



# Tecnológico de Monterrey

Inteligencia artificial avanzada para la ciencia de datos 2

Gpo 501

## **Docentes**

Dr. Benjamín Valdés Aguirre

Ma. Eduardo Daniel Juárez Pineda

Dr. Ismael Solis Moreno

Dr. José Antonio Cantoral Ceballos

Dr. Carlos Alberto Dorantes Dosamantes

## **Integrantes**

Carlos Rodrigo Salguero Alcántara	A00833341
Diego Perdomo Salcedo	A01709150
Dafne Fernández Hernández	A01369230
José Emiliano Riosmena Castañón	A01704245
Luis Arturo Rendón Iñarritu	A01703572

<b>1.0 Introducción</b>	<b>3</b>
1.1 Propósito	3
<b>2.0 Diseño de pruebas</b>	<b>3</b>
2.1 Separación del Dataset (entrenamiento, validación y prueba)	3
<b>3.0 Problema, técnicas de modelado, supuestos</b>	<b>4</b>
3.1 Técnica utilizada	4
3.2 Razonamiento	5
3.3 Consideraciones de usar otras técnicas	5
3.4 Supuestos:	6
<b>4.0 Modelado</b>	<b>7</b>
4.1 Descripción/Arquitectura del modelo	7
4.2 Capa de Aumentación de Datos	7
4.2.1 Modelo Diurno	7
4.2.2 Modelo Nocturno	8
4.3 Pipeline de procesamiento	8
4.4 Hiperparámetros	8
<b>5.0 Resultados del Modelo</b>	<b>9</b>
5.1 Rendimiento Diurno	9
5.1.1 Métricas de Entrenamiento	9
5.1.2 Métricas de Evaluación	10
5.1.3 Matriz de Confusión	11
5.1.4 Resultados en Conjunto de Pruebas	11
5.1.5 Resumen	12
5.2 Rendimiento Nocturno	12
5.2.1 Métricas de Entrenamiento	12
5.2.2 Métricas de Evaluación	13
5.2.3 Matriz de Confusión	14
5.2.4 Resultados en Conjunto de Pruebas	14
5.2.5 Resumen	15
<b>5.3 Limitaciones y Consideraciones</b>	<b>15</b>
<b>6.0 Evaluando los resultados</b>	<b>16</b>
6.1 Validación de Objetivos	16
6.2 Sigüientes pasos	16

## 1.0 Introducción

Este documento técnico describe el proceso de desarrollo, implementación y evaluación de un sistema de detección y conteo automatizado de ganado bovino mediante técnicas de visión computacional. Este documento describe el modelo ResNet50V2 que se optimizó el para cada condición específica.

### 1.1 Propósito

El propósito principal de este documento es establecer y documentar el proceso de selección, implementación y evaluación de técnicas de modelado para lograr la detección y conteo preciso de vacas en imágenes de ganado bovino.

Este documento sirve como guía técnica y referencia para el equipo de desarrollo, stakeholders y futuros mantenedores del sistema, asegurando la transparencia en el proceso de desarrollo y facilitando la comprensión de las decisiones técnicas tomadas durante el proyecto.

## 2.0 Diseño de pruebas

Para probar la calidad del modelo hemos tomado la decisión de dividir el dataset en entrenamiento, validación y prueba. De esta forma nos aseguramos de validar y probar adecuadamente la precisión del modelo respecto a nuestro objetivo de minería de datos..

### 2.1 Separación del Dataset (entrenamiento, validación y prueba)

Generamos un script que hace una división estratificada del conjunto de datos en tres subconjuntos: entrenamiento, validación y prueba, con proporciones de 75%, 15% y 10% respectivamente. El programa mantiene la distribución proporcional de las categorías en cada subconjunto, lo que significa que conserva la representación relativa de cada clase en los tres splits. La división se realiza de manera aleatoria y procesa los archivos en paralelo utilizando múltiples hilos para mejorar el rendimiento. Los datos divididos se guardan en directorios separados manteniendo la estructura original de las categorías.

## 2.2 Validación del diseño de pruebas con objetivos de minería de datos

Esta división también nos permite asegurarnos que cumplamos con nuestro objetivo de minería de datos.

### Objetivo de Minería de datos:

- ❖ Determinar la cantidad de vacas en cada imagen en cualquier condición.
  - Un modelo para condiciones diurnas con un 80% de precisión.
  - Un modelo para condiciones nocturnas con un 50% de precisión.

Nuestros criterios de éxito listan que requerimos un 50% de precisión en el modelo con imágenes nocturnas, 80% con imágenes diurnas. Con el diseño de pruebas actual podemos validar estos criterios usando el subconjunto de prueba y así evaluar los resultados.

