

DESARROLLO Y APLICACIÓN DE UN SISTEMA DE RECONOCIMIENTO DE EMOCIONES EN TIEMPO REAL USANDO APRENDIZAJE PROFUNDO Y VISIÓN POR COMPUTADORA

DEVELOPMENT AND APPLICATION OF A REAL-TIME EMOTION RECOGNITION SYSTEM USING DEEP LEARNING AND COMPUTER VISION

José de Jesús Castillo Nolasco¹

jcastillon001@alumno.uaemex.mx

ORCID: 0009-0000-8070-3280

Doricela Gutiérrez Cruz²

dgutierrezcr@uaemex.mx

ORCID: 0000-0003-2843-3273

Resumen

Este estudio presenta el desarrollo y aplicación de un sistema avanzado para el reconocimiento de emociones en videos pregrabados, que no solo identifica emociones a partir de transmisiones de video, sino que también determina la identidad, género, raza y rango de edad del sujeto. La problemática radica en la necesidad de mejorar la interacción humano-máquina mediante la identificación precisa de emociones y características demográficas en videos. Utilizando técnicas de aprendizaje profundo y visión por computadora, el sistema emplea un modelo de red neuronal convolucional (CNN) entrenado con el conjunto de datos FER-2013. Además, se integran modelos adicionales para el reconocimiento facial y la clasificación demográfica, demostrando precisión en variadas condiciones operativas. La metodología incluye un enfoque modular que permite la especialización en tareas específicas: detección facial con clasificadores Haar, clasificación de emociones con CNN, y predicción de edad y raza utilizando modelos preentrenados en

¹ Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, UAEMEX.

² Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, UAEMEX.



arquitecturas Caffé y GoogleNet. La evaluación se realizó en dos fases: una interna, utilizando un conjunto de datos de validación separado para medir la precisión y robustez del modelo en un entorno controlado, y una externa, probando el sistema con videos pregrabados de personas famosas para evaluar su rendimiento en condiciones más realistas y variadas. Los resultados resaltan la efectividad del sistema en la identificación de emociones y características demográficas, lo que abre nuevas vías para aplicaciones en seguridad, marketing personalizado y estudios comportamentales. Las conclusiones destacan la robustez del modelo propuesto y sugieren investigaciones futuras para explorar su aplicación en entornos multiculturales y dinámicos.

Palabras clave: Reconocimiento de emociones, aprendizaje profundo, FER-2013, procesamiento en tiempo real, visión por computadora, análisis emocional.

Abstract

This study presents the development and application of an advanced system for emotion recognition in prerecorded videos, which not only identifies emotions from video streams but also determines the subject's identity, gender, race, and age range. The challenge lies in the need to improve human-machine interaction through accurate identification of emotions and demographic features in videos. Using deep learning and computer vision techniques, the system employs a Convolutional Neural Network (CNN) model trained with the FER-2013 dataset. Additionally, facial recognition and demographic classification models are integrated, demonstrating accuracy under various operational conditions. The methodology includes a modular approach that specializes in specific tasks: facial detection using Haar classifiers, emotion classification with CNN, and age and race prediction using pretrained models based on Caffé and GoogleNet architectures. The evaluation was conducted in two phases: an internal phase, using a separate validation dataset to measure the model's precision and robustness in a controlled environment, and an external phase, testing the system with prerecorded videos of famous individuals to assess its performance in more realistic and varied conditions. The results highlight the system's effectiveness in identifying emotions and demographic features, opening new avenues for applications in security, personalized marketing, and behavioral studies. The conclusions emphasize the robustness of the proposed model and suggest future research to explore its application in multicultural and dynamic environments.



Keywords: Emotion recognition, deep learning, FER-2013, real-time processing, computer vision, emotional analysis.

Fecha de envío: 05/06/2024

Fecha de aprobación: 10/11/2024

Fecha de publicación: 01/01/2025

Introducción

En la sociedad actual, donde la interacción entre humanos y máquinas es cada vez más prevalente, comprender las emociones humanas a través de tecnologías automatizadas se ha convertido en un campo de investigación esencial. Los sistemas de reconocimiento de emociones, que inicialmente se limitaban a análisis en imágenes estáticas, han evolucionado hacia la interpretación de video en tiempo real, ofreciendo nuevas oportunidades para aplicaciones en sectores como la salud mental, la educación y la seguridad. Domínguez Jalili (2022) destaca que la capacidad de estos sistemas para adaptarse y responder en tiempo real es crucial para su eficacia en aplicaciones prácticas.

Este estudio tiene como objetivo principal desarrollar un sistema avanzado para el reconocimiento de emociones en tiempo real que también pueda identificar características demográficas como identidad, género, raza y rango de edad. La hipótesis central es que la integración de estas capacidades multifacéticas mejorará significativamente la precisión y la utilidad de los sistemas de reconocimiento emocional.

La relevancia de este trabajo radica en su enfoque holístico, que aborda tanto los aspectos técnicos del procesamiento de imágenes y aprendizaje automático como las aplicaciones prácticas de esta tecnología en contextos reales. Al incorporar múltiples atributos de reconocimiento, el sistema propuesto no solo aumenta su aplicabilidad en áreas diversas, sino que también contribuye al estudio de la dinámica emocional en diferentes grupos demográficos, abriendo nuevas vías para la personalización de servicios y productos tecnológicos.



Estado del Arte

Metodologías y Técnicas

- **Redes Neuronales Convolucionales (CNN):** Las CNN son fundamentales en FER debido a su capacidad para aprender características directamente de las imágenes. Estudios recientes han empleado arquitecturas avanzadas como VGGNet, ResNet y GoogleNet para mejorar la precisión y robustez de los modelos de FER. Estos modelos se entrenan utilizando grandes conjuntos de datos etiquetados, como FER-2013, AffectNet y CK+ (Zheng et al., 2020; Canal *et al.*, 2022).
- **Transfer Learning y Modelos Híbridos:** El uso de *transfer learning* ha permitido mejorar los modelos de FER al aprovechar redes preentrenadas en grandes conjuntos de datos generales y ajustarlas a tareas específicas de reconocimiento de emociones. Los modelos híbridos que combinan CNN con otras técnicas, como las Redes Neuronales Recurrentes (RNN), también han mostrado mejoras en la precisión del reconocimiento en entornos no controlados (Miragaia *et al.*, 2024).

