

# **Predicción de Fotogramas con Redes Neuronales Profundas**

Máster Universitario en Visión Artificial

Autor: Nuria Oyaga de Frutos

Tutor: José María Cañas Plaza

Cotutor: Inmaculada Mora Jiménez

# ÍNDICE

---

1. Introducción

---

2. Objetivos

---

3. Infraestructura

---

4. Generación de secuencias

---

5. Predicción con imágenes modeladas

---

6. Predicción con imágenes crudas

---

7. Conclusiones

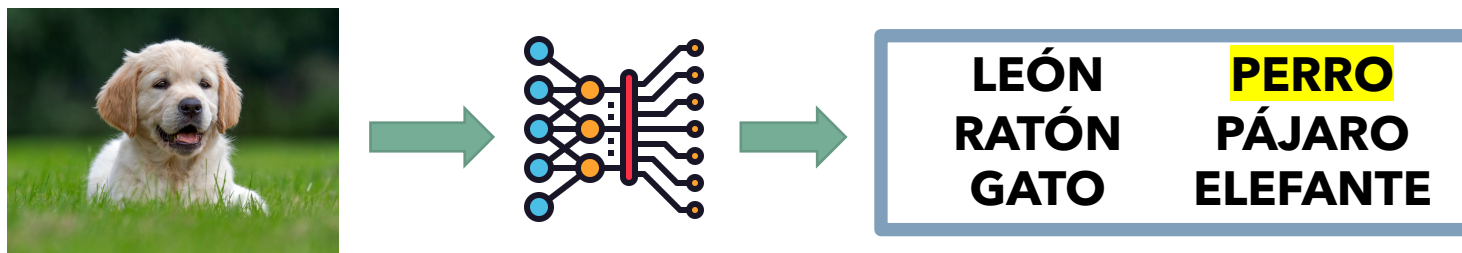
The background features abstract geometric shapes. A large blue semi-circle occupies the right side. A green circle is in the upper left. A green square outline is on the left. Various green dashed lines and a solid green line are scattered in the upper left and center.

# INTRODUCCIÓN

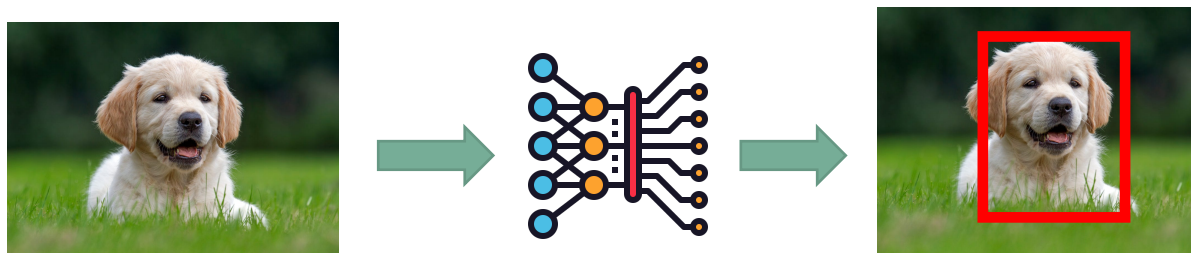
# 1. Introducción

## Tareas de visión artificial con Redes Neuronales

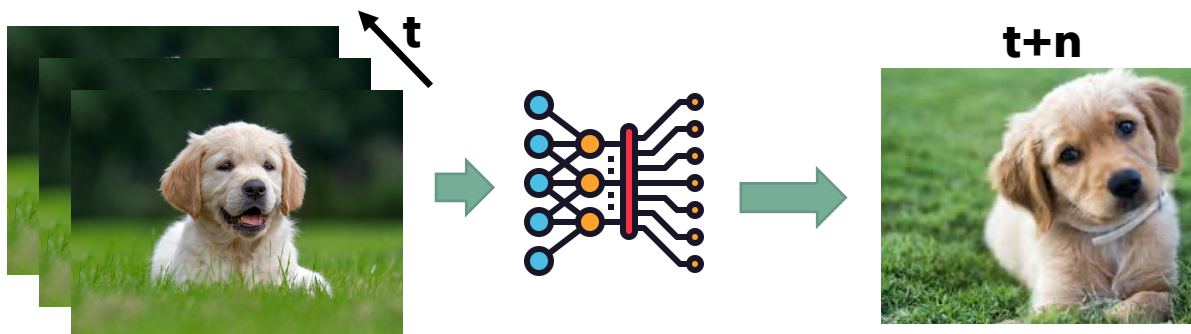
- Clasificación:



- Detección:



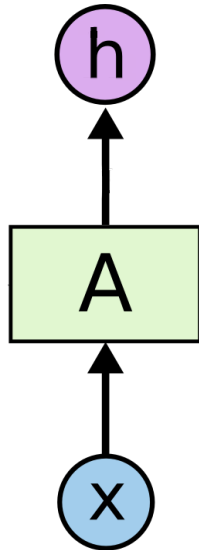
- Predicción :



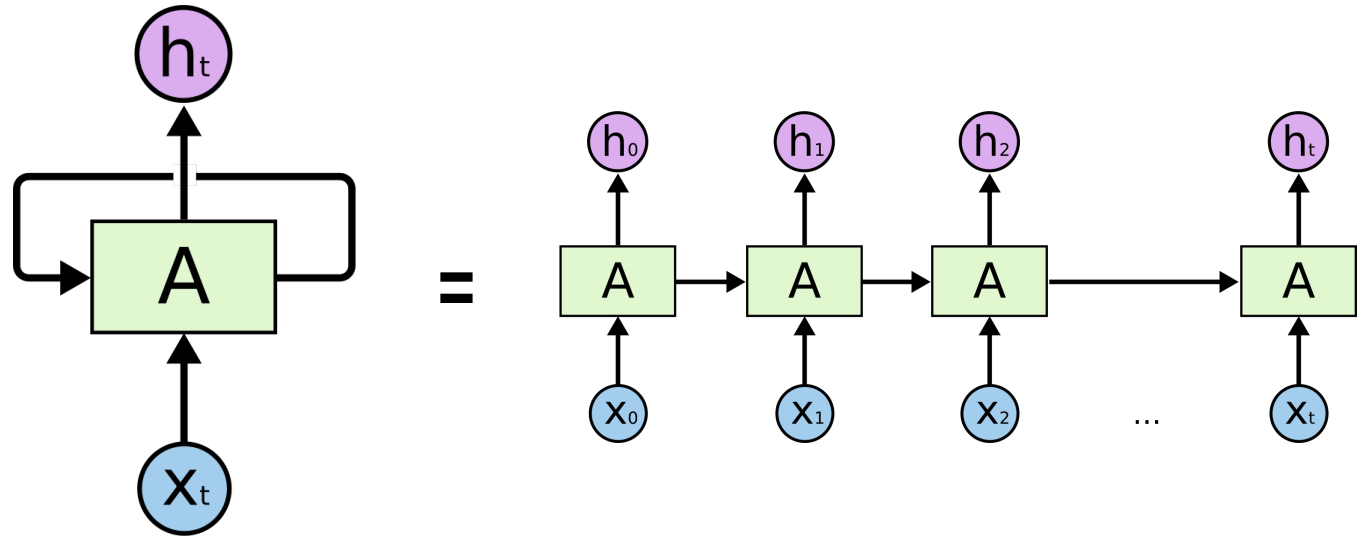
# 1. Introducción

## Redes no recurrentes VS recurrentes

No recurrente



Recurrente



Persistencia del conocimiento



The background features a large blue circle on the right side. To its left is a smaller green circle. Further left are several green dashed lines of varying lengths and orientations. At the top, there are green line segments forming a partial square and a triangle. A green square outline is visible on the left edge.

# OBJETIVOS

## 2. Objetivos

**Diseño y el análisis** de distintas redes neuronales como **predictores visuales** con secuencias de **vídeo**.

- **Desarrollo software** para **ejecución y evaluación de redes** neuronales.
- **Creación de las bases de datos.**
- Estudio y evaluación de redes para la **predicción con imágenes modeladas.**
- Estudio y evaluación de redes para la **predicción con imágenes crudas.**

The background features abstract geometric shapes. A large blue semi-circle occupies the right half of the frame. On the left, there are several green elements: a solid circle, a dashed vertical line, a solid square outline, and several short dashed line segments. A green line also extends from the top center towards the right.

