



# **DATA SCIENCE PROJECT**

# CONTENT

- 01** CONJUNTO DE DATOS
- 02** PROCESAMIENTO DE DATOS
- 03** TARGET Y MISSING VALUES
- 04** MODELOS
- 05** RESULTADOS
- 06** TECNICAS
- 07** RESULTADO FINAL

# DATOS:

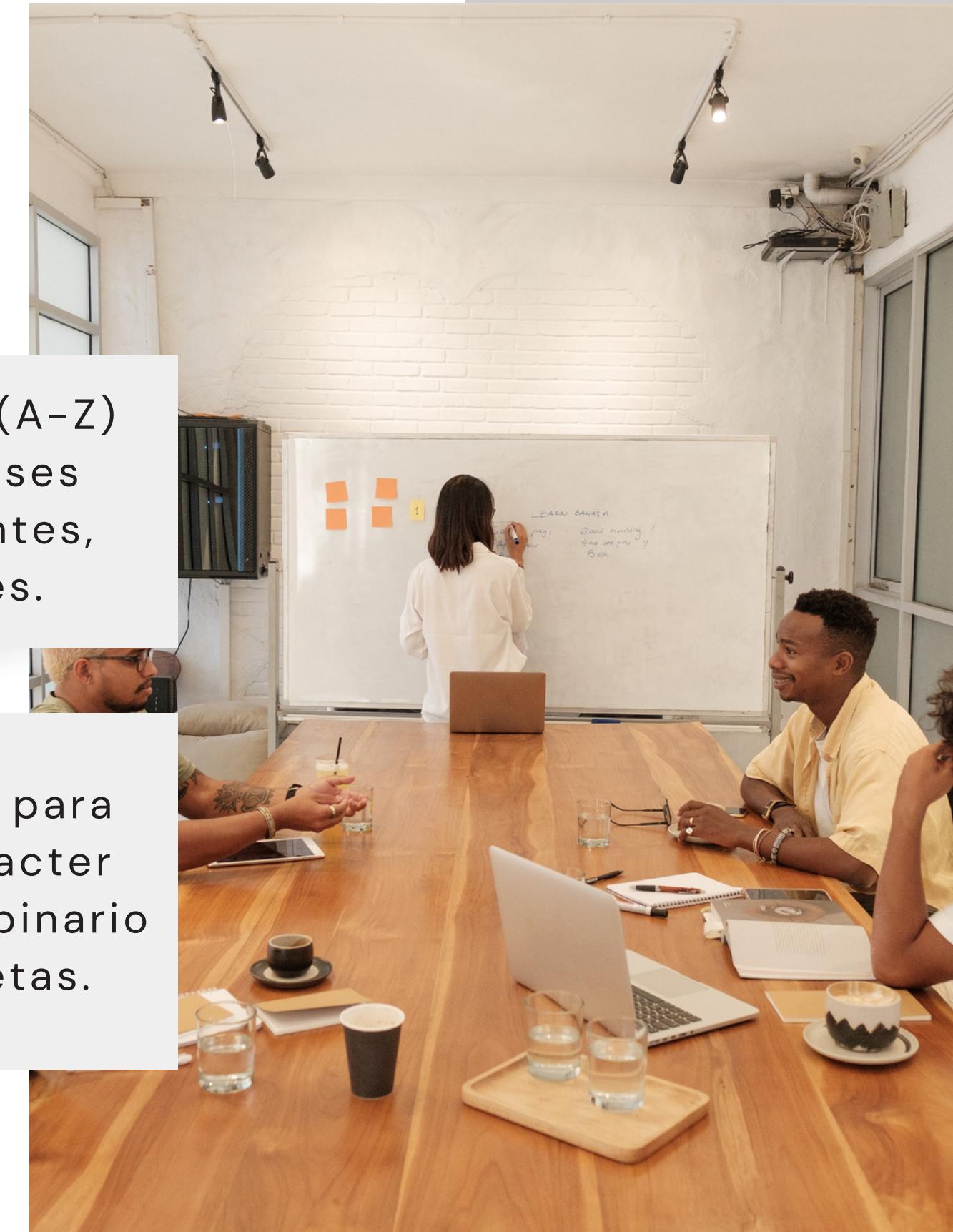
Conjunto de datos de fuentes de caracteres alfabéticos.



