

# DESEMPEÑO DE RNN EN FUNCIÓN DE LARGOS DE SECUENCIA

Integrantes: Benjamín Bautista  
Samuel Bugueño  
Grupo: 7b  
Tutor: Jorge Espejo

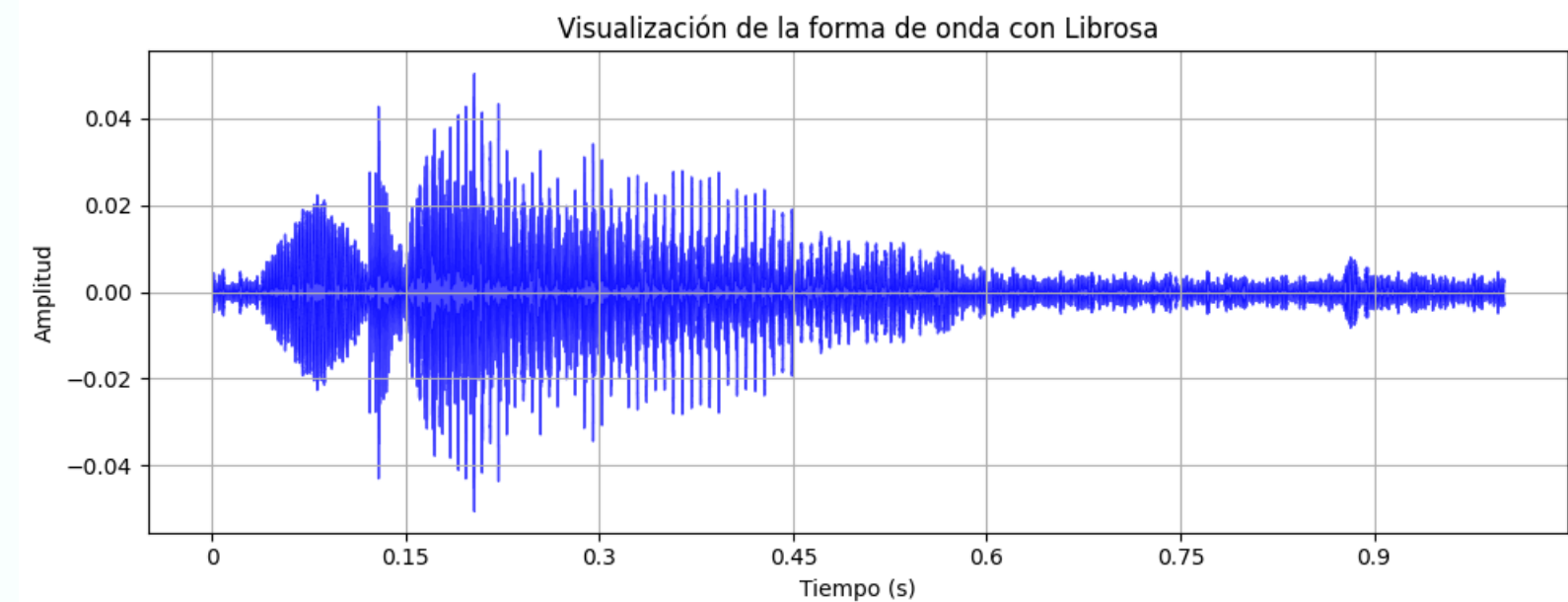
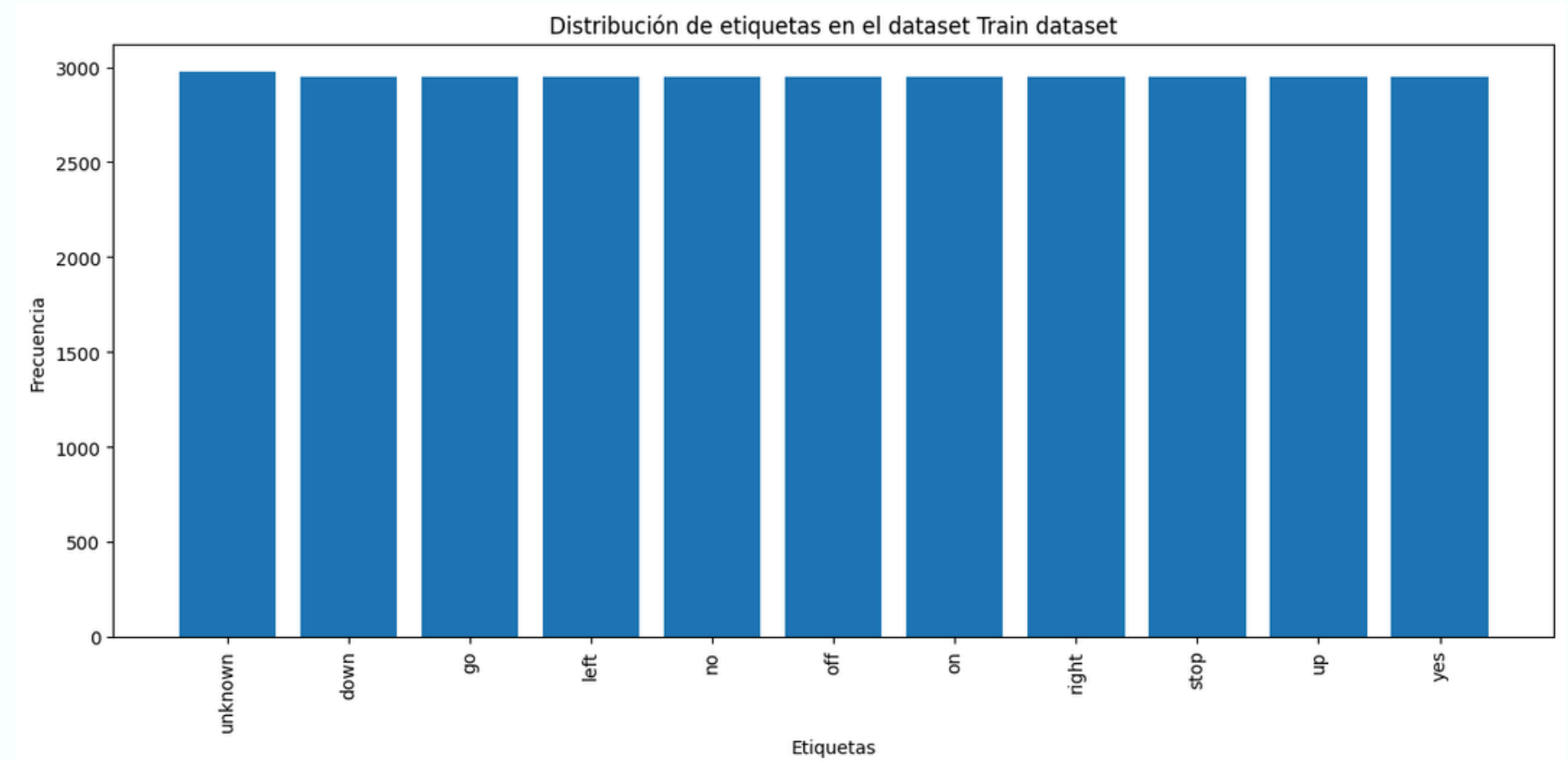
Sobre el Proyecto

