	reLU	55,64 %
Arma	(10)	reLU/Identity	100 %
Raza	(30, 35, 25)	reLU	87,92 %

Tabla 7: Parámetros de las mejores redes neuronales obtenidas en el estudio.

	Exactitud	Precisión	F1
Relación	0,4954	0,4954	0,4954
Arma	1,0000	1,0000	1,0000
Raza	0,8780	0,8780	0,8780

Tabla 8: Métricas de rendimiento para las redes neuronales

obteniendo unos índices de aciertos del 100 % y del 87,92 %, respectivamente, aunque para la red neuronal que clasifica relación entre criminal y víctima, denominada por la UACC-USAL como de mayor prioridad, los resultados obtenidos son bastante mejorables, obteniendo un ratio de aciertos del 55,64 %, obtenido después de llevar a cabo el preprocesado en el que se purgaban los valores cuya presencia fuera menor al 1 % de los casos. La diferencia de rendimiento entre antes y después de efectuar este limpiado, como se observa en la tabla 9, no son muy significativos, mejorando el rendimiento en 5 % aproximadamente. Esto quizás podría haberse solventado haciendo un purgado con un umbral mínimo de presencia algo mayor, quizás en un 2 % o un 5 %, porque posiblemente este bajo rendimiento sea debido a que, aún habiendo reducido a la mitad el número de parámetros sobre los que entrenar, 11 sigue siendo un número relativamente elevado, sobre todo en comparación a la cantidad de datos disponibles en el momento del estudio.

Estructura	Con purgado		Sin purgado	
	Relu	Identity	Relu	Identity
(17,15,10,5,3)	0,3870	0,4815	0,4737	0,4411
(15, 10, 5, 3)	0,3870	0,4759	0,4105	0,4385
(30, 35, 25)	0,5564	0,4981	0,5009	0,4540
(10, 5, 3)	0,3870	0,4840	0,4322	0,4399
(10, 5)	0,3870	0,4995	0,3464	0,4483
(14, 8)	0,5072	0,4949	0,4766	0,4578
35	0,5270	0,4882	0,4902	0,4518
30	0,5390	0,4978	0,4642	0,4482
25	0,5388	0,4987	0,4799	0,4523
10	0,5169	0,4985	0,4639	0,4531

Tabla 9: Diferencia de rendimiento al aplicar purgado

Para visualizar mejor los resultados obtenidos, además de las puntuaciones numéricas y las métricas, se incluyen en las figuras 11 a 13 las matrices de confusión extraídas de las tres redes neuronales entrenadas con el mejor modelo dentro del estudio.

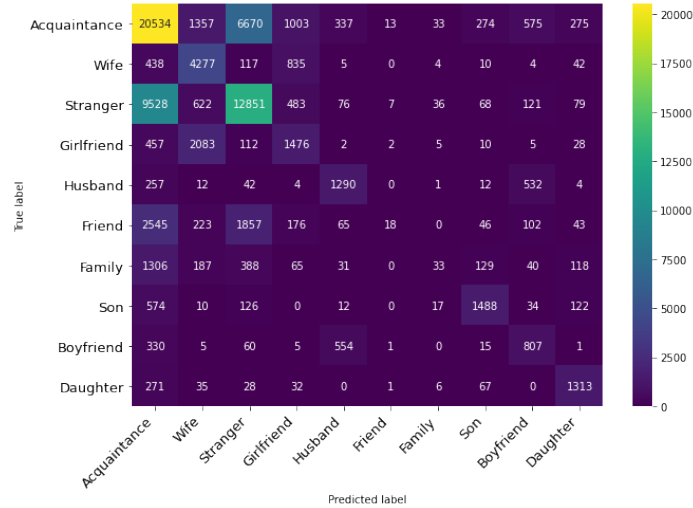


Figura 11: Matriz de confusión para relación víctima-criminal

En la figura 14 se muestra la comparativa entre la matriz de confusión de la red entrenada para relación antes y después de llevar a cabo el purgado.

Como se puede observar, los valores obtenidos caen ligeramente más sobre la diagonal, pero no hay una mejoría radical. De todas formas, analizando es-

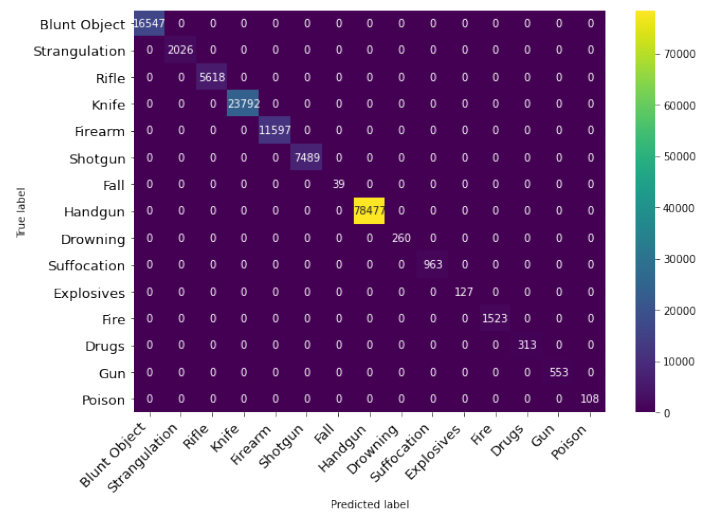


Figura 12: Matriz de confusi3n para arma homicida

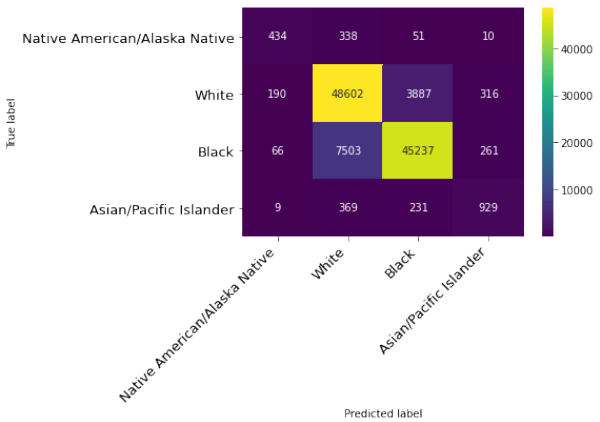
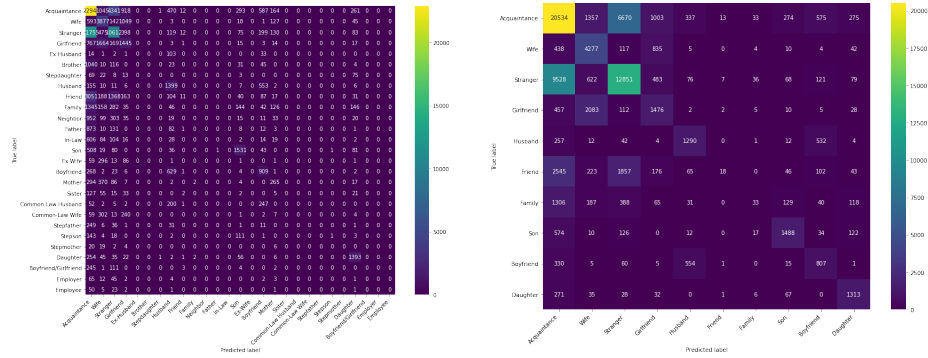


Figura 13: Matriz de confusi3n para Race



(a) Matriz de confusión de relación antes de llevar a cabo el purgado (b) Matriz de confusión de relación después de llevar a cabo el purgado

Figura 14: Comparación de matrices de confusión de relación antes y después del purgado

tos resultados se tiene que, para predicciones de “Acquaintance” (Conocidos) y “Stranger” (Extraños), tiene una precisión razonablemente buena.

6.2.2 Reglas de asociación De forma similar al caso de la UACC-USAL, se muestran las primeras reglas calculadas en las tablas 10 y 11. Como en este caso no existe el problema de la variedad de casos, dado que la base de datos cuenta con más de 600.000 entradas distintas, no es necesario llevar a cabo el filtrado por soporte. Se observa que, en este caso, el lift obtenido de las reglas es bastante mayor al caso de la UACC-USAL, también debido a la mayor riqueza de la base de datos, obteniendo valores de lift incluso mayores de 4. Las reglas obtenidas de esta base pueden ser tomadas para consideración en futuras investigaciones, si se filtran de la forma debida.

Antecedent	Consequent	Confidence	Support	Lift
Perp. Ethnicity: Not Hispanic	Perp. Race: Black	0,4918	0,1112	4,3662
Perp. Ethnicity: Not Hispanic	Crime Solved: Yes	0,9258	0,2094	4,0814
Perp. Ethnicity: Not Hispanic	Perp. Sex: Male	0,8051	0,1821	4,0514
Victim Ethnicity: Not Hispanic	Perp. Ethnicity: Not Hispanic	0,3383	0,1046	3,2128
Victim Ethnicity: Not Hispanic	Crime Solved: Yes	0,6772	0,2094	2,9986

Tabla 10: Reglas de asociación con un antecedente para la base de datos de Kaggle

Como los resultados de entreno de la red que clasifica la relación fueron bastante mejorables, se intenta abordar ese problema desde otro ángulo: Intentando

Antecedent 1	Antecedent 2	Consequent	Confidence	Support	Lift
Perp. Race: Black	Victim Ethnicity: Not Hispanic	Perp. Ethnicity: Not Hispanic	0,9876	0,1112	4,3662
Perp. Sex: Male	Victim Ethnicity: Not Hispanic	Perp. Ethnicity: Not Hispanic	0,9164	0,1821	4,0514
Perp. Ethnicity: Not Hispanic	Victim Race: Black	Victim Ethnicity: Not Hispanic	0,9935	0,1046	3,2128
Addtl. Perps.: 0	Perp. Ethnicity: Not Hispanic	Victim Ethnicity: Not Hispanic	0,9355	0,1835	3,0252
Perp. Ethnicity: Not Hispanic	Perp. Race: Black	Victim Ethnicity: Not Hispanic	0,9347	0,1112	3,0228

Tabla 11: Reglas de asociación con 2 antecedentes para la base de datos de Kaggle.

encontrar reglas de asociación cuyo consecuente indique la relación entre víctima y criminal. Algunas de las reglas obtenidas se muestran en la tabla 12. Como se puede observar, se encuentran reglas variadas, pero que todas presentan unas métricas de utilidad bastante bajas, teniendo en el mejor de los casos una confianza de poco más del 30 %, un soporte de menos del 20 %, y un *lift* no lo suficientemente alto como para compensar los valores pobres de los otros parámetros.

Antecedent	Consequent	Confidence	Support	Lift
Perp. Race: Black	Relationship: Acquaintance	0,3193	0,1076	1,6156
Perp. Sex: Male	Relationship: Acquaintance	0,2492	0,1553	1,4826
Perp. Sex: Male	Relationship: Stranger	0,1889	0,1177	1,4128
Crime Solved: Yes	Relationship: Acquaintance	0,2781	0,1942	1,4072
Crime Solved: Yes	Relationship: Stranger	0,1736	0,1212	1,2984
Victim Sex: Male	Relationship: Stranger	0,1727	0,1337	1,1333
Victim Race: Black	Relationship: Acquaintance	0,2209	0,1043	1,1178
Victim Sex: Male	Relationship: Acquaintance	0,2171	0,1681	1,0984
Agency Type: Municipal Police	Relationship: Stranger	0,1431	0,1108	1,0702
Addtl. Victims: 0	Relationship: Acquaintance	0,1723	0,1581	1,0251
Addtl. Victims: 0	Relationship: Stranger	0,1366	0,1254	1,0220

Tabla 12: Reglas de asociación de Kaggle con un antecedente y con consecuente “Relación”

Además, se ha realizado una búsqueda para obtener reglas cuyo consecuente arroja información sobre el arma homicida y sobre el grupo racial del criminal, de la misma forma que las redes entrenadas, para poder comparar la utilidad de un método frente al otro. Los resultados obtenidos se muestran en las tablas 13 a 15. Se observa, para el caso del arma homicida, que la mayoría de las reglas calculadas informan del uso de pistolas o cuchillos, y rara vez de otras armas, debido a que en la inmensa mayoría de los casos el arma homicida coincidía con éstas, siendo las demás mucho menos utilizadas. En el caso del grupo racial se observa que, para reglas con un solo antecedente el nivel de confianza obtenido es mucho menor que con dos antecedentes. Cabe destacar que, en el caso de

las reglas del grupo racial se obtienen unos valores de lift superiores al resto de reglas específicas para una categoría.

Antecedent	Consequent	Confidence	Support	Lift
Victim Race: Black	Weapon: Handgun	0,5668	0,2677	1,1366
Crime Solved: Yes	Weapon: Knife	0,1698	0,1185	1,1270
Victim Sex: Male	Weapon: Handgun	0,5445	0,4216	1,0918
Perpetrator Race: Black	Weapon: Handgun	0,5433	0,1831	1,0894
Crime Solved: No	Weapon: Handgun	0,5290	0,1596	1,0608
Additional Victims: 0	Weapon: Blunt Object	0,1118	0,1026	1,0557
Agency Type: Municipal Police	Weapon: Handgun	0,5207	0,4031	1,0442
Additional Perpetrators: 0	Weapon: Knife	0,1562	0,1366	1,0369
Additional Victims: 0	Weapon: Knife	0,1556	0,1427	1,0327
Agency Type: Municipal Police	Weapon: Knife	0,1553	0,1203	1,0312
Perpetrator Sex: Male	Weapon: Handgun	0,5007	0,3121	1,0042

Tabla 13: Reglas obtenidas de la base de datos de Kaggle referentes al arma homicida.

Antecedent	Consequent	Confidence	Support	Lift
Perp. Ethnicity: Not Hispanic	Perp. Race: Black	0,4918	0,1112	4,3662
Victim Race: White	Perp. Race: White	0,6156	0,3045	1,8259
Victim Race: Black	Perp. Race: Black	0,3280	0,1549	1,7916
Relationship: Acquaintance	Perp. Race: Black	0,5445	0,1076	1,6156
Perp. Sex: Male	Perp. Race: Black	0,2722	0,1697	1,4866
Perp. Sex: Male	Perp. Race: White	0,2123	0,1324	1,4617
Crime Solved: Yes	Perp. Race: White	0,4340	0,3031	1,4320
Crime Solved: Yes	Perp. Race: Black	0,4273	0,2983	1,4318
Weapon: Handgun	Perp. Race: Black	0,3123	0,1558	1,1672
Agency Type: Municipal Police	Perp. Race: Black	0,2042	0,1581	1,1154
Victim Sex: Male	Perp. Race: Black	0,2012	0,1558	1,0987
Victim Ethnicity: Not Hispanic	Perp. Race: Black	0,3643	0,1126	1,0808
Additional Victims: 0	Perp. Race: Black	0,1168	0,1071	1,0366
Victim Ethnicity: Not Hispanic	Perp. Race: White	0,3412	0,1055	1,0120

Tabla 14: Reglas de la base de datos de Kaggle con un antecedente referentes al grupo racial del criminal.

Finalmente, se llevó a cabo una búsqueda de reglas que aportasen información sobre cuándo un caso **no** es resuelto, para aprender qué factores pueden influir a que un caso quede estancado. Estas reglas se muestran en la tabla 16.

Se observa, en las reglas para determinar si un caso es resuelto o no, que ninguna supera un nivel de confianza del 40 %, ni se obtiene un *lift* muy alto,

Antecedent 1	Antecedent 2	Consequent	Confidence	Support	Lift
Perp. Sex: Male	Victim Race: Black	Perp. Race: Black	0,8976	0,2454	2,6633
Perp. Sex: Male	Victim Race: White	Perp. Race: White	0,8277	0,2730	2,4548
Additional Victims: 0	Victim Race: White	Perp. Race: White	0,6116	0,2723	1,8141
Agency Type: Municipal Police	Perp. Ethnicity: Not Hispanic	Perp. Race: Black	0,6066	0,1031	1,7997
Additional Perp.s: 0	Victim Race: White	Perp. Race: White	0,6045	0,2598	1,7929
Additional Victims: 0	Victim Race: Black	Perp. Race: Black	0,5969	0,2644	1,7710
Victim Race: White	Weapon: Handgun	Perp. Race: White	0,5908	0,1280	1,7521
Victim Race: White	Victim Sex: Male	Perp. Race: White	0,5898	0,2128	1,7494
Agency Type: Municipal Police	Victim Race: Black	Perp. Race: Black	0,5875	0,2412	1,7431
Victim Race: Black	Weapon: Handgun	Perp. Race: Black	0,5787	0,1549	1,7172

Tabla 15: Reglas de la base de datos de Kaggle con dos antecedentes referentes al grupo racial del criminal (10 primeras).

Antecedent	Consequent	Confidence	Support	Lift
Victim Race: Black	Crime Solved: No	0,3534	0,1380	1,1713
Agency Type: Municipal Police	Crime Solved: No	0,3518	0,1444	1,1660
Victim Sex: Male	Crime Solved: No	0,3454	0,1456	1,1448
Weapon: Handgun	Crime Solved: No	0,3201	0,1596	1,0608
Victim Race: White	Crime Solved: No	0,2841	0,1025	0,9415

Tabla 16: Reglas de la base de datos de Kaggle con un antecedente sobre la resolución del caso.

por lo que, en caso de querer aplicar estas reglas para predecir futuros casos no resueltos, habría que tomar las precauciones necesarias, dado que no presentan una suficiente robustez como para ser consideradas fiables, por lo que no deberían ser tomadas al pie de la letra.

7. Conclusiones y futuras líneas de trabajo.

En este trabajo se ha llevado a cabo el desarrollo de un sistema de búsqueda de reglas de asociación basadas en el algoritmo Apriori y se ha aplicado a dos bases de datos: Una perteneciente a la Unidad de Análisis de Conducta Criminal de la USAL, conteniendo datos sobre perfiles psicológicos de asesinos en serie; y otra disponible en Kaggle, conteniendo datos sobre todos los asesinatos y homicidios cometidos en Estados Unidos entre los años 1980 y 2014. Con las reglas obtenidas, se ha efectuado una serie de filtrados para optimizar la cantidad de información obtenible de las reglas proporcionadas. Se ha utilizado además un sistema basado en expresiones regulares para poder realizar búsquedas de forma sencilla, y en base a esto se ha creado un sistema de búsqueda interactivo para cualquier conjunto de reglas de asociación. Las reglas obtenidas del primer conjunto de datos aportan información de interés, pero hasta que no se pueda ampliar dicho conjunto, no deberán tomarse como nada más que una prueba de concepto sobre la que continuar trabajando en el futuro.

Para el segundo conjunto de datos se han implementado tres redes neuronales entrenadas sobre tres características distintas: Relación víctima-criminal, arma homicida y grupo racial del criminal. Se obtienen para estas dos últimas unos ratios de acierto suficientemente elevados como para plantearse el aplicarlo a casos reales (88 % en caso del grupo racial y un 100 % de aciertos para el arma homicida), pero en el caso de la predicción de la relación entre víctima y criminal se obtienen resultados cercanos al 56 %, por lo cual queda patente que es necesario llevar a cabo mayor investigación en esta aplicación. Sobre estas redes neuronales se ha llevado a cabo un estudio de estructuras internas para determinar cuál proporciona mejores resultados en cada caso, y se han obtenido métricas de rendimiento así como sus matrices de confusión para poder interpretar mejor los resultados obtenidos. Mediante estos métodos de optimización se ha conseguido aumentar en un 10 % el rendimiento de la red que predice relación entre víctima y criminal. Además, se le ha aplicado a este conjunto de datos el sistema de obtención de reglas desarrollado para el primer caso, obteniendo en la mayoría de los casos reglas con un valor de lift mayor de 2, lo que las convierte en reglas de interés para su análisis.

Como futuras líneas de trabajo, los objetivos prioritarios son:

- Continuar proyecto de la UACC-USAL, desarrollando una interfaz intuitiva para el programa y facilitando el acceso a la base de forma remota.
- Desarrollar un sistema de introducción interactivo de casos para la base de datos de la UACC-USAL, facilitando así su crecimiento.
- Aplicar un sistema de redes neuronales, clasificando para los distintas tipificaciones del perfil.
- Mejorar los sistemas de filtrado de reglas, estudiando la introducción de nuevas métricas como el factor de potencia de regla (RPF).
- Ampliar el estudio de estructuras de las redes neuronales para mejorar el rendimiento de la red que clasifica para relación entre víctima y criminal.

8. Agradecimientos

En primer lugar, agradecer mis tutoras, la profesora Dra. Angélica González Arrieta, a la Dra. Ana Gil González y a la Dra. Ana de Luis Reboredo por brindarme esta oportunidad de indagar en la investigación y aplicación de aprendizaje automático en una herramienta que, en un futuro, pueda ser de utilidad para las investigaciones policiales. Agradecer también a la Unidad de Análisis de Conducta Criminal de la USAL por todo su apoyo y colaboración al proporcionar la base de datos utilizada en el primer caso, y por su ayuda continuada con todas las cuestiones de índole criminalístico que surgieron a lo largo del desarrollo de este trabajo.

Referencias

- [1] Rakesh Agrawal, Tomasz Imieliński y Arun Swami. “Mining association rules between sets of items in large databases”. En: *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*. 1993, págs. 207-216.
- [2] Jaber Alwidian, Bassam H Hammo y Nadim Obeid. “WCBA: Weighted classification based on association rules algorithm for breast cancer disease”. En: *Applied Soft Computing* 62 (2018), págs. 536-549.
- [3] Scott Bonn. *Criminal Profiling: The Original Mind Hunter*. Dic. de 2017. URL: <https://www.psychologytoday.com/gb/blog/wicked-deeds/201712/criminal-profiling-the-original-mind-hunter>.
- [4] Adolfo Crespo Márquez, Antonio de la Fuente Carmona y Sara Antomaroni. “A Process to Implement an Artificial Neural Network and Association Rules Techniques to Improve Asset Performance and Energy Efficiency”. En: *Energies* 12.18 (2019), pág. 3454.
- [5] Jeremy Elson y col. “Asirra: A CAPTCHA that Exploits Interest-Aligned Manual Image Categorization”. En: *Proceedings of 14th ACM Conference on Computer and Communications Security (CCS)*. Association for Computing Machinery, Inc., oct. de 2007. URL: <https://www.microsoft.com/en-us/research/publication/asirra-a-captcha-that-exploits-interest-aligned-manual-image-categorization/>.
- [6] Richard HR Hahnloser y col. “Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit”. En: *Nature* 405.6789 (2000), págs. 947-951.
- [7] Petr Hájek, Ivan Havel y Michal Chytil. “The GUHA method of automatic hypotheses determination”. En: *Computing* 1.4 (1966), págs. 293-308.
- [8] Javier Herrero y Luis Sánchez-Gil. *Proyecto de investigación de asesinos en serie*. Ene. de 2016. ISBN: 978-84-8196-371-7.
- [9] Ronald M Holmes, James De Burger y Stephen T Holmes. “Inside the mind of the serial murder”. En: *American Journal of Criminal Justice* 13.1 (1988), págs. 1-9.
- [10] Suresh Kumar, Nisheeth Joshi y col. “Rule power factor: a new interest measure in associative classification”. En: *Procedia Computer Science* 93 (2016), págs. 12-18.
- [11] Yann LeCun y col. “Gradient-based learning applied to document recognition”. En: *Proceedings of the IEEE* 86.11 (1998), págs. 2278-2324.
- [12] Titus Livius. *The history of Rome*. 1896.
- [13] Warren S McCulloch y Walter Pitts. “A logical calculus of the ideas immanent in nervous activity”. En: *The bulletin of mathematical biophysics* 5.4 (1943), págs. 115-133.
- [14] RN Pamela Fox Levin MS, RN Jeanne Beauchamp Hewitt MS y RN Susan Terry Misner MS. “Female workplace homicides: An integrative research review”. En: *Workplace Health & Safety* 40.5 (1992), pág. 229.

- [15] C Gabrielle Salfati y David V Canter. "Differentiating stranger murders: Profiling offender characteristics from behavioral styles". En: *Behavioral Sciences & the Law* 17.3 (1999), págs. 391-406.
- [16] Qian Sima. *Records of the grand historian: Han dynasty*. Vol. 65. Columbia University Press, 1993.
- [17] Carsten Stephan y col. "Multicenter evaluation of an artificial neural network to increase the prostate cancer detection rate and reduce unnecessary biopsies". En: *Clinical chemistry* 48.8 (2002), págs. 1279-1287.
- [18] Petri Toiviainen. "Modeling the target-note technique of bebop-style jazz improvisation: An artificial neural network approach". En: *Music Perception: An Interdisciplinary Journal* 12.4 (1995), págs. 399-413.
- [19] Elif Derya Übeyli. "Implementing automated diagnostic systems for breast cancer detection". En: *Expert systems with Applications* 33.4 (2007), págs. 1054-1062.
- [20] Liming Yao y col. "Synergies Between Association Rules and Collaborative Filtering in Recommender System: An Application to Auto Industry". En: *Data Science and Digital Business*. Springer, 2019, págs. 65-80.
- [21] Leonid N Yasnitsky y col. "THE USE OF ARTIFICIAL INTELLIGENCE METHODS IN THE ANALYSIS OF SERIAL KILLERS' PERSONAL CHARACTERISTICS". En: *RUSSIAN JOURNAL OF CRIMINOLOGY* 9.3 (2015), págs. 423-430.

Sistemas de reconocimiento y detección de poses

Francisco García Encinas, María N. Moreno García, and Juan Francisco de Paz Santana

Departamento de Informática y Automática, Universidad de Salamanca.
Plaza de los Caídos s/n. 37008, Salamanca España
{frangaren, mmg, fcofds}@usal.es

Resumen Este artículo se centra en realizar una revisión del estado del arte en el campo de la estimación de poses y en la propuesta de un método de interacción gestual a cuerpo completo. Para ello, partiendo de los errores de otras alternativas anteriores, se desarrolla una propuesta basada en un algoritmo de estimación de poses bidimensional sobre imágenes RGB. Para mejorar la usabilidad del sistema, todos los gestos se crean a partir de una metáfora y se prioriza la sencillez por delante de la utilización de gestos poco intuitivos. La evaluación de los resultados de esta interacción se ha realizado mediante encuestas a usuarios sobre los gestos utilizados y probando con ellos un prototipo funcional de una pizarra que utiliza este sistema interactivo.

Keywords: Interacción gestual, Estimación de poses, Detección de poses.

1. Introducción

En la actualidad, la mayoría de las interfaces persona-máquina se basan en mecanismos de interacción que utilizan el teclado, el ratón o paneles táctiles como entradas. Sin embargo, no son los únicos mecanismos posibles y, posiblemente, tampoco los mejores, al menos, en algunos casos de uso. Por ejemplo, el uso de una interfaz gestual a cuerpo completo, en lugar de una interfaz táctil, en las pantallas utilizadas para realizar pedidos en los restaurantes de comida rápida podría ser una alternativa más higiénica.

Este tipo de interacción se ha visto anteriormente en sistemas de entrenamiento videoconsolas como la Nintendo Wii o la Microsoft Xbox 360 o, incluso, en sistemas médicos como GestSure¹. En el primer caso, este tipo de interacción habilita la introducción de nuevas mecánicas que eran, anteriormente, irrealizables. En el segundo caso, permite que los cirujanos puedan controlar equipos de ayuda sin arriesgarse a interactuar con otras entradas, como ratón y teclado, que puedan no estar estériles. En el futuro, la utilización de este tipo de mecanismos, podría dar lugar a acciones tan naturales como señalar a un dispositivo y pedirle que se encienda. Sin embargo, pese a las promesas de futuro, los métodos expuestos tienen sus limitaciones.

¹<https://www.gestsure.com/>

Pese al sentido de la propiocepción que permite que conozcamos la posición de las partes de nuestro propio cuerpo, incluso cuando no podemos verlas, a veces es complicado establecer la correspondencia entre la posición de la mano en el mundo real y la posición en la pantalla sobre la que se está actuando. Este problema se resuelve en estos sistemas mediante el uso de cursores.

Uno de los problemas más fáciles de ver es la necesidad de periféricos específicos. En el caso de Xbox se hace necesario el uso de Kinect que consta de una cámara y un sensor de profundidad, en el de Wii se utiliza un mando con sensores infrarrojos y en el de GestSure un “conjunto de sensores avanzados”. La utilización de dispositivos específicos hace que una interacción de este tipo no esté al alcance de todo el mundo.

Jacob Nielsen [17] menciona algunos otros problemas de este tipo de sistemas, en concreto, de Kinect [21]. Entre estos destacan los de visibilidad, los de consistencia y estándares, y los de confianza y activación accidental.

Los primeros hacen referencia a que las posibles acciones que se pueden realizar no quedan claras sólo con ver la pantalla. Es necesario introducir previamente una pantalla con instrucciones que explique los mecanismos de interacción. Esto aumenta la carga cognitiva del usuario y hace más incómodo el uso de este tipo de interfaces.

Los segundos se refieren a que, debido a que este tipo de mecanismos de interacción aún no se han popularizado, no existe un estándar, ni de *facto* ni de *jure*, que marque una serie de comportamientos que el usuario pueda esperar. Por ello, cada aplicación funciona de una forma distinta dando lugar a una inconsistencia entre ellas. Además, el usuario tienen que aprender a utilizar cada interfaz individualmente perdiendo eficiencia.

Con respecto al tercero, el sistema de detección de poses de la Xbox 360 no es capaz de detectar gestos en manos. Por ello, el mecanismo utilizado para seleccionar opciones es mantener la mano sobre una región durante un cierto período de tiempo. Esto da lugar a errores e interacciones indeseadas cuando, involuntariamente, se deja la mano en una posición por demasiado tiempo.

Debido a que los sistemas existentes cuentan con diversos problemas, en este trabajo se realizará una propuesta que mejore las alternativas ya existentes. Para detallar esta propuesta, se expondrá un apartado con los métodos y técnicas utilizadas, otro con los resultados obtenidos y, por último, un apartado con las conclusiones y líneas de trabajo futuras.

2. Estado del arte

La mayoría de métodos de estimación de poses del estado del arte utilizan redes neuronales profundas formadas, principalmente, por capas convolucionales. La arquitectura de estos métodos sigue una tendencia distinta en función del conjunto de datos que se tome como referencia.

Los métodos que obtienen mejores resultados sobre *MPII* [2], generalmente, son variantes de la arquitectura *hourglass* [16, 8, 5, 15, 14]. Esta recibe este nombre dado que la resolución de los mapas de características varía siguiendo un

patrón con forma de reloj de arena. En otras palabras, se comienza disminuyendo la resolución de los mapas de características utilizando capas de *pooling* para, después, aumentarla utilizando capas de *upsampling* hasta llegar a generar unos mapas de calor de la posición de cada articulación. Esta red además cuenta con conexiones residuales entre capas con el mismo tamaño. Estas conexiones permiten evitar problemas de explosión y desvanecimiento del gradiente. Este patrón se puede repetir varias veces en cascada para mejorar los resultados.

En el caso de los métodos que destacan en *COCO keypoints* [13] existe una mayor variedad. Una de las tendencias son las propuestas [6, 25] basadas en las redes piramidales de características [12]. Su funcionamiento es similar al de la arquitectura *hourglass* pero en este caso se realizan predicciones a distintas resoluciones y no únicamente al final. Otra de las tendencias, bastante diferente a las otras dos, es la familia *HRNet* [23, 7]. En esta, se apuesta por mantener en todo momento una rama que genera mapas de características de alta resolución. De esta, se ramificaría otra que genere mapas de una menor resolución. Aplicando este mecanismo recursivamente, se puede generar una red que encuentre características a distintas resoluciones. Finalmente, todas las ramas se unen a la rama de la que parten realizando un *upsampling* para dar lugar a un mapa de calor que constituirá el resultado.

También es importante destacar *Mask R-CNN* [11]. Este, es un método de estimación de detección y segmentación de objetos con suficiente flexibilidad como para poder utilizarse para abordar el problema de la estimación de poses. Pese a no estar diseñado específicamente para este problema, obtiene resultados comparables a otros métodos del estado del arte sobre el conjunto de datos *COCO keypoints*.

Los métodos de estimación de poses cada vez son más complejos y se hace difícil comprender cuáles son las características que mejoran realmente los resultados. Para intentar resolver este problema, en [24] se propone un método basado en *ResNet* [10], muy utilizada en otros campos de la visión artificial, que pretende proporcionar una alternativa más sencilla, al menos conceptualmente, con la que comparar otros métodos.

3. Comparativa de métodos

En este apartado se pretenden mostrar los resultados de distintos métodos de estimación de poses bidimensionales del estado del arte sobre los dos conjuntos de datos más importantes: *MPII* [2] y *COCO keypoints* [13]. Los autores de estos dos conjuntos de datos particionan cada uno en tres subconjuntos (entrenamiento, validación y prueba). De esta forma, cualquier trabajo que utilice estos conjuntos de datos se entrenan, validan y prueban sobre los mismos subconjuntos y se pueden comparar sus resultados. Las anotaciones de la partición de prueba no son públicas (los autores del conjunto se encargan de calcular las métricas correspondientes a partir de los resultados obtenidos si se solicita) para garantizar que no se tiene en cuenta durante el entrenamiento. Por ello, es común en el campo realizar comparativas sobre el conjunto de datos de validación,

aunque esta aproximación tenga sus problemas. En concreto, se evaluarán distintas configuraciones de las siguientes propuestas: *HRNet* [23], *HigherHRNet* [7], la arquitectura de referencia propuesta por Xiao *et al.* [24], *PyraNet* [25] y la propuesta de Chu *et al.* [8].

Cada uno de las configuraciones se entrenarán sobre la partición de entrenamiento del conjunto de datos sobre el que se realizará la evaluación. Este entrenamiento se realizará siguiendo las directrices proporcionadas por los autores de cada uno de los métodos en los artículos en los que se proponen. Una vez entrenado el modelo se obtendrán las predicciones de las poses que aparecen en las imágenes de la partición de validación del mismo conjunto de datos sobre el que se entrenó. Por último, teniendo en cuenta estas predicciones y las poses reales, se calcularán una serie de métricas que permitan evaluar la calidad de las predicciones del modelo.

En caso de que se quiera evaluar la misma configuración sobre los dos conjuntos de datos, se realizará el entrenamiento por separado en ambos obteniendo dos modelos distintos. La validación de cada uno de los modelos se realizará únicamente sobre la partición de validación del conjunto de datos sobre el que se ha entrenado.

El coste, tanto temporal como económico, de entrenar cada una de las configuraciones de cada uno de los métodos sobre ambos conjuntos de datos es muy alto. Por ejemplo, han sido necesarios dos días y medio de entrenamiento utilizando una tarjeta gráfica *NVIDIA Tesla P100* para *HRNet-W48* sobre el conjunto de datos *MPII*. Para remediarlo, no todas las configuraciones se evaluarán en ambos conjuntos de datos y, además, en los casos en los que los autores del propio método los proporcionen, se utilizarán modelos preentrenados en lugar de realizar todo el proceso de entrenamiento.

3.1 Comparativa sobre MPII

Para la comparación de los resultados de los distintos algoritmos, se utilizará la métrica estándar para este conjunto de datos [2], PCKh (head-normalized probability of correct keypoint). Esta métrica considerará que una articulación se ha predicho correctamente si la predicción está a una distancia del valor real menor al tamaño de la cabeza multiplicado por un factor α . En concreto, se considera el tamaño de la cabeza como el 60 % de la diagonal de la caja que bordea la cabeza. Para esta comparativa, $\alpha = 0,5$.

Los resultados de esta comparativa se recogen en la tabla 1.

Como se puede observar en la comparativa (tabla 1), todos los métodos obtienen resultados muy similares. Llama especialmente la atención el experimento realizado utilizando *ResNet152* 256×256 con pesos entrenados para este trabajo ya que obtiene unos resultados bastante peores al resto de experimentos. Incluso, con respecto al experimento realizado utilizando los pesos preentrenados por los autores del método, la diferencia es del 1.65 %.

² Resultados obtenidos utilizando un modelo preentrenado por los autores del método.

³ Resultados obtenidos utilizando un modelo entrenado para este trabajo.

Tabla 1: Comparativa de resultados entre métodos del estado del arte sobre el conjunto de datos MPII. Se muestran los valores de la métrica PCKh totales y para cada una de las articulaciones: cabeza (Cab.), hombros (Hom.), codos (Cod.), muñecas (Muñ.), cadera (Cad.), rodillas (Rod.) y tobillos (Tob.).

Método	Cab.	Hom.	Cod.	Muñ.	Cad.	Rod.	Tob.	Total
HRNet-W32 [23] ²	97.10	95.94	90.34	86.45	89.09	87.08	83.28	90.33
HRNet-W48 [23] ²	97.00	95.64	90.23	85.15	88.20	84.89	81.37	89.43
HRNet-W48 [23] ³	97.00	95.64	90.17	85.08	87.80	84.71	81.06	89.29
ResNet50 256 × 256 [24] ²	96.35	95.33	88.99	83.18	88.42	83.96	79.59	88.53
ResNet50 384 × 384 [24] ²	96.66	95.75	89.79	84.61	88.52	84.67	79.29	89.07
ResNet101 256 × 256 [24] ²	96.86	95.87	89.52	84.38	88.44	84.49	80.70	89.13
ResNet101 384 × 384 [24] ²	96.97	95.91	90.27	85.78	89.60	85.94	82.10	90.00
ResNet152 256 × 256 [24] ²	97.03	95.94	90.05	84.98	89.16	85.31	81.27	89.62
ResNet152 256 × 256 [24] ³	96.59	94.99	88.24	82.10	88.21	82.97	78.63	87.97
ResNet152 384 × 384 [24] ²	96.79	95.62	90.08	86.23	89.70	86.86	82.85	90.20
Chu <i>et al.</i> [8] ²	97.22	95.52	89.69	85.04	87.69	84.24	81.01	89.07
PyraNet [25] ²	97.41	96.16	91.10	86.88	90.05	86.00	83.89	90.27

Es curioso como *PyraNet*, pese a obtener los mejores resultados en todas las articulaciones a excepción de en las rodillas, termina obteniendo una métrica PCKh total menor al de *HRNet-W32* (el método que mayor PCKh total obtiene). Se puede afirmar que ambos son los métodos que obtienen mejores resultados.

En el caso de *ResNet* tanto aumentar el número de capas de la red como aumentar la resolución ha mejorado los resultados. Sin embargo, el aumento del número de mapas de características en *HRNet* ha sido perjudicial.

En cuanto a las articulaciones, los tobillos, las rodillas y las muñecas suelen ser las más difíciles de detectar. Por el contrario, la cabeza y los hombros son las más sencillas.

3.2 Comparativa sobre COCO

Al igual que en el conjunto de datos MPII, se utilizarán las métricas estándar recomendadas por los autores del conjunto de datos y que usan todos los artículos que realizan evaluaciones de métodos sobre este conjunto de datos (algo que es bastante común en el campo). De esta forma, se puede garantizar que los resultados de los métodos de esta comparativa sean comparables a los de otros métodos evaluados por otros autores en otros trabajos. Estas métricas son AP (*Average Precision*) y AR (*Average Recall*). En ambas, se utilizará la métrica OKS para calcular la similitud entre las poses predichas y las poses de referencia.

La métrica OKS viene dada por la ecuación 1. En esta ecuación, n es el número de articulaciones, d_i es la distancia entre la posición predicha y la de referencia, s es la escala del objeto calculada como la raíz cuadrada del área que la persona ocupa en la foto y k_i es una constante calculada empíricamente para cada articulación.

$$\text{OKS} = \frac{\sum_i^n \exp(\frac{-d_i^2}{2 \cdot s^2 \cdot k_i})}{n} \quad (1)$$

En concreto, se utilizarán AP@0,50, AP@0,75 y AP@[0,50:0,05:0,95]. Y, de forma similar a AP, AR@0,50, AR@0,75 y AR@[0,50:0,05:0,95].

Los resultados de esta comparativa se recogen en la tabla 2.

Tabla 2: Comparativa de resultados entre métodos del estado del arte sobre el conjunto de datos COCO.

Método	AP@			AR@		
	0.5	0.75	0.5:0.95	0.5	0.75	0.5:0.95
HigherHRNet-W32 512 [7] ⁴	0.86	0.73	0.67	0.89	0.77	0.72
HigherHRNet-W32 640 [7] ⁴	0.87	0.75	0.69	0.89	0.78	0.73
HigherHRNet-W64 640 [7] ⁴	0.87	0.76	0.70	0.89	0.79	0.74
HRNet-W32 256 × 192 [23] ⁴	0.91	0.82	0.74	0.94	0.87	0.80
HRNet-W32 384 × 288 [23] ⁴	0.91	0.83	0.76	0.94	0.87	0.81
HRNet-W48 256 × 192 [23] ⁴	0.91	0.82	0.75	0.94	0.87	0.80
HRNet-W48 256 × 192 [23] ⁵	0.90	0.82	0.74	0.94	0.86	0.80
HRNet-W48 384 × 288 [23] ⁴	0.91	0.83	0.76	0.94	0.87	0.81
ResNet50 256 × 192 [24] ⁴	0.89	0.78	0.70	0.93	0.83	0.76
ResNet50 384 × 288 [24] ⁴	0.89	0.79	0.72	0.93	0.84	0.78
ResNet101 256 × 192 [24] ⁴	0.89	0.79	0.71	0.93	0.84	0.77
ResNet101 384 × 288 [24] ⁴	0.90	0.81	0.74	0.93	0.85	0.79
ResNet152 256 × 192 [24] ⁴	0.89	0.80	0.72	0.93	0.85	0.78
ResNet152 384 × 288 [24] ⁴	0.90	0.81	0.74	0.94	0.86	0.80

En el caso de la comparativa de la tabla 2, a diferencia de la de la tabla 1, se ha realizado una evaluación considerando múltiples personas en lugar de una única persona por imagen.

Se puede observar muy claramente como los algoritmos con detección de múltiples personas de arriba a abajo (*ResNet* y *HRNet*) obtienen mejores resultados que los de abajo a arriba (*HigherHRNet*). Los algoritmos de arriba a abajo utilizan un detector de personas (en este caso uno [24] que obtenía un valor de la métrica AP de 56.4 sobre COCO) para encontrar a las personas y recortarlas para que el algoritmo de estimación de poses pueda procesarlas individualmente. En el caso de los algoritmos de abajo a arriba, el algoritmo de estimación de poses detecta todas las articulaciones y, en un segundo paso, las agrupa en distintas personas. Por ello, estos últimos tienen que ser capaces

⁴ Resultados obtenidos utilizando un modelo preentrenado por los autores del método.

⁵ Resultados obtenidos utilizando un modelo entrenado para este trabajo.

de encontrar articulaciones a distintas escalas y, por esa razón, suelen obtener peores resultados.

En esta comparativa, las diferencias en los resultados entre todos los experimentos sobre *HRNet* y *ResNet* son muy pequeñas y no se pueden apreciar con claridad las tendencias observadas en la comparativa sobre MPII. En este caso el método que obtiene consistentemente mejores resultados es *HRNet-W48* 384×288 pero, la diferencia con respecto a la mayoría de alternativas es prácticamente insignificante (del orden de milésimas).

4. Método

En líneas generales, este método consiste en un mecanismo de interacción que utiliza como entradas la posición de las articulaciones detectadas con un algoritmo de estimación de poses y la pulsación de botones en un mando. A partir de estas entradas, se detectarán una serie de gestos que permitirán realizar distintas acciones sobre una interfaz gráfica.

A continuación, se detallará el sistema profundizando en tres aspectos: cómo se detecta la posición de las articulaciones, qué tipo de gestos se detectan y qué características especiales contiene la interfaz gráfica.

4.1 La posición de las articulaciones

En este apartado se pretende explicar cómo trasladar la posición de las articulaciones en el mundo físico a unas coordenadas en la pantalla sobre las que se realizará la interacción. Por ejemplo, en el caso de utilizar un ratón, movimientos de este sobre un plano horizontal se trasladan por medio de un sensor a movimientos de un cursor en el plano vertical de la pantalla que indica la localización sobre la que se realizará la acción. También, en el caso de los dispositivos táctiles, la interacción se produce sobre la posición de pantalla sobre la que se ha producido el toque detectado por el panel.

En el caso de mecanismos de interacción similares al propuesto, como ya se ha mencionado, se suelen utilizar dispositivos específicos como mandos diseñados para el caso de uso, cámaras RGB-D o, incluso, complejas colecciones de sensores. En esta propuesta, la detección de la posición de las articulaciones requiere únicamente de una cámara web.

El uso de un hardware más sencillo hace que el software necesario para realizar la detección de la posición sea más complejo. En concreto, se hace necesario el uso de un algoritmo de detección de poses.

Existen una ingente cantidad de métodos capaces de estimar una pose bidimensional a partir de una imagen RGB [7, 23, 6, 11]. De hecho, prácticamente la totalidad de ellos consiguen resultados excelentes sobre los distintos conjuntos de referencia en el campo. Pero, la mayoría de ellos cuentan con el mismo problema que limita su campo de aplicación: requieren un gran coste computacional y sus implementaciones únicamente son capaces de obtener resultados en tiempo real en los equipos de más alta gama.

En esta propuesta es necesario un método cuya implementación permita obtener resultados en tiempo real incluso en los dispositivos más modestos [3, 19, 18], como los equipos móviles. En concreto, el método utilizado será una implementación de los trabajos [19, 18] proporcionada por *TensorFlow* [1] y que recibe el nombre de *PoseNet* ⁶.

Las predicciones de *PoseNet* no son consistentes y tienen ruido. Aunque se mantenga una articulación fija en una posición en dos instantes de tiempo, la posición predicha varía ligeramente. Para resolver esto, se hace necesaria la utilización de un filtro que suavice los cambios entre fotogramas continuos. En general, cualquier filtro utilizado en el procesamiento de señales para eliminar ruido puede dar buenos resultados. En este caso, tras considerar otros métodos cómo Savitzky-Golay [20] o una media móvil, se llegó a la conclusión de que el más adecuado es el filtro de 1€ [4], ya que, a diferencia de los anteriores, es capaz de eliminar el ruido en movimientos a bajas velocidades sin añadir un retardo en movimientos a altas velocidades.

En ocasiones, la detección del algoritmo de estimación de poses intercambia las posiciones de las articulaciones a la derecha con las de la izquierda. Y, aunque en un primer momento se consideró la utilización de algún algoritmo de seguimiento de las articulaciones que evitara esto, la idea se acabó desestimando. Los errores de este tipo son muy esporádicos y, normalmente, en la próxima detección se corrigen. Además, con el algoritmo probado, si, pese a contar con el algoritmo, en algún caso las articulaciones se intercambiaran, este mismo algoritmo dificulta que puedan volver a su posición original.

4.2 Los gestos

Cuando se utiliza un ratón, aunque el cursor se mueva por la pantalla, no se produce ninguna acción hasta que se presiona alguno de los botones. Además, en función del botón presionado se produce una acción u otra (hay una diferencia entre el clic izquierdo y el clic derecho). Cuando se usan paneles táctiles, la acción se produce en el momento de tocar sobre una zona de la pantalla. En función de la cantidad de dedos utilizados y el movimiento realizado, se realiza una acción u otra.

En el caso de Xbox 360, para realizar una acción es necesario mantener la mano sobre una posición durante un tiempo para realizar una actuación sobre ella. En ocasiones, dependiendo del programa utilizado, pueden utilizarse poses específicas para realizar ciertas acciones como, por ejemplo, pausar un juego. En la Wii, las acciones se realizan cuando se pulsan ciertos botones en el mando y, dependiendo del botón pulsado, se realizaba una acción u otra. Tanto en la Xbox 360 como en la Wii, fuera de los juegos, se tiene en cuenta únicamente la posición de una de las manos para interactuar con la interfaz.

En este caso, al igual que en la Wii, se utilizan pulsaciones de botones para desencadenar distintas acciones pero, con la diferencia de que en este caso no

⁶<https://github.com/tensorflow/tfjs-models/tree/master/posenet>

se requiere de un mando específico (se puede utilizar cualquier mando similar a los que se utilizan para jugar a videojuegos). Además, a diferencia de los dos métodos anteriores se tienen en cuenta ambas manos. De esta forma, ambas manos podrían estar realizando distintas acciones al mismo tiempo o, incluso, podrían realizar una única acción combinada.

Además, para evitar caer en el problema que tenían los sistemas anteriores como eran los gestos poco intuitivos y la necesidad de explicaciones sobre las acciones disponibles en cada momento, se utiliza una aproximación más simple en la que se usan únicamente dos botones por mano (asociados a las acciones “tocar” y “agarrar”) y un conjunto de gestos mínimo.

En este caso, en el prototipo de pizarra interactiva se utilizan tres gestos distintos: zum, girar y desplazar. Los tres permitirán modificar la posición del lienzo de la pizarra. El primero de ellos se realiza agarrando con las dos manos y separándoles o juntándolas en función de si se desea acercarse o alejarse. El segundo de ellos se realiza agarrando con ambas manos y moviéndolas realizando una trayectoria circular en torno al centro del giro. El tercero se realiza agarrando y moviendo ambas manos la misma distancia en el mismo sentido. Además, utilizando el gesto “interactuar” se realizan acciones que dependen del componente como, por ejemplo, pulsar si se trata de un botón.

4.3 La interfaz

La interfaz del prototipo (ver ilustración 1) se asemeja bastante a otras interfaces que se pueden ver en ordenadores. La mayor diferencia es el uso de botones ligeramente más grandes y que las barras de herramientas aparecen por duplicado a ambos lados de la pantalla para que sea más sencillo interactuar sobre ellas.



Figura 1: Interfaz del prototipo.

Uno de los problemas de este tipo de sistemas es que el usuario tiene dificultades para conocer la posición sobre la que se está interactuando debido a

que los gestos no se realizan sobre la superficie de la pantalla. Para resolverlo, en esta propuesta se utilizan dos cursores, uno por cada mano, que se sitúan sobre la posición de interacción. Estos, a mayores, indican la mano a la que se corresponden mediante la posición del pulgar. En concreto, el pulgar está en la posición en la que lo vería una persona si estuviera observando una vista dorsal de su propia mano. Además, se simboliza la interacción que se está realizando mediante la posición de los dedos: una mano completamente cerrada para agarrar, una mano con el índice estirado para tocar y todos los dedos estirados mientras no se realice ninguna acción.

En la ilustración 2 se puede ver en detalle el cursor utilizado. En la fila superior se encuentran los distintos estados correspondientes a la mano izquierda y, en la inferior, los correspondientes a la mano derecha. En la primera columna se encuentra el cursor utilizado para simbolizar que no se está realizando ninguna interacción. En la segunda columna el cursor que se muestra cuando se está tocando. Y, en la tercera columna, el cursor que se usa cuando se está agarrando.

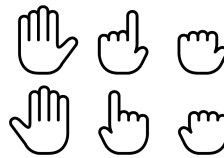


Figura 2: Cursor utilizado para señalar la interacción.

5. Prototipo

Se ha desarrollado un prototipo con dos objetivos. En primer lugar, comprobar si se puede llevar a la práctica con la tecnología actual. Y, en segundo lugar, evaluar la experiencia de usuario del sistema. En esta sección se pretende explicar el funcionamiento y las conclusiones extraídas de este prototipo.

5.1 Funcionamiento técnico del sistema

Este prototipo se ha desarrollado únicamente utilizando tecnologías web. En concreto, se ha utilizado el lenguaje *Typescript* y el *framework React* para el desarrollo de la interfaz. Además, se ha utilizado la biblioteca *Tensorflow.js* y *PoseNet* para la detección de poses. En la ilustración 3 se puede observar un esquema del funcionamiento.