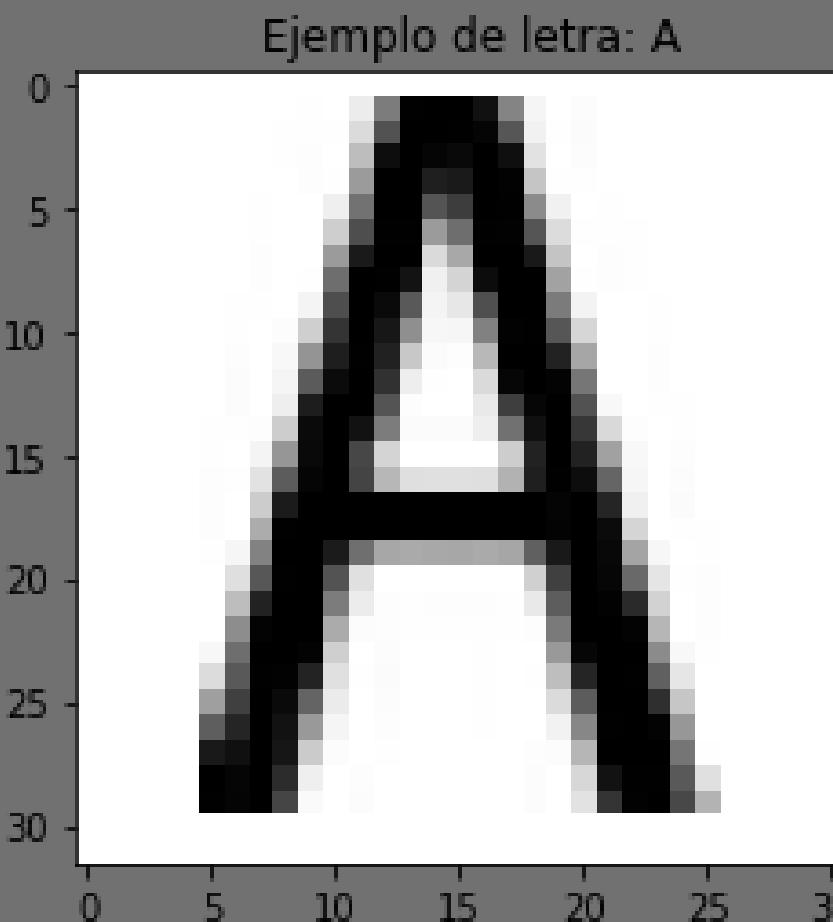
El conjunto de datos contiene 26 carpetas (A-Z) de caracteres alfabéticos en escala de grises renderizados utilizando más de 14900 fuentes, cada una con un tamaño de 32×32 píxeles.



También se proporciona un archivo binario numpy para facilitar la carga de datos. El kernel "Loading Character Dataset" contiene una demo para cargar el archivo binario numpy en matrices de píxeles de imagen y etiquetas.