# INTRODUCCIÓN

El reconocimiento de comandos de voz es una característica importante en distintos dispositivos, siendo uno de los más comunes los teléfonos. En este proyecto, se investigará cómo diferentes longitudes de secuencia afectan el desempeño de las redes recurrentes

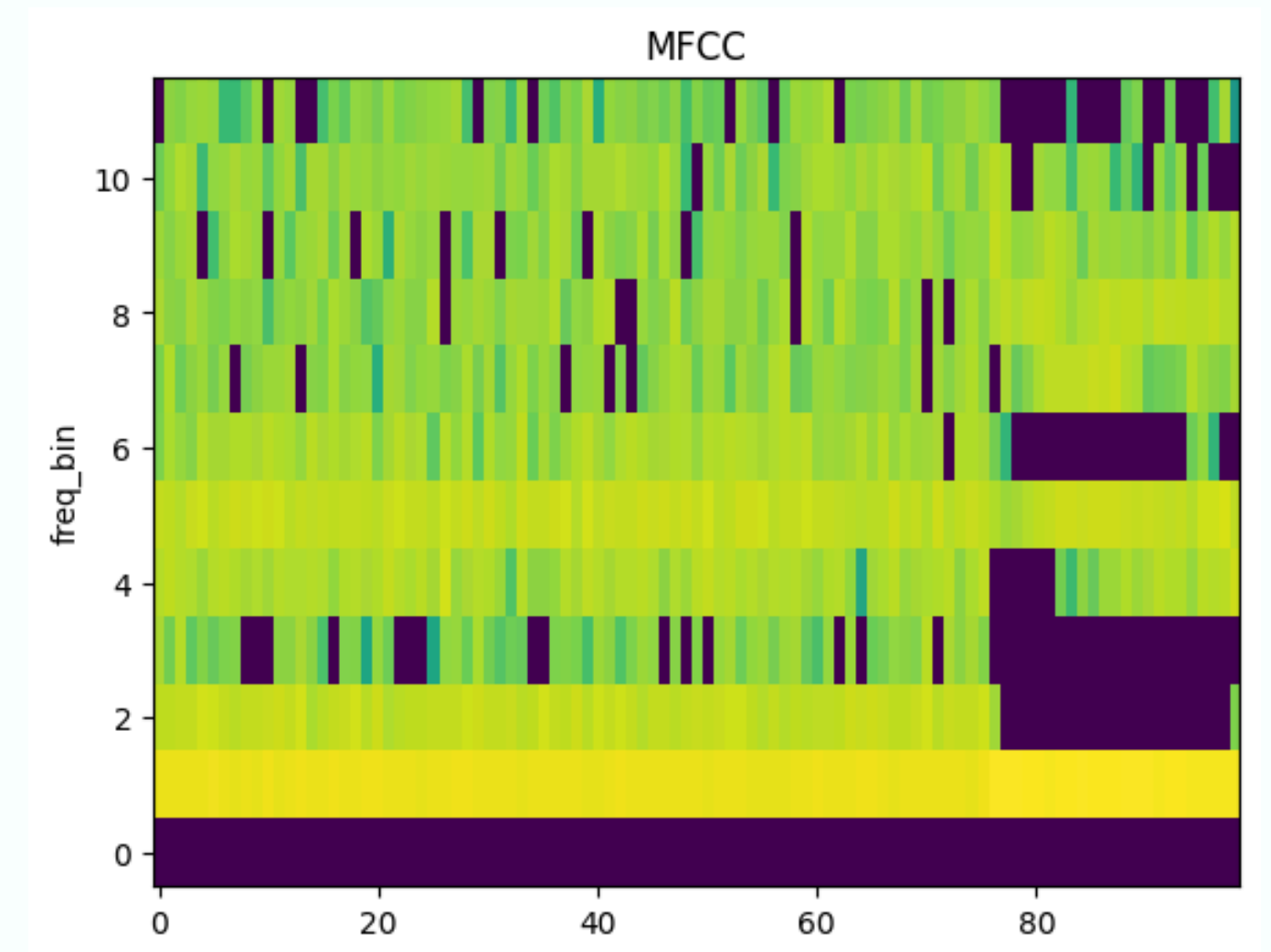
# DATASET

- Para entrenamiento y prueba se va a usar el dataset Speech Commands.
- Almacena grabaciones de 1 segundo con varias clases de palabras comunes en inglés.
- Existe poca variación entre las muestras usadas.



# MFCC

- Se procesa y condensa información del audio en una escala similar a cómo el oído humano procesa el sonido.
- Se obtienen coeficientes para una serie de ventanas de tiempo del audio.
- Se usan coeficientes de 0 al 12 al contener la mayor parte de la información relevante.