Por un lado, utilizando la *Media Streams API* se recupera la secuencia de vídeo de la cámara web. Esta secuencia, espejada, se le proporciona a PoseNet que dará una predicción de la posición de cada una de las articulaciones. A continuación, utilizando el filtro de 1€ se suaviza la predicción para eliminar el

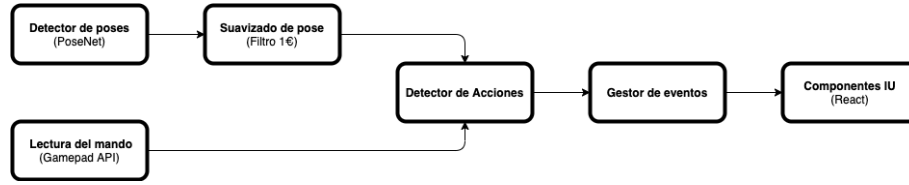


Figura 3: Esquema de funcionamiento del prototipo.

ruido. Y, por otro lado, utilizando la *Gamepad API* se recuperan los botones pulsados en el mando.

Utilizando la pose suavizada y los botones pulsados en el mando se genera una acción. Esto es, por ejemplo, a partir de “la mano izquierda está en la posición P y se está pulsando el botón ZL” se pasa a “la mano izquierda agarra en P ”. Aunque en este prototipo únicamente se utilice la posición de las manos, se detecta la posición de todas las articulaciones porque son útiles para detectar acciones como “levantar la mano” que podrían utilizarse en futuras versiones. Estas acciones se pasan al gestor de eventos que es el responsable de detectar sobre qué componente de la interfaz se está produciendo la interacción y notificarlo con una llamada a un *callback*. Existen eventos para cuando una mano entra, sale o se mueve sobre un elemento, cuando se empieza o se deja de interactuar y para cuando se empieza o deja de agarrar.

Para poder ser el objetivo de una interacción de este sistema, los componentes deben suscribirse a ellas. En el momento de realizar la suscripción, se puede indicar una prioridad de manera que si se hace una acción sobre una posición, el componente que recibirá el evento será el que mayor prioridad tenga en esa posición.

Todo el proceso, desde la detección de las articulaciones y las pulsaciones de botones hasta la generación de los eventos correspondientes, se realiza cada 75 milisegundos (unas 13 veces por segundo).

5.2 Instrucciones de uso del sistema

Conocido el funcionamiento técnico del sistema, con la ayuda de la figura 4, se explicará el funcionamiento a nivel de usuario.

Al mover las manos, sus cursores (**3** y **4** en la figura 4) se mueven en la dirección en la que se movieron estas. En concreto, la posición relativa de una mano sobre la imagen de la cámara web, ignorando unos márgenes ya que en los bordes la predicción es mucho menos precisa, será la posición relativa del cursor sobre la pantalla.

Si se sitúa uno de los cursores sobre alguno de los botones de las barras de herramientas (**1** y **2** en la figura 4) y se presiona el botón asociado a esa mano la acción “interactuar” en el mando, se desencadenará la acción que se corresponda. En el caso de pulsar sobre el primero, se seleccionará la herramienta lápiz para la mano izquierda si se pulsa en la barra de herramientas izquierda o para la

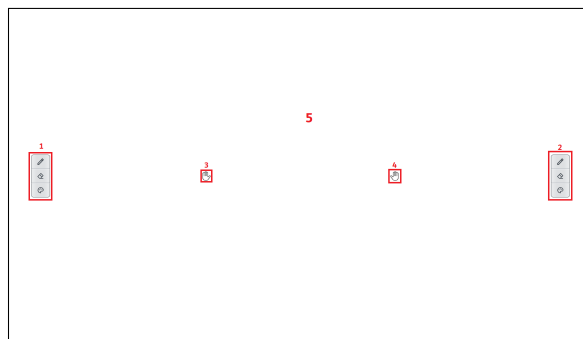


Figura 4: Prototipo del sistema. **1** y **2** son las barras de herramientas para la manos izquierda y derecha, respectivamente. **3** y **4** son los cursores de las manos izquierda y derecha, respectivamente. Y **5** es el lienzo de la pizarra.

derecha si se pulsa en la barra de herramientas derecha. Si se pulsa en el segundo, se activa la herramienta goma. Si se pulsa sobre el tercero, se abre el diálogo de selección de color.

En el caso de pulsar el botón del mando asociado a “interactuar” sobre el lienzo (**5** en la figura 4), se realizará una acción distinta en función de la herramienta seleccionada. Si se está usando la herramienta lápiz, se dibujara una línea por las posiciones por las que pase el cursor mientras el botón este pulsado. Y, si la acción borrar está seleccionada, se borran todas las líneas por las que pase el cursor.

En el caso de pulsar el botón del mando asociado a “agarrar” en ambas manos sobre el lienzo, se realizará una acción de cambio de vista (zum, giro y traslación combinados) siguiendo los principios mencionados en la sección 2.

5.3 Evaluación del prototipo

En primer lugar, se ha evaluado hasta qué punto los gestos son intuitivos y consistentes con esta. Para ello se han realizado entrevistas a cinco usuarios: tres de ellos de entre 18 y 25 años y dos de ellos de entre 45 y 55 años. En esta entrevista se han realizado preguntas orientadas a dos focos. Primero, para evaluar los gestos de una única mano con preguntas como “¿Qué gesto asocia con tocar algo?” o “¿Qué gesto asocia con agarrar algo?”. Tras esto, se evaluaron los gestos de dos manos utilizando preguntas como “Si tuviera en sus manos un folio de un material elástico, ¿cómo haría que el tamaño de su contenido aumentara?”, “Si tuviera en sus manos un folio de un material elástico, ¿cómo lo movería a otra posición?” o “Si tuviera en sus manos un folio de un material elástico, ¿cómo lo giraría?”.

Sorprendentemente, todos los usuarios dieron respuestas muy similares. Todos identificaban “agarrar” con una mano completamente cerrada y “tocar” con

una mano completamente cerrada y el índice estirado. También, todos los usuarios asociaban “aumentar de tamaño” con agarrar el folio en dos posiciones y separar las manos, “desplazar” con agarrar el folio y mover las manos a la posición deseada y “girar” con agarrar el folio y girar las manos en el ángulo adecuado.

La evaluación funcional del prototipo comenzó con una explicación sobre qué botones del mando realizaban el gesto de tocar y agarrar de cada mano. Tras esto, se encomendaron distintas tareas a cada usuario, entre ellas dibujar una línea, hacer zum o desplazar la vista.

En este caso, los usuarios más jóvenes no tuvieron ningún problema en llevar a cabo las tareas descritas. Por el contrario, los usuarios con mayor edad tuvieron dificultades en los primeros instantes por tener que utilizar un mando. Esto puede deberse a que los primeros habían utilizado mandos similares en ocasiones anteriores, mientras que los segundos no.

Al preguntar a los usuarios por mejoras para futuras versiones del prototipo, todos ellos mencionaron el poder realizar los gestos con las manos, sin necesidad del mando. Entre otras sugerencias se encuentran contar con una herramienta para insertar imágenes (propuesta por tres de los usuarios), poder escribir texto (propuesta por dos de los usuarios) o poder cargar y guardar el contenido de la pizarra de alguna forma (propuesta por dos usuarios).

Pese a que el prototipo sea intuitivo para los usuarios, aún cuenta con sus desventajas. Por ejemplo, el sistema de detección de poses no es lo suficientemente preciso como para poder realizar trazos con precisión. Para mejorar esto, existen distintas alternativas: desarrollo de nuevos métodos de estimación de poses en tiempo real que obtengan mejores resultados, investigación en métodos de seguimiento de poses capaces de funcionar en tiempo real o recopilación de un conjunto de datos cuyas imágenes se asemejen más a las imágenes que tendrá que procesar el método de estimación de poses durante un uso normal del sistema.

Además, sería deseable la utilización de algún algoritmo de detección de gestos en manos [22, 9] para sustituir a los mandos. Y, permitir definir al usuario un rango de trabajo para evitar que paredes y muebles puedan molestar al intentar llegar a los bordes de la pizarra.

Sin embargo, pese a que haya cosas que se pueden mejorar, el prototipo es bastante prometedor. No sólo las pruebas con usuarios han demostrado que su uso es intuitivo. Sino que, además, en las pruebas realizadas sobre un ordenador con una tarjeta gráfica NVIDIA GTX 960 la respuesta del sistema es rápida y no se notan demasiados retardos.

6. Conclusiones

El campo de la estimación de poses bidimensionales cuenta con una enorme cantidad de métodos que obtienen muy buenos resultados. Sin embargo, la mayoría de ellos cuentan con un enorme coste computacional. Propuestas con métodos más ligeros cuyas implementaciones pudieran ejecutarse en tiempo real, incluso, en dispositivos con prestaciones más modestas abrirían muchas puertas en la interacción persona-máquina, el entretenimiento o la realidad virtual. Investigar

en arquitecturas para la detección de poses basadas en el algoritmo de detección de objetos YOLOv3 podría ser una buena aproximación a la consecución de una propuesta de estas características.

En la comparativa sobre MPII, la mayoría de los métodos obtienen resultados muy similares, de en torno al 90 % PCKh y una diferencia menor al 2 % entre el mejor y el peor. Es decir, la mayoría de los métodos obtienen muy buenos resultados y a penas hay diferencias entre ellos. En el caso de la comparativa sobre COCO, los resultados son bastante más dispares entre unos métodos y otros, aunque, no son directamente comparables a MPII dado que se utilizan métricas distintas. La diferencia en los resultados en los experimentos con un mismo método cuando se utilizan los pesos preentrenados y los pesos entrenados para este trabajo no es significativa y puede deberse a diversas razones, entre ellas aspectos como la inicialización de los pesos a la hora del entrenamiento pueden afectar a si se converge a un mínimo u otro.

Con la propuesta se ha conseguido desarrollar un método de interacción bastante intuitivo y natural que puede ser de gran utilidad en ciertos casos de uso. También, ha permitido definir qué aspectos de los algoritmos de estimación de poses actuales es necesario mejorar para un mejor funcionamiento de este tipo de métodos de interacción. Por ejemplo, algoritmos de detección de poses con mayor precisión y un menor tiempo de inferencia conseguirían que la interacción fuera más agradable para el usuario. Esto no es únicamente posible mediante el desarrollo de nuevos algoritmos. La creación de conjuntos de datos más representativos de este caso de uso o el uso de algoritmos de seguimiento de poses más rápidos y precisos también ayudarían.

Por otro lado, esta propuesta tiene ciertos aspectos que se podrían mejorar en futuros trabajos. El principal es la realización de una evaluación de los resultados más exhaustiva y con un mayor número de usuarios. También, sería útil investigar en el campo de la detección de gestos en manos para poder eliminar la necesidad del uso de mandos y permitir optar por una alternativa más natural. Además, se debería garantizar que todas las acciones pueden realizarse con una única mano para mejorar la accesibilidad.

7. Agradecimientos

Me gustaría agradecer a la Universidad de Salamanca y a la Fundación Salamanca Ciudad de Cultura y Saberes por la beca que se me concedió mientras estudiaba el Máster en Sistemas Inteligentes. También, a los profesores del Grado en Ingeniería Informática y el Máster en Sistemas Inteligentes de la Universidad de Salamanca que me han ayudado a adquirir los conocimientos para llegar hasta aquí, especialmente, a mis tutores, Juan Francisco de Paz Santana y María N. Moreno García, por ayudarme a hacer este trabajo posible. Finalmente, a mis familiares, amigos y compañeros que me han animado cuando más lo he necesitado.

Referencias

- [1] Martin Abadi y col. “TensorFlow: A system for large-scale machine learning”. En: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. 2016, págs. 265-283. URL: <https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf>.
- [2] Mykhaylo Andriluka y col. “2D Human Pose Estimation: New Benchmark and State of the Art Analysis”. En: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Jun. de 2014.
- [3] Z. Cao y col. “Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields”. En: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, págs. 1302-1310.
- [4] Géry Casiez, Nicolas Roussel y Daniel Vogel. “1 €Filter: A Simple Speed-Based Low-Pass Filter for Noisy Input in Interactive Systems”. En: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. Austin, Texas, USA: Association for Computing Machinery, 2012, págs. 2527-2530. ISBN: 9781450310154. DOI: 10.1145/2207676.2208639. URL: <https://doi.org/10.1145/2207676.2208639>.
- [5] Y. Chen y col. “Adversarial PoseNet: A Structure-Aware Convolutional Network for Human Pose Estimation”. En: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, págs. 1221-1230.
- [6] Y. Chen y col. “Cascaded Pyramid Network for Multi-person Pose Estimation”. En: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, págs. 7103-7112.
- [7] Bowen Cheng y col. “HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation”. En: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020.
- [8] Xiao Chu y col. “Multi-context Attention for Human Pose Estimation”. En: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Jul. de 2017, págs. 5669-5678. DOI: 10.1109/CVPR.2017.601.
- [9] X. Fang y X. Lei. “Hand pose estimation on hybrid CNN-AE model”. En: *2017 IEEE International Conference on Information and Automation (ICIA)*. 2017, págs. 1018-1022.
- [10] K. He y col. “Deep Residual Learning for Image Recognition”. En: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, págs. 770-778.
- [11] K. He y col. “Mask R-CNN”. En: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, págs. 2980-2988.
- [12] T. Lin y col. “Feature Pyramid Networks for Object Detection”. En: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, págs. 936-944.
- [13] Tsung-Yi Lin y col. “Microsoft COCO: Common Objects in Context”. En: *Computer Vision – ECCV 2014*. Ed. por David Fleet y col. Cham: Springer International Publishing, 2014, págs. 740-755. ISBN: 978-3-319-10602-1.

- [14] Diogo C. Luvizon, Hedi Tabia y David Picard. “Human pose regression by combining indirect part detection and contextual information”. En: *Computers & Graphics* 85 (2019), págs. 15-22. ISSN: 0097-8493. DOI: <https://doi.org/10.1016/j.cag.2019.09.002>. URL: <http://www.sciencedirect.com/science/article/pii/S0097849319301475>.
- [15] Alejandro Newell, Zhiao Huang y Jia Deng. “Associative embedding: End-to-end learning for joint detection and grouping”. English (US). En: *Advances in Neural Information Processing Systems* 2017-December (ene. de 2017). 31st Annual Conference on Neural Information Processing Systems, NIPS 2017 ; Conference date: 04-12-2017 Through 09-12-2017, págs. 2278-2288. ISSN: 1049-5258.
- [16] Alejandro Newell, Kaiyu Yang y Jia Deng. “Stacked Hourglass Networks for Human Pose Estimation”. En: *Computer Vision – ECCV 2016*. Ed. por Bastian Leibe y col. Cham: Springer International Publishing, 2016, págs. 483-499. ISBN: 978-3-319-46484-8.
- [17] Jakob Nielsen. *Kinect Gestural UI: First Impressions*. <https://www.nngroup.com/articles/kinect-gestural-ui-first-impressions/>. 2010.
- [18] George Papandreou y col. “PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model”. En: *Computer Vision – ECCV 2018*. Ed. por Vittorio Ferrari y col. Cham: Springer International Publishing, 2018, págs. 282-299. ISBN: 978-3-030-01264-9.
- [19] George Papandreou y col. “Towards Accurate Multi-person Pose Estimation in the Wild”. En: 2017. URL: <https://arxiv.org/abs/1701.01779>.
- [20] Abraham. Savitzky y M. J. E. Golay. “Smoothing and Differentiation of Data by Simplified Least Squares Procedures.” En: *Analytical Chemistry* 36.8 (jul. de 1964), págs. 1627-1639. DOI: 10.1021/ac60214a047. URL: <https://doi.org/10.1021/ac60214a047>.
- [21] Jamie Shotton y col. “Real-Time Human Pose Recognition in Parts from Single Depth Images”. En: *Communications of the ACM*. Vol. 56. Jun. de 2011, págs. 1297-1304. DOI: 10.1109/CVPR.2011.5995316.
- [22] Y. Son y O. Choi. “Image-based hand pose classification using faster R-CNN”. En: *2017 17th International Conference on Control, Automation and Systems (ICCAS)*. 2017, págs. 1569-1573.
- [23] Ke Sun y col. “Deep High-Resolution Representation Learning for Human Pose Estimation”. En: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Jun. de 2019.
- [24] Bin Xiao, Haiping Wu y Yichen Wei. “Simple Baselines for Human Pose Estimation and Tracking”. En: *Computer Vision – ECCV 2018*. Ed. por Vittorio Ferrari y col. Cham: Springer International Publishing, 2018, págs. 472-487. ISBN: 978-3-030-01231-1.
- [25] Wei Yang y col. “Learning Feature Pyramids for Human Pose Estimation”. En: *ICCV*. Oct. de 2017, págs. 1290-1299. DOI: 10.1109/ICCV.2017.144.

Estudio sobre viabilidad de interacción gestual en el tratamiento de mapas en entornos móviles

David Peral García, José Rafael García-Bermejo Giner

Departamento de Informática y Automática, Facultad de Ciencias Plaza de los
Caídos s/n, 37008, Salamanca, España
{daveral,coti}@usal.es
<http://diaweb.usal.es/>

Resumen La interacción gestual junto con el reconocimiento y síntesis de voz son unas de las áreas más estudiadas y con mayor proyección en la actualidad, en especial, debido a la proliferación y mejora de los dispositivos móviles. Este artículo comienza analizando el estado del arte de ambos campos de conocimiento, mediante una revisión sistemática de literatura obtenemos los artículos relevantes y acto seguido, analizamos y extraemos las características más relevantes de los mismos, permitiéndonos componer así un estado del arte de dichos campos. Se implementa una aplicación para el sistema operativo iOS, la cual nos permite hacer uso de bibliotecas tanto de interacción gestual como de reconocimiento y síntesis de voz. Centrando así el foco de la implementación en el manejo y adaptación para la innovación. A través de esta aplicación utilizamos la tecnología de interacción gestual para permitir al usuario poder realizar las diferentes acciones que la aplicación permite, pero utilizando gestos sencillos y accesibles. De igual modo, se da la posibilidad de revertir acciones fácilmente en los casos que la carga cognitiva sea mayor. En suma, se incorpora tecnología de reconocimiento y síntesis de voz, la cual es aplicada para la creación de rutas. El reconocimiento de voz no se encuentra activo mientras la aplicación se encuentre activa, mejorando así la privacidad del usuario pero haciendo esta funcionalidad accesible en instantes de tiempo en los que no se pueda manipular el dispositivo.

1. Introducción

La interacción gestual es una de las áreas que han tenido un gran auge debido a la proliferación y extensión de los dispositivos móviles en la mayoría de los ámbitos de nuestra vida. Esto explica la mayor necesidad de innovación en este terreno.

Mediante la interacción gestual se trata de reflejar una determinada acción para transmitir un mensaje que será interpretado posteriormente. Los diferentes tipos de movimientos son ampliamente utilizados en diferentes ámbitos, en este artículo serán utilizados como método de control de una aplicación de un dispositivo móvil.

Muchas de las aplicaciones que utilizan interacción gestual demandan un usuario experto, que entienda y sepa utilizar los diferentes gestos, en nuestro *framework* trataremos de mejorar la experiencia del usuario a través del uso de gestos sencillos, complementarios y que no requieran de un usuario experto para su uso, expandiendo así el número de usuarios capaces de utilizarla de forma correcta.

Otro de los ámbitos tecnológicos de los cuales se ha incrementado su uso exponencialmente es el reconocimiento y síntesis de voz. Este será un aspecto estudiado e implementado, el cual, combinado junto a la interacción gestual, dotará a la aplicación de una gran facilidad de uso, sin importar la situación externa del usuario.

El paso previo a la implementación de nuestra aplicación es la recopilación de artículos relacionados tanto con la interacción gestual como con la síntesis y el reconocimiento de voz, con el objetivo de conocer las principales metodologías y técnicas utilizadas en los mismos.

El resultado es el desarrollo de una aplicación de mapas para los sistemas operativos *iOS* y *iPadOS*, con funcionalidades tanto del ámbito de la interacción gestual como del *text-to-speech* y *speech-to-text*, la cual es completamente funcional y accesible para la mayoría de los usuarios.

Algunos de los gestos implementados son: *swipe left* y *swipe right*, utilizados para el manejo de los menús y controles de la aplicación; *tap* y *double tap* para la gestión de los marcadores del mapa o *pinch and zoom* para controlar la altitud a la que se visualiza el mapa desde el dispositivo.

La creación de rutas se realiza mediante el reconocimiento de voz, el cual se activa con un gesto de *double tap*, con esto conseguimos que el reconocimiento de voz no se encuentre en escucha mientras la aplicación está activa, preservando la privacidad del usuario a la vez que se hace accesible en situaciones que no se pueda prestar mucha atención al dispositivo, como en la conducción.

2. Objetivos

Mediante este trabajo se tiene como objetivo principal el diseño de una aplicación relativa al tratamiento de mapas y a la gestión de marcadores y puntos de interés. Se intenta conseguir una mejor ergonomía al reducir los movimientos

de interacción gestual así como incorporar reconocimiento y síntesis de voz para mejorar la experiencia de usuario. Los principales objetivos son:

- Realizar una revisión sistemática de los diferentes métodos y técnicas de interacción gestual.
- Realizar un estudio de las diferentes tecnologías y dominios de aplicación en el ámbito del reconocimiento y la síntesis de voz.
- Seleccionar la tecnología mediante la cual se desarrollará nuestro proyecto.
- Utilizar gestos accesibles y sencillos para el usuario.
- Implementar herramientas de reconocimiento y síntesis de voz.
- Crear una herramienta capaz de trazar rutas entre diferentes puntos de un mapa.
- Permitir la adición, traslación y borrado de marcadores en un mapa.
- Implementar y aplicar la propuesta citada.

3. Estado del Arte

3.1 Interacción Gestual

Con el desarrollo de la tecnología, en los últimos años se ha intensificado el uso de la interacción gestual en el ámbito de las manos y la cara, estos ámbitos de aplicación ocupan un lugar muy importante dentro de la interacción persona-ordenador. A través de ella, se permite a los computadores capturar e interpretar los movimientos gestuales y ejecutar una acción posteriormente.

La representación de los gestos sirve tanto para reflejar una acción determinada como para transmitir un mensaje que será interpretado posteriormente. Los diferentes movimientos son ampliamente utilizados en diferentes ámbitos, como facilitar la interpretación a personas con movilidad reducida o trastornos psicológicos, aplicaciones de ocio [20] o control industrial [32] [33]. Por otra parte, muchas de las aplicaciones que utilizan interacción gestual demandan un usuario experto, que entienda y sepa utilizar los diferentes gestos, de lo contrario, la experiencia del usuario puede verse afectada a la hora de realizar algunas tareas [65].

Observamos aplicaciones que permiten diferenciar varios gestos con una mano, como [45], [1] que identifican el número de dedos estirados, así como el pulgar y el puño. Existen aplicaciones que se centran en los gestos dinámicos, como son *swipe left/right/up/down*, *tap* y *pinch* [48],[11]. Incluso se identifica un *double swipe* con ambos brazos [46] [fig:1]. Otros estudios como [19], [28],[29] diferencian todos los gestos del lenguaje de gestos, siendo estos mucho más completos.

Para el proceso de clasificación de gestos se utilizan diferentes tecnologías, como pueden ser redes neuronales artificiales (ANN) en [7], [64],[54] y [54]; redes neuronales convolucionales (CNN) en [1], [45], [61], [19],[9] y, en [66] se utiliza a su vez dos redes convolucionales 3D interconectadas (3D-CNN); redes neuronales progresivas (PNN) [9]; redes neuronales recurrentes (RNN) [3] y tipos derivados de la misma, como redes neuronales recurrentes cerradas (GRU-RNN) [53] y redes de memoria corta-larga (LTSM) [50].

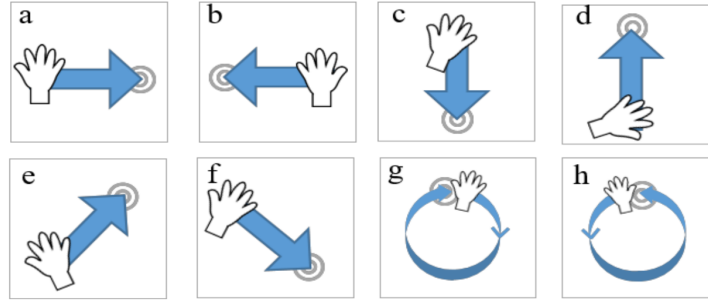


Figura 1: Gestos manuales dinámicos[1]

También es muy utilizado el modelo oculto de Markov (HMM) [14], [44], [28], [62], el modelo Bayesiano (IBCC) [28], el algoritmo de k vecinos más próximos (KNN) [54], [20], [54], [58] y [36]; las máquinas de soporte vectorial (SVM) [45], [48], [54], [66], [62], [54], [58]; el algoritmo Adaboost en [21],[31] y [62]; el algoritmo SDD [34]; Random Forest (RF) [58] y las distribuciones Gaussianas [62]. Por último se destaca el uso de sistemas expertos [63], el algoritmo de expectativa maximizada (EM) [25], [24] y el algoritmo de análisis discriminante (LDA) que permite encontrar una combinación lineal de rasgos que caracterizan o separan dos o más clases de objetos [36].

En cuanto al porcentaje de éxito de los estudios, encontramos que todos ellos se encuentran por encima de un 85 % de éxito en el reconocimiento de gestos, lo que supone un porcentaje muy alto de acierto. Entre los mismos, un 28 % de los artículos se encuentran en torno a un 85-90 % de acierto, un 32 % entre 90-95 % y un 40 % entre 95-100 %.

3.2 Tecnologías del habla

La tecnología de síntesis de voz transforma texto a voz, la cual es usada principalmente como tecnología de asistencia al usuario, donde se utiliza una voz interactiva como forma de recepción de información. La mejora de los dispositivos portátiles en términos del rendimiento y costo computacional permiten al usuario que utiliza la tecnología de síntesis de voz, por ejemplo, la navegación de las rutas en Google Maps [6], asistentes virtuales, como Siri o Amazon Alexa [10], o e-readers como Amazon Kindle o Sony Reader que utilizan la síntesis de voz para la lectura de cómics, los cuales diferencian entre las voces de los personajes y su estado de ánimo [60].

Entre las técnicas y modelos de síntesis mas utilizados encontramos las redes neuronales [22] [Zhu2019645] [59] [35] [35], el modelo oculto de Markov [6] [27] [43] [49] [47] [2], los vocoder [2] y las WaveNet [26].

Los dominios de aplicación en los que se utilizan los sistemas *text-to-speech* son: las aplicaciones de propósito general basadas en la generación de voz, voz cantada [17], síntesis del discurso expresivo [Zhu2019645] [56], detección de

la personalidad y del estado de ánimo [4] e, incluso percepción del discurso silencioso [5].

Tabla 1: Porcentaje de acierto de los artículos recopilados interacción gestual

Artículo	Porcentaje de éxito
[58]	85 %
[31]	86-92 % ¹
[65] ²	86,20 %
[44]	86,67 %
[20]	87,3 %
[42]	88,5 %
[53]	88 %
[14]	90-100 % ¹
[23]	90 %
[29]	90,80 %
[54]	92,23 %
[66]	90-95 % ¹
[7]	92,86 %
[38]	92,5 %
[9]	94-98,31 % ¹
[50]	95 %
[65] ³	95.28 %
[3]	96 %
[62]	96 %
[28]	96.05 %
[33]	96.5 % - 97.3 % ¹
[41]	97.2 %
[46]	97,98 %
[18]	98.6 %
[34]	99.2 %

La escala MOS es un sistema de medida numérica que permite evaluar la calidad media de un discurso de voz, utilizada en evaluación directa o singular. Fue utilizada en primer lugar para medir la calidad de las llamadas telefónicas en el ámbito de la telecomunicación, extendiéndose posteriormente a otros ámbitos, como el de la síntesis de voz. Se evalúa en una escala del 1 (malo) al 5 (excelente).

Los factores que afectan a la misma son: la latencia, es decir, tiempo de generación y de emisión de la síntesis de voz; el *jitter*, que es la variación de latencia entre muestreos, contra mayor sea la variación, más alto el *jitter* y peor

¹Si se indica rango de porcentaje se toma el limite inferior como indicador

²Usuario inexperto

³Usuario experto

la calidad del audio. En lo referente a los *CODECs* utilizados existen *CODECs* con y sin compresión. Para evaluaciones comparativas, el rango de calificación también oscila entre uno y cinco, pero se define una escala equivalente conocida como DMOS (Degradation MOS).

Observamos que un 13 % de los *frameworks* se encuentran entre 1 y 3 de la escala MOS/DMOS, lo que implica que se requiere un esfuerzo considerado o resulta imposible de entender la propia síntesis. La mayoría de los artículos, un 65 %, se escala entre 3 y 4, lo que significa que se requiere de un esfuerzo moderado a la hora de la escucha. Por último, un 22 % de los mismos solo necesita de un esfuerzo aceptable para tratar de escuchar el discurso de voz.

Tabla 2: Porcentaje de acierto de los artículos recopilados tecnologías del habla

Artículo	Porcentaje de éxito en escala MOS
[51]	1.97
[51]	2.59
[26]	2.98
[43]	3.01
[43]	3.04
[6]	3.14
[49]	3,15
[8]	3,18
[51]	3.32
[43]	3.36
[47]	3.43
[51]	3.46
[47]	3.58
[6]	3.71
[51]	3.83
[43]	3.91
[30]	4.07
[8]	4,23
[30]	4.40
[30]	4.48
[52]	3.35 ⁴
[52]	3.32 ⁴
[2]	4.19 ⁴

⁴Escala DMOS

4. Arquitectura propuesta

4.1 Introducción

El propósito de este trabajo radica en crear una aplicación relativa al tratamiento de mapas junto a la gestión de marcadores y puntos de interés. Se utilizará MapKit [13], [12], [57] como biblioteca para el gestor de mapas y Firebase [16], [55] como gestor de base de datos. La principal interacción con el usuario se realiza a través de un menú lateral [39], [40], [15], [37] y mediante gestos en el propio mapa. En este apartado se explicará el funcionamiento general de la aplicación desde la perspectiva del programador.

4.2 Aspectos relevantes de programación

Se comentan los aspectos relevantes de programación implementados en el sistema, prestando atención al código que permite utilizar la interacción gestual y le síntesis y el reconocimiento de voz. Se detalla por archivos y funciones que engloban funcionalidades completas, más concretamente las cabeceras de las funciones y, en algún caso oportuno, fragmentos de código.

4.2.1 ViewController.swift Este es el archivo principal y en el que se encuentran la mayoría de detectores gestuales. Mediante un Pan Gesture Recognizer permitimos desplazar región del mapa visible, modificando la latitud y longitud de la región que puede visualizar el usuario. Se aplica una traslación al mapa según el movimiento del dedo y se reescala el polígono que forman los puntos en caso de estar dibujado.

```
@IBAction func panAction(_ sender:
UIPanGestureRecognizer) {
    ...
    mapa.centerCoordinate = CLLocationCoordinate2D
        .init(latitude:newLat, longitude:newLon)
    menuController.updatePolygonFromGesture()
}
```

Con un RotationGesture, permite al usuario rotar el mapa más allá de 360° grados a derecha o izquierda. La limitación existe cuando el zoom es lo suficientemente pequeño no permite la rotación, esta viene dada por el propio mapa, dado que el la figura del mapa es rectangular y aparecerían bordes negros en los extremos. El valor de sender.rotation se resetea en cada ejecución de la función para prevenir la acumulación del mismo.

A través de los gestos SwipeLeft y SwipeRight se permite al usuario mostrar u ocultar el menú lateral, adaptándose al mismo si este se encuentra en la parte izquierda o derecha del mapa respectivamente. Si el menú está oculto en la parte izquierda y detecta SwipeRight, lo muestra, y si detecta SwipeLeft, lo oculta, y viceversa en caso de encontrarse en la parte derecha.

El gesto `PinchandZoom` es detectado con el objetivo de hacer zoom $\pm$ sobre el mapa. El valor de `sender.scale` se resetea en cada ejecución de la función.

Se han implementado tres `TapGestures`, el primero sirve para, una vez se ha seleccionado un mapa y se hace un doble tap en un punto del mapa, se sitúa un punto sobre el mismo. El segundo `TapGesture` permite al usuario, al hacer click sobre uno de los puntos previamente situados con la función anterior, eliminarlo del mapa. El tercero se activa mediante la pulsación con dos dedos dos veces seguidas. Mediante el mismo se activa la función de *text-to-speech* para la creación de rutas.

En cuanto al `LongPressGesture`, es utilizado para permitir al usuario mover uno de los puntos situados a otra zona del mapa. En cuanto el usuario selecciona el punto se destaca y se comienza a actualizar la información de su situación si el usuario lo desplaza, guardando la misma cuando este suelta el punto.

4.2.2 NavBarController Es un controlador en el que están recogidas todas las *constraints* necesarias para la ubicación y reubicación de los elementos en caso de rotación del dispositivo, aparición del menú lateral y cambio de lado del menú así como rotación del dispositivo.

4.2.3 MyMenuController Es el archivo controlador de las funciones del menú lateral, tanto del menú principal como del menú auxiliar de cada mapa. Para ambos menús se ha utilizado un `UITableView`. Los valores del menú principal son obtenidos de forma dinámica ya que, en parte, está compuesto por los diferentes mapas que el usuario ha creado y estos pueden cambiar a gusto del mismo. Cada elemento del menú principal pertenece a la clase `PointsMenuOption`, recogida en `PointsMenuOption.swift`, mientras que los elementos del menú auxiliar son estáticos y están recogidos en el archivo `MenuOption.swift`.

Mediante la función `putPoints()` se obtienen los datos a mostrar y `translatePoints()`, se ubican en el mapa, una vez están fijados se reescala el mapa para mostrar mejor dichos puntos. La función `updatePolygon()` dibuja el polígono que crean los puntos, si estos se encuentran mostrados.

Para la obtención de los datos remotos que se encuentran en la base de datos, se utilizan las dos funciones siguientes, en `necessaryFirebaseUpdate()` se detecta mediante un *timestamp* de la base de datos si este coincide con el *timestamp* local, para actualizar o no los valores. En caso de ser necesaria una actualización, se llama a la función `obtenerMarcadores()`. El atributo `'completed'` es un *closure* que impide que las funciones posteriores se ejecuten hasta que esta no ha acabado, permitiendo al programador asegurarse de que los datos han sido descargados y almacenados antes de realizar ninguna acción con ellos.

```
typealias FinishedDownload = () -> ()
func necessaryFirebaseUpdate(
    completed: @escaping FinishedDownload)
```

```

func necessaryFirebaseUpdate(
  completed: @escaping FinishedDownload) {
  ref.child("timeStamp").observeSingleEvent(
    of: .value, with: { (snapshot) in
      let values = snapshot.value as? NSDictionary
      let timeStampFirebase = values?.value(forKey: "time")
      as? Int64
      ...
    }
  )
}

```

Para manejar la adición de puntos implementamos la función `addPointAndUpdateFirebase()`, así como `deletePoint()` para el borrado de los mismos por separado y `updatePointAndUpdateFirebase()` para su traslación. Con `deleteMap()` crea una alerta para confirmar la acción. Si la respuesta es afirmativa, borra el mapa de la base de datos, actualiza el *timestamp* y carga el menú principal. La función `deleteAllPoints()` sigue un funcionamiento similar a `deleteMap()`, pero la diferencia radica en que borra todos los puntos del mapa y no vuelve al menú principal.

Mediante `createListOfPoints()` permitimos crear un mapa nuevo, para ello instancia un objeto de la clase `AlertView` en el que se genera un cuadro de diálogo para introducir el nombre del mapa. Es comprobado que no exista un nombre ya existente, ni uno vacío, tampoco se permite la un nombre con más de 8 caracteres.

4.2.4 SpeechToText.swift En `SpeechToText.swift` se incorporan la funciones que permiten tanto capturar, como ejercer de controlador de las funciones *speech-to-text*. La función `recordingController()` se encarga de obtener la mejor transcripción posible así como de manejar la propia transcripción y transformarla en la ruta en cuestión.

4.3 Consecución de objetivos

Se detallará la utilidad de nuestro programa:

En visualizamos la pantalla principal, en la parte en la que se encuentra el mapa podemos desplazar el mapa a la posición que deseemos mediante un *pan gesture*. A través del mismo mapa también se permite hacer zoom mediante *pinch and zoom* y una rotación, esto último solo cuando existe un zoom suficiente para poder realizar la rotación del mapa. Para mostrar el menú lateral se puede hacer mediante el botón de la esquina superior izquierda o mediante el gesto *swipe right* (en caso de encontrarse el menú en la parte izquierda, el cual esta así por defecto).

Una vez se encuentra mostrado [2] podemos ocultarlo mediante el gesto *swipe left* o mediante el mismo botón que lo mostramos.

Los elementos del menú principal son los siguientes: 1º Item: Bar Side. Cambio de lado del menú lateral, se actualizan todas las restricciones para su correcta visualización. 2º Item: Map Type. Cambiar tipo de mapa, mediante la pulsación

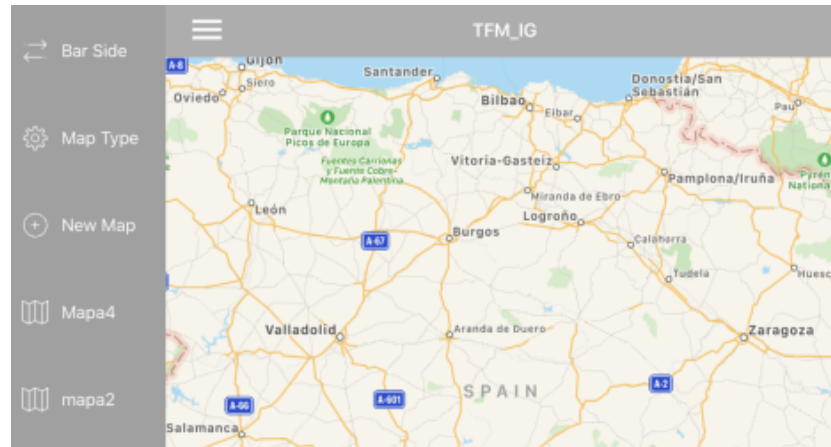


Figura 2: Menú Lateral Principal

del mismo seguiremos este cambio en los tipos de mapa político - físico - híbrido - político. 3º Item: New Map. Crear Mapa: permite crear un mapa nuevo, se genera un cuadro de diálogo para introducir el nombre del mapa. Es comprobado que no exista un nombre ya existente, ni uno vacío, tampoco se permite la un nombre con más de 8 caracteres. 4º Item y sucesivos. Mapas creados por el usuario, se muestran dos creados a modo de ejemplo. Se accede al menú auxiliar de mapas detallado a continuación.

En este apartado [3], una vez se ha seleccionado el mapa que se quiere editar, el mapa permite la adición de puntos mediante un *double tap* en cualquier punto del mapa, el movimiento de los mismos mediante un *longpress* desplazando y, por último, la eliminación de los mismos mediante un *tap* en el punto en el que se encuentre el elemento.

La funcionalidad de *speech-to-text* para la creación de rutas se activa con el gesto *double tap* dos veces seguidas. Al activarse se generará un audio de la forma *Hi, -nombre de usuario-*. Mientras se encuentre activa la escucha se superpondrá un filtro de color naranja en el mapa.

Se implementan tres funcionalidades posibles, la primera, la cual permite generar la ruta y permite al usuario utilizarla de la forma *Go + From + -nombre de la ciudad de origen- + To + -nombre de la ciudad de destino-[fig:4b]*. La segunda, mediante la palabra *Stop* provoca que se dehabilite tanto la función de escucha como la de *speech-to-text*. Por último, a través de la palabra *Clear*, se reinicia la función de *speech-to-text*, eliminando así las palabras reconocidas por el dispositivo.

5. Resultados

El propósito principal de esta aplicación radica en permitir al usuario la realización de las diferentes funcionalidades en el ámbito de la interacción ges-

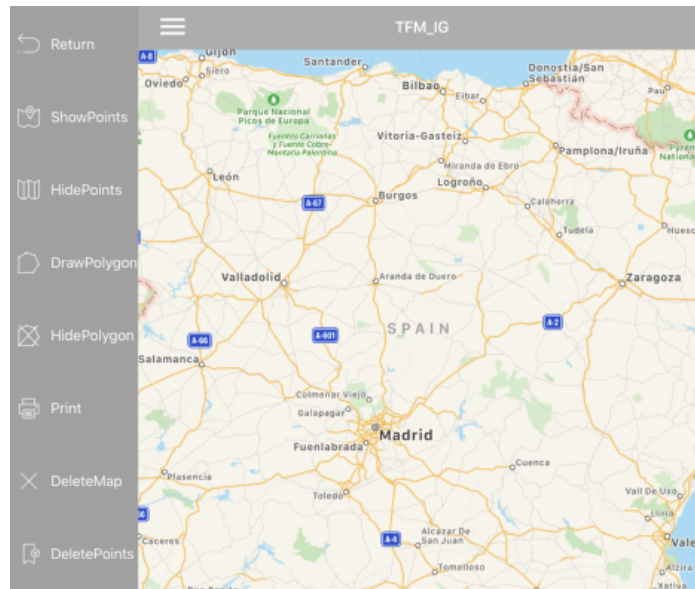
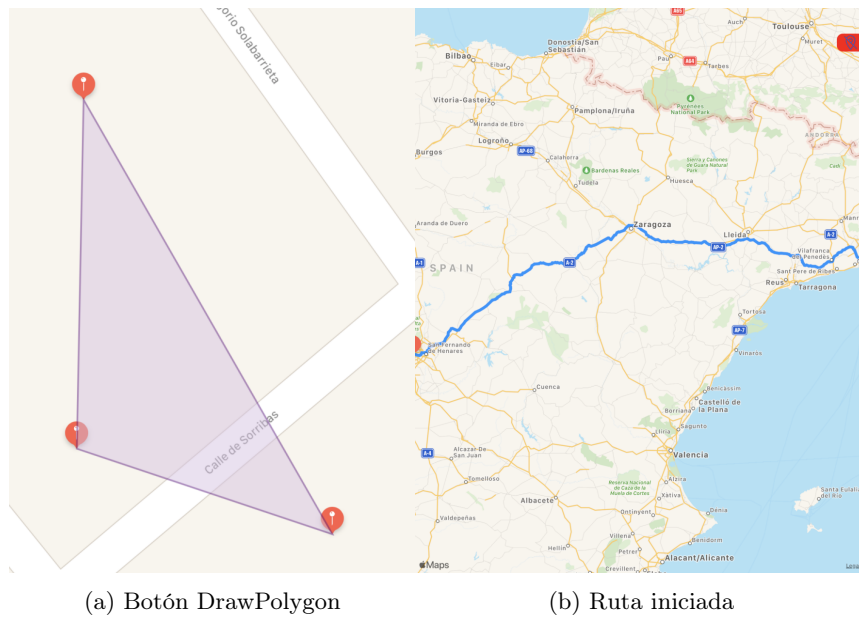


Figura 3: Menú Lateral Auxiliar



(a) Botón DrawPolygon

(b) Ruta iniciada

Figura 4: Funcionalidad de ruta

tual. Aunque la carga cognitiva se incrementa al existir un mayor número de gestos que memorizar y aprender, se trata de solventar ofreciendo al usuario la posibilidad de revertir la acción, ya sea con el mismo gesto o con el inverso.

Por ejemplo, la aparición y desaparición del menú lateral se realiza mediante los gestos *swipe left* y *swipe right* respectivamente, así como mediante el botón que lo expande o contrae. El cambio de tipo de mapa puede realizarse en el menú lateral o mediante un *swipe up/down* de forma cíclica, permitiendo llegar al mapa en el que se encontraba a través el gesto inverso o realizando el mismo gesto dos veces. Se incorpora también la posibilidad de revertir el zoom una vez se ha realizado esta acción, en caso de que el usuario haga zoom por error. En cuanto a la funcionalidad de *speech-to-text* para la creación de rutas se puede habilitar y deshabilitar con el mismo gesto, el cual se explicará en el siguiente apartado.

El establecimiento, borrado y traslado de puntos hacen que se ajuste al mapa cada vez se produce uno de estos eventos, además, los gestos utilizados son un *double tap*, *single tap* y *long press*, respectivamente, los cuales se tratan son gestos familiares para el usuario utilizados en numerosas aplicaciones.

Con el objetivo de adaptar la aplicación y hacer extensible los métodos de interacción gestual para todos los usuarios se incorpora la funcionalidad decidir la mano dominante, ya sea para personas diestras o zurdas, la cual invierte todos los elementos interactivables sobre el eje x según sus preferencias o mano dominante. Por otra parte, en el terreno de las acciones irreversibles, estas requieren de una doble confirmación para realizarlas, dificultando así llevarlas a cabo por error.

Otro de los aspectos destacables en este apartado radica en la posibilidad de creación de rutas de una forma rápida y efectiva para un usuario, ya que se permite iniciar una ruta únicamente con el gesto *double-tap* dos veces seguidas en cualquier punto del mapa y, a continuación, se activará el reconocimiento de voz, para que el usuario introduzca la ruta que desee. A diferencia de otras aplicaciones, esta no requiere de un reconocimiento de voz continuado desde que se inicia la aplicación, ni tampoco que el usuario busque y seleccione el botón adecuado, lo que en situaciones como en la conducción puede provocar despistes y distracciones graves.

En el aspecto cuantitativo podemos observar como nuestra aplicación es más rápida calculando rutas cortas, interprovinciales (rutas 11-14), intraprovinciales (rutas 15-18) y rutas sin solución (rutas 19-20) pero, salvo algún caso, tarda más tiempo que Google Maps en el cálculo de rutas internacionales (rutas 1-9).

Tabla 3: Tiempo de cálculo de rutas (en segundos)

Ruta	Tiempo Google Maps	Tiempo aplicación propia
Madrid - Stuttgart	3,90	2,50
Lisboa - Lyon	3,21	1,60
Salamanca - Florencia	3,37	4,00
Nueva York - Montreal	3,47	4,96

Dallas - Guatemala	5,07	4,58
Buenos Aires - Caracas	4,81	4,97
Bamako - Burkina Faso	2,62	2,97
El Cairo - Praga	3,54	3,42
Kabul - Calcuta	2,56	3,21
Moscú - Oslo	3,91	4,32
Medina del Campo - El Burgo de Osma	2,47	1,04
Calatayud - Villarrobledo	2,29	1,78
Aroche - Écija	2,51	1,29
Ezcaray - Carcastillo	2,40	1,44
Tórtoles de Esgueva - Sotillo de la Ribera	1,66	0,74
Torrecilla del Pinar - Sepúlveda	2,07	1,25
Ortigueira - Loiba	1,88	1,26
Barahona - La Riba de Escalote	2,03	1,43
Boston - Oporto	2,08	1,56
San Diego - Tokyo	1,37	1,06

6. Conclusiones y líneas futuras de trabajo

Este artículo describe una aplicación de mapas para el sistema operativo *iOS*, con funcionalidades tanto del ámbito de la interacción gestual como del *text-to-speech* y *speech-to-text*. El resultado es una aplicación completamente funcional y accesible para la mayoría de los usuarios.

Finalizado este artículo y habiendo utilizado la aplicación, se observa la posibilidad de realizar gestos sencillos y accesibles para el usuario. De igual modo incorpora la posibilidad de revertir acciones fácilmente.

Otro de los aspectos importantes es la implementación de la creación de rutas mediante el reconocimiento de voz, el cual se activa con un gesto de *double tap*. Con esto conseguimos que el reconocimiento de voz no se encuentre activo mientras la aplicación se encuentre activa, mejorando así la privacidad pero haciendo esta funcionalidad accesible en instantes de tiempo en los que no se pueda manejar el dispositivo correctamente. El tiempo de cálculo de rutas interprovinciales, intraprovinciales y rutas sin solución es significativamente menor respecto a aplicaciones similares.

Para este trabajo se ha realizado una investigación y desarrollo de una aplicación de mapas para los sistemas operativos *iOS* y *iPadOS*, con funcionalidades tanto del ámbito de la interacción gestual como del *text-to-speech* y *speech-to-text*. El resultado es una aplicación completamente funcional y accesible para la mayoría de los usuarios.

En primer lugar, se ha llevado a cabo una revisión sistemática de literatura de los avances en el campo de la interacción gestual en los años comprendidos entre

2016 y 2020. Se identifican los principales tipos de gestos reconocidos, la tecnología utilizada, los métodos de preprocesamiento, identificación y clasificación, así como los dominios de aplicación en los que se aplica.

Dado que también se incorpora tecnología de reconocimiento y síntesis de voz, se ha realizado una segunda revisión sistemática de los artículos publicados entre 2018 y 2020. Se especifican las diferentes técnicas y modelos de síntesis, así como los procesos de optimización que se encuentran en desarrollo y los dominios de aplicación de los mismos.

Para el desarrollo de esta aplicación se han hecho uso de bibliotecas y tecnología existente, los cuales implementan por sí mismos el reconocimiento de gestos y de voz, logrando así una capa de abstracción adicional y permitiéndonos centrar nuestro esfuerzo en la utilización de la tecnología existente. Estas bibliotecas están optimizadas para el hardware que utilizamos, en este caso *iPhone* o *iPad*.

6.1 Líneas de trabajo futuras

Entre las posibles líneas de trabajo futuras encontramos:

- Mejorar el reconocimiento de voz, en este apartado se engloba tanto el aspecto propio de la metodología de reconocimiento, así como su posible extensión a otros ámbitos y funcionalidades de la aplicación, creación de mapas o polígonos mediante voz.
- Mejorar la ergonomía, acceso a funciones de la aplicación con un menor número de gestos y carga cognitiva. Esta sería llevada a cabo mediante pruebas con usuarios.
- Añadir mayor número de funcionalidades y características a la aplicación, como puede ser la exportación de mapas a un mayor número de formatos, tamaño visible de los polígonos mostrados o mejora del tiempo de respuesta.
- Reducir el tiempo en el cálculo de rutas, se propone una aproximación del cálculo de ruta óptima mediante computación cuántica y el uso de la biblioteca *Qiskit*, desarrollada por IBM [5].

