

Informe de Investigación: Detección de Enfermedades en Plantas

- Mesias Mariscal
- Denise Rea
- Julio Viche

1. Tema

Análisis comparativo de modelos de aprendizaje automático para la detección de enfermedades en plantas: evaluación de un modelo desarrollado en Python versus un modelo implementado en Edge Impulse, utilizando un dataset de imágenes de Kaggle.

2. Objetivos

Objetivo General

Evaluar y comparar el rendimiento de dos enfoques de aprendizaje automático (un modelo personalizado desarrollado en Python y un modelo basado en Edge Impulse) en la detección automática de enfermedades en plantas.

Objetivos Específicos

1. Entrenar un modelo de clasificación de imágenes en Python utilizando un dataset de plantas enfermas de Kaggle.
2. Desarrollar y entrenar un modelo equivalente utilizando la plataforma Edge Impulse con el mismo dataset de entrenamiento.
3. Evaluar el rendimiento de ambos modelos utilizando métricas estándar (precisión, recall, F1-score) sobre un conjunto de imágenes de prueba.
4. Comparar los resultados obtenidos por ambos modelos en términos de:
 - Exactitud en la detección de enfermedades
 - Tiempo de inferencia
 - Facilidad de implementación y despliegue
 - Recursos computacionales requeridos

5. Determinar cuál de los dos enfoques es más efectivo para la detección de enfermedades en plantas en un contexto académico.

3. Pregunta de Investigación

Pregunta Principal:

¿Cuál modelo de aprendizaje automático presenta mejor desempeño en la detección de enfermedades en plantas: un modelo personalizado desarrollado en Python o un modelo implementado mediante Edge Impulse?

Preguntas Secundarias:

- ¿Qué diferencias existen en la precisión de clasificación entre ambos modelos al evaluar imágenes nuevas?
- ¿Cómo se compara el tiempo de inferencia entre el modelo en Python y el modelo de Edge Impulse?
- ¿Qué ventajas y desventajas presenta cada enfoque en términos de implementación y usabilidad?
- ¿Cuál modelo es más adecuado para un despliegue en dispositivos con recursos limitados?

4. Enfoque

Enfoque Metodológico

Esta investigación adoptará un **enfoque cuantitativo experimental comparativo**, donde se implementarán y evaluarán dos modelos de clasificación de imágenes bajo condiciones controladas.

Tipo de Investigación

- **Investigación aplicada:** Se busca resolver un problema práctico de detección de enfermedades en plantas.
- **Investigación experimental:** Se manipularán variables (tipo de modelo) y se medirán los resultados de manera controlada.
- **Estudio comparativo:** Se contrastarán dos métodos diferentes para una misma tarea.

Metodología

Fase 1: Preparación de Datos

- Selección y descarga del dataset de Kaggle con imágenes de plantas sanas y enfermas
- Preprocesamiento de imágenes (normalización, redimensionamiento, augmentación)
- División del dataset en conjuntos de entrenamiento, validación y prueba (70-15-15)
- Creación de un conjunto adicional de imágenes nuevas para pruebas finales

Fase 2: Desarrollo del Modelo en Python

- Diseño e implementación de una red neuronal convolucional (CNN) utilizando frameworks como TensorFlow/Keras o PyTorch
- Entrenamiento del modelo con el dataset preparado
- Optimización de hiperparámetros
- Validación del modelo

Fase 3: Desarrollo del Modelo en Edge Impulse

- Carga del mismo dataset a la plataforma Edge Impulse
- Configuración y entrenamiento del modelo utilizando las herramientas de la plataforma
- Ajuste de parámetros dentro de las opciones disponibles
- Validación del modelo

Fase 4: Evaluación Comparativa

- Prueba de ambos modelos con el conjunto de imágenes de prueba
- Evaluación con imágenes completamente nuevas (no incluidas en el dataset original)
- Medición de métricas de rendimiento:
 - Matriz de confusión
 - Precisión (accuracy)
 - Precisión por clase (precision)
 - Recall
 - F1-score
 - Tiempo de inferencia
- Análisis de recursos computacionales utilizados

Fase 5: Análisis y Conclusiones

- Análisis estadístico de los resultados obtenidos
- Comparación detallada entre ambos modelos

- Identificación de fortalezas y debilidades de cada enfoque
- Formulación de conclusiones y recomendaciones

Herramientas y Tecnologías

- **Python:** TensorFlow/Keras o PyTorch, NumPy, Pandas, Matplotlib, Scikit-learn
- **Edge Impulse:** Plataforma online para desarrollo de modelos de ML
- **Dataset:** Kaggle - Plant Disease Dataset (u otro dataset relevante)
- **Hardware:** Computadora con GPU (para entrenamiento del modelo Python)

Criterios de Éxito

El modelo superior será aquel que demuestre:

1. Mayor precisión en la clasificación de enfermedades
2. Mejor generalización con imágenes nuevas
3. Balance óptimo entre rendimiento y eficiencia computacional
4. Facilidad de implementación para proyectos académicos