Técnicas de Preprocesamiento

El preprocesamiento de imágenes es crucial para el rendimiento de los modelos de FER. Las técnicas comunes incluyen la normalización de iluminación, la alineación de rostros y el aumento de datos para crear variaciones en las imágenes de entrenamiento, mejorando así *la capacidad del modelo para generalizar en condiciones variadas* (Miragaia et al., 2024).

Integración de Características Demográficas

- **Predicción de Edad, Género y Raza:** La integración de características demográficas junto con el reconocimiento de emociones mejora la aplicabilidad del sistema. Modelos preentrenados en arquitecturas como Caffé y GoogleNet se utilizan para predecir la edad, género y raza de los sujetos. Esto permite aplicaciones más personalizadas y



contextualmente relevantes en áreas como la publicidad y la seguridad pública (Canal *et al.*, 2022).

Evaluación de Desempeño

- **Métricas de Evaluación:** Las métricas estándar para evaluar los sistemas de FER incluyen precisión, sensibilidad y especificidad. Los estudios recientes han destacado la importancia de evaluar los modelos en entornos tanto controlados como no controlados para asegurar su robustez y aplicabilidad en el mundo real (Miragaia *et al.*, 2024).

Aplicaciones Prácticas

- **Seguridad Pública:** Los sistemas de FER se utilizan en vigilancia para detectar comportamientos sospechosos y prevenir incidentes en lugares públicos como aeropuertos y estaciones de tren (Miragaia *et al.*, 2024).
- **Atención Médica:** En la salud, estos sistemas pueden monitorizar el bienestar emocional de los pacientes, facilitando intervenciones más oportunas y personalizadas, especialmente en el campo de la salud mental (Miragaia *et al.*, 2024).
- **Marketing:** Las empresas pueden analizar las reacciones emocionales de los consumidores a productos o anuncios, permitiendo estrategias de marketing más efectivas y mejorando la satisfacción del cliente (Canal *et al.*, 2022).

Desafíos y Futuras Direcciones

- **Robustez y Precisión:** La precisión de los sistemas de FER puede verse afectada por variaciones en la iluminación, poses faciales y expresiones parciales. Se requiere una mejora continua en los algoritmos para reducir los falsos positivos y negativos en estas condiciones (Canal *et al.*, 2022).
- **Diversidad de Datos:** La inclusión de conjuntos de datos con mayor diversidad étnica y emocional es crucial para mejorar la generalización de los modelos y reducir los sesgos inherentes (Zheng *et al.*, 2020).



- Tecnologías Emergentes: El uso de nuevas arquitecturas de redes neuronales, como los Transformadores de Visión (ViT), y tecnologías emergentes como el aprendizaje federado, puede ofrecer mejoras significativas en la eficiencia y privacidad del procesamiento de datos en tiempo real (Miragaia *et al.*, 2024).

Diseño del Sistema

En el contexto del desarrollo de un sistema avanzado para el reconocimiento de emociones y características demográficas en videos pregrabados, esta investigación se dirige a superar los desafíos inherentes a la precisión y la velocidad del reconocimiento. Para lograr esto, se ha seguido un enfoque metódico que incluye las siguientes fases:

1. **Diseño del Sistema:** Se utilizó un enfoque modular que permite una especialización en cada tarea (detección facial, clasificación de emociones y determinación de características demográficas). Este diseño modular no solo permite un enfoque especializado en cada tarea, sino que también mejora la escalabilidad y la eficiencia del sistema al procesar datos en videos pregrabados.
2. **Preparación de los Datos:** Los datos de entrenamiento y prueba se almacenaron en directorios separados. Se utilizaron técnicas de aumento de datos para mejorar la robustez del modelo, aplicando transformaciones como rotación, desplazamiento, corte, zoom y volteo horizontal.
3. **Entrenamiento del Modelo:** Se entrenó una red neuronal convolucional (CNN) con varias capas para la clasificación de emociones, utilizando el conjunto de datos FER-2013. El modelo se compiló utilizando el optimizador Adam y la pérdida de entropía cruzada categórica. Callbacks como 'ModelCheckpoint' y 'EarlyStopping' se emplearon para guardar el mejor modelo y evitar el sobreajuste.
4. **Integración de Módulos Adicionales:** Se integraron modelos preentrenados para la detección de edad y raza, utilizando arquitecturas Caffe y GoogleNet. Estos modelos se cargaron y utilizaron para predecir características demográficas en imágenes de rostros detectados.



5. Evaluación del Desempeño: La evaluación se realizó en dos fases: una interna, utilizando un conjunto de datos de validación separado para medir la precisión y robustez del modelo en un entorno controlado, y una externa, probando el sistema con videos pregrabados de personas famosas para evaluar su rendimiento en condiciones más realistas y variadas.

Este enfoque estableció un balance entre rendimiento y eficiencia que es crucial para la implementación práctica de tales sistemas.

Metodología

La metodología de este estudio abarca desde el diseño del sistema hasta las técnicas específicas utilizadas para el entrenamiento del modelo y la evaluación de su desempeño. Se siguieron metodologías ágiles de desarrollo de software, como Scrum, para la gestión del proyecto, y se adoptó el enfoque CRISP-DM (Cross-Industry Standard Process for Data Mining) para la estructuración de la investigación.

Se emplearon enfoques de aprendizaje profundo, especialmente redes neuronales convolucionales (CNN), para el reconocimiento de emociones y características demográficas. Según Leal Hernández (2017), la implementación de CNN para el reconocimiento de emociones facilita una alta precisión en la detección y clasificación de emociones a partir de expresiones faciales, lo que fundamenta la elección de esta tecnología para nuestro sistema.

El sistema se diseñó para procesar video. La arquitectura del sistema se compone de múltiples módulos: detección facial, clasificación de emociones y determinación de características demográficas (edad, género y raza). Este diseño modular no solo permite un enfoque especializado en cada tarea, sino que también mejora la escalabilidad y la eficiencia del sistema al procesar datos en tiempo real.