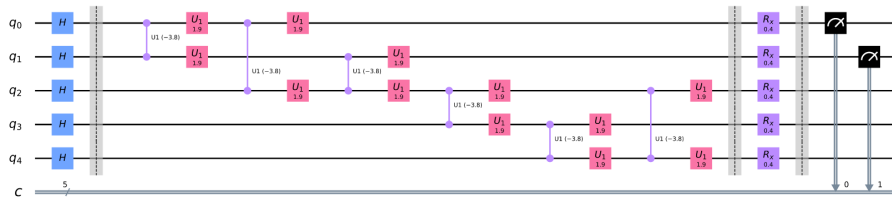


Figura 5: Circuito cuántico computacional

Referencias

- [1] Shahzad Ahmed y Sung Ho Cho. “*Hand Gesture Recognition Using an IR-UWB Radar with an Inception Module-Based Classifier*”. En: *Sensors* 20.2 (2020). DOI: 10.3390/s20020564. URL: %3CGo%20to%20ISI%3E://WOS:000517790100238.
- [2] M. Airaksinen y col. “*A Comparison Between STRAIGHT, Glottal, and Sinusoidal Vocoding in Statistical Parametric Speech Synthesis*”. En: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26.9 (sep. de 2018), págs. 1658-1670. ISSN: 2329-9304. DOI: 10.1109/TASLP.2018.2835720.
- [3] D. Avola y col. “*Exploiting Recurrent Neural Networks and Leap Motion Controller for the Recognition of Sign Language and Semaphoric Hand Gestures*”. En: *IEEE Transactions on Multimedia* 21.1 (ene. de 2019), págs. 234-245. ISSN: 1941-0077. DOI: 10.1109/TMM.2018.2856094.
- [4] M. P. Aylett, A. Vinciarelli y M. Wester. *Speech Synthesis for the Generation of Artificial Personality*. Abr. de 2020. DOI: 10.1109/TAFFC.2017.2763134.
- [5] J.S. Brumberg, K.M. Pitt y J.D. Burnison. “*A Noninvasive Brain-Computer Interface for Real-Time Speech Synthesis: The Importance of Multimodal Feedback*”. English. En: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26.4 (2018). cited By 10, págs. 874-881. ISSN: 15344320. DOI: 10.1109/TNSRE.2018.2808425. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85042376864&doi=10.1109%5C%2fTNSRE.2018.2808425&partnerID=40&md5=99821bcbf90b1b7cb5df045523a680bb>.
- [6] S. Chanjaradwichai, A. Suchato y P. Punyabukkana. “*A multi model HMM based speech synthesis*”. English. En: *Engineering Journal* 22.1 (2018). cited By 0, págs. 187-203. ISSN: 01258281. DOI: 10.4186/ej.2018.22.1.187. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85042654137&doi=10.4186%5C%2fej.2018.22.1.187&partnerID=40&md5=7953a31a44232ce53b04397195c764b6>.
- [7] Ankit Chaudhary y J. L. Raheja. “*Light invariant real-time robust hand gesture recognition*”. En: *Optik* 159 (2018), págs. 283-294. ISSN: 0030-4026. DOI: 10.1016/j.ijleo.2017.11.158. URL: %3CGo%20to%20ISI%3E://WOS:000430761300035.
- [8] C. Chiang. “*Cross-Dialect Adaptation Framework for Constructing Prosodic Models for Chinese Dialect Text-to-Speech Systems*”. En: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26.1 (ene. de 2018), págs. 108-121. ISSN: 2329-9304. DOI: 10.1109/TASLP.2017.2762432.
- [9] U. Côté-Allard y col. “*Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning*”. En: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.4 (abr. de 2019), págs. 760-771. ISSN: 1558-0210. DOI: 10.1109/TNSRE.2019.2896269.

- [10] M. Cotescu y col. “*Voice Conversion for Whispered Speech Synthesis*”. English. En: *IEEE Signal Processing Letters* 27 (2020). cited By 0, págs. 186-190. ISSN: 10709908. DOI: 10.1109/LSP.2019.2961213. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079096163&doi=10.1109%5C%2fLSP.2019.2961213&partnerID=40&md5=e61000c37a3d018b375cb634ab5940e7>.
- [11] Marta Sylvia Del Rio Guerra y col. “*Hand Gestures in Virtual and Augmented 3D Environments for Down Syndrome Users*”. En: *Applied Sciences-Basel* 9.13 (2019). DOI: 10.3390/app9132641. URL: %3CGo%20to%20ISI%3E://WOS:000477031900060.
- [12] Akif Demirezen. *Mapkit tutorial*. <https://medium.com/@maddemirezen/how-to-get-and-display-location-using-mapkit-view-swift-4-2274182bff73>.
- [13] Matthew Devfright. *Mapkit tutorial*. <https://www.devfright.com/mkmapview-and-mkmapview-delegate-tutorial/>.
- [14] Zeyu Ding y col. “*Similar Hand Gesture Recognition by Automatically Extracting Distinctive Features*”. En: *International Journal of Control Automation and Systems* 15.4 (2017), págs. 1770-1778. ISSN: 1598-6446. DOI: 10.1007/s12555-015-0403-6. URL: %3CGo%20to%20ISI%3E://WOS:000406822800032.
- [15] Stephan Dowless. *Mapkit tutorial*. <https://www.youtube.com/watch?v=dB-vB9uDRCl>.
- [16] Google Firebase. *Mapkit tutorial*. <https://console.firebase.google.com/>.
- [17] M. Freixes, F. Alías y J.C. Socoró. “*A unit selection text-to-speech-and-singing synthesis framework from neutral speech: proof of concept*”. English. En: *Eurasip Journal on Audio, Speech, and Music Processing* 2019.1 (2019). cited By 0. ISSN: 16874714. DOI: 10.1186/s13636-019-0163-y. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85076496195&doi=10.1186%5C%2fs13636-019-0163-y&partnerID=40&md5=d0debda3b8b6ba3603a60c18ad3aeld4>.
- [18] Sang-Ho Han y col. “*Natural Hand Gesture Recognition with an Electronic Textile Goniometer*”. En: *Sensors and Materials* 31.5 (2019), págs. 1387-1395. ISSN: 0914-4935. DOI: 10.18494/sam.2019.2261. URL: %3CGo%20to%20ISI%3E://WOS:000468176500002.
- [19] B. Hemanth, K. Sai Venkat e Y. Bhaskar Rao. “*Hand gesture recognition system using deep learning*”. En: *International Journal of Engineering and Advanced Technology* 8.6 Special Issue (2019). cited By 0, págs. 240-245. DOI: 10.35940/ijeat.F1048.0886S19. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85073377411%5C&doi=10.35940%5C%2fijeat.F1048.0886S19%5C&partnerID=40%5C&md5=da0e89a8c75814f751aa62997de36eae>.
- [20] Q. Hu y col. “*ViHand: Gesture Recognition with Ambient Light*”. En: *2019 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCCom/IOP/Ago. de 2019*, págs. 468-474. DOI: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00122.

- [21] J. Huang, Y. Shang y H. Chen. “*Improved Viola-Jones face detection algorithm based on HoloLens*”. En: *Eurasip Journal on Image and Video Processing* 2019.1 (2019). cited By 2. DOI: 10.1186/s13640-019-0435-6. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061351190%5C&doi=10.1186%5C%2Fs13640-019-0435-6%5C&partnerID=40%5C&md5=00e4c8e570329eac3c53ec45b63ef520>.
- [22] L. Huang, S. Zhuang y K. Wang. “*A Text Normalization Method for Speech Synthesis Based on Local Attention Mechanism*”. En: *IEEE Access* 8 (2020), págs. 36202-36209. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2974674.
- [23] S. Imura y H. Hosobe. “*A hand gesture-based method for biometric authentication*”. En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10901 LNCS (2018). cited By 1, págs. 554-566. DOI: 10.1007/978-3-319-91238-7_43. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061554446%5C&doi=10.1007%5C%2F978-3-319-91238-7_43%5C&partnerID=40%5C&md5=d97f7e275949e5dd845712d671075a50.
- [24] Z. Ju y col. “*An Integrative Framework of Human Hand Gesture Segmentation for Human-Robot Interaction*”. En: *IEEE Systems Journal* 11.3 (sep. de 2017), págs. 1326-1336. ISSN: 1937-9234. DOI: 10.1109/JSYST.2015.2468231.
- [25] Zhaojie Ju y col. “*An Integrative Framework of Human Hand Gesture Segmentation for Human-Robot Interaction*”. En: *Ieee Systems Journal* 11.3 (2017), págs. 1326-1336. ISSN: 1932-8184. DOI: 10.1109/jsyst.2015.2468231. URL: %3CGo%20to%20ISI%3E://WOS:000417373200013.
- [26] L. Juvela y col. “*GlottNet - A Raw Waveform Model for the Glottal Excitation in Statistical Parametric Speech Synthesis*”. English. En: *IEEE/ACM Transactions on Audio Speech and Language Processing* 27.6 (2019). cited By 3, págs. 1019-1030. ISSN: 23299290. DOI: 10.1109/TASLP.2019.2906484. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85064621868&doi=10.1109%5C%2FTASLP.2019.2906484&partnerID=40&md5=0de9b4871bc14db1f4726b6aa29b5dc1>.
- [27] M. Kiran Reddy y K. Sreenivasa Rao. “*Excitation modeling method based on inverse filtering for HMM-based speech synthesis*”. English. En: *Advances in Intelligent Systems and Computing* 748 (2019). Ed. por Pachori R.B. Tanveer M. cited By 0; Conference of International conference on Machine Intelligence and Signal Processing, MISIP 2017 ; Conference Date: 22 December 2017 Through 24 December 2017; Conference Code:216999, págs. 85-91. ISSN: 21945357. DOI: 10.1007/978-981-13-0923-6_8. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85051970254&doi=10.1007%5C%2F978-981-13-0923-6_8&partnerID=40&md5=6074fdf5db26f0e6a79b15be32cf08da.
- [28] Pradeep Kumar, Partha Pratim Roy y Debi Prosad Dogra. “*Independent Bayesian classifier combination based sign language recognition using facial expression*”. En: *Information Sciences* 428 (2018), págs. 30-48. ISSN: 0020-

0255. DOI: 10.1016/j.ins.2017.10.046. URL: %3CGo%20to%20ISI%3E://WOS:000416495500002.
- [29] Pradeep Kumar y col. “*Coupled HMM-based multi-sensor data fusion for sign language recognition*”. En: *Pattern Recognition Letters* 86 (2017), págs. 1-8. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2016.12.004. URL: %3CGo%20to%20ISI%3E://WOS:000394400200001.
- [30] O. Kwon y col. “*An Effective Style Token Weight Control Technique for End-to-End Emotional Speech Synthesis*”. En: *IEEE Signal Processing Letters* 26.9 (sep. de 2019), págs. 1383-1387. ISSN: 1558-2361. DOI: 10.1109/LSP.2019.2931673.
- [31] H. Lahiani, M. Kherallah y M. Neji. “*Hand Pose Estimation System Based on a Cascade Approach for Mobile Devices*”. En: *Advances in Intelligent Systems and Computing* 736 (2018). cited By 2, págs. 619-629. DOI: 10.1007/978-3-319-76348-4_60. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85044475420%5C&doi=10.1007%5C%2F978-3-319-76348-4_60%5C&partnerID=40%5C&md5=e7867624cce9d69914527bcf66667bb5.
- [32] C. Linqin y col. “*Dynamic hand gesture recognition using RGB-D data for natural human-computer interaction*”. En: *Journal of Intelligent and Fuzzy Systems* 32.5 (2017). cited By 6, págs. 3495-3507. DOI: 10.3233/JIFS-169287. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85018673257%5C&doi=10.3233%5C%2FJIFS-169287%5C&partnerID=40%5C&md5=e870098e91914e6f52f4519561000d5b>.
- [33] L. Liu e Y. Huai. “*Dynamic Hand Gesture Recognition Using LMC for Flower and Plant Interaction*”. En: *International Journal of Pattern Recognition and Artificial Intelligence* 33.1 (2019). cited By 3. DOI: 10.1142/S0218001419500034. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85052921427%5C&doi=10.1142%5C%2FS0218001419500034%5C&partnerID=40%5C&md5=74f114c7fa532f5a5ea7731bf669099e>.
- [34] Peng Liu y col. “*Hand Gesture Recognition Based on Single-Shot Multibox Detector Deep Learning*”. En: *Mobile Information Systems* 2019 (2019). ISSN: 1574-017X. DOI: 10.1155/2019/3410348. URL: %3CGo%20to%20ISI%3E://WOS:000507260200001.
- [35] Q. Liu, P.J.B. Jackson y W. Wang. “*A Speech Synthesis Approach for High Quality Speech Separation and Generation*”. English. En: *IEEE Signal Processing Letters* 26.12 (2019). cited By 0, págs. 1872-1876. ISSN: 10709908. DOI: 10.1109/LSP.2019.2951894. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077778272&doi=10.1109%5C%2FLSP.2019.2951894&partnerID=40&md5=1a7853160150cd5eac3c59d9ac7dd542>.
- [36] Yanqiu Liu, Xiuhui Wang y Ke Yan. “*Hand gesture recognition based on concentric circular scan lines and weighted K-nearest neighbor algorithm*”. En: *Multimedia Tools and Applications* 77.1 (2018), págs. 209-223. ISSN: 1380-7501. DOI: 10.1007/s11042-016-4265-6. URL: %3CGo%20to%20ISI%3E://WOS:000419995400010.

- [37] Icons8 LLC. *Mapkit tutorial*. <https://iconos8.es/icon/set/mapa/ios>.
- [38] J. H. Mosquera y col. "Identifying facial gestures to emulate a mouse: navigation application on Facebook". En: *Ieee Latin America Transactions* 15.1 (2017), págs. 121-128. ISSN: 1548-0992. DOI: 10.1109/tla.2017.7827915. URL: %3CGo%20to%20ISI%3E://WOS:000394600100004.
- [39] Simon Ng. *Mapkit tutorial*. <https://www.appcoda.com/ios-programming-sidebar-navigation-menu/>.
- [40] Simon Ng. *Mapkit tutorial*. <https://github.com/appcoda/Slideout-Sidebar-Demo>.
- [41] J. Pansare y M. Ingle. "2D hand gesture for numeric devnagari sign language analyzer based on two cameras". En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10127 LNCS (2017). cited By 1, págs. 148-160. DOI: 10.1007/978-3-319-52503-7_12. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85011383010%5C&doi=10.1007%5C%2F978-3-319-52503-7_12%5C&partnerID=40%5C&md5=608837fe70fe3d4d3d727ff57956e7a6.
- [42] G. Pomboza-Junez, J.A. Holgado-Terriza y N. Medina-Medina. "Toward the gestural interface: comparative analysis between touch user interfaces versus gesture-based user interfaces on mobile devices". En: *Universal Access in the Information Society* 18.1 (2019). cited By 2, págs. 107-126. DOI: 10.1007/s10209-017-0580-6. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85032031904%5C&doi=10.1007%5C%2Fs10209-017-0580-6%5C&partnerID=40%5C&md5=d2dc932ac48a8ff574acd0d8daf2d903>.
- [43] J. J. Prakash y H. A. Murthy. "Analysis of Inter-Pausal Units in Indian Languages and Its Application to Text-to-Speech Synthesis". En: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 27.10 (oct. de 2019), págs. 1616-1628. ISSN: 2329-9304. DOI: 10.1109/TASLP.2019.2924534.
- [44] Z. Qiu-Yu y col. "Dynamic hand gesture trajectory recognition based on block feature and skin-color clustering". En: *International Journal of Multimedia and Ubiquitous Engineering* 11.12 (2016). cited By 0, págs. 21-34. DOI: 10.14257/ijmue.2016.11.12.03. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85018212720%5C&doi=10.14257%5C%2Fijmue.2016.11.12.03%5C&partnerID=40%5C&md5=208b77ca6a5ce3351f2cb9275abe1185>.
- [45] M.A. Rahim, J. Shin y M.R. Islam. "Hand gesture recognition-based non-touch character writing system on a virtual keyboard". En: *Multimedia Tools and Applications* (2020). cited By 0. DOI: 10.1007/s11042-019-08448-6. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077612394%5C&doi=10.1007%5C%2Fs11042-019-08448-6%5C&partnerID=40%5C&md5=51750e292b299cc1743bab893577c1ae>.
- [46] Md Abdur Rahim, Jungpil Shin y Md Rashedul Islam. "Gestural flick input-based non-touch interface for character input". En: *Visual Computer*

- (2019). ISSN: 0178-2789. DOI: 10.1007/s00371-019-01758-8. URL: %3CGo%20to%20ISI%3E://WOS:000488921500001.
- [47] M.K. Reddy y K.S. Rao. “*Excitation modelling using epoch features for statistical parametric speech synthesis*”. English. En: *Computer Speech and Language* 60 (2020). cited By 2. ISSN: 08852308. DOI: 10.1016/j.csl.2019.101029. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85073953481&doi=10.1016%5C%2Fj.csl.2019.101029&partnerID=40&md5=a0e8cbc978129908cd78b1063d62eb22>.
- [48] M. Saad y col. “*Ultrasonic hand gesture recognition for mobile devices*”. En: *Journal on Multimodal User Interfaces* 12.1 (2018). cited By 2, págs. 31-39. DOI: 10.1007/s12193-017-0257-8. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85037614772%5C&doi=10.1007%5C%2Fs12193-017-0257-8%5C&partnerID=40%5C&md5=7b0c7b1d9c5263a0d51382810d09503f>.
- [49] Y. Saito, S. Takamichi y H. Saruwatari. “*Statistical Parametric Speech Synthesis Incorporating Generative Adversarial Networks*”. English. En: *IEEE/ACM Transactions on Audio Speech and Language Processing* 26.1 (2018). cited By 53, págs. 84-96. ISSN: 23299290. DOI: 10.1109/TASLP.2017.2761547. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85031781820&doi=10.1109%5C%2Ftaslp.2017.2761547&partnerID=40&md5=715f6b764783bcbbb07a7e05cf2cc6f2>.
- [50] A. Sarkar y col. “*Dynamic hand gesture recognition for mobile systems using deep LSTM*”. En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10688 LNCS (2017). cited By 7, págs. 19-31. DOI: 10.1007/978-3-319-72038-8_3. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85038211140%5C&doi=10.1007%5C%2F978-3-319-72038-8_3%5C&partnerID=40%5C&md5=1ed532e984b9e9978752288bad8e8697.
- [51] K. Sawada y col. “*Constructing text-to-speech systems for languages with unknown pronunciations*”. English. En: *Acoustical Science and Technology* 39.2 (2018). cited By 1, págs. 119-129. ISSN: 13463969. DOI: 10.1250/ast.39.119. URL: [https://www.scopus.com/inward/record.uri?eid=2-s2.0-85043497716&doi=10.1250%5C%2fast.39.119&partnerID=40&md5=861f0f1c477d84f278f19de90e5cc21d](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85043497716&doi=10.1250%5C%2Fast.39.119&partnerID=40&md5=861f0f1c477d84f278f19de90e5cc21d).
- [52] P. Sharma y col. “*Reducing footprint of unit selection based text-to-speech system using compressed sensing and sparse representation*”. English. En: *Computer Speech and Language* 52 (2018). cited By 3, págs. 191-208. ISSN: 08852308. DOI: 10.1016/j.csl.2018.05.003. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85049898621&doi=10.1016%5C%2Fj.csl.2018.05.003&partnerID=40&md5=c639ad0b716df2186bd895805eba97fe>.
- [53] Seunghyeok Shin y Whoi-Yul Kim. “*Skeleton-Based Dynamic Hand Gesture Recognition Using a Part-Based GRU-RNN for Gesture-Based Interface*”. En: *Ieee Access* 8 (2020), págs. 50236-50243. ISSN: 2169-3536. DOI:

- 10.1109/access.2020.2980128. URL: %3CGo%20to%20ISI%3E://WOS:000524898700004.
- [54] Joyeeta Singha, Amarjit Roy y Rabul Hussain Laskar. “*Dynamic hand gesture recognition using vision-based approach for human-computer interaction*”. En: *Neural Computing & Applications* 29.4 (2018), págs. 1129-1141. ISSN: 0941-0643. DOI: 10.1007/s00521-016-2525-z. URL: %3CGo%20to%20ISI%3E://WOS:000424000000017.
- [55] Doug Stevenson. *Mapkit tutorial*. <https://medium.com/firebase-developers/should-i-query-my-firebase-database-directly-or-use-cloud-functions-fbb3cd14118c>.
- [56] M. Tahon, G. Lecorve y D. Lolive. “*Can we Generate Emotional Pronunciations for Expressive Speech Synthesis?*” English. En: *IEEE Transactions on Affective Computing* (2018). cited By 2; Article in Press. ISSN: 19493045. DOI: 10.1109/TAFFC.2018.2828429. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045725677&doi=10.1109%5C%2fTAFFC.2018.2828429&partnerID=40&md5=1cbacd3bcb534a984965309015437c89>.
- [57] Shobhakar Tiwari. *Mapkit tutorial*. <https://stackoverflow.com/questions/41189147/mapkit-zoom-to-user-current-location>.
- [58] M.F. Wahid y col. “*Subject-independent hand gesture recognition using normalization and machine learning algorithms*”. En: *Journal of Computational Science* 27 (2018). cited By 10, págs. 69-76. DOI: 10.1016/j.jocs.2018.04.019. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85047092579%5C&doi=10.1016%5C%2fj.jocs.2018.04.019%5C&partnerID=40%5C&md5=04714ff450aad9e11ae5b47173e89be7>.
- [59] X. Wang, S. Takaki y J. Yamagishi. “*Autoregressive Neural F0 Model for Statistical Parametric Speech Synthesis*”. En: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26.8 (ago. de 2018), págs. 1406-1419. ISSN: 2329-9304. DOI: 10.1109/TASLP.2018.2828650.
- [60] Y. Wang y col. “*Comic-guided speech synthesis*”. English. En: *ACM Transactions on Graphics* 38.6 (2019). cited By 1. ISSN: 07300301. DOI: 10.1145/3355089.3356487. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85078882730&doi=10.1145%5C%2f3355089.3356487&partnerID=40&md5=e6c992079ce471b7e29721657074e851>.
- [61] C. Xu y col. “*Accurate hand detection from single-color images by reconstructing hand appearances*”. En: *Sensors (Switzerland)* 20.1 (2020). cited By 0. DOI: 10.3390/s20010192. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077510517%5C&doi=10.3390%5C%2fs20010192%5C&partnerID=40%5C&md5=3c1a4ddbd7dbc8d76e992e622bd25c8f>.
- [62] Kapil Yadav y Jhilik Bhattacharya. “*Real-Time Hand Gesture Detection and Recognition for Human Computer Interaction*”. En: *Advances in Intelligent Systems and Computing* 384 (2016). Ed. por S. Berretti, S. M. Thampi y P. R. Srivastava, págs. 559-567. DOI: 10.1007/978-3-319-23036-8_49. URL: %3CGo%20to%20ISI%3E://WOS:000381427400049.

- [63] K.N. Yassen y col. “*Hand detection and segmentation using smart path tracking fingers as features and expert system classifier*”. En: *International Journal of Electrical and Computer Engineering* 9.6 (2019). cited By 0, págs. 5277-5285. DOI: 10.11591/ijece.v9i6.pp5277-5285. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85071677444%5C&doi=10.11591%5C%2fijece.v9i6.pp5277-5285%5C&partnerID=40%5C&md5=847d47d4c1209337697cdb6fc03e59cf>.
- [64] Yifan Zhang y col. “*EgoGesture: A New Dataset and Benchmark for Ego-centric Hand Gesture Recognition*”. En: *Ieee Transactions on Multimedia* 20.5 (2018), págs. 1038-1050. ISSN: 1520-9210. DOI: 10.1109/tmm.2018.2808769. URL: %3CGo%20to%20ISI%3E://WOS:000430728400002.
- [65] Yufei Zhang, Bin Liu y Zhiqiang Liu. “*Recognizing Hand Gestures With Pressure-Sensor-Based Motion Sensing*”. En: *Ieee Transactions on Biomedical Circuits and Systems* 13.6 (2019), págs. 1425-1436. ISSN: 1932-4545. DOI: 10.1109/tbcas.2019.2940030. URL: %3CGo%20to%20ISI%3E://WOS:000507321400029.
- [66] Z. Zhang, Z. Tian y M. Zhou. “*HandSense: smart multimodal hand gesture recognition based on deep neural networks*”. En: *Journal of Ambient Intelligence and Humanized Computing* (2018). cited By 6; Article in Press. DOI: 10.1007/s12652-018-0989-7. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85052634594%5C&doi=10.1007%5C%2fs12652-018-0989-7%5C&partnerID=40%5C&md5=0e9dee5f37fd222ffe28388b12290958>.

Sistema de Clasificación Emocional basado en Estímulos Externos

Nuria Mateos, Belén Pérez, Ana Belén Gil, and
Ana de Luis

Departamento de Informática y Automática,
Facultad de Ciencias, Universidad de Salamanca
Plaza de los Caídos, 37008 Salamanca, España
{nurimateos, abg, lancho, adeluis}@usal.es

Resumen Los grandes avances en la tecnología han permitido el desarrollo de dispositivos portátiles capaces de monitorizar distintas medidas fisiológicas de una manera económica, no invasiva y eficiente. La Realidad Virtual (VR) también ha evolucionado logrando una experiencia inmersiva muy realista en diferentes contextos. La combinación de los dispositivos de adquisición de señales y la VR permite generar conocimiento útil incluso en situaciones desafiantes de la vida diaria. En el campo de la Inteligencia Artificial (IA), la información recopilada es clave para desarrollar sistemas de reconocimiento de emociones vitales para controlar la salud humana. Se investiga la viabilidad de detectar estrés en individuos utilizando señales fisiológicas recolectadas con un sensor de fotoplethysmografía (PPG) incorporado en un reloj de muñeca. Las características adquiridas durante la simulación en VR se toman como entrada a un modelo basado en algoritmos de aprendizaje automático, que informará si el individuo presenta o no estrés.

Keywords: Fotoplethysmografía; Reconocimiento de emociones; Realidad Virtual; Dispositivos inteligentes móviles; Redes Neuronales.

1. Introducción

Las personas aprenden a diario en una amplia variedad de contextos, como en el trabajo, la universidad o en la vida privada [15]. En ocasiones, la adquisición de conocimiento puede ser un desafío, como identificar las reacciones y los comportamientos de los conductores ante ciertos eventos externos visuales y acústicos [36] o imprevistos como un accidente de tráfico, y conocer qué emociones fueron provocadas [37]. Cada vez más investigadores en el campo de la IA han llevado a cabo estudios de computación afectiva en general, y reconocimiento de emociones en particular, convirtiéndolos en un área de investigación emergente y prometedora [37].

Análisis estadísticos recientes revelan que el comportamiento humano representa una de las causas principales de los accidentes de tráfico [16]. Varias

investigaciones han comprobado cómo el estado anímico del conductor y las distintas situaciones que ocurren durante un trayecto en carretera, como la fatiga, los atascos, los cruces y los semáforos afectan a la salud y en concreto, al estrés de los pasajeros [5], [13], [16], [21]. Hoy en día se sabe que el estrés es uno de los problemas de salud mental más padecidos por la población y que tiene efectos muy perjudiciales para la salud [18], por lo que su detección puede conducir a su mitigación [18]. Por este motivo, monitorear las condiciones en tiempo real del usuario puede ser clave para detectar sus conductas [2], [33], [36].

Se han desarrollado sistemas capaces de adaptarse al estado emocional de los usuarios para mejorar el aprendizaje y la experiencia [15], [20], [32], [34], fusionando datos recopilados de múltiples sensores que utilizan diferentes canales de expresión emocional humana (audio [2], [3], [38], vídeo y postura [29], registros informáticos [6] y expresión facial [22], [26], [27], [29]). El problema de todos estos canales es que son propensos al enmascaramiento social [17], y muchos de los métodos planteados atentan contra la privacidad y seguridad, como las grabaciones de vídeo o del discurso, o el registro de teclas pulsadas. Por el contrario, las señales fisiológicas, que son resultados directos de la actividad del Sistema Nervioso Autónomo (SNA) y que reflejan el estado emocional intrínseco de la persona, evitan los encubrimientos afectivos, siendo más confiables y objetivos [17]. Además, la literatura científica evidencia la utilidad concreta de posibles enfoques basados en el análisis de señales fisiológicas para la detección y clasificación de emociones [8], [9], [12], [14], [15], [17], [24], [32] y [34].

La *Photoplethysmography* o Fotoplethysmografía (PPG) es una técnica que permite detectar signos vitales mediante la medición de los cambios en la absorción de la luz en la piel [15]. Es un método utilizado muy a menudo para medir la Frecuencia Cardíaca (FC) de forma no invasiva, portable y a bajo costo, pero con buenos resultados. Con una luz infrarroja mide las variaciones volumétricas de la circulación sanguínea, información muy importante en el tema de la salud. El proceso de funcionamiento se basa en la proyección de luz en la piel a través de unos LEDs verdes del dispositivo, que llegan a la sangre que circula por el individuo, rebota, y un foto-diodo colocado bajo el dispositivo captura esa luz. Cuanto mayor cantidad de sangre circule por los vasos sanguíneos mayor será la luz reflejada y viceversa, conociendo así la FC. Investigaciones recientes han utilizado estos dispositivos para clasificar el estrés de los individuos con una precisión del 75.56 % en [40] y del 81.2 % en [19] aplicando algoritmos de *Machine Learning* (ML).

La evolución tecnológica ha provocado grandes desarrollos en sensores portátiles capaces de monitorizar distintas medidas fisiológicas humanas de manera económica, no invasiva y eficiente, al igual que con otros métodos tradicionales, como la Electroencefalografía (EEG) [4], donde la monitorización con estos dispositivos restringe la actividad, es costosa e incómoda [2]. Por ello, se investiga la viabilidad de detectar el estado de estrés de un individuo utilizando señales fisiológicas recolectadas con un sensor de PPG incorporado en un reloj de mu-

ñeca ampliamente utilizados hoy en día.

La motivación de este trabajo es combinar las posibilidades que ofrece la VR para la realización de un sistema de reconocimiento del estrés de un individuo durante distintas situaciones de conducción de un vehículo. La combinación de estímulos visuales, de audio y la posibilidad de interacción e inmersión total del usuario en el entorno tridimensional, potencia el realismo y el impacto de los estímulos emocionales [12], [23], [28]. De esta manera, se plantea una innovadora forma de aprendizaje y de inducción de estímulos, que de manera segura sorteas las limitaciones de espacio y tiempo a través del Head Mounted Display (HMD)¹ de VR. Esta tecnología permite sumergirse a los usuarios en espacios virtuales diseñados específicamente con un fin concreto, recrear escenarios muy difíciles e incluso imposibles en la vida real, como en este caso un accidente, y los usuarios pueden incluso interactuar y manipular los objetos libremente y sin peligro, mediante controladores o con otro tipo de tecnologías, como el seguimiento ocular o con sensores de movimiento [23]. En [11] se comprobó cómo el uso de la tecnología de VR para provocar las emociones aumenta y estimula las respuestas fisiológicas de los sujetos, ya que los resultados obtenidos consistieron en un aumento de la Actividad Electrodermica de la piel (EDA) media y la FC respecto a una simulación estándar o sin inmersión (2D) durante una simulación de conducción.

El sistema propuesto en el presente trabajo es un modelo de aprendizaje automático capaz de determinar mediante clasificación con alta precisión y en tiempo real si los usuarios tienen estrés, utilizando datos de señales fisiológicas basadas en la Variación de la Frecuencia Cardíaca (VFC), obtenidas de las señales PPG capturadas por los relojes portátiles de uso común y de bajo coste. Se implementan distintos algoritmos y se comprueba que la VFC es válida para clasificar el estrés del usuario. El reloj se utilizó para adquirir datos de PPG y respiración, que se ha visto que son efectivos para el reconocimiento de estados emocionales básicos [16], [40]. El *smartwatch* recopila información de la FC, la Saturación de Oxígeno (SO) y la VFC, de las cuales se extraen características de tiempo y frecuencia para luego seleccionar las más adecuadas para el algoritmo de clasificación emocional. También se utilizó el dispositivo Leap Motion de seguimiento de manos para observar los movimientos de los participantes durante la simulación. Los datos recogidos de los sensores biomédicos se grafican, se integran en un sistema móvil de contexto más amplio y se almacenan en el teléfono inteligente en algún formato accesible desde cualquier computadora, en este caso CSV, para llevar a cabo una monitorización en tiempo real y la visualización de las características seleccionadas, como en [25], o las tareas de procesamiento y clasificación. La primera de estas actividades consiste en procesar las señales y extraer las características relevantes para los clasificadores, tareas y conocimiento útil que puede ser obtenido a través de las técnicas de minería de Big Data. La segunda fase implica el uso de las tecnologías de ML, ampliamente utilizadas con éxito para la resolución de cuestiones del mundo real. Cada vez hay más

¹HMD es un dispositivo de visualización en 3D.

investigadores que abordan problemas de toma de decisiones y clasificación con ayuda de algoritmos bioinspirados en el cerebro humano, a saber, las redes neuronales [1], [15], [32], [35], [34], ya que son algoritmos muy flexibles y aplicables a fuentes diversas de distintos tipos.

2. Objetivos

Actualmente, la investigación del reconocimiento de emociones se centra en los siguientes temas: (1) la correlación entre los distintos tipos de señales fisiológicas y las emociones; (2) los métodos de selección de estímulos para la inducción de los estados emocionales deseados; (3) la aplicación de algoritmos que extraigan los rasgos que caracterizan los estados emocionales; y (4) las técnicas de reconocimiento de emociones basadas en la fusión de la información multimodal [37]. Por ello, los objetivos principales y parciales planteados en este trabajo son los siguientes:

- **O1:** Desarrollar un sistema de reconocimiento automático de estrés, información de mucha importancia para la salud humana.
 - **O1.1:** Investigar los métodos de medición de los estados emocionales humanos. La finalidad es ver qué dispositivos de uso común pueden utilizarse para adquirir señales valiosas para la entrada al modelo.
 - **O1.2:** Búsqueda de un conjunto de datos válido para clasificar la emoción del usuario en función de los datos recogidos por los sensores, es decir, ver qué *datasets* contienen como atributos información de las señales capturadas por el sensor.
 - **O1.3:** Aplicar varios algoritmos de aprendizaje automático para construir el clasificador, con el fin comparar y validar los resultados con las métricas principales, y ver cuál es el mejor para el modelo.
 - **O1.4:** Desarrollar un entorno de experimentación para validar el modelo utilizando un simulador de realidad virtual.
 - **O1.5:** Desarrollar una herramienta para un dispositivo móvil que monitorice en tiempo real los resultados obtenidos. De esta manera, el usuario podrá conocer en todo momento su estado de estrés.

3. Métodos

La metodología seguida para conseguir los objetivos del proyecto y desarrollar un sistema de reconocimiento automático de emociones consistió en llevar a cabo varios procedimientos: (1) La búsqueda y selección de una base de datos relevante, (2) El desarrollo de una aplicación para la inducción de estados emocionales a los usuarios, (3) Elegir y utilizar dispositivos capaces de capturar las señales emitidas como respuesta a los estímulos, (4) Introducir la información recopilada por los sensores como entrada a la red neuronal, (5) Aplicar distintos algoritmos de ML para comparar los resultados y validarlos en función de

las distintas métricas existentes, y (6) el desarrollo de una aplicación móvil que integre el sistema clasificador o red neuronal para la determinación del estado emocional del usuario y que permita visualizar características relevantes para el usuario.

El lenguaje de alto nivel utilizado para el desarrollo del proyecto es Python. Su elección se debe a la gran cantidad de bibliotecas existentes y bien optimizadas para llevar a cabo el cálculo científico y las tareas comunes de ML, como Numpy, Pandas, Scikit-learn, entre otras. Además, Python es independiente de la plataforma, por lo que es compatible y se puede ejecutar en la mayoría de los dispositivos. Se utiliza el IDE Jupyter Notebook que proporciona la plataforma Anaconda para el desarrollo del código, ya que es modular, extensible, gratuito y fácil de usar.

3.1 Conjunto de datos

El conjunto de datos utilizado fue creado en el año 2007 y titulado “*Stress Recognition in Automobile Drivers*” o *Reconocimiento de estrés en conductores de automóviles*. Se encuentra disponible de forma gratuita en la plataforma de PysioNet² y en Google Cloud.³ Se realizó una búsqueda exhaustiva de todas las bases de datos relacionadas con el reconocimiento automático de emociones, y tras puntuar los conjuntos de datos disponibles siguiendo ciertos criterios de evaluación, el *dataset* de PysioNet obtuvo la máxima puntuación. Además, también fue elegido debido a la similitud de intenciones del *dataset* con los objetivos del proyecto. El conjunto recopila información de sensores de Electrocardiograma (ECG), de la Respuesta Galvánica de la Piel (GSR) y de la respiración durante la conducción de un vehículo real con el fin de determinar el nivel de estrés del conductor [18]. Además, el monitoreo de las señales se realiza en una posición relativamente inmóvil, ya que está sentado, por lo que las señales pueden ser más claras y similares a las recopiladas en el experimento propuesto en este trabajo.

3.2 Modelo clasificador de estrés

A partir del conjunto de datos se han entrenado diferentes modelos a través de las técnicas de ML. Una vez entrenados los modelos se utilizan con el nuevo conjunto de datos recopilados con el reloj. La información y los datos se adquieren durante el experimento en los escenarios de realidad virtual, es decir, el simulador de conducción y de accidentes, con el fin de determinar si el usuario tiene o no estrés y monitorizar las distintas características.

²PhysioNet: recurso de investigación para señales fisiológicas complejas. Disponible en: <https://physionet.org/content/drivedb/1.0.0/>

³<https://console.cloud.google.com/storage/browser/drivedb-1.0.0.physionet.org>.

3.2.1 Preprocesamiento de los datos. El primer paso realizado fue una exploración inicial al *dataset* para observar qué características o atributos se manejan, qué tipo de datos son y obtener una descripción estadística de todos ellos. El conjunto de datos tiene un total de 23 atributos y 4129 instancias. Todos ellos se corresponden a las señales de ECG, EMG y GSR. La Tabla 1 muestra las características que no se han utilizado para la realización del modelo y el motivo que impulsó la decisión, y las que por el contrario, se han utilizado para la construcción del clasificador.

Características eliminadas		Características válidas	
footGSR handGSR	Sensor no disponible	HR	Frecuencia cardíaca instantánea
EMG ECG	Dificultad de adquisición	RESP	Número de respiraciones por minuto
ULF VLF	Media = 0 Grabaciones largas	Seconds newtime	Tiempo entre intervalos de onda
LF_HF	Datos faltantes	LF HF	Dominio de frecuencia (FC)
interval in seconds time	Correlación 1 con <i>AVNN</i> Correlación con <i>seconds</i>	AVNN SDNN rMSSD pNN50	Dominio de tiempo (FC)
marker	Irrelevante	TP	Segmento desde el final de la onda T hasta el inicio de la siguiente onda P

Tabla 1: Estudio de las características del *dataset*.

Respecto a las características que se pueden extraer de la FC en el dominio de frecuencia, para determinar la ULF el tiempo de grabación tiene que ser a largo plazo, de unas 24 horas [24], por ello este tipo de señal no se suele usar en la práctica y también se descarta para este proyecto, ya que gran parte de los valores que presenta son nulos, probablemente debido a los cortos intervalos de tiempo del experimento. Lo mismo ocurre con la información de la banda de VLF, que prácticamente todos los valores toman valor nulo, por lo que se descarta esta característica para el desarrollo del modelo. Se llega a resultados similares a otros autores [17], que han demostrado que la banda VLF es una medida muy poco confiable en lecturas de menos de 5 minutos. Al realizar una revisión de los *outliers* y de los valores faltantes en cada atributo, se observa que hay una variable, $LF - HF$, que es la relación de potencia de baja a alta frecuencia, que tiene el 99 % de los datos vacíos, por lo cual esta variable se elimina. De los atributos restantes, únicamente el atributo *AVNN* tiene valores faltantes, desconociendo el valor de un total de 122 instancias (el 3 % de las instancias totales). Como es un porcentaje mínimo, se realiza una función para reemplazar dichos valores por la media, ya que se trata de un atributo numérico.

Por lo tanto, el conjunto de características para la construcción del modelo son la FC, la respiración (*RESP*), el tiempo entre intervalos de onda (*seconds*, *newtime* y *TP*), las características de la FC en el dominio de tiempo (*AVNN*, *SDNN*, *rMSSD* y *PNN50*) y en el dominio de frecuencia (*LF* y *HF*). Los intervalos NN toman los valores medios de la FC en una ventana o intervalo de tiempo de 30 segundos. Todas estas características son requeridas para la extracción de características de la VFC. Los indicadores obtenidos tras el análisis en los dominios de tiempo y frecuencia reflejan características fisiológicas e información relacionada con el estrés [7].

Tras el tratamiento inicial de los datos se estudia la correlación entre las variables con el fin de ver si hay atributos altamente correlacionados que puedan ser eliminados. Las variables *interval_in_seconds* y *AVNN*, que se corresponden con el intervalo de tiempo en segundos entre dos latidos y el promedio de todos los intervalos entre latidos, respectivamente, tienen un valor de correlación 1 obtenido mediante el coeficiente de correlación de Pearson, por tanto se elimina una de ellas (*interval_in_seconds*), ya que no aporta información adicional, puesto que miden la misma característica. Lo mismo ocurre con los atributos *time* y *seconds*, en el que fue descartado el primero de ellos. Pese a que hay otras dos variables con valores superiores a 0.7 (*SDNN* y *RMSSD*), no se eliminan en la etapa de preprocesamiento, ya que puede suponer la pérdida de información.

Para finalizar la preparación de los datos para el aprendizaje automático, se realiza una normalización para evitar que las variables medidas en diferentes escalas contribuyan de forma diferente a la función de ajuste y aprendizaje del modelo, pudiendo crear un sesgo. Se utilizó el escalado de variables *StandardScaler* de *scikit-learn* para transformar los datos de manera que la distribución de cada atributo (χ) tenga un valor medio $\mu = 0$ y una desviación estándar $\sigma = 1$, donde cada entrada se normaliza entre unos límites definidos.

La clase considerada es el nivel de estrés del usuario. El interés del problema es conocer si el usuario se encuentra o no con estado de estrés, por lo que se trata de un modelo de clasificación binaria, donde la etiqueta puede tomar dos valores, 1 = *estrés* o 0 = *no estrés*. También se observa que no hay desequilibrio entre las clases, ya que hay 2170 muestras etiquetadas con estrés y 1959 sin estrés, por lo que la diferencia de instancias entre ambas clases, con un valor de 211, no debería afectar a la clasificación. Aún así, en algunos algoritmos, como en los árboles, en algunos parámetros (como *class_weight*) se le da un mínimo valor mayor a la clase minoritaria, para así compensar la diferencia de valores.

3.2.2 Extracción de características. La detección de los estados emocionales requiere una extracción adecuada de las características de las señales, que se correlacionan con los estados emocionales registrados por los participantes en la autoevaluación. En las señales de ECG, que son similares a las de PPG, se extraen características en el dominio del tiempo y de frecuencia para determinar la VFC. Las mediciones paramétricas de las señales ECG en el dominio

del tiempo cuantifican la variabilidad de Intervalos entre Latidos (IBI) sucesivos [31], y son muy importantes para el análisis de las grabaciones a corto plazo [24]. En este método, se determina la FC tomada en cualquier instante o entre intervalos, determinando el Intervalo entre dos Latidos Normales (NN) o la FC instantánea. El dominio de frecuencia determina la distribución de la potencia, ya que distingue las señales de FC según su frecuencia e intensidad. Proporciona información sobre cuánto cambia el valor de la FC aprovechando las oscilaciones periódicas de la FC en varias frecuencias. Los principales componentes espectrales calculados se describen como ULF, VLF, LF y HF. En la Tabla 2 se describen las características generadas a partir de los picos de las señales ECG.

Dominio de tiempo		Dominio de frecuencia	
AVNN	Promedio de todos los intervalos NN.	ULF	Frecuencia ultrabaja: potencia espectral total de todos los intervalos NN hasta 0.003Hz.
SDNN	Desviación estándar de todos los intervalos NN.	VLF	Muy baja frecuencia: potencia espectral total de todos los intervalos NN entre 0.003 y 0.04 Hz.
rMSSD	Raíz media cuadrática de diferencias entre intervalos NN adyacentes.	LF	Baja frecuencia: potencia espectral total de todos los intervalos NN entre 0.04 y 0.15 Hz.
pNN50	Porcentaje de diferencias entre intervalos NN adyacentes que son mayores de 50 ms.	HF	Alta frecuencia: potencia espectral total de todos los intervalos NN entre 0.15 y 0.4Hz.

Tabla 2: Medidas de dominio de tiempo y frecuencia de la señal.

3.2.3 Selección de características. Se pueden utilizar distintos métodos de selección de funciones para analizar la relevancia de una característica y seleccionar un subconjunto de atributos. El objetivo principal de la extracción y selección de características es combinar aquellas que mejor representen el conjunto de datos y ver cuáles son las más importantes. Se ha realizado un método de filtrado de características independiente del clasificador para ordenar las características y dar una puntuación a cada una de ellas para estimar la clase. Se utilizó la clase *SelectKBest* de la librería *Sklearn* para la puntuación de las características, que elimina todas excepto los K atributos con la puntuación más alta, que en la práctica se le estableció un valor de $K = 5$. El resultado devuelve una lista con los mejores atributos: $[HR, NNRR, AVNN, SDNN, RMSSD]$.

3.2.4 División del conjunto de datos. Una vez preprocesados los datos y realizada la extracción y selección de características, se divide el conjunto de datos en dos subconjuntos diferentes, uno de entrenamiento y otro de prueba, para el correcto desarrollo y aprendizaje del modelo de ML. Para la división de datos se utiliza la función *train_test_split* de *Sklearn*, que divide los datos en

subconjuntos aleatorios de conjuntos de entrenamiento (*train*) y prueba (*test*). Durante la aplicación de la función se estableció el parámetro *test_size* con un valor de 0.2, lo que significa que el 20 % de los datos se utilizan para la prueba y el 80 % restante para el entrenamiento.

Se implementaron distintos algoritmos supervisados, y algunos de ellos se probaron con distintos parámetros, con el fin de encontrar el modelo que mejor realice la clasificación. Independientemente del algoritmo utilizado, se realiza una validación cruzada durante el entrenamiento para comprobar si el modelo es válido introduciéndole entradas de un nuevo conjunto de datos.

3.2.5 Entrenamiento y validación del modelo.

Clasificador gaussiano. El primer modelo desarrollado implementó el algoritmo NB basado en probabilidades y utilizado para la clasificación. Se utilizan las variables que almacenan los datos de entrenamiento y de prueba, ya que la clase debe estar separada. Se crea una instancia del clasificador *GaussianNB()* mediante la librería de Sklearn, y se entrena el modelo pasándole como entrada los valores de la lista de características seleccionadas obtenidas en 4 y las etiquetas o clases correspondientes a cada una de las instancias. Una vez entrenado el modelo, se comprueba la precisión en la clasificación con los datos de entrenamiento, comparando la predicción realizada por el modelo con los datos de entrenamiento con las etiquetas reales. Se obtuvo un 56 % de precisión para el conjunto de entrenamiento. La precisión del modelo con el conjunto de datos de prueba fue del 55 %. Los resultados de las distintas métricas de evaluación indican que el modelo no clasifica correctamente, ya que se obtiene un valor de *accuracy* del 56 %, con *Recall* un 55 %, *F1 Score* 56 % y *ROC* 56 %.

KNN. El algoritmo KNN o K vecinos más cercanos está basado en instancias y sirve para la clasificación de valores buscando los datos más similares por cercanía aprendidos durante el entrenamiento. Por este motivo es muy importante elegir qué valor o número de elementos más cercanos se va a seleccionar. Por ello, previamente se llevó a cabo una transformación de las características, escalándolas dentro de un rango con *MinMaxScaler* y se buscó qué valor dar al parámetro *k*, para ello se hace una comparación de la tasa de error con distintos valores de *k*, eligiendo el valor final 7, que es el que mejor precisión logra. Antes de realizar el modelo, se realiza un preprocesamiento previo para la aplicación del algoritmo, que consiste en transformar las características escalando cada una de ellas dentro de un rango con *MinMaxScaler*, para que así todas las variables tengan el mismo peso. Una vez que se crea el modelo *KNeighborsClassifier(n_neighbors = k)*, éste se entrena y se valida el resultado de la clasificación. Se obtiene una precisión del 80 % en el conjunto de entrenamiento y del 71 % en los datos de prueba. La matriz de confusión devuelve que el modelo clasifica 327 instancias correctamente como verdaderos positivos (VP), y 141 de forma incorrecta como falsos

negativos (FN). Por otro lado, el modelo clasifica 409 instancias correctamente como verdaderos negativos (VN), mientras que clasifica 156 instancias como falsos positivos (FP). El valor devuelto por *Recall* es del 71 %, al igual que *F1 Score*.

		PREDICCIÓN	
		Positivos	Negativos
REALIDAD	Positivos	327	141
	Negativos	156	409

Tabla 3: Matriz de confusión con KNN.

Árboles de Decisión. También se implementaron varios algoritmos de aprendizaje basados en árbol para la clasificación, como DT. Los árboles de decisión son representaciones gráficas de posibles soluciones a una decisión basadas en ciertas condiciones, y es uno de los algoritmos más utilizados para la clasificación. Para el DT, se estableció la entropía como parámetro para medir el número de divisiones para la profundidad máxima del árbol. Se utiliza la función *KFold* para ver cuál es la precisión en función de cada valor de la variable. Esta función crea varios subgrupos a partir de los datos iniciales de entrada para valorar y validar los árboles con distintos niveles de profundidad, para poder elegir el mejor resultado. En nuestro caso se eligió el valor 20. Se crea el árbol *tree.DecisionTreeClassifier* para la clasificación, y se establecen los parámetros *criterion = entropy*, eligiendo la entropía como medida de incertidumbre, *min_samples_split = 20*, estableciendo en 20 el valor mínimo de muestras que el nodo debe tener para poder subdividir, *min_samples_leaf = 5*, por lo que 5 es la cantidad mínima que puede tener una hoja final, *max_depth = depth*, ya que la variable *depth* contiene el mejor resultado encontrado de niveles de profundidad del árbol, y *class_weight = 1 : 1,2*, con peso 1 para la clase estrés y peso 1.2 para la variable no estrés, para así compensar el número de instancias de más que tiene primera clase. En algunos de estos parámetros se fueron haciendo pruebas para ver qué valores eran los óptimos para el modelo y para la clasificación. La precisión alcanzada con DT es del 90.34 %, y las métricas de validación indican que el modelo clasifica de forma correcta, obteniendo un valor de *accuracy* del 90.34 %, *Recall* un 88 %, *F1 Score* 86 % y *ROC* 89 %.

Random Forest. Random Forest se basa en una gran cantidad de árboles de decisión individuales que operan como conjunto. Cada árbol devuelve una predicción de la clase dentro del bosque, y la predicción del modelo será la clase con mayor votos. Se crea el modelo con *RandomForestClassifier* y se establece una cantidad de 100 árboles como parámetro, y *sqr*t como método para la selección de la cantidad máxima de características para cada árbol, de las cuales el número de las que serán consideradas en cada división se realiza de forma

automática. Se usan muestras para la construcción de los árboles, se entrena el modelo y se observan los resultados de la clasificación. La *accuracy* del modelo en la clasificación es del 84 %, *Recall* devuelve un valor del 82 %, *F1 Score* del 83 %.

Bagging. El algoritmo *Bagging* se ajusta a clasificadores en subconjuntos aleatorios del conjunto de datos original agregando luego sus predicciones individuales. Se utiliza para reducir la varianza del árbol de decisión. Se realiza un entrenamiento de los datos mediante *10-fold cross-validation* o validación cruzada, para garantizar que los resultados son independientes de la partición entre datos de entrenamiento y prueba. Se obtiene una precisión del 87.47 %, ya que clasifica 3612 instancias de forma correcta y 517 de forma incorrecta. *Recall* y *F1 Score* devuelven un valor del 87.5 %, mientras que *ROC* devuelve un 94.4 %, por lo que es un valor muy cercano al 100 %.

Perceptrón multicapa. Se crea una red neuronal basada en el perceptrón de cuatro capas ocultas. Tras instanciar el modelo *MLPClassifier*, durante el entrenamiento se ajustan los pesos de cada una de las neuronas. Se parametriza o limita el algoritmo estableciendo el número de iteraciones o cantidad de pasos del entrenamiento *num_steps* = 3000 que tendrá la tasa de aprendizaje *learning_rate_init* = 0,03, ya que en cada una de ellas se debe ver una mejora incremental, y se entrena el modelo mediante la función *fit* para que aprenda del conjunto de datos etiquetados. El modelo obtiene una precisión del 78.56 %, clasificando 3244 instancias de forma correcta y 885 de manera incorrecta. *Recall* y *F1 Score* devuelven un valor del 78.4 % y ROC 85 %. Respecto a la matriz de confusión, clasifica 1407 instancias como estrés de forma correcta, y 561 como estrés sin serlo. Mientras que por otro lado clasifica 1837 instancias como no estrés de forma correcta, y 324 instancias como estrés que realmente no lo son.

