

# **Sistema computacional para la identificación de la enfermedad de Parkinson mediante la prueba del espiral y visión computarizada**

Proyecto Intermedio

**José Alfredo Torres González**

**Maestría en Ciencias de la Computación**

Matemáticas para las ciencias de la computación

Profesor: Dr. Ricardo Barrón Fernández

Centro de Investigación en Computación, IPN

November 3, 2021

# Contents

	Page
<b>1 Sección</b>	<b>2</b>
1.1 Justificación . . . . .	2
1.2 Objetivos . . . . .	2
1.3 Aportaciones . . . . .	2
<b>2 Sección</b>	<b>3</b>
2.1 Estado del arte . . . . .	3
<b>3 Sección</b>	<b>4</b>
3.1 Fundamentos . . . . .	4
3.2 Solución Propuesta . . . . .	5
<b>4 Sección</b>	<b>6</b>
4.1 Resultados . . . . .	6
4.2 Conclusión . . . . .	8
<b>References</b>	<b>9</b>

# 1 Sección

## 1.1 Justificación

El origen de la medicina se remonta a la propia aparición del ser humano. En sus inicios el arte de curar se definía por conocimientos obtenidos de forma subjetiva y rudimentaria, pero tras la aparición del método científico el progreso tecnológico se ha visto beneficiado y a su vez el área de la salud. Resulta indispensable, para la continuidad de nuestra especie, que las metodologías y técnicas diagnosticas en la medicina continúen evolucionando y mejorando a la par del desarrollo tecnológico. Aportando un poco a la exploración de nuevas metodologías para diagnóstico y tratamiento de enfermedades, se ha decidido desarrollar un sistema computacional de reconocimiento para la enfermedad de Parkinson, que sirva como filtro para la detección de esta.

## 1.2 Objetivos

**General:** Desarrollar sistema computacional de reconocimiento para la enfermedad de Parkinson.

**Específicos:**

- Utilizar Python como lenguaje de programación y Google Colab como entorno de desarrollo para el proyecto.
- Preprocesar los datos para su correcta implementación.
- Analizar imagenes mediante visión computarizada y OpenCV.
- Utilizar HOG y el algoritmo de Bosque aleatorio para el entrenamiento del modelo predictivo.
- Determinar y representar los aciertos y errores del modelo mediante una matriz de confusión

## 1.3 Aportaciones

El propósito del proyecto es meramente exploratorio y de implementación, utilizando técnicas ya existentes, con fines académicos y de experimentación. Se busca generar información que a futuro sirva de base para el desarrollo de nuevos y mejores proyectos e investigaciones, que contribuyan al área de la salud con nuevas herramientas que permitan dar mejores diagnósticos a diversos padecimientos, como en este caso la enfermedad Parkinson.

