



EVENTO VIRTUAL
PRIMERA CONVENCION CIENTIFICA
INTERNACIONAL
"ISLACIENCIA 2021"

APLICACION PARA LA CLASIFICACION DE PACIENTES CON PARKINSON
CON EL EMPLEO DE APRENDIZAJE AUTOMATICO

Javier Crespo Damota ¹ , Arlety Leticia Garcia Garcia ² , Alberto Salazar Peña ³

1 Cuba, Estudiante, Facultad de Ciencias Técnicas, Carrera Ingeniería informática, UIJ jcdamota@uij.edu.cu

2 Cuba, Profesora Auxiliar UIJ, ingeniera en Comunicaciones, Máster en dirección., algarciag@uij.edu.cu

3 Cuba, Profesor Asistente UIJ, licenciado en ciencia de la Computación, aspena@uij.edu.cu

Informatización para el desarrollo sostenible

Resumen

El diagnóstico y la predicción de enfermedades mediante técnicas de Aprendizaje Automático se perfilan como un área de investigación con diferentes herramientas y técnicas que ha despertado el interés de la comunidad científica en los últimos tres años. En la esfera de la salud en la última década se ha desarrollado la llamada medicina de precisión con el empleo de los sensores disponibles en el mercado (que van desde los llamados llevables en la ropa hasta los que se encuentran incorporados en los Smartphone o teléfonos inteligentes), con técnicas de internet de las cosas para la extracción de señales biomecánicas y técnicas de aprendizaje automático para la clasificación, monitoreo y seguimiento de pacientes en diversas áreas de la salud. La enfermedad de Parkinson es uno de los principales trastornos del sistema nervioso central, hasta el momento se desconoce la causa de la enfermedad, es por ello que no existe un examen biomédico que pueda diagnosticarla, demora varios años la aparición de los síntomas para clasificarla pues en ocasiones puede confundirse con los de otras enfermedades del sistema nervioso. Su diagnóstico y clasificación ha sido de interés para el área del aprendizaje automático en los últimos años, por la cantidad de datos que genera y la dificultad adicional que generan los múltiples síntomas que dan inicio a la enfermedad. Los investigadores utilizan muchas tecnologías como el aprendizaje automático, el análisis de imágenes, el análisis de señales y técnicas basadas en dispositivos para examinar los datos de esta enfermedad desde diversas fuentes. Existen varias herramientas y técnicas para lograrlo; en el presente trabajo mediante el uso de las tecnologías como Anaconda, PyCharm y Eclipse, se implementaron los algoritmos Random Forest, Red Neuronal Convolucional, KNN, Gauss Process y Support Vector Machine, con resultados de 80 % de exactitud en la clasificación de pacientes con Parkinson, de todos los algoritmos implementados.

Palabras clave: aprendizaje automático, Parkinson, algoritmos.

Introducción

El aprendizaje automático es un campo de inteligencia artificial que utiliza técnicas estadísticas para que los sistemas informáticos puedan aprender a partir de los datos, sin ser programados explícitamente.

Tiene como función explorar el estudio y la construcción de algoritmos que pueden aprender y hacer predicciones sobre los datos; dichos algoritmos se superan siguiendo instrucciones estrictamente estáticas del programa al hacer predicciones o decisiones basadas en datos, mediante la construcción de un modelo a partir de entradas de muestra. Se emplea en una variedad de tareas informáticas en las que diseñar y programar algoritmos explícitos con buen rendimiento es difícil o inviable.

En la actualidad, es uno de los pilares sobre los que descansa la transformación digital. Se está utilizando para encontrar nuevas soluciones en diferentes campos como vehículos inteligentes, procesamiento de Lenguaje Natural, búsquedas, medicina, redes sociales, entre otros. Hoy por hoy tiene una sostenida y continua participación en la ciberseguridad; los nuevos antivirus y motores de detección de malware ya se sirven del aprendizaje automático para potenciar el escaneado, acelerar la detección y mejorar la habilidad de reconocer anomalías.

Durante un estudio exhaustivo acerca de las funcionalidades de este método (M.L), se ha determinado que tiene una amplia utilización en los diversos campos de investigación, siendo uno de ellos, la medicina de precisión, en este caso la enfermedad del Parkinson. Conocida como la segunda enfermedad neurodegenerativa, tan solo por detrás del Alzheimer; la cual afecta mayormente a nivel mundial a las personas de la 3ra edad.