SVM. Por último, la máquina de vectores de soporte obtuvo precisiones muy bajas en la clasificación, con una precisión del 63.98 %. *Recall* obtiene un valor del 63.9 % y *F1 Score* y *ROC* del 0.637 %. La matriz de confusión devuelve que el modelo clasifica 1126 instancias correctamente como estrés siendo estrés, y 842 de forma incorrecta como no estrés, siendo estrés. Por otro lado, el modelo clasifica 1512 instancias correctamente como no estrés siendo no estrés, mientras que clasifica 649 instancias como estrés cuando realmente no lo es.

3.3 Diseño del experimento

Una vez que está entrenado el modelo, para validar el clasificador, éste se prueba en un escenario de VR. Para ello se inducen emociones por medio de distintos estímulos provocados durante una simulación con el objetivo de obtener los datos de la respuesta fisiológica con el reloj, como muestra la Figura 1. Se

monitoriza a las personas mientras están realizando la simulación de una conducción natural y dos situaciones que están diseñadas con intención de generar estrés, como es un choque frontal con un coche y un vuelco del vehículo. Se utiliza un reloj Garmin Forerunner 235 para recopilar las respuestas fisiológicas de los usuarios, información que dará lugar a un nuevo conjunto de datos que será aplicado para la detección de estrés y la validación del modelo, ya que prueba la efectividad del clasificador que, a través de esos datos de entrada, aplica el algoritmo de ML y clasifica el estado del usuario, en este caso, si le ha provocado o no estrés. Los modelos resultantes de la aplicación de los distintos algoritmos se utilizan para etiquetar el estado estresado o no estresado en el nuevo conjunto de datos sin etiquetar.

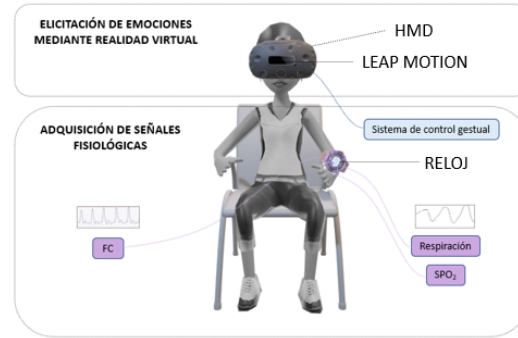


Figura 1: Estructura del sistema propuesto.

La Figura 2 muestra las fases llevadas a cabo durante la experimentación.

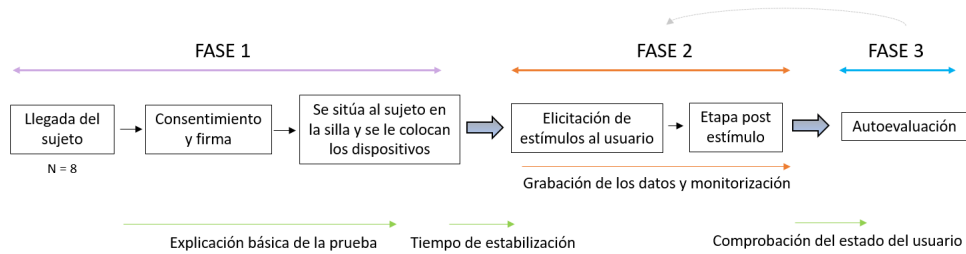


Figura 2: Diseño y protocolo experimental.

Fase 1. Para la selección de participantes se realizó un estudio de los perfiles de los sujetos adecuados para el experimento, ya que se ha descubierto que hay diferencias significativas en la experimentación de emociones en los distintos in-

dividuos [30]. Se reveló que la edad, el género, la personalidad, o la salud de los participantes, especialmente los que siguen algún tratamiento o consumen de forma habitual medicamentos, pueden afectar a los resultados de la prueba [10]. En nuestro caso se excluyeron perfiles de sujetos con enfermedades o problemas cardíacos, ya que podrían inferir en las señales fisiológicas, y a las personas con vértigos o mareos, que pueden verse potenciados al llevar a cabo la simulación en VR. En este experimento participaron 8 personas sanas (4 hombres y 4 mujeres), con un rango de edades que abarca entre los 24 y los 50 años, de forma voluntaria y que firmaron un consentimiento previo. Pese al estudio preliminar de los sujetos que iban a participar en el experimento, antes de comenzar la grabación de cada participante se le hizo rellenar un cuestionario que recoge información general sobre su estado de salud. Los usuarios tienen que afirmar que no han consumido bebidas alcohólicas o café, que no han realizado ejercicio físico antes de la prueba y que carecen de enfermedades o patologías previas, con el fin de evitar alteraciones en los datos. Además, los usuarios firmaron el consentimiento de la realización de la prueba y los riesgos que puede provocar el desarrollo de la misma, como posibles mareos, ataques de ansiedad u otro tipo de posibles factores provocados por el uso del HMD, y más aún por el elevado carácter de la prueba, ya que la simulación de un accidente es una situación de elevado sobresalto. A continuación, los participantes fueron provistos de un conjunto de instrucciones que explican el protocolo experimental, sin revelar lo que sucede en la simulación, además, el conductor interactúa e informa al usuario durante la simulación. Una vez proporcionadas las instrucciones de seguridad y la información necesaria para la realización de la prueba, se manda al usuario sentar en una silla, se le coloca la pulsera en la muñeca y se le pone el HMD a la altura de los ojos sobre la cabeza, junto con los cascos de sonido. Para evitar las distracciones y aumentar la concentración, sólo el usuario y la persona que supervisa la prueba están en la sala. Para la captura de las señales fisiológicas y, en concreto, las relacionadas con el sensor PPG que se encuentra integrado en algunos relojes y que obtiene información sobre la FC del usuario, se revisaron los dispositivos más utilizados por otros autores para obtener la VFC. Las señales de PPG se registraron mediante el sensor óptico del reloj con una frecuencia de muestreo entre 1 y 2 segundos. El rango de la FC varía entre 50 lpm y 150 lpm. El reloj consta de 3 LEDs verdes y un LED infrarrojo, como se puede observar en la Figura 3, que son los que permiten calcular la FC a través de las variaciones volumétricas en la circulación sanguínea.



Figura 3: Dispositivo *Garmin* de captura de señal.

También se utilizó el dispositivo Leap Motion⁴ para capturar y realizar un seguimiento en tiempo real de los movimientos de los brazos realizados por el usuario. De esta manera, los usuarios pueden ver sus manos durante la simulación e interaccionar con otros objetos del entorno, aportando una mayor credibilidad. Además, con este dispositivo se controlan los cambios de posición y el movimiento de las manos durante toda la prueba. Esto es clave, ya que en este sentido los datos del acelerómetro no son tan importantes como en otros estudios debido a que el individuo realiza todo el experimento sentado, por lo que la precisión en el etiquetado se mantiene, ya que no se malinterpretan los posibles movimientos del usuario como estrés físico. En cambio, los gestos o movimientos realizados con las manos pueden aportar información interesante.

Fase 2. Para la elicitación de emociones se desarrolla una aplicación de VR mediante el motor de juego creado por la empresa Epic Games con el nombre de Unreal Engine 4⁵, con la versión 4.24, ya que es un software muy potente para VR. Está basado en el lenguaje C++ e incluye características de nivel avanzado, como un sistema de iluminación dinámico y en tiempo real, y es un motor gráfico muy potente. Se utilizó el HMD HTC Vive Pro⁶. El software desarrollado consta de un menú principal que abarca dos escenarios tridimensionales diferentes, la simulación de un choque frontal y un vuelco con un vehículo, buscando provocar una emoción en dos situaciones extremas. Durante la ejecución del programa se emitirán los sonidos pertinentes a las distintas acciones. El usuario comienza el juego en el asiento del copiloto, en el interior del coche, junto con el conductor, quien comenzará a explicar al usuario la prueba a la que se va a enfrentar y las acciones de seguridad requeridas. Durante esa breve introducción, el copiloto puede visualizar a través del HMD todo el entorno. Posteriormente, el vehículo comienza su marcha, incrementando progresivamente la velocidad y viajando con normalidad. Se busca inducir un estado emocional base o neutro. En este punto, el desarrollo del juego varía en función de los dos escenarios existentes para inducir estrés al usuario con el choque del vehículo (Figura 4), o con el vuelco (Figura 5).

⁴LeapMotion: <https://www.ultraleap.com/product/leap-motion-controller/>

⁵Unreal Engine 4: <https://www.unrealengine.com/en-US/>

⁶HTC Vive: <https://www.vive.com/mx/product/vive-pro/>



Figura 4: Escenario de choque frontal.



Figura 5: Escenario de vuelco del vehículo.

El desarrollo de la aplicación engloba esquemáticamente varios procesos con el fin de estimular al usuario: el diseño y modelado del entorno tridimensional y de las animaciones, la programación de eventos en la simulación y la configuración del contexto ambiental. El reconocimiento de emociones se centró en las reacciones del participante ocasionadas por los estímulos preparados y el ambiente desarrollado en la simulación. Por cada participante se registró una sesión de 90 segundos, obteniendo un total de 8 sesiones. La mitad de los participantes hicieron la prueba con un escenario, y la otra mitad con el otro. En los primeros minutos de grabación se mostró una pantalla en blanco con un estímulo musical que pretende inducir un estado emocional neutro, y una vez transcurrido el minuto, se sumerge a los usuarios en los escenarios propuestos. El protocolo de adquisición de los datos fue similar para cada participante. Al terminar, cada usuario realiza una autoevaluación informativa que permite localizar aproximadamente los períodos de tiempo donde el usuario se sintió estresado, ya que la percepción del estado emocional es subjetivo o dependiente del usuario, y que la VFC varía en función de la prueba y la tolerancia de cada persona.

3.3.1 Visualización de los resultados. La visualización de los resultados se realiza mediante una aplicación móvil desarrollada para Android, cuya interfaz se muestra en la Figura 6. El software desarrollado contiene el sistema reconocedor de estrés, por lo que al acceder a la información proporcionada por el sensor del reloj, introduce las características como entrada al sistema, realizando una conexión entre los datos recopilados de los sensores y la aplicación móvil, que devuelve el estado del usuario, con estrés o sin estrés.



Figura 6: Interfaz de la aplicación móvil.

4. Resultados

Los resultados tras aplicar los distintos algoritmos se recogen en la Tabla 4. Se muestra la precisión del algoritmo en la clasificación y el valor devuelto por las distintas métricas de evaluación. El algoritmo KNN fue elegido para el clasificador de la aplicación móvil, ya que es uno de los algoritmos de clasificación más simples y que, pese a no ser el que mejores resultados obtiene, éstos son muy competitivos, como se muestra en la Tabla 4, con una precisión en la clasificación del 87.02 %. El rendimiento general del clasificador resumido sobre todos los umbrales posibles viene dado por la curva ROC. Esta métrica obtiene un valor del 87 %, lo que significa que el clasificador es bueno separando instancias de las dos clases e identificando el umbral que mejor las separa. La elección del algoritmo KNN se debe al tipo de problema a resolver y el *dataset* utilizado. Como el número de entradas del modelo no es muy alto, ya que únicamente se dispone del sensor PPG para la captura de los datos, se puede tomar como una ventaja que el algoritmo KNN esté basado en instancias, ya que de esta manera, a la hora de realizar la predicción el algoritmo, se basará en las instancias entrenadas para realizar la clasificación, utilizando esos datos para generar la respuesta. Por lo tanto, aunque no construye ningún modelo, como por ejemplo si ocurre con los DT, clasifica cuando llega la instancia de test, sin suponer la distribución de los datos. Además, la baja dimensionalidad o el pequeño número de atributos tomados como entrada al algoritmo y la selección de características realizadas como preprocesamiento de los datos, evita que la precisión se vea afectada por las características irrelevantes o el ruido. Por otro lado, aunque el coste computacional de este algoritmo es elevado, ya que almacena todos los datos de entrenamiento, necesitando muchos recursos de procesamiento (CPU) y mucha memoria, al tener pocas características de entrada y tratarse de grabaciones de tiempo cortas, no es un problema en este caso de estudio ni para la implementación en el dispositivo móvil, donde no se encuentra retardo en la clasificación ni produce ningún problema de almacenamiento, debido a que no se manejan grandes cantidades de datos, ya que no ha sido necesario llevar a cabo ningún procedimiento de reducción de características (como PCA). Por estos

motivos se elige el algoritmo KNN como clasificador, entendiendo que el mejor modelo de los datos son los propios datos, sin buscar un modelo optimizado, sino comparar cada una de las instancias con los datos de entrenamiento para tener un criterio y medir la similitud de cada instancia a clasificar. No se eligió el algoritmo DT para el clasificador pese a obtener la mayor precisión (90.34 %), ya que en muchos casos presentan inestabilidad en la clasificación. Las pequeñas variaciones en los datos pueden provocar un gran cambio en la estructura del árbol. En la detección de estrés durante la simulación, los datos son cambiantes y con un amplio rango de valores, por lo que esto puede generar oscilaciones en la clasificación. Por otro lado, el RF y el *bagging* suelen mejorar los resultados del DT, y no ha sido el caso de los resultados obtenidos, como se observa en la Tabla 4. Esto puede interpretarse como un sobreajuste del modelo, es decir, puede ser que el algoritmo DT no generalice bien a partir de los datos de entrenamiento, y que la precisión en la predicción con los datos recopilados por el sensor no devuelva resultados precisos. Asimismo, una de las grandes ventajas del DT es su capacidad de identificar variables importantes en problemas de alta dimensionalidad, o la gran cantidad de valores que puede tomar la variable de destino, que no es el caso del *dataset* utilizado, por lo que no son relevantes en nuestro estudio. Además, los árboles de decisión toman las decisiones localmente óptimas en cada nodo, pero no garantizan que el árbol global devuelto sea el más óptimo. El algoritmo bagging, que suele utilizarse para reducir la varianza en los DT, ya que el valor de la varianza puede significar que, al hacer una división aleatoria de los datos de entrenamiento en dos grupos, si se ajusta un DT a cada mitad, los resultados obtenidos pueden ser bastante diferentes, obtiene una precisión muy semejante al algoritmo KNN.

ALGORITMO	RESULTADO			
	Precisión	Recall	F1 Score	ROC
Decision Tree	90.34 %	0.88	0.86	0.89
Bagging	87.47 %	0.87	0.87	0.94
Random Forest	84.13 %	0.82	0.85	0.83
K-Nearest Neighbors	87.02 %	0.71	0.71	0.87
Multilayer Perceptron	78.56 %	0.78	0.78	0.85
SVM	63.98 %	0.63	0.63	0.63
Gaussian Näive Bayes	56.00 %	0.55	0.56	0.56

Tabla 4: Resultados en la clasificación con los distintos algoritmos.

Tras realizar el modelo se llevan a cabo las pruebas con los datos recopilados con el reloj. El modelo recibe como entradas la información adquirida de las señales fisiológicas y devuelve el resultado de la clasificación a través del dispositivo móvil, indicando si el sujeto tiene o no estrés. El resultado del modelo se comprueba con la autoevaluación realizada por el usuario, que indica las etapas

que sintió estresantes y las que estuvo en un estado neutral. Todos los sujetos aseguraron sentirse estresados durante el estímulo más notable, el accidente, de hecho, la Tabla 5 muestra los valores medios de la FC obtenidos como resultado mediante el sensor PPG entre todos los participantes en cada etapa experimental, y la Figura 7 muestra la VFC a lo largo del tiempo en función de cada una de las etapas de presentación de estímulos. En ellas se observa claramente que el estímulo provoca una respuesta fisiológica en los participantes. Sin embargo, otras escenas vividas durante la simulación también les causaron estrés y fueron detectadas de forma correcta por el modelo. Algunos participantes señalaron los primeros segundos como estresantes, mientras el conductor del vehículo explica la prueba y cuando el coche aumenta la velocidad notablemente. Otros señalan que la elevación del tono de voz del conductor y los momentos previos al accidente también les ocasionaron estrés. El modelo devuelve una serie temporal con el resultado del estado emocional del usuario, y se comprueba con las etapas informadas por los participantes. El modelo es capaz de marcar como estrés las etapas informadas por el usuario como estrés, y de detectar la clase sin estrés en los momentos en los que el usuario se encuentra en un estado emocional neutro.

	NORMAL 1	CHOQUE	NORMAL 2	VUELCO
PPG	53 lpm	103 lpm	55 lpm	97 lpm

Tabla 5: Frecuencia cardíaca media obtenida en cada etapa experimental.

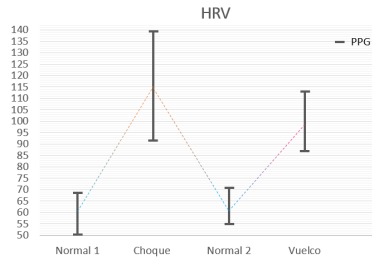


Figura 7: Variación de la FC durante las etapas de presentación de estímulos.

5. Discusión

Tras investigar los métodos de medición de los estados emocionales humanos, se observa que con la tecnología de consumo actual, simplemente capturando información sobre la VFC y aplicando las técnicas de aprendizaje automático, es posible desarrollar un sistema capaz de detectar el estrés de manera fiable, como

muestran los resultados obtenidos en la Sección 4. Los resultados muestran que la FC se correlaciona estrechamente con el nivel de estrés de los ocupantes del vehículo virtual. Las señales fisiológicas capturadas a través de un reloj de uso común puede proporcionar una métrica del estrés del conductor y la posibilidad de monitorizar a las personas en los automóviles, pudiendo recopilar información útil de cómo las diferentes condiciones de la carretera afectan a los conductores. Hay que tener en cuenta que al realizar el experimento sentado se minimizan los riesgos de fallo del sensor PPG por condiciones externas, como los movimientos irregulares, pero en otras condiciones de trabajo los resultados pueden verse alterados. También hay que añadir que muchos valores de los atributos eran nulos, ya que el experimento se realizó en un corto intervalo de tiempo. Además, la VFC puede variar en función de los sujetos, ya que no es equiparable para cualquier individuo, sino que dependen factores como la edad, el sexo, el estado de salud, el consumo, etc. y que cada persona es subjetiva o responde de forma diferente ante estímulos similares. Aún así, la VFC puede convertirse en un método no invasivo valioso para la evaluación diaria del estado de salud de las personas. Pese a que NB suele dar buenos resultados en la clasificación, en este caso ha sido el algoritmo que peor ha realizado la clasificación de las clases. Lo mismo ocurre con la SVM que a pesar de los buenos resultados obtenidos en otros trabajos con objetivos similares, la precisión alcanzada en nuestro caso es del 63.98 %. Por el contrario, tanto RF como DT clasificaron con una precisión superior al 84 % las clases de estrés. El mejor resultado se obtuvo con el DT, con una precisión del 90.34 %, lo que demuestra que el sistema es capaz de detectar el estrés, incluso cuando se prueba el modelo con los datos recopilados a través del sensor PPG de un reloj durante la simulación. El conjunto de datos demuestra ser válido para clasificar la emoción del usuario en función de los datos recogidos por los sensores, pese a que en el trabajo realizado se han manejado pocas características de entrada. Los resultados obtenidos evidencian también que los estímulos inducidos mediante la tecnología de VR provocan respuestas fisiológicas humanas.

La tabla 6 muestra una comparativa de este estudio con otros trabajos con objetivos similares. Se puede observar la gran precisión en la clasificación permitiendo una monitorización diaria, con libertad de movimientos, a bajo coste, con estímulos inmersivos y que pueden ser cambiados en cualquier momento y sin requerir una instalación compleja.

	Monitorización diaria	Bajo coste	Estímulo inmersivo	Libertad de movimiento	Instalación fácil	Sin laboratorio	Precisión >85 %
Trabajo Propuesto	V	V	V	V	V	V	V
[39]	V	V	X	V	V	V	X
[1]	V	V	X	X	V	X	V
[35]	X	X	X	X	X	X	V
[40]	V	V	X	V	V	V	X
[33]	X	V	V	X	X	X	X

Tabla 6: Comparativa de trabajos.

Aún así, como este trabajo es una investigación preliminar, los resultados y la discusión del sistema clasificador son muy limitados, ya que se han hecho sobre un pequeño número de individuos, y solo se ha analizado una emoción en dos situaciones extremas.

6. Conclusiones y líneas de trabajo futuras

Las tareas abordadas durante el presente trabajo de investigación han permitido alcanzar los objetivos principales planteados en el proyecto.

- Se realiza una revisión exhaustiva del estado del arte.
- Se lleva a cabo un estudio y un análisis de los distintos métodos de clasificación basados en técnicas de ML, y de las métricas de evaluación de los modelos.
- Se entrenan varios modelos con nuestro conjunto de datos.
- Se desarrolla un software de VR para inducir estados de estrés al usuario.
- Se utiliza un reloj inteligente para la captura y adquisición de las señales fisiológicas.
- Se realiza una aplicación para visualizar los resultados en un dispositivo móvil.
- Se diseña y se realiza el experimento para validar el clasificador.
- Se lleva a cabo un análisis de los resultados obtenidos.

Como trabajo futuro, se pueden recopilar datos de más sensores no invasivos para poder registrar más información y poder detectar más estados emocionales, lo que puede ser muy beneficioso en muchos ámbitos de la vida diaria. Siguiendo la línea del trabajo, se podría potenciar la inducción de estímulos cambiando la silla de oficina por una plataforma mecánica que acompañe y dé más realismo a las acciones realizadas por el jugador en la aplicación de VR. Se realizaría una conexión entre la aplicación de VR y el autómata programable de la plataforma, y se podrían reproducir las sensaciones físicas que puedan estar asociadas al entorno. También se puede considerar realizar un sistema que adquiera directamente las señales del usuario mientras realiza las acciones cotidianas, de tal manera que las características sin procesar puedan ser entradas para el modelo reconecedor de emociones, es decir, pasar a formar parte directamente del

conjunto de datos de entrenamiento, y visualizar los resultados directamente en el dispositivo móvil. De esta manera, se podría analizar la emoción en múltiples situaciones. Por otro lado, también sería interesante realizar la experimentación con un mayor número de individuos, de manera que no se limiten tanto los resultados del sistema clasificador.

En proyectos más ambiciosos puede ser interesante realizar recomendaciones a los usuarios en función de sus estados anímicos, ya que hay correlación entre algunos estados afectivos y algunos lugares, o personas, por ejemplo. Aunque se debe de considerar que el sistema está desarrollado para ejecutarse en un sistema móvil, por lo que el modelo no debe de tener demasiado coste computacional.

HMD	Head-Mounted Display
VR	Realidad Virtual
FC	Frecuencia Cardíaca
VFC	Variabilidad de la Frecuencia Cardíaca
PPG	Fotoplestismografía
ML	Machine Learning
KNN	K-Nearest Neighbor
NB	Naïve Bayes
DT	Decision Tree
MP	Multilayer Perceptron
RF	Random Forest
IA	Inteligencia Artificial
SNA	Sistema Nervioso Autónomo
GSR	Respuesta Galvánica de la Piel
LF	Baja frecuencia
HF	Alta frecuencia
ULF	Frecuencia ultrabaja
VLF	Muy baja frecuencia

Referencias

- [1] Fatma Patlar Akbulut, Baris Ikitimur y Aydin Akan. “Wearable sensor-based evaluation of psychosocial stress in patients with metabolic syndrome”. En: *Artificial Intelligence in Medicine* 104 (2020), pág. 101824. ISSN: 0933-3657. DOI: <https://doi.org/10.1016/j.artmed.2020.101824>. URL: <http://www.sciencedirect.com/science/article/pii/S0933365719311777>.
- [2] Mouhannad Ali y col. “Emotion Recognition Involving Physiological and Speech Signals: A Comprehensive Review”. En: ene. de 2018, págs. 287-302. ISBN: 978-3-319-58995-4. DOI: 10.1007/978-3-319-58996-1_13.
- [3] Andreia Artífice y col. “Student’s Attention Improvement Supported by Physiological Measurements Analysis”. En: *Technological Innovation for Smart Systems*. Ed. por Luis M. Camarinha-Matos, Mafalda Parreira-Rocha y Javaneh Ramezani. Cham: Springer International Publishing, 2017, págs. 93-102.
- [4] J. Birjandtalab y col. “A Non EEG Biosignals Dataset for Assessment and Visualization of Neurological Status”. En: *2016 IEEE International Workshop on Signal Processing Systems (SiPS)*. 2016, págs. 110-114.
- [5] Karel A Brookhuis y Dick De Waard. “Monitoring drivers’ mental workload in driving simulators using physiological measures”. En: *Accident Analysis & Prevention* 42.3 (2010), págs. 898-903.

- [6] Davide Carneiro y Paulo Novais. "Quantifying the effects of external factors on individual performance". En: *Future Generation Computer Systems* 66 (2017), págs. 171-186. ISSN: 0167-739X. DOI: <https://doi.org/10.1016/j.future.2016.05.019>. URL: <http://www.sciencedirect.com/science/article/pii/S0167739X16301303>.
- [7] Hsiu-Sen Chiang. "ECG-based Mental Stress Assessment Using Fuzzy Computing and Associative Petri Net". En: *Journal of Medical and Biological Engineering* 35 (nov. de 2015). DOI: 10.1007/s40846-015-0095-7.
- [8] L. Cominelli y col. "A Multimodal Perception Framework for Users Emotional State Assessment in Social Robotics". En: *FUTURE INTERNET* 9 (2017). DOI: 10.3390/fi9030042.
- [9] J.A. Domínguez-Jiménez y col. "A machine learning model for emotion recognition from physiological signals". En: *Biomedical Signal Processing and Control* 55 (2020), pág. 101646. ISSN: 1746-8094. DOI: <https://doi.org/10.1016/j.bspc.2019.101646>. URL: <http://www.sciencedirect.com/science/article/pii/S1746809419302277>.
- [10] Maria Egger, Matthias Ley y Sten Hanke. "Emotion Recognition from Physiological Signal Analysis: A Review". En: *Electronic Notes in Theoretical Computer Science* 343 (2019). The proceedings of AmI, the 2018 European Conference on Ambient Intelligence., págs. 35-55. ISSN: 1571-0661. DOI: <https://doi.org/10.1016/j.entcs.2019.04.009>. URL: <http://www.sciencedirect.com/science/article/pii/S157106611930009X>.
- [11] L. Eudave y M. Valencia. "Physiological response while driving in an immersive virtual environment". En: *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 2017, págs. 145-148.
- [12] Joana F. Pinto, Ana Fred y Hugo Plácido da Silva. "Biosignal-Based Multimodal Emotion Recognition in a Valence-Arousal Affective Framework Applied to Immersive Video Visualization". En: *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2019 (jul. de 2019). DOI: 10.1109/EMBC.2019.8857852.
- [13] Zhenhai Gao y col. "Experimal study of young male drivers' responses to vehicle collision using EMG of lower extremity". En: *Bio-medical materials and engineering* 26.s1 (2015), S563-S573.
- [14] Philip Gouverneur y col. "Classification of Physiological Data for Emotion Recognition". En: *Artificial Intelligence and Soft Computing*. Ed. por Leszek Rutkowski y col. Cham: Springer International Publishing, 2017, págs. 619-627.
- [15] A. Gruenewald y col. "[Regular Paper] Biomedical Data Acquisition and Processing to Recognize Emotions for Affective Learning". En: *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BI-BE)*. Oct. de 2018, págs. 126-132. DOI: 10.1109/BIBE.2018.00031.
- [16] Neska Haouij y col. "AffectiveROAD System and Database to Assess Driver's Attention". En: abr. de 2018.

- [17] S. Hassani y col. "Physiological signal-based emotion recognition system". En: *2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS)*. Nov. de 2017, págs. 1-5. DOI: 10.1109/ICETAS.2017.8277912.
- [18] J. A. Healey y R. W. Picard. "Detecting stress during real-world driving tasks using physiological sensors". En: *IEEE Transactions on Intelligent Transportation Systems* 6.2 (2005), págs. 156-166.
- [19] Karen Hovsepian y col. "cStress: towards a gold standard for continuous stress assessment in the mobile environment". En: *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 2015, págs. 493-504.
- [20] Jia-Lien Hsu y col. "Affective content analysis of music emotion through EEG". En: *Multimedia Systems* 24 (feb. de 2017), págs. 1-16. DOI: 10.1007/s00530-017-0542-0.
- [21] Shinji Kajiwara. "Evaluation of driver's mental workload by facial temperature and electrodermal activity under simulated driving conditions". En: *International journal of Automotive technology* 15.1 (2014), págs. 65-70.
- [22] Elena Lozano-Monasor y col. "Facial expression recognition in ageing adults: from lab to ambient assisted living". En: *Journal of Ambient Intelligence and Humanized Computing* 8 (2017), págs. 567-578.
- [23] I. Mavridou y col. "FACETEQ interface demo for emotion expression in VR". En: *2017 IEEE Virtual Reality (VR)*. Mar. de 2017, págs. 441-442. DOI: 10.1109/VR.2017.7892369.
- [24] V. Montesinos y col. "Multi-Modal Acute Stress Recognition Using Off-the-Shelf Wearable Devices". En: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Jul. de 2019, págs. 2196-2201. DOI: 10.1109/EMBC.2019.8857130.
- [25] Sang-Hun Nam, Ji Yong Lee y Jung-Yoon Kim. "Biological-Signal-Based User-Interface System for Virtual-Reality Applications for Healthcare". En: *J. Sensors* 2018 (2018), 9054758:1-9054758:10.
- [26] Abdoul Matine Ousmane, Tahirou Djara, Antoine Vianou y col. "Automatic recognition system of emotions expressed through the face using machine learning: Application to police interrogation simulation". En: *2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART)*. IEEE. 2019, págs. 1-4. DOI: 10.1109/BIOSMART.2019.8734245.
- [27] Yassine Rabhi, Makrem Mrabet y Farhat Fnaiech. "A facial expression controlled wheelchair for people with disabilities". En: *Computer Methods and Programs in Biomedicine* 165 (2018), págs. 89-105. ISSN: 0169-2607. DOI: <https://doi.org/10.1016/j.cmpb.2018.08.013>. URL: <http://www.sciencedirect.com/science/article/pii/S0169260717315456>.
- [28] Jonas Reichenberger y col. "Men Scare Me More: Gender Differences in Social Fear Conditioning in Virtual Reality". En: *Frontiers in Psychology* 10 (2019), pág. 1617. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2019.

01617. URL: <https://www.frontiersin.org/article/10.3389/fpsyg.2019.01617>.
- [29] Hamilton Rivera y col. "Development of a Toolkit for Online Analysis of Facial Emotion". En: *XXVI Brazilian Congress on Biomedical Engineering*. Ed. por Rodrigo Costa-Felix, João Carlos Machado y André Victor Alvarenga. Singapore: Springer Singapore, 2019, págs. 619-625.
 - [30] Stanislaw Saganowski y col. "Emotion Recognition Using Wearables: A Systematic Literature Review Work in progress". En: *ArXiv* abs/1912.10528 (2019).
 - [31] L. Santamaria-Granados y col. "Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)". En: *IEEE Access* 7 (2019), págs. 57-67.
 - [32] Jungryul Seo, Teemu Laine y Kyung-Ah Sohn. "An Exploration of Machine Learning Methods for Robust Boredom Classification Using EEG and GSR Data". En: *Sensors* 19 (oct. de 2019). DOI: 10.3390/s19204561.
 - [33] Lixin Yan, Ping Wan y Dunyao Zhu. "The Induction and Detection Method of Angry Driving: Evidences from EEG and Physiological Signals". En: *Discrete Dynamics in Nature and Society* 2018 (ago. de 2018), págs. 1-16. DOI: 10.1155/2018/3702795.
 - [34] Heekyung Yang, Jongdae Han y Kyungha Min. "A Multi-Column CNN Model for Emotion Recognition from EEG Signals". En: *Sensors*. 2019.
 - [35] Heekyung Yang, Jongdae Han y Kyungha Min. "Distinguishing Emotional Responses to Photographs and Artwork Using a Deep Learning-Based Approach". En: *Sensors (Basel, Switzerland)* 19 (2019).
 - [36] Enrico Zero y col. "Towards real-time monitoring of fear in driving sessions". En: *IFAC-PapersOnLine* 52.19 (2019). 14th IFAC Symposium on Analysis, Design, and Evaluation of Human Machine Systems HMS 2019, págs. 299-304. ISSN: 2405-8963. DOI: <https://doi.org/10.1016/j.ifacol.2019.12.068>. URL: <http://www.sciencedirect.com/science/article/pii/S2405896319318981>.
 - [37] Jianhua Zhang y col. "Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review". En: *Information Fusion* 59 (2020), págs. 103-126. ISSN: 1566-2535. DOI: <https://doi.org/10.1016/j.inffus.2020.01.011>. URL: <http://www.sciencedirect.com/science/article/pii/S1566253519302532>.
 - [38] Kejun Zhang y col. "The PMEmo Dataset for Music Emotion Recognition". En: jun. de 2018, págs. 135-142. DOI: 10.1145/3206025.3206037.
 - [39] Zhan Zhang y col. "Emotion recognition based on customized smart bracelet with built-in accelerometer". En: *PeerJ* 4 (jul. de 2016), e2258. DOI: 10.7717/peerj.2258.
 - [40] B. Zhao y col. "EmotionSense: Emotion Recognition Based on Wearable Wristband". En: *2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. 2018, págs. 346-355.

Industrial Internet of Things, Big Data y Sistemas inteligentes bajo un modelo Edge Computing

V. Gabriel^{1,2}, Juan M. Corchado^{1,2}, and Ana Belén Gil^{1,2}

¹ Máster de Sistemas Inteligentes, Universidad de Salamanca

² BISITE Research Group, Calle Espejo,s/n,37007 Salamanca, Spain
{JorgeVicenteGabriel}@usal.es

Resumen El mantenimiento predictivo es un campo de investigación que ha surgido de la necesidad de mejorar los sistemas instalados. Esa tarea se centra en el control de la degradación de módulos solares para mejorar la fiabilidad, el funcionamiento y el rendimiento a través de la monitorización. En este trabajo se implementó un sistema capaz de predecir la potencia de salida de un modulo solar haciendo uso de técnicas de aprendizaje automático. Aparte, se aplicaron unos márgenes de confianza para poder detectar posibles degradaciones en placas solares bajo un conjunto de datos previamente tratados. Por otro lado, el algoritmo de aprendizaje automático fue evaluado con diferentes métricas para optimizar y encontrar la mejor configuración que englobase el problema. Se añadió una propuesta arquitectónica basada en *Edge Computing* para aplicar el sistema propuesto, que ofrece grandes ventajas como la monitorización individual de cada una de las placas, la optimización de la respuesta del sistema y la velocidad de comunicación con el *Cloud*.

Keywords: Aprendizaje automático, mantenimiento predictivo, redes LSTM, edge computing, paneles solares

1. Introducción

En la actualidad nos encontramos en pleno auge de la industria 4.0, conocida como la “cuarta revolución industrial”. Se basa principalmente en el uso de sensores IoT (*Internet of Things*) en instalaciones que están interconectados, recogiendo un gran volumen de datos para convertirlo en un *Big Data* y procesarlo con técnicas de aprendizaje automático [guerrero4_0].

En este artículo nos vamos a centrar en el mantenimiento predictivo en placas solares. Estamos en una época de cambio donde la energía a partir de los combustibles fósiles ha pasado a un segundo plano frente a las energías renovables como la energía solar. Leaman recoge en su estudio [12] los beneficios de la energía solar mostrando un análisis del impacto y el crecimiento en EEUU en

los últimos años. Tan solo en 2014 la industria de la energía solar creó más puestos de trabajo que juntando las grandes compañías tecnológicas como Apple, Google, Facebook o Twitter. Además, ha llegado en un momento crucial de la economía energética donde el encarecimiento de los precios de la electricidad no cesan. Hacer operativo este tipo de energía, es uno de los motivos por lo que se busca un abaratamiento de los costes en el mantenimiento de estas instalaciones.

El artículo se estructura de la siguiente manera: la sección 5 hace una revisión del estado del arte sobre el mantenimiento predictivo en placas solares. A continuación, en la sección 5 se presenta la propuesta para la monitorización de placas solares y detección de fallos. La sección 5 presenta los resultados obtenidos del caso de estudio. Por último, en la sección 5 se presentan las conclusiones y una posible propuesta de trabajo futuro.

2. Análisis del estado del arte

Entre los modelos que han logrado la detección de fallos en módulos fotovoltaicos, se encuentra el uso de arquitecturas con modelos predictivos ayudados en gran parte por la monitorización de sistemas en tiempo real y la utilización de heurísticas matemáticas para detectar comportamientos anómalos. Existen arquitecturas como el *Edge Computing* [19] [17] que mejora la comunicación con la nube y libera carga de cómputo haciendo los cálculos directamente en el nodo. Es una técnica que tiene muchos beneficios y que verá su estandarización dentro de poco tiempo.

Dentro de los algoritmos utilizados cabe destacar que la mayoría son soluciones específicas a un problema dado utilizando modelos probabilísticos [22] [16] [21]. Al ser soluciones específicas quedan en un segundo plano, ya que no ofrecen una solución que se pueda ajustar a la casuística. Otros algoritmos a tener en cuenta son las redes neuronales, muy utilizadas en los últimos años y en concreto aquellas que son convolucionales [5] [1] [23]. Aparte, otro tipo de algoritmos a tener en cuenta son las máquinas de vectores de soporte (MVS) [18] [7] [24]. Por otro lado, dentro del reconocimiento de imágenes [14] [3] [8] [2] los trabajos se dividen en dos vertientes. Una se centra en el reconocimiento de nubes para predecir la potencia de salida de módulo fotovoltaico y la otra en el reconocimiento de puntos de calor o roturas en las placas a través de imágenes captadas por un dron o similar. Pero quedan descartados ya que se necesitan recursos muy específicos y que quedan fuera del alcance de este estudio.

Los trabajos citados a continuación se consideran de gran interés para el estudio ya que reúnen muchos puntos en común con el enfoque del presente trabajo [15] [5] [7] [17] [20] [1] [10] [13]. En la tabla 1, se puede observar un resumen de las metodologías adoptadas por los artículos mencionados.

Dentro de las metodologías estudiadas, se encuentra la predicción de potencia fotovoltaica de módulos solares. En el artículo [15] construyen un algoritmo

Tabla 1: Metodologías adoptadas por los artículos seleccionados

Ref	Año	Algoritmo	Edge Comp	Predicción	D.Fallos	Monitorizar
[17]	2019	Prob Model	Sí	Sí	Sí	Sí
[22]	2019	Prob Model	No	Sí	Sí	Sí
[16]	2016	Prob Model	No	Sí	No	Sí
[21]	2016	Prob Model	No	Sí	Sí	Sí
[5]	2020	CNN	No	Sí	No	No
[1]	2019	CNN	No	Sí	Sí	Sí
[23]	2018	CNN	No	Sí	No	No
[18]	2020	ANN, Fuzzy logic, Otros, SVM	No	Sí	Sí	Sí
[7]	2019	SVM	No	Sí	No	No
[24]	2019	SVM	No	Sí	No	No
[3]	2019	Img Rec	No	Sí	Sí	No
[8]	2019	Img Rec	No	Sí	Sí	No
[2]	2018	Img Rec	No	Sí	Sí	Sí
[15]	2019	ANN	No	Sí	No	No
[20]	2019	ANN	No	Sí	Sí	Sí
[10]	2017	Prob Model	No	Sí	Sí	No
[13]	2019	Fuzzy logic	No	Sí	Sí	Sí

predictivo que es capaz de predecir la potencia de salida de un panel a través de predicciones de datos meteorológicos. Como resultado obtuvieron la mejor configuración de una red neuronal “*Long Short, Term Memory*” (LSTM) a través de métricas estadísticas como el error cuadrático medio, error absoluto medio, etc. Además de hacer predicciones 24 horas a futuro. Otro estudio muy parecido es el que propone el artículo [1], pero en este caso, el predictor se basó en una red convolucional. Obtuvieron un sistema para predecir picos de consumo en comunidades de viviendas a través de predicciones eólicas y predicciones solares. Al comparar el error medio de la raíz cuadrada de las predicciones solares obtuvieron un rendimiento superior del 7,6 % respecto otros estudios. Pero estos trabajos citados no aplican ningún método de predicción de fallos.

Por otro lado, otra de las metodologías estudiadas son los análisis de fallos como propone el artículo [10]. Para realizarlo analiza la ventana de tiempo de la potencia de salida de los módulos fotovoltaicos. Esta curva temporal la clasifican en un correcto funcionamiento o un fallo en concreto. Como resultado, al tener N paneles, el peor resultado que podrían obtener es de $2N - 1$ fallos. Un punto fuerte es la capacidad que tuvieron de detectar fallos en partes concretas del conjunto de módulos solares. Otro enfoque a tener en cuenta es la monitorización de placas solares a través de una red neuronal como realizó el artículo [20]. En este estudio obtuvo un predictor de potencia con una red cuya precisión no fue inferior al 98 % y una diferencia inferior al 0,02 entre el valor real y el valor predicho. Esta diferencia les permitió detectar degradaciones en los paneles solares y su mal funcionamiento.

Por último, los anteriores estudios no aplican sus investigaciones en una arquitectura descentralizada. Sin embargo, el artículo [17] propuso una arquitectura donde se monitorizan diferentes nodos IoT. Como resultado obtuvo un análisis del rendimiento de los paneles solares a lo largo del año, con los picos de producción de potencia más altos que fueron en septiembre y octubre. También observaron los picos de menor producción que fueron mayo y marzo. El mayor problema de esta arquitectura es que no implementa ningún algoritmo de aprendizaje automático para mejorar el rendimiento de la estación. Pero da un caso de estudio de como desplegar una arquitectura basada en IoT donde se puede hacer uso del *Edge Computing* para reducir los tiempos de latencia y reducir el trafico de datos en las comunicaciones.

3. Metodología

La propuesta metodológica del presente trabajo para el mantenimiento predictivo consiste en la elaboración de una sistema capaz de resolver un problema de series temporales. Los problemas de series temporales consisten en el estudio a lo largo del tiempo de como observaciones o patrones que se repiten. Así se determinan valores en un futuro próximo y posibilitan el realizar una toma decisiones en consecuencia. Se basa en los problemas de regresión pero tienen algunas diferencias como:

- Las variables dependen del tiempo por lo que no cumplen que todas las observaciones sean independientes.
- Muestra tendencias de estacionalidad, crecimiento o decrecimiento.

Se plantea un sistema capaz de obtener la predicción meteorológica de la zona y con ella obtener qué potencia de salida van a tener los módulos solares. Para ello se hace uso de las redes LSTM que se explicaran en la siguiente sección 5. Por último, en tiempo real se calcula una zona de confianza capaz de determinar degradaciones en los sistemas fotovoltaicos descrita en la sección 5.

3.1 Definición de las redes LSTM

Las redes “Long Short-Term Memory” (LSTM) o redes de memoria larga a corto plazo [6] son una adaptación de las redes neuronales recurrentes, en concreto el cambio se centra en la capa oculta de la red. Esta capa oculta se denomina “celda LSTM” constituida por tres puertas, una de entrada, una de salida y otra de olvido que controlan el flujo de datos de la red. En la Fig. 1 se puede ver el esquema general de la celda que constituye las redes LSTM. Cuando se toma un instante de tiempo t con una entrada X_t , la salida de la capa oculta H_t , su salida anterior H_{t-1} , se obtiene que el estado de entrada es $\tilde{C}_t$, el estado de salida es C_t y el estado anterior C_{t-1} . Aparte, los estado de las puertas toman los valores de i_t , f_t y o_t . Se debe observar a partir del esquema que los valores C_t y H_t se van propagando por la red. Para obtener en concreto estos dos valores se debe de seguir una serie de ecuaciones.

En primer lugar, se debe de calcular los estados de las tres puertas y el estado de la entrada de la celda:

$$i_t = \sigma(W_1^i \cdot x_t + W_h^i \cdot h_{t-1} + b_i) \quad \text{Puerta de entrada} \quad (1)$$

$$f_t = \sigma(W_1^f \cdot x_t + W_h^f \cdot h_{t-1} + b_f) \quad \text{Puerta de olvido} \quad (2)$$

$$o_t = \sigma(W_1^o \cdot x_t + W_h^o \cdot h_{t-1} + b_o) \quad \text{Puerta de salida} \quad (3)$$

$$\tilde{C}_t = \tanh(W_1^C \cdot x_t + W_h^C \cdot h_{t-1} + b_C) \quad \text{Entrada de la celda} \quad (4)$$

donde $W_1^i, W_1^f, W_1^o, W_1^C$ son las matrices de los pesos que conectan X_t a las tres puertas y la entrada de la celda, $W_h^i, W_h^f, W_h^o, W_h^C$ son las matrices de peso que conectan h_{t-1} a las tres puertas y a la entrada de la celda, b_i, b_f, b_o, b_C son los términos de sesgo de las tres puertas y la entrada de la celda, θ representa la función sigmoidea $\frac{1}{1+\exp(-x)}$ y $\tanh$ representa la función tangente hiperbólica $\frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}$. En segundo lugar, se calcula el estado de salida de la celda:

$$C_t = i_t * \tilde{C}_t + f_t * C_{t-1} \quad (5)$$

donde $i_t, f_t, \tilde{C}_t, C_t, C_{t-1}$ tienen la misma dimensión. En tercer lugar se calcula la salida de la capa oculta:

$$h_t = o_t * \tanh(C_t) \quad (6)$$

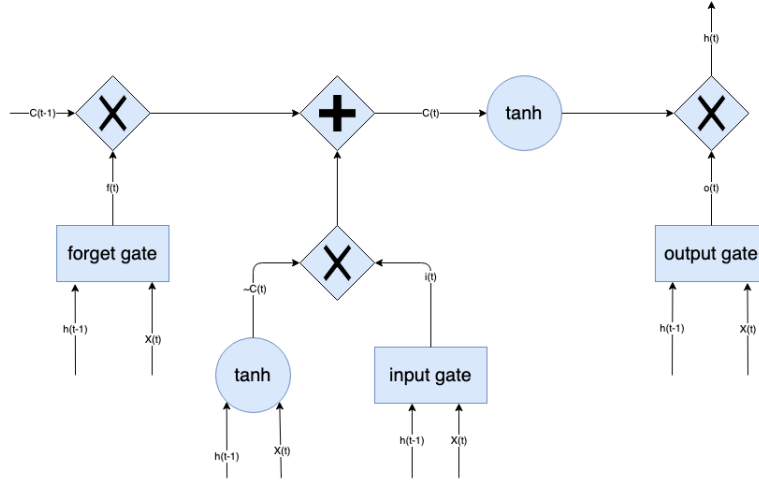


Figura 1: Esquema de una celda LSTM

La salida de la celda se define como:

$$\tilde{x}_{t+1} = W_2 \cdot h_t + b \quad (7)$$

donde W_2 es la matriz de pesos entre la capa de entrada y la capa de salida, b es el sesgo de la capa de salida.

Se debe de tener en cuenta el tamaño de la ventana ya que al no tener un histórico infinito limita la ventana de la red. Estos datos históricos recorrerán toda la red cambiando sus estados para obtener la predicción buscada.

3.2 Descripción del Sistema

A continuación, se define la configuración de la red que se utilizó en el estudio para desarrollar el predictor de potencia. El diseño de la red neuronal se basó en ensayo y error hasta encontrar la mejor configuración que englobase al problema. La estructura que se seleccionó como resultado final fue:

- En la primera capa se especifica el número de tiempos que se van a pasar y el número de parámetros que lo forma. En el diseño se especificó que recibiría 120 instantes de tiempo y tres parámetros de entrada.
- Se definen dos capas ocultas con 24 y 48 neuronas cuya función de activación es “relu”. Esta función de activación es la más utilizada en la redes recurrentes, de ahí su elección.
- Como salida se usó la capa densa con 72 neuronas. Para poder obtener 72 predicciones que formarían la ventana a futuro que se quiere obtener.

Al basarse en un problema de regresión la función de pérdida que se utilizó fue el “Error Cuadrático Medio” que mide el promedio de errores al cuadrado.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - Y_i)^2 \quad (8)$$

Por un lado, se definió el algoritmo de descenso de gradiente. Se utilizó el algoritmo “Adam” [11] caracterizado por sus tasas de aprendizaje notablemente decrecientes. Por otro lado, se especificó para el entrenamiento que el número de épocas fuese 50 y el tamaño de los lotes fuese 120.

Una vez entrenada la red se obtiene como resultado una ventana de predicciones. El objetivo con estos resultados es analizar las posibles anomalías que pueda tener el módulo fotovoltaico. Se compara la potencia real que produce la placa con la potencia de predicción realizada con la red neuronal. Se calculan unos márgenes de confianza como los utilizados en el artículo [20], una desviación del 10 % del valor real. Surgen los márgenes de confianza que están comprendidos entre un límite superior que es el 10 % más que el valor real y el límite inferior que es un 10 % menos que el valor real. Si el valor predicho de la red está fuera de esos márgenes, se considera que el módulo puede haber entrado en un proceso de degradación. No significa que el módulo esté roto sino que tiene un comportamiento anómalo que en un futuro puede convertirse en un fallo más grave.

$$P_{red} > P_{real} * 1,1 \quad \text{Sobrepasa el límite superior} \quad (9)$$

$$P_{red} < P_{real} * 0,9 \quad \text{Por debajo del límite inferior} \quad (10)$$

siendo P_{red} el valor que se obtiene como resultado de la red neuronal y P_{real} el valor real de potencia que tiene la placa fotovoltaica.

4. Resultados

A continuación, se introduce el caso de estudio que se utilizó tanto para probar el funcionamiento de la red neuronal como para evaluar su rendimiento.

4.1 Caso de estudio

Para el caso de estudio se utilizan los datos que ofrece el centro solar DKC [9], el proyecto Alice Sprint ubicado en Australia. La estación seleccionada es la numero 25 [4] que fue instalada en el 2016. Pertenece a la marca “Hanwha Solar” con tecnología de silicio policristalino. En la tabla 2 se puede observar las principales características.

Tabla 2: Información de las características principales del modulo solar

Nombre	Valor
Array Rating	5.83kW
Panel Rating	265W
Number Of Panels	22
Panel Type	HSL 60S
Array Area	36.74
Inverter Size / Type	SMA SMC 6000A
Installation Completed	Sat, 2 Jul 2016
Array Tilt/Azimuth	Tilt=20, Azi=0 (Solar North)

El conjunto de datos está constituido por mediciones cada 5 minutos desde julio del 2016 hasta febrero del 2020. Estos datos están grabados en un excel cuyo acceso es público desde la página de la organización. El fichero está constituido por doce columnas pero se extrajeron las seis columnas más importantes teniendo como referencia el artículo [20] :

- La variable “Timestamp” se utiliza para agrupar el conjunto de datos ya que representa el instante de tiempo cuando se tomó el dato.
- La variable “Rainfall” representa la precipitación media medida en mililitros.
- La variable “ActPower” representa la potencia de activación del módulo solar, medida en kilovatios y es la variable objetivo que se predice.
- La variable “Temperature” representa la temperatura que se registra en la zona medida en Celsius.
- La variable “GlobalRadiation” representa la radiación directa del sol sobre la placa, está medida en W/m^2 .
- La variable “DiffuseRadation” representa la radiación difusa, es decir, la radiación que no incide directamente en la placa. Se mide en W/m^2 .