## 2 Sección

### 2.1 Estado del arte

No.	Título	Autores	Publicación que emite	Año	Ideas Principales
1	Random Forests	Breiman, L.	SpringerLink	2001	Se describe lo que son los Random forest y se profundiza en sus capacidades como una efectiva herramienta para la predicción.
2	Parkinson's Disease Detection from Spiral and Wave Drawings using Convolutional Neural Networks: A Multistage Classifier Approach.	Chakraborty, S., et.al.	ICAAC	2020	Estudio sobre la correlación que existe entre la severidad en la enfermedad de Parkinson y las irregularidades de la escritura a mano. Se describe y propone un sistema de análisis basado en espirales.
3	Histograms of Oriented Gradients for Human Detection	Dalal, N., Triggs, B.	IEEE Computer Society Conference on Computer Vision and Pattern Recognition	2005	Se enlistan diferentes descriptores de gradiente y bordes y se destaca la la capacidad de HOG sobre estos y se profundiza en lo mismo, demostrando el argumento principal.
4	A Complete Guide to the Random Forest Algorithm	Donges, N.	Built In Beta	2019	Random Forest es un algoritmo de Machine Learning muy sencillo de utilizar que puede ser utilizado para tareas de clasificación y regresión. Se explica en qué destaca de otros algoritmos y cómo puede ser utilizado
5	Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking	Hasnain, M.	IEEEExplore	2020	El artículo se enfoca en una evaluación de servicios web, pero aporta mucha información necesaria sobre las matrices de confusión y su relación con la confiabilidad de las predicciones.
6	Understanding Random Forests from Theory to Practice	Louppe, G.	University of Liège	2014	Un profunda revision del tema árboles de decisión y bosques aleatorios, destacando sus capacidades de aprendizaje, procesos internos e interpretabilidad de los mismos.
7	Digitized Spiral Drawing: A Possible Biomarker for Early Parkinson's Disease	San Luciano, M., et.al.	PLoS ONE	2016	Un estudio diferente sobre la relación existente de la enfermedad de Parkinson con la fidelidad de los trazos dibujados a mano.
8	Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS® Implementations	Zhu, W., et.al.	NESUG	2010	Se rescata mucha información sobre los conceptos de sensibilidad, especificidad y precisión en el contexto de diagnóstico de enfermedades.

## 3 Sección

### 3.1 Fundamentos

La enfermedad de Parkinson es una enfermedad del sistema nervioso que afecta al movimiento. Se trata de una enfermedad progresiva que se puede categorizar en 5 etapas:

- Síntomas leves, que incluyen temblores y problemas de movilidad en un solo lado del cuerpo, pero que no interfieren con tu vida diaria.
- Síntomas que van empeorando y que ahora afectan ambos lados del cuerpo.
- Pérdida del equilibrio y caídas frecuentes por las dificultades en movilidad.
- Los síntomas son severos y la persona ya es incapaz de vivir sola.
- Es imposible caminar o mantenerse en pie, es posible que la persona necesite de una silla de ruedas y que incluso llegue a presentar alucinaciones.

Existe una clara relación entre la gravedad en la enfermedad y la movilidad de la persona, pero su diagnóstico temprano puede hacer que se controle mejor la enfermedad, ya que esta no tiene cura. Gracias a diferentes estudios, se ha desarrollado una prueba basada en el spiral de Arquímedes, que tiene como finalidad determinar si una persona tiene la enfermedad, analizando las variaciones en el trazo de dicho espiral.