Descripción del Funcionamiento del Código

1. Preparación y Generación de Datos de Entrenamiento:

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator  
# Directorios de datos de entrenamiento y prueba
```



```
train_dir = 'archivos/data/dataset_emociones/train'
test_dir = 'archivos/data/dataset_emociones/test'
# Aumento de datos para el conjunto de entrenamiento
train_datagen = ImageDataGenerator(
    rescale=1./255,      # Normalización de los píxeles de las imágenes
    rotation_range=20,   # Rotación aleatoria de imágenes
    width_shift_range=0.2, # Desplazamiento horizontal aleatorio
    height_shift_range=0.2, # Desplazamiento vertical aleatorio
    shear_range=0.2,     # Corte aleatorio
    zoom_range=0.2,      # Zoom aleatorio
    horizontal_flip=True, # Volteo horizontal
    fill_mode='nearest'  # Modo de relleno
)
# Normalización de datos para el conjunto de prueba
test_datagen = ImageDataGenerator(rescale=1./255)

# Generadores de datos de entrenamiento y validación
train_generator = train_datagen.flow_from_directory(
    train_dir,
    target_size=(48, 48), # Tamaño de las imágenes
    batch_size=32,        # Tamaño del lote
    color_mode='grayscale', # Imágenes en escala de grises
    class_mode='categorical' # Clasificación categórica
)
validation_generator = test_datagen.flow_from_directory(
    test_dir,
    target_size=(48, 48),
    batch_size=32,
    color_mode='grayscale',
    class_mode='categorical'
)
```



2. Entrenamiento del Modelo de Clasificación de Emociones:

Se creó y entrenó un modelo CNN para la clasificación de emociones.

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout,
BatchNormalization
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import ModelCheckpoint

# Definición del modelo CNN
model = Sequential([
    Conv2D(64, (3, 3), activation='relu', input_shape=(48, 48, 1)), # Capa convolucional
    BatchNormalization(), # Normalización por lotes
    MaxPooling2D(2, 2), # Capa de agrupación
    Conv2D(128, (3, 3), activation='relu'), # Otra capa convolucional
    BatchNormalization(),
    MaxPooling2D(2, 2),
    Conv2D(256, (3, 3), activation='relu'),
    BatchNormalization(),
    MaxPooling2D(2, 2),
    Conv2D(512, (3, 3), activation='relu'),
    BatchNormalization(),
    MaxPooling2D(2, 2),
    Flatten(), # Aplanamiento de los datos
    Dense(1024, activation='relu'), # Capa totalmente conectada
    Dropout(0.5), # Regularización
    Dense(512, activation='relu'),
    Dropout(0.5),
```



```
Dense(7, activation='softmax')                                # Capa de salida
])

# Compilación del modelo
model.compile(optimizer=Adam(learning_rate=0.0001), loss='categorical_crossentropy',
metrics=['accuracy'])

# Configuración de callbacks
checkpoint = ModelCheckpoint(
    'mejor_modelo_emociones.keras',
    monitor='val_accuracy',
    verbose=1,
    save_best_only=True,
    mode='max'
)

# Entrenamiento del modelo
history = model.fit(
    train_generator,
    epochs=60,
    validation_data=validation_generator,
    callbacks=[checkpoint, tf.keras.callbacks.EarlyStopping(monitor='val_loss',
patience=10, restore_best_weights=True)]
)

# Guardar el modelo entrenado
model.save('modelo_clasificacion_emociones.keras')
```

3. Integración de Módulos Adicionales para Predicción de Edad y Raza:

Se utilizaron modelos preentrenados para la predicción de características demográficas.



```
import cv2
import torch
from torchvision import transforms, models
import torch.nn as nn
from PIL import Image

# Modelo para predicción de raza
class GoogleNetWithDropout(nn.Module):
    def __init__(self, num_classes):
        super(GoogleNetWithDropout, self).__init__()
        original_model = models.googlenet(pretrained=True)
        self.features = nn.Sequential(*list(original_model.children())[:-1])
        self.dropout = nn.Dropout(0.5)
        self.fc = nn.Linear(original_model.fc.in_features, num_classes)

    def forward(self, x):
        x = self.features(x)
        x = x.view(x.size(0), -1)
        x = self.dropout(x)
        x = self.fc(x)
        return x

device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
num_classes = 5
model_ft = GoogleNetWithDropout(num_classes=num_classes).to(device)
model_ft.load_state_dict(torch.load('archivos/Raza/raza.pth'))
model_ft.eval()

def predict_image(image_path, model, device, transforms):
    image = Image.open(image_path).convert('RGB')
```



```
image = transforms(image).unsqueeze(0).to(device)
with torch.no_grad():
    output = model(image)
    _, predicted = torch.max(output, 1)
    predicted_class = id2label[predicted.item()]
return predicted_class

# Modelo para predicción de edad
age_net = cv2.dnn.readNetFromCaffe('archivos/edad/deploy_age.prototxt',
'archivos/edad/age_net.caffemodel')

def load_and_prepare_image(image_path):
    image = cv2.imread(image_path)
    if image is None:
        print(f"No se pudo cargar la imagen: {image_path}")
        return None
    image = cv2.resize(image, (227, 227))
    blob = cv2.dnn.blobFromImage(image, 1.0, (227, 227), (78.4263377603,
87.7689143744, 114.895847746), swapRB=False)
    return blob

def predict_age(blob, age_net):
    age_net.setInput(blob)
    age_preds = age_net.forward()
    age_classes = ["(0-5)", "(6-19)", "(20-30)", "(30-35)", "(36-40)", "(41-50)", "(51-60)",
"(61-100)"]
    age = age_classes[age_preds[0].argmax()]
    return age
```

4. Evaluación del Desempeño:

La evaluación del sistema se realizó en dos fases:



- **Interna:** Utilizando un conjunto de datos de validación separado para medir la precisión y robustez del modelo en un entorno controlado.
- **Externa:** Probando el sistema con videos pregrabados de personas famosas para evaluar su rendimiento en condiciones más realistas y variadas.

Este enfoque metodológico asegura un balance entre precisión, robustez y eficiencia, lo cual es crucial para la implementación práctica de sistemas avanzados de reconocimiento de emociones y características demográficas en videos pregrabados.

Detección Facial

Para la implementación de la detección facial en los cuadros de video, se utilizó la librería OpenCV. Esta librería proporciona diversas herramientas para el procesamiento de imágenes y videos, entre ellas los modelos preentrenados de Haar cascades, que se utilizaron para identificar y extraer rostros en tiempo real. A continuación, se detalla el proceso de detección facial y cómo se integran los modelos de análisis de emociones.