# MODELOS

## 01 RNN (Vanilla)

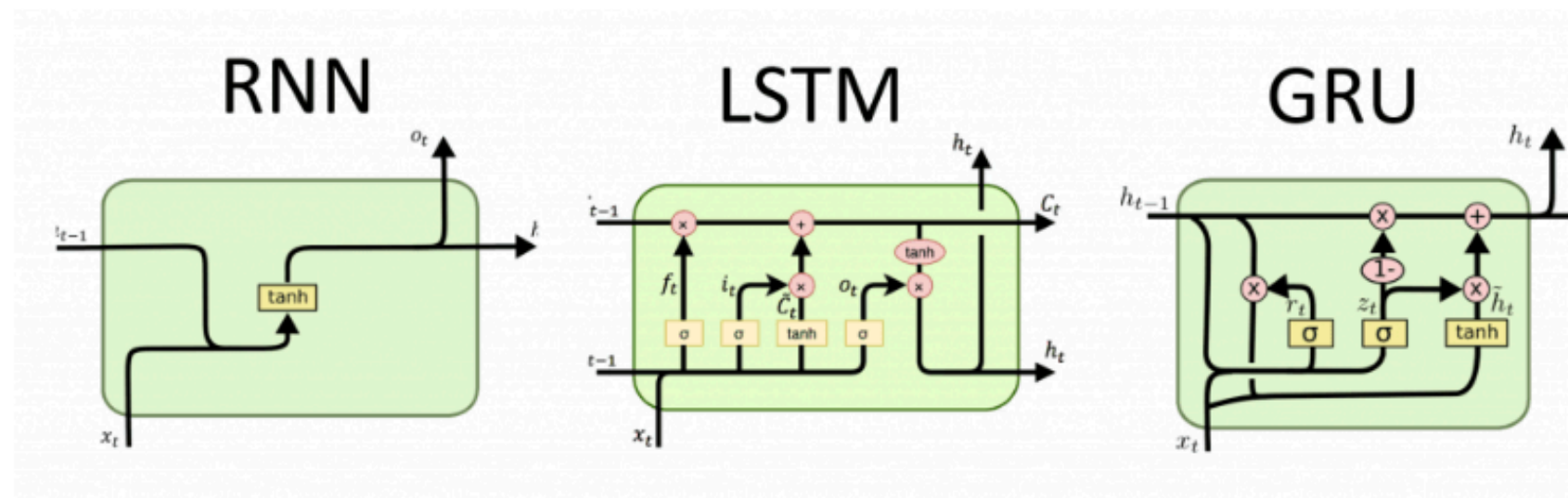
- Arquitectura más simple.
- Salida de un paso de tiempo es entrada para el siguiente.
- Rápida y poco costo computacional.
- Poca memoria y desvanecimiento.

## 02 LSTM

- Variante de RNN para secuencias más largas.
- Introduce 3 compuertas para controlar flujo de información
- Mayor precisión para secuencias largas, pero más costoso.

## 03 GRU

- Variante de LSTM.
- En lugar de 3 compuertas usa solo 2 compuertas.
- Menor costo computacional y desempeño comparable a LSTM.



# EXPERIMENTOS

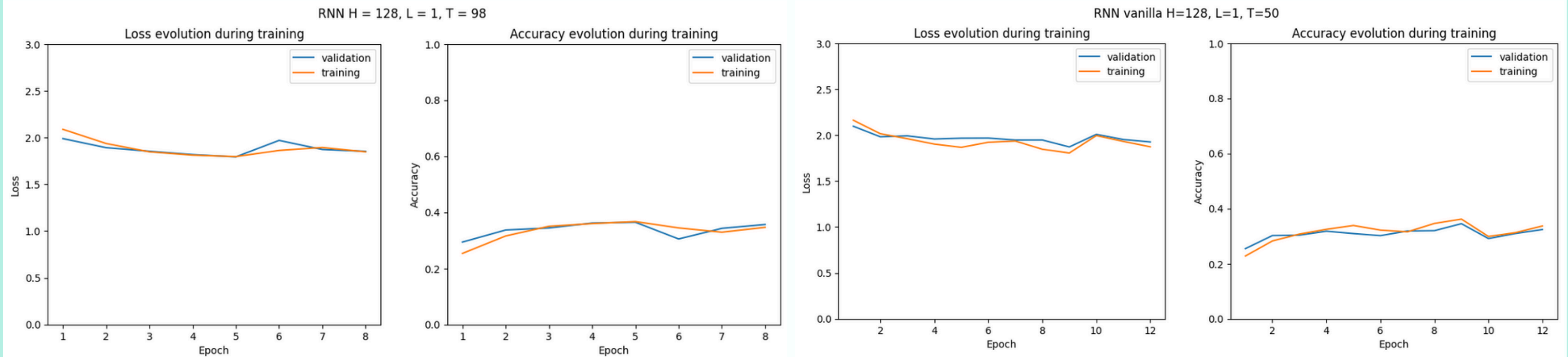
Modelo	Parámetro variado	Condición fija	Valores probados
RNN Vanilla	Neuronas por capa	1 capa oculta, secuencia = 98	64, 128, 256
RNN Vanilla	Capas ocultas	128 neuronas, secuencia = 98	1, 2, 3
RNN Vanilla	Largo de secuencia	1 capa oculta, 128 neuronas	50, 98, 120
GRU	Neuronas por capa	1 capa oculta, secuencia = 98	64, 128, 256
GRU	Capas ocultas	128 neuronas, secuencia = 98	1, 2
GRU	Largo de secuencia	1 capa oculta, 128 neuronas	50, 98 , 160
LSTM	Neuronas por capa	1 capa oculta, secuencia = 98	128, 256
LSTM	Largo de secuencia	1 capa oculta, 128 neuronas	50, 98, 150

Parámetros de entrenamiento:

- Épocas: 20
- Early stopping patience: 3.
- Función de costos: Cross Entropy
- Optimizador: Adam