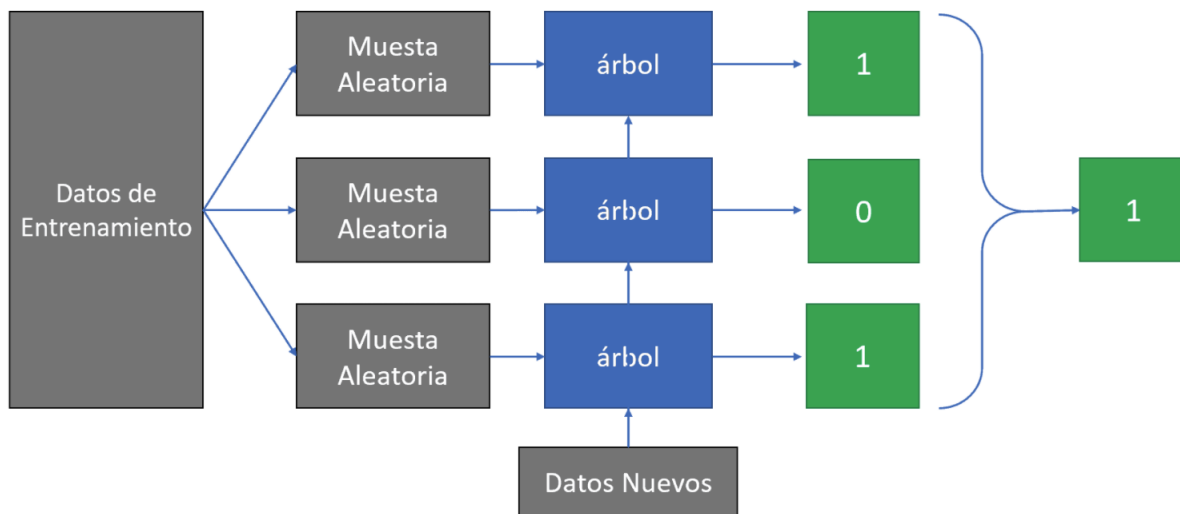


## 3.2 Solución Propuesta

Basado en lo anterior se tiene la intención de desarrollar un sistema que reciba diferentes imágenes con dibujos de espirales hechos por personas sanas y personas con Parkinson. Para obtener los datos necesarios, se pretende utilizar un dataset obtenido del perfil de Scott Mader en Kaggle, de donde no sólo se obtienen dibujos de espirales, si no también de ondas.

Para entrenar el modelo de clasificación se utilizarían Histogramas de Gradientes Orientados (HOG), en conjunto con el algoritmo de Machine Learning Random Forest Classifier. Se ha optado por utilizar esta estrategia y no Redes Neuronales Convolucionales (CNNs), ya que no se tienen muchos datos de entrenamiento y la primera estrategia funciona mejor con pocos datos.

HOG es un descriptor estructural con el que se van a cuantificar las imágenes para poder determinar que tantas variaciones hay en el trazo del espiral, ya que funcionan muy bien para analizar forma y textura. Por otro lado un Random Forest es un algoritmo de aprendizaje automático que se relaciona con los árboles de decisión, donde distintos árboles ven diferentes porciones de los datos, lo que al combinar sus resultados, ayuda a que la predicción se generalice mucho mejor.



## 4 Sección

### 4.1 Resultados

Aunque los resultados no fueron tan precisos como se esperaba, fueron satisfactorios, ya que en la prueba del espiral se obtuvo una precisión del 80 por ciento. Algo también importante de notar es que la sensibilidad fue de 0.7067, lo que nos dice que el modelo fue capaz de predecir positivos verdaderos el 70.67 por ciento del total de predicciones. Al aumentar el número de 'trials', que para esta prueba estuvo en 5, también aumenta la precisión. Estaría interesante entrenar el modelo con muchos más 'trials' y datos de entrada, para determinar que tanto aumenta dicha precisión.

```
✓ 1s ▶ # ciclo de procesamiento con base al número de 'trials'
for i in range(0, args["trials"]):
    # entrenar el modelo
    print("Entrenando modelo {} de {}".format(i + 1,
        args["trials"]))
    model = RandomForestClassifier(n_estimators=100)
    model.fit(trainX, trainY)

    # Hacer predicciones en con los datos de testeo y guardar las métricas
    predictions = model.predict(testX)
    metrics = {}

    # generar matriz de confusión y derivarla en precisión, sensibilidad y especificidad
    cm = confusion_matrix(testY, predictions).flatten()
    (tn, fp, fn, tp) = cm
    metrics["acc"] = (tp + tn) / float(cm.sum())
    metrics["sensitivity"] = tp / float(tp + fn)
    metrics["specificity"] = tn / float(tn + fp)

    # ciclo con las métricas
    for (k, v) in metrics.items():
        # actualizar el diccionario 'trials' con la lista de valores de la métrica actual
        l = trials.get(k, [])
        l.append(v)
        trials[k] = l

Entrenando modelo 1 de 5...
Entrenando modelo 2 de 5...
Entrenando modelo 3 de 5...
Entrenando modelo 4 de 5...
Entrenando modelo 5 de 5...
```

```

✓ [28] # definir rutas a los directorios del dataset
1s trainingPath = os.path.sep.join([args["dataset"], "training"])
testingPath = os.path.sep.join([args["dataset"], "testing"])

# cargar datos
print("Preparando datos...")
(trainX, trainY) = load_split(trainingPath)
(testX, testY) = load_split(testingPath)

# codificar etiquetas como datos numéricos
le = LabelEncoder()
trainY = le.fit_transform(trainY)
testY = le.transform(testY)

trials = {}

Preparando datos...