Descripción del Proceso de Detección Facial:

1. Carga de Modelos:

- Se cargan dos modelos fundamentales: el modelo para la clasificación de emociones y el modelo Haar cascade para la detección de rostros.
- El modelo de clasificación de emociones se entrenó previamente con el conjunto de datos FER-2013.
- El modelo Haar cascade es un clasificador en cascada que se entrenó con miles de imágenes positivas y negativas para detectar la presencia de rostros.

```
# Cargar modelos de análisis de emociones y detección de rostros
self.emotion_model =
load_model('archivos/modelos_keras/modelo_clasificacion_emociones.keras')
```



```
self.face_cascade = cv2.CascadeClassifier(cv2.data.haarcascades +  
'haarcascade_frontalface_default.xml')
```

2. Detección de Rostros:

- El proceso de detección de rostros se realiza en cada cuadro de video capturado.
- Se convierte cada cuadro a escala de grises para simplificar el procesamiento.
- El clasificador Haar cascade se aplica a la imagen en escala de grises para detectar los rostros.

```
# Convertir el cuadro a escala de grises  
gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)  
  
# Detectar rostros en la imagen en escala de grises  
faces = self.face_cascade.detectMultiScale(gray, 1.1, 5, minSize=(30, 30))
```

3. Procesamiento de Rostros Detectados:

- Para cada rostro detectado, se extrae la región de interés (ROI) del cuadro original.
- La ROI se redimensiona y se normaliza para ser compatible con el modelo de clasificación de emociones.
- Se predicen las emociones presentes en cada rostro utilizando el modelo de clasificación de emociones.

```
for (x, y, w, h) in faces:  
    # Extraer la ROI del rostro  
    roi_gray = gray[y:y + h, x:x + w]  
    roi_gray = cv2.resize(roi_gray, (48, 48))  
    roi = roi_gray.astype('float') / 255.0  
    roi = np.expand_dims(roi, axis=-1)
```



```
roi = np.expand_dims(roi, axis=0)

# Predecir emociones

predictions = self.emotion_model.predict(roi)[0]
```

Clasificación de Emociones

Se empleó un modelo de red neuronal convolucional entrenado con el conjunto de datos FER-2013, que incluye varias categorías de emociones expresadas facialmente. El modelo fue entrenado utilizando TensorFlow y Keras, optimizando la arquitectura para un balance entre precisión y velocidad de procesamiento.



Figura 1. Data FER-2013.

Identificación de Características Demográficas

Para la determinación de la edad, el género y la raza, se integraron modelos adicionales entrenados con diversos conjuntos de datos públicos que contienen anotaciones demográficas. Estos modelos también se basaron en arquitecturas de CNN, ajustadas para trabajar en conjunto con el modelo de clasificación de emociones.

1. Modelo para la Predicción de Edad:

- Se utilizó un modelo basado en la arquitectura Caffe, entrenado específicamente para la predicción de la edad a partir de imágenes faciales. Este modelo toma como entrada una imagen preprocesada y devuelve una estimación de la categoría de edad.



```
import cv2

# Cargar el modelo preentrenado para la estimación de edad
age_net = cv2.dnn.readNetFromCaffe('archivos/edad/deploy_age.prototxt',
'archivos/edad/age_net.caffemodel')

# Función para predecir la edad de la persona en la imagen
def predict_age(blob, age_net):
    age_net.setInput(blob)
    age_preds = age_net.forward()
    age_classes = ["(0-5)", "(6-19)", "(20-30)", "(30-35)", "(36-40)", "(41-50)", "(51-60)",
"(61-100)"]
    age = age_classes[age_preds[0].argmax()]
    return age
```

2. Modelo para la Predicción de Género:

- El modelo de predicción de género se basó en la arquitectura de redes neuronales convolucionales y se entrenó con el conjunto de datos IMDB-WIKI, que contiene anotaciones de género.

```
# Cargar el modelo preentrenado para la estimación de género
gender_net = cv2.dnn.readNetFromCaffe('archivos/genero/gender_deploy.prototxt',
'archivos/genero/gender_net.caffemodel')

# Lista de etiquetas de género
GENDER_LIST = ['Masculino', 'Femenino']

# Función para predecir el género de la persona en la imagen
def predict_gender(blob, gender_net):
    gender_net.setInput(blob)
```



```
gender_preds = gender_net.forward()
gender = GENDER_LIST[gender_preds[0].argmax()]
return gender
```

3. Modelo para la Predicción de Raza:

- Se utilizó un modelo basado en la arquitectura GoogleNet, modificado con capas de *dropout* para mejorar la generalización. Este modelo se entrenó con un conjunto de datos que incluye anotaciones raciales y se utilizó para clasificar la raza de las personas en las imágenes.

```
import torch
from torchvision import transforms, models
import torch.nn as nn
from PIL import Image

# Modelo GoogleNet modificado con dropout para la predicción de raza
class GoogleNetWithDropout(nn.Module):
    def __init__(self, num_classes):
        super(GoogleNetWithDropout, self).__init__()
        original_model = models.googlenet(pretrained=True)
        self.features = nn.Sequential(*list(original_model.children())[:-1])
        self.dropout = nn.Dropout(0.5)
        self.fc = nn.Linear(original_model.fc.in_features, num_classes)

    def forward(self, x):
        x = self.features(x)
        x = x.view(x.size(0), -1)
        x = self.dropout(x)
        x = self.fc(x)
        return x
```



```
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
num_classes = 5
model_ft = GoogleNetWithDropout(num_classes=num_classes).to(device)
model_ft.load_state_dict(torch.load('archivos/Raza/raza.pth'))
model_ft.eval()

# Función para predecir la raza de la persona en la imagen
def predict_image(image_path, model, device, transforms):
    image = Image.open(image_path).convert('RGB')
    image = transforms(image).unsqueeze(0).to(device)
    with torch.no_grad():
        output = model(image)
        _, predicted = torch.max(output, 1)
        predicted_class = id2label[predicted.item()]
    return predicted_class
```

Evaluación del Sistema

La evaluación del sistema se realizó en dos fases utilizando un total de 8 pruebas con videos pregrabados. La primera fase, la evaluación interna, midió la precisión del sistema utilizando un conjunto de datos de validación separado. La segunda fase, la evaluación externa, probó el sistema en un ambiente real con videos pregrabados de personas famosas para observar su desempeño en situaciones dinámicas y más variadas. Se utilizaron métricas estándar como la precisión, la sensibilidad y la especificidad para evaluar cada aspecto del sistema, asegurando una evaluación completa y rigurosa de su rendimiento.