En Cuba, aproximadamente el 22% de la población de la 3^{ra} edad padece de dicho trastorno, por lo que dentro de los programas priorizados del país está el envejecimiento, conocido por el nombre "Calidad de vida y salud". En el Municipio especial Isla de la Juventud, también se pone en práctica dicho programa, y se les brinda atención a esas personas cada cierto periodo de tiempo, dándole así seguimiento a la evolución de su enfermedad.

Actualmente, en el mundo, existen numerosas tecnologías y aplicaciones para clasificar pacientes con Parkinson. Cuba no cuenta con tantas tecnologías ni aplicaciones, por lo que se hace necesario también el desarrollo de aplicaciones mediante el uso de herramientas de la vida cotidiana como Smartphone para la extracción de los datos, que puedan ser validadas para el apoyo a los especialistas en la toma de decisiones clínicas.

La cantidad de información que un solo paciente genera a lo largo del tiempo en que se le da seguimiento hasta el momento en que se diagnostica, y el posterior tratamiento dificulta el seguimiento a los datos, genera grandes volúmenes de datos distribuidos en historias clínicas y en el mejor de los casos en bases de datos.

Además de que el diagnóstico actual consta de un test de acerca de 100 preguntas, las cuales dependen de un especialista de neurología. Dicha evaluación se basa en una escala donde 0 es sano y 5 es grave, lo que hace muy subjetiva la evaluación.

La situación problemática que se nos presenta está caracterizada por la existencia de grandes volúmenes de flujos de datos de pacientes con Parkinson que necesitan ser extraídos y analizados en tiempo real, la diversidad de métricas para los datos a ser extraídos según los síntomas presentes y la necesidad de evaluación de tendencias en el cuidado clínico. El objetivo del trabajo es: Desarrollar una herramienta informática capaz de clasificar a personas con Parkinson mediante el empleo de los Algoritmos de Aprendizaje Automático.

Desarrollo

El aprendizaje automatizado (Machine Learning, ML) es un subcampo de la Inteligencia artificial encargado de estudiar el problema del aprendizaje en las máquinas, es decir, su objeto de estudio es el problema de cómo las máquinas pueden adquirir el conocimiento que las capacite para resolver problemas determinados (Mitchell, 1997).

El aprendizaje automático tiene como resultado un modelo para resolver una tarea dada. De esta forma se presenta una forma de clasificar los algoritmos en función de su matemática o cimientos. Si bien la discusión se simplifica, proporciona una forma integral de explorar algoritmos desde los primeros principios (Flach, 2012).

Dos de las técnicas de aprendizaje basado en máquina más ampliamente adoptados son aprendizaje supervisado y aprendizaje no supervisado, pero además de estos, existen también otras técnicas de machine learning.

El aprendizaje supervisado

Son entrenados mediante ejemplos etiquetados, como una entrada donde se conoce el resultado deseado. El algoritmo de aprendizaje recibe un conjunto de entradas junto con los resultados correctos correspondientes, y el algoritmo aprende al comparar su resultado real con resultados correctos para encontrar errores, luego modifica el modelo en consecuencia (Sandoval, 2018).

Los métodos de clasificación se encuentran dentro de la categoría de ML supervisado. Predicen o explican un valor de clase. Por ejemplo, pueden ayudar a predecir si un cliente en línea comprará o no un producto. La salida puede ser sí o no: comprador o no comprador. Pero los métodos de clasificación no se limitan a dos clases. Por ejemplo, un método de clasificación podría ayudar a evaluar si una imagen dada contiene un automóvil o un camión. El algoritmo de clasificación más común es logístico, Regresión.

Métodos de clasificación

Algunos de los métodos que más se describen en la literatura consultada y que se emplean en la presente investigación son abordados a continuación.

Máquinas de Vectores de Soporte

Surgen para encontrar la manera óptima de clasificar. Son una técnica de machine learning que encuentra la mejor separación posible entre clases. En vez de encontrar la línea óptima, el SVM encuentra el hiperplano que maximiza el margen de separación entre clases.