# DESEMPEÑO

## RNN



Tiempo de entrenamiento : 22 segundos

Val loss: 1.8522, Val acc: 35.66%

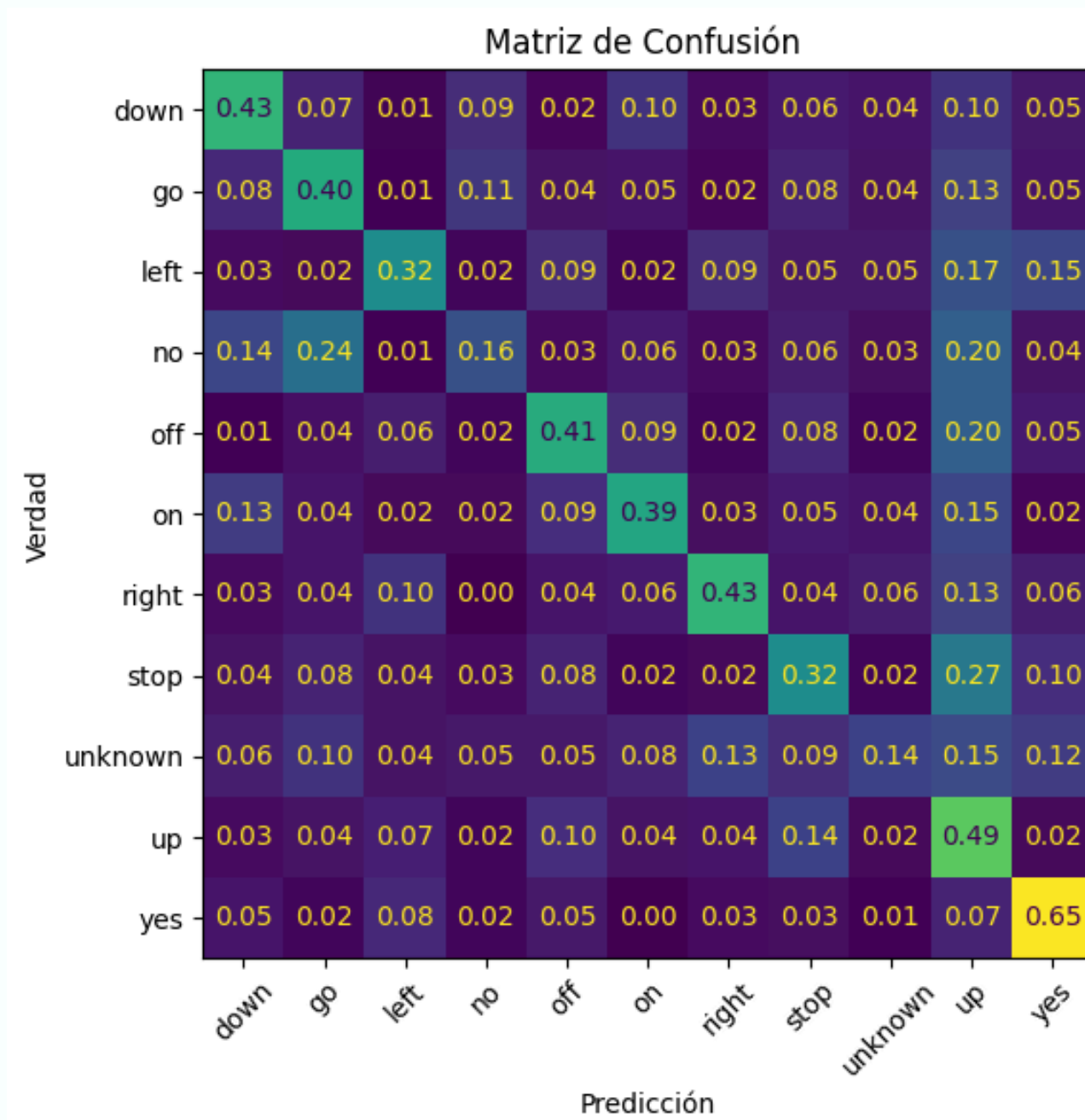
Epoch 8/20

Tiempo de entrenamiento : 31 segundos

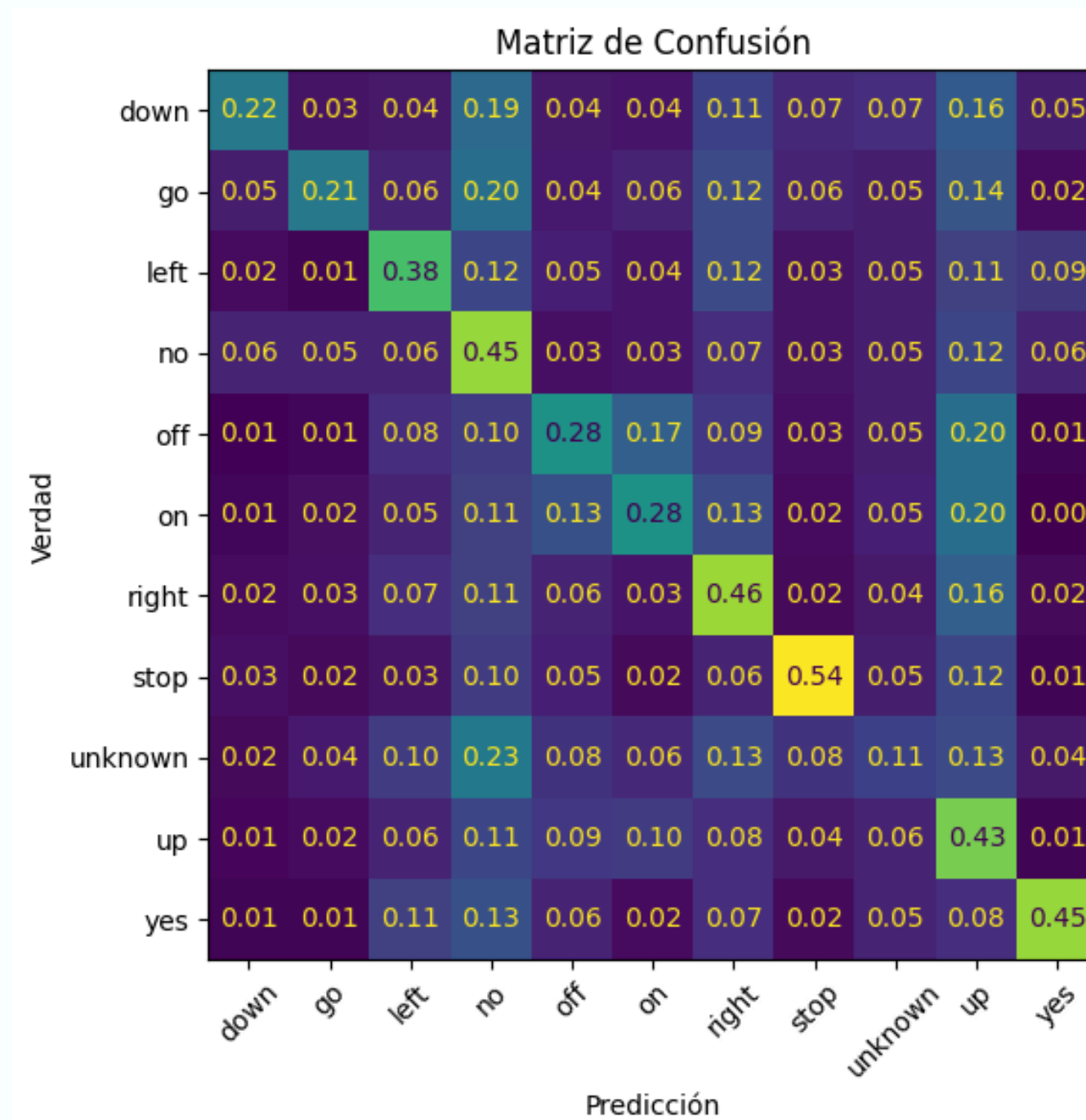
Val loss: 1.9254, Val acc: 32.41%

Epoch 12/20\$

# RNN



Largo de secuencia: 98



Largo de secuencia: 50



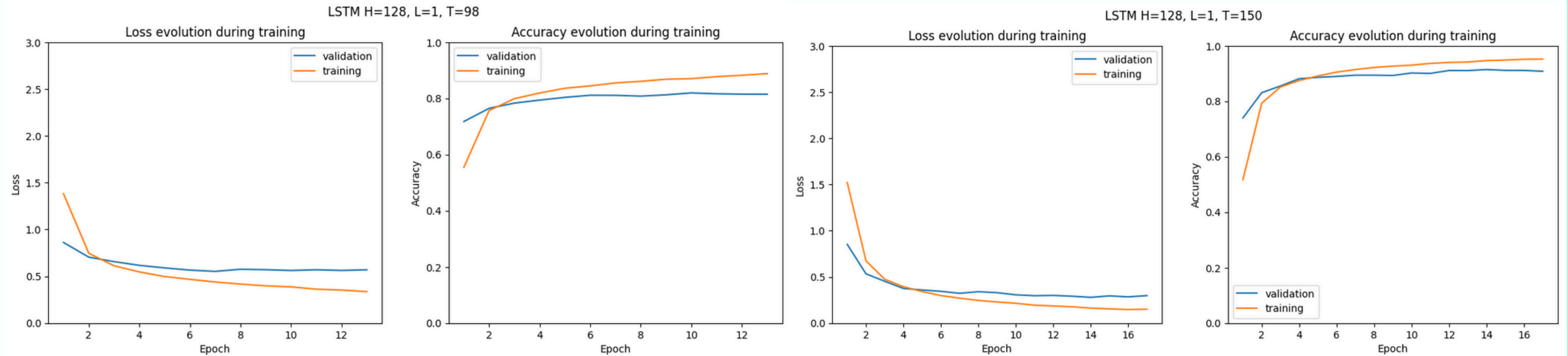
## RNN

Modelo	Capas	Neuronas	T	Accuracy	Precision	Recall	F1-score
RNN Vanilla	1	128	98	0.3759	0.38	0.38	0.37
RNN Vanilla	1	128	50	0.3444	0.38	0.35	0.34