En la Fig. 2 se puede ver una representación de los datos procesados a lo largo del tiempo. Además, se puede apreciar cómo los valores de la radiación son

semejantes al de la producción de potencia de la planta, por lo que se consideró que es una variable muy importante para el modelo. Aparte, la variable que representa la precipitación se eliminó ya que no aporta valor para el modelo. La estación se encuentra en Australia por lo que tiene un clima muy cálido, llegando a ser desértico en la mayoría de su territorio. Si la localización de la planta tuviese otro clima, se podría valorar la importancia que tendría en el modelo.

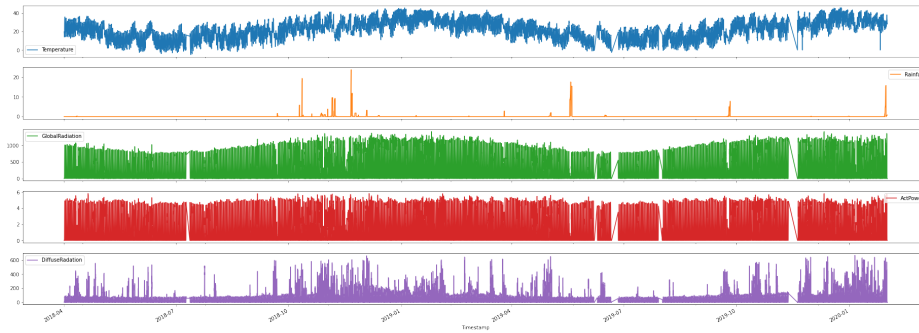


Figura 2: Selección de variables procesadas frente al tiempo

4.2 Evaluación del modelo

Se diseñó un entorno en el que se introdujo una ventana de tiempo que estaba compuesta por un histórico de cinco días de mediciones y un día de previsión futura. Se obtuvo como resultado la potencia de salida que tendría la placa solar. Este resultado se comparó con el valor real que proporciona el conjunto de datos. La comparación se puede ver tanto en la Fig. 3 como en la Fig. 4.

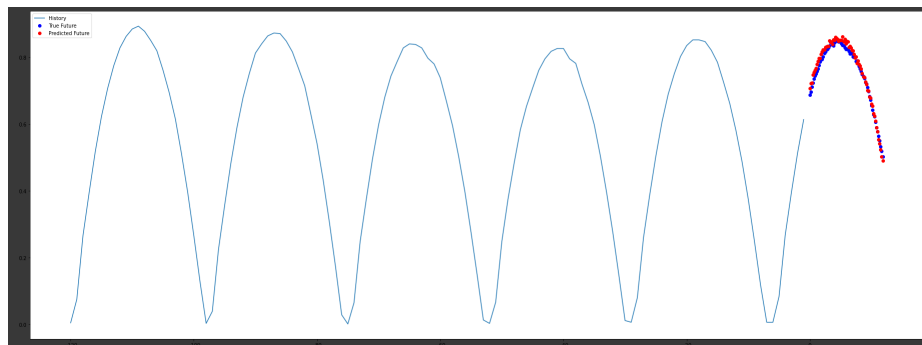


Figura 3: Potencia de salida del modulo solar de la primera prueba.

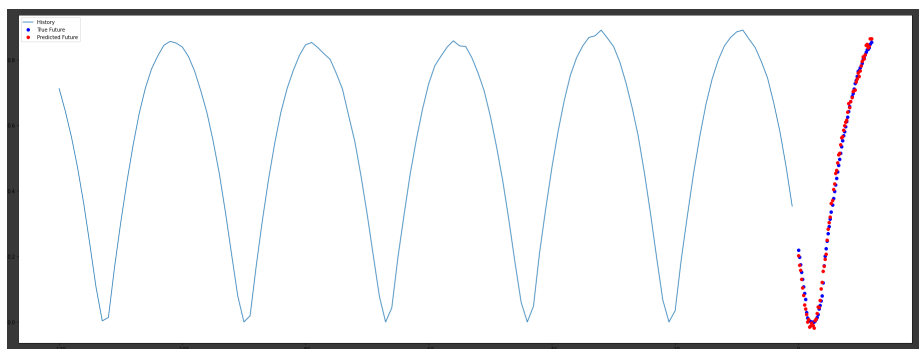


Figura 4: Potencia de salida del modulo solar de la segunda prueba.

Los puntos rojos representan las predicciones que ha hecho la red neuronal sobre los datos de entrada. La línea azul continua representa la potencia que ha tenido en el pasado en la ventana de tiempo. Y por último, los puntos azules son los valores reales que tiene el modulo solar en ese instante de tiempo.

Por otro lado, se aplicaron márgenes de confianza detallados en la descripción del sistema. Estos límites se representan gráficamente para poder ver que las predicciones entran dentro de los valores normales. En la Fig. 5 y en la Fig. 6 se puede ver el resultado que se obtuvieron.

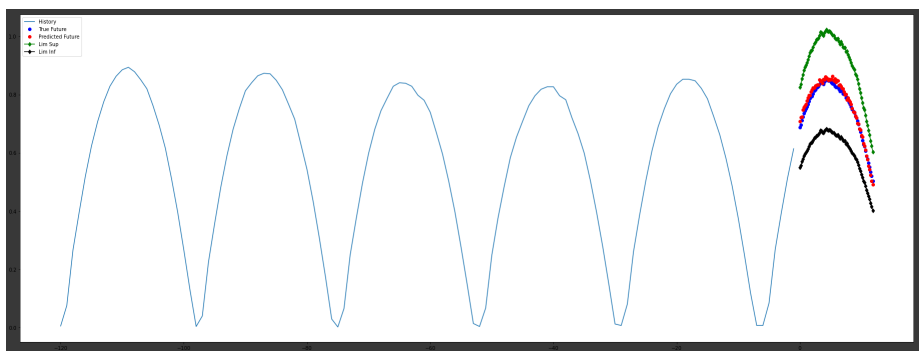


Figura 5: Potencia de salida del modulo solar de la tercera prueba.

El límite inferior se representa con la línea negra y el límite superior se representa con la línea verde. Se observa que las predicciones entran dentro de los límites calculados como una zona de rendimiento óptimo de la placa.

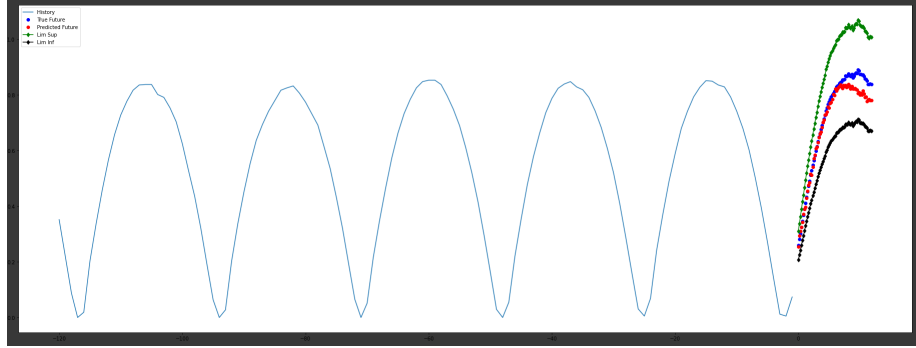


Figura 6: Potencia de salida del modulo solar de la cuarta prueba.

4.3 Evaluación del sistema

A la hora de evaluar el sistema se utilizaron métricas como el “error absoluto medio” (MAE), el “error cuadrático medio” (MSE) y la “raíz del error cuadrático medio” (RMSE) ya que se basa en un problema de regresión:

- El error absoluto medio calcula el promedio de los errores del conjunto de predicciones. Esta métrica tiene en cuenta todas las diferencias individuales por el mismo peso. Además, su valor es absoluto, así elimina los signos en los errores.

$$MAE = \frac{1}{n} \sum_{j=1}^n |Y_i - \tilde{Y}_i| \quad (11)$$

- El error cuadrático medio mide el promedio de errores al cuadrado del valor real del estimado. Esta métrica se utilizó como el valor de pérdidas en la configuración de la red neuronal.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - Y_i)^2 \quad (12)$$

- La raíz del error cuadrático medio es la regla de puntuación cuadrática que mide el promedio de errores. Calcula el valor a través de la raíz cuadrada del promedio de las diferencias al cuadrado entre las predicciones y las observaciones reales.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (Y_i - \tilde{Y}_i)^2} \quad (13)$$

Cuando se realizaron las pruebas se utilizó un conjunto de validación de 735 instancias. Se probaron diversas configuraciones añadiendo un número determinado de neuronas en diversas capas de la red. Además, se hicieron dos conjuntos de pruebas con diferentes algoritmos de optimización. El primer algoritmo de

optimización en el que se realizaron las pruebas se llama “Root Mean Square Propagation” (RMSprop). El algoritmo “RMSprop” intenta buscar una menor oscilación en la bajada del gradiente, eligiendo una única tasa de aprendizaje para cada parámetro en específico. En la tabla 3 se puede ver el resultado que se obtuvo.

Tabla 3: Pruebas con el algoritmo RMSprop

Nombre	N° Neuronas	MAE	MSE	RMSE
Prueba 0	Capa1: 6	0,0899	0,0178	0,1334
Prueba 1	Capa1: 12	0,0904	0,0196	0,1483
Prueba 3	Capa1: 24	0,0821	0,0165	0,1284
Prueba 4	Capa1: 48	0,0731	0,0157	0,1254
Prueba 5	Capa1: 72	0,0744	0,0162	0,1273
Prueba 6	Capa1: 6 Capa2: 3	0,1160	0,0253	0,1590
Prueba 7	Capa1: 12 Capa2: 6	0,0751	0,0154	0,1241
Prueba 8	Capa1: 48 Capa2: 12	0,0730	0,0168	0,1298
Prueba 9	Capa1: 48 Capa2: 24	0,0749	0,0170	0,1304
Prueba 10	Capa1: 12 Capa2: 48	0,0745	0,0163	0,1275
Prueba 11	Capa1: 24 Capa2: 48	0,0740	0,0169	0,1298
Prueba 12	Capa1: 12 Capa2: 6 Capa3: 3	0,0867	0,0173	0,1315
Prueba 13	Capa1: 48 Capa2: 12 Capa3: 6	0,0850	0,0186	0,1363

En cada una de las pruebas se fue aumentando el número de neuronas en la capa oculta de la red para observar si los resultados mejoraban o no. Se llegó a probar hasta con 72 neuronas donde se observó que el resultado obtenido no mejoraba con la prueba anterior. Entonces, se probó a aumentar el número de capas. Se iba aumentando el número de neuronas en la primera capa y disminuyendo en la segunda y viceversa. Pero partiendo de los mejores resultados de las anteriores pruebas como la 0, la 1 y la 3. Las mejores configuraciones que se obtuvieron fueron en la prueba 7 y en la prueba 8. Se aumentó el número de capas ocultas a 3, pero no se obtuvo mejora en el resultado. Finalmente se obtuvo que la configuración más óptima sería con dos capas teniendo 12 neuronas en la primera capa y 6 neuronas en la segunda capa. Se seleccionó la prueba 7 aunque tuviese un MAE superior a la prueba 8, ya que en el RMSE tiene un valor considerablemente menor.

Por otra parte, el segundo algoritmo con el que se realizaron las pruebas fue con el algoritmo “Adam”. Este algoritmo se basa en otros algoritmos como el “RMSprop”, “Adadelta” y “Momentum”. Se utilizó la misma configuración de neuronas que en las pruebas anteriores para poder comparar el mejor resultado. La tabla 4, muestra los resultados obtenidos.

Como resultado se obtuvo que la prueba 11 tuvo el mejor resultado incluso si lo comparamos con los resultados de la anterior prueba. El diseño final de la red se basó en dos capas ocultas, la primera con 24 neuronas y la segunda con 48 neuronas. Además de utilizar el algoritmo “Adam” como algoritmo de

Tabla 4: Pruebas con el algoritmo Adam

Nombre	Nº Neuronas	MAE	MSE	RMSE
Prueba 0	Capa1: 6	0,1035	0,0220	0,1483
Prueba 1	Capa1: 12	0,0738	0,0155	0,1245
Prueba 3	Capa1: 24	0,0847	0,0189	0,1374
Prueba 4	Capa1: 48	0,0704	0,0159	0,1261
Prueba 5	Capa1: 72	0,0708	0,0159	0,1262
Prueba 6	Capa1: 6 Capa2: 3	0,2634	0,0875	0,2958
Prueba 7	Capa1: 12 Capa2: 6	0,0757	0,0154	0,1240
Prueba 8	Capa1: 48 Capa2: 12	0,0716	0,0157	0,1251
Prueba 9	Capa1: 48 Capa2: 24	0,0712	0,0162	0,1272
Prueba 10	Capa1: 12 Capa2: 48	0,0768	0,0165	0,1284
Prueba 11	Capa1: 24 Capa2: 48	0,0684	0,0141	0,1188
Prueba 12	Capa1: 12 Capa2: 6 Capa3: 3	0,0762	0,0158	0,1256
Prueba 13	Capa1: 48 Capa2: 12 Capa3: 6	0,0850	0,0164	0,1280

optimización de gradiente. Por otro lado, se hizo una comparación con un trabajo analizado en el estado del arte. En este artículo [1] propone un modelo basado en redes convolucionales con múltiple cabeza. En la tabla 5 se puede ver el porcentaje de error al comparar las métricas obtenidas con el estudio del artículo.

Tabla 5: Medición de mejora del modelo propuesto con el modelo comparado

Nombre	MAE	RMSE
Modelo Comparado	0,0917	0,1339
Modelo Propuesto	0,0684	0,1188
Mejora %	25,4	11,27

Se seleccionó este artículo ya que utiliza una red neuronal como sistema predictivo y usa las mismas métricas de evaluación que se utilizaron en este trabajo. Como resultado se obtuvo que el modelo que se propone en este trabajo es mejor que el del artículo referenciado. En concreto se consiguió una mejora del 11,27 % en RMSE y en MAE un 25,4 %.

4.4 Propuesta arquitectónica

Para finalizar se plantea una posible aplicación del sistema dentro de una arquitectura basada en el *Edge Computing*, así como la ingesta de datos *Big Data* llegando a mostrar un control de fallos en tiempo real.

La idea se basa en que cada módulo fotovoltaico esté asociado a un nodo (*Edge*) donde se realizan las operaciones del sistema. La respuesta sería única para el módulo asociado y así se ahorraría costes de comunicación y latencia. Aparte de obtener los datos de la placa, el nodo recibe los datos meteorológicos de la zona para poder hacer la predicción. El flujo de datos de comunicación se realizaría a través de comunicaciones MQTT o similares. Estas comunicaciones estarán programadas cada cierto tiempo con la nube para almacenar todos los eventos que suceden. La nube sería la encargada de gestionar todas las operaciones que realiza la arquitectura, a parte de registrar toda la actividad de la planta. A mayores se desarrollaría una plataforma basada en servicios donde se podría monitorizar en tiempo real toda la estación. Aparte, se mostrarían las predicciones que se realizarían en los nodos. Estos nodos a su vez gestionarían una serie de actuadores que gestionan el flujo de corriente de las placas solares en caso de que la red tenga un fallo crítico. En la Fig. 7 se puede ver un esquema de la propuesta.

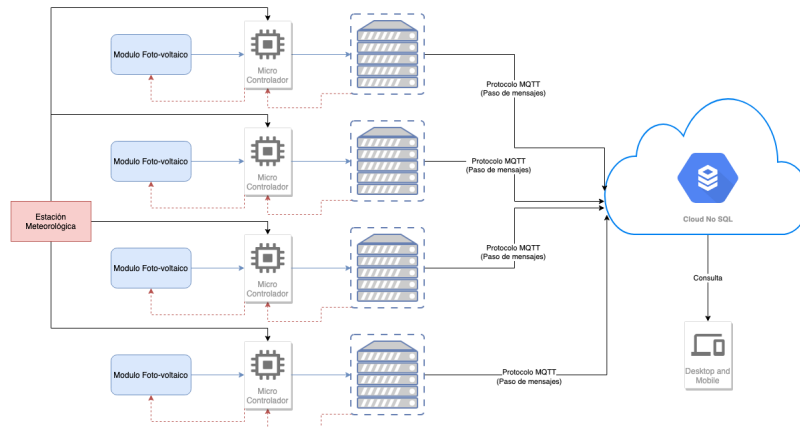


Figura 7: Esquema de componente de la propuesta arquitectónica.

5. Conclusiones

A lo largo del estudio se ha hecho un mapeo sistemático sobre la tecnología de mantenimiento predictivo enfocado principalmente en la energía solar. Se definió una propuesta acorde con ese estudio capaz de detectar degradaciones en los módulos solares.

Se desarrolló una red neuronal capaz de predecir la potencia de salida de placas solares a través de predicciones meteorológicas. Una vez obtenida la ventana de tiempo era comparada con el valor real del módulo. Aplicando unos límites de confianza se puede determinar el posible mal funcionamiento del panel. Como resultado se obtuvo un modelo predictivo con un $MAE = 0,0684$ y un $RMSE = 0,1188$ teniendo un 25,4 % y un 11,27 % mejor rendimiento que con algunos trabajos publicados. Adicionalmente, se propuso el diseño de una arquitectura basada en el *Edge Computing* donde se aplicaría el sistema desarrollado. Esta arquitectura conectaría los dispositivos IoT y sería capaz de almacenar grandes volúmenes de datos en la nube.

Como trabajo futuro se plantea el estudio del rendimiento del sistema con los cambios de estación y los efectos en rendimiento bajo condiciones climáticas adversas. Además, se realizaría la arquitectura propuesta para detectar degradaciones en módulos solares.

Referencias

- [1] M Alhussein, S I Haider y K Aurangzeb. “Microgrid-level energy management approach based on short-term forecasting of wind speed and solar irradiance”. En: *Energies* 12.8 (2019). DOI: 10.3390/en12081487.
- [2] Moath Alsafasfeh y col. “Unsupervised Fault Detection and Analysis for Large Photovoltaic Systems Using Drones and Machine Vision”. En: *ENERGIES* 11.9 (sep. de 2018). DOI: 10.3390/en11092252.
- [3] Mahmoud Dhinish, Peter Mather y Violeta Holmes. “Novel Photovoltaic Hot-Spotting Fault Detection Algorithm”. En: *IEEE TRANSACTIONS ON DEVICE AND MATERIALS RELIABILITY* 19.2 (jun. de 2019), págs. 378-386. ISSN: 1530-4388. DOI: 10.1109/TDMR.2019.2910196.
- [4] DKASC, Alice Springs 25 Hanwha Solar poly-Si Fixed 2016 / DKA Solar Centre. Recuperado el 25 de Junio 2020.
- [5] Na Dong y col. “A novel convolutional neural network framework based solar irradiance prediction method”. En: *INTERNATIONAL JOURNAL OF ELECTRICAL POWER & ENERGY SYSTEMS* 114 (2020). ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2019.105411.
- [6] Yanjie Duan, Yisheng Lv y Fei-Yue Wang. *Travel time prediction with LSTM neural network*. IEEE, 2016.

- [7] M Farhadi y N Mollayi. "Application of the least square support vector machine for point-to-point forecasting of the PV power". En: *International Journal of Electrical and Computer Engineering* 9.4 (2019), págs. 2205-2211. DOI: 10.11591/ijece.v9i4.pp2205-2211.
- [8] Lei Fu y col. "A Regional Photovoltaic Output Prediction Method Based on Hierarchical Clustering and the mRMR Criterion". En: *ENERGIES* 12.20 (oct. de 2019). DOI: 10.3390/en12203817.
- [9] *Home Page / DKA Solar Centre*. Recuperado el 25 de Junio 2020, website: <http://dkasolarcentre.com.au/>.
- [10] Dong Ji y col. "Photovoltaic Array Fault Detection by Automatic Reconfiguration". En: *ENERGIES* 10.5 (mayo de 2017). ISSN: 1996-1073. DOI: 10.3390/en10050699.
- [11] Diederik P Kingma y Jimmy Ba. "Adam: A method for stochastic optimization". En: *arXiv preprint arXiv:1412.6980* (2014).
- [12] C. Leaman. *The benefits of solar energy*. 2015.
- [13] L. Liu e Y. Huai. "Dynamic Hand Gesture Recognition Using LMC for Flower and Plant Interaction". En: *International Journal of Pattern Recognition and Artificial Intelligence* 33.1 (2019). cited By 3. DOI: 10.1142/S0218001419500034. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85052921427%5C&doi=10.1142%5C%2fS0218001419500034%5C&partnerID=40%5C&md5=74f114c7fa532f5a1ea7731bf669099e>.
- [14] L López-Fernández y col. "Automatic evaluation of photovoltaic power stations from high-density RGB-T 3D point clouds". En: *Remote Sensing* 9.6 (2017). DOI: 10.3390/rs9060631.
- [15] N Maitanova y col. "A machine learning approach to low-cost photovoltaic power prediction based on publicly available weather reports". En: *Energies* 13.3 (2020). DOI: 10.3390/en13030735.
- [16] Isabel M Moreno-Garcia y col. "Real-Time Monitoring System for a Utility-Scale Photovoltaic Power Plant". En: *SENSORS* 16.6 (jun. de 2016). ISSN: 1424-8220. DOI: 10.3390/s16060770.
- [17] R I S Pereira y col. "Comparative analysis of photovoltaic modules center and edge temperature using iot embedded system". En: *Renewable Energy and Power Quality Journal* 17 (2019), págs. 198-207. DOI: 10.24084/repqj17.261.
- [18] S Rodrigues y col. "Machine learning photovoltaic string analyzer". En: *Entropy* 22.2 (2020). DOI: 10.3390/e22020205.
- [19] A H Sabry y col. "Wireless monitoring prototype for photovoltaic parameters". En: *Indonesian Journal of Electrical Engineering and Computer Science* 11.1 (2018), págs. 9-17. DOI: 10.11591/ijeecs.v11.i1.pp9-17.
- [20] Sufyan Samara y Emad Natsheh. "Intelligent Real-Time Photovoltaic Panel Monitoring System Using Artificial Neural Networks". En: *IEEE ACCESS* 7 (2019), págs. 50287-50299. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2911250.

- [21] Younes Seyedi, Houshang Karimi y Santiago Grijalva. "Distributed Generation Monitoring for Hierarchical Control Applications in Smart Microgrids". En: *IEEE TRANSACTIONS ON POWER SYSTEMS* 32.3 (mayo de 2017), págs. 2305-2314. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2016.2610322.
- [22] Xingshu Sun, Raghu Vamsi Krishna Chavali y Muhammad Ashraful Alam. "Real-time monitoring and diagnosis of photovoltaic system degradation only using maximum power point-the Suns-Vmp method". En: *PROGRESS IN PHOTOVOLTAICS* 27.1 (2019), págs. 55-66. ISSN: 1062-7995. DOI: 10.1002/pip.3043.
- [23] Fei Wang y col. "Wavelet Decomposition and Convolutional LSTM Networks Based Improved Deep Learning Model for Solar Irradiance Forecasting". En: *APPLIED SCIENCES-BASEL* 8.8 (2018). DOI: 10.3390/app8081286.
- [24] Z M Yasin y col. "Prediction of solar irradiance using grey Wolf optimizer least square support vector machine". En: *Indonesian Journal of Electrical Engineering and Computer Science* 17.1 (2019), págs. 10-17. DOI: 10.11591/ijeecs.v17.i1.pp10-17.

Machine Learning y Puntuación de Riesgo Poligénico para la Enfermedad de Alzheimer

Nibeth Mena Mamani

Departamento de Informática y Automática, Facultad de Ciencias
Plaza de los Caídos s/n, 37008, Salamanca, España
n.mena@usal.es
<http://diaweb.usal.es/>

Resumen Durante los últimos 10 años y después de importantes descubrimientos, como la secuenciación del ADN de todo el genoma humano, ha habido un aumento considerable en el interés en las investigaciones de modelos de predicción de riesgos de enfermedades de origen genético a través de dos enfoques principales: la puntuación de riesgo poligénico y las técnicas de aprendizaje automático. El objetivo de este trabajo se centró en el estudio de la aplicación de estas dos técnicas para la predicción de la enfermedad de Alzheimer. Para ello, inicialmente se realizó un mapeo sistemático de literatura donde se seleccionaron 41 artículos y durante la revisión de los mismos se identificaron diversos métodos de cálculo de la puntuación de riesgo poligénico. A partir de los datos (exoma) de un individuo se encontraron 27 variantes genéticas asociadas a la enfermedad del Alzheimer. Posteriormente, con ayuda de las herramientas PLINK y PRSice2 se calculó la puntuación de riesgo poligénico con distintos modelos obtenidos por medio de estudios asociativos previos (GWAS ¹). La interpretación de los resultados obtenidos se realizó en base a la prevalencia establecida para la enfermedad de Alzheimer. Por otra parte, también se intentó construir un modelo de predicción usando una técnica de aprendizaje automático a partir de datos genéticos. La importancia del presente trabajo, es que el “pipeline” que se ha desarrollado podrá ser utilizado para el análisis del exoma de cualquier individuo, dando como resultado final su puntuación de riesgo poligénico para padecer la enfermedad de Alzheimer.

Keywords: Puntuación de riesgo poligénico, aprendizaje automático, PLINK, PRSice2, predicción, Alzheimer

1. Introducción

Las enfermedades complejas, como el cáncer, las cardiovasculares, respiratorias y los trastornos neurológicos causan enormes problemas de salud pública.

¹Estudios de asociación de genoma completo, son estudios que examinan todas las variaciones existentes en el genoma, estos datos son utilizados para identificar genes que pueden contribuir al riesgo de desarrollar una determinada enfermedad.

La Organización Mundial de la Salud (OMS) a través del informe *Neurological Disorders: Public health challenges* anunció que cientos de millones de personas en todo el mundo sufren trastornos neurológicos, más de 50 millones de personas en todo el mundo padecen demencia, y se estima que cada año se diagnostican 7.7 millones de casos nuevos y mueren 6.8 millones de personas como consecuencia de los trastornos neurológicos [26]. Una de las enfermedades más común de demencia es la enfermedad de Alzheimer (EA), donde según las estadísticas de la Asociación de Alzheimer, la EA representa del 60 % al 80 % de los casos de demencia [1]. La EA es una enfermedad neurodegenerativa eso significa que empeora con el tiempo y ataca progresivamente las neuronas hasta destruirlas, se manifiesta lentamente a través del tiempo, implica la degeneración irreversible del cerebro y gradualmente termina con la insuficiencia cerebral completa, es decir es una pérdida lenta y progresiva de la memoria y otras áreas cognitivas como lenguaje, la orientación del espacio y tiempo, entre otros. Se cree que la EA solo afecta a la población de edad avanzada. Sin embargo, esta enfermedad también puede manifestarse a una edad temprana como a los 30 años debido a que la persona hereda uno de los genes asociado a la EA. En ambos casos existen tratamientos que ayudan a retrasar el proceso de desarrollo de la EA, pero es necesario e importante predecir, diagnosticar y realizar una intervención temprana para proteger al cerebro de sufrir el mayor daño posible.

La mayoría de las enfermedades tienen como origen mutaciones genéticas en el ADN, la EA también es causado por mutaciones genéticas. Para poder predecir estas enfermedades de origen genético existen dos enfoques principales que se manejan hoy en día: la primera es la puntuación de riesgo poligénico - Polygenic Risk Score (PRS), conocido como un enfoque tradicional, que es un dato probabilístico obtenido a través de técnicas estadísticas que predicen la probabilidad de contraer alguna enfermedad genética en función de sus genes, el segundo enfoque es las técnicas de aprendizaje automático -Machine Learning (ML), que a través de un conjunto de algoritmos supervisados y no supervisados predicen la probabilidad de contraer la enfermedad genética.

El objetivo principal de este trabajo se centra en el estudio de estos dos enfoques para la predicción temprana de la EA. Para ello se realiza un mapeo sistemático con el fin de conocer el estado del arte, se estudian las técnicas utilizadas y en base a ello se realiza un caso de estudio donde a partir de los datos (exoma) de un individuo y siguiendo un flujo de trabajo ("pipeline") definido se calcula el PRS y se interpreta los resultados en base a la prevalencia establecida para la EA.

La estructurada del trabajo se encuentra de la siguiente forma: En la sección 2 describe el estado del arte por medio de los resultados del mapeo sistemático. La sección 3 presenta los materiales utilizados. En la sección 4 describe el caso de estudio, la interpretación de resultados y la aplicación de técnicas de aprendizaje automático. En la sección 5 presenta las discusiones y en la sección 6 se presenta las conclusiones alcanzadas y las líneas futuras de trabajo.

2. Estado del arte

Los avances recientes en la tecnología genética, principalmente la secuenciación de próxima generación llamado en inglés Next-Generation Sequencing (NGS), redujo enormemente el costo y la velocidad de pruebas genéticas y con ello la comprensión de las causas genéticas de enfermedades neurológicas. Los estudios de asociación de genoma completo conocido en inglés Genome-wide Association Studies (GWAS) examinan todas las variaciones en el genoma humano incrementando la probabilidad de encontrar el gen o genes causantes de una determinada enfermedad. Las variantes genéticas se los conoce como polimorfismos genéticos, estas pueden clasificarse en benignas (no asociadas con la enfermedad) o patógenas (asociadas con la enfermedad). Las variantes de un solo nucleótido son las más frecuentes en nuestro genoma, son las sustituciones que afectan solo a un par de bases, denominadas polimorfismos de un solo nucleótido, en inglés Single Nucleotide Polymorphism (SNP) [24], se estima que hay al menos 11 millones de SNPs en el genoma humano.

2.1 Puntuación de riesgo poligénico y aprendizaje automático

La puntuación de riesgo poligénico (PRS) es una técnica que combina la prueba de ADN con las técnicas de modelado matemático para determinar si las enfermedades tienen algún componente genético. Utiliza un enfoque de modelo fijo para sumar la contribución de un conjunto de alelos de riesgo a una enfermedad compleja. El 2007 se propuso un método para examinar la influencia agregada de múltiples marcadores genéticos [27]. El método implicaba generar un PRS basada en los resultados de un GWAS. Después de ejecutar un GWAS en una muestra se seleccionan los SNP asociados al fenotipo, el PRS se calcula como una suma de esos SNPs asociados que a menudo son ponderados por los coeficientes específicos del GWAS. La mayoría de las investigaciones utilizaron este método para obtener un PRS, en la tabla X se muestra la agrupación de artículos según el tipo de datos que utilizaron en sus estudios.

Las técnicas de PRS se han vuelto cada vez más populares facilitando descubrimientos genéticos para rasgos complejos. El programa más conocido y usado para realizar la estimación de PRS es PRSice2 [10] el cual se utilizará en el caso de estudio acompañado de la herramienta PLINK [6]

Por otra parte, la aplicación de técnicas de ML a partir de su conjunto de algoritmos supervisados y no supervisados ofrece varios enfoques de aplicación como diagnóstico de enfermedades [4], desarrollo de fármacos [23], etc. Las técnicas que han sido aplicadas con éxito la predicción de la EA se muestran en la tabla XX donde se observa cierta preferencia a algunos métodos como las máquina de vectores de soporte (SVM), Random forests y Regresión logística, entre otros.

2.2 Patología neurológica – Alzheimer

La EA se divide en 2 subtipos, según la edad de inicio: EA de inicio temprano y EA de inicio tardío, ambos con características peculiares y son investigadas

independientemente. La EA se presenta debido a un daño lento y continuo de diferentes partes del cerebro. La comunidad científica maneja la hipótesis de que este daño que afecta a las neuronas es el resultado de dos procesos biológicos diferentes que se producen en el cerebro, la proteína Beta Aminoide y la proteína Tau que son los biomarcadores biológicos para la EA [3]. La EA también es causada por mutaciones genéticas, es decir tiene un componente genético muy fuerte, cuya caracterización se ha convertido en una parte esencial de los esfuerzos para comprenderla. Existen tres genes de la EA de inicio temprano con mutaciones causales la APP ubicado en el cromosoma 21, la PSEN1 en el cromosoma 14, la PSEN2 en el cromosoma 1 y un gen de inicio tardío la ApoE ubicado en el cromosoma 19[58], las personas que heredan uno de estos genes presentan altos riesgos de desarrollar la EA ya sea a una edad temprana o tardía. Sin embargo, se ha demostrado existen factores de riesgo asociados fuertemente a la reducción de riesgo de desarrollar la EA como la actividad física regular, el manejo de factores de riesgo cardiovascular (especialmente diabetes, obesidad, fumar e hipertensión), la dieta saludable, aprendizaje permanente y el entrenamiento cognitivo [28].

2.3 Resultados del mapeo sistemático

El mapeo sistemático se llevó a cabo siguiendo el proceso descrito por [21]. Las búsquedas sistemáticas se realizaron en tres bases de datos (PubMed, Scopus, Science Direct), se tiene un total de 41 artículos que cumplieron los criterios de selección. La información más detallada se encuentra almacenada en un repositorio de GitHub de acceso libre [19]. Como resultado del mapeo se tiene la Tabla 1 que muestra las técnicas utilizadas por más de 2 artículos (la información completa de la tabla se encuentra en el repositorio)

Tabla 1: Técnicas utilizadas por los artículos seleccionados

Técnicas	Total de Artículos
Algoritmos genéticos	3
LASSO	5
Step-wise (BSWiMS)	2
Gradient boosted	2
Support vector Machine (SVM)	17
Support vector regressor (SVR)	2
Naïve Bayesian	2
k-Nearest Neighbors (kNN)	3
Elastic Net	2
Gaussian processes (GP)	3
Logistic Regression	17
Random Forest	10
Linear Regression	8
Neural Network	6
Cross-Validation	13

En la Tabla 2 se tiene los diferentes tipos de datos que utilizaron los autores para realizar sus respectivos estudios. Algunos estudios utilizaron los tres tipos de datos.

Tabla 2: Tipo de dato utilizados en los artículos seleccionados

Tipo de dato	Total de Artículos
Datos genéticos (GWAS, exoma, genotipo, fenotipo, etc.)	41
Datos demográficos	19
Imágenes de resonancia magnética (MRI)	20

3. Métodos

3.1 Análisis estadístico

El análisis estadístico dentro el cálculo de PRS se desarrolla con la finalidad de verificar el mejor modelo predictivo, también usualmente se analiza la sensibilidad, especificidad y previsibilidad del modelo para identificar correctamente

los casos y controles. Para encontrar los mejores predictores se debe probar una variedad de modelos de regresión hasta encontrar una mejor precisión de predicción para ello se utiliza el análisis de regresión logística para establecer valores predictivos (sensibilidad, especificidad, AUC, PPV, NPV) mediante un modelo de regresión logística definiendo variables predictoras como la edad, sexo, educación, etc. Otro análisis importante que proporciona más detalles de la tendencia de datos es la puntuación de deciles, cuartiles, percentiles, etc.

El análisis estadístico usado particularmente en el campo de la genética conlleva realizar varias actividades a partir de una muestra genética. Suponiendo que se inicia de una muestra de GWAS, se puede realizar un análisis de regresión logística para examinar las asociaciones entre riesgo de EA, donde se pueden usar diferentes enfoques analíticos, por ejemplo: Utilizar una muestra completa de GWAS, utilizar solo controles de GWAS [7]. Shun en su estudio sobre la obtención de PRS para la depresión y ansiedad en mujeres, modela de dos maneras el PRS haciendo uso de quintiles para probar una posible relación no lineal y por aumento de desviación estándar. El ajuste del modelo se da en base a ciertas variables como la edad, la estratificación de población, índice de masa corporal, años de fumar, actividad física.

3.2 Algoritmo de puntuación de riesgo poligénico

Para determinar el PRS Choi et al recomienda considerar varios aspectos que lo detalla en su artículo [9] como también recomienda realizar un control de calidad a los datos antes de calcular el PRS. Estas recomendaciones son aplicadas en el caso de estudio.

Las puntuaciones de riesgo poligénico se calculan sumando todos los SNP que cumplen un umbral de valor p , o un conjunto de umbrales. Para cada individuo, se suma el número de alelos de riesgo transportados en cada variante (0, 1 o 2). Según [9] creador de la herramienta PRSice2, indica que fórmula para calcular el PRS es:

$$PRS = \sum_i^m \beta_i \left(\sum_{j=0}^2 w_{i,j} \times j \right) \quad (1)$$

Donde la $w_{i,j}$ es la probabilidad de observar el genotipo j , donde $j \in \{0,1,2\}$, para el i ésimo SNP; y la β_i es el tamaño del efecto del i ésimo SNP estimado a partir de los datos relevantes del estudio de asociación de genoma de base (GWAS). Para herramienta PLINK [6] la fórmula predeterminada para el cálculo de PRS en PLINK es:

$$PRS_j = \frac{\sum_i^N S_i \times G_{i,j}}{P \times M_j} \quad (2)$$

Donde el tamaño del efecto de SNP i es S_i ; el número de alelos de efecto observados en la muestra j es $G_{i,j}$; el número total de SNP incluidos en el PRS es N y el número de SNP no faltantes observados en la muestra j es M_j .

4. Caso de estudio

A lo largo de esta sección se describe el caso de estudio del presente artículo, partiendo con la identificación del conjunto de datos para realizar el cálculo de PRS, seguidamente la identificación de variantes y el genotipado, luego realizar el cálculo del PRS con las herramientas de PRSice2 [10], PLINK [6] y con distintos modelos obtenidos por medio de estudios asociativos previos (GWAS). En la figura 1 se muestra el flujo de trabajo descrito anteriormente y además se denota que se debe realizar otras acciones intermedias que se describe en las siguientes subsecciones.

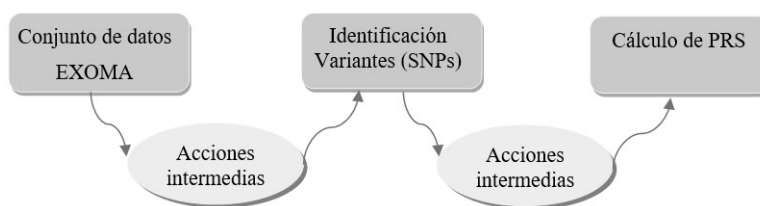


Figura 1: Flujo de trabajo del caso de estudio. Fuente: Elaboración propia

4.1 Conjunto de datos

Los PRSs requieren dos conjuntos de datos de entrada: El primero Datos base (GWAS): estadísticas resumida y el segundo datos objetivo que son los genotipos y fenotipos del individuo o muestra. Para el conjunto de datos objetivo se decidió utilizar un exoma de un individuo descargado del Proyecto 1000 Genomas [2]. A continuación, se describe la información conocida del conjunto de datos objeto: Muestra HG00126:

Sex: Masculino

Población: Británico Inglaterra y Escocia, ascendencia europea

ID Bio muestra: SAME122872

ID individuo: SRR707196

La información que contiene este archivo es muestra en la figura 2 que el código genético del individuo compuesto por los 23 cromosomas y cada cromosoma compuesto por los cuatro nucleótidos ACGT (adenina, citosina, guanina, timina).

El segundo conjunto de datos que se va utilizar es datos base (GWAS). Se descargo del sitio de IGAP (International Genomics of Alzheimer's Project) [16]. El archivo contiene 11632 polimorfismos de un solo nucleótido (SNP) agrupadas por el número de cromosoma y cada SNP tiene su posición base (Position), el identificador de SNP (MarkerName), el alelo de referencia (Effectallele), el alelo sin referencia (NonEffectallele), el tamaño total estimado del efecto para el alelo efecto (Beta), Error estándar general para la estimación del tamaño del efecto (SE) y el valor P de asociación entre los genotipos SNP y el fenotipo base (Pvalue).

Para la identificación de variantes se sigue el flujo de trabajo ("pipeline") definido en la figura 3, estas acciones pueden ser ejecutadas por diferentes herramientas automatizadas tal es el caso Galaxy, varstation, SNPalyzer, piMASS, entre otros. Inicialmente se preparó el entorno de trabajo sobre la plataforma Unix y tras instalar los diferentes paquetes que se requerían se procedió con la alineación del exoma.



Figura 3: Flujo de trabajo – alineamiento de exoma. Fuente: Elaboración propia

La examinación de la secuencia permitió ver la calidad de secuenciación y en base a ello realizar el control de calidad. Para iniciar con la alineación del exoma se requiere un genoma de referencia, para este caso se utilizó la referencia hg38.fa [15]. Para realizar la alineación por medio de los paquetes instalados como bwa, samtools, etc. se utilizó una secuencia de comandos de la guía [25]. Se han identificado 44669 SNPs que se encuentran en el archivo VCF, inicialmente no contaba con el identificador del SNP, usando la herramienta snpEff [11] se completó el identificador, este proceso es conocido como "Variant Annotation". En la figura 4 se muestra un pequeño fragmento del archivo VCF con los identificadores de cada SNP.

Con los SNP del individuo ya se puede calcular el PRS. Sin embargo, se debe realizar primeramente el control de calidad de los datos base y objetivo de acuerdo con las recomendaciones de Choi et al [9]. Los siguientes filtros se realizaron sobre los datos base usando el comando awk. Luego de aplicar los filtros se tiene 5816 SNP de 44669 SNP en el conjunto base.

1. Filtrar los SNP de acuerdo con la puntuación MAF, con la opción
2. Eliminar todos los SNPs ambiguos
3. Filtrar SNPs duplicados

Para los datos objetivo, se realizó la eliminación de SNP de baja frecuencia de alelos menores (MAF) y el equilibrio de Hardy-Weinberg (HWE) utilizando el comando `- - maf` y `- - hwe` en PLINK. inicialmente se tenía 44669 SNPs, tras aplicar los filtros de MAF y HWE se han eliminado 18274 SNP debido a la baja

frecuencia de alelos. Finalmente se tiene un total de 26395 SNP en el conjunto de datos objetivo.

```

root@nmena-VirtualBox: /home/nmena/Documentos/dataset
##INFO=<ID=SAO,Number=1,Type=Integer,Description="Variant Allele Origin: 0 - unspecified, 1 -
##INFO=<ID=GENEINFO,Number=1,Type=String,Description="Pairs each of gene symbol:gene id. The
##INFO=<ID=INT,Number=0,Type=Flag,Description="In Intron FxnCode = 6">
##INFO=<ID=G5,Number=0,Type=Flag,Description=">5% minor allele frequency in 1+ populations">
##INFO=<ID=OM,Number=0,Type=Flag,Description="Has OMIM/OMIA">
##INFO=<ID=PMC,Number=0,Type=Flag,Description="Links exist to PubMed Central article">
##INFO=<ID=SSR,Number=1,Type=Integer,Description="Variant Suspect Reason Codes (may be more t
##INFO=<ID=RSPOS,Number=1,Type=Integer,Description="Chr position reported in dbSNP">
##INFO=<ID=HD,Number=0,Type=Flag,Description="Marker is on high density genotyping kit (50K d
##INFO=<ID=PM,Number=0,Type=Flag,Description="Variant is Precious(Clinical.Pubmed.Cited)">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT sampleSRR sor
chr1 930939 rs9988021 G A 180.0 PASS DP=11;VDB=0.163983;SGB=-0.662
chr1 941119 rs4372192 A G 225.0 PASS DP=17;VDB=0.0226682;SGB=-0.68
chr1 952180 rs3748595 A C 225.0 PASS DP=56;VDB=0.284063;SGB=-0.693
chr1 953778 rs13303056 G C 209.0 PASS DP=12;VDB=0.235908;SGB=-0.686
chr1 953779 rs13302945 A C 215.0 PASS DP=12;VDB=0.235908;SGB=-0.686
chr1 959193 rs13303010 G A 225.0 PASS DP=25;VDB=0.543918;SGB=-0.692
chr1 961945 rs4970441 G C 225.0 PASS DP=56;VDB=0.00287564;SGB=-0.6
chr1 964548 rs6677386 G C 225.0 PASS DP=19;VDB=0.00107536;SGB=-0.6
chr1 964557 rs143296006 G T 210.0 PASS DP=15;VDB=0.000291538;SGB=-0.6
chr1 964558 rs147467971 G C 215.0 PASS DP=15;VDB=0.000291538;SGB=-0.6
chr1 964562 rs71509448 G A 220.0 PASS DP=15;VDB=0.000291538;SGB=-0.6
(base) root@nmena-VirtualBox: /home/nmena/Documentos/dataset#

```

Figura 4: Variantes del individuo - archivo VCF anotado

Una vez concluido con el control de calidad se procede a generar los archivos binarios (bed, bim, fam) con el comando `-make-bed` en PLINK para luego utilizar como datos de entrada para el cálculo de PRS.

4.3 Cálculo de puntuación de riesgo poligénico

Al momento de realizar el calculo de PRS se debe analizar el desequilibrio de enlace que corresponde a la correlación entre los genotipos de las variantes genéticas en todo el genoma, donde es extremadamente difícil la identificación de la contribución de las variantes genéticas causales. Una forma de capturar aproximadamente el nivel correcto de señal causales es realizar una agrupación (clumping), ya que elimina los SNP de manera que solo se retienen los SNP débilmente correlacionados, pero preferentemente retiene los SNP más asociados con el Alzheimer. El agrupamiento (clumping) se puede realizar utilizando el comando `-clump` para los parámetros con los valores de: `-clump-kb 250`, `-clump-p1 1`, `-clump-r2 0.1`. Después de realizar el clumping se observa que de 26395 SNPs solo 27 SNPs pasaron los filtros y el control de calidad, por tanto con 27 SNPs se calcula el PRS. Se genera el PRS con los parámetros del estudio de Cruchaga et al [12] en PLINK con un umbral de $1.21e-03$ usando el comando `-score`. El valor de PRS generado es: 0.02934375.

Usando la herramienta de PRSice, se procede a calcular el PRS siguiendo el método de Chandler et al [Chanfler] donde utilizó los siguientes parámetros:

MAF 0.01, HWE $1e-06$, clump-kb 500, clump-r2 0.25 y un umbral de $5e-08$. Para este caso se tiene 25 SNP para el cálculo del PRS, usando el comando expuesto abajo el valor de PRS es de 0.0363. Se observa que varía el valor de PRS comparado con el método de Cruchaga et al, debido a los umbrales utilizados es diferente. El comando completo para calcular el PRS en PRSice es:

```
Rscript PRSice.R --dir . --prsice ./PRSice_linux
--base baseQC.txt
--target sampleready.QC
--thread 1 --snp MarkerName
--A1 Effect_allele
--stat Beta --beta
--pvalue Pvalue
--binary-target F
--chr Chromosome --fastscore
--no-regress --bar-level 5e-08
--clump-kp 500 --clump-r2 0.25
--out prsresultHannah
```

Una vez obtenido el segundo valor de PRS con un método diferente se procede a calcular el PRS con los diferentes métodos, parámetros y umbrales que definieron los estudios revisados en el mapeo sistemático. En la tabla 3 se exponen los diferentes PRS calculados con la herramienta PRSice2 y los valores de los distintos parámetros que definieron los autores.

4.4 Interpretación de resultados

La prevalencia de la enfermedad es un parámetro, que debe especificarse para la estimación de AUC. La prevalencia de EA en la población depende mucho particularmente de la edad. Estimaciones recientes muestran una prevalencia del 3 % 17 % y 33 % en los grupos de edad 65-74, 75-84 y 85+, respectivamente [14]. Según [13] indica que la prevalencia de vida de AD 2 %, por lo que la influencia del riesgo poligénico varía según la función de la edad. Lewis et al [18] define por medio de una distribución normal de puntuaciones de riesgo poligénico los valores a considerar que dependen de la prevalencia de la enfermedad y del valor “t” que devuelve el cálculo de PRS $\text{medio} = -\text{prev} \times t / (1 - \text{prev})$ lo que representa a los controles no afectados por el trastorno, también describen el análisis individual de PRS se basa en percentiles, siendo las medidas más utilizadas para evaluar la capacidad predictiva de un PRS.

Tabla 3: PRS obtenido por diferentes métodos con PRSice2

Autor	PRS	Umbral	SNP	R2	KB	MAF	HWE
Cruchaga [12]	0.02934375	1.21e -03	27	0.1	250	0.01	1e -06-
Chandler [5]	0.0363	5e -08	25	0.25	500	0.01	-
Scott Price [13]	0.028502	0.5	25	0.2	500	-	-
Chaudhury [8]	0.0377	1e -06	27	0.1	250	-	1.03e -07
Leonenko [17]	0.03025	0.001	25	0.1	1000	0.1	1e -06
Rosales [20]	0.0363	5e -08	25	-	-	-	1e -05

En base a los criterios de interpretación de resultados se procede a interpretar los resultados presentados en la tabla 3. Revisando los estudios donde han obtenido el predictor de PRS para Alzheimer señalaron que se maneja el 2 % de prevalencia y Cohens D = 0.583 y un $t = 0.28$ [5] por lo cual los controles no afectados por el trastorno de Alzheimer son - 0.0057. Con el valor de PRS 0.0363 indicamos que el individuo no presenta una alta probabilidad de contraer EA. De la misma forma para el estudio de ScottPrice et al [13] la prevalencia es del 2 % y $t = 0.24$, entonces los controles no afectados por el trastorno serían - 0.0049 para un PRS = 0.028502 se indica que el individuo no presenta alta probabilidad de contraer la EA. Además, revisando la lista de 27 SNP que se tomó en cuenta para el cálculo de PRS no presenta ninguno de los genes más comunes asociado con la EA como el APOE, PSEN1 y PSEN2, por lo cual se llega a la conclusión que el individuo con los PRS calculados por diferentes métodos no presenta alta probabilidad de contraer la EA.

4.5 Técnicas de aprendizaje automático

La aplicación de técnicas de ML basadas en datos genéticos se ha popularizado para identificar los factores de riesgo asociados con varias enfermedades, ya que han demostrado ser robustas cuando el problema involucra cientos de miles de predictores [20]. La aplicación de ML permitió construir y mejorar modelos poligénicos para predecir la EA a partir de datos GWAS. En el presente trabajo se implementó una técnica de ML siguiendo los estudios de Rosales et al [20], con los 27 SNPs se intentó reproducir los resultados para el modelo LASSO usando la herramienta de FRESA.CAD. Sin embargo, no se logró en lo mínimo aproximarse al resultado. Evidente luego de hacer muchas pruebas e investigaciones no fue claro entender que parámetros utilizados del conjunto de datos para realizar la clasificación de las variantes y luego generar la curva de ROC (especificidad versus sensibilidad). Para generar la curva ROC con el algoritmo de LASSO se utilizó el paquete pglmet sobre la plataforma R. Se generó los coeficientes y la curva de ROC (Figura 5) pero estos no son los valores esperados tener un

$AUC = 1$ es totalmente erróneo, no se puede tener modelo de predicción con una precisión del 100

```

Coefficients:
(Intercept)  SNPrs11209  SNPrs117187697  SNPrs12286721  SNPrs1473261  SNPrs1697  SNPrs16973457
-12.3430      4.4863      10.0109      2.6352      7.9152      3.0930      4.1942
SNPrs17030969  SNPrs17188228  SNPrs1804020  SNPrs2293579  SNPrs2297128  SNPrs2303697  SNPrs3115446
12.3390      5.7441      0.4415      -5.0061      4.7346      -0.1777      5.5407
SNPrs35251323  SNPrs3750013  SNPrs3764648  SNPrs405509  SNPrs439401  SNPrs4463136  SNPrs4810988
-3.9506      6.4784      7.2464      -11.2230      -11.2230      6.7924      3.9831
SNPrs4980533  SNPrs56356701  SNPrs5918  SNPrs61729628  SNPrs6451173  SNPrs7201952  SNPrs74064442
6.3048      7.9323      5.5586      5.8815      4.5847      3.4835      2.6612
SNPrs7830986
8.5716

Degrees of Freedom: 28 Total (i.e. Null); 0 Residual
Null Deviance: 3.496
Residual Deviance: 2.33e-10  AIC: NA

```

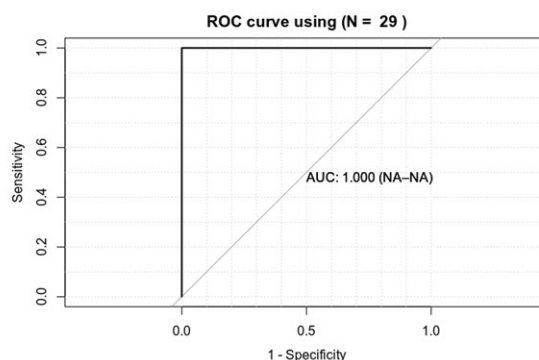


Figura 5: Coeficientes y curva ROC del modelo LASSO (resultado erróneo)

Un punto a resaltar en este contexto la cantidad de datos tiene un gran impacto en la precisión del modelo de predicción, se debe contar con dos conjuntos de datos uno para entrenamiento del modelo y otro para el testeo. Evidentemente no solo se aplica en el contexto de la genética sino en cualquier campo donde se utilice técnicas de ML para determinar modelos de predicción. En este caso de estudio se requiere una gran cantidad de SNPs para construir el modelo, se ha intentado con 27 SNPs un número bastante bajo comparando con los datos usado por [20] que son 1000 SNPs.