## 3.0 Problema, técnicas de modelado, supuestos

Nuestro problema se considera de clasificación. Buscamos clasificar las imágenes a base de la cantidad de vacas que se encuentran en una imagen en un momento dado. En esta sección se detallan las técnicas de modelado posibles, la técnica utilizada para resolver el problema.

Las técnicas principales para esto incluyen:

- Redes neuronales convolucionales (CNN)
- Arquitecturas basadas en regiones (R-CNN, Fast R-CNN)
- Single Shot Detectors (SSD)
- Cascade Detectors
- Feature Pyramid Networks (FPN)

### 3.1 Técnica utilizada

Para abordar este problema de clasificación, implementamos una solución basada en deep learning utilizando una arquitectura de red neuronal profunda con transfer learning. **ResNet50V2** es una arquitectura de red neuronal convolucional profunda que introduce conexiones residuales (skip connections) para resolver el problema de la degradación del gradiente en redes profundas. Estas conexiones permiten que la información fluya directamente a través de la red, facilitando el entrenamiento de arquitecturas más profundas.

La versión V2 mejora la arquitectura original reordenando las capas de normalización y activación, lo que resulta en un mejor flujo de gradientes durante el entrenamiento. Esta arquitectura es particularmente efectiva en tareas de clasificación de imágenes gracias a su capacidad para aprender jerarquías complejas de características mientras mantiene la estabilidad durante el entrenamiento

## 3.2 Razonamiento

Razones principales que soportan esta decisión:

### 1. Eficiencia del Transfer Learning

- ❖ Aprovecha conocimiento pre-entrenado sobre detección de formas y patrones
- ❖ Permite fine-tuning específico para condiciones nocturnas
- ❖ Optimiza recursos computacionales

### 2. Arquitectura adecuada para baja iluminación

- ❖ Las conexiones residuales ayudan a preservar detalles en condiciones de bajo contraste
- ❖ Capacidad de aprender características en diferentes escalas de intensidad
- ❖ Robustez ante ruido y variaciones en la calidad de imagen
- ❖ Sistema de normalización que ayuda con rangos de intensidad bajos

Esta elección representa un balance óptimo entre capacidad de procesamiento, precisión en condiciones adversas y eficiencia computacional, específicamente adaptada para el desafío de conteo de vacas en condiciones subóptimas.

## 3.3 Consideraciones de usar otras técnicas

Las demás técnicas fueron descartadas por las siguientes razones:

### 1. CNNs básicas:

- Requieren más capas y parámetros para alcanzar el mismo nivel de precisión
- Menor capacidad para manejar variaciones en iluminación
- Mayor tiempo de entrenamiento desde cero

### 2. Arquitecturas R-CNN:

- Complejidad computacional excesiva para nuestro hardware
- Proceso de dos etapas que ralentiza la inferencia
- Mayor dificultad de optimización para nuestro caso específico

### 3. Single Shot Detectors (SSD):

- Menor precisión en condiciones de baja iluminación
- Dificultad para manejar variaciones extremas en el tamaño de los objetos
- Requieren más ajustes para mantener la estabilidad del entrenamiento

### 4. Cascade Detectors

- Demasiado pesados para implementación en Raspberry Pi
- Tiempo de inferencia no compatible con requisitos en tiempo real
- Complejidad innecesaria para nuestro problema específico

### 5. Feature Pyramid Networks:

- Mayor consumo de recursos computacionales
- Beneficios no justifican la complejidad adicional
- Menos efectivos en condiciones de iluminación variable

## 3.4 Supuestos:

### Supuestos sobre el Entorno:

- ❖ Las condiciones de iluminación permiten distinguir objetos del fondo.
- ❖ No hay factores ambientales extremos (niebla densa).
- ❖ La distancia entre la cámara y los objetos está dentro del rango de entrenamiento.
- ❖ El ángulo de visión es similar al de las imágenes de entrenamiento.

### Supuestos sobre los Datos de Entrada:

- ❖ Las imágenes de entrada tienen una resolución mínima suficiente para distinguir vacas (al menos 224x224 píxeles).
- ❖ Los objetos a detectar (vacas) ocupan al menos un 5% del área total de la imagen.
- ❖ Las imágenes son a color (3 canales RGB).

### Supuestos sobre los Objetos:

- ❖ Las vacas aparecen en posiciones naturales (de pie, pastando, caminando).
- ❖ Los objetos son distinguibles del fondo.

### Supuestos sobre las Condiciones de Operación:

- ❖ La Raspberry Pi opera dentro de su rango de temperatura normal.
- ❖ Hay suficiente memoria disponible para el procesamiento.

- ❖ La fuente de alimentación es estable y suficiente.

## 4.0 Modelado

En esta sección se detallarán las características específicas del modelo implementado, incluyendo su funcionamiento, parámetros, comportamiento y limitaciones.