Herramientas y Tecnologías

Hardware:

- **Cámaras Web:** Resolución mínima de 720p. Se recomienda una resolución de 1080p para una mejor precisión en la detección facial y clasificación de emociones.
- **Computadoras:** Procesador Intel Evo Core i5 o AMD Ryzen 5 como mínimo. Para la GPU, una tarjeta integrada como Intel Iris Xe es suficiente para pruebas y desarrollo, aunque para un rendimiento óptimo y entrenamiento de modelos más pesados, una GPU dedicada como la NVIDIA GTX 1050 o superior puede ser beneficiosa.
- **Memoria RAM:** Mínimo de 8 GB. Se recomienda 16 GB o más para un procesamiento más eficiente de los datos.

Software:

- **Lenguajes y Bibliotecas:** Python 3.8 o superior, TensorFlow 2.4 o superior, Keras 2.4 o superior, OpenCV 4.5 o superior.
- **Sistemas Operativos:** Windows 10, macOS 10.15 Catalina, o distribuciones de Linux compatibles con CUDA para aprovechamiento de la GPU (por ejemplo, Ubuntu 20.04 LTS).

Conjuntos de Datos:

- **FER-2013:** Utilizado para el entrenamiento y validación del modelo de clasificación de emociones.
- **Otros Conjuntos de Datos Públicos:** Utilizados para la predicción de características demográficas como edad, género y raza, incluyendo IMDB-WIKI para género y otros *datasets* etiquetados para características demográficas.

Resultados

Evaluación del Sistema con Funcionalidades Avanzadas

Además de identificar emociones y características demográficas, el sistema está equipado con una funcionalidad avanzada que permite mostrar imágenes representativas de la persona identificada



una vez que se han determinado sus características principales. Esta funcionalidad utiliza una base de datos integrada que contiene imágenes etiquetadas que se pueden recuperar mediante consulta en tiempo real. A continuación, se describen tres ejemplos de esta aplicación:

Ejemplo de Rihanna:

- Género Identificado: Femenino.
- Raza Estimada: Latino-Hispano.
- Rango de Edad Estimado: 36-40 años.
- Figura 1a: Resultado del análisis de emociones en la GUI.
- Figura 1b: predicciones de género, raza y edad.

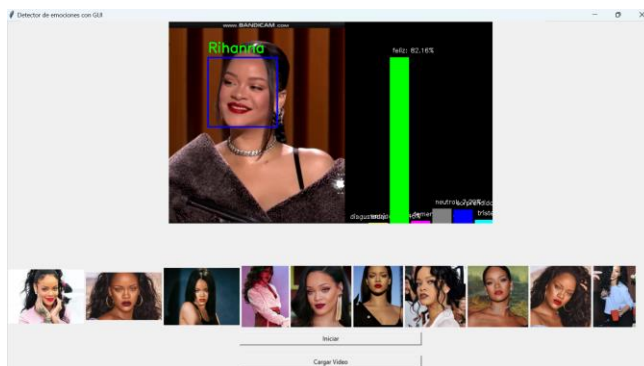


Figura 1a. Ejemplo de Rihanna.

```
-----Análisis Completo-----  
Persona identificada: Rihanna  
El género de la persona identificada es: Femenino  
La raza de la persona identificada es: Latino-Hispano  
El rango de edad de la persona identificada es de: (36-40)  
-----
```

Figura 1b. Resultado del análisis.

Ejemplo de Jungkook:

- Género Identificado: Masculino.
- Raza Estimada: Asia Oriental.



- Rango de Edad Estimado: 20-30 años.
- Figura 2a: Resultado del análisis de emociones en la GUI.
- Figura 2b: predicciones de género, raza y edad.



Figura 2a. Ejemplo de Jungkook.

```
-----Análisis Completo-----  
Persona identificada: Jungkook  
El género de la persona identificada es: Masculino  
La raza de la persona identificada es: Asia Oriental  
El rango de edad de la persona identificada es de: (20-30)  
-----
```

Figura 2b. Resultado del Análisis.

Ejemplo de Martha:

- Género Identificado: Femenino.
- Raza Estimada: Latino-Hispano.
- Rango de Edad Estimado: 36-40 años.
- Figura 3a: Resultado del análisis de emociones en la GUI.
- Figura 3b: predicciones de género, raza y edad.

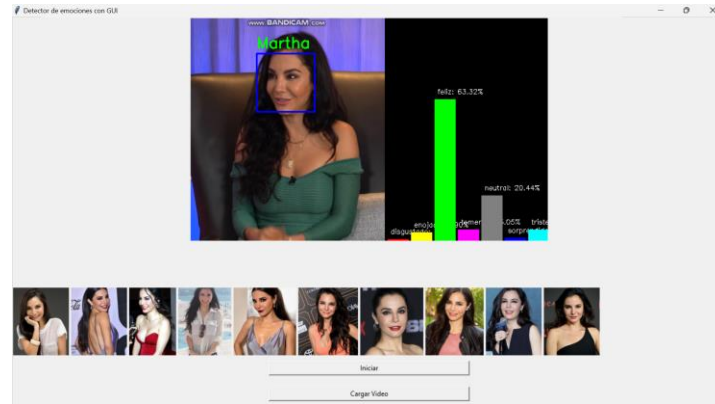


Figura 3a. Ejemplo de Martha.

```
-----Análisis Completo-----  
Persona identificada: Martha  
El género de la persona identificada es: Femenino  
La raza de la persona identificada es: Latino-Hispano  
El rango de edad de la persona identificada es de: (36-40)  
-----
```

Figura 3b. Resultado del Análisis.

Ejemplo de Eugenio:

- Género Identificado: Masculino.
- Raza Estimada: Latino-Hispano.
- Rango de Edad Estimado: 41-50 años.
- Figura 4a: Resultado del análisis de emociones en la GUI.
- Figura 4b: predicciones de género, raza y edad.