# INFRAESTRUCTURA



### 3. Infraestructura

#### Software



python™

matplotlib



Keras

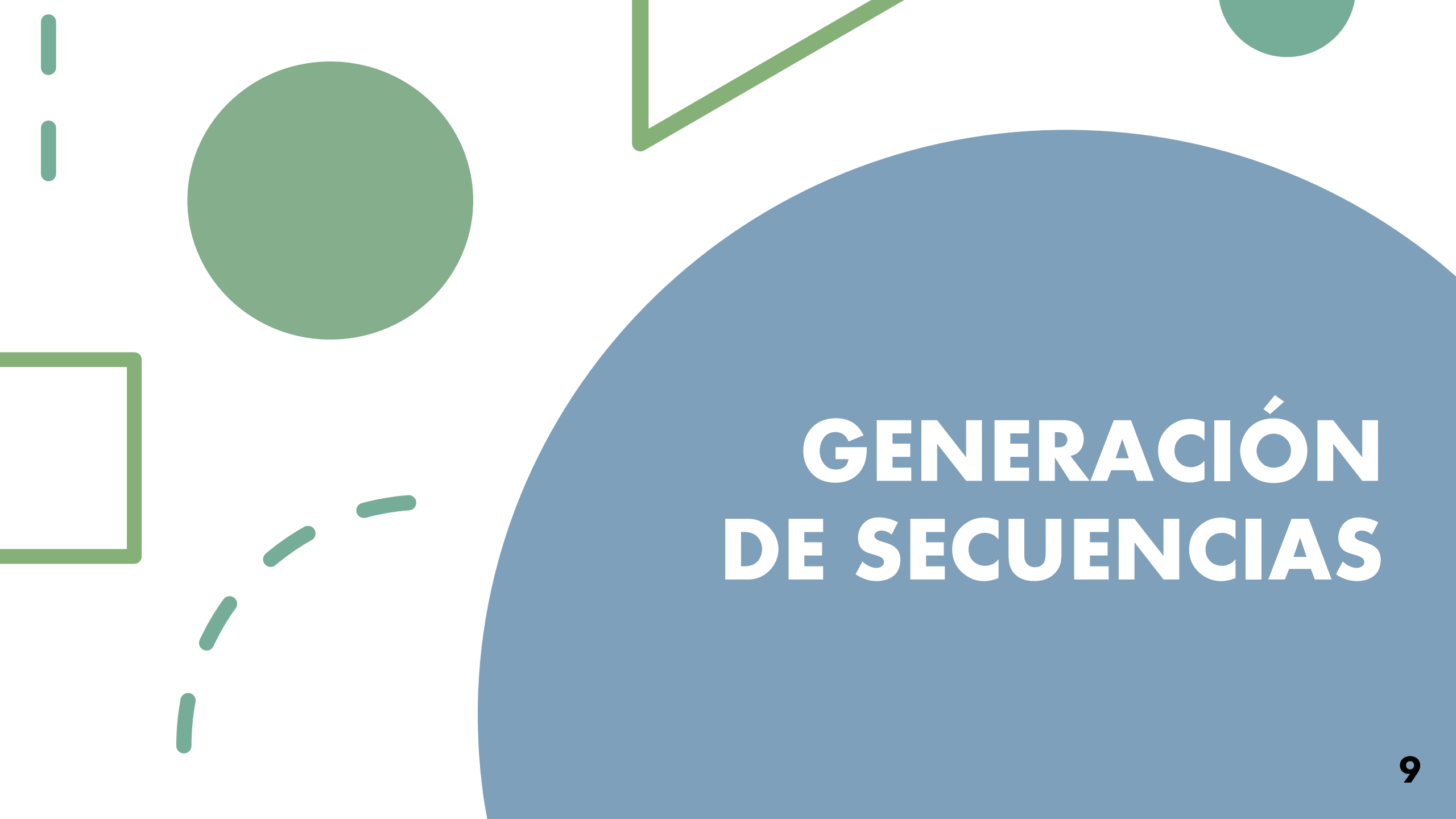


OpenCV

#### Hardware



- Intel(R) Xeon(R) CPU E5-2609 v4 @ 1.70GHz
- 8 cores
- 64GB
- GeForce GTX 1080

The background features abstract geometric shapes. A large blue semi-circle occupies the right side. A green circle is in the upper left. A green square outline is on the left. Green dashed lines and a green line segment are also present.

# **GENERACIÓN DE SECUENCIAS**

## 4. Generación de secuencias

### Limitaciones del trabajo

- Imágenes muy sencillas:
  - Tamaño 80x120.
  - Píxel blanco (activo) que se desplaza sobre fondo negro.
- Muestreo regular:
  - Velocidad constante.
  - No faltan muestras.
- Ausencia de ruido.

# 4. Generación de secuencias

## Tipos de imágenes

**CRUDAS**



| x  | y  |
|----|----|
| 0  | 17 |
| 3  | 18 |
| 6  | 2  |
| 9  | 22 |
| 12 | 23 |
| 15 | 25 |
| 18 | 27 |
| 21 | 28 |
| 24 | 3  |
| 27 | 32 |
| 30 | 34 |
| 33 | 35 |
| 36 | 37 |
| 39 | 39 |
| 42 | 4  |
| 45 | 42 |
| 48 | 44 |
| 51 | 46 |
| 54 | 47 |
| 57 | 49 |
| 87 | 66 |

**MODELADAS**

## 4. Generación de secuencias

### Tipos de dinámicas - Dinámica lineal

$$y = mx + n$$

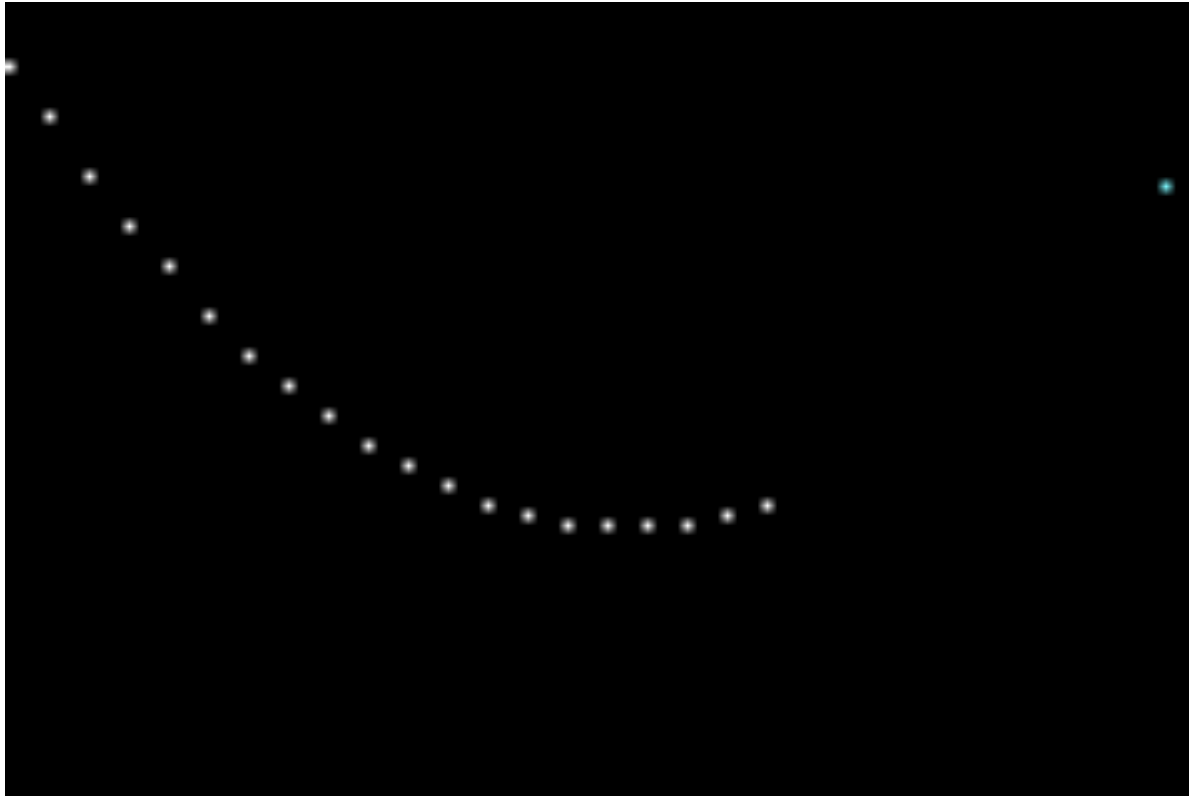


| DOF      | Parámetro |
|----------|-----------|
| <b>1</b> | m         |
| <b>2</b> | n         |

## 4. Generación de secuencias

### Tipos de dinámicas - Dinámica parabólica

$$y = ax^2 + bx + c$$



| DOF      | Parámetro |
|----------|-----------|
| <b>1</b> | a         |
| <b>2</b> | c         |
| <b>3</b> | b         |