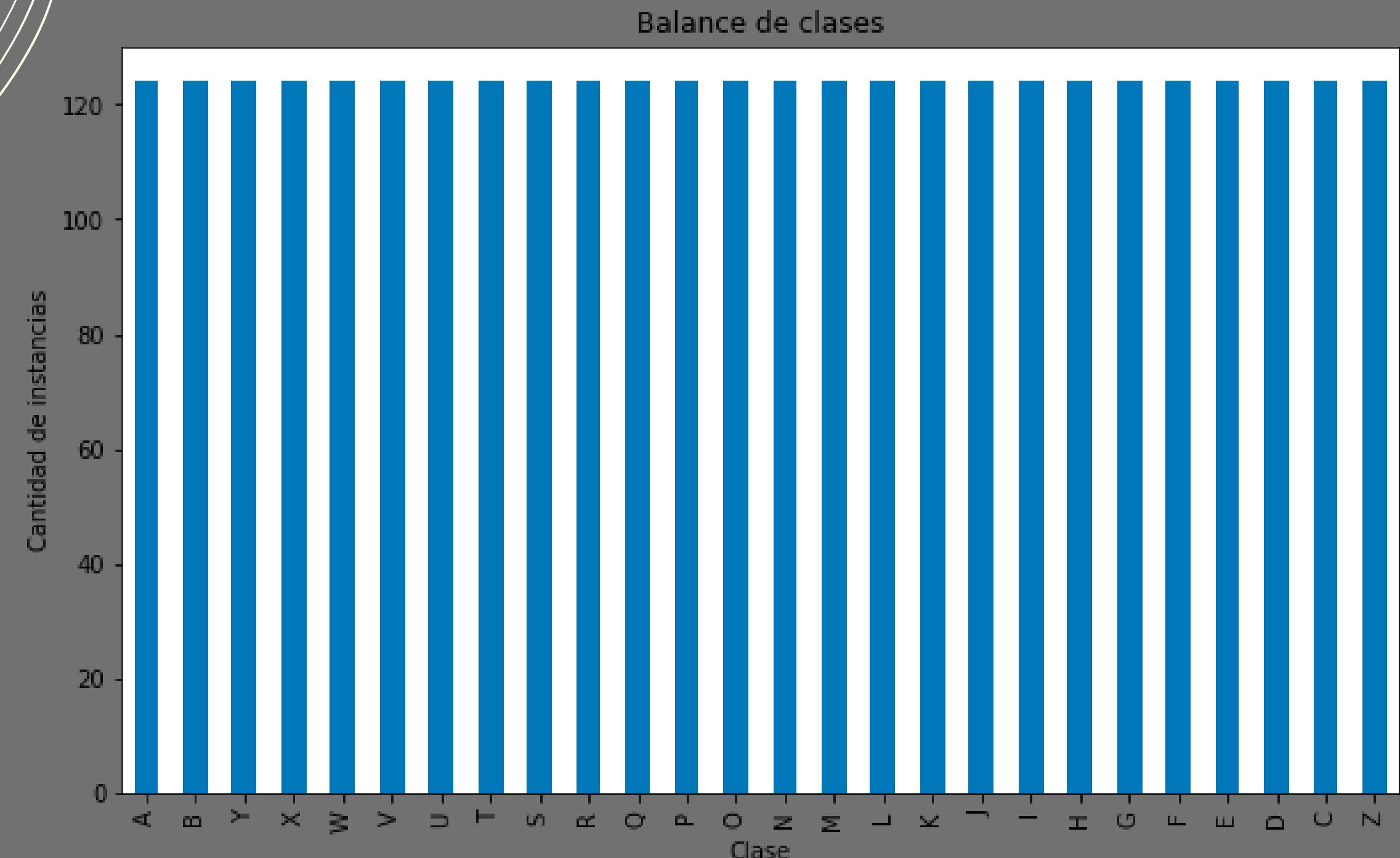


# PROCESAMIENTO DE DATOS



DIMENSIONES DE X\_RESHAPE: (3224, 1024)

# TARGET



**MISSING VALUES:** Series([], dtype: int64)



# MODELOS

# REGRESIÓN LOGÍSTICA

## Métricas Regresión Logística

Precisión: 0.85

Recall: 0.83

F1-score: 0.84

Accuracy: 0.83

# DECISION TREE

## Métricas Decision Tree

Precisión: 0.61

Recall: 0.60

F1-score: 0.60

Accuracy: 0.60

# SVC

Métricas SVC:

Precisión: 0.86

Recall: 0.84

F1-score: 0.85

Accuracy: 0.84

# KNN

Métricas KNN

Precisión: 0.78

Recall: 0.74

F1-score: 0.75

Accuracy: 0.74

# ESAMBLE

## Métricas de clasificación

Precisión: 0.80  
Recall: 0.78  
F1-score: 0.79  
Accuracy: 0.78

## Métricas Random Forest

Precisión: 0.83  
Recall: 0.81  
F1-score: 0.82  
Accuracy: 0.81

## Métricas XGboost

Precisión: 0.81  
Recall: 0.80  
F1-score: 0.80  
Accuracy: 0.80

## Métricas ADAboost

Precisión: 0.42  
Recall: 0.34  
F1-score: 0.34  
Accuracy: 0.34

# REDES CONVOLUCIONALES

## Métricas de Redes convolucionales

Precisión: 0.88

Recall: 0.87

F1-score: 0.87

Accuracy: 0.87

# DEEP LEARNING

## Métricas MLP

Precisión: 0.86

Recall: 0.84

F1-score: 0.85

Accuracy: 0.84

# RESULTADOS

Modelo	Precisión	Recall	F1-score	Accuracy
Regresión Logística	0.85	0.83	0.84	0.83
Decision Tree	0.61	0.60	0.60	0.60
SVC	0.86	0.84	0.85	0.84
KNN	0.78	0.74	0.75	0.74
Random Forest	0.83	0.81	0.82	0.81
Gradient Boosting	0.80	0.78	0.79	0.78
XGboost	0.81	0.80	0.80	0.80
ADABOOST	0.42	0.34	0.34	0.34
MLP	0.86	0.84	0.85	0.84
Redes convolucionales	0.88	0.87	0.87	0.87