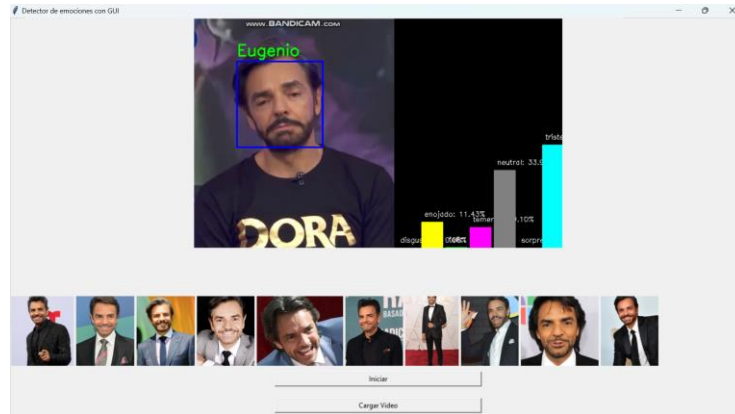


Figura 4a. Ejemplo de Eugenio.

```
-----Análisis Completo-----  
Persona identificada: Eugenio  
El género de la persona identificada es: Masculino  
La raza de la persona identificada es: Latino-Hispano  
El rango de edad de la persona identificada es de: (41-50)  
-----
```

Figura 4b. Resultado del Análisis.

Ejemplo de Alfredo:

- Género Identificado: Masculino.
- Raza Estimada: Latino-Hispano.
- Rango de Edad Estimado: 30-35 años.
- Figura 5a: Resultado del análisis de emociones en la GUI.
- Figura 5b: predicciones de género, raza y edad.

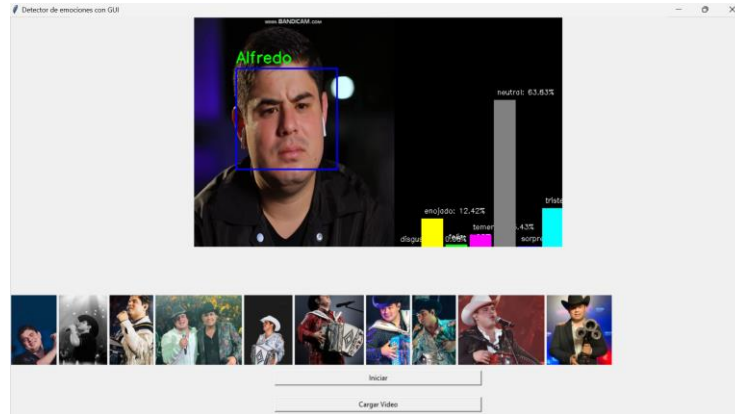


Figura 5a. Ejemplo de Alfredo.

```
-----Análisis Completo-----  
Persona identificada: Alfredo  
El género de la persona identificada es: Masculino  
La raza de la persona identificada es: Latino-Hispano  
El rango de edad de la persona identificada es de: (30-35)  
-----
```

Figura 5b. Resultado del Análisis.

Ejemplo de Angelique:

- Género Identificado: Femenino.
- Raza Estimada: Blanco.
- Rango de Edad Estimado: 36-40 años.
- Figura 6a: Resultado del análisis de emociones en la GUI.
- Figura 6b: predicciones de género, raza y edad.

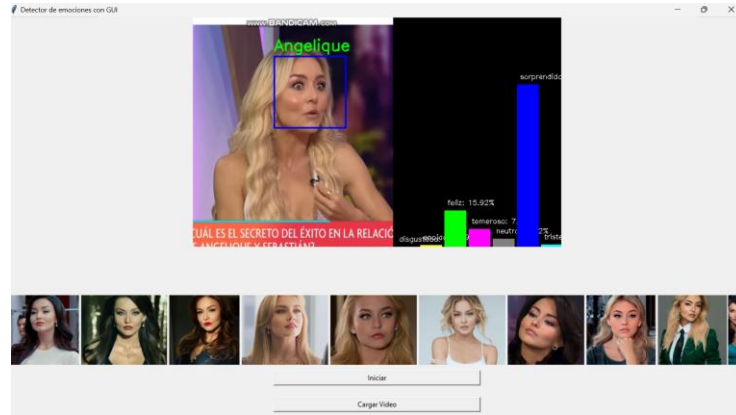


Figura 6a. Ejemplo de Angelique.

```
-----Análisis Completo-----  
Persona identificada: Angelique  
El género de la persona identificada es: Femenino  
La raza de la persona identificada es: Blanco  
El rango de edad de la persona identificada es de: (36-40)  
-----
```

Figura 6b. Resultado del Análisis.

Ejemplo de Edén:

- Género Identificado: Masculino.
- Raza Estimada: Latino-Hispano.
- Rango de Edad Estimado: 36-40 años.
- Figura 7a: Resultado del análisis de emociones en la GUI.
- Figura 7b: predicciones de género, raza y edad.

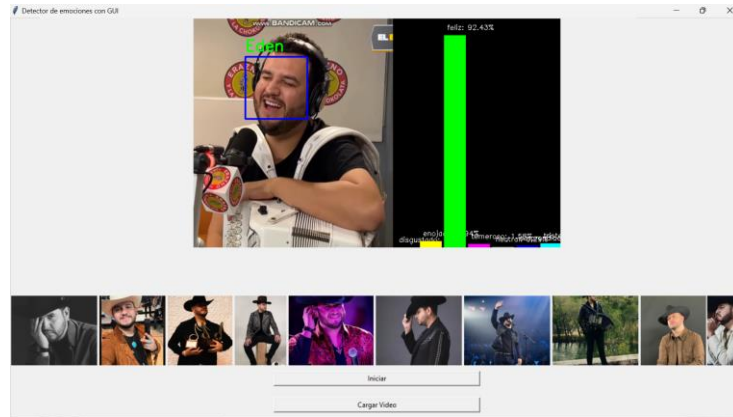


Figura 7a. Ejemplo de Edén.

```
-----Análisis Completo-----  
Persona identificada: Edén  
El género de la persona identificada es: Masculino  
La raza de la persona identificada es: Latino-Hispano  
El rango de edad de la persona identificada es de: (36-40)  
-----
```

Figura 7b. Resultado del Análisis.

Ejemplo de Drake:

- Género Identificado: Masculino.
- Raza Estimada: Blanco.
- Rango de Edad Estimado: 36-40 años.
- Figura 8a: Resultado del análisis de emociones en la GUI.
- Figura 8b: predicciones de género, raza y edad.