## 4. Generación de secuencias

### Tipos de dinámicas - Dinámica sinusoidal

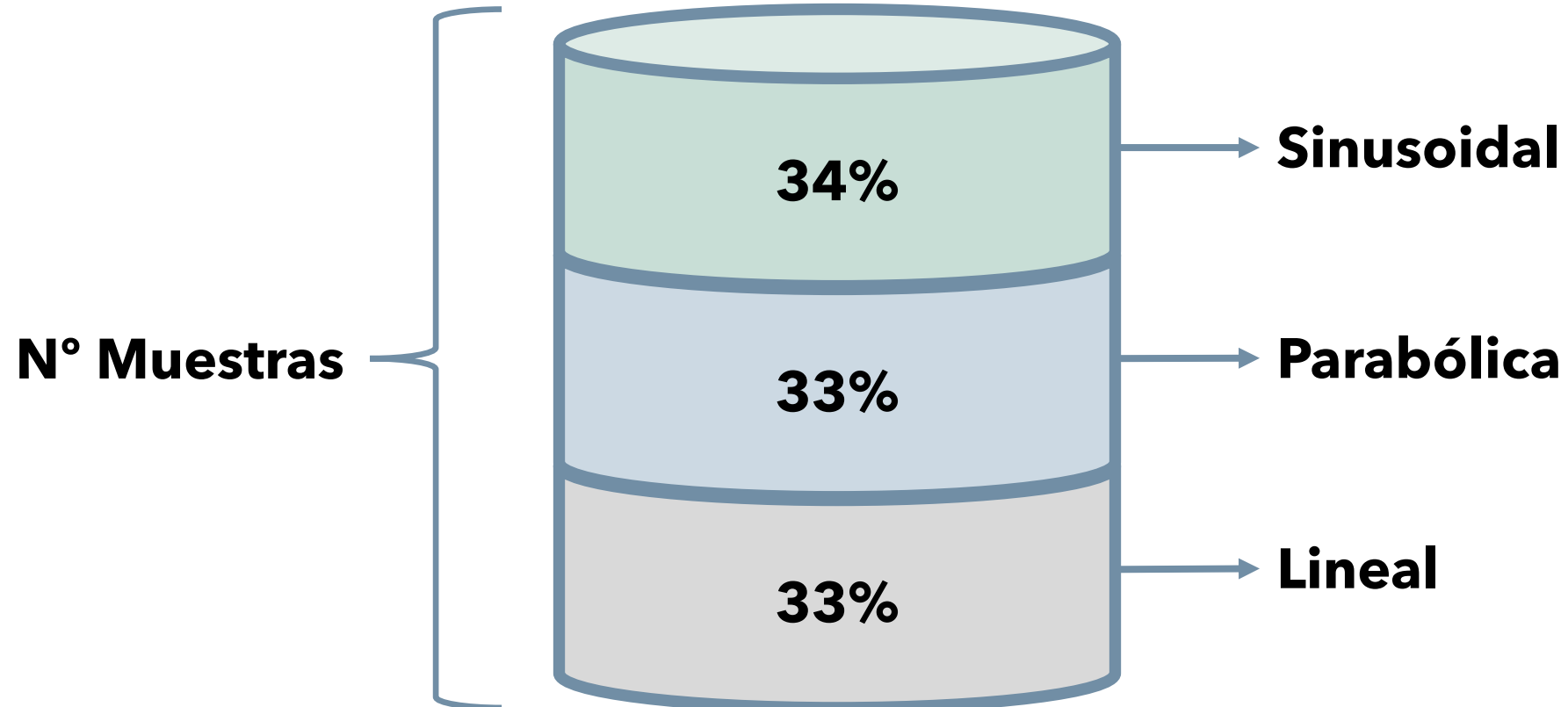
$$y = A \times \sin(2\pi f x + \theta) + c$$



| DOF      | Parámetro |
|----------|-----------|
| <b>1</b> | f         |
| <b>2</b> | c         |
| <b>3</b> | A         |
| <b>4</b> | $\theta$  |