Largo de secuencia: 98

Largo de secuencia: 50

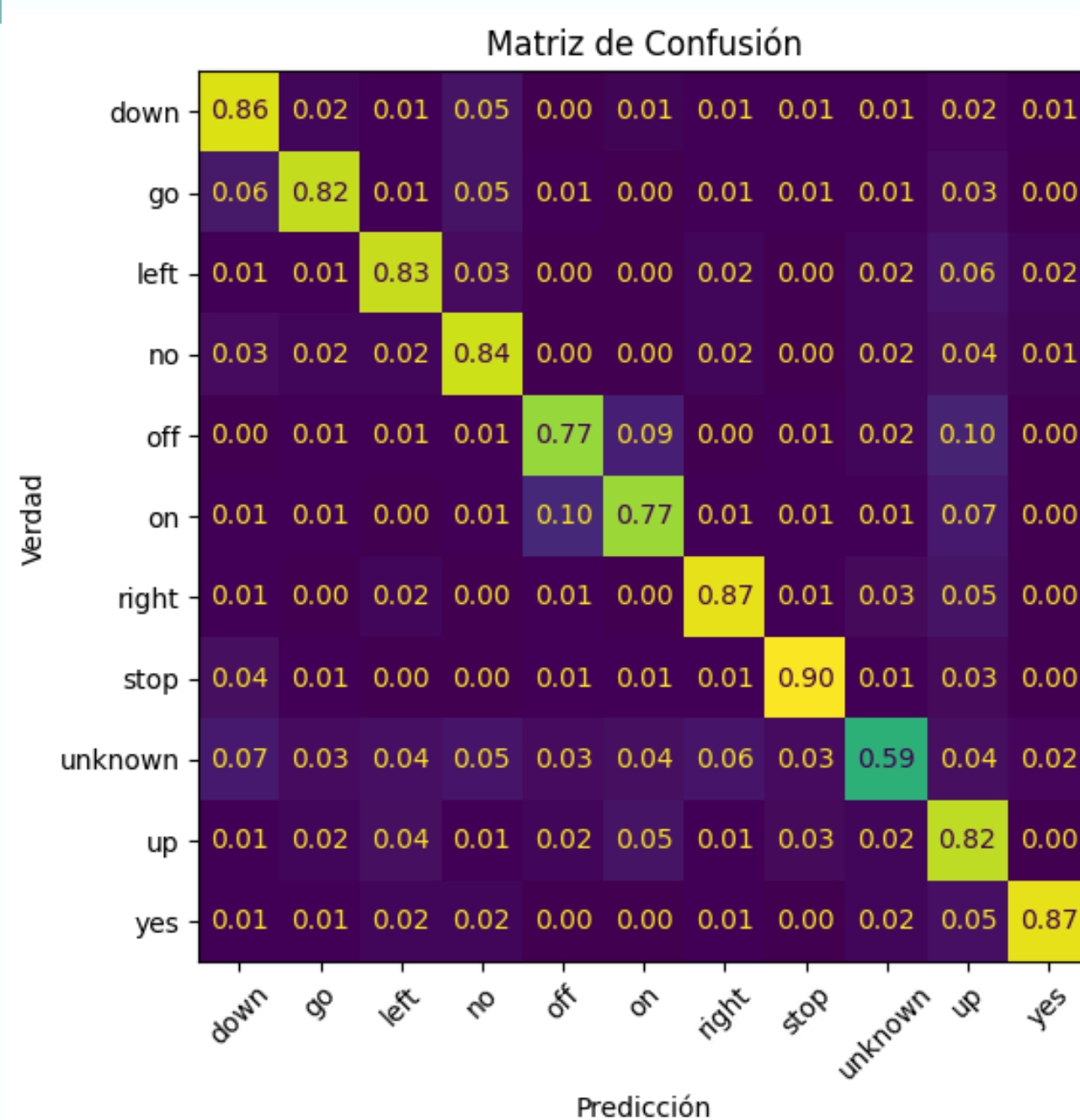
# LSTM



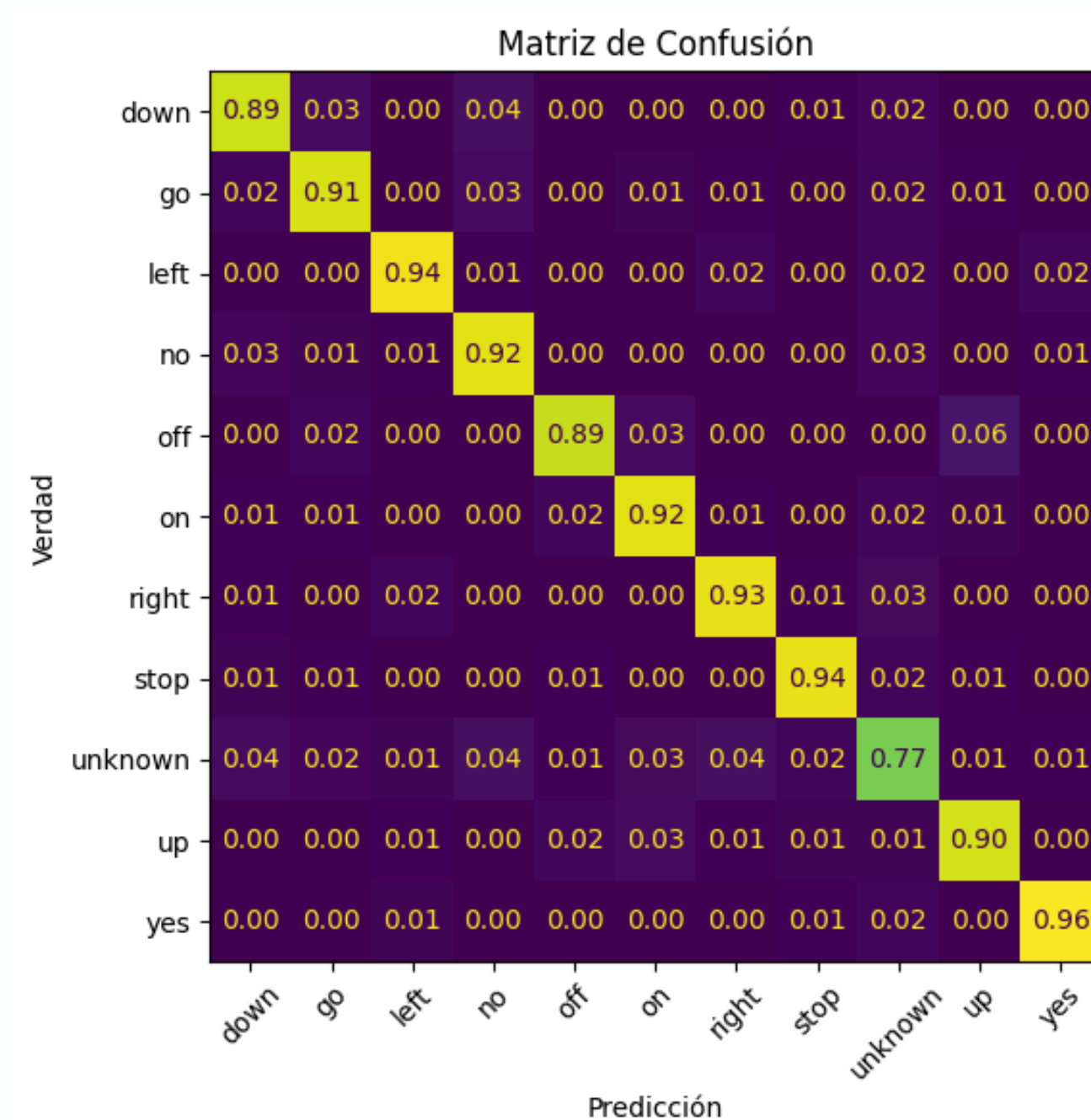
**Tiempo de entrenamiento : 39 segundos**  
**Val loss: 0.5693, Val acc: 81.47%**  
**Epoch 13/20**

**Tiempo de entrenamiento : 60 segundos**  
**Val