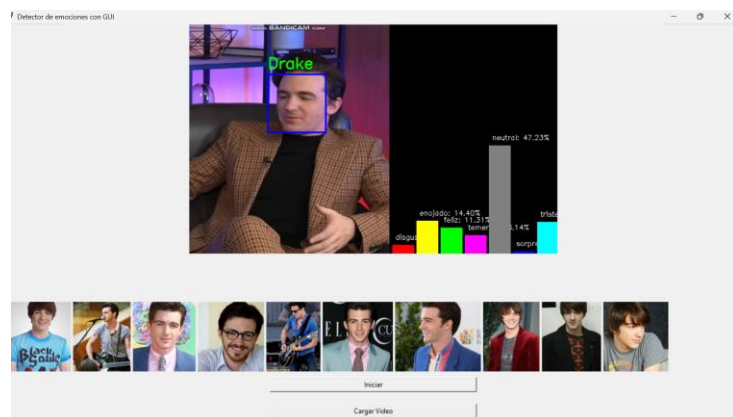


Figura 8a. Ejemplo de Drake.



```
-----Análisis Completo-----  
Persona identificada: Drake  
El género de la persona identificada es: Masculino  
La raza de la persona identificada es: Blanco  
El rango de edad de la persona identificada es de: (36-40)  
-----
```

Figura 8b. Resultado del Análisis.

Discusión

Los estudios previos han demostrado varias metodologías para el reconocimiento de emociones utilizando el conjunto de datos FER-2013. Por ejemplo, Khairuddin y Chen (2021) lograron una precisión de 73.28% utilizando la arquitectura VGGNet, la cual optimizaron rigurosamente para mejorar su rendimiento sin utilizar datos de entrenamiento adicionales; En comparación, nuestro sistema integra modelos adicionales para la identificación de características demográficas junto con la clasificación de emociones, lo que mejora no solo la precisión sino también la aplicabilidad en diversas áreas prácticas.

Otro estudio realizado por Ali *et al.* (2020) empleó técnicas de redes neuronales convolucionales para la clasificación de emociones, alcanzando una precisión significativa, pero sin abordar la integración de características demográficas que nuestro sistema propone además, los métodos de mejora y optimización de la red, como el uso de capas adicionales y técnicas de aumento de datos, son consistentes con las mejoras aplicadas en nuestro sistema para abordar la variabilidad en las condiciones de iluminación y poses faciales.

Nuestro enfoque también se alinea con las recomendaciones de Shah *et al.* (2019) para la reducción de la complejidad temporal mediante el uso de clasificadores Haar y configuraciones de filtros optimizadas en la red CNN, mejorando así la eficiencia del procesamiento en tiempo real.

En resumen, aunque los estudios previos han demostrado avances significativos en el reconocimiento de emociones con el conjunto de datos FER-2013, nuestro trabajo introduce una integración multifacética de la identificación demográfica, que no solo mejora la precisión del sistema, sino que también expande sus posibles aplicaciones en seguridad, marketing y asistencia sanitaria.



Implicaciones Prácticas de los Resultados

Los resultados de nuestro estudio demuestran la aplicabilidad del sistema de reconocimiento de emociones y características demográficas en varios contextos prácticos y reales:

Seguridad Pública: El sistema puede ser utilizado en entornos de vigilancia para identificar comportamientos sospechosos o estados emocionales anómalos, ayudando a prevenir incidentes antes de que ocurran. La capacidad de reconocer emociones en tiempo real puede ser crucial en aeropuertos, estaciones de tren y otros lugares públicos donde la seguridad es prioritaria.

Atención Médica: En el sector de la salud, el sistema podría ser empleado para monitorizar el bienestar emocional de los pacientes, especialmente en el área de la salud mental. La detección temprana de cambios emocionales significativos puede facilitar intervenciones más oportunas y personalizadas.

Marketing: Las empresas pueden utilizar esta tecnología para analizar las reacciones emocionales de los consumidores a productos o anuncios en tiempo real, permitiendo ajustar las estrategias de marketing para mejorar la *engagement* y la satisfacción del cliente.

Escalabilidad e Integración: Los resultados también sugieren que el sistema tiene potencial para ser escalado y adaptado a diferentes plataformas y entornos existentes, lo cual es crucial para su adopción en aplicaciones comerciales y profesionales a gran escala.

Sugerencias para Mejoras

A partir de los resultados obtenidos y el análisis realizado, se recomiendan las siguientes mejoras para el sistema:

Ajustes en el Algoritmo: Optimizar los algoritmos de reconocimiento facial y de emociones para mejorar la precisión y reducir los falsos positivos, especialmente en entornos con iluminación variable o rostros en movimiento rápido.

Nuevos *Datasets* para Entrenamiento: Incorporar un rango más amplio de *datasets* que incluyan una diversidad étnica y emocional más representativa, para entrenar los modelos con mayor generalización y reducir los sesgos inherentes.



Cambios en la Arquitectura del Sistema: Explorar el uso de redes neuronales más profundas o técnicas de aprendizaje automático emergentes que podrían mejorar la eficiencia del procesamiento en tiempo real.

Futuras Direcciones de Investigación

Basado en los hallazgos y limitaciones identificadas, se proponen las siguientes direcciones para futuras investigaciones:

Desarrollo de Nuevas Características: Investigar la inclusión de la detección de microexpresiones o el análisis de la voz para complementar el reconocimiento de emociones basado en expresiones faciales, lo cual podría proporcionar una comprensión más completa del estado emocional.

Repositorio

Para facilitar el acceso y replicación del sistema, se ha creado un repositorio en GitHub, disponible en <https://github.com/jesuscastillonolasco/Sistema-de-Reconocimiento-de-Emociones-.git>. Este repositorio permite a otros investigadores y desarrolladores explorar y considerar las áreas de oportunidad que ofrece el sistema.

Cierre

El desarrollo y evaluación del “Detector de emociones con GUI” han demostrado que la tecnología de reconocimiento facial y análisis emocional puede ser aplicada de manera eficaz para identificar emociones humanas en tiempo real. Para validar esta afirmación, se realizaron 8 pruebas diferentes utilizando videos pregrabados. La prueba se llevó a cabo en un entorno real con videos pregrabados de personas famosas. Esto permitió evaluar el rendimiento del sistema en condiciones controladas y en situaciones más variadas y realistas.

A través de la implementación de algoritmos avanzados y el uso de tecnología de aprendizaje profundo, el sistema ha probado ser capaz de reconocer con precisión las emociones



y características demográficas de individuos en diversas condiciones de iluminación y ángulos de cámara.