## 4. Generación de secuencias

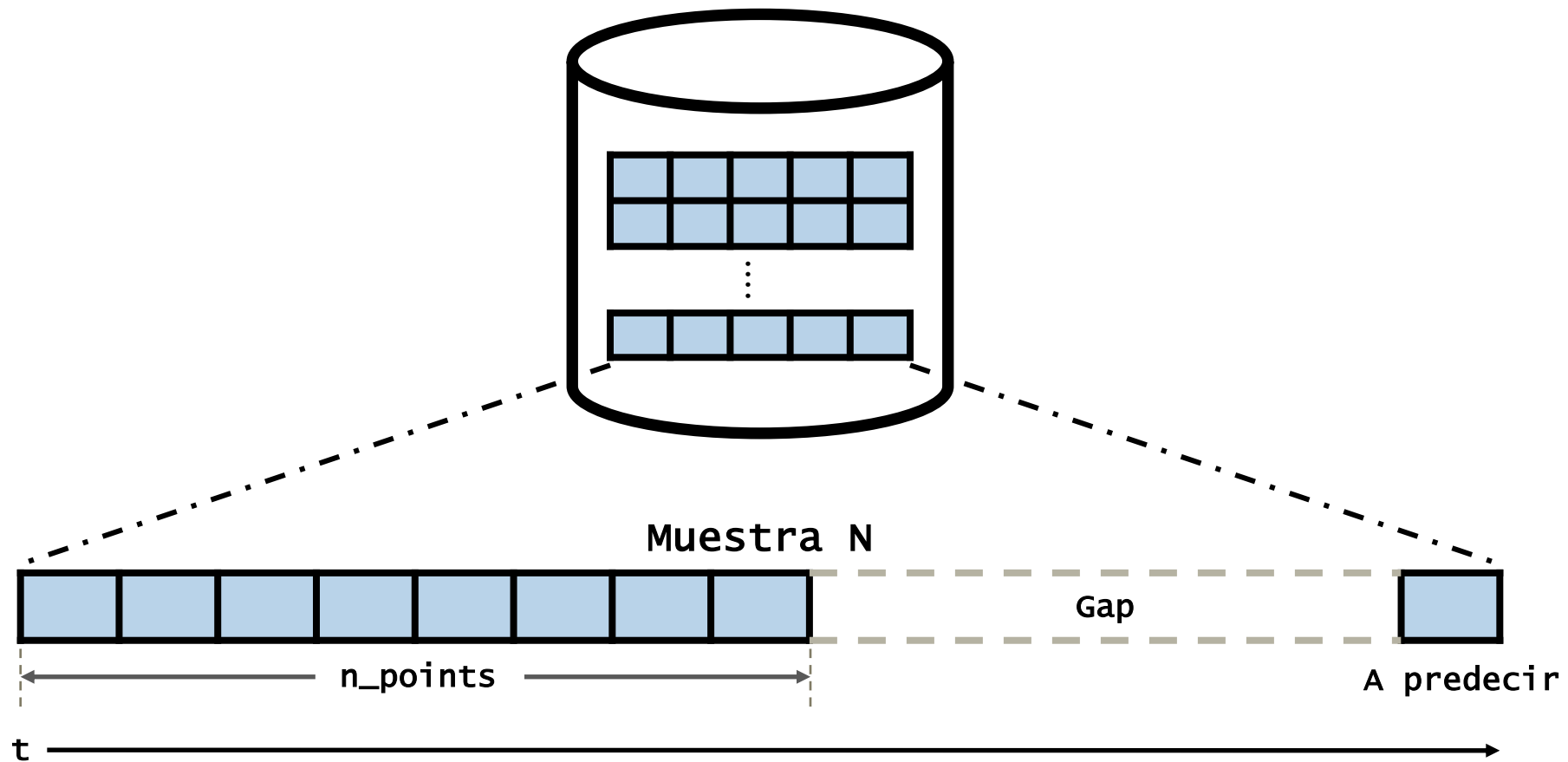
Tipos de dinámicas - Dinámica combinada





## 4. Generación de secuencias

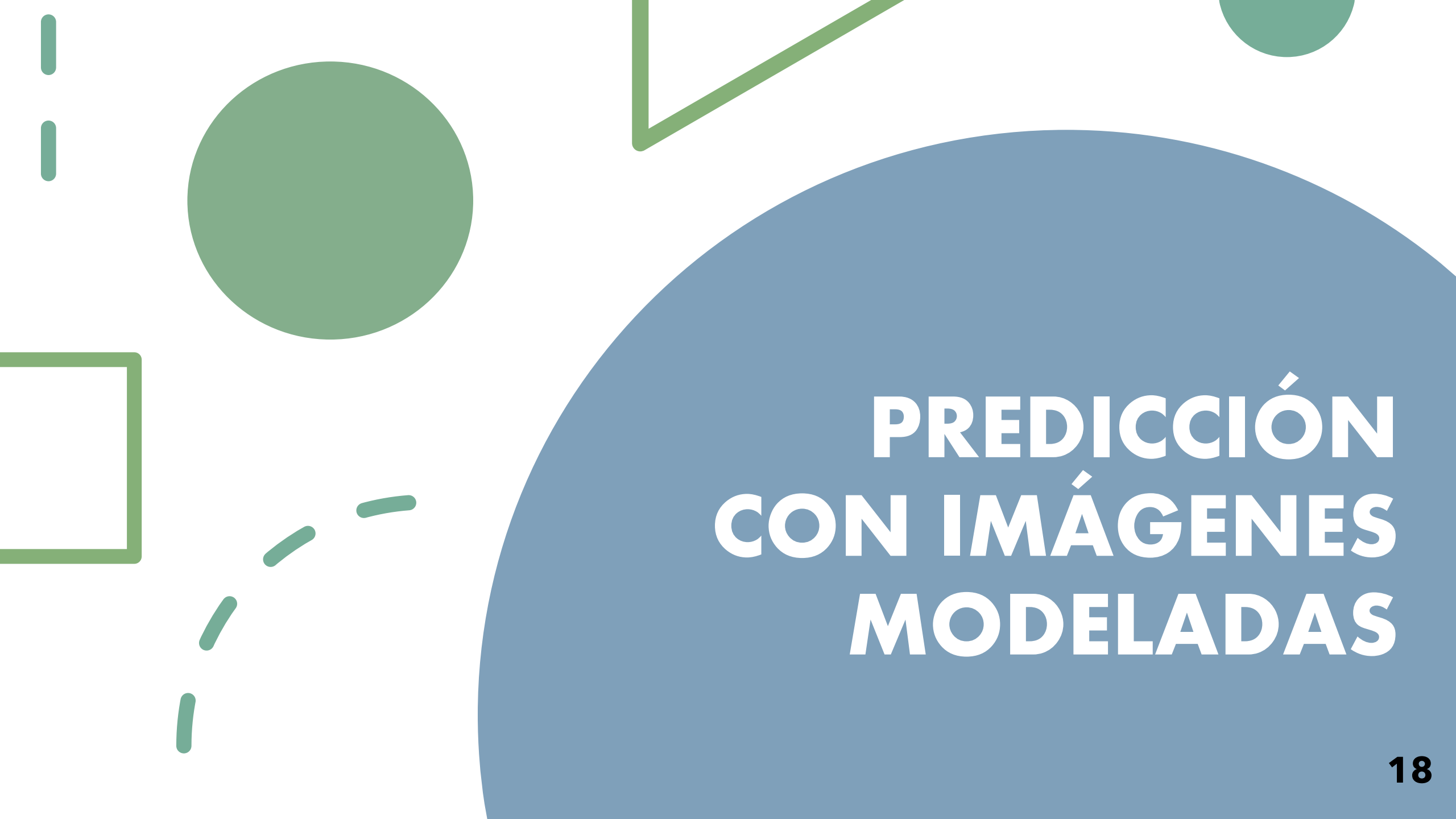
### Estructura de la base de datos



## 4. Generación de secuencias

### Características comunes

| Parámetro                |            | Valor                   |
|--------------------------|------------|-------------------------|
| Gap                      |            | 10 Instantes temporales |
| N_points                 |            | 20 Instantes temporales |
| División de subconjuntos | Train      | 80%                     |
|                          | Validation | 10%                     |
|                          | Test       | 10%                     |

The background features several abstract geometric elements: a large blue semi-circle on the right side, a medium green circle in the upper left, a green square outline on the left, and various green dashed lines and line segments scattered across the white background.