Red Neuronal convolucional

Es un tipo de red neuronal artificial donde las neuronas corresponden a campos receptivos de una manera muy similar a las neuronas en la corteza visual primaria (V1) de un cerebro biológico. Este tipo de red es una variación de un perceptrón multicapa, pero su funcionamiento las hace mucho más efectivas para tareas de visión artificial, especialmente en la clasificación de imágenes.

Métodos de ensamble

Aunque hay diversas formas de ensamblar o unir algoritmos débiles para formar otros, las más usadas y populares son el bagging y el boosting.

Bagging: Los métodos de bagging son métodos donde los algoritmos simples son usados en paralelo. El principal objetivo de los métodos en paralelo es el de aprovecharse de la independencia que hay entre los algoritmos simples, ya que el error se puede reducir bastante al promediar las salidas de los modelos simples.

Una técnica mejorada del Bagging es el Random Forest, que además de elegir un grupo aleatorio de individuos, también elige un grupo aleatorio de variables.

Bosque aleatorio (Random Forest)

El algoritmo Random Forest es una técnica de aprendizaje supervisado que genera múltiples árboles de decisión sobre un conjunto de datos de entrenamiento: los resultados obtenidos se combinan a fin de obtener un modelo único más robusto en comparación con los resultados de cada árbol por separado (Zuñiga, 2020).

El aprendizaje automático es un campo único, que evoluciona rápidamente a medida que avanzan los años; los algoritmos a menudo se reemplazan para dar paso a métodos más eficientes. El método de detección Histograma de gradientes orientados (HOG para abreviar) es uno de estos algoritmos antiguos, que tiene casi una década; sin embargo, una cosa que lo diferencia de otros es que todavía se usa mucho hoy con resultados fantásticos.

Histogramas de Gradientes Orientados

HOG es un tipo de “descriptor de características”. El objetivo de un descriptor de características es generalizar el objeto de tal forma que el mismo objeto produzca lo más cerca posible del mismo descriptor de características cuando se le vea bajo diferentes condiciones. Esto hace que la tarea de clasificación sea más fácil.

También a través de ellos se convierte la imagen (ancho x alto x canales) en un vector de características de longitud n elegido por el usuario. Aunque puede ser difícil ver estas imágenes, estas imágenes serán perfectas para algoritmos de clasificación de imágenes con el fin de producir buenos resultados (Adrián, 2015).

En la investigación será empleada con el objetivo de mejorar la calidad de las imágenes de ola o espiral empleadas por los diferentes algoritmos.

Antecedentes sobre el tema

En la revisión de los artículos encontrados sobre el empleo de algoritmos del aprendizaje automático para clasificar la enfermedad del Parkinson se denota que existe interés de parte de la comunidad científica en desarrollar investigaciones de esta naturaleza con el empleo de Smartphone.

En el 2019 Feng Tian, Xiangmin Fan, Junjun Fan y otras seis personas, realizaron una investigación llamada "What Can Gestures Tell? Detecting Motor Impairment in Early Parkinson's from Common Touch Gestural Interactions". De un grupo de 102 personas (35 pacientes y 67 sujetos de control) se evaluaron movimientos de las manos: pellizco, arrastre, escritura a mano, tecleo, tecleo alterno entre dedos.

Se usó como dispositivo un Smartphone Huawei P9 Plus, con una versión Android de 7.0. Se desarrolló un modelo de aprendizaje automático (Support Vector Machine) que separe los pacientes con Parkinson de las personas sanas a partir de censar y analizar el comportamiento de los movimientos de los dedos.

En el mismo año, Murtadha D. Hssayeni, Joohi Jimenez-Shahed y Behnaz Ghoraani desarrollaron una investigación llamada "Hybrid Feature Extraction for Detection of Degree of Motor Fluctuation Severity in Parkinson's Disease Patients". El grupo evaluado fue de 24 personas con parkinsonismo idiopático, de los que se tomó la medida de los sensores en sus actividades diarias. Utilizaron Sensores como herramientas, ubicados en las extremidades superiores e inferiores. Se desarrolló un algoritmo para estimar el grado de severidad de la fluctuación motora, después se desarrolló un modelo de clasificación con el empleo del algoritmo Random Forest (Algoritmo perteneciente al Machine Learning).