```

```

✓ 0s for metric in ("acc", "sensitivity", "specificity"):
    # acceder a los valores de cada métrica y después generar su promedio y desviación estandar
    values = trials[metric]
    mean = np.mean(values)
    std = np.std(values)

    # mostrar métricas
    print(metric)
    print("-" * len(metric))
    print("u={:.4f}, o={:.4f}".format(mean, std))
    print("")

└─ acc
   ---
   u=0.8000, o=0.0211

   sensitivity
   -----
   u=0.7067, o=0.0327

   specificity
   -----
   u=0.8933, o=0.0327

```





## 4.2 Conclusión

Para este proyecto, fue posible apreciar el potencial de las ciencias computaciones como complemento y herramienta en el área de la salud. Mediante visión computarizada y algoritmos de Machine Learning fue posible analizar imágenes y determinar si se trataba de un dibujo hecho por una persona con Parkinson y pesar de que la precisión no fue tan alta como se esperaba, si que deja un punto de partida para futuras investigaciones. Como trabajo futuro sería interesante juntar más datos y comparar la precisión de este modelo, con uno de Deep Learning, pero por lo pronto se logró desarrollar un modelo funcional que se puede utilizar como filtro diagnóstico, auxiliar en la identificación de la enfermedad.

# References

- Breiman, L. (2001). Random Forests. *Machine Learning* [online]. 45, pp. 5–32. DOI: 10.1023/a:1010933404324. Available from: <https://link.springer.com/article/10.1023/A:1010933404324>.
- Chakraborty, S., Aich, S., Jong-Seong-Sim, Han, E., Park, J., and Kim, H.-C. (Feb. 2020). *Parkinson’s Disease Detection from Spiral and Wave Drawings using Convolutional Neural Networks: A Multi-stage Classifier Approach*. IEEE Xplore. DOI: 10.23919/ICACT48636.2020.9061497. Available from: <https://ieeexplore.ieee.org/document/9061497> [Accessed Nov. 3, 2021].
- Dalal, N. and Triggs, B. (2005). *Histograms of Oriented Gradients for Human Detection*. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). DOI: 10.1109/cvpr.2005.177. Available from: <https://ieeexplore.ieee.org/document/1467360>.
- Donges, N. (2019). *A Complete Guide to the Random Forest Algorithm*. Built in. Available from: <https://builtin.com/data-science/random-forest-algorithm>.
- Hasnain, M. (May 2020). *Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking*. IEEEExplore. Available from: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9091880> [Accessed Nov. 2, 2021].
- Louppe, G. (July 2014). “Understanding Random Forests From Theory To Practice”. PhD thesis. Available from: <https://arxiv.org/pdf/1407.7502.pdf> [Accessed Nov. 2, 2021].
- San Luciano, M., Wang, C., Ortega, R. A., Yu, Q., Boschung, S., Soto-Valencia, J., Bressman, S. B., Lipton, R. B., Pullman, S., and Saunders-Pullman, R. (Oct. 2016). Digitized Spiral Drawing: A Possible Biomarker for Early Parkinson’s Disease. *PLoS ONE* [online]. 11, e0162799. DOI: 10.1371/journal.pone.0162799. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5061372/> [Accessed Nov. 3, 2021].
- Zhu, W., Zeng, N., and Wang, N. (2010). *Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS® Implementations*. Available from: <https://www.lexjansen.com/nesug/nesug10/h1/h107.pdf>.

(Chakraborty et al., 2020) (San Luciano et al., 2016) (Louppe, 2014). (Dalal and Triggs, 2005) (Donges, 2019) (Breiman, 2001) (Zhu, Zeng, and Wang, 2010) (Hasnain, 2020)