5. Discusión

Realizar el cálculo de PRS no fue complicado ya que se cuenta con los programas como PLINK y PRSice que facilitan el cálculo por medio de los comandos de línea, además se cuenta con una gran comunidad que brinda soporte ante cualquier problema. La interpretación de los resultados fue el uno de los obstáculos que se tuvo, el hecho de ver números como resultado de un análisis predictivo lleva a la intuición de querer interpretar los resultados. Sin embargo, tras investigar

y revisar las guías [22] de cómo realizar la interpretación los valores del PRS se llegó a comprender con más seguridad y claridad el uso del PRS y importancia de la interpretación de los resultados genéticos en el contexto clínico.

Se ha demostrado que a partir de un exoma se identifico las variaciones del individuo para luego usarlo como datos de entrada para el cálculo de PRS, dicho cálculo fue en base a los métodos de precisión de predicción construidos en los estudios seleccionados. Los resultados obtenidos fueron interpretados según la prevalencia de EA, llegando a concluir que el individuo estudiado no presenta altos riesgos de contraer la EA. En segundo lugar, se ha analizado la aplicación de técnicas de ML con datos GWAS, donde a partir de 27 SNPs se intentó construir un modelo de clasificación basado en el algoritmo de LASSO, sin embargo, debido a la poca cantidad de SNP no se logró calcular un modelo correcto.

Una posible limitación en el estudio fue la incompleta información de los parámetros requeridos para el cálculo del PRS (tabla 3), así mismo la poca información del individuo, solo se tenía conocimiento del sexo y la ascendencia, con la poca información que se tiene fue complicado llegar a interpretar los resultados del PRS ya que sea visto que la edad es un factor muy importante que permite predecir el riesgo específico de la EA. También otra limitante fue el poco tiempo que se tuvo para abarcar una inmensa área de investigación, con varios aspectos internos por investigar de manera independiente. Sin embargo, se ha logrado cumplir con los objetivos planteados en el presente estudio. Además de diseñar un “pipeline” que podrá ser utilizado para el análisis del exoma de cualquier individuo dando como resultado final su puntuación de riesgo poligénico para padecer la enfermedad de Alzheimer.

6. Conclusiones y líneas futuras de trabajo

Una vez calculo el PRS, el puntaje simplemente representa una representación de la propensión genética de la muestra a una determinada enfermedad y no se sabe cuál es el riesgo relativo en comparación con otros. Actualmente, el PRS sigue siendo el más adecuado para el análisis basado en la población, pero todavía está bastante lejos de poder hacer predicciones a nivel individual.

Llegar a interpretar un $PRS = 0.02934375$ (tabla 3) de alguna manera sigue siendo una predicción basada en la población ya sea que los estudios trabajaron en cohortes o estratificaciones específicas. A partir de ahora se comprende que el uso de PRS es para realizar algún tipo de análisis de estratificación, estratificando muestras en cuantiles / deciles. O para cada cuantil / decil se puede calcular en relación con una referencia. Entonces se puede proporcionar un riesgo.^a las muestras en diferentes cuantiles. Sin embargo, se debe tener en cuenta que esto todavía está en el nivel de la población, es decir, el riesgo es para las muestras dentro del cuantil de la muestra, cada una de esas muestras puede tener un riesgo real" diferente, por lo que todavía no se puede decir que la muestra A tiene una probabilidad del 10 % para desarrollar la EA en el futuro. De la misma forma en las técnicas de ML la construcción de modelos de predicción o clasificación se basan en una muestra específica en este caso la cantidad de datos en mucho mayor sin embargo la precisión de predicción es a nivel de población o muestra.

Como futuro trabajo se propone llegar a construir un PRS con datos GWAS (manejar casos y controles) en base a alguno de los estudios seleccionados [20]. Luego de la construcción con los mismos datos construir un modelo de predicción con las técnicas de ML y realizar la comparación respectiva, concluyendo si para esa cohorte las técnicas de ML proporcionan una mejor precisión. Es posible encontrar varios factores externos desde obviar un simple filtro de control de calidad en los datos base o objetivo el cual podría tener una relación colateral en los resultados. Así mismo se sugiere investigar los diferentes métodos y modelos de cálculo de PRS, por ejemplo, la herramienta PRSice2 [30] ofrece varios modelos –model add, dom, rec, het y métodos –score avg, sum, con-std, std que pueden ser utilizado al momento de realizar el cálculo e incluso con estudio de estos métodos llegar a proponer un nuevo método de cálculo de PRS.

Referencias

- [1] Alzheimer’s Association. “2019 Alzheimer’s disease facts and figures”. En: *Alzheimer’s & Dementia* 15.3 (mar. de 2019), págs. 321-387. ISSN: 15525260. DOI: 10.1016/j.jalz.2019.01.010. URL: <http://doi.wiley.com/10.1016/j.jalz.2019.01.010>.
- [2] Adam Auton y col. *A global reference for human genetic variation*. Sep. de 2015. DOI: 10.1038/nature15393. URL: <https://www.nature.com/articles/nature15393>.
- [3] Lynn M. Bekris y col. *Genetics of Alzheimer disease*. Dic. de 2010. DOI: 10.1177/0891988710383571. URL: <http://journals.sagepub.com/doi/10.1177/0891988710383571>.
- [4] Kristy A. Carpenter y Xudong Huang. “Machine Learning-based Virtual Screening and Its Applications to Alzheimer’s Drug Discovery: A Review”. En: *Current Pharmaceutical Design* 24.28 (dic. de 2018), págs. 3347-3358. ISSN: 13816128. DOI: 10.2174/1381612824666180607124038.
- [5] Hannah L. Chandler y col. “Polygenic risk for Alzheimer’s disease shapes hippocampal scene-selectivity”. En: *Neuropsychopharmacology* 45.7 (2020), pág. 1171. ISSN: 1740634X. DOI: 10.1038/s41386-019-0595-1.
- [6] Christopher C Chang y col. “Second-generation PLINK: rising to the challenge of larger and richer datasets”. En: *GigaScience* 4.1 (dic. de 2015), pág. 7. ISSN: 2047-217X. DOI: 10.1186/s13742-015-0047-8. URL: <https://academic.oup.com/gigascience/article-lookup/doi/10.1186/s13742-015-0047-8>.
- [7] Shun Chiao Chang y col. “Polygenic risk score of shorter telomere length and risk of depression and anxiety in women”. En: *Journal of Psychiatric Research* 103 (ago. de 2018), págs. 182-188. ISSN: 18791379. DOI: 10.1016/j.jpsychires.2018.05.021.
- [8] Sultan Chaudhury y col. “Alzheimer’s disease polygenic risk score as a predictor of conversion from mild-cognitive impairment”. En: *Translational Psychiatry* 9.1 (dic. de 2019), pág. 154. ISSN: 2158-3188. DOI: 10.1038/s41398-019-0485-7. URL: <http://www.nature.com/articles/s41398-019-0485-7>.

- [9] Shing Wan Choi, Timothy Shin Heng Mak y Paul F. O'Reilly. "A guide to performing Polygenic Risk Score analyses". En: *bioRxiv* 2 (2018), pág. 416545. DOI: 10.1101/416545. URL: <https://www.biorxiv.org/content/10.1101/416545v1>.
- [10] Shing Wan Choi y Paul F O'reilly. "PRSice-2: Polygenic Risk Score software for biobank-scale data". En: 8 (2019), págs. 1-6. DOI: 10.1093/gigascience/giz082. URL: <http://orcid.org/0000-0003-2215-3238>;
- [11] Pablo Cingolani y col. "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3". En: *Fly* 6.2 (2012), págs. 80-92. ISSN: 19336942. DOI: 10.4161/fly.19695. URL: [/pmc/articles/PMC3679285/?report=abstract](http://pmc/articles/PMC3679285/?report=abstract) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3679285/>.
- [12] Carlos Cruchaga y col. "Polygenic risk score of sporadic late-onset Alzheimer's disease reveals a shared architecture with the familial and early-onset forms". En: *Alzheimer's and Dementia* 14.2 (2018), págs. 205-214. ISSN: 15525279. DOI: 10.1016/j.jalz.2017.08.013. URL: http://adni.loni.usc.edu/wpcontent/uploads/how%7B%5C_%7Dto%7B%5C_%7Dapply/ADNI%7B%5C_%7DAcknowledgement%7B%5C_%7DList.pdf <http://dx.doi.org/10.1016/j.jalz.2017.08.013>..
- [13] Valentina Escott-Price y col. "Common polygenic variation enhances risk prediction for Alzheimer's disease". En: *Brain a journal of neurology* 138 (2015). ISSN: 3673-3684. DOI: 10.1093/brain/awv268.
- [14] Liesi E. Hebert y col. "Alzheimer disease in the United States (2010-2050) estimated using the 2010 census". En: *Neurology* 80.19 (mayo de 2013), págs. 1778-1783. ISSN: 00283878. DOI: 10.1212/WNL.0b013e31828726f5.
- [15] W. J. Kent y col. "The Human Genome Browser at UCSC". En: *Genome Research* 12.6 (mayo de 2002), págs. 996-1006. ISSN: 1088-9051. DOI: 10.1101/gr.229102. URL: [/pmc/articles/PMC186604/?report=abstract](http://pmc/articles/PMC186604/?report=abstract) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC186604/>.
- [16] Jean Charles Lambert y col. "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease". En: *Nature Genetics* 45.12 (dic. de 2013), págs. 1452-1458. ISSN: 15461718. DOI: 10.1038/ng.2802.
- [17] Ganna Leonenko y col. "Polygenic risk and hazard scores for Alzheimer's disease prediction". En: *Annals of Clinical and Translational Neurology* 6.3 (mar. de 2019), págs. 456-465. ISSN: 23289503. DOI: 10.1002/acn3.716.
- [18] Cathryn M. Lewis y Evangelos Vassos. *Polygenic risk scores: From research tools to clinical instruments*. Mayo de 2020. DOI: 10.1186/s13073-020-00742-5. URL: <https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-020-00742-5>.
- [19] *Repositorio de información, Mapeo sistemático*. URL: <https://tinyurl.com/y9vwnzhu>.

- [20] Brissa Lizbeth Romero-Rosales y col. "Improving predictive models for Alzheimer's disease using GWAS data by incorporating misclassified samples modeling". En: *PLoS ONE* 15.4 (2020), págs. 1-15. ISSN: 19326203. DOI: 10.1371/journal.pone.0232103.
- [21] Cécilia Samieri y col. "Association of cardiovascular health level in older age with cognitive decline and incident dementia". En: *JAMA - Journal of the American Medical Association* 320.7 (ago. de 2018), págs. 657-664. ISSN: 15383598. DOI: 10.1001/jama.2018.11499.
- [22] Chin Hong Tan y Rahul S. Desikan. *Interpreting Alzheimer disease polygenic scores*. Mar. de 2018. DOI: 10.1002/ana.25164. URL: <http://doi.wiley.com/10.1002/ana.25164>.
- [23] M Tanveer y col. "Machine Learning Techniques for the Diagnosis of Alzheimer's Disease: A Review". En: *ACM Trans. Multimedia Comput. Commun. Appl* 35.1s (2020), pág. 30. DOI: 10.1145/3344998. URL: <https://doi.org/10.1145/3344998>.
- [24] *The genetic basis of disease*. Dic. de 2018. DOI: 10.1042/EBC20170053.
- [25] Peter J. Ulintz, Weisheng Wu y Chris M. Gates. "Bioinformatics Analysis of Whole Exome Sequencing Data". En: *Methods in Molecular Biology*. Vol. 1881. Humana Press Inc., 2019, págs. 277-318. DOI: 10.1007/978-1-4939-8876-1_21. URL: http://link.springer.com/10.1007/978-1-4939-8876-1_21.
- [26] World Health Organization. "Neurological Disorders: Public Health Challenges". En: *Archives of Neurology* 65.1 (2008), pág. 154. ISSN: 0003-9942. DOI: 10.1001/archneurol.2007.19.
- [27] Naomi R. Wray, Michael E. Goddard y Peter M. Visscher. "Prediction of individual genetic risk to disease from genome-wide association studies". En: *Genome Research* 17.10 (oct. de 2007), págs. 1520-1528. ISSN: 10889051. DOI: 10.1101/gr.6665407.
- [28] Rachel Aine Yotter, Xiao Da y Christos Davatzikos. "LIFESTYLE FACTORS MAY DELAY THE AGE OF CLINICAL DIAGNOSIS OF ALZHEIMER'S DISEASE". En: *Alzheimer's & Dementia* 10.4 (jul. de 2014), P804-P805. ISSN: 1552-5260. DOI: 10.1016/j.jalz.2014.05.1576.

Visualización de incertidumbre en grafos sociales para la exploración de bibliotecas digitales

Iñaki Arberas, Roberto Therón, and Alejandro Benito Santos

Departamento de Informática y Automática,
Facultad de Ciencias, Universidad de Salamanca
Plaza de los Caídos, 37008 Salamanca, España
{idul2003,theron,
abenito}@usal.es
<http://www.usal.es>

Resumen Las humanidades y la computación se entrelazan para formar un único campo de investigación, denominado como Humanidades Digitales (DH). Los avances tecnológicos han servido para facilitar las diversas tareas que les puedan surgir a los usuarios de las humanidades y, a su vez, dichas tareas pueden resultar un foco de investigación para los expertos en computación centrados en la creación de soluciones que las puedan solventar. Las diversas soluciones presentadas por los expertos en computación deben ser entendibles y acordes al conocimiento computacional que una persona no necesariamente experta en dicho campo, como podría ser el caso de los humanistas, podría tener. De entre las diversas tareas a desempeñar en las humanidades, en la presente investigación se ha tratado de generar una solución numérica y visual a un problema típico, denominado como “Desambiguación de nombres”, con el objetivo de presentar dicha solución, intuitiva a la vez que efectiva, a usuarios no necesariamente expertos en el ámbito matemático y computacional.

Keywords: Humanidades Digitales, Analítica Visual, Visualización de la información, Desambiguación de nombres

1. Introducción

El campo de las Humanidades Digitales [7][4] (DH) avanza a medida que avanza también la tecnología y surgen nuevos problemas en el campo de las humanidades que motivan la búsqueda de nuevas soluciones computacionales. Los métodos que los expertos en computación deben presentar a los expertos en humanidades deben permitir el análisis de los datos y la interacción con respecto a éstos, para permitir a los humanistas obtener conclusiones propias y sin sesgo. Esto último, correspondiente al marco de la Analítica Visual, está contenido en el campo de la Visualización de la información, en el cual, el principal objetivo es el de transmitir conocimiento sin que éste sea enturbiado por los diferentes focos de incertidumbre[1][25][9][25][18][25] que se puedan generar.

El rango de tareas que se puede realizar dentro del campo de DH [2] es amplio, siendo la “Desambiguación de nombres” una de ellas. Este es un problema común en la historia de la investigación, ya que la necesidad de diferenciar a dos personas por su nombre es vital. En [22], por ejemplo, se trata el problema de asociar una publicación científica a una persona de forma incorrecta debido a que ésta tiene el mismo nombre y apellidos o usa las mismas abreviaturas que el autor original de la publicación. En consecuencia existen dos tipos de ambigüedad de nombres: los nombres sinónimos y los homónimos. El problema de la sinonimia en nombres viene dada por nombres que se escriben de forma diferente debido a las reglas lingüísticas que lo permiten, pero hacen referencia a la misma persona. En particular una persona con el nombre de “Barbara Masson” puede firmar sus publicaciones con su nombre y apellidos completos o por “B. Masson”. En cuanto a la homonimia de nombres, ocurre cuando un mismo nombre hace referencia a dos personas. Para este caso, en DBLP¹, una conocida web bibliográfica en el ámbito de la computación, si se hace una búsqueda del nombre “Rui Wang”, se encuentran un total de 15 investigadores diferentes con esta composición de nombre y apellido [22]. Se debe remarcar que las diferentes soluciones al problema dependerán en gran parte las características de éste y a las diferentes variables que puedan ser extraídas del contexto en el que esté centrado.

En particular, la tarea de la desambiguación de nombres que se ha tratado de realizar en esta investigación se ha centrado en una biblioteca humanística, de carácter histórico en particular. Se conoce con el nombre de “1641 depositions”² y es un conjunto de archivos o *corpora* históricos en los que se recogen declaraciones de personas que sufrieron diferentes incidentes o ataques en la rebelión Irlandesa de 1641. Se ha podido tener acceso a dicho conjunto de datos ya que esta investigación se enmarca en un proyecto de investigación internacional de PROVIDEDH³ (“*PROgressive Visual DEcision-Making in Digital Humanities*”), coordinado por el grupo de investigación en Analítica Visual y Visualiación de la información de la Universidad de Salamanca⁴.

2. Objetivos del trabajo

La tarea principal que se va a resolver, como ya se ha determinado, es la Desambiguación de nombres sobre la base de datos de “1641 depositions”. Para ello, se creará una ecuación, basada en las variables que se puedan extraer de la base de datos mencionada, que devuelva un valor entre 0 y 1, correspondiente a la probabilidad de que dos nombres puedan hacer referencia a la misma persona. Para ello se deberán investigar diferentes procedimientos que permitan obtener soluciones numéricas a este tipo de problemas.

A parte, esta tarea ha sido seleccionada ya que la creación de una ecuación personalizada resultaría de interés para enlazarla con el objetivo visual que per-

¹<https://dblp.uni-trier.de/>

²<https://1641.tcd.ie/>

³<https://providedh.eu/>

⁴<http://vis.usal.es/>

sigue, también, en esta investigación. En concreto lo que se busca es que, tras generar dicha ecuación, los resultados puedan ser planteados de forma visual a los usuarios, los cuales serán historiadores principalmente que busquen resolver una tarea de desambiguación sobre la base de datos comentada, siendo esto de vital importancia a lo largo del desarrollo de la solución visual, para que ellos mismos puedan inferir su propio resultado sin necesidad de entender la base matemática que habrá detrás del resultado numérico final. Es decir, la tarea de la desambiguación centrada en la base de datos histórica de 1641 podría resultar de gran interés para algún historiador experto en dicha época, sin embargo, puede que su conocimiento en el campo de las matemáticas no sea el suficiente como para que los datos le resulten fiables. Por lo tanto, lo que se busca conseguir es que una vez se haya realizado el estudio en términos matemáticos, cualquier persona no experta en dicho campo pueda inferir su propio resultado sobre los casos de las personas desambiguadas, en particular los historiadores, que podrán imponer, en cierto modo, su criterio de forma más estricta a la hora de analizar los resultados.

El proceso comentado queda dentro del campo de la Analítica Visual, lo cual implicará el estudio de los procedimientos aplicados en dicho campo con el objetivo de crear un proceso completo que permita a los usuarios entender, desde la base de datos de forma visual, hasta los resultados numéricos finales de las desambiguaciones. Para ello, se debe estudiar la incertidumbre que puede aparecer en cada una de las fases del proceso, así como las diferentes técnicas implementables para solventarla.

3. Trabajo relacionado

Este problema puede ser abordado con técnicas tanto matemáticas como visuales. En cuanto a resolver el problema de forma matemática, la estrategia utilizada es analizar las diferentes variables que permita sacar el problema y obtener valores numéricos que den una percepción de proximidad entre dos nombres. En [22] se da una solución al problema numérico asociado, valiéndose de diferentes términos como el de la similitud de Jaccard [14], por ejemplo.

Sobre la parte de la visualización para presentar al usuario únicamente los resultados asociados al problema, hay diversos artículos que muestran varias soluciones posibles. En [22] se dan un conjunto de visualizaciones, entre las que cabe destacar un grafo. Este tipo de visualización permite modelar una gran variedad de relaciones(en bioinformática [20][21], en ingeniería del software [10][11], en cienciometría [19] o en análisis de rendimiento deportivo [3][26] y particularmente en visualización en humanidades digitales [5][6][8], entre otros), codificando información tanto en sus aristas, como en sus nodos [12]. Así mismo, los algoritmos de representación de grafos pueden ser de gran ayuda, ya que se deben evitar problemas como el de solapamiento de nodos, por ejemplo. En [15] se da una explicación a un algoritmo que crea representaciones de grafos minimizando posibles errores o solapamientos, basando dicha visualización en un grafo dirigido por fuerzas [28].

Para la parte en la que se le presentan al usuario las diferentes visualizaciones, se debe tener en cuenta que éstas conviene que sean interactivas, ya que permiten al usuario realizar diferentes acciones que le permitirán analizar mejor la base de datos a tratar. Se debe destacar en este punto el concepto de Manipulación directa de Shneiderman [13][23], la cual trata con este tipo de relaciones hombre-computadora, que permiten a los usuarios una exploración de los objetos representados sencilla e intuitiva. Entre las estudiadas, se podría destacar la del [17], que genera un recuadro modificable por el usuario que permite hacer selecciones para, por ejemplo, filtrar los datos de un conjunto.

Finalmente se ha estudiado la incertidumbre que pueden surgir en las colecciones culturales históricas, debido a que el conjunto de datos hace referencia a un *corpora* antiguo. En [30] se da una clasificación detallada de dichos tipos, que será tenida en cuenta para la investigación llevada a cabo en este artículo.

El proceso completo tratará de ser construido sobre el mantra de Keim [16], que trata de modelizar los procesos asociados a la Analítica Visual y dice lo siguiente: “Analizar primero, mostrar lo importante, zoom, filtrado y estudio a fondo, detalles bajo demanda (*Analyze first, Show the Important, Zoom, filter and analyze further, Details on demand*)”. Dicho mantra es una modificación del mantra de Visualización de la información de Shneiderman [24].

4. Propuesta

4.1 Solución numérica

Lo primero que se debe determinar es una medida que cuantifique la similitud para cada par de nombres, basándose en determinadas características obtenidas de los testimonios que están contenidos en la base de datos de “1641 depositions”.

4.2 Variables

Tras analizar la información que aparece en los documentos del *corpus*, se observó que podían definirse dos tipos de variables principales: las relacionadas con cada par de participantes a analizar y las relacionadas a los testimonios a los que pertenecen dichos participantes. De este punto en adelante se utilizará participante y nombre de forma intercambiable, haciendo referencia a las personas que aparecen en cada testimonio.

4.2.1 Variables asociadas a cada par de participantes. Se han definido una serie de variables para cada par de nombres acotadas entre 0 y 1. Cuanto más cercanas estén de 1, más probable será que los nombres hagan referencia a la misma persona y menos probable de que lo sean en caso contrario. Las variables definidas son las siguientes:

- $ratio_{i,j}$ = ratio de Levenshtein evaluado sobre los nombres de los participantes i y j . Como ya se sabe el ratio de Levenshtein devuelve un valor entre 0 y 1, que hace referencia a la proporción de letras coincidentes entre ambos nombres.

$$\blacksquare S_{i,j} = \begin{cases} 1 & \text{si} \quad \text{sexo de } i = \text{sexo de } j \\ 0,35 & \text{en caso contrario} \end{cases}$$

Al no tener un subconjunto de datos correctamente clasificado y verificado, los valores de esta variable han sido seleccionados bajo un criterio personal. El hecho de haber seleccionado un valor de 0,3 a la ecuación en vez de un 0 en caso de no coincidir el sexo es debido a que se han podido cometer errores a la hora de recoger los datos por parte de los testigos. En caso de haber tenido un subconjunto clasificado, se podría haber sacado la proporción de personas a las que se les ha asignado de forma errónea el sexo y haber sustituido el valor de 0,3 por dicha proporción.

$$\blacksquare R_{i,j} = \begin{cases} 1 & \text{si} \quad \text{rol de } i = \text{rol de } j \\ 0,65 & \text{en caso contrario} \end{cases}$$

Al igual que para el caso de la variable anterior, tampoco se sabe la proporción de fallos que se han cometido a la hora de clasificar los roles de las personas. Se ha seleccionado un valor de 0,65 en caso de no coincidir en el rol, en vez de 0,3 como en el caso del sexo. Para esta modificación se ha supuesto que es más improbable que una persona sea la misma en dos testimonios cuando aparece con dos sexos diferentes, a si aparece con dos roles diferentes. A parte, como se ha podido observar en el estudio del dataset, existen roles cuya función es similar.

$$\blacksquare O_{i,j} = \begin{cases} 1 & \text{si} \quad \text{occupation de } i = \text{occupation de } j \\ 0,65 & \text{en caso contrario} \end{cases}$$

Esta variable viene determinada de forma análoga a explicada en el punto anterior.

4.2.2 Variables asociadas a cada par de testimonios. No solo las características particulares de cada persona resultan relevantes a la hora de desambiguar dos nombres, si no que también son relevantes datos aportados por el conjunto de personas al que pertenecen.

Para reducir la notación utilizada en la explicación de las variables se tendrán las siguientes notaciones:

- k = número de testimonios totales contenidos en el corpus
- $d = \{d_1, d_2, \dots, d_k\}$ = conjunto de todos los testimonios del corpus.
- d_{n_i} = testimonio n , donde aparece el participante i .
- $p(d_n)$ = personas que participan en el testimonio n
- $p_{dep} = \cup_{z=1}^k p(d_z)$ t.q. $dep \in p(d_z)$. Es decir, p_{dep} es el conjunto de personas que aparecen en los testimonios hechos por el testigo dep .

Teniendo en cuenta lo anterior se han definido las siguientes variables:

- $G_{i,j} = \frac{|p(d_{n_i}) \cap p(d_{n_j})|}{|p(d_{n_i}) \cup p(d_{n_j})|}$ = índice de Jaccard [14] asociado a la proporción de personas que coinciden en los testimonios de i y j .

Esta variable asocia un valor de probabilidad a dos nombres de que hagan referencia a la misma persona en función al número de participantes coincidentes en ambos testimonios. Son variables particulares del testimonio, ya que cualquier par de nombres de dos testimonios diferentes tendrán el mismo índice, debido a que las personas a tener en cuenta por la variable G son las mismas. Únicamente variará el valor en caso de sacar de los conjuntos $p(d_{n_i})$ y $p(d_{n_j})$ las personas i y j , teniendo en cuenta que si hay un número reducido de participantes por testimonio la variación en el índice G podría ser significativo.

- $D_{i,j} = \frac{|p_{dep_i} \cap p_{dep_j}|}{|p_{dep_i} \cup p_{dep_j}|}$ =índice de Jaccard asociado a la proporción de personas que coinciden en todos los testimonios de los autores de los testimonios a los que pertenecen i y j .

Esta variable funciona de forma análoga a la explicada en el punto anterior, con la diferencia de que los grupos de personas son generados por los autores de los testimonios a los que pertenecen i y j . La importancia de dicho término reside en la suposición de que dos personas tienen mayor probabilidad de ser la misma si los testigos han recogido testimonios a personas del mismo entorno.

La variable $D_{i,j}$ tendrá un valor diferente si se quitan de los conjuntos p_{dep_i} y p_{dep_j} los propios nombres a analizar, pudiendo ser un cambio de valor significativo en caso de constar de conjuntos pequeños.

- $Gauss_{i,j} = 1 - \exp(-(\frac{(tiempo-1)^2}{2} + \frac{(distancia-1)^2}{2}))$ t.q. $tiempo, distancia \in [0, 1]$.

Para obtener el valor de la variable *tiempo* se debe calcular el intervalo temporal, dado en días, entre los dos testimonios de los nombres que se van a analizar. Después, se deben obtener el mínimo y máximo intervalo temporal, dado también en días, para todos los pares de testimonios. Finalmente, se debe aplicar la normalización min-max teniendo en cuenta los tres valores obtenidos previamente, dando como resultado un valor acotado entre 0 y 1. Debido a las características de la función campana de Gauss, para obtener la relación deseada entre la variable *tiempo* y *distancia* que se explicará más adelante en este mismo apartado, se deberá realizar la resta de 1 menos el valor obtenido previamente para finalizar con el cálculo de *tiempo*.

El valor de la variable *distancia* se calcula de forma análoga a *tiempo*, teniendo en cuenta que en vez de intervalos temporales tendremos distancias, dadas en kilómetros. En este caso, no se debe realizar la resta final con 1.

Para explicar la relación que se obtiene de aplicar lo anterior, se va a proceder explicando el objetivo de dicha ecuación. Al tratar de desambiguar dos nombres, una de las relaciones más significativa es la asociada a las variables geo-temporales. Se puede considerar altamente probable que dos nombres hagan referencia a la misma persona si los lugares en los que se han recogido los testimonios es el mismo. También se puede considerar igualmente probable que dos nombres hagan referencia a la misma persona si, pese a no haberse recogido los testimonios en el mismo lugar ni cerca, ha pasado un periodo temporal suficientemente significativo como para poder suponer que dicha persona ha viajado de un lugar a otro. Para esta suposición hay

que tener en cuenta las dificultades de la época para desplazarse a largas distancias. El máximo valor se alcanza cuando dos testimonios han sido recogidos en el mismo lugar, en un intervalo temporal prácticamente nulo, es decir el mismo día, y el mínimo se obtendría en el caso opuesto a ambas afirmaciones. Sin embargo, no se trata de una relación lineal, teniendo en cuenta los siguientes dos factores: el valor final de la variable será cercano a 1 siempre que la distancia sea cercana a 0 independientemente del término temporal y será cercano a 1 también siempre que el intervalo temporal sea cercano al máximo valor para esta variable, independientemente del término espacial.

Con lo descrito en el párrafo anterior tenemos una relación entre dos variables que se podría modelizar mediante una función bivalente. Tras tratar de graficar la superficie tridimensional que conformaría, se puede ver la similitud con la campana de Gauss invertida, la cual se puede ver en la Figura 1 sin invertir. Por lo tanto, únicamente se tendría que aplicar la función de Gauss a los valores *tiempo* y *distancia* pero restando al resultado el valor de 1, para obtener dicha función invertida.

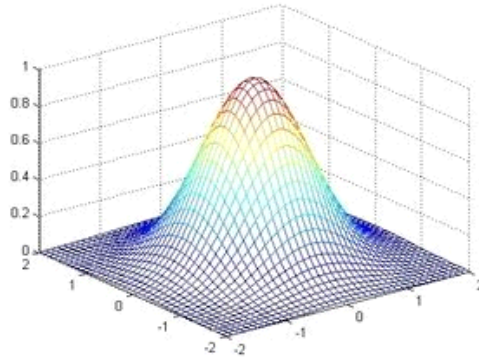


Figura 1: Función de Gauss centrada en (0,0)

4.3 Ecuación

La ecuación se ha determinado con el objetivo de facilitar un resultado final al estudio conjunto de todas las variables anteriores, tratando de resumir en un único valor todo lo que se debe tener en cuenta en este caso para tratar de desambiguar dos nombres. Por lo tanto, la ecuación es la siguiente:

$$M_{i,j} = (\alpha * ratio_{i,j} + \beta * Gauss_{i,j} + \gamma * (\gamma_1 * G_{i,j} + \gamma_2 * D_{i,j})) * S_{i,j} * R_{i,j} * O_{i,j} \quad (1)$$

$\alpha, \beta, \gamma, \gamma_1$ y γ_2 son parámetros para ponderar el valor de las variables a las que están asociadas, con el objetivo de que el usuario pueda decidir el peso que determinan cada una de ellas en la ecuación principal. Es decir, si alguien considera que coincidir en los apellidos debería tener mayor relevancia que las variables asociadas a los índices de Jaccard, entonces el parámetro α tendrá un valor mayor asociado con respecto a γ . Como se tiene que cumplir que $M_{i,j} \in [0, 1] \forall i, j \in \{\text{participantes}\}$, entonces $\alpha, \beta, \gamma, \gamma_1, \gamma_2 \in [0, 1]$, $\alpha + \beta + \gamma = 1$ y $\gamma_1 + \gamma_2 = 1$.

4.4 Proceso analítico

4.4.1 Mostrar lo importante Una vez que se ha creado la ecuación que devuelve el valor numérico se debe determinar un proceso mediante el que se le va a facilitar al usuario tanto la exploración sobre el conjunto de datos, como el entendimiento del resultado final que devuelve la ecuación para cada par de nombres. Una vez se le ha explicado el problema de la desambiguación a tratar, se le debe mostrar la dimensionalidad de la base de datos sobre la que se va a trabajar. En este caso, dicha dimensionalidad se puede pensar que está directamente relacionada con la cantidad de personas diferentes o posiblemente diferentes que constituyen el conjunto de datos. En consecuencia, para focalizar la atención del usuario en lo comentado se han seleccionado dos vistas enlazadas: un *choropleth* [31] y un diagrama de barras, como se puede ver en la Figura 2.

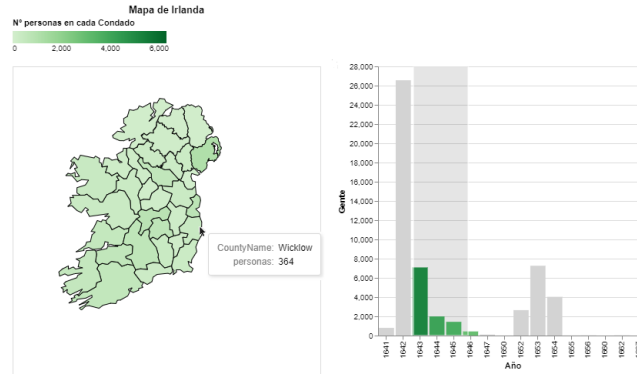


Figura 2: *Choropleth* y diagrama de barras filtrado aplicando un brush entre los años 1963 y 1966

Un *choropleth* es un mapa que codifica por áreas cierta información mediante la coloración de éstas en función a un determinado valor. En el implementado para este trabajo la coloración de las áreas viene determinada por la cantidad de personas acumuladas dentro del bruhs implementado en el diagrama de barras, que se puede observar en la 2. Para saber el nombre de cierta región y el valor de personas exacto asociado a dicha región se ha implementado un *tooltip*, que es

una ventana emergente según se pasa el cursor por encima de cualquier elemento que devuelve el valor requerido, como, también se puede observar en la 2.

Una vez que se ha explorado el conjunto de nombres por región y año para tener una primera noción del *dataset*, se le debe mostrar al usuario los nombres entre los que puede elegir para realizar la tarea de la desambiguación. Para ello se mostrará una tabla a la que habrá que determinarle dos dígitos a modo de entrada. Esto se debe a que los nombres de personas vienen codificadas mediante un número para facilitar su tratamiento, por lo que al introducir dos valores numéricos se mostraran todos los nombres con un identificador asociado dentro de dicho rango. En la Figura 3 se pueden ver los resultados se los nombres de la búsqueda asociada a los valores 350 y 750.

1 to 25 of 400 entries Filter ?

index	Persona	Apariciones
0	Keene Cullen	1
1	Robert Davye	1
2	John Sephton	4
3	William Roen	2
4	Bartholmew walsh	1
5	John Copland	1
6	Richard Delahide	7
7	Mulmory None	1
8	Brian modder mc Quire	1
9	Patrick Luttrell	1
10	Richard Nugent	8
11	Mathew Brenan	1
12	Margret Rely	1
13	Hue Rely	1
14	Owen Boylan	1
15	Donogh mc Brian	1
16	Brian Ocadan	1
17	Donogh mc Gillerroe	1
18	Owen o Brislan	1
19	Loughlin mc Caffery	1
20	Katherin mc Caffery	1
21	William Cole	11
22	Cahill mc Scullock	1
23	William Parsons	20
24	Jo Borlase	1

Show 25 per page 1 2 10 16

Figura 3: Tabla interactiva de nombres con índices correspondientes al rango de valores que queda entre 350 y 750

4.4.2 Zoom, filtrado y estudio a fondo Una vez que se le han presentado al usuario todos los datos, se le debe dar la opción de seleccionar un pequeño subconjunto de ellos. En este estudio se le van a pedir al usuario dos valores: un nombre de una de las personas del conjunto de datos, el cuál tendrá que introducir manualmente en una casilla prefijada, y un valor numérico que deberá

introducir mediante la modificación de un *slider*, que es un elemento de control que permite filtrar la información mediante un indicador que se mueve a través de una barra, generalmente de forma horizontal. Dicho valor estará entre 0 y 1 y se corresponderá con mínimo ratio de Levenshtein al que estarán los nombres del subconjunto que se va a representar con respecto al nombre escrito de forma manual.

Figura 4: Formato en el que el usuario deberá introducir la información para realizar el filtrado de los datos.

Tras introducir el nombre y el ratio correspondientes, los resultados de la búsqueda realizada se presentarán en una tabla antes de pasar al siguiente paso. En dicha tabla aparecerán los nombres que han sido encontrados en función a los parámetros de búsqueda seleccionados y el número de apariciones de cada uno.

4.4.3 Detalles bajo demanda En cuanto a la vista completa del resultado, se puede observar en la Figura 5 que la visualización está formada por 8 gráficas: 1 grafo, dirigido por fuerzas, que modela los nombres y el valor de M en cada una de sus aristas y 7 mapas de calor o *heatmaps*⁵, los cuales modelan uno a uno el valor de las variables utilizadas en la ecuación. Los heatmaps en esta visualización son de menor tamaño y cada uno da una información diferente, a esta técnica se conoce como *small multiples*⁶ y es ampliamente utilizada.

El grafo (vista 1 en Figura 5) ha sido elegido para tratar de modelar las personas que aparecen en el dataset y su relación de cercanía en cuanto a la ecuación (1). En el dataset hay un total de 36584 nombres diferentes, sabiendo que por nombre se hace referencia al resultado de la concatenación que viene dada en los apartados de nombre y apellido de cada persona. Dichos nombres puede hacer referencia a más de una persona, ya que si, por ejemplo, dos nombres iguales aparecen en dos testimonios diferentes, se partirá bajo la suposición de que dicha persona puede no ser la misma. Por lo que el número total de posibles personas diferentes en el dataset asciende a 53922, cada una de ellas representada mediante un nodo diferente. El nombre de cada nodo vendrá dado por el nombre de cada persona y el número del testimonio en el que aparece, el cual será un

⁵“Mapa de calor” es un término relativamente nuevo, pero la práctica de sombrear matrices existe desde hace más de un siglo: Loua (1873) utilizó una matriz de sombreado para visualizar las estadísticas sociales en los distritos de París [29]

⁶Tamién conocida como trellis chart, lattice chart, grid chart, o panel chart. El término snall múltiples fue popularizado por Tufte [27]

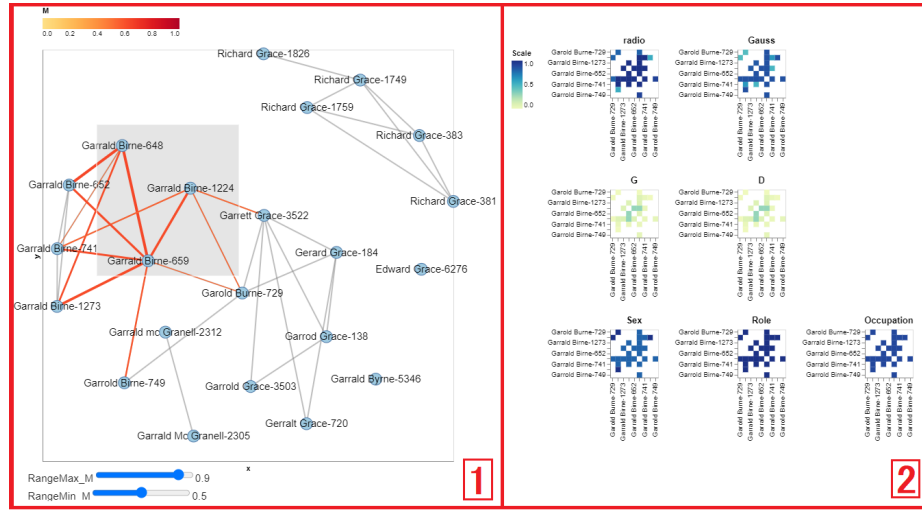


Figura 5: Vista completa creada para la tarea de la desambiguación de nombres sobre el conjunto de datos de “1641 Depositions”. 1) Grafo dirigido por fuerzas y 2) conjunto de *heatmaps*

número de 4 dígitos como máximo, entre 0 y 7011. Entonces, si se tiene el nombre “Garrol Grace” y aparece en el testimonio 3503, entre otros, el nombre que aparecerá en el nodo será “Garrol Grace-3503”, como se puede ver en el ejemplo de 5.

La primera técnica visual que se ha aplicado para codificar dicha información es la coloración de la arista en función del valor $M_{i,j}$ que toma el par de nodos (nombres) que une. Se ha seleccionado una escala de colores cálidos, como se puede observar en la Figura 5, codificando los valores de $M_{i,j}$ más altos con los colores más cálidos de la escala y los más bajos con los menos cálidos. La segunda técnica visual aplicada, la cual también se puede observar en la figura 5, vendrá determinada en la anchura de la arista, la cual será, de forma intuitiva, más ancha cuanto mayor sea el valor de $M_{i,j}$ que codifica y viceversa. Al contrario que la escala de colores, que resultará la misma sea cual sea el grafo, la anchura de la arista no será fija para todas las posibles combinaciones de visualizaciones. Es decir, si tenemos una arista de cierta anchura en una visualización e introducimos otra nueva arista, la anchura de la primera deberá adecuarse al nuevo orden que ocupa en función a la arista introducida.

La ecuación (1) es aplicada para cada par de nombres, por lo tanto, todos los nodos pueden estar conectados entre sí mediante una arista, por lo que se ha decidido implementar dos sliders. El slider de arriba hace referencia a la cota superior del intervalo y el de abajo a la cota inferior. La información que filtran es el valor de la variable $M_{i,j}$, sacando por pantalla todas las aristas asociadas a los valores que estén dentro del rango determinado. La técnica visual de la anchura irá también cambiando en función a los valores de los sliders, es decir,

las anchuras de las aristas se recomputarán en cada iteración con los sliders, ya que al filtrar los datos se añadirán o se quitarán aristas de la visualización y las que ya estaban ocuparan un nuevo orden con respecto a sus anchuras, que deberán ser modificadas visualmente. Las vistas de la Figura 5 están enlazadas por lo que cualquier variación en los sliders del grafo supondrá la variación del subconjunto de datos mostrados en los heatmaps.

El *brush* es una técnica de filtrado que va a permitir realizar selecciones de subconjuntos de nodos que se representen en la visualización, siempre y cuando al menos una de las aristas que los une con el resto quede dentro del rango del slider. El brush en concreto genera un rectángulo de dimensiones modificables con el que quedarán seleccionados todos los nodos que queden dentro de éste. Cuando un nodo queda dentro del brush, se sabrá que ha sido seleccionado ya que éste quedará coloreado, así como todas sus aristas adyacentes, mientras que el resto de nodos y aristas no seleccionados tomarán un color gris. En la vista 1 de la Figura 5 se puede ver un caso particular del brush comentado, enfocado únicamente en 3 de los nodos la visualización. De nuevo, las vistas de los heatmaps varían en función a dicho selector aplicado en el grafo.

Esta técnica es especialmente útil ya que si se quiere analizar un único nodo (nombre) o un subconjunto más pequeño aun de ellos, permite hacerlo de forma fácil e intuitiva.

Cada uno de los *heatmaps* (vista 2 de la Figura 5) codifica una de las variables de la siguiente forma: en los ejes vienen determinados los identificadores de cada persona (nombre más la declaración en la que aparece) y en la coloración de la celda de cada par de nombres correspondiente el valor de la variable a la que hagan referencia. Al pasar el cursor sobre cada uno de los cuadrados aparecerá un tooltip similar al del choropleth la Figura 2, mostrando por pantalla el resultado numérico exacto.

5. Caso de estudio

Se va realizar un caso de estudio con el objetivo de mostrar la aplicabilidad del proceso creado para la desambiguación de dos nombres en el conjunto de las Declaraciones de 1641. Hay que tener en cuenta que la solución creada debe facilitar a los historiadores la tarea de desambiguar nombres en el *dataset*, lo cual será tratado de demostrar en las explicaciones que vienen a continuación sobre el procedimiento que se debe llevar a cabo.

5.1 Mostrar lo importante

El paso correspondiente a mostrar la información relevante será omitida, ya que no suponen una dificultad mayor por la que deban ser explicadas en este artículo. Únicamente se determinará que el nombre seleccionado para realizar la tarea de la desambiguación será el de Anthony Willoughby.

5.2 Zoom, filtrado y estudio a fondo

Una vez introducidos el nombre y el ratio en los campos correspondientes a la Figura 4, aparecerá un resumen de la Figura 6, sobre los nombres resultantes de dicha búsqueda. En el caso de Anthony Willoughby y ratio 0.76 se puede observar que aparecen 4 resultados diferentes, de entre los que caben destacar Anthony willoughby y Anthony Willoby. Son nombres muy similares con respecto al seleccionado, lo que incita a pensar que podrían ser posibles errores de transcripción. Se deberá continuar con el estudio para saber si, en efecto, se cumple dicha suposición, si no se cumple bajo ningún criterio o si no se puede llegar a un resultado concluyente debido a la falta de información. Así, se podrá certificar, por ejemplo, de que el paso de la digitalización de los manuscritos no ha sufrido errores o de que el nombre, a lo largo del tiempo, no ha sido modificado en la base de datos.

1 to 4 of 4 entries ?

index	Nombres	Apariciones
0	Anthony Willoughby	4
1	Anthony Colclough	5
2	Anthony willoughby	1
3	Anthony Willoby	1

Show per page

Figura 6: Resultados del filtrado aplicado.

5.3 Detalles bajo demanda

Al ejecutar el código correspondiente a las visualizaciones finales y tras filtrar los valores más altos de $M_{i,j}$, para valores iguales o superiores a 0.75, se obtienen los dos grafos de la Figura 7. Se le ha aplicado también un brush sobre los cuatro nombres que siguen quedando completamente conectados tras realizar el filtrado comentado. El hecho de que formen un grafo completamente conectado sobre valores altos de $M_{i,j}$ indica que hay una probabilidad alta de que se puedan considerar como la misma persona, sin necesidad de mirar el resto de variables en los *heatmaps*. Esto se debe a que se puede hacer una relación en cadena circular cerrada, es decir, se podrían ir concatenando suposiciones de coincidencia partiendo de un nombre y llegar de nuevo hasta el nombre de partida. Dicha suposición en cadena se cumple para cualquier combinación posible, de entre todas las existentes, inicializando dicha cadena en cualquiera de los 4 nombres. Un ejemplo de cadena sería el siguiente: se podría partir del nodo “Anthony Willoughby-5119” y mediante el valor de la arista que lo une con el nodo “Anthony Willoughby-5128” suponer que hacen referencia a la misma persona. Siguiendo de nuevo el mismo razonamiento se podría suponer que “Anthony Willoughby-5128” y “Anthony Willoughby-5123” y que “Anthony Willoughby-5123” y “Anthony Willoughby-544” también son las mismas personas. Finalmente, llegamos a que “Anthony Willoughby-544” y “Anthony Willoughby-5119” (el

nodo de partida) son la misma persona. Se ha creado un círculo cerrado de suposiciones que se cumple para cualquiera que sea el nodo inicial y dicha relación se cumplirá también para las diferentes combinaciones de secuencias que se pueden realizar.

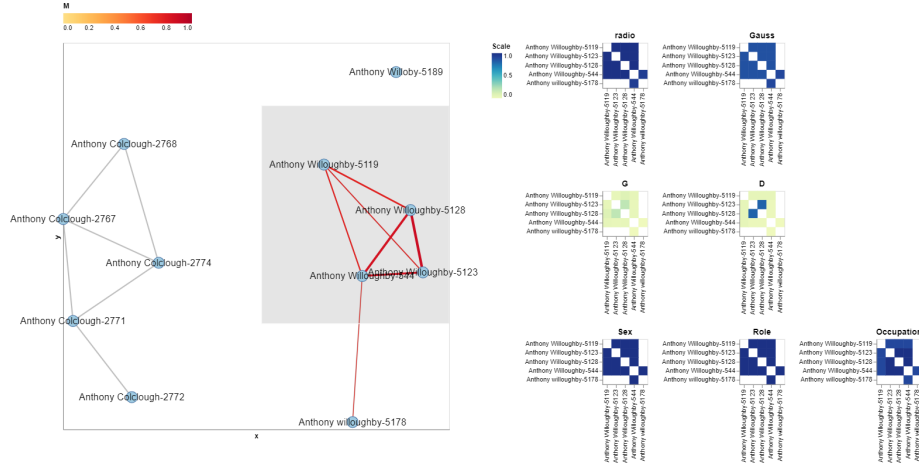


Figura 7: Visualización del grafo y los *heatmaps* con un rango de $M_{i,j}$ entre 0.75 y 1 y un *brush* aplicado.

Si, además, se analizan los *heatmaps* sobre los 4 nodos comentados, se puede observar cómo la variable de Gauss tiene valores superiores a 0.85 en todos los casos, el sexo y el role coinciden como se ha explicado anteriormente y coinciden en 3 casos las ocupaciones también (entre “Anthony Willoughby-5128”, “Anthony Willoughby-5123” y “Anthony Willoughby-544”, a excepción de “Anthony Willoughby-5119”, sobre el que no viene determinada su ocupación, debido a que todos los valores con respecto a los tres restantes son de 0.9.). Las variables G y D únicamente dan valores altos y, por lo tanto, relevantes en el caso entre “Anthony Willoughby-5123” y “Anthony Willoughby-5128”.

Por lo tanto, para este caso se podría suponer con una probabilidad alta que los cuatro nombres sobre los que se ha aplicado el *brush* hacen referencia a la misma persona. Se cumple el mismo patrón para todos los nodos con el nombre Anthony Colclough, con valores de $M_{i,j}$ superiores o iguales a 0.7. Pese a no ser un caso tan claro como el primero analizado, también se podría suponer que dichos nodos hacen referencia a la misma persona.

Antes de determinar si las dos personas definidas hasta el momento podrían unificarse o no, se va a tratar de analizar si los nodos “Anthony willoughby-5178” y “Anthony Willoby-5189” se pueden determinar como una dichas personas.

Para ello se debe bajar el valor del slider correspondiente a la cota inferior de $M_{i,j}$ hasta que “Anthony willoughby-5178” y “Anthony Willoby-5189” estén enlazadas con alguno de los grupos de nodos desambiguados hasta el momento.

Para “Anthony willoughby-5178” se puede ver en la Figura 8 que es para un valor igual a 0.75 y para “Anthony Willoby-5189” un valor igual al 0.7 mostrado en la Figura 8.

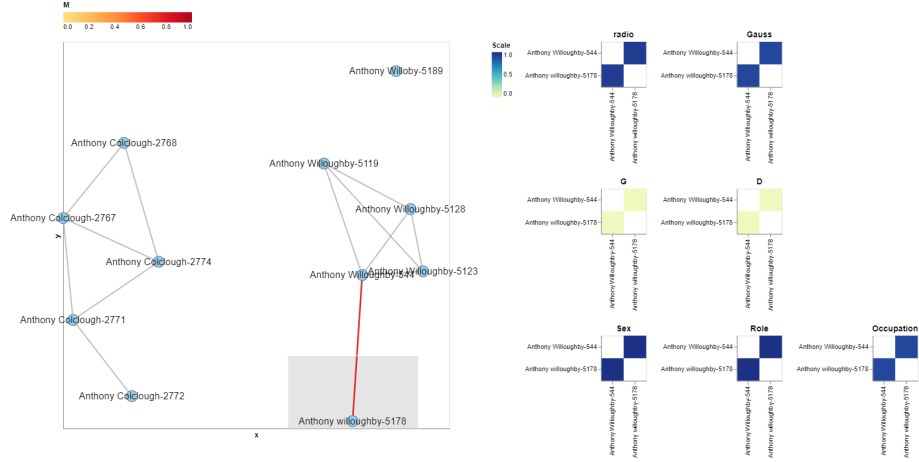


Figura 8: Visualización del grafo y los *heatmaps* con un rango de $M_{i,j}$ entre 0.75 y 1 y un *brush* sobre el nodo “Anthony willoughby-5178”.

Para el caso de “Anthony willoughby-5178”, si se consigue desambiguar el caso mostrado en el brush de la Figura 8, el cual únicamente hace referencia a la relación entre dicho nodo y “Anthony Willoughby-544”, se podría suponer que dicho nodo y el grupo de nodos que hacen referencia al nombre de “Anthony Willoughby” hacen referencia a la misma persona, debido a la relación de “Anthony Willoughby-544” con el resto de nodos de dicho grupo. Entonces, una vez que se analizan los *heatmaps* correspondientes siguiendo la misma estrategia aplicada a las ambigüaciones realizadas hasta el momento se puede determinar que ambos nombres hacen referencia a la misma persona. En particular se ha conseguido determinar que había un fallo en escritura del nombre, de carácter otrográfico, ya que el apellido se debe iniciar también con una letra mayúscula, al igual que el resto de nombres del conjunto de datos.

Con un procedimiento análogo se puede determinar el mismo resultado para el nodo “Anthony Willoby-5189”, dando como resultado una posible detección de un error tipográfico cometido por el deponente a la hora de recoger el nombre.

Finalmente quedaría saber si las dos personas determinadas hasta el momento se podrían unificar o no. Para ello se va a realizar el filtrado de los valores de $M_{i,j}$ partiendo de 0 hasta 0.45. En el resultado del grafo, el cual se puede ver en la Figura 9, se puede ver cómo cada uno de los nodos correspondientes a una de las dos personas se une con todos y cada uno de los nodos de la otra. Esto indica que se puede afirmar que las dos personas desambiguadas hasta el momento no hacen referencia a la misma persona, de forma contundente. Se ha

trazado una línea roja que separa las dos personas y, en efecto, se puede ver como dicha línea corta todas las aristas entre cada par de nodos referentes a una persona cada uno. Como cada grupo de nodos separados por la línea roja hace referencia a una misma persona y el valor de $M_{i,j}$ que se está representando es suficientemente bajo *a priori* como para suponer que los nodos que aparecen unidos en la visualización no hacen referencia a la misma persona, es lógico que las aristas que aparecen unan todos y cada uno de los nodos de un grupo con el resto de nodos del otro.

Para enfatizar la veracidad del resultado y reducir el nivel de incertidumbre que puede surgir en dicha suposición, se deben mirar de nuevo los valores de los *heatmaps*. En efecto, sin necesidad de aplicar el brush y con la única ayuda del *tooltip* se puede ver que en lo único que coinciden cada par de nodos es en el sexo. El resto de variables están asociadas a valores bajos, siendo especialmente destacables los valores en general de la variable de Gauss, que llega en alguno de los casos a alcanzar el valor de 0.38 (“Anthony Colclough-2772” y “Anthony Willoughby-5178”), el valor asociado a la variable del role, que indica que no coinciden en ningún par de nodos y algunos de los valores asociados a la ocupación, que como es el caso entre “Anthony Colclough-2771” y “Anthony Willoughby-5178” no coincidirían.

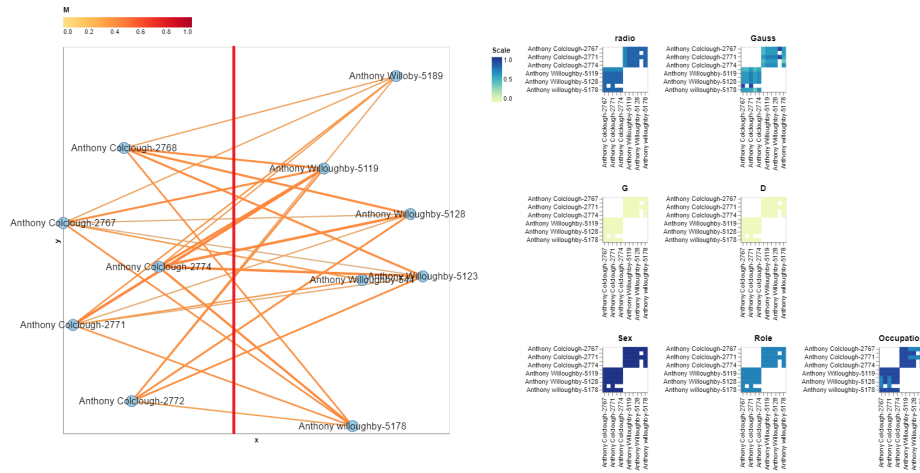


Figura 9: Visualización del grafo y los *heatmaps* con un rango de $M_{i,j}$ entre 0 y 0.45.

6. Conclusiones y trabajo futuro

Se ha obtenido una ecuación que intuye de forma aceptable la probabilidad de que dos nombres hagan referencia a la misma persona con respecto al criterio

del autor. Si bien es cierto que los parámetros son modificables, la base construida sobre las variables seleccionadas ha resultado ser sólida y ha permitido variar dichos parámetros hasta ajustarlos sobre unos resultados deseados. Una vez ajustados, las comprobaciones para cada par de nombres realizadas dieron resultados favorables o acordes al criterio, una vez más, del autor del trabajo. Por lo tanto, se tuvo una buena base matemática sobre la que tratar de construir la solución visual, para que los resultados fuesen lo más precisos posibles.

En cuanto a las visualizaciones seleccionadas para presentar la base de datos y para dar solución al problema de la desambiguación, resulten de forma efectiva las tareas para las que han sido seleccionadas. En particular, la representación en forma de grafo resulta de gran ayuda a la hora de transmitir la información del valor obtenido en la ecuación (1). Se ha comprobado cómo se pueden ver relaciones de nombres que hacen referencia a la misma persona solo con fijar la vista en las relaciones que cumplen los nodos dentro de los grafos, como la que constituyen los grafos completamente conectados en los valores filtrados de $M_{i,j}$ altos. Al ser un trabajo centrado en transmitir esta información más allá de un único resultado numérico, se ha podido comprobar también cómo los heatmaps, no solo ayudan al usuario no experto a entender de donde sale el valor final de $M_{i,j}$, si no que también aportan información sobre los nombres que se están tratando de desambiguar al instante, lo cual puede resultar de gran ayuda.

Independientemente del resultado numérico, se ha realizado una combinación de vistas que permiten al usuario obtener toda la información, en principio, necesaria para desambiguar un nombre con respecto a otro o a más de uno a la vez. En consecuencia, siguiendo la línea de las DH explicada en la introducción, este trabajo aparentemente cumple los requisitos de aplicar soluciones facilitadas por una computadora a la exploración de bases de datos históricas, estando a la espera de verificación por expertos en el campo de las humanidades.

Dejando a un lado si la tarea ha podido ser realizada con éxito o no, se debe tener en cuenta que los campos que han intervenido en la investigación realizada son demasiado amplios como para suponer que en un periodo de tiempo tan reducido se ha conseguido obtener la información necesaria de cada uno de ellos para que la solución propuesta haya resultado óptima. En conclusión, la línea de trabajo principal será la de tratar de mejorar el proceso completo, analizando e investigando las áreas de los diversos campos comentados, con el objetivo de obtener un resultado cada vez mejor.

Referencias

- [1] “Diccionario de la Real Academia Española”. <https://dle.rae.es/>. Última entrada: 07-07-2020.
- [2] “Introducción a las DH”. <https://libguides.lib.cuhk.edu.hk/Intro-DH/>. Última entrada: 07-07-2020.
- [3] Alejandro Benito Santos y col. “Data-driven visual performance analysis in soccer: An exploratory prototype”. En: *Frontiers in Psychology* 9 (2018), pág. 2416.

- [4] A Benito-Santos. *Visualización de datos en Humanidades Digitales*. 2016.
- [5] Alejandro Benito-Santos y Roberto Therón Sánchez. “A Data-Driven Introduction to Authors, Readings, and Techniques in Visualization for the Digital Humanities”. En: *IEEE Computer Graphics and Applications* 40.3 (2020), págs. 45-57.
- [6] Alejandro Benito-Santos y Roberto Therón Sánchez. “Cross-domain visual exploration of academic corpora via the latent meaning of user-authored keywords”. En: *IEEE Access* 7 (2019), págs. 98144-98160.
- [7] Kristin A Cook y James J Thomas. *Illuminating the path: The research and development agenda for visual analytics*. Inf. téc. Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2005.
- [8] Alejandro Rodríguez y col. “Intuitive Ontology-Based SPARQL Queries for RDF Data Exploration”. En: *IEEE Access* 7 (2019), págs. 156272-156286.
- [9] Peter F Fisher. “Models of uncertainty in spatial data”. En: *Geographical information systems* 1 (1999), págs. 191-205.
- [10] Juan García y col. “A visual analytics tool for software project structure and relationships among classes”. En: *International Symposium on Smart Graphics*. Springer. 2009, págs. 203-212.
- [11] Antonio Gonzalez y col. “Combined visualization of structural and metric information for software evolution analysis”. En: *Proceedings of the joint international and annual ERCIM workshops on Principles of software evolution (IWPSE) and software evolution (Evol) workshops*. 2009, págs. 25-30.
- [12] Hua Guo, Jeff Huang y David H Laidlaw. “Representing uncertainty in graph edges: An evaluation of paired visual variables”. En: *IEEE transactions on visualization and computer graphics* 21.10 (2015), págs. 1173-1186.
- [13] Edwin L Hutchins, James D Hollan y Donald A Norman. “Direct manipulation interfaces”. En: *Human-computer interaction* 1.4 (1985), págs. 311-338.
- [14] Paul Jaccard. “Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines”. En: *Bull Soc Vaudoise Sci Nat* 37 (1901), págs. 241-272.
- [15] Tomihisa Kamada, Satoru Kawai y col. “An algorithm for drawing general undirected graphs”. En: *Information processing letters* 31.1 (1989), págs. 7-15.
- [16] Daniel Keim y col. “Visual analytics: Definition, process, and challenges”. En: *Information visualization*. Springer, 2008, págs. 154-175.
- [17] Allen Richard Martin y Matthew O Ward. “High-dimensional brushing for interactive exploration of multivariate data.” Tesis de mtría. Worcester Polytechnic Institute., 1995.
- [18] Alex T Pang, Craig M Wittenbrink, Suresh K Lodha y col. “Approaches to uncertainty visualization”. En: *The Visual Computer* 13.8 (1997), págs. 370-390.
- [19] Rodrigo Santamari*****a y Roberto Therón. “Overlapping clustered graphs: Co-authorship networks visuali-

- zation". En: *International Symposium on Smart Graphics*. Springer. 2008, págs. 190-199.
- [20] Rodrigo Santamaría y Roberto Therón. "Treevolution: visual analysis of phylogenetic trees". En: *Bioinformatics* 25.15 (2009), págs. 1970-1971.
 - [21] Rodrigo Santamaría, Roberto Therón y Luis Quintales. "BicOverlapper 2.0: visual analysis for gene expression". En: *Bioinformatics* 30.12 (2014), págs. 1785-1786.
 - [22] Qiaomu Shen y col. "Nameclarifier: A visual analytics system for author name disambiguation". En: *IEEE transactions on visualization and computer graphics* 23.1 (2016), págs. 141-150.
 - [23] Ben Shneiderman. "1.1 direct manipulation: a step beyond programming languages". En: *Sparks of innovation in human-computer interaction* 17 (1993), pág. 1993.
 - [24] Ben Shneiderman. "The eyes have it: A task by data type taxonomy for information visualizations". En: *Proceedings 1996 IEEE symposium on visual languages*. IEEE. 1996, págs. 336-343.
 - [25] Roberto Therón Sánchez y col. "Towards an Uncertainty-Aware Visualization in the Digital Humanities". En: *Informatics*. Vol. 6. 3. Multidisciplinary Digital Publishing Institute. 2019, pág. 31.
 - [26] Roberto Therón y Laura Casares. "Visual analysis of time-motion in basketball games". En: *International Symposium on Smart Graphics*. Springer. 2010, págs. 196-207.
 - [27] Edward R Tufte, Nora Hillman Goeler y Richard Benson. *Envisioning information*. Vol. 126. Graphics press Cheshire, CT, 1990.
 - [28] William Thomas Tutte. "How to draw a graph". En: *Proceedings of the London Mathematical Society* 3.1 (1963), págs. 743-767.
 - [29] Leland Wilkinson y Michael Friendly. "The history of the cluster heat map". En: *The American Statistician* 63.2 (2009), págs. 179-184.
 - [30] Florian Windhager, Saminu Salisu y Eva Mayr. "Exhibiting uncertainty: Visualizing data quality indicators for cultural collections". En: *Informatics*. Vol. 6. 3. Multidisciplinary Digital Publishing Institute. 2019, pág. 29.
 - [31] John K Wright. "Problems in population mapping". En: *Notes on Statistical Mapping With Special Reference to the Mapping of Population Phenomena* (1938), págs. 1-18.

Sistema inteligente en torno al mundo del grafiti

Adilson Isaac González, Angélica González Arrieta, Daniel López Sánchez

Departamento de Informática y Automática, Universidad de Salamanca,
37008 Salamanca, España

adilson.gonzalez@fec.uni.edu.ni, {angelica, lope}@usal.es

Resumen Desde un enfoque de Inteligencia Artificial el reconocimiento de imágenes es uno de los usos más importante de la visión por computadora, el cual permite la capacidad de interpretar lo que una computadora visualiza y poder entre otras cosas clasificarlo. Este trabajo hace una comparativa teórica y práctica de técnicas para el reconocimiento de imágenes, utilizando la arquitectura de red neuronal preentrenada VGG16 y los algoritmos SURF, SIFT y ORB de la librería Open Computer Vision (OpenCV). La implementación de los algoritmos y la red neuronal convolutional preentrenada se hace en un set de imágenes de grafitis¹; dichas imágenes son las mismas que han motivado el trabajo para poder contribuir a esta problemática social, apuntando al desarrollo de un sistema el cual asocie una nueva imagen de grafiti a partir de otras ya registradas indubitadas, es decir de los que se conoce su autoría.

Keywords: Redes convolucionales, aprendizaje profundo, reconocimiento de imágenes, detección de características.

1. Introduction

En las últimas décadas, el grafiti vandálico se ha convertido en un gran problema que se extiende desde las ciudades más grandes a otros lugares. Debido a su creciente prevalencia en muchas áreas y el alto coste de limpieza y el trabajo de prevención de la pintada, se está ante un problema persistente, casi intratable, del que muchos ayuntamientos están haciendo un esfuerzo por erradicarlo.

En la actualidad se registran estos casos asociando la imagen de los grafitis con sus posibles autores y otros detalles de acuerdo a los resultados del peritaje e investigaciones pertinentes de cada caso, sin embargo la identificación y clasificación de los grafitis ha sido hasta el momento trabajo humano y de expertos en la materia, es por ello y con el fin de contribuir a esta problemática se propone un trabajo con un enfoque investigativo el cual permita presentar técnicas para el reconocimiento de imágenes, de tal manera se puedan asociar grafitis a partir de otros ya registrados y de esta manera considerar a sus posibles autores.

Se presenta el uso de redes neuronales artificiales con aprendizaje profundo; particularmente el trabajo propone el empleo de una arquitectura de red neuronal

¹Grafitis pueden verse en: <https://cutt.ly/ppwRimN>

convolucional preentrenada VGG16 y la implementación de los algoritmos de la biblioteca libre de visión artificial mas conocida por sus siglas en inglés OpenCV, en particular:

- SIFT: Scale-Invariant Feature Transform
- SURF: Speeded-Up Robust Features
- ORB: Oriented FAST¹ and Rotated BRIEF²

Las pruebas de los algoritmos se realizaron utilizando imágenes de grafitis haciendo comparativas de las técnicas de la librería OpenCV con la red neuronal pre entrenada VGG16.

La motivación de este trabajo tiene la visión de desarrollar un sistema inteligente capaz de asociar a través de un grafiti a posibles autores a partir de comparativas de otros grafitis indubitados.

La propuesta de este trabajo de fin de máster se presenta en este documento describiendo los objetivos, antecedentes, marco teórico, la metodología utilizada para la implementación de los algoritmos y la arquitectura de red, finalmente la sección de resultados y discusiones.

2. Metodología

Para el desarrollo del trabajo se planteó el objetivo principal de aplicar algoritmos de la librería OpenCV² y una arquitectura de red neuronal artificial para el reconocimiento de grafitis similares o iguales, el cual consideró; aplicar algoritmos SIFT³, SURF⁴ y ORB⁵ para el reconocimiento de grafitis a partir de un set de imágenes de grafitis, utilizar una arquitectura de red neuronal artificial que permita la identificación de grafitis similares o iguales a partir de otros, analizar los resultados obtenidos para la reconocimiento tanto por los algoritmos de la librería OpenCV como por la arquitectura de red neuronal artificial.

2.1 Entorno y herramientas de trabajo

OpenCV Python⁶

- SO Windows 10 Enterprise N 64-bit.
- Microsoft Visual C++. Microsoft Visual Studio 2019.
- Python interpreter 3.8 64 bits.
- CMake 3.17.2 64 bits.
- PyPi Package.
- Visual Studio Code.

¹Features from Accelerated Segment Test.

²Binary Robust Independent Elementary Features.

²OpenCV: Open Computer Visión.

³SIFT: Scale-Invariant Feature Transform.

⁴SURF: Speeded-Up Robust Features.

⁵ORB: Oriented FAST and Rotated BRIEF.

⁶Proceso de Instalacion del entorno <https://drthitirat.wordpress.com/2019/01/20/opencv-python-build-opencv-4-0-1-dev-contrib-non-free-siftsurf-from-sources-on-windows-10-64-bit-os/>

2.2 Conjunto de Datos

El set de datos es el conjunto de grafitis proporcionados por investigadores policiales para desarrollar el trabajo de fin de máster, puede accederse en la dirección: <https://cutt.ly/ppwRimN>.

2.3 Experimentos a realizar

2.3.1 Algoritmos OpenCV Para los algoritmos de la librería OpenCV se consideran los siguientes experimentos:

1. Del set de datos proporcionado se formarán seis pares de grafitis
2. A cada par se aplicará las técnicas SIFT, SURF y ORB
3. la estrategia de matching será FLANN, "Fast Library for Approximate Nearest Neighbors"
4. Se realizará una comparación cuantitativa de los algoritmos de acuerdo a:
 - a) Número de características.
 - b) Número de feature-points.
 - c) Tiempo de descripción de detección de características.
 - d) Tiempo de matching de características.
5. Se buscará un grafiti dentro de todo el set de datos en donde se medirá:
 - a) Tiempo por cada algoritmo en buscar el grafiti más parecido.
 - b) Número de feature-point.
 - c) Número de características.

2.3.2 Arquitectura de red neuronal artificial El mecanismo de trabajo será definido en las siguientes etapas:

1. **Extracción de características:** Esta etapa se emplea una CNN⁷ **VGG16** una red formada por 13 capas convolucionales y 3 densas. La extracción de características se hará dos momentos una para registrar todas las imágenes del set de grafitis y se guarda con cada imagen correspondiente, el segundo momento será para obtener las características de la imagen de consulta.
2. **Comparación:** para determinar cual grafiti es más parecido dentro del set de grafitis dado.
3. **Resultado:** se mostrará el grafiti más similar como resultado de la fase anterior de acuerdo a la métrica de distancia establecida.

Se considerarán los siguientes experimentos:

1. Se buscará 5 imágenes dentro del set de datos aplicando la medida de desimilitud Euclideana.
2. Se medirá tiempo de ejecución según la medida y la desimilitud entre las imágenes a partir de la medida establecida.

⁷CNN: Convolutional Neural Network

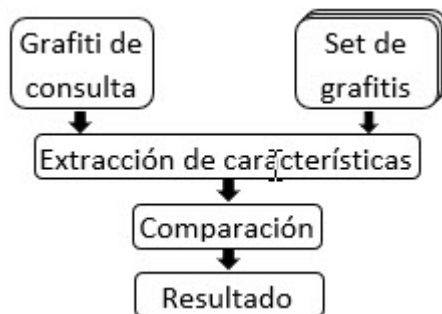


Figura 1: Descripción de metodología con red neuronal artificial preentrenada

3. Trabajos relacionados

La realización del trabajo se basó principalmente en aquellos artículos que describen los algoritmos que se han utilizado para cumplir con el objetivo planteado en éste artículo. SIFT, SURF y ORB de la librería OpenCV y redes neuronales convolucionales profundas.

3.1 ORB: Oriented FAST and Rotated BRIEF

Este algoritmo fue presentado por Ethan Rublee, Vincent Rabaud, Kurt Konolige y Gary R. Bradski en su artículo ORB: An efficient alternative to SIFT or SURF en 2011.

Se basa en el detector de puntos clave FAST [13] y el descriptor BRIEF [G6]; Por este motivo llamado ORB (Oriented FAST y Rotated BRIEF). Ambas técnicas de buen rendimiento y bajo costo.

FAST y sus variantes [12], [2] son el método de elección para encontrar puntos clave en sistemas en tiempo real que coinciden con características visuales, por ejemplo, seguimiento paralelo y mapeo[6]. Es eficiente y encuentra puntos clave de esquina razonables, aunque debe aumentarse con esquemas piramidales para la escala [6], y en nuestro caso, un filtro de esquina de Harris [4] para rechazar bordes y proporcionar una puntuación razonable.

Muchos detectores de puntos clave incluyen un operador de orientación (SIFT y SURF son dos ejemplos), pero FAST no. Hay varias formas de describir la orientación de un punto clave; muchos de estos involucran histogramas de cálculos de gradiente, por ejemplo en SIFT [7] y la aproximación por patrones de bloque en SURF [3]. Estos métodos son computacionalmente exigentes o, en el caso de SURF, producen aproximaciones pobres. El documento de referencia de Rosin [11] ofrece un análisis de varias formas de medir la orientación de las esquinas, y tomamos prestado de su técnica de centroide. A diferencia del operador de orientación en SIFT, que puede tener un valor múltiple en un solo punto clave, el operador centroides da un único resultado dominante.

Descriptores BRIEF [G6] es un descriptor de características reciente que utiliza pruebas binarias simples entre píxeles en un parche de imagen suavizado.

Su rendimiento es similar al SIFT en muchos aspectos, incluida la robustez de la iluminación, el desenfoque y la distorsión de la perspectiva. Sin embargo, es muy sensible a la rotación en el plano.

BRIEF surgió de la investigación que utiliza pruebas binarias para entrenar un conjunto de árboles de clasificación [9]. Una vez entrenados en un conjunto de aproximadamente 500 puntos clave típicos, los árboles se pueden usar para devolver una firma para cualquier punto clave arbitrario [G15]. De manera similar, buscamos las pruebas menos sensibles a la orientación. El método clásico para encontrar pruebas no correlacionadas es el análisis de componentes principales; Por ejemplo, se ha demostrado que PCA para SIFT puede ayudar a eliminar una gran cantidad de información redundante [5]. Sin embargo, el espacio de posibles pruebas binarias es demasiado grande para realizar PCA y en su lugar se utiliza una búsqueda exhaustiva.

Los métodos de vocabulario visual [10], [14] utilizan la agrupación fuera de línea para encontrar ejemplos que no están correlacionados y se pueden utilizar en la correspondencia. Estas técnicas también pueden ser útiles para encontrar pruebas binarias no correlacionadas.

El sistema más cercano a ORB es [8], que propone un punto clave de Harris a escala múltiple y un descriptor de parche orientado. Este descriptor se utiliza para unir imágenes y muestra una buena invariancia rotacional y de escala.

3.1.1 SIFT: Scale-Invariant Feature Transform En 2004, D.Lowe, de la Universidad de Columbia Británica, presentó un nuevo algoritmo, Scale Invariant Feature Transform (SIFT) en su artículo, Distinctive Image Features from Scale-Invariant Keypoints, que extrae puntos clave y calcula sus descriptores.

Hay principalmente cuatro pasos involucrados en el algoritmo SIFT.

1. Detección de escala-espacio extremo
2. Localización de puntos clave
3. Asignación de orientación
4. Descriptor de puntos claves

Detección de escala-espacio extremo: La primera etapa de la computación busca en todas las escalas y ubicaciones de imágenes. Se implementa de manera eficiente mediante el uso de una función de diferencia de Gauss para identificar posibles puntos de interés que son invariables para la escala y la orientación.

Localización de puntos clave: En cada ubicación candidata, se ajusta un modelo detallado para determinar la ubicación y la escala. Los puntos clave se seleccionan en función de las medidas de su estabilidad.

Asignación de orientación: Se asignan una o más orientaciones a cada ubicación de punto clave en función de las direcciones de gradiente de imagen local. Todas las operaciones futuras se realizan en datos de imagen que se han transformado en relación con la orientación, la escala y la ubicación asignadas para cada característica, lo que proporciona invariabilidad a estas transformaciones.

Descriptor de punto clave: los gradientes de imagen locales se miden en la escala seleccionada en la región alrededor de cada punto clave. Estos se transforman en una representación que permite niveles significativos de distorsión de forma local y cambio en la iluminación.

Los puntos clave SIFT son particularmente útiles debido a su carácter distintivo, que permite seleccionar la coincidencia correcta para un punto clave de una gran base de datos de otros puntos clave. Este carácter distintivo se logra ensamblando un vector de alta dimensión que representa los gradientes de imagen dentro de una región local de la imagen. Los puntos clave son invariables para la rotación y escala de la imagen y robustos en un rango sustancial de distorsión afín, adición de ruido y cambio en la iluminación. Se pueden extraer grandes cantidades de puntos clave de imágenes típicas, lo que conduce a la robustez en la extracción de objetos pequeños entre el desorden. El hecho de que los puntos clave se detecten en un rango completo de escalas significa que hay pequeñas características locales disponibles para hacer coincidir objetos pequeños y altamente ocluidos, mientras que los puntos clave grandes funcionan bien para imágenes sujetas a ruido y desenfoco. Su cálculo es eficiente, por lo que se pueden extraer varios miles de puntos clave de una imagen típica con un rendimiento casi en tiempo real en hardware de PC estándar.

El enfoque que hemos descrito utiliza una búsqueda aproximada del vecino más cercano, una transformación de Hough para identificar grupos que coinciden en la pose del objeto, la determinación de la pose de los mínimos cuadrados y la verificación final.

3.1.2 SURF:Speeded Up Robust Features: El descriptor SURF tiene complejidad reducida aún más. El primer paso consiste en fijar una orientación reproducible basada en información de una región circular alrededor del punto de interés. Luego, se construye una región cuadrada alineada a la orientación seleccionada, y se extrae el descriptor SURF de ella.

Para la asignación de orientación, SURF usa respuestas wavelet en dirección horizontal y vertical para una vecindad de tamaño 6s. También se le aplican pesos gaussianos adecuados. Luego se trazan en un espacio como se muestra en la imagen a continuación. La orientación dominante se calcula calculando la suma de todas las respuestas dentro de una ventana de orientación deslizante de ángulo de 60 grados. Lo interesante es que, la respuesta wavelet se puede encontrar usando imágenes integrales muy fácilmente a cualquier escala. Para muchas aplicaciones, la invariancia de rotación no es necesaria, por lo que no es necesario encontrar esta orientación, lo que acelera el proceso. SURF proporciona una funcionalidad llamada Upright-SURF o U-SURF.

Para la descripción de la característica, SURF usa respuestas Wavelet en dirección horizontal y vertical (nuevamente, el uso de imágenes integrales facilita las cosas). Se toma una vecindad de tamaño 20sX20s alrededor del punto clave donde s es el tamaño. Se divide en subregiones 4x4.

Otra mejora importante es el uso del signo de Laplacian (rastros de la matriz de Hesse) para el punto de interés subyacente. No agrega ningún costo de cálculo

ya que ya se calcula durante la detección. El signo del laplaciano distingue las manchas brillantes sobre fondos oscuros de la situación inversa. En la etapa de coincidencia, solo comparamos características si tienen el mismo tipo de contraste (como se muestra en la imagen a continuación). Esta información mínima permite una coincidencia más rápida, sin reducir el rendimiento del descriptor.

3.2 Redes neuronales profundas

3.3 El paradigma del aprendizaje profundo

Las redes neuronales convolucionales profundas Deep Convolutional Neural Networks, DCNN, son quizás las más representativas dentro del panorama actual del deep learning y muchos modelos avanzados se basan en combinar este tipo de redes con otros modelos.

3.4 Redes neuronales convolucionales profundas

Una convolutional Neural Network o también conocida por ConvNet, es uno de los más populares algoritmos para Deep Learning con imágenes y videos. Como otras redes una Convolutional Neural Net-work está compuesta con una capa de entrada, y muchas capas de salidas entre ellas.

Las redes convolucionales usan un tipo de neurona especialmente diseñada para procesar imágenes sin necesidad de convertirlas en un vector. Este tipo de neuronas, toman una imagen como entrada y producen otra imagen a su salida, que suele llamarse mapa de características. En lugar de mediante una serie de pesos, el comportamiento de cada neurona convolucional viene determinado por una serie de matrices llamados filtros de convolución, que suelen ser de tamaño mucho menor que las imágenes de entrada y producen otra imagen a su salida, que suele llamarse mapa de características.

En lugar de mediante una serie de pesos, el comportamiento de cada neurona convolucional viene determinado por una serie de matrices llamados filtros de convolución, que suelen ser de tamaño mucho menor que las imágenes de entrada. Supongamos que tenemos una imagen de entrada I con c canales (normalmente los tres canales RGB) de forma que cada canal se representa como una matriz bidimensional $I_1, I_2, \dots, I_c$. Entonces, una neurona convolucional que opere sobre esta imagen deberá tener c filtros del tamaño deseado (frecuentemente 3×3) que denotaremos como $K_1, K_2, \dots, K_c$. Además, la neurona tienen un valor de bias análogo al de las neuronas convencionales que denotaremos como b . Entonces, la salida de la neurona convolucional serán un mapa de características representado como una matriz M que se calcula como:

$$M = g(b + \sum_c I_c \otimes k_c) \quad (1)$$

donde $\otimes$ denota la operación de convolución que explicaremos más adelante, el bias b se suma a cada elemento de la matriz resultante del sumatorio de convoluciones y $g(\cdot)$ es la función de activación elegida que se aplica elemento a

elemento sobre la matriz que recibe. En el contexto de las redes convolucionales, la función de activación más frecuentemente utilizada es la rectified linear unit o relu:

$$relu(x) = x^+ = \max(0, x) \quad (2)$$

A pesar de su aparente sencillez, la función de activación relu ha demostrado ser muy efectiva y tener numerosas ventajas, siendo hoy en día el estándar de facto en el campo de las redes convolucionales. La figura 2 muestra visualmente el efecto de esta función de activación.

$$relu\left(\text{img}\right) = \text{img}_{\text{relu}}$$

Figura 2: Efecto de la función de activación relu sobre el resultado de una operación de convolución.

En cuanto a la operación de convolución, esta consiste en ir aplicando un filtro sobre la matriz de entrada, desplazando el mismo por toda la superficie de la imagen para obtener la matriz de salida. En cada posición, se multiplican los valores del filtro por los valores de la matriz de entrada que quedan debajo del mismo, para luego sumarlos dando lugar a un único valor que se anota en la matriz de salida de la convolución. A medida que el filtro se desplaza sobre la imagen original, se van llenando los huecos hasta completar la matriz de salida (ver figura 3). Volviendo a la ecuación, vemos que cuando la imagen de entrada tiene varios canales, el filtro de convolución debe estar formado por un número igual de canales, y la salida de la operación es la suma de aplicar cada canal del filtro de convolución a cada canal de la imagen de entrada (ver figura 4).

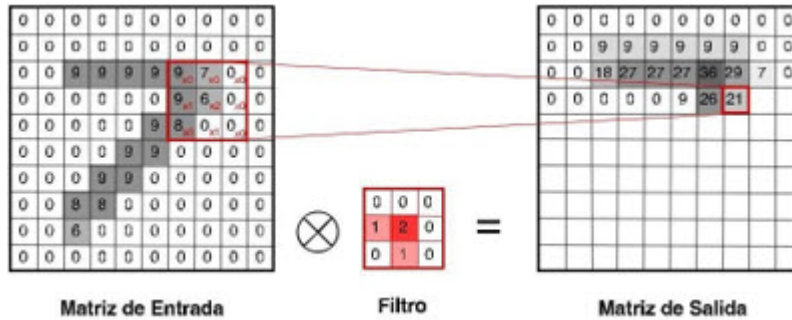


Figura 3: Ejemplo de operación de convolución actuando sobre una imagen de un solo canal. En este caso la matriz de entrada corresponde a una imagen de un dígito escrito a mano en escala de grises. Se usa un filtro 3 x 3.

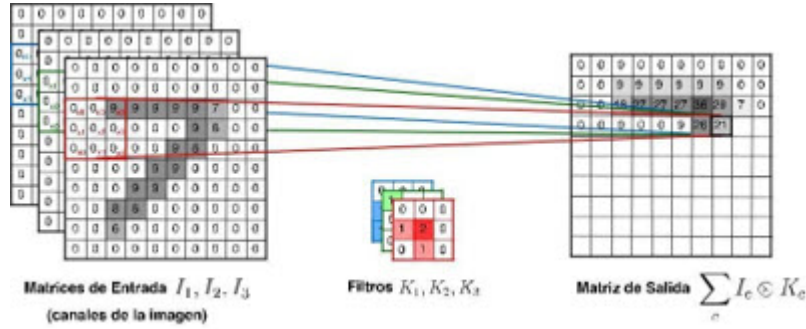


Figura 4: Ejemplo de operación de convolución actuando sobre varios canales. Se usan filtros 3×3 .

Lo mejor para entender la naturaleza de esta operación y su papel en las neuronas convolucionales es visualizar su funcionamiento de forma interactiva⁸.

Si nos fijamos bien, veremos que cuando el filtro se desplaza cerca de los bordes de la matriz de entrada, algunos de los elementos del filtro quedan fuera de la matriz. Existen diferentes opciones para manejar este inconveniente. Una opción es simplemente no desplazar el filtro tan lejos del centro de la imagen como para que los bordes del filtro salgan fuera de los límites de la matriz de entrada. Como resultado, la matriz de salida no tendrá las mismas dimensiones que las de entrada, sino que será ligeramente menor. Otra opción más frecuentemente utilizada consiste en considerar el espacio alrededor de la matriz de entrada como lleno de ceros, de forma que cuando el filtro se aplique parcialmente fuera de la matriz, los valores faltantes serán simplemente completados con ceros, asegurando de esta forma que la matriz de salida tiene el mismo tamaño que la de entrada. A esta técnica se le conoce como zero-padding (ver figura 5).

Otro aspecto que podemos controlar con respecto a la operación de convolución es el salto o desplazamiento que en cada paso realiza el filtro sobre la matriz de entrada. Por defecto, el filtro se mueve una unidad, pasando de esta forma por todas las posiciones de la matriz de entrada. Sin embargo, es posible establecer un salto mayor para el filtro, indicando que debe desplazarse dos o más unidades en cada paso. A este hiper-parámetro se le conoce como stride. Debemos tener en cuenta que usar un valor de stride superior a uno reducirá el tamaño de la matriz resultante de la operación de convolución, al saltarse algunas de las posiciones posibles del filtro sobre la matriz de entrada. Hasta este punto, hemos visto como una neurona convolucional toma a su entrada una imagen, posiblemente con varios canales, y genera a su salida una imagen con un solo canal (mapa de características) en base a una serie de filtros que determinan el comportamiento de la neurona. Tal como las neuronas convencionales, las neuronas convolucionales se agrupan en capas. De esta forma, la salida de una capa de neuronas convolucionales que actúan sobre una imagen será una serie

⁸Página web: <http://cs231n.github.io/convolutional-networks/>

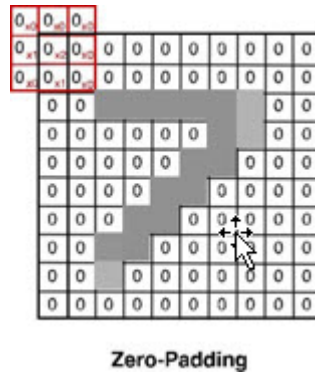


Figura 5: Ejemplo del funcionamiento del zero-padding. Cuando parte del filtro queda fuera de la matriz de entrada, los valores faltantes se consideran como ceros.

mapas de características del mismo tamaño (tantos como neuronas). La figura 6 muestra un ejemplo de imagen de entrada a una capa de convolución y los mapas de características generados por algunas de las neuronas.

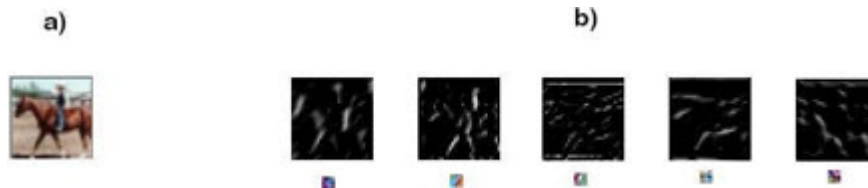


Figura 6: Ejemplo de a) imagen de entrada a una capa de neuronas convolucionales y b) mapas de características generados por algunas de las neuronas y sus filtros correspondientes. Nótese que, dado que las neuronas de la primera capa tienen un filtro por cada canal de entrada, estos pueden ser visualizados juntos como una imagen RGB.

Convenientemente, es posible enlazar varias capas de convolución una tras otra. Como hemos visto, las neuronas de convolución toman como entrada una imagen con un número arbitrario de canales. Por este motivo, es posible considerar los mapas de características generados por una capa como los diferentes canales de una misma imagen y alimentar así la entrada de la siguiente capa convolucional. De esta forma, se van apilando las capas para formar una red convolucional profunda. Además de capas de convolución, se suele hacer uso de otros tipos de capas. En particular, se suelen intercalar las capas de neuronas convolucionales con las llamadas capas de Pooling. El objetivo de las capas de tipo pooling es reducir de forma progresiva el tamaño espacial de las imágenes

procesadas por la red a medida que progresan hacia las capas más profundas. De esta forma, se reduce la cantidad de parámetros que la red debe mantener y con ellos el coste computacional. Además, la paulatina reducción de la dimensión espacial de las imágenes permite a la red extraer características de nivel creciente de abstracción, tal como describimos al principio de la sección. Las capas de pooling operan de forma independiente en cada mapa de características generado por la capa anterior, redimensionándolos a un tamaño menor. Esto se suele conseguir aplicando filtros especiales 2×2 con un stride de 2 unidades sobre los mapas de características, de forma que cada filtro devuelve únicamente el valor máximo encontrado en cada posición. De nuevo, lo mejor para entender esta operación es verla con una imagen, por lo que se recomienda ver la figura 6.

La red neuronal convolucional típica se construye alternando parejas de capas de convolución con capas de tipo max-pooling. Adicionalmente, las redes convolucionales suelen incluir algunas capas finales formadas por neuronas convencionales, que permiten usar la función de activación softmax ya estudiada para emitir predicciones acerca de la categoría a la que pertenecen las imágenes de entrada. En el argot del deep learning a este tipo de capas se les conoce como fully connected. La figura 7 muestra la arquitectura típica de este tipo de redes.

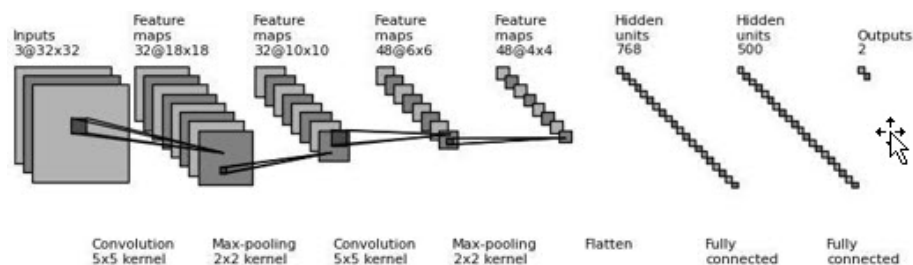


Figura 7: Arquitectura típica de una red neuronal convolucional

Si nos fijamos en la imagen, veremos que las anotaciones de la parte de arriba nos indican la salida de cada una de las capas del modelo. Para los mapas de características se usa la notación $3@32 \times 32$ para indicar que se tienen tres canales, cada uno de los cuales representado por una matriz 32×32 . Las anotaciones de la parte de abajo nos indican el tipo de capa que actúa en cada caso y sus hiperparámetros más importantes.

Finalmente, debemos mencionar una técnica que por su eficacia se ha vuelto esencial para entrenar de forma eficaz modelos del aprendizaje profundo: el dropout [1]. Esta técnica busca evitar el sobre-entrenamiento de las redes, permitiendo una correcta generalización de los modelos. En esencia, la técnica de regularización dropout consiste en fijar a cero la salida de algunas neuronas elegidas al azar durante cada iteración del entrenamiento de la red (ver figura 8). Una vez finalizada la iteración, estas neuronas vuelven a funcionar de forma

normal y se selecciona otro conjunto aleatorio de neuronas para ser desactivadas en la siguiente iteración. De esta forma, se previene la co-adaptación de unas neuronas con otras, manteniendo la generalidad del modelo (es decir, su capacidad de funcionar correctamente sobre datos no vistos durante su entrenamiento). La mayoría de librerías de redes neuronales artificiales nos permiten elegir si deseamos aplicar dropout en cada una de las capas de nuestros modelos, estableciendo en cada caso el porcentaje de neuronas a desactivar en cada iteración. Convenientemente, esta técnica puede aplicarse tanto sobre capas del tipo convolución/pooling como sobre capas de neuronas convencionales.

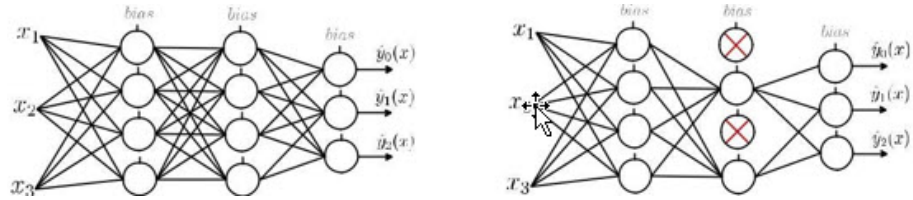


Figura 8: Representación visual del efecto de dropout en la segunda capa oculta de un perceptrón multi-capas.

4. Implementación de algoritmos y arquitectura de red neuronal artificial

4.1 Arquitectura de red neuronal artificial

Siguiendo la metodología definida para éste trabajo se realizará la búsqueda de un grafiti dentro de un set de grafiti.

Para ello se ha utilizado un set de 7 imágenes de grafitis y el grafiti a Buscar (ver figura 9).

El experimento muestra da como resultado la imagen mas parecida a partir de sus características, en esta caso el Grafiti 3.

El segundo experimento considera la búsqueda de un grafiti a partir de una firma. Para ello se ha utilizado el mismo set de experimento 1, que se visualizan en la figura 9 y cuyo resultado se muestra en la figura junto con la firma que se intenta asociar en el set de imágenes 10, mostrando la imagen en donde encuentra la firma.

4.2 Algoritmos de la librería OpenCV

Los experimentos con los algoritmos de la librería OpenCV se han formado seis pares de imágenes de grafitis la cual cada par de imagen es evaluada por cada uno de los tres algoritmos considerados para éste trabajo (SIFT, SURF y ORB).



Figura 9: Búsqueda de una imagen con VGG16



(a) Resultado con VGG16

Figura 10: Resultado búsqueda de una imagen con VGG16

Graffiti	Características
Graffiti 1	60010.945
Graffiti 2	53583.94
Graffiti 3	53386.105
Graffiti 4	57868.99
Graffiti 5	59973.37
Graffiti 7	0
Tiempo	9 seg.

Tabla 1: Resultado búsqueda de una imagen con VGG16



(a) A buscar



(b) Salida

Figura 11: Resultado 2 búsqueda de una imagen con VGG16

Grafiti	Características
Grafiti 1	60010.945
Grafiti 2	53583.94
Grafiti 3	61839.293
Grafiti 4	53386.105
Grafiti 5	57868.99
Grafiti 6	59973.37
Grafiti 7	0
Tiempo	11 seg.

Tabla 2: Resultado 2 búsqueda de una imagen con VGG16

Los pares de imágenes y los resultados al evaluarlas en cada algoritmos se muestran en la figuras de la 12 a la 17. Por cada par de imágenes se muestra por cada algoritmos las características por cada imagen coincidencias y el tiempo.

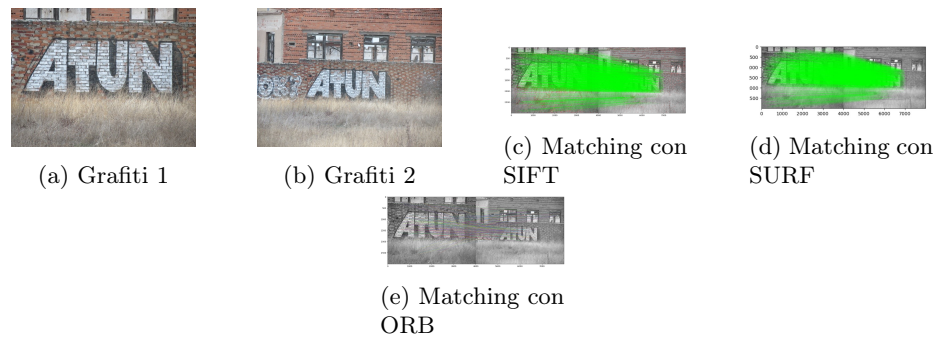


Figura 12: Comparación 1 algoritmos OpenCV

El segundo experimento consiste en buscar un grafiti específico en un set de grafitis con los algoritmos SIFT y SURF. Para ello se ha considerado un set de imágenes que se muestran en la figura 18 junto con el grafiti a buscar.

La figura 19 muestra el resultado de buscar el grafiti más parecido y mostrando además el resultado al hacer comparada con el resto de imágenes, tanto con el algoritmo SIFT como con el algoritmo SURF

Algoritmo	Características		Coincidencias	Tiempo
	Imágen 1	Imágen 2		
SIFT	37516	24470	3406	35
SURF	91618	78435	5224	44
ORB	500	500	500	2

Tabla 3: Resultados comparación 1 algoritmos OpenCV

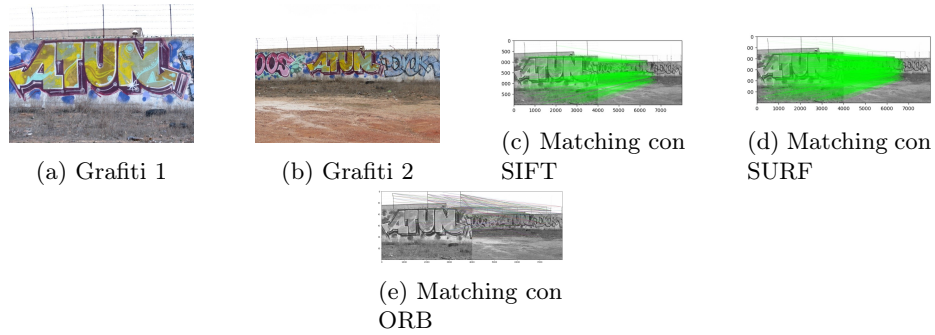


Figura 13: Comparación 2 algoritmos OpenCV

Algoritmo	Características		Coincidencias	Tiempo
	Imágen 1	Imágen 2		
SIFT	7995	37353	1758	29
SURF	41832	66239	3252	41
ORB	500	500	500	1

Tabla 4: Resultados comparación 2 algoritmos OpenCV

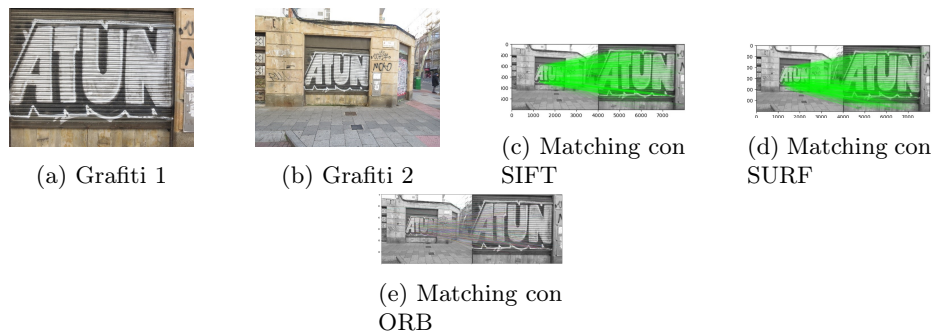


Figura 14: Comparación 3 algoritmos OpenCV

Algoritmo	Características		Coincidencias	Tiempo
	Imágen 1	Imágen 2		
SIFT	36633	17759	2282	48
SURF	64246	78066	2696	186
ORB	500	500	500	1

Tabla 5: Resultados comparación 3 algoritmos OpenCV

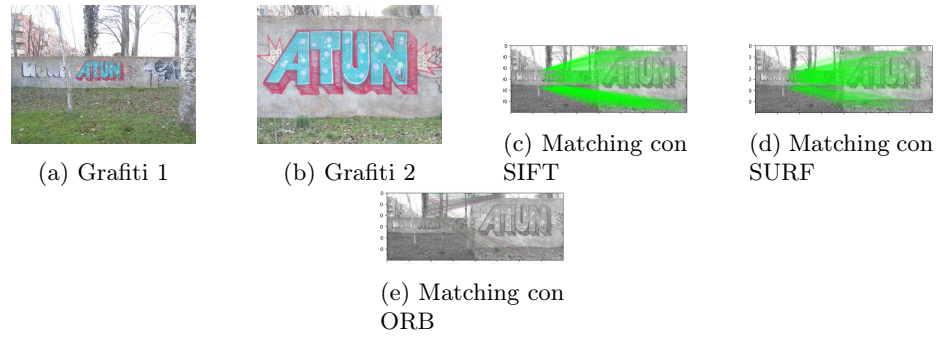


Figura 15: Comparación 4 algoritmos OpenCV

Algoritmo	Características		Coincidencias	Tiempo
	Imágen 1	Imágen 2		
SIFT	95473	25664	2338	174
SURF	97215	69247	2035	231
ORB	500	500	500	1

Tabla 6: Resultados comparación 4 algoritmos OpenCV

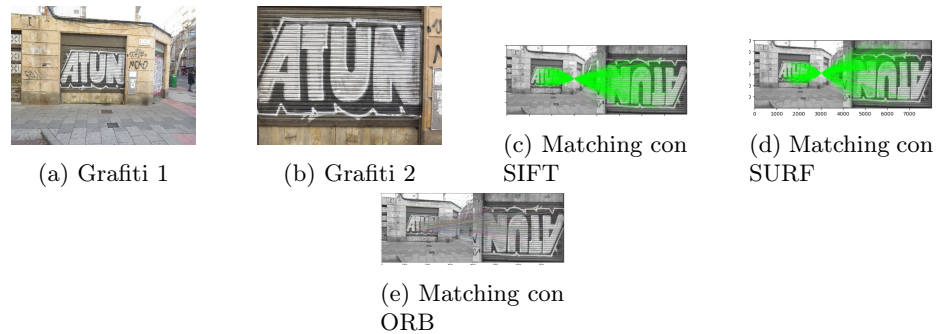


Figura 16: Comparación 5 algoritmos OpenCV

Algoritmo	Características		Coincidencias	Tiempo
	Imágen 1	Imágen 2		
SIFT	36633	17711	2219	50
SURF	64246	78066	2683	180
ORB	500	500	500	1

Tabla 7: Resultados comparación 5 algoritmos OpenCV

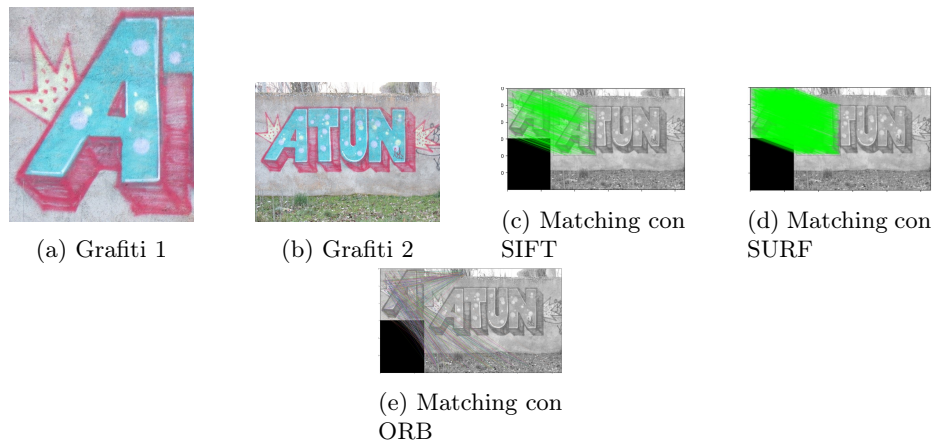


Figura 17: Comparación 6 algoritmos OpenCV

Algoritmo	Características		Coincidencias	Tiempo
	Imágen 1	Imágen 2		
SIFT	2257	25664	1732	46
SURF	9703	69247	7530	28
ORB	500	500	500	1

Tabla 8: Resultados comparación 6 algoritmos OpenCV

Graffiti	Características	
	SIFT	SURF
1	18	126
2	4	148
3	165	255
4	9	143
5	6	101
6	16	136
Tiempo	144 seg.	562 seg.

Tabla 9: Características de búsqueda de una imagen con SIFT y SURF



Figura 18: Búsqueda de una imagen SIFT y SURF



Figura 19: Resultado búsqueda de una imagen SIFT y SURF

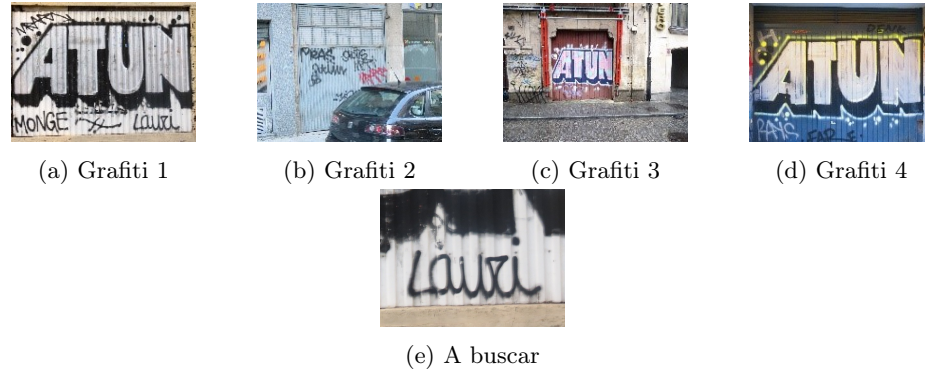


Figura 20: Búsqueda 2 de una imagen SIFT y SURF



Figura 21: Resultado 2 búsqueda de una imagen SIFT y SURF

	Características	
Grafiti	SIFT	SURF
1	1263	4207
2	7	37
3	6	39
4	8	27
Tiempo	41 seg.	63 seg.

Tabla 10: Características de búsqueda 2 de una imagen con SIFT y SURF

5. Conclusiones

Motivado con aportar a la problemática social que genera el grafiti vandálico, se planteó y realizó el presente trabajo, el cual consistió en aplicar los algoritmos: SIFT, SURF y ORB de la librería OpenCV y una arquitectura de red neuronal, el propósito fue comparar grafitis y poder buscar el grafiti más parecido dentro de un set de grafitis proporcionado por autoridades policiales. Entre los tres algoritmos de la librería OpenCV se pudo observar que ORB es el algoritmo más rápido, pero mucho mejor el rendimiento en cuanto a tiempo utilizando la red neuronal artificial preentrenada. Para lograr la asociación de grafitis con la red neuronal artificial se utilizó una red preentrenada VGG16 que permitió obtener el mapa de características por cada imagen y de esta manera hacer una comparación a partir de una distancia euclidiana.

EL trabajo, también de carácter investigativo, permitió profundizar en conocimientos relacionado a redes neuronales convolucionales su arquitectura y mecanismo de trabajo.

Como línea de trabajo futuro se pretende realizar un sistema de información automático que apoye al proceso investigativo que se realizan antes tales actos delictivos, y que a partir de una imagen de grafiti se pueda realizar una búsqueda el cual asocie con otros grafitis ya registrados por la autoridades pertinentes y asociar posibles autores.

Referencias

- [1] Geoffrey E Hinton et al. "Improving neural networks by preventing co-adaptation of feature detectors." En: ". In: *arXiv preprint arXiv:1207.0580* (2012).
- [2] R. Porter E. Rosten y T. Drummond. "Faster and better: A machine learning approach to corner detection". En: *EEE Trans. Pattern Analysis and Machine Intelligence* (2010).
- [3] T. Tuytelaars H. Bay y L. Van Gool. "Surf: Speeded up robust features." En: *International Journal of Computer Vision* (mayo de 2006).
- [4] c. Harris y M. Stephens. "A combined corner and edge detector. " En: *In Alvey Vision Conference* (1988).
- [5] Y. Ke y R. Sukthankar. "Pca-sift: A more distinctive representation for local image descriptors." En: *In Computer Vision and Pattern Recognition* (2004).
- [6] G. Klein y D. Murray. "Improving the agility of keyframebased SLAM." En: *In European Conference on Computer Vision* (2008).
- [7] D. G. Lowe. "Distinctive image features from scale-invariant keypoints." En: *International Journal of Computer Vision* (2004).
- [8] S. Winder M. Brown y R. Szeliski. "Multi-image matching using multi-scale oriented patches." En: *In Computer Vision and Pattern Recognition* (2005).
- [9] V. Lepetit M. Calonder y P. Fua. "Keypoint signatures for fast learning and recognition. " En: *In European Conference on Computer Vision* (2008).
- [10] D. Nister y H. Stewenius. "Scalable recognition with a vocabulary tree." En: *In CVPR* (2006).
- [11] P. L. Rosin. Measuring corner properties. "Computer Vision and Image Understanding". En: "" (1999).
- [12] E. Rosten y T. Drummond. "Machine learning for highspeed corner detection". En: *In European Conference on Computer Vision* (2006).
- [13] E. Rosten y T. Drummond. "Machine learning for highspeed corner detection". En: *Redes de computadoras, 4ta Edicion*. Vol. 1. In European Conference on Computer Vision, 2006.
- [14] J. Sivic y A. Zisserman. "Video google: A text retrieval approach to object matching in videos". En: *International Conference on Computer Vision* (2003).

Design and implementation of a Pipeline for advanced DNA analysis (NSG) by using Machine Learning Techniques

Zakieh Alizadehsani, Juan M. Corchado

Departamento de Informática y Automática,
Facultad de Ciencias, Universidad de Salamanca
Plaza de los Caídos, 37008 Salamanca, España
<http://www.usal.es>

Abstract

DNA sequencing is a lab method to determine the sequence of a DNA molecule. It could be used to discover human diversity and disease. Current sequencing technologies give large amounts of DNA sequence data, which are applied in a wide area of biological applications including disease discovery, genome expression analysis, and detection of sequence variants.

For the disease identification base on NGS data, these technologies use variant calling. The variant is Differences in human DNA sequence, it affects the way human body functions or some of the variant disrupts the body's function. Variant calling finds all type of variants, most of the variants are benign and some of them due to disease. Regarding it, we need approaches to filter out benign and pathogenic variants. Most of the researches applied to many tools to diagnose Chromosomal abnormalities to diagnose disease.

We implement a platform that presents a combined predictor model to detect potentially pathogenic variants. This method attempts to use the best and most effective features to identify disease variants. The features used are divided into two categories, the first one based on the well-known variant pathogenic prediction tools such as SIFT and the second one are biological features. To detect pathogenic variant, we work on DNA bases changes analysis on mutation characterizes. We used clinically significant variants data-set called ClinVar database to train our model. The predictor model, use a configurable Ensemble Strategy to achieve a more accurate model and reduce over fitting. The results demonstrate that our model has reclassified uncertain or not-provided variants as pathogenic or benign.

Keywords: DNA variants, Machine learning, Variant classification, variant pathogenic prediction.

1. Introduction

DNA sequencing is a lab method used to determine the sequence of a DNA molecule. It determines the order of the four bases, i.e. thymine (T), adenine (A), cytosine (C) and guanine (G) in a single DNA strand. DNA sequencing

may be used to determine the sequence of specific genes, large genome regions, the entire chromosome, and the entire genome. Increasing demand to reduce the cost of sequencing has led to the development of high-performance sequencing technologies (the new generation of sequencing) that can sequence thousands or millions of sequences, simultaneously. The new generation of sequencing is capable of rapidly sequencing large pieces of DNA throughout the genome, and with minimal tools, can generate open data at each sequencing stage. New generation sequencing (NGS) methods include a wide range of applications such as genome sequencing.

Specially using machine learning approaches to predict and classify variants are common in rare disease and common genetic conditions. more cancers are being sequenced so if patients have tumor, in most of the time, Clinical staff get a piece of their tumor and then sequence it. But Clinical users usually have not enough time, motivation, and skill to collect, control and work with all variant data, but we can use automatic approaches to prioritize the variants and help to classify tumor correctly and then that can lead to better treatment.

With whole-genome sequencing, researchers can assemble genomes, compare the genome of sample organism to a reference genome, explore the molecular evolution of a species or population, obtain the whole genome the sequence of a single cell, etc. Genome sequencing can be used to represent expressed regions of the genome and can be used to find protein coding. But in this work, we are going to use track pathogen outbreaks application of genome sequencing.

The structure of the this work is as follows. In section 2 the state of the Art has been analyzed. Then, in section 3, the methodologies presented. In section 4 the system implementation details. In section 5, the results obtained are provided. Finally, section 6 concludes the thesis.

2. State of the art

A common strategy used to identify disease base on NGS data includes several technologies. These technologies are applied clinically by their workflows: 1) Blood draw 2) Sequencing 3) Alignment 4) Variant calling 5) Variant annotation 6) Variant filtering 7) Variant Classification

One of the objectives that were explicitly defined in the TFM proposal was to review the existing approaches for the application of machine learning or AI algorithms in identify disease by variant Classification. Regarding it, in this part, we provide a systematic review for variant Classification section.

2.1 Definitions

In this work, there are new concepts, most of them are Biology concepts. So, first of all we review those new concepts for computer science researchers.

The first concept is DNA Sequencing, The normal structure of DNA includes the four bases of thymine (T), adenine (A), cytosine (C) and guanine (G). DNA sequencing is the determination of the physical order of these bases in a DNA molecule.

Sequencing may be used to determine the sequence of specific genes, large genomics regions, the entire chromosome, and the entire genome. high-performance technologies have been developed to generate data or a new generation of sequencing that can sequence thousands or millions of sequences simultaneously. In this work we are going to focus on variant classification.

the second concept is Variant calling, it is the method of naming genetic variation in sequencing data. For instance, single nucleotide variants (SNV) and structural variants, such as indels and inversions, and fusion genes.

the common format for variant classification is VCF file. VCF file stands for variant call format. It is a simple file tab delimited format. VCF used for storage and description of genomic variants.

VCF file was developed for the 1000 Genomes Project [21], but it was used by other projects. Currently, this format is well-known in bioinformatics field. It can represent variants in a single sample (one person) or multiple samples. VCF is divided into three

2.2 Research Questions

In this study we are going to answer following questions by Systematic review of variant classification.

1. Which technique or algorithms are used for variant classification?
2. What approach are used for scaled data?

After investigate the related work we answered them as follow:

AQ1: Which algorithms are used for variant classification? In general, there are two methods to discover the pathogenic variant. The first one is based on well-known tools that usually give a continuous score to variants. For example, CADD defines 0-1 range for its scoring. These kinds of tools usually use regression algorithms. Other studies usually use SVM, random forest, and deep

learning approaches.

Also, the trend of using algorithms base on year are showed in figure 1, this seems the Deep learning are one of the common algorithms in trend, the primary reason can be the Deep variant [20] Technique by google in 2016. This research was done by Google Brain team [7]. After it most of well-known tools switched from statically analysis approach to this deep learning algorithms. in their tools. Also, biology researchers applied deep learning more than before.

AQ2: What technique are used for scaled data?

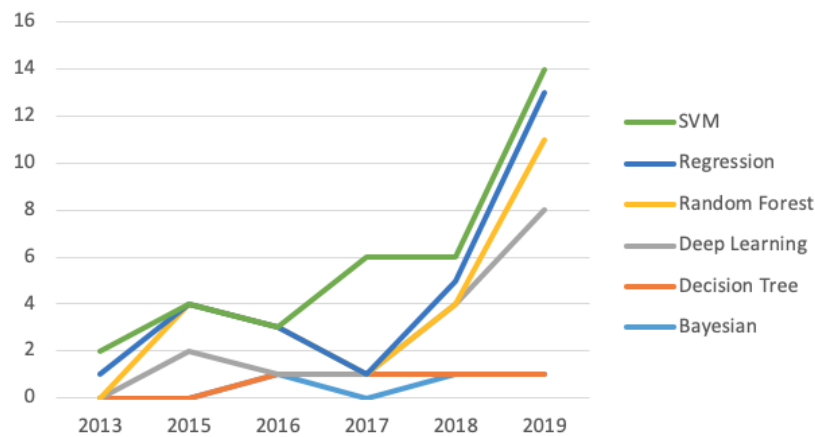


Figura 1: Algorithm base on years

A large number of data in NGS technology leads to handling scaled data. One way to overcome this problem is to use cloud technologies such as EC2 or S3 in amazon services or other cloud company services. Also, Filtering variants base on their quality of reads can be applied to decrease the size of data.

2.3 Related work

In order to analysis genomics data we can have following big picture of process [25].

1. Sample preparation: this step contain of blood draw.
2. Sequencing: Sequencing may be used to determine the sequence of specific genes.

3. Quality control: After Quality control phase, if reference genome is available next step is alignment NGS data base on existing reference. But if it is new genome, we have to assembly it.
4. Alignment: Methods of arranging RNA, DNA, and Protein sequences.

After Alignment there are several applications for alimented genome data such as:

- Genome expression analysis.
- Peak finding.
- Variant detection.

Most of these tools provide prioritizing causal variants to boost discovery power. Most of biomedical tools have a similar theoretical basis [22]. For instance in SIFT [16] and PolyPhen [1] they just consider protein-coding metrics, but as we mentioned before Non-coding DNA can play important role in some disease. Also, these method use known pathogenic mutations and cannot accept for all cases. Regarding these challenges there are some models that used above tools to provide powerful features to train ML or AI models [27], [5]. One of current methods limitation is extract a single information type. But, CADD model [11] consider this issue and provided Combined annotation-dependent depletion [8]. In CADD model, they used huge dataset to variants and identification of pathogenic mutations. As we know SVM is one of machine learning algorithms that work well by huge data. they applied SVM algorithm on high-frequency human-derived alleles from 14.7 million simulated variants. CADD used Ensemble learning[2]. In Ensemble strategy we can combining the.[24], [17], [12].

As we know if there are several tips to select classifier algorithm, nonlinear classifier can make over-fit the data or linear provide classifier under-fit the data[9], [10]. Deep neural networks approaches are good model to cover none-liner data [18], [19]. Investigating the Bias and Variance of data can help us to select linear or nonlinear classifier. Also, this paper [23] provided valuable methods to estimate classification complexity.

Regarding CADD used a linear kernel support vector machine, its limitation is that it does not capture non-linear relationships, DANN [21] use a deep neural network, which is capable of detecting non-linear relationships. they introduced DANN for annotation of variants and identification of pathogenic mutations. Based on the algorithm of CADD. The only different is they trained deep neural network model.

As we know deep learning are proper for domain problems with large and complex data. There are some other models that use sequence-based deep learning. There are some other models that applied deep learning and NN on their study [15], [26].

Multiple models or multiple algorithms to get better performance in predictions. CADD trains SVM on 10 different subsets. This model can detect selected input and outputs. But the problem of CADD is it cannot detect non-linear relationships [4].

In [13] they investigate to annotated effect on prediction results. We had several options to annotate our data. AnnoVar, snpEff “ SNP effect prediction tool. - and VEP- Variant Effect Predictor -are the most common tools for variant annotation.

In [14] they provided a comparison between those three tools. Input file type, output, covering coding and -non-coding are the some of features that they investigated.

3. Methodologies

3.1 Proposed Framework

Base on TFM definition we need to provide the UI base application for our pipeline. We implemented web base application base on Django/python. In this platform users can upload the patient information and the related genomics data in VCF format. After upload the required input they can see the statistic result of their VCF file and predict the variants by our classifier. The architect of our platform are displayed in figure 2.

As the interface base presentation user follow this step:

1. Enter the patient information such as name, last name, description of their history and upload the patient VCF file.
2. visit the overview page of input data.
3. predict each variant type in variant detail page.

3.2 Proposed Dataset

We used clinically significant variants dataset called ClinVar database to train our model. The database contains variety of clinical laboratories. Also, some of research groups add their result of observed in their samples.

Our data-set contains 305983 observed variants. The key problem with the variant data-set is unbalanced data. Because there is not enough balance between benign and pathogenic data counts. To overcome data, here we solve this problem by re-sample our data.

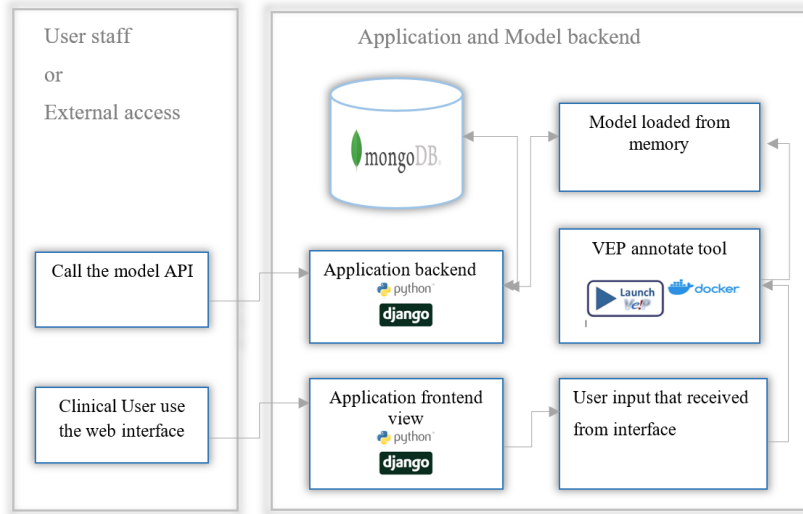


Figura 2: Proposed Framework

3.3 Proposed Learning Model

In this work, after data preprocessing, we have applied some machine learning algorithms. Based on our systematic review, Random forest and deep learning are common algorithms that are applied in related work. As we know decision Tree, random forest and gradient boosting are type of decision maker algorithms. random forest and gradient boosting are ensemble methods, which combines several decision trees to provide accurate prediction. Random forest algorithms contain large number of trees. Also, it uses different samples for training models. In this work we tested several algorithms to select best one. we will discuss about details in Implementation section. Regarding the clinical decision is a sensitive decision making, our main goal was to aim acceptable predict and provide reliable model, so we try to add ensemble learning strategy to reduce miss-classified and improve observations. It can cover weak learning by vote majority.

So, we decided to provide flexible ensemble strategy for our model, this model is configurable to have a one or more training algorithm in it's ensemble. This configuration is embedded in UI interface. So the final design of our model is displayed in figure 3.

3.4 User interface

Base on TFM definition we need to provide interactive application to help user to make better clinical decision. It Base on this requirement we implement

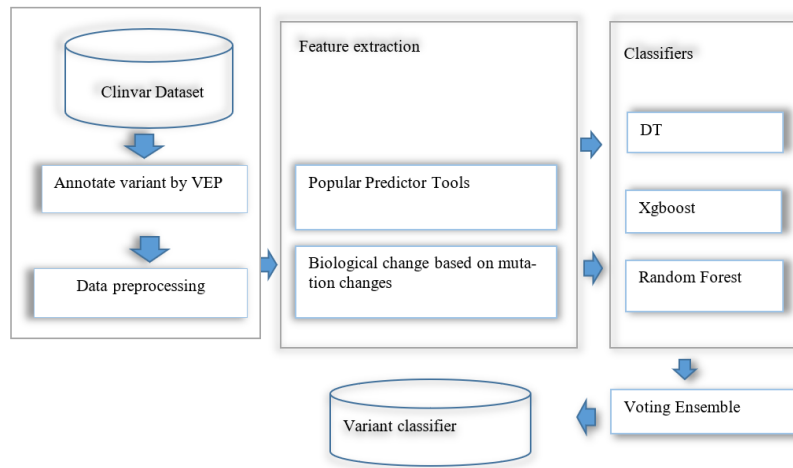


Figura 3: Proposed Learning Model

web base platform. This application used MVC structure. MVC stand on MOdel-View-Controller structure. The most important advantage of this method is that it can provide well-structure platform. learning model and classes that working with database are put in Model section. The user input will handle by controller section. Finally View are responsible for user interactive part.

4. Proposed System

In this section we are going to discuss about implementation details and technology which used in our work. The structure separated to five part:

- Proposed variant annotation procedure
- Feature extraction
- Learning algorithms
- Web application technologies
- Cloud base usage and Container technologies

4.1 Proposed Variant Annotation Procedure

One of important things in machine learning is providing informative training data. The common solution for gaining informative data is using annotating tools. Annotation data provide a powerful effect to train data and due to improve the disease diagnosis. But select high-reliable tools are important because Incorrect or incomplete annotations cause models to train incorrectly and provide false positive or false negative results.

In variant calling, there are two type of variants:

- Rare variants of large effect,
- Common variants of weak effect,

Generally, variant can happen in two section of DNA:

- protein-coding: A few parts of DNA has critical roles that is make Protein. Mutation variant in this part can cause a changing amino acid sequence and provide large effect on human health.
- Non-coding DNA: Around 99 of DNA. Protein coding genes alone make up a small portion of the human genome, and most genomics sequences are Non-coding DNA. They have some important task such as translating and transcribing protein-coding sequence rules or control of gene activity. Recent studies emphasize the increasing role of these Non-coding DNA in the parthenogenesis of various diseases, and challenge the fact that protein-encoding genes are the only factor in the development of human disease.

So, providing proper variant annotation and variant prioritization can be help to dis-ease gene discovery.

Most of these tools provide prioritizing causal variants to boost discovery power. For instance in SIFT [16] and PolyPhen [1] they just consider protein-coding metrics, but as we mentioned before Non-coding DNA can play important role in

some dis-ease. Also, these method use known pathogenic mutations and cannot accept for all cases. Regarding these challenges there are some models that used above tools to provide powerful features to train ML or AI models. One of current methods limitation is extract a single information type. But, CADD model [11] consider this issue and provided Combined annotation-dependent depletion. In CADD model, they used huge data-set to variants and identification of pathogenic mutations. As we know SVM is one of machine learning algorithms that work well by huge data. they applied SVM algorithm on high-frequency human-derived alleles from 14.7 million simulated variants. In [13] they investigate to annotated effect on prediction results. We had several options to annotate our data. AnnoVar, snpEff and VEP prediction tool. In this work we use the VEP command tools to anotate our data.

4.2 Feature Extraction

By ignoring the irrelevant variables, or selecting the ones that improve accuracy, we reduce the amount of strain on the system and produce better results. In previous section, we annotated our data with VEP tools. In this part we are going to do Feature extraction.

After annotation we have several features, we have number of features, so we need Feature selection, this method attempts to use the best and most effective features to identify disease variants. The features used are divided into two categories, the first one based on the in-silico prediction tools such as Polyphen, SIFT, CADD and the second one is biological features. To detect pathogenic variant features, we work on DNA bases changes analysis on codon, amino-acids, and mutation characterizes.

4.2.1 Tools Score features Most of these tools provide prioritizing causal variants to boost discovery power. Regarding Variant Annotation section, In our work we are going to consider coding and non-coding variant, so, we need to select scoring tools from both categories. We selected SIFT, CADD and polyphen. In order annotating variants, we used VEP command tools base on [6] instruction

4.2.2 Biological features In this part we are going discusses about biological features, there are several aspects in biological that can due to disease. For example, we have following categories: 1)Rare variants of large effect2) Common variants of weak effect Also, we have, Coding or non-coding variant that has different effect. Following biological features are selected to provide informative features for our model.

Allele frequency: Allele frequency refers to how common an allele is in a population. It is determined by counting how many times the allele appears in the

population then dividing by the total number of copies of the gene. Also, this paper investigate [3] impact of rare and low-frequency genetic variants in common disease.

In our data we use allele frequency from 3 database:

Also, this paper investigate [3] impact of rare and low-frequency genetic variants

$$\text{Allele frequency} = \frac{\text{number of copies of allele } A \text{ in population}}{\text{Total number of copies of gene } A \text{ in population}}$$

in common disease.

In our data we use allele frequency from 3 database:

- AF_ESP: This field display Allele frequencies or Gene frequency from GO-ESP, Go ESP is NHLBI GO Exome Sequencing Project (ESP) that includes genes and mechanisms contributing to heart, lung and blood disorders.
- AF_EXAC: Allele frequencies from ExAC. ExAC database includes the frequency of the alternative allele, which is usually but not always the minor allele
- AF_TGP: Allele frequencies from the 1000 genomes project

Introns and Exons: Generally, variant can happen in two section of DNA: protein-coding and Non-coding DNA. protein-coding do main job to made Protein. As other concept, we have Exon and Intron. Introns are nucleotide sequences which do not impact on protein creation, But Exon effect directly on protein creation process. So we extract Boolean variable as Is_Exon. So, regarding importance of Exon section we provide more features base on Ex-on length and Exon position. The figure 4 display the position of pathogenic class base on Exon length and Exon position.

Codons: most effect of disease caused by stop mutation. It will disrupt function of protein and due to diseases. Regarding it, we provide machine understandable feature base on codon change.

Amino_acids: This field contains variant affects the protein-coding sequence. It contains two part. This column is important for photogenic prediction. At the end Base on feature correlation we reduction the selected feature.

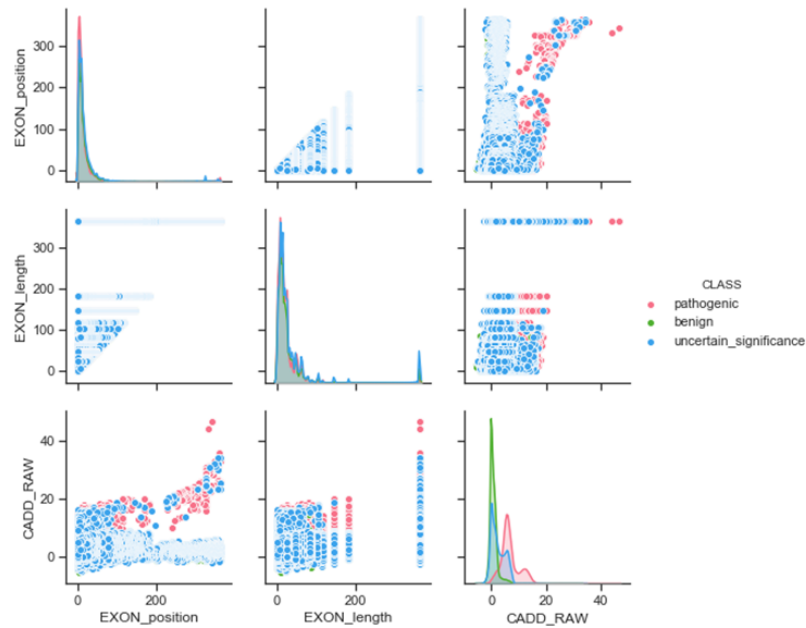


Figura 4: Exon and Calss correlation

4.3 Learning algorithms

In this work, after data preprocessing, we have applied some machine learning algorithms. Based on our systematic review, Random forest and deep learning are common algorithms that are applied in papers. As we know deep learning is strong approach for most of domain problems, but it needs large Data-set, for example in DANN they provided huge real and simulated data for training their deep learning model, but regarding we have annotating step in our work, So providing huge data-set needs more resource. So, in this work we focus on machine learning algorithms.

4.4 Web Application Technologies

Base on TFM definition we need to provide the UI base application for our pipeline. We implemented web base application base on Django/python. In this platform users can upload the patient information and the related genomics data in VCF format. After upload the required input they can see the statistic result of their VCF file and predict the variants by our classifier.

As we know clinical staff needs to overview all of related data to variant, so we have provided variant detail page that display the variant information, Also user in this page can predict the type of variant.

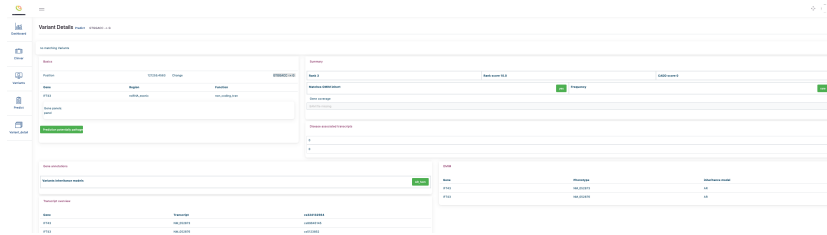


Figure 5: Statistics visualization of VCF dataset

4.5 Cloud base usage and Container technologies

NGS provide huge data that are categorised as high outputs. In order handle this data, we used cloud base services and NOSQL databases, Also to provide fast installation of tools we used docker containers.

5. Results

One of well-known challenges for genomics data is unbalanced data, In unbalanced data for best classifier we obtained 82 % accuracy, but as we know, the Accuracy parameter is a tempty parameter, so according to the values of recall and f1-score, we needed to improve the model. So we re-sample the data and we got 85 % accuracy, it seems it is almost the same as before , but an f1-score and recall increased for all classes.

*****Balanced Vriant data Result*****					
=====					
Decision Tree Classifier Detail :					
	precision	recall	f1-score	support	
benign	0.86	0.83	0.85	29963	
pathogenic	0.77	0.88	0.82	29963	
uncertain_significance	0.74	0.65	0.69	29963	
accuracy			0.79	89889	
macro avg	0.79	0.79	0.79	89889	
weighted avg	0.79	0.79	0.79	89889	
=====					
Logistic Regression classifier Detail :					
	precision	recall	f1-score	support	
benign	0.79	0.83	0.81	29963	
pathogenic	0.69	0.82	0.75	29963	
uncertain_significance	0.63	0.48	0.54	29963	
accuracy			0.71	89889	
macro avg	0.70	0.71	0.70	89889	
weighted avg	0.70	0.71	0.70	89889	
=====					
Random forest classifier Detail :					
	precision	recall	f1-score	support	
benign	0.92	0.89	0.90	29963	
pathogenic	0.85	0.85	0.85	29963	
uncertain_significance	0.78	0.80	0.79	29963	
accuracy			0.85	89889	
macro avg	0.85	0.85	0.85	89889	
weighted avg	0.85	0.85	0.85	89889	
=====					
Gradient Boost Classifier classifier Detail :					
	precision	recall	f1-score	support	
benign	0.91	0.89	0.90	29963	
pathogenic	0.83	0.88	0.85	29963	
uncertain_significance	0.79	0.76	0.78	29963	
accuracy			0.84	89889	
macro avg	0.84	0.84	0.84	89889	
weighted avg	0.84	0.84	0.84	89889	
=====					
Ensemble Classifier classifier Detail :					
	precision	recall	f1-score	support	
benign	0.91	0.90	0.91	29963	
pathogenic	0.85	0.87	0.86	29963	
uncertain_significance	0.80	0.79	0.80	29963	
accuracy			0.85	89889	
macro avg	0.85	0.85	0.85	89889	
weighted avg	0.85	0.85	0.85	89889	
=====					

Figura 6: Balanced data results

Base on Figure 21, Random forest has best result.

Regarding the clinical decision is a sensitive decision making, our main goal was to aim acceptable predict and provide reliable model, so we try to add en-

semble learning strategy to reduce mis-classified and improve observations. It can cover weak learning by vote majority. We selected ensemble approaches and we obtain almost same result. Base on figure 21 The result did not show the sensitive change but the accuracy for pathogenic and uncertain significant variant are changed.

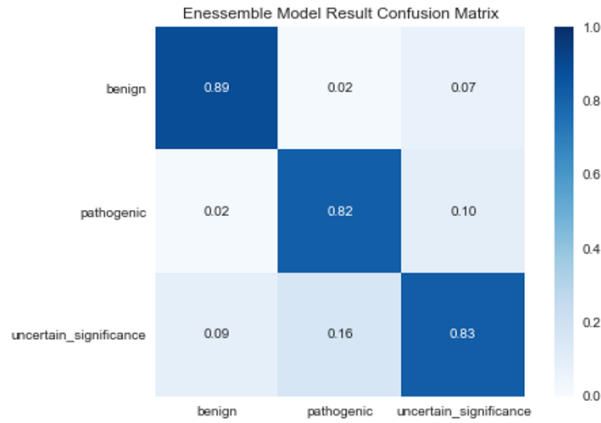


Figura 7: Ensemble model result

6. Conclusions and Future Work

In this section we discuss about the conclusion and the future works that can apply to the proposed structure to improve it.

6.1 Conclusions

In this thesis, We have implemented potential pathogenic variant platform that benefit from machine learning algorithm. This method attempts to use the best and most effective features to identify disease variants. The features used are divided into two categories:

1. variant pathogenic prediction tools:the well-known variant pathogenic prediction tools such as SIFT, CADD, Pholyphen.

2. Biological features.

The main conclusion of the thesis could be listed as follows:

- For increasing the accuracy of the proposed model the CADD predictor has been used that consider functional and non-functional effect of variants.
- To detect pathogenic variant, we work on DNA bases changes analysis on mutation characterizes.
- For training model we have used clinically significant variants data-set called ClinVar database.
- One of main common challenges in genomics data is they are unbalanced data, So we re-sample the data and after it we increase the model metric results.
- In unbalanced data for best classifier we obtained 82% accuracy, but as we know, the accuracy parameter is a tempty parameter, so according to the values of recall and f1-score, we needed to improve the model. So we re-sample the data and we got 85% accuracy, it seems it is almost the same as before , but an f1-score and recall increased for all classes.
- In order to train our model, Base on our systematic literature review, we select trend algorithms related our data. Random forest has best result on our data.
- Regarding the clinical decision is a sensitive decision making, our main goal was to aim acceptable predict and provide reliable model.
- The reason of selecting flexible and configurable data is this model should work in interactive platform so we need fast training approach. regarding this issue we used cloud base platforms to handle big data.
- The results demonstrate that our model has reclassified uncertain or not-provided variants as pathogenic or benign.

6.2 Future Work

Regarding our review, more researchers used random forst and deep learning algorithms. regarding, it, we implemented deep neural network using the Keras library. But the results were not satisfied. The problem is deep neural network will work accrue with large amount data. if we used raw clinvar dataset it was fine. but we had two chalenge:

1-we had one more step for annotating raw clinvar dataset. as we know annotating is a time and resource consuming process and regarding our resource and time we used 305983 data

2-second challenge is we use web base platform that performance is important, so we need faster approaches

So for future work we can deal with those challenges and use deep learning approaches to capturing nonlinear relation in input and output data.

Referencias

- [1] Ivan A Adzhubei y col. "A method and server for predicting damaging missense mutations". En: *Nature methods* 7.4 (2010), págs. 248-249.
- [2] Luisa Azevedo y col. "Improving the in silico assessment of pathogenicity for compensated variants". En: *European Journal of Human Genetics* 25.1 (2017), págs. 2-7.
- [3] Lorenzo Bomba, Klaudia Walter y Nicole Soranzo. "The impact of rare and low-frequency genetic variants in common disease". En: *Genome biology* 18.1 (2017), pág. 77.
- [4] Mark S DeFrancesco y col. "Hereditary cancer risk assessment and genetic testing in the community-practice setting". En: *Obstetrics & Gynecology* 132.5 (2018), págs. 1121-1129.
- [5] Jean-Pierre Desvignes y col. "VarAFT: a variant annotation and filtration system for human next generation sequencing data". En: *Nucleic acids research* 46.W1 (2018), W545-W553.
- [6] ensembl. *Variant Effect Predictor*. https://www.ensembl.org/info/docs/tools/vep/script/vep_download.html/. [Online; accessed 10-July-2020]. 2020.
- [7] health. *health*. <https://research.google/teams/health/>. [Online; accessed 10-July-2020]. 2020.
- [8] James M Holt y col. "VarSight: prioritizing clinically reported variants with binary classification algorithms". En: *BMC bioinformatics* 20.1 (2019), págs. 1-10.
- [9] Xuyun Hu y col. "Proband-only medical exome sequencing as a cost-effective first-tier genetic diagnostic test for patients without prior molecular tests and clinical diagnosis in a developing country: the China experience". En: *Genetics in Medicine* 20.9 (2018), págs. 1045-1053.
- [10] Panagiotis Katsonis y Olivier Lichtarge. "Objective assessment of the evolutionary action equation for the fitness effect of missense mutations across CAGI-blinded contests". En: *Human mutation* 38.9 (2017), págs. 1072-1084.
- [11] Martin Kircher y col. "A general framework for estimating the relative pathogenicity of human genetic variants". En: *Nature genetics* 46.3 (2014), págs. 310-315.
- [12] Qigang Li y col. "Xrare: a machine learning method jointly modeling phenotypes and genetic evidence for rare disease diagnosis". En: *Genetics in Medicine* 21.9 (2019), págs. 2126-2134.
- [13] Davis J McCarthy y col. "Choice of transcripts and software has a large effect on variant annotation". En: *Genome medicine* 6.3 (2014), pág. 26.
- [14] W McLaren y col. *The Ensembl Variant Effect Predictor Genome biology* 17, 122. 2016.
- [15] Enrico Mossotto y col. "GenePy-a score for estimating gene pathogenicity in individuals using next-generation sequencing data". En: *BMC bioinformatics* 20.1 (2019), pág. 254.

- [16] Pauline C Ng y Steven Henikoff. "SIFT: Predicting amino acid changes that affect protein function". En: *Nucleic acids research* 31.13 (2003), págs. 3812-3814.
- [17] Abhishek Niroula y Mauno Vihinen. "Predicting severity of disease-causing variants". En: *Human mutation* 38.4 (2017), págs. 357-364.
- [18] Kymberleigh A Pagel y col. "Pathogenicity and functional impact of non-frameshifting insertion/deletion variation in the human genome". En: *PLoS computational biology* 15.6 (2019), e1007112.
- [19] Sofia Papadimitriou y col. "Predicting disease-causing variant combinations". En: *Proceedings of the National Academy of Sciences* 116.24 (2019), págs. 11878-11887.
- [20] Ryan Poplin y col. "Creating a universal SNP and small indel variant caller with deep neural networks". En: *BioRxiv* (2016), pág. 092890.
- [21] Daniel Quang, Yifei Chen y Xiaohui Xie. "DANN: a deep learning approach for annotating the pathogenicity of genetic variants". En: *Bioinformatics* 31.5 (2015), págs. 761-763.
- [22] Rosalba E Sacca y col. "Trans-counseling: A case series of transgender individuals at high risk for BRCA1 pathogenic variants". En: *Journal of Genetic Counseling* 28.3 (2019), págs. 708-716.
- [23] Sameer Singh. "Multiresolution estimates of classification complexity". En: *IEEE Transactions on pattern analysis and machine intelligence* 25.12 (2003), págs. 1534-1539.
- [24] Damian Smedley y col. "A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease". En: *The American Journal of Human Genetics* 99.3 (2016), págs. 595-606.
- [25] Alex Wells y col. "Ranking of non-coding pathogenic variants and putative essential regions of the human genome". En: *Nature communications* 10.1 (2019), págs. 1-9.
- [26] Alison Luk Young y col. "Talking across generations: Family communication about BRCA1 and BRCA2 genetic cancer risk". En: *Journal of Genetic Counseling* 28.3 (2019), págs. 516-532.
- [27] Lin Zhou y Fangqing Zhao. "Prioritization and functional assessment of noncoding variants associated with complex diseases". En: *Genome medicine* 10.1 (2018), pág. 53.

Autores



Alberto Botana Lopez, nació en Santiago de Compostela, donde creció y cursó la carrera de Física. Desde pequeño tuvo una pasión abrumadora por las historias de ciencia ficción con robots y, especialmente, en aquellas donde había inteligencias artificiales. Es por ello que dedicó su Trabajo de Fin de grado a un primer encuentro con este campo y posteriormente estudió el Máster en Sistemas Inteligentes por la Universidad de Salamanca. Ahora mismo trabaja como Técnico de Robótica y Control en el Área de Inteligencia Artificial en AIMEN (Asociación de Investigación Metalúrgica del Noroeste) en O Porriño, Galicia.

Nuria Mateos, cursó el grado en Ingeniería Informática en Sistemas de Información en la Universidad de Salamanca. Posteriormente, estudió el Máster en Sistemas Inteligentes, también en la Universidad de Salamanca. Actualmente ha obtenido una beca otorgada de forma conjunta por CaixaBank y Microsoft, por lo que se encuentra trabajando en el departamento de Auditoría Continua, realizando proyectos de detección de fraude mediante algoritmos de Deep Learning.



Nibeth Mena Mamani, nacida en Potosí, Bolivia. Graduada en Ingeniería de Sistemas por la universidad Mayor de San Simón de Cochabamba. Trabajo como ingeniera de proyectos en el área de telecomunicaciones en diferentes proyectos a nivel regional y nacional, a la par realizó un Diplomado en Educación superior basado en competencias en la universidad Mayor de San Simón el cual le permitió impartir cátedra, ser tutora y miembro de tribunal de pregrado y posgrado en el área de ciencias de la computación en diferentes instituciones educativas de Cochabamba. El 2016 obtuvo una beca otorgada por la embajada de los EEUU en el programa “WeAmericas.”^{el} cual le permitió adquirir capacitaciones sobre emprendimientos tecnológicos en los EEUU. Posteriormente trabajó como Scrum Master, Project Manager y Agile coach en una innovadora empresa de desarrollo de software que aplica inteligencia artificial a sus productos. Ha participado también en diferentes congresos, conferencias y eventos tecnológicos a nivel regional, nacional e internacional, promoviendo la participación de la mujer en la tecnología. El 2019 obtuvo una beca otorgada por la Fundación Simón I. Patiño

que le permitió realizar el Máster en Sistemas Inteligentes en la Universidad de Salamanca (USAL). Actualmente trabaja como investigadora colaboradora en el grupo de investigación GRIAL y cursa el Doctorado en Formación en la Sociedad del Conocimiento de la USAL.

Francisco García Encinas, estudió el Grado en Ingeniería Informática y el Máster en Sistemas Inteligentes en la Universidad de Salamanca. En la actualidad, se encuentra matriculado en el programa de doctorado de esta misma universidad y participa en varios proyectos de investigación relacionados con la inteligencia artificial y, más concretamente, con el aprendizaje profundo dentro del grupo de investigación ESALab.



Belén Perez Lancho es Licenciada en Ciencias Físicas y Doctora por la Universidad de Salamanca desde 1995. Realizó su formación posdoctoral en la Universidad Paris VI durante el curso académico 1996-97 y desde 1998 es Profesora del Área de Ingeniería de Sistemas y Automática en la Universidad de Salamanca. Imparte docencia en las titulaciones de Grado en Ingeniería Informática, Grado en Física y en el Máster en Sistemas Inteligentes. Ha participado en más de 20 proyectos de investigación, principalmente en aplicaciones de sistemas de control inteligente y sistemas multiagente, y ha ocupado cargos de gestión como Secretaria Académica y Vicedecana de la facultad de Ciencias.



David Peral García, graduado en Ingeniería Informática por la Universidad de Salamanca en la que realizó el Máster en Sistemas Inteligentes. Su labor se ha centrado principalmente en el desarrollo de aplicaciones en realidad virtual y aumentada así como aplicaciones en iOS. Beneficiario de la beca TCUE relativa a la Universidad de Salamanca y de la beca Digitaliza impartida por CISCO en el campo de la programación de redes. Actualmente acompaña la realización del doctorado con el trabajo en el grupo de investigación ESALAB participando en proyectos de desarrollo e investigación en el área del guiado y mapeo de robots..

Belén Curto Diego, es Licenciada en Ciencias Físicas en la rama de Electrónica y Doctora en Ciencias Físicas. Actualmente es Profesora Titular de Universidad. Pertenecer al Dpto de Informática y Automática de la Universidad de Salamanca. Imparte clase en las titulaciones de Grado en Ingeniería Informática, Grado en Ingeniería Química, Máster en Ingeniería Informática y Máster en Sistemas Inteligentes sobre Robótica, Arquitectura de Computadores, Control de Procesos y Automatización Industrial. Pertenecer al Grupo de Investigación “Robótica y Sociedad” de la Universidad de Salamanca donde realizan trabajos sobre Inteligencia Artificial-Robótica y Automatización Industrial con empresas del sector alimentario, formación médica, nuclear, etc. Es autora de más de 30 artículos en revistas científicas especializadas. Ha dirigido números trabajos de Tesis Doctoral, Trabajos Fin de Máster y Grado a lo largo de su intensa labor docente e investigadora



Iñaki Arberas Vicente natural de Durango, es Graduado en Matemáticas (2019), por la Universidad del País Vasco. Posteriormente, realizó el Máster en Sistemas Inteligentes (2020), en la Universidad de Salamanca, complementando los estudios con unas prácticas en el Grupo de Investigación GRIAL, también de la Universidad de Salamanca.

Roberto Therón Sánchez cursó sus estudios de Informática en la Universidad de Salamanca (Diplomatura) y la Universidad de la Coruña (Licenciatura). Tras entrar a formar parte del Grupo de Investigación en Robótica de la Universidad de Salamanca, presentó su trabajo de Tesis recibiendo el Premio Extraordinario de Doctorado. Posteriormente ha obtenido los títulos de Licenciado en Comunicación Audiovisual (Universidad de Salamanca) y Licenciado en Humanidades (Universidad de Salamanca).



En la misma Universidad de Salamanca continúa realizando su trabajo de investigador, como encargado del grupo VisUsal (dentro del Grupo de Investigación Reconocido MiDa) que se centra en la combinación de enfoques procedentes de la Informática, Estadística, Diseño Gráfico y Visualización de Información, para obtener una adecuada comprensión de conjuntos de datos complejos. En los últimos años, se ha dedicado al desarrollo de herramientas de visualización avanzada para datos multidimensionales. En el área de Analítica Visual desarrolla productivas colaboraciones con grupos e instituciones de reconocido prestigio internacional, como el Laboratorio de Ciencias del Clima y del Medio Ambiente (París) o el Centro de Analítica Visual Avanzada de la ONU (Suiza). Es autor de más de 70 artículos en revistas y congresos internacionales.



Jose Luís Alonso Berrocal, es Doctor en Informática por la Universidad de Salamanca. Actualmente desarrolla su actividad académica como Profesor Titular del Departamento de Informática y Automática de la Universidad de Salamanca (Área de Lenguajes y Sistemas Informáticos). En el apartado de investigación cuenta con dos tramos de investigación y pertenece al grupo de investigación REINA (REcuperación de INformación Avanzada). Es autor

de un importante número de artículos en revistas y congresos internacionales y ha participado en diferentes proyectos de investigación de carácter regional, nacional. Sus líneas de investigación son: Recuperación de Información, Extracción Automática de Conocimiento, Cibermetría, Redes Sociales, Recuperación de Información en el Web. Imparte docencia en estudios de grado y en 3 másteres. Es miembro del Instituto de Estudios de la Ciencia y la Tecnología

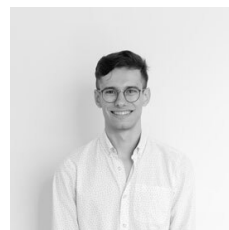
Pastora Isabel Vega Cruz es Doctora por la Universidad de Salamanca y Catedrática del área de conocimiento de Ingeniería de Sistemas y Automática.





Adilson Isaac González López, es Ingeniero en computación. Actualmente labora como Docente de la Universidad Nacional de Ingeniería Nicaragua para el departamento de lenguajes y simulación. En el ámbito de la formación de post grado es especialista en Administración TIC. Intereses a futuro es aportar en su institución de labores en la línea de investigación relacionado al campo de sistemas Inteligentes.

Jorge Vicente Gabriel es Graduado en Ingeniería Informática por la Universidad de Salamanca (2019), cuyo trabajo de fin de grado fue premiado por la beca de prototipos del TCUE (2018-2019). Además, ha estudiado el Máster de Sistemas Inteligentes de la Universidad de Salamanca (2020) y continua sus estudios en el Máster de Big Data y Data Science en la Universidad Internacional de Valencia. Aparte, ha formado parte del equipo del Grupo de Investigación BISITE de la Universidad de Salamanca donde pudo orientar su carrera profesional.



Zakieh Alizadehsani received The B.Sc. degree in 2010 from Ferdowsi University, Mashhad, Iran. She was involved in several projects as a developer and Penetration tester. Also, she Taught web application security vulnerability scanning and validation techniques. Since 2019, She has started working in Bitsie, the project that she was involved in includes big data engineering, web development. She got her Master's degree in the Artificial intelligent field in

2020.

Angel Luis Sánchez Lázaro, Doctor y Profesor Titular en el área de conocimiento de Lenguajes y Sistemas Informáticos de la Universidad de Salamanca.

renewcommandÍndice alfabéticoAutores