Shounak Ray, en el año 2019 bajo el nombre "A Predictive Diagnosis for Parkinson's Disease Through Machine Learning" desarrolló la investigación. En este estudio, se analizaron datos demográficos, de movimiento y del habla humana para determinar si un individuo tiene la enfermedad de Parkinson, lo que resulta en un problema de clasificación binaria. Como dispositivo usó un Smartphone Android y datasets de repositorios.

"Deep Reinforcement Scheduling for Mobile Crowdsensing in Fog Computing", investigación realizada por HE LI, KAORU OTA, y MIANXIONG DONG en el año 2019. Usó como algoritmo una Red Neuronal Convolutiva de cuatro capas. Como ejercicio se consideró un juego de nodos de nube y un grupo de usuarios, la tarea de programar el problema intenta adjudicar nodos de la nube a los pasos de cada tarea presentada por los usuarios.

Gennaro Vessio, 2019, "Dynamic Handwriting Analysis for Neurodegenerative Disease Assessment: A Literary Review". Investigación donde se plantea por lo general el uso de Tablet o lápices electrónicos, cada uno con sus características. Las variables significativas que se adquieren son la posición en los ejes X y Y, y el tiempo de stamp.

En el 2017 Clayton R. Pereira junto a su equipo de trabajo llevan a cabo una investigación que tenía como título "Deep Learning-aided Parkinson's Disease Diagnosis from Handwritten Dynamics". La sección experimental comprendía diferentes arquitecturas de CNN (Convolutional Neural Networks), así como imágenes con diferentes resoluciones y distintos tamaños de conjuntos de entrenamiento. Se propone modelar el problema de distinguir la EP y los individuos de control como una tarea de reconocimiento de imágenes mediante CNN.

En 2019, Manuel Gil-Martín, Juan Manuel Montero y Rubén San-Segundo, realizaron una investigación con el nombre "Parkinson's Disease Detection from Drawing Movements Using Convolutional Neural Networks". En este se analizó la capacidad de discriminación de direcciones diferentes durante el dibujo de movimientos obteniendo los mejores resultados para X y Y. El estudio se desarrolló con el dataset público: Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet dataset(77 personas, 62 con PD y 15 sanos).

Después del análisis de los artículos sobre el tema se decidió emplear para el desarrollo de la aplicación, los algoritmos que mejor se adaptan a las tareas de clasificación de imágenes, Random Forest, Support Vector Machine, Convolutional Neural Network, KNN, Gauss Process.

Solución propuesta

Se decide desarrollar una aplicación para Android para la interacción con el cliente desde un Smartphone en el que pueda dibujar una espiral o una ola, y otra que se mantenga leyendo por medio de la wifi desde una pc con la que interactuará el médico y donde se implementarán los algoritmos de clasificación, de manera que este último podrá recibir en tiempo real la imagen de los pacientes y tomar decisiones en base a la clasificación de los algoritmos.

Tecnologías

A partir del estudio de las tecnologías, se concluyó tomar como lenguaje de desarrollo principal Python para el desarrollo de la aplicación de desktop y como lenguaje secundario Java para la aplicación androide, como sistema de gestión de paquetes Anaconda, como Entorno de Desarrollo Integrado (IDE) Pycharm y Eclipse. Por ser necesario el empleo de algoritmos como la red neuronal convolucional, ya sea para su construcción como para su interpretación se usaron librerías externas de Python, Tensorflow, Keras, Scikit-Learn y OpenCV.

Arquitectura del sistema

El programa cuenta con una aplicación androide que será usada por el paciente y una aplicación de desktop, la que estará bajo supervisión del doctor. Ambas

aplicaciones están conectadas entre sí a través de una conexión wifi (Ver Anexo, figura.1). Se establece la conexión otorgándole a la aplicación android el IP de la computadora donde se encuentra el software. El paciente dibuja una espiral o una ola, depende de las instrucciones del doctor, en el Smartphone y envía las coordenadas a la aplicación de la computadora, dibujándose así, en tiempo real, en la aplicación de desktop. El neurólogo es el encargado de clasificar la imagen desarrolla por el paciente al presionar el botón diagnosticar; automáticamente aparecen reflejados los resultados obtenidos en una tabla después de haber sido clasificada la imagen por los distintos algoritmos empleados por el autor. La imagen puede ser guardada, junto a ella se guardará un documento con el mismo nombre dado a la imagen, con algunos datos necesarios de la imagen. También se puede abrir nuevamente la imagen y volver a clasificarla.