loss: 0.2979, Val acc: 90.81%**  
**Epoch 17/20**

# LSTM



Largo de secuencia: 98



Largo de secuencia: 150

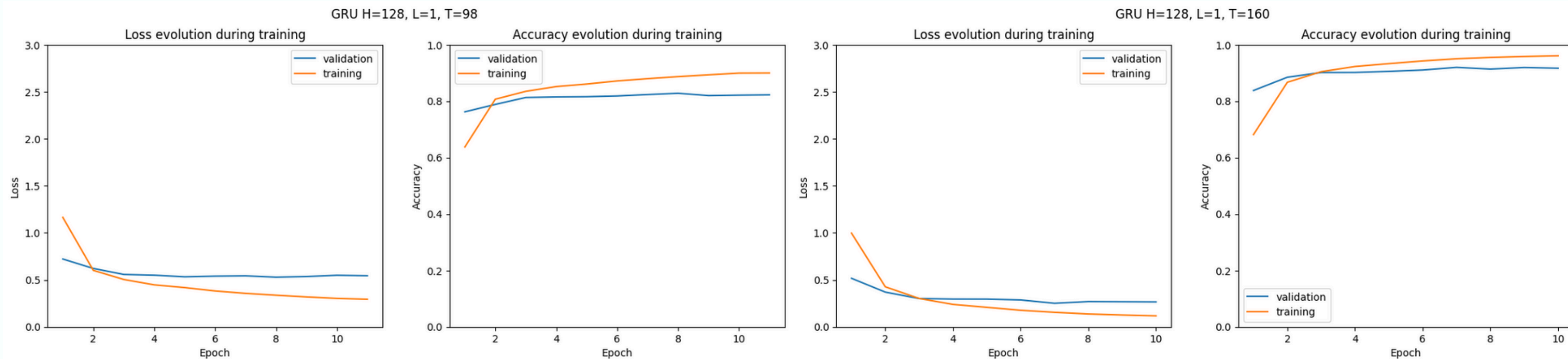
## LSTM

Modelo	Capas	Neuronas	T	Accuracy	Precision	Recall	F1-score
LSTM	1	128	98	0.8105	0.82	0.81	0.81
LSTM	1	128	150	0.9064	0.91	0.91	0.91