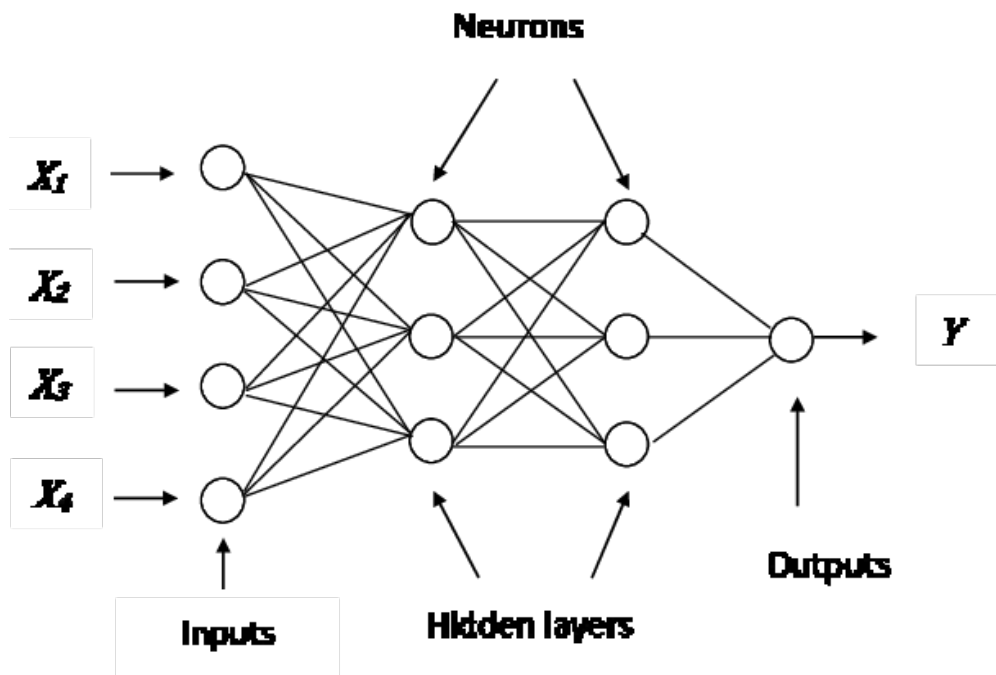
# **PREDICCIÓN CON IMÁGENES MODELADAS**

# 5. Predicción con imágenes modeladas

## Tipos de redes

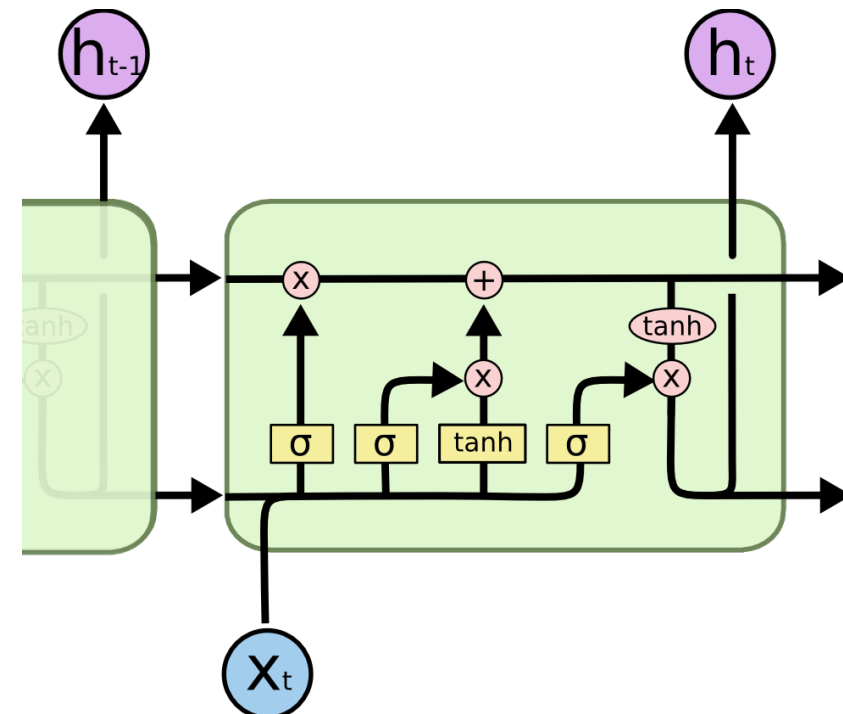
Perceptrón multicapa

**MLP**



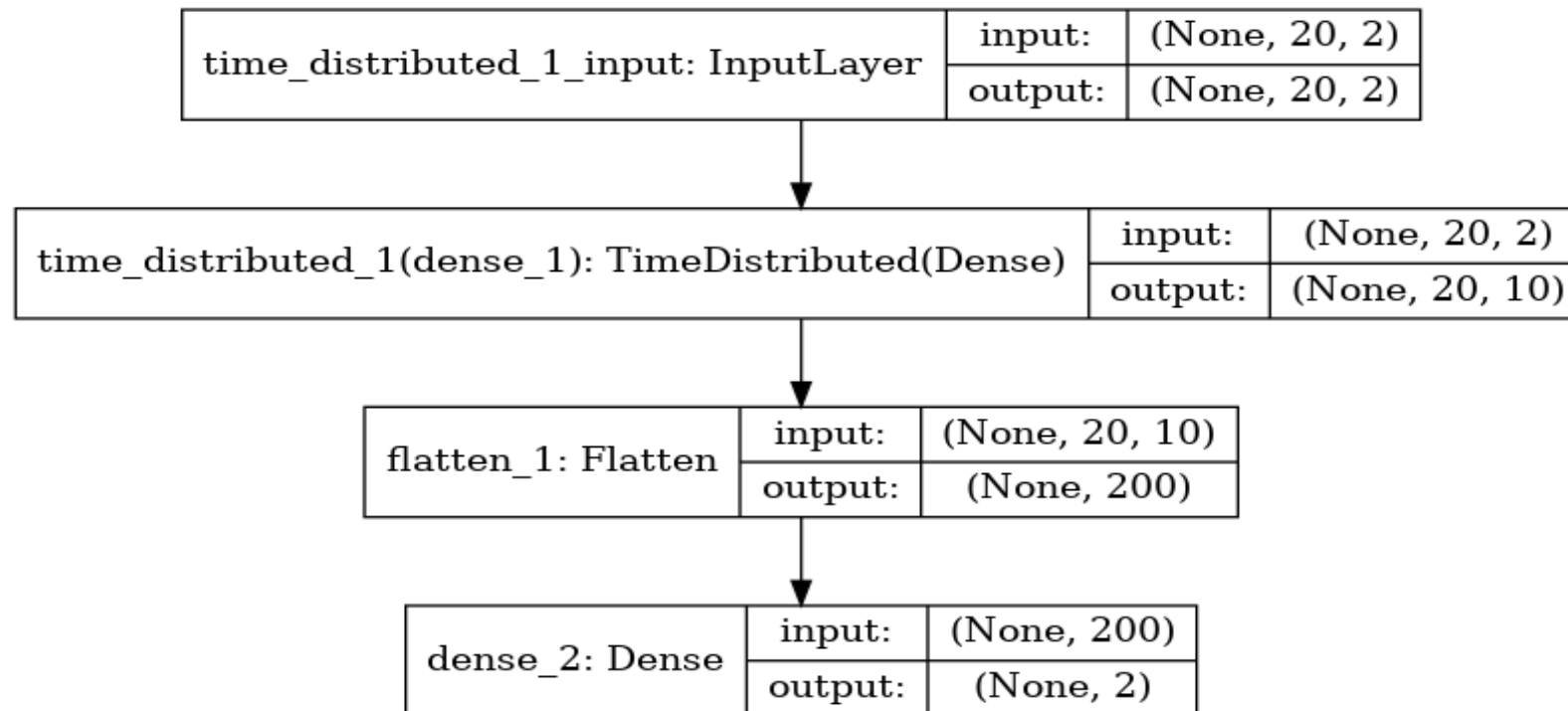
Long-Short Term Memory

**LSTM**



# 5. Predicción con imágenes modeladas

## Estructuras propuestas - MLP



- 1 capa oculta con 10 neuronas.

# 5. Predicción con imágenes modeladas

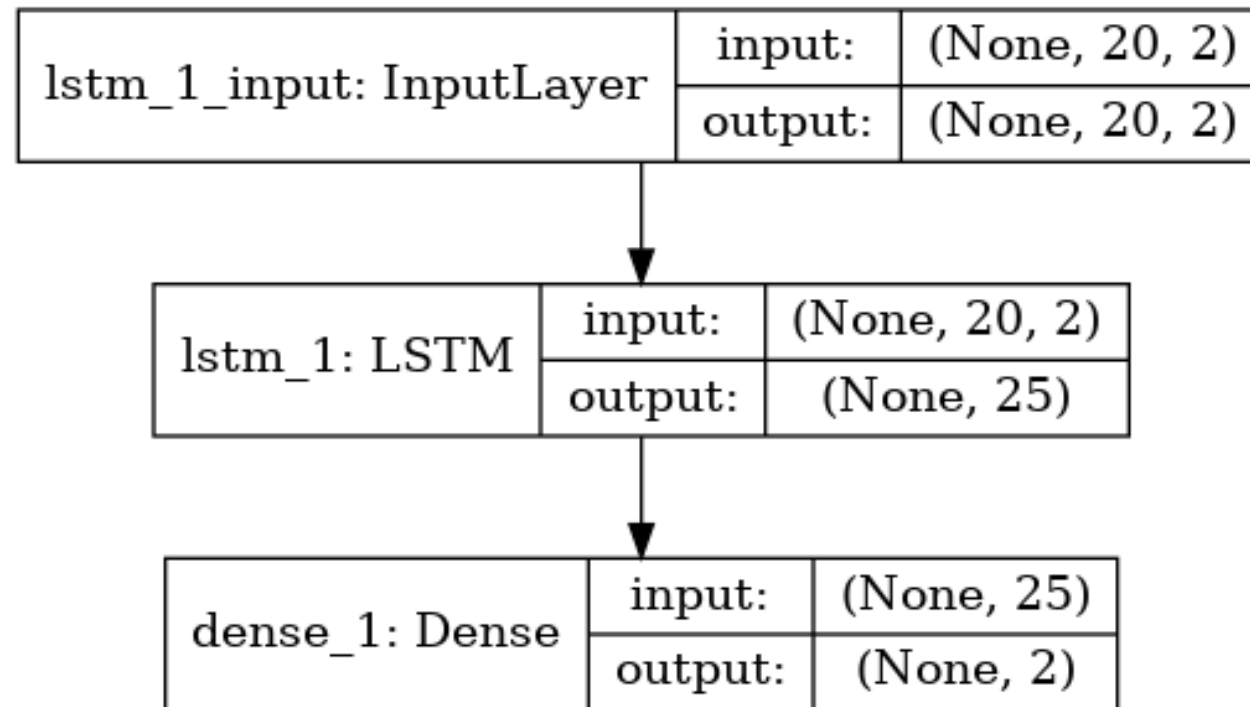
## Resultados - MLP

- 10000 muestras de test.
- Límite en sinusoidal de 2 DOF.

| DINÁMICA   |       | MEDIA ERROR RELATIVO |
|------------|-------|----------------------|
| Lineal     | 1 DOF | 0.21 %               |
|            | 2 DOF | 0.31 %               |
| Parabólica | 1 DOF | 0.28 %               |
|            | 2 DOF | 0.42 %               |
|            | 3 DOF | 0.65 %               |
| Sinusoidal | 1 DOF | 0.54 %               |
|            | 2 DOF | 3.89 %               |

# 5. Predicción con imágenes modeladas

## Estructuras propuestas - LSTM-1



- 1 capa LSTM con 25 neuronas.

# 5. Predicción con imágenes modeladas

## Resultados - LSTM-1

- 10000 muestras de test.
- Límite en sinusoidal de 4 DOF.

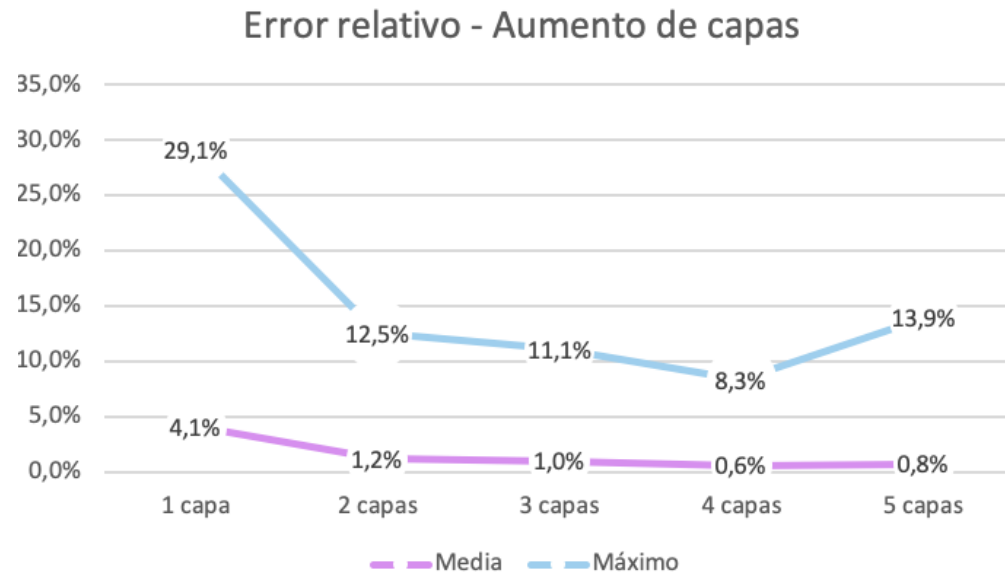
| DINÁMICA   |       | MEDIA ERROR RELATIVO |
|------------|-------|----------------------|
| Lineal     | 1 DOF | 0.16 %               |
|            | 2 DOF | 0.25 %               |
| Parabólica | 1 DOF | 0.12 %               |
|            | 2 DOF | 0.35 %               |
|            | 3 DOF | 0.58%                |
| Sinusoidal | 1 DOF | 0.42 %               |
|            | 2 DOF | 0.89 %               |
|            | 3 DOF | 0.84 %               |
|            | 4 DOF | 4.1 %                |



# 5. Predicción con imágenes modeladas

## LSTM-4 - Mejora de la red

### Aumento n° capas



- Mejora las prestaciones.
- Límite en 4 capas.