### 4.1 Descripción/Arquitectura del modelo

El modelo implementado utiliza ResNet50V2 como arquitectura base, aprovechando su sistema innovador de conexiones residuales que permite el entrenamiento efectivo de redes profundas al resolver el problema de la degradación del gradiente. El modelo se configuró específicamente para nuestro caso de uso mediante la adaptación de sus capas finales, complementandose con capas densas para el procesamiento de características específicas de las vacas, normalización por lotes para estabilizar el entrenamiento, y dropout para prevenir el sobreajuste.

La arquitectura se especializa en dos variantes optimizadas para diferentes condiciones de iluminación:

- Modelo Diurno: Configurado para manejar imágenes con alto contraste, enfatizando la detección de bordes y formas definidas.
- Modelo Nocturno: Adaptado para procesar imágenes con bajo contraste, con mayor énfasis en la detección de patrones en condiciones de iluminación limitada.

Ambas variantes mantienen la estructura base de ResNet50V2, pero difieren en sus capas de adaptación final y parámetros de procesamiento, utilizando el optimizador AdamW y funciones de activación ReLU y Softmax para optimizar el rendimiento en sus respectivas condiciones de iluminación.

### 4.2 Capa de Aumentación de Datos

#### 4.2.1 Modelo Diurno

Aumentación	Valor	Probabilidad
Normalización	0.1	100%

Brillo	+ 0.2	80%
Contraste (factor)	1.2	80%
Volteo horizontal		0.5%
Ruido gaussiano	mu de 0, omega de 0.01	0.30%

**Razón:** Se eligieron estos parámetros para simular las variaciones naturales en iluminación diurna y mejorar la robustez del modelo ante diferentes condiciones de luz solar.

#### 4.2.2 Modelo Nocturno

Aumentación	Valor	Probabilidad
Normalización	0.1	100%
Brillo	+ 0.2	80%
Contraste (factor)	1.2	80%
Volteo horizontal		0.5%
Ruido gaussiano	mu de 0, omega de 0.01	0.30%

**Razón:** Los parámetros se ajustaron para manejar las condiciones de baja luminosidad y el ruido característico de las imágenes nocturnas.

### 4.3 Pipeline de procesamiento

1. Capa de aumentación de datos
2. Ruido gaussiano
3. Rescalado (-1. 1)
4. ResNet50V2
5. Global Average Pooling
6. Capa densa final (activación de softmax)

**Razón:** Este pipeline se diseñó para garantizar un procesamiento eficiente y consistente de las imágenes, maximizando la extracción de características relevantes en cada etapa.

### 4.4 Hiperparámetros

- ❖ Learning rate: 5e-4



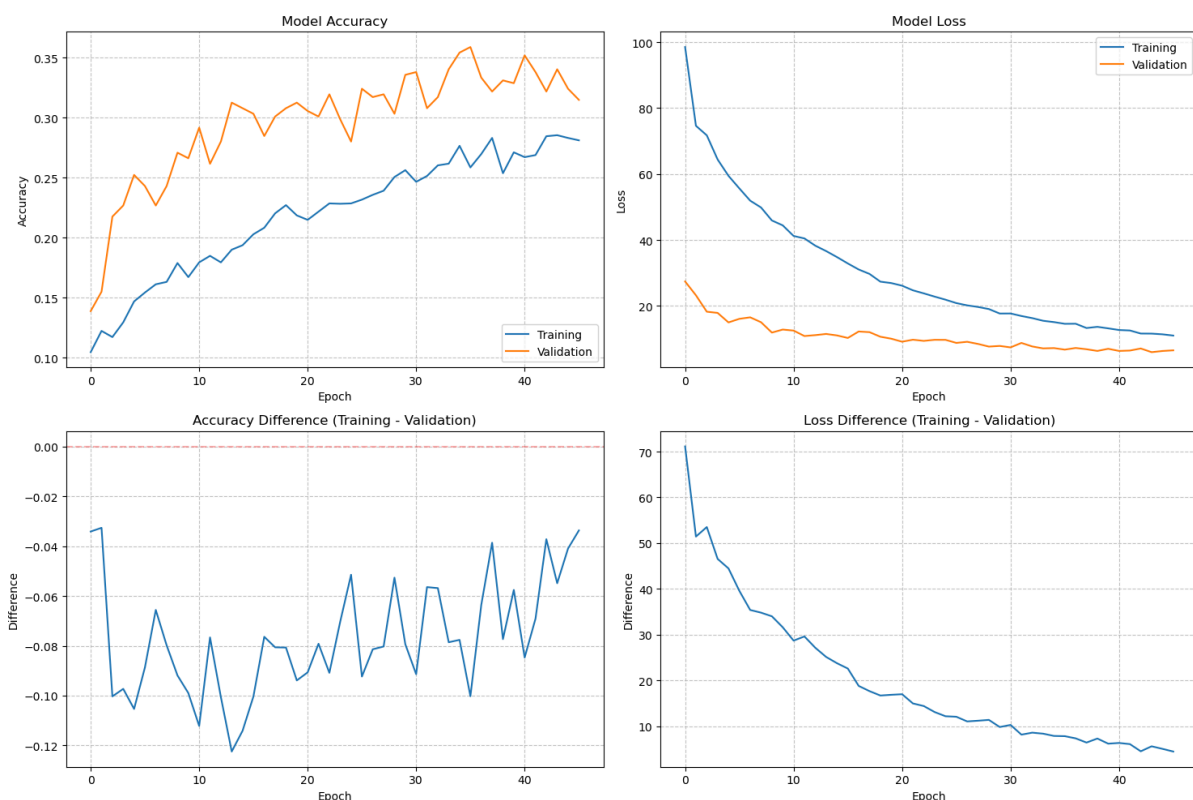
- ❖ Weight decay: 0.01
- ❖ Dropout rate: 0.3
- ❖ Batch size: 32
- ❖ Épocas: 30
- ❖ Optimizadores: AdamW