Largo de secuencia: 98

Largo de secuencia: 150

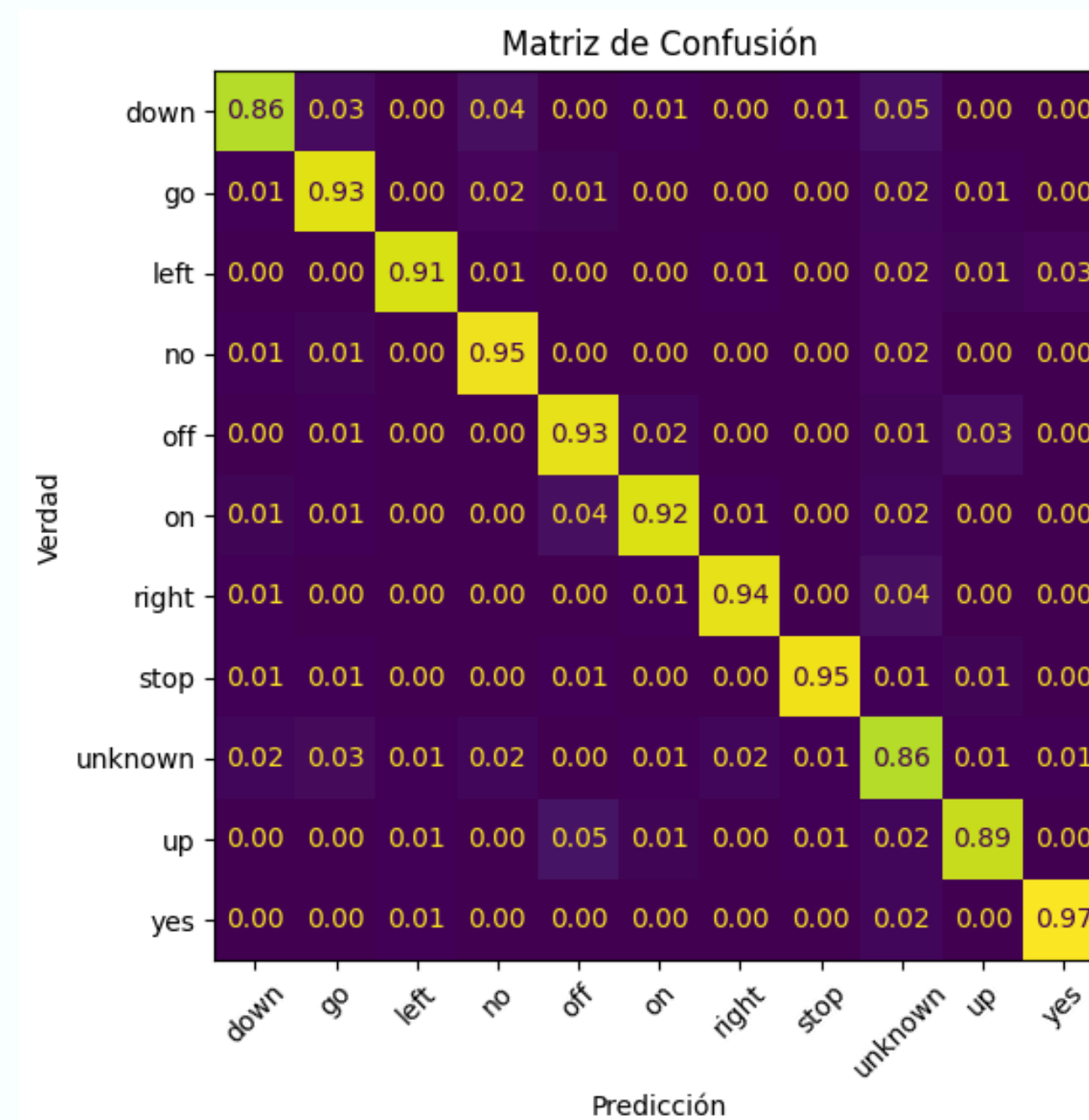
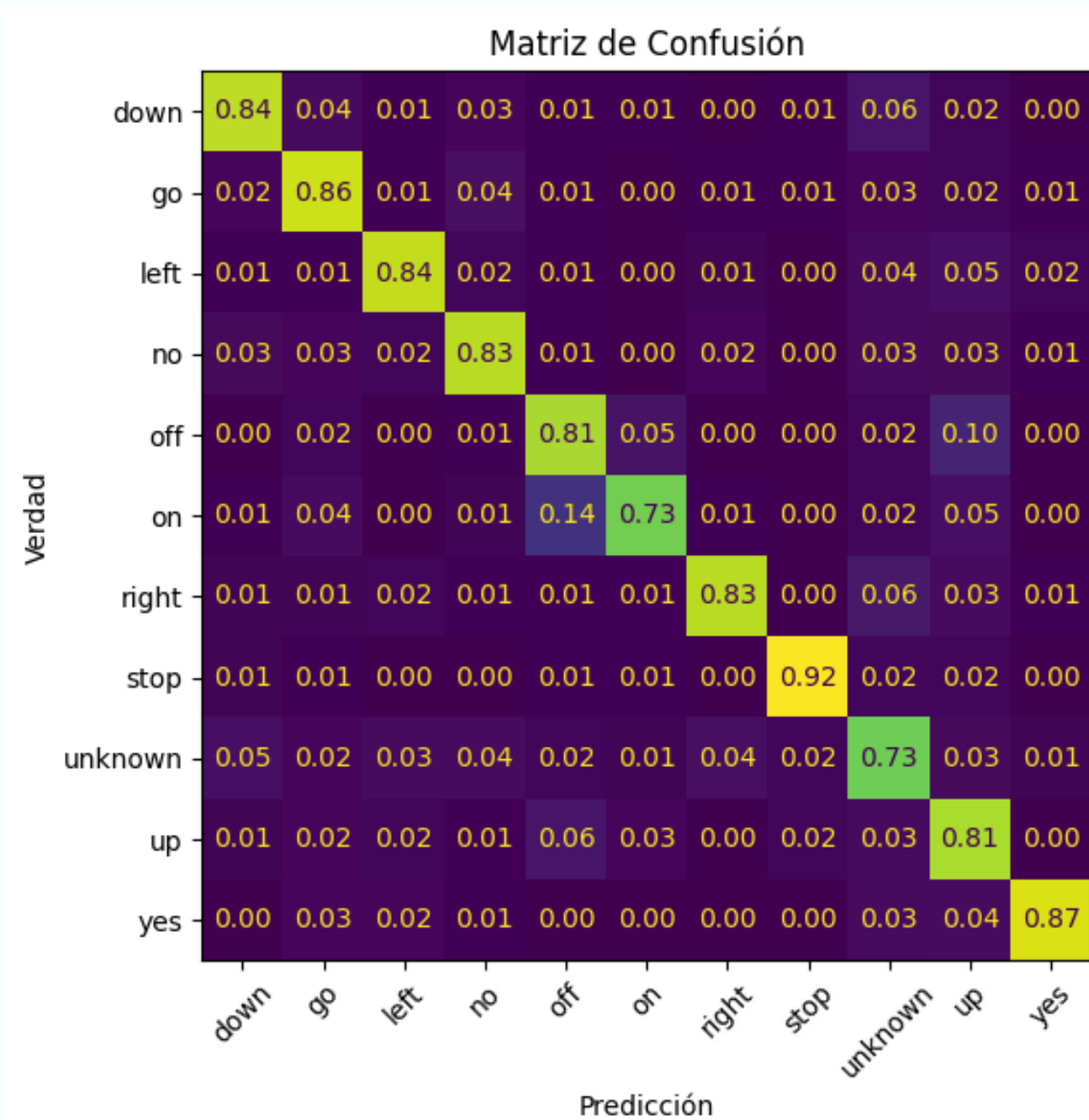
# GRU