# TECNICAS

# AUMENTANDO LA COMPLEJIDAD DEL MODELO

Métricas de Redes Convolucionales

Precisión: 0.94

Recall: 0.93

F1-score: 0.94

Accuracy: 0.93

# DROP OUT

Métricas de Redes Convolucionales con Dropout

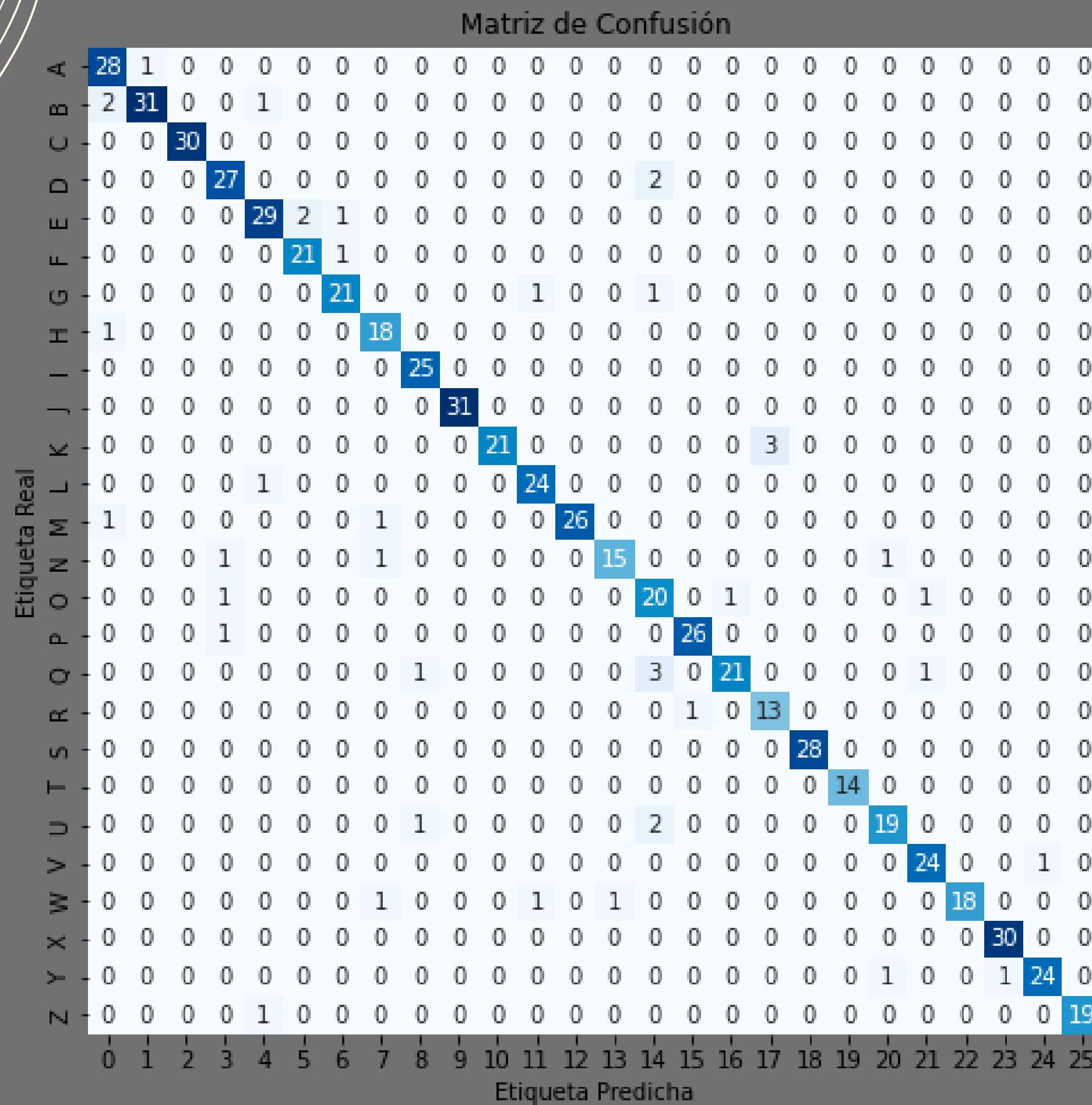
Precisión: 0.89

Recall: 0.88