**Razón:** Estos hiperparámetros fueron seleccionados para balancear la velocidad de convergencia con la estabilidad del entrenamiento, previniendo el sobreajuste mientras se mantiene una tasa de aprendizaje efectiva.

## 5.0 Resultados del Modelo

### 5.1 Rendimiento Diurno

#### 5.1.1 Métricas de Entrenamiento

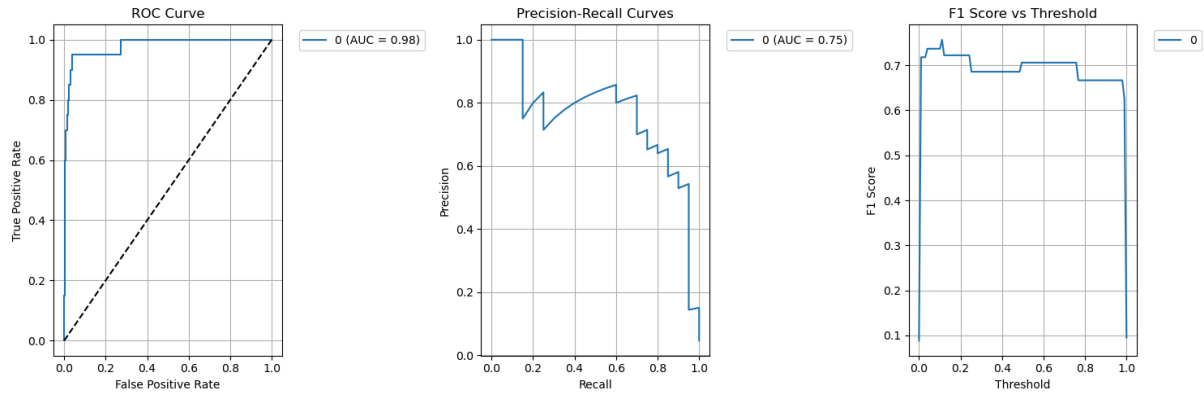


- ❖ Precisión del modelo
  - Muestra cómo mejora la precisión de la imagen a lo largo de las épocas
  - El modelo llega a un aproximado de 32% de precisión en validación y 28% en entrenamiento

#### ❖ Pérdida del modelo

- Muestra como disminuye el error del modelo
- Comienza alto (aproximadamente 100) y se estabiliza cerca de 10

### 5.1.2 Métricas de Evaluación



#### ❖ ROC Curve

- Muestra la relación entre verdaderos positivos y falsos positivos
- AUC de 0.98 indica excelente capacidad discriminativa
- La curva muy por encima de la diagonal es altamente positiva

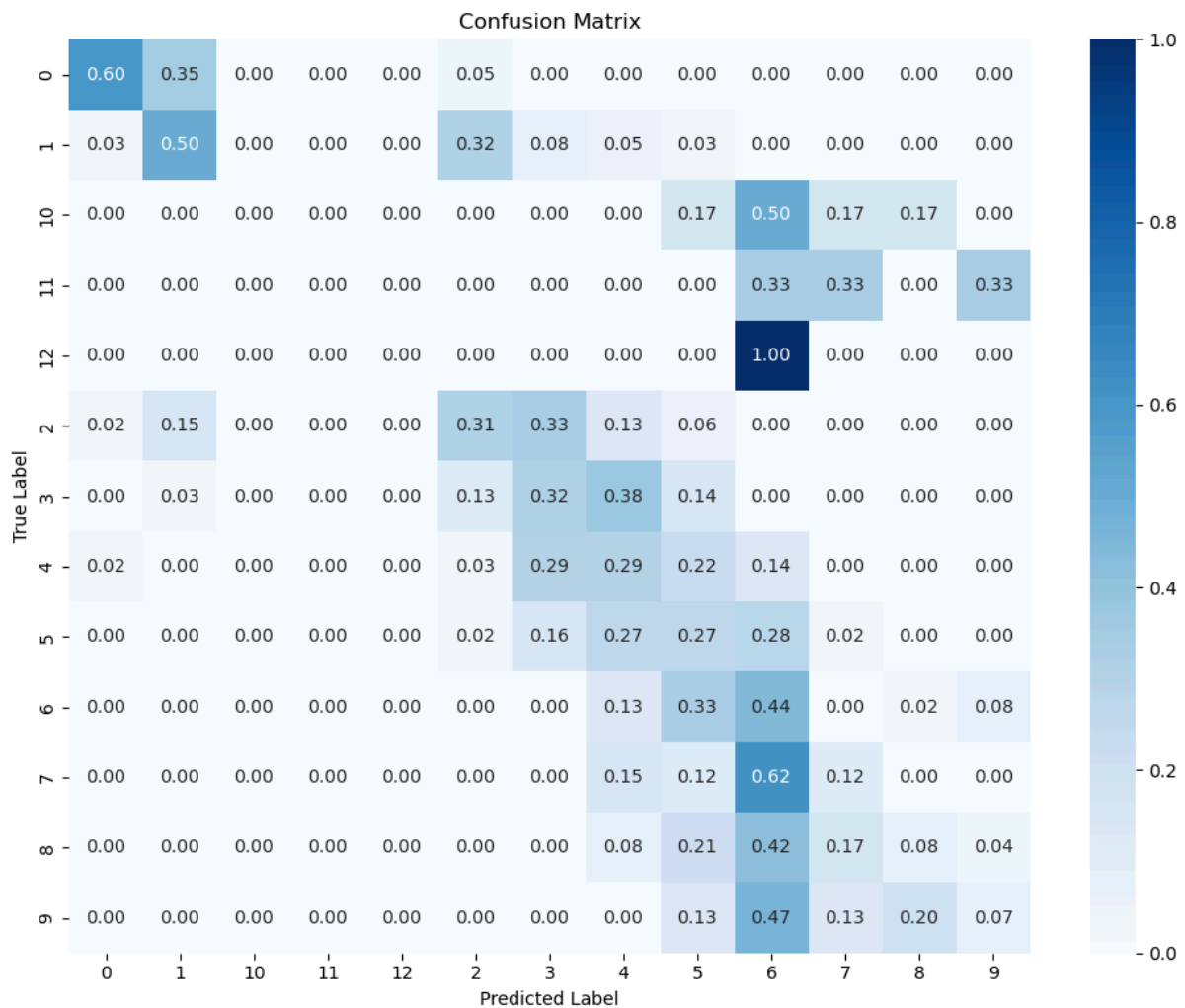
#### ❖ Precision-Recall Curve

- Muestra el balance entre precisión y exhaustividad
- AUC de 0.75 indica buen rendimiento
- La curva muestra estabilidad

#### ❖ F1 Score vs Threshold

- Muestra cómo cambia el F1-Score según el umbral de decisión
- El mejor F1-Score es aproximadamente 0.757
- Mantiene estabilidad alrededor de 0.7 en un amplio rango de umbrales

### 5.1.3 Matriz de Confusión



Muestra cómo el modelo clasifica cada clase. Los números en la diagonal son los aciertos y los colores más oscuros indican mayor proporción de predicciones. Observaciones clave:

- ❖ Clase 0 tiene el mejor desempeño (0.60 de precisión)
- ❖ Clase 1 muestra buen rendimiento (0.50 de precisión)
- ❖ Hay confusión significativa entre clases cercanas
- ❖ Las clases 10, 11 y 12 muestran muy bajo rendimiento

### 5.1.4 Resultados en Conjunto de Pruebas

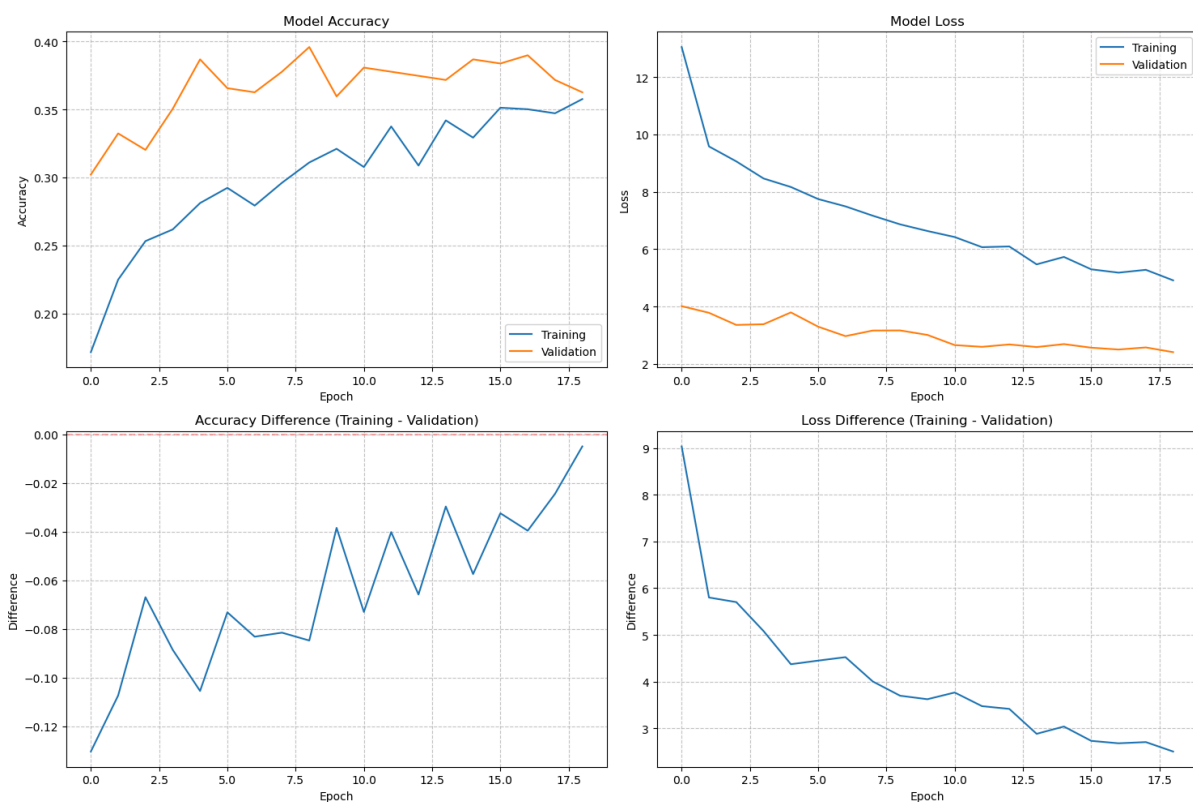
En el conjunto de pruebas, el modelo obtuvo una precisión de 30.56% y un f1-score de 0.22. Estos resultados sugieren que, aunque el modelo puede identificar algunas clases con relativa precisión, su rendimiento general no es adecuado para una aplicación práctica de clasificación de ganado.

### 5.1.5 Resumen

El modelo tiene un rendimiento moderado, con problemas de desbalance de clases y confusión entre categorías similares. El ROC AUC es muy bueno (0.977) pero la precisión general (30.56%) sugiere problemas significativos con la clasificación multiclase. La matriz de confusión muestra que algunas clases se predicen bien (0 con 60%, 1 con 50%) mientras otras, especialmente las clases 10, 11 y 12, tienen un rendimiento muy pobre.

## 5.2 Rendimiento Nocturno

### 5.2.1 Métricas de Entrenamiento



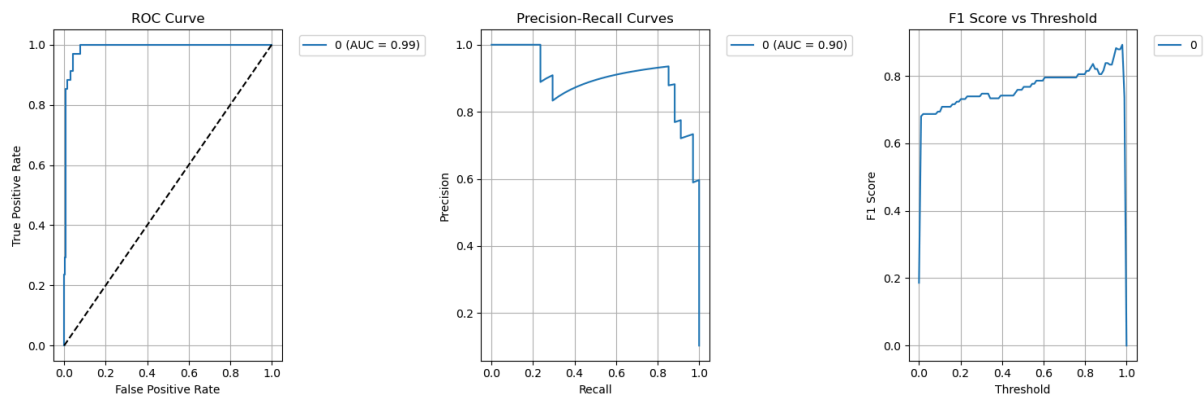
#### ❖ Precisión del modelo

- La precisión en entrenamiento alcanza un aproximado de 35%
- La precisión en validación se estabiliza en 37%
- La curva de validación muestra fluctuaciones pero se mantiene por encima del entrenamiento