**Tiempo de entrenamiento : 31 segundos**  
**Val loss: 0.5440, Val acc: 82.22%**  
**Epoch 11/20**

**Tiempo de entrenamiento : 32 segundos**  
**Val loss: 0.2298, Val acc: 93.32%**  
**Epoch 9/20**

# GRU



Largo de secuencia: 98

Largo de secuencia: 160

## GRU

Modelo	Capas	Neuronas	T	Accuracy	Precision	Recall	F1-score
GRU	1	128	98	0.8242	0.83	0.82	0.83
GRU	1	128	150	0.9187	0.92	0.92	0.92

Largo de secuencia: 98

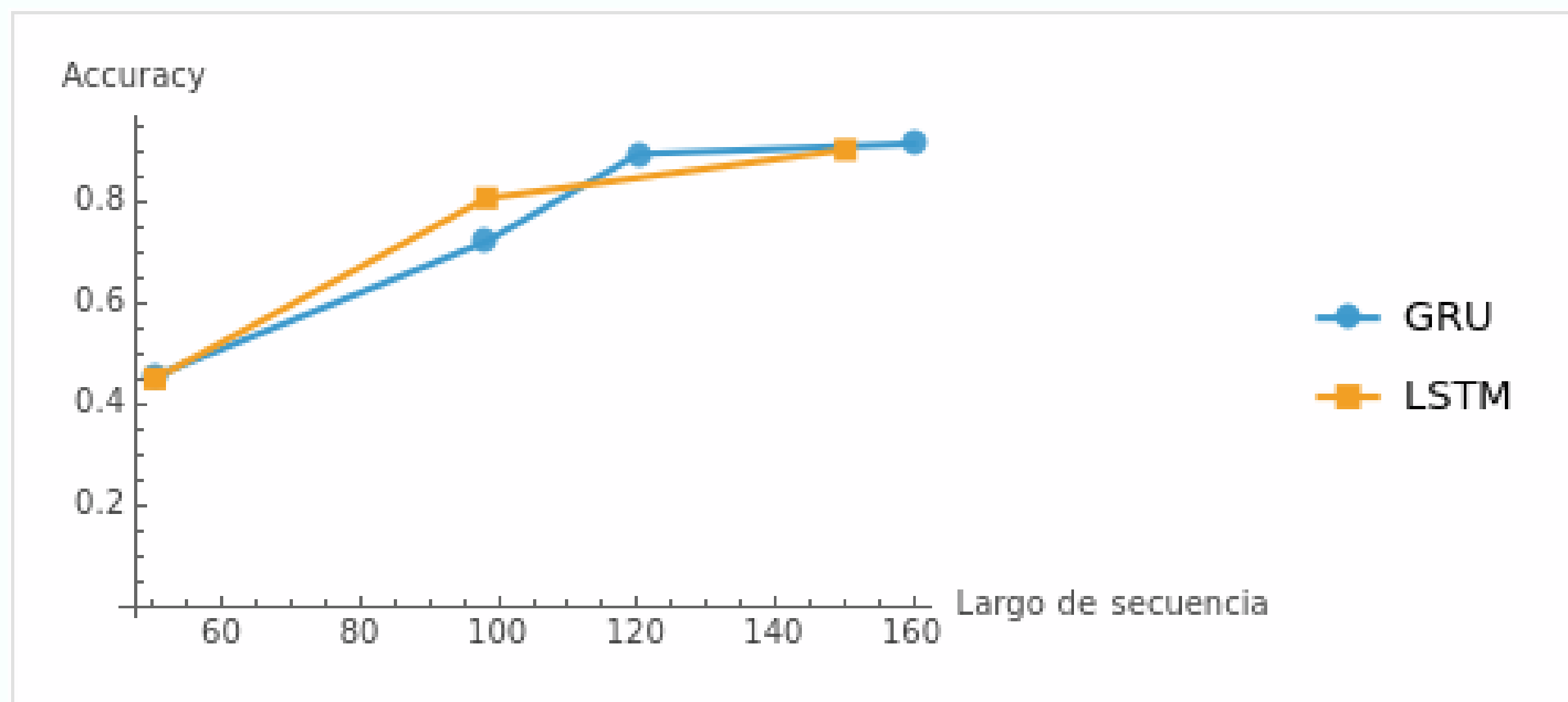
Largo de secuencia: 160

# CONCLUSIONES PRELIMINARES

Modelo	Capas	Neuronas	T	Accuracy	Precision	Recall	F1-score
RNN Vanilla	1	128	98	0.3759	0.38	0.38	0.37
RNN Vanilla	1	128	50	0.3444	0.38	0.35	0.34
LSTM	1	128	98	0.8105	0.82	0.81	0.81
LSTM	1	128	150	0.9064	0.91	0.91	0.91
GRU	1	128	98	0.8242	0.83	0.82	0.83
GRU	1	128	150	0.9187	0.92	0.92	0.92



# CONCLUSIONES PRELIMINARES



# CONCLUSIONES PRELIMINARES

Modelo	Largo de secuencia (T)	Época final	Tiempo (s)
RNN Vanilla	98	8	22
RNN Vanilla	50	12	31
LSTM	98	13	39
LSTM	150	17	60
GRU	98	11	31
GRU	160	9	32