Estructura del dataset

El dataset cuenta con doscientas cuatro imágenes en total, ciento dos imágenes de espirales y ciento dos imágenes de ola. Dentro de las carpetas de ola y espiral, se encuentra la carpeta testing que divide el total de imágenes de las carpetas anteriores en treinta imágenes y training que cuenta con setenta y dos imágenes.

En el nivel más bajo dentro de las carpetas anteriores se encuentran clasificadas las imágenes en saludables o con parkinson. (Ver Anexo, figura 2.)

Implementación del algoritmo Red Neuronal Convolutiva

El algoritmo de entrenamiento cuenta con una estructura de cuatro convoluciones y a cada convolución se le hace un max-pooling, posteriormente el entrenamiento se realiza mediante ciento veinticinco iteraciones en cincuenta épocas. Como resultado del entrenamiento se obtuvo 0.83 de efectividad.

Implementación de los algoritmos restantes

La implementación del algoritmo se realiza en cinco pasos:

- Se redimensionan a doscientos por doscientos las imágenes del conjunto de datos y la imagen recibida.
- Se les realiza un threshold a cada una de las imágenes, tanto las del dataset como la imagen recibida, lo que elimina el ruido.
- Se les aplica como descriptor de imágenes el Histograma de Gradiente Orientado a cada una de las imágenes, para convertir las imágenes en vectores.
- Se entrena el conjunto de datos del dataset.

Diagrama de despliegue

Para el funcionamiento de la herramienta propuesta es necesario que el dispositivo Smartphone y la computadora se encuentren conectados dentro de la misma red o conectados en red local a través de la wifi. (Ver Anexo, figura 3.)

El ordenador debe tener:

- Un procesador i3 o mayor para el óptimo funcionamiento del tensorflow.
- El procesador debe tener instrucciones AVX.
- Espacio disponible 200MB
- Mínimo 4 GB RAM.

El Smartphone debe tener:

- Pantalla Super AMOLED
- Version android 4.4.2 o superior
- Espacio disponible 50MB

Pruebas de validación

Se realizará de dos formas, por la tabla de contingencia para la validación de pruebas diagnósticas donde se determina la especificidad y la sensibilidad.

El segundo método es el cross validation, será empleado para validar los algoritmos específicamente. En este caso se evaluará entre los 4 clasificadores a emplear. Ej. Si hay un total de 20 datos (imágenes), por el método 4-fold cross validation, se llevarán a cabo cuatro iteraciones, y en cada una se utilizarán unos datos de entrenamiento diferentes, que serán analizadas por cuatro clasificadores, que posteriormente evaluarán los datos de prueba.

En este caso, para este tipo de estudio se busca que esté por encima del 80%, que se considera aceptable en ambos casos. La tabla 1 muestra los resultados. (ver Anexos)

Conclusiones

En la revisión de artículos de los últimos tres años se pudo constatar la utilidad del empleo del aprendizaje automático en el desarrollo de tareas de clasificación de pacientes con Parkinson, los algoritmos que más se utilizan son: Random Forest, Support Vector Machine, K Nearest Neighbour, red Neuronal Convolutiva y Gauss Process.

Para la validación de la aplicación se emplearon los métodos de las tablas de contingencia y el método de cross validation, obteniendo resultados de 80% de efectividad en ambos casos. Los algoritmos que mejores resultados obtuvieron para este conjunto de datos con ambos métodos fueron Random Forest(85%, 84%) y Support Vector Machine(85%, 83%).