#### ❖ Pérdida del modelo

- Comienza en 12 y disminuye consistentemente
- La pérdida de entrenamiento se estabiliza en 5
- La pérdida de validación se estabiliza en 2.5

## 5.2.2 Métricas de Evaluación



### ❖ ROC Curve

- AUC excepcional de 0.99
- Curva casi perfecta, indicando excelente capacidad discriminativa
- Muestra un rendimiento robusto en diferentes umbrales de clasificación

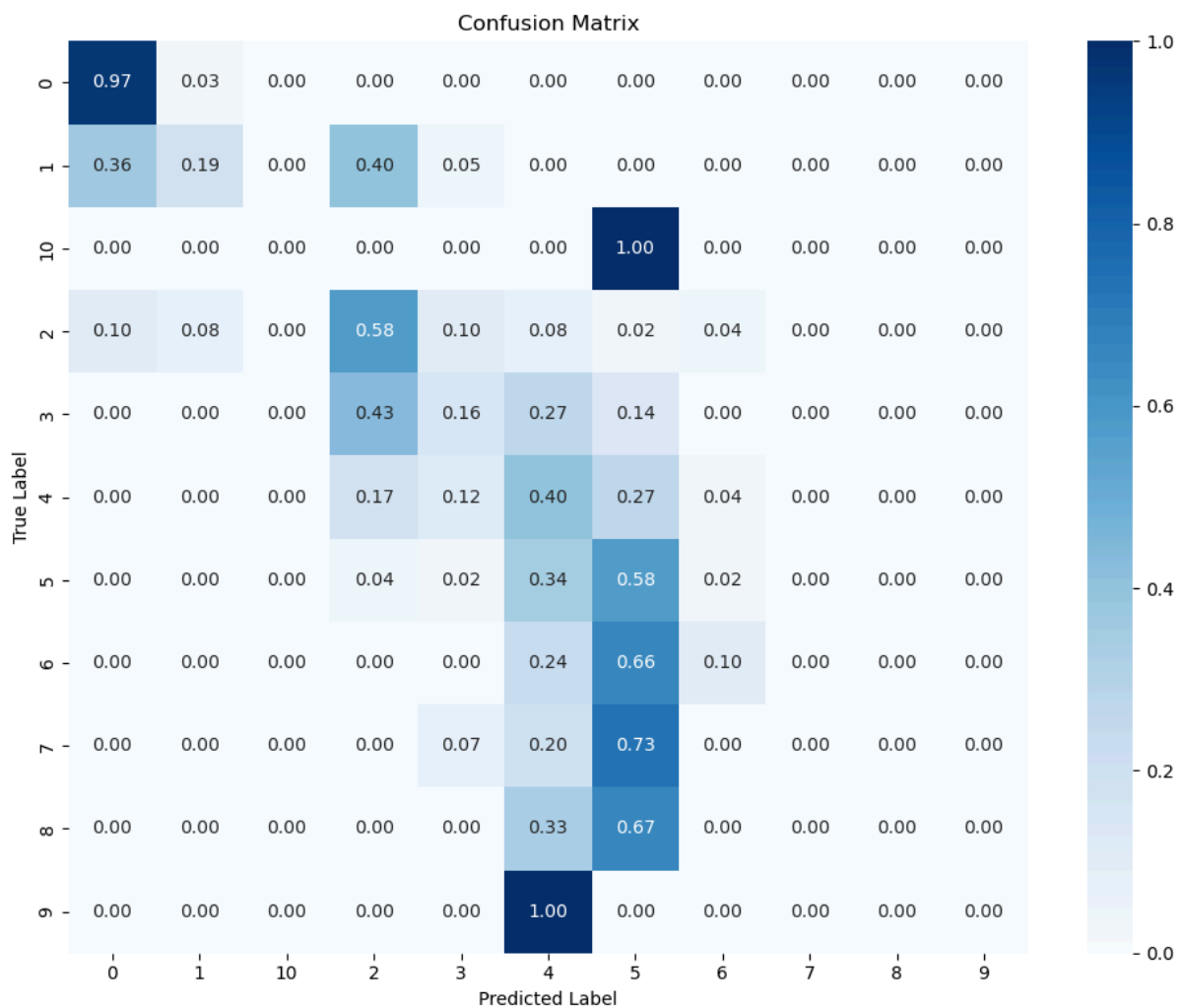
### ❖ Precision-Recall Curve

- AUC de 0.90, indicando muy buen rendimiento
- Mantiene alta precisión hasta un recall de 0.8
- Caída pronunciada después de 0.8 de recall