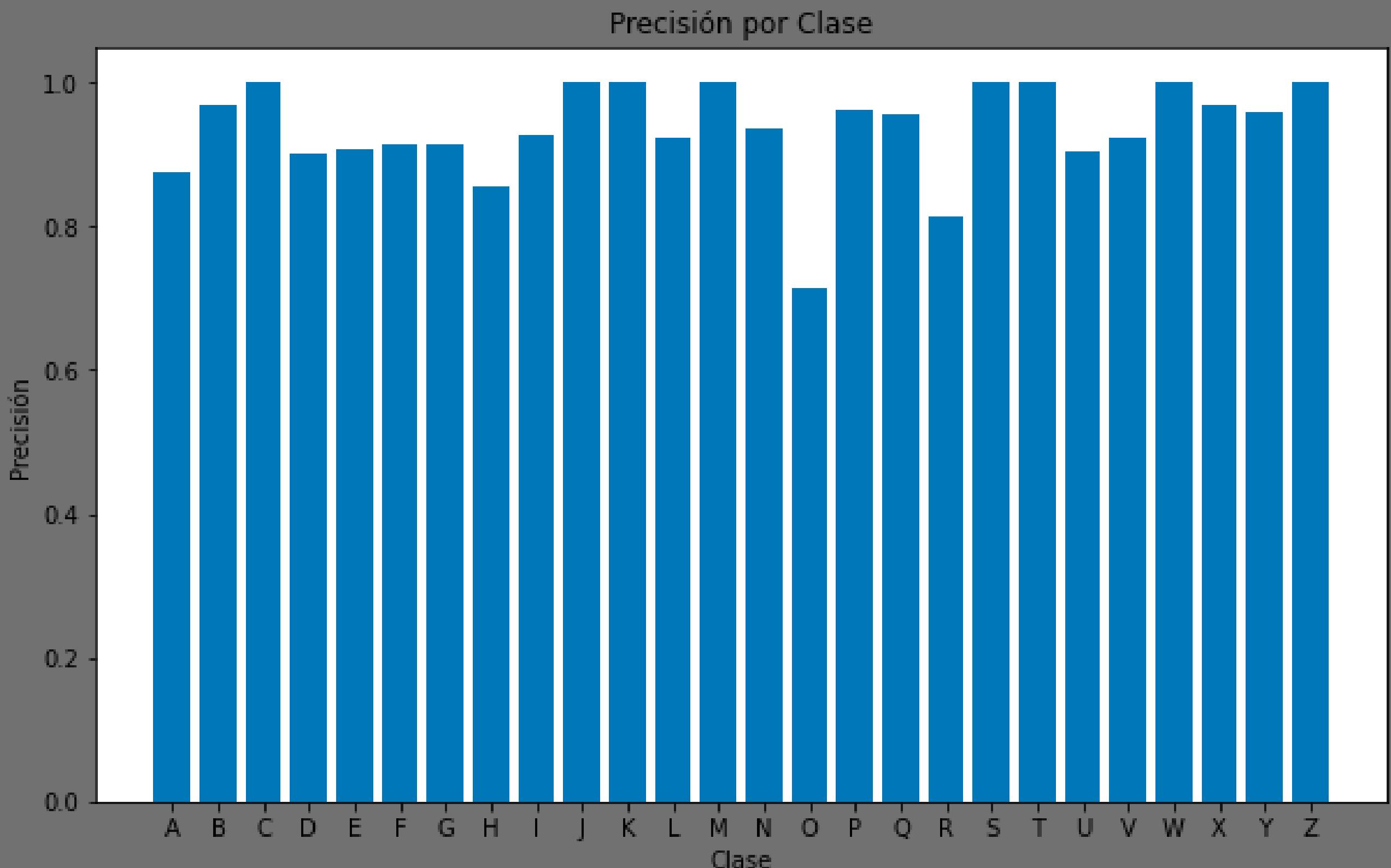
F1-score: 0.88

Accuracy: 0.88

# MATRIZ DE CONFUSIÓN



# PRESICIÓN POR CLASE



# PREPROCESADO DE IMÁGENES



IMAGEN ORIGINAL

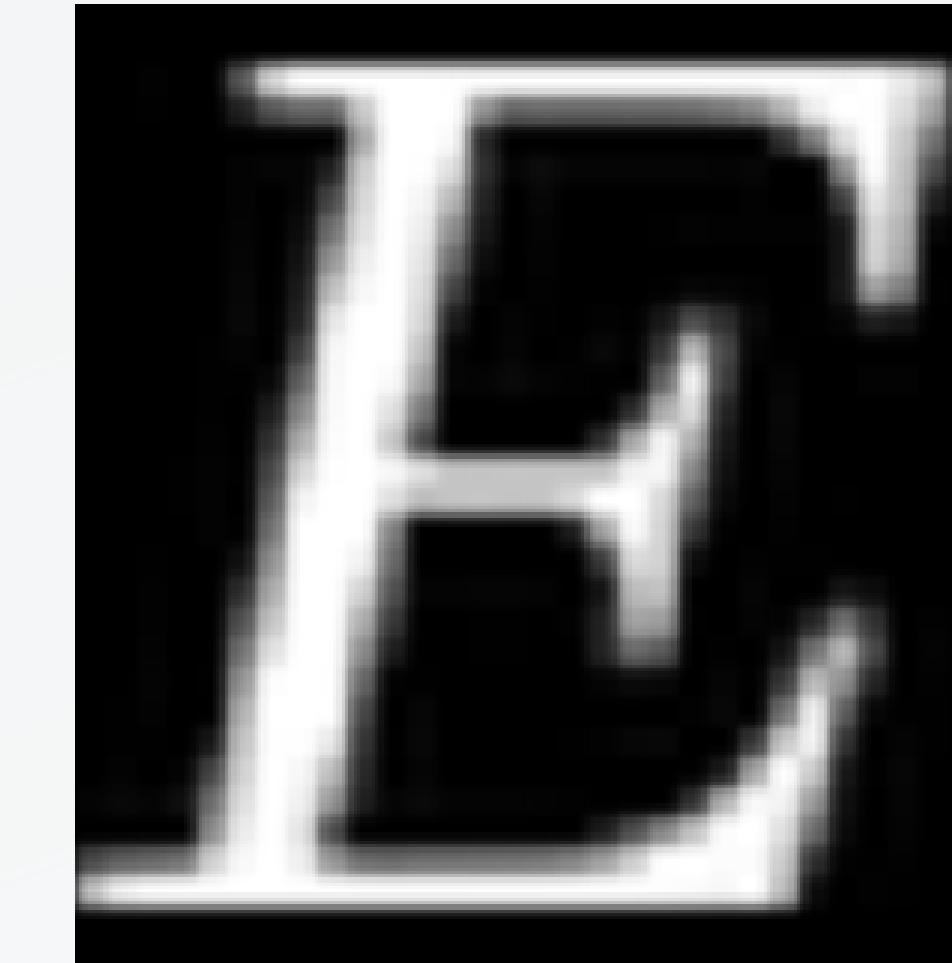
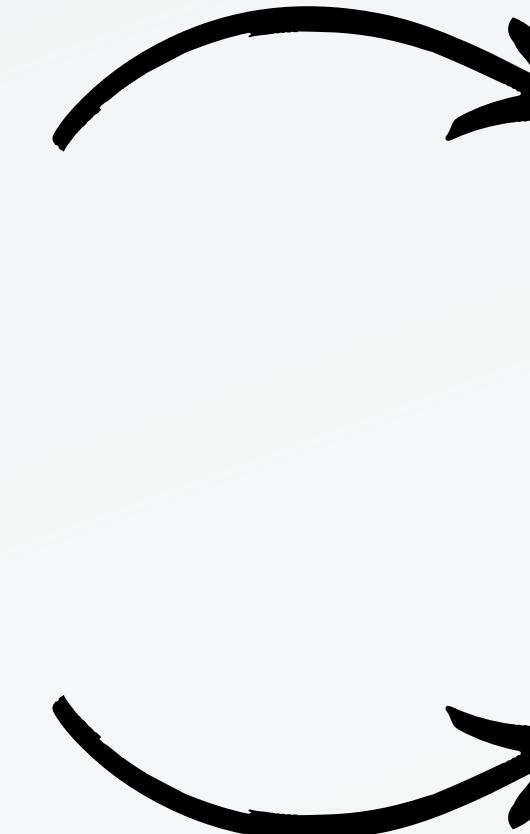


IMAGEN ENTRENAMIENTO

## PRERREQUISITOS PARA EL CORRECTO PROCESADO DE IMÁGENES

- En la imagen solo puede aparecer una letra.
- El fondo debe de ser de un color homogéneo.
- Debe existir un claro contraste entre el fondo y la letra.
- Puede aparecer algo de ruido, pero nunca más grande que la letra.