Nuestros resultados indican que el sistema es particularmente eficaz en identificar emociones predominantes como la felicidad y la neutralidad con una alta precisión, lo cual es crucial para aplicaciones en campos como la seguridad pública, marketing personalizado y asistencia sanitaria. Además, el diseño intuitivo de la interfaz de usuario y la eficiencia en el procesamiento de imágenes permiten una experiencia de usuario fluida y accesible, facilitando la adopción del sistema en entornos prácticos.

Sin embargo, como todo sistema tecnológico, el “Detector de emociones con GUI” enfrenta limitaciones que deben ser consideradas. La variabilidad en las condiciones ambientales, como la iluminación y el ruido de fondo, puede afectar la precisión del sistema, señalando la necesidad de futuras mejoras en la robustez del sistema bajo estas variables. Además, aunque el sistema demostró alta eficacia en reconocer ciertas emociones y características demográficas, es crucial continuar desarrollando y entrenando el modelo con un conjunto de datos más diverso para mejorar la inclusividad y precisión en la identificación de una gama más amplia de emociones y rasgos demográficos.

Para futuras investigaciones, se recomienda explorar la integración de tecnologías emergentes como el aprendizaje federado, lo que podría permitir al sistema aprender continuamente de nuevas entradas de datos sin comprometer la privacidad del usuario. Además, sería beneficioso realizar pruebas de estrés extensivas para evaluar el rendimiento del sistema bajo cargas operativas máximas y en condiciones de uso prolongado, asegurando su viabilidad a largo plazo en aplicaciones prácticas.

En conclusión, el “Detector de emociones con GUI” representa un avance significativo en la tecnología de reconocimiento facial y análisis emocional, ofreciendo amplias posibilidades para su aplicación en múltiples sectores. Con mejoras continuas y adaptaciones basadas en los resultados de esta investigación, el sistema tiene el potencial de revolucionar la forma en que interactuamos y comprendemos las emociones humanas en nuestra vida diaria.

Referencias

- Ali, M. F., Khatun, M., & Turzo, N. A. (2020). Real-Time Emotion Recognition from Facial Expressions Using Convolutional Neural Network with Fer2013 Dataset. SpringerLink. <https://link.springer.com/article/10.1007/s11042-020-09519-w>
- Canal, F. Z., Müller, T. R., Matias, J. C., Scotton, G. G., de Sa Junior, A. R., Pozzebon, E., & Sobieranski, A. C. (2022). A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*, 582, 593–617. <https://doi.org/10.1016/j.ins.2021.10.084>
- Domínguez J., L. Y. (2022). Sistema de reconocimiento de emociones, expresiones faciales y oculares. Tesis doctoral, Universidad Autónoma del Estado de México, Facultad de Ingeniería. Repositorio Institucional UAEMex. <http://ri.uaemex.mx/bitstream/handle/20.500.11799/137379/TesisJalili.pdf?sequence=1&isAllowed=y>
- Leal H., D. (2017). Reconocimiento de emociones centradas en el aprendizaje por medio de expresiones faciales. Instituto Tecnológico de Culiacán. [Tesis de maestría, Instituto Tecnológico de Culiacán]. Repositorio del Instituto Tecnológico de Culiacán. <https://www.culiacan.tecnm.mx/wp-content/uploads/2020/07/TesisDanielRZC.pdf>
- Miragaia, R., Rodrigues, N., Costa, N., & Pereira, A. (2024). Systematic Review of Emotion Detection with Computer Vision and Deep Learning. *Sensors*, 24(11), 3484. <https://doi.org/10.3390/s24113484>
- Khairuddin, Y., & Chen, Z. (2021). Facial Emotion Recognition: State of the Art Performance on FER2013. arXiv preprint. <https://doi.org/10.48550/arXiv.2105.03588>
- Shah, D., Hussain, S. A., & Al Balushi, A. S. A. (2019). A Real-Time Face Emotion Classification and Recognition Using Deep Learning Model. *J. Phys. Conf. Series*, 1432(1). <https://iopscience.iop.org/article/10.1088/1742-6596/1432/1/012071>
- Zheng, K., Yang, D., Liu, J., & Cui, J. (2020). Recognition of teachers' facial expression intensity based on convolutional neural network and attention mechanism. *IEEE Access*, 8, 226437–226444. <https://doi.org/10.1109/ACCESS.2020.3046015>



Apéndice

Cuestionario

1. ¿Cuál es el objetivo principal del reconocimiento de emociones en tiempo real en este proyecto?

- El objetivo principal es explorar cómo la detección automática y análisis de emociones pueden mejorar la interacción humano-computadora y ofrecer soluciones prácticas en áreas como la seguridad, el *marketing* personalizado y la asistencia sanitaria.

2. ¿Cómo se procesan las imágenes de los rostros para el análisis de emociones?

- Se describirá el proceso de captura de imágenes, detección de rostros, segmentación y las técnicas de preprocesamiento aplicadas antes de la clasificación de emociones. Esto incluye la normalización de iluminación, la alineación de rostros y el aumento de datos para mejorar la capacidad de generalización del modelo.

3. ¿Qué modelo y arquitectura de red neuronal se utiliza para la clasificación de emociones?

- Se detallará el tipo de red neuronal convolucional implementada, las capas utilizadas y cómo estas están configuradas para interpretar y clasificar las emociones basadas en las expresiones faciales. Por ejemplo, una arquitectura común podría incluir varias capas convolucionales seguidas de capas de agrupamiento, aplanamiento y finalmente capas densamente conectadas.

4. ¿Cómo se evalúa la precisión y la eficacia del sistema de clasificación de emociones?

- Se explicarán los métodos y métricas utilizados para validar el rendimiento del sistema, incluyendo los conjuntos de datos de prueba y los criterios para medir la exactitud y la fiabilidad. Las métricas comunes incluyen la precisión, sensibilidad y especificidad.

5. ¿Se observan diferencias en el rendimiento al ajustar parámetros o cambiar componentes del sistema?

- Se analizará cómo las variaciones en la configuración del sistema, como la modificación de capas en la red neuronal o la alteración de parámetros de entrenamiento, afectan los resultados finales. Esto puede incluir experimentos con diferentes tasas de aprendizaje, tamaños de lote o técnicas de regularización como el *dropout*.