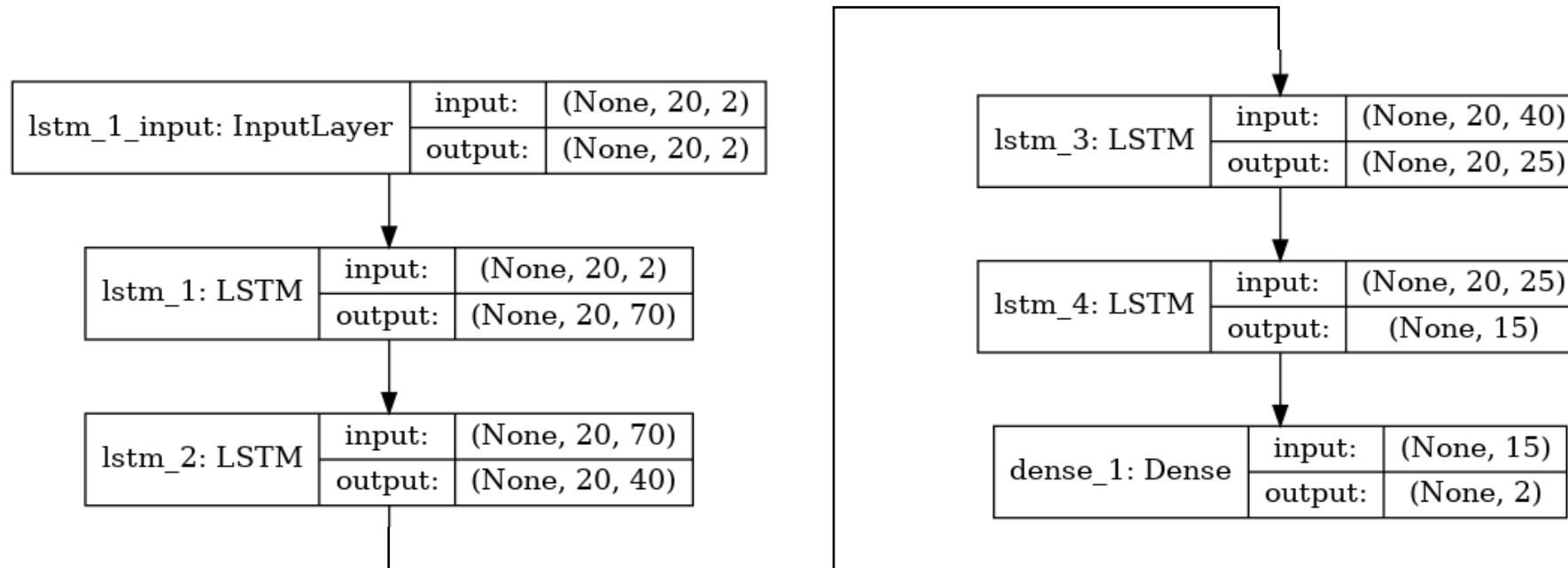
### Aumento n° neuronas

- 25 → 50 neuronas.
- 4% → 2.5% media error relativo.
- Mejora muy poco las prestaciones.

- Dinámica sinusoidal 4 DOF.
- 80000 muestras entrenamiento.
- 10000 muestras test.

# 5. Predicción con imágenes modeladas

## Estructuras propuestas - LSTM-4



- 4 capas LSTM con 70, 40, 25 y 15 neuronas.

# 5. Predicción con imágenes modeladas

## Resultados - LSTM-4

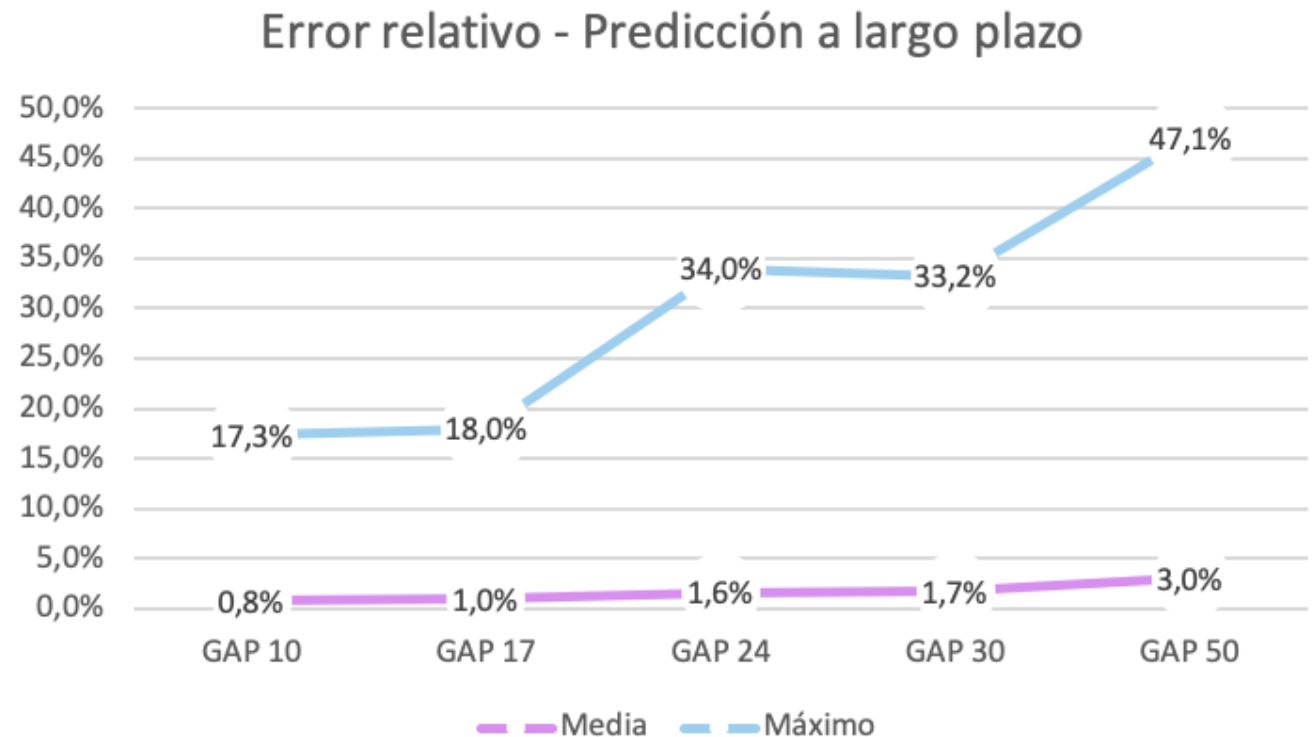
- 10000 muestras de test.
- Se logra predecir.


| DINÁMICA   |       | MEDIA ERROR RELATIVO |
|------------|-------|----------------------|
| Lineal     | 2 DOF | 0.12 %               |
| Parabólica | 3 DOF | 0.5%                 |
| Sinusoidal | 4 DOF | 0.61 %               |
| Combinada  |       | 0.81 %               |

# 5. Predicción con imágenes modeladas

## LSTM-4 - Predicción a largo plazo

- Conjunto combinado.
- 10000 muestras de test.
- Pérdida de capacidad predictiva con umbrales admisibles.
- Imagen 640x480:
  - 14 píxeles de media a 30 fotogramas (1,7 %).
  - 24 píxeles de media a 50 fotogramas (3 %).





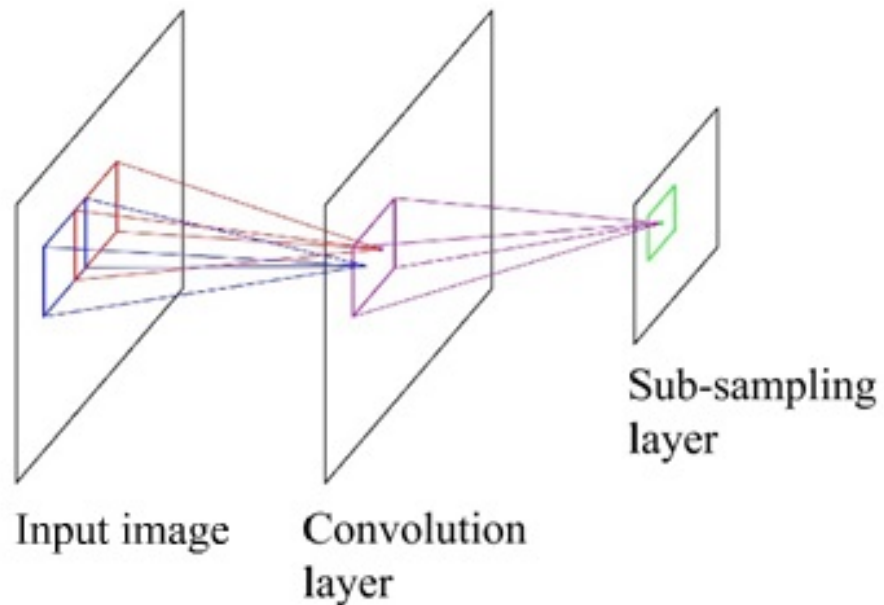
# PREDICCIÓN CON IMÁGENES CRUDAS

# 6. Predicción con imágenes crudas

## Tipos de redes

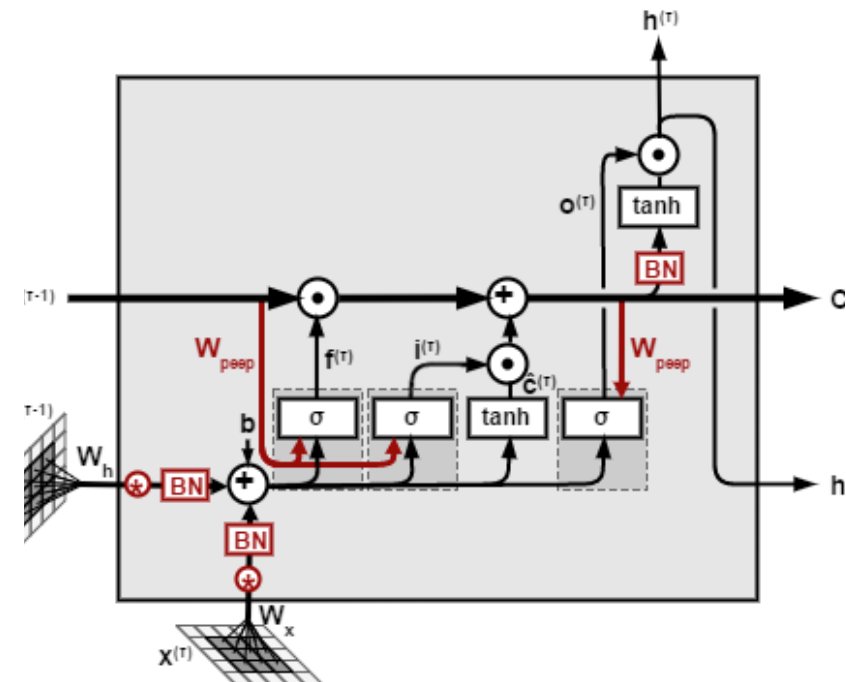
Convolutional Neural Network

**CNN**



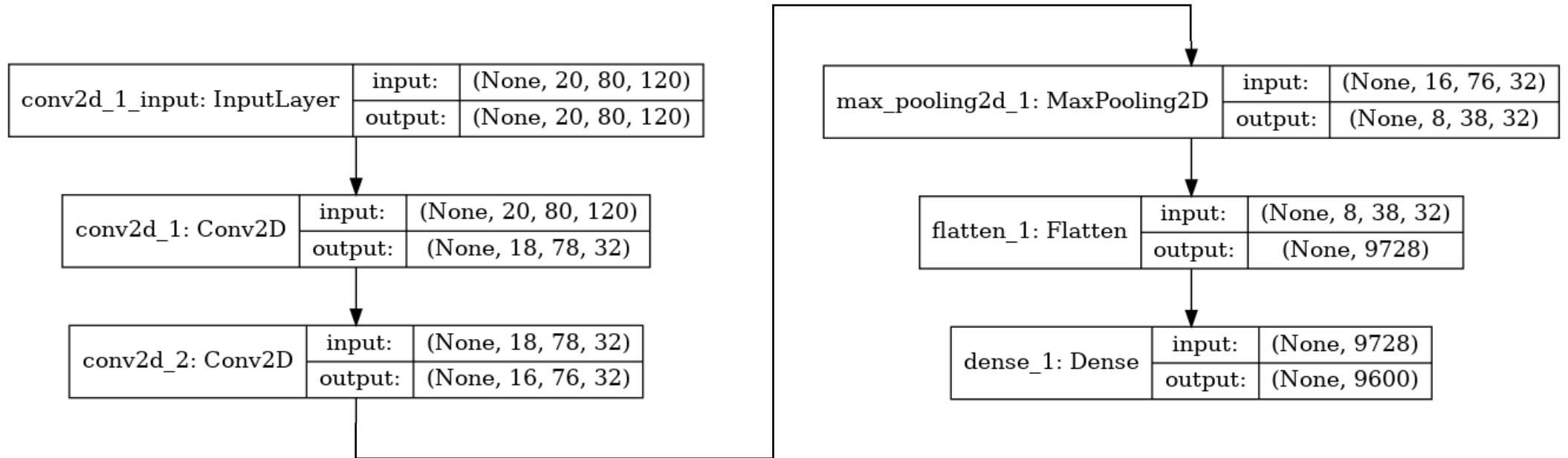
Convolutional LSTM

**ConvLSTM**



# 6. Predicción con imágenes crudas

## Estructuras propuestas - CNN

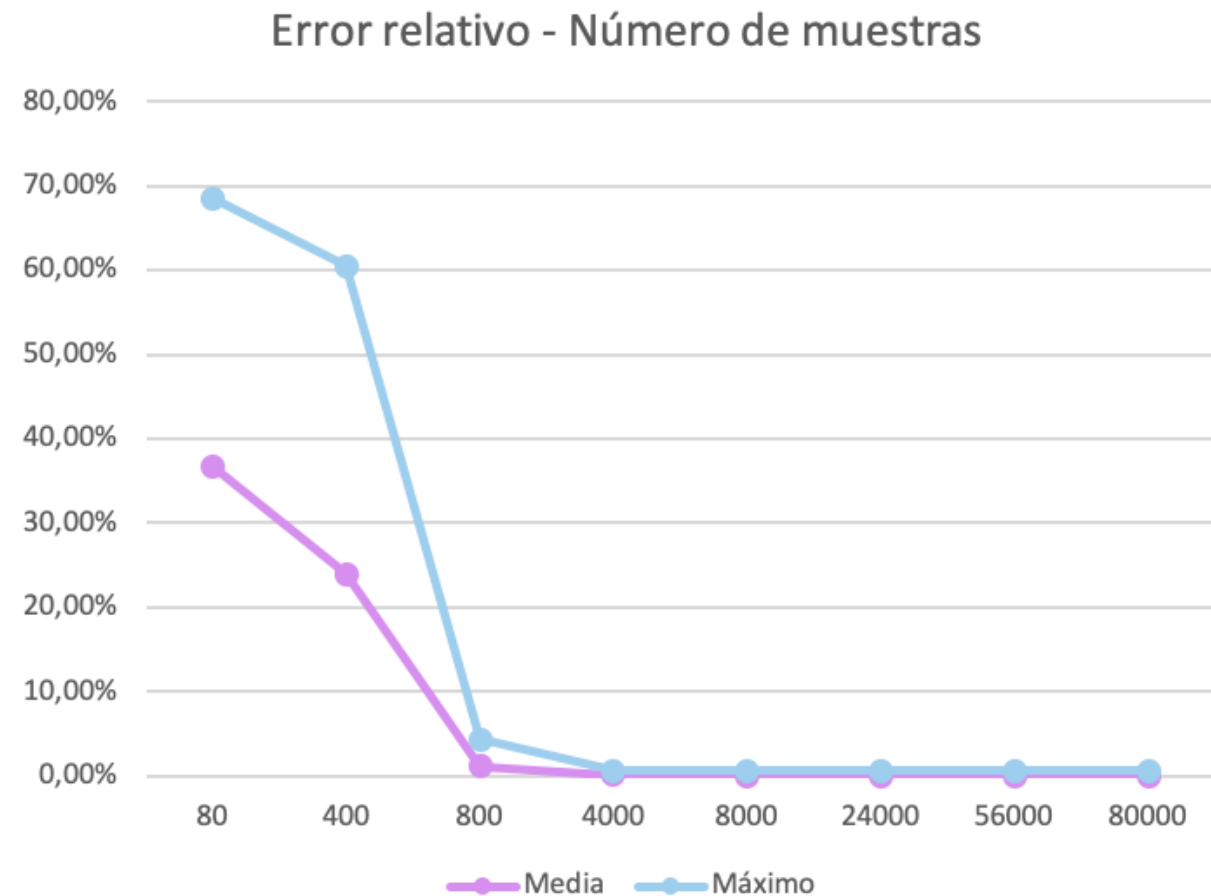


- 2 capas convolucionales con 32 neuronas.
- 1 capa de MaxPooling 2x2.

# 6. Predicción con imágenes crudas

## CNN - Influencia del número de muestras de entrenamiento

- Dinámica lineal 1DOF.
- Evaluación con 1000 muestras.
- Mejora en prestaciones hasta estabilización.
- Equilibrio entre número de muestras y complejidad.





## 6. Predicción con imágenes crudas

### Resultados - CNN

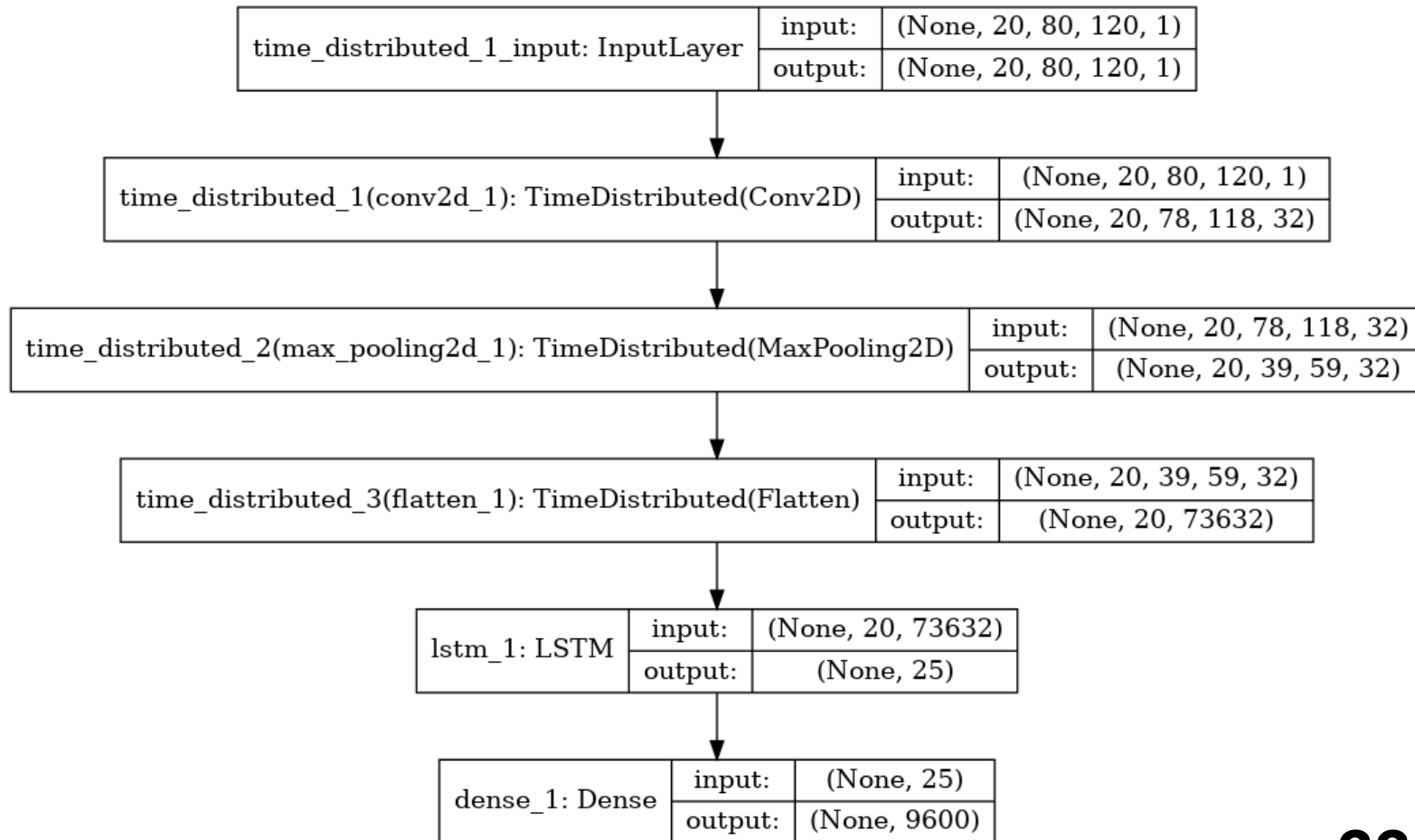
- 10000 muestras de test.
- Valores altos de máximo.
- Muchos *outliers*.

| DINÁMICA   |       | MEDIA ERROR RELATIVO |
|------------|-------|----------------------|
| Lineal     | 1 DOF | 0.07 %               |
|            | 2 DOF | 0.39 %               |
| Parabólica | 1 DOF | 0.01 %               |
|            | 2 DOF | 0.07 %               |
|            | 3 DOF | 4.4 %                |
| Sinusoidal | 1 DOF | 0.003 %              |
|            | 2 DOF | 1.12 %               |

# 6. Predicción con imágenes crudas

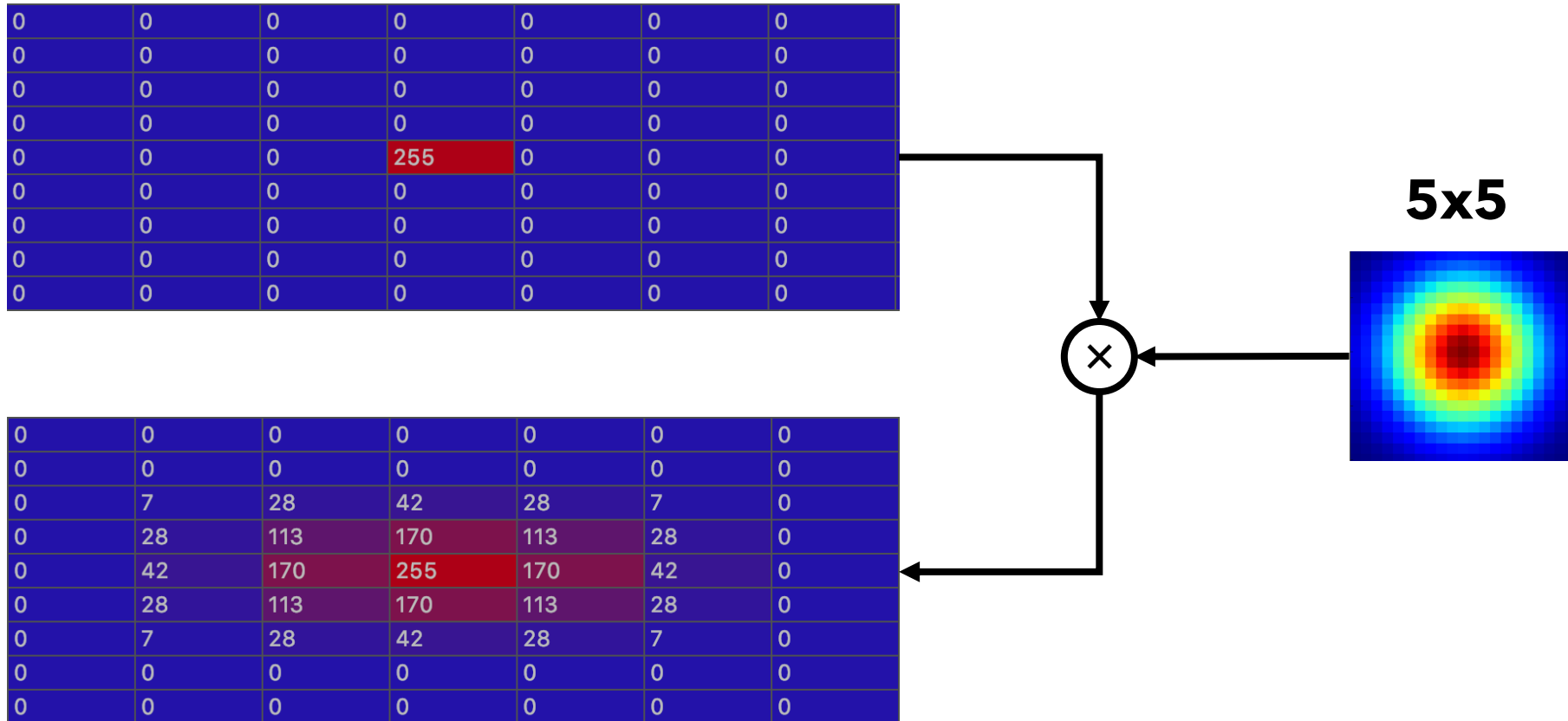
## Estructuras propuestas - CNN + LSTM

- 1 capa convolucional con 32 neuronas.
- 1 capa de MaxPooling 2x2.
- 1 capa LSTM con 25 neuronas.



## 6. Predicción con imágenes crudas

CNN+LSTM - Píxel discreto VS extendido



## 6. Predicción con imágenes crudas

### Resultados - CNN+LSTM