### ❖ F1 Score vs Threshold

- F1-Score máximo de 0.892
- Rendimiento estable entre thresholds de 0.2 a 0.8
- Mejores resultados en thresholds altos (0.8-0.9)

### 5.2.3 Matriz de Confusión



La matriz de confusión muestra patrones claros de predicción:

- ❖ Clase 0: Excelente precisión (97%)
- ❖ Clase 10: 100% de precisión en sus predicciones
- ❖ Clase 9: 100% de precisión en la clasificación
- ❖ Desafíos principales:
  - Clase 1: Solo 19% de precisión
  - Clases 3-5: Precisión moderada (40-58%)
  - Confusión significativa entre clases adyacentes

### 5.2.4 Resultados en Conjunto de Pruebas

En el conjunto de pruebas, el modelo obtuvo una precisión de 39.57% y un f1-score de 0.22. Estos resultados sugieren que, aunque el modelo tiene buena capacidad discriminativa

para ciertas clases específicas, su rendimiento general aún necesita mejoras para una aplicación práctica robusta.

### 5.2.5 Resumen

El modelo nocturno muestra un rendimiento mixto, con métricas de evaluación contrastantes. Por un lado, exhibe excelentes métricas de discriminación binaria (ROC AUC 0.99, PR AUC 0.90) y F1-Score máximo de 0.892. La matriz de confusión revela un rendimiento excepcional en algunas clases específicas (97% para clase 0, 100% para clases 9 y 10), pero muestra debilidades significativas en otras. Con una precisión general de 30.56%, el modelo demuestra capacidad para clasificar correctamente ciertas categorías, pero necesita mejoras para lograr un rendimiento más consistente a través de todas las clases.

## 5.3 Limitaciones y Consideraciones

Categoría	Limitación	Justificación
Rendimiento general	Bajo rendimiento y sobreajuste	Precisión diurna de 30.04% y nocturna de 36.22%, con evidencia de sobreajuste en ambos casos
Entrenamiento	Inestabilidad en métricas	Fluctuaciones significativas en validación, especialmente en modo nocturno
Generalización	Pobre capacidad de adaptación	Brecha creciente entre precisión de entrenamiento y validación, hasta 6% en época 40
Consistencia	Variabilidad entre condiciones	Diferencia notable entre rendimiento diurno (30.04%) y nocturno (36.22%)
Estabilidad	Fluctuaciones en validación	Picos significativos y volatilidad en métricas de validación, especialmente en modo nocturno
Complejidad	Modelo potencialmente sobredimensionado	El comportamiento sugiere que el modelo es demasiado complejo para el

		problema actual
--	--	-----------------

## 6.0 Evaluando los resultados

Los modelos especializados por condiciones de iluminación alcanzaron la precisión de:

- ❖ Modelo diurno: 30.56%
- ❖ Modelo nocturno: 39.57%

Ambos modelos presentan problemas de sobreajuste moderado y bajo rendimiento en la generalización de las características del conjunto de datos.

### 6.1 Validación de Objetivos

#### NO LOGRAMOS EL OBJETIVO

##### Objetivo de Minería de Datos

- ❖ Determinar la cantidad de vacas en cada imagen en cualquier condición.
  - Un modelo para condiciones diurnas con un 80% de precisión.
  - Un modelo para condiciones nocturnas con un 50% de precisión.

Los resultados de los modelos, tanto diurno como nocturno, quedaron por debajo de nuestras expectativas iniciales, no alcanzando el umbral objetivo del 50% de precisión. El modelo diurno alcanzó un 30.56% de precisión, mientras que el nocturno logró un 39.57%. A pesar de que estos resultados están por debajo del umbral esperado, proporcionaron información valiosa sobre los desafíos específicos en la clasificación de ganado bajo diferentes condiciones de iluminación.

### 6.2 Siguiendo pasos

Estos resultados, aunque no alcanzaron el umbral deseado, proporcionan una base importante para futuras mejoras. Las lecciones aprendidas sugieren varias áreas clave de mejora:

- ❖ Necesidad de técnicas más robustas para el manejo de clases minoritarias
- ❖ Importancia de reducir la confusión entre clases adyacentes
- ❖ Potencial para mejorar la preparación y aumentación de datos
- ❖ Oportunidad para explorar arquitecturas de modelo más especializadas



La siguiente fase del proyecto se centrará en abordar estas áreas de mejora, implementando técnicas más avanzadas de balanceo de datos, explorando arquitecturas de modelo alternativas y mejorando las estrategias de preprocesamiento de imágenes. El objetivo será construir sobre los aspectos positivos identificados en esta primera iteración para desarrollar modelos más robustos y precisos en futuras implementaciones.