Referencias bibliográficas

- Adrián, R. C. (2015). *Sistema de detección y rastreo de personas en tiempo real para el robot Golem-II+*. México.
- Arango, M. (2019). *Introducción al Aprendizaje por Refuerzo*.
- Challenger-Pérez, I., Díaz-Ricardo, Y., & Becerra-García, R. A. (2014). El lenguaje de programación Python . *Ciencias Holguín*, 1-13 .
- Clinic, H. n. (2020-2021). *Parkinsons Disease*. Obtenido de <https://www.mayoclinic.org/es-es/diseases-conditions/parkinsons-disease/diagnosis-treatment/drc-20376062>
- Ecured*. (s.f.). Obtenido de Pycharm: <https://www.ecured.cu/Pycharm>
- Flach, P. (2012). *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*.
- Fox, P. (2021). *Khan Academy*. Obtenido de Algoritmos de Aprendizaje Automático: <https://es.khanacademy.org/computing/ap-computer-science-principles/data-analysis-101/x2d2f703b37b450a3:machine-learning-and-bias/a/machine-learning-algorithms>
- García, M. P. (2018). *Técnicas Boosting*. Sevilla.
- Heras, J. M. (28 de Mayo de 2019). *IArtificial.net*. Obtenido de Máquinas de Vectores de Soporte (SVM): <https://www.iartificial.net/maquinas-de-vectores-de-soporte-svm/>
- Ignacio, J. (13 de Mayo de 2018). *Aprende Machine Learning*. Obtenido de Regresión Lineal en Español con Python: <https://www.aprendemachinlearning.com/regresion-lineal-en-espanol-con-python/#more-5722>
- Microsoft Azure*. (2021). Obtenido de Algoritmos de Aprendizaje Automático: <https://azure.microsoft.com/es-es/overview/machine-learning-algorithms/#techniques>
- Mitchell. (1997). *T.M Machine Learning*. McGrawHill.
- Noren, A. (24 de diciembre de 2019). *Sitio Big Data*. Obtenido de Árbol de decisión en Aprendizaje Automático : <https://sitiobigdata.com/2019/12/24/arbore-de-decision-en-aprendizaje-automatice/>
- Pérez, F. P. (2016). *Aprendizaje Automatico*. Madrid.
- Pressman, R. S. (2001). *Ingeniería del software: Un enfoque práctico*. New York.
- R. Domínguez-López, L. A.-d.-I.-C.-y.-M.-C.-R.-S.-M. (2016). *Multiclasificadores basados en aprendizaje automático como herramienta para la evaluación del perfil neurotóxico de líquidos iónicos*.
- Rodríguez, D. (23 de julio de 2018). *Analytics Lane*. Obtenido de La Regresión Logística: <https://www.analyticslane.com/2018/07/23/la-regresion-logistica/>
- Roman, V. (25 de Abril de 2019). *Ciencias & Datos*. Obtenido de Algoritmos Naive Bayes: Fundamentos e Implementación: <https://medium.com/datos-y-ciencia/algoritmos-naive-bayes-fudamentos-e-implementaci%C3%B3n-4bcb24b307f>

Salgueiro, A. P. (2017). Aprendizaje Automatico. *CIBERNÉTICA APLICADA*, (pág. 33).

Sandoval, L. J. (2018). ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS. *REVISTA TECNOLÓGICA N° 11*, 5.

Toro, L. (15 de Septiembre de 2017). *Desde Linux*. Obtenido de Anaconda Distribution: La Suite más completa para la Ciencia de datos con Python: https://blog.desdelinux.net/ciencia-de-datos-con-python/?utm_content=buffer5431d&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer

Yolvi Ocaña-Fernández, L. A.-F.-A. (2019). *Inteligencia artificial y sus implicaciones en la educación superior*. Peru.

Zuñiga, J. J. (2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. *Ingeniería Investigación y tecnología* , 1-16.

Anexos

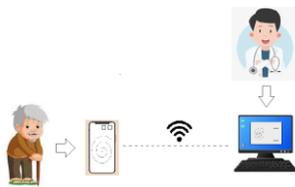


Figura 1Arquitectura del sistema propuesto

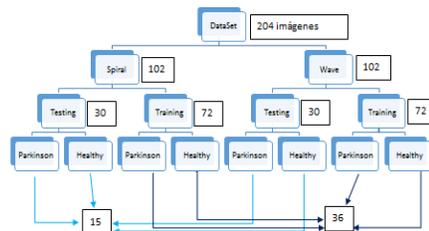


Figura 2Estructura del dataset



Figura 3 Diagrama de despliegue

Tabla 1Resultados de la validación por ambos métodos

Algoritmos	Resultados Tabla de contingencia	Resultados cross validation
Random Forest Espiral	0.85	0.84
Random Forest Ola	0.80	0.81
SVM Espiral	0.85	0.83
SVM Ola	0.80	0.79
KNN Espiral	0.80	0.82
KNN Ola	0.80	0.80
Gauss Process Espiral	0.80	0.83
Gauss Process Ola	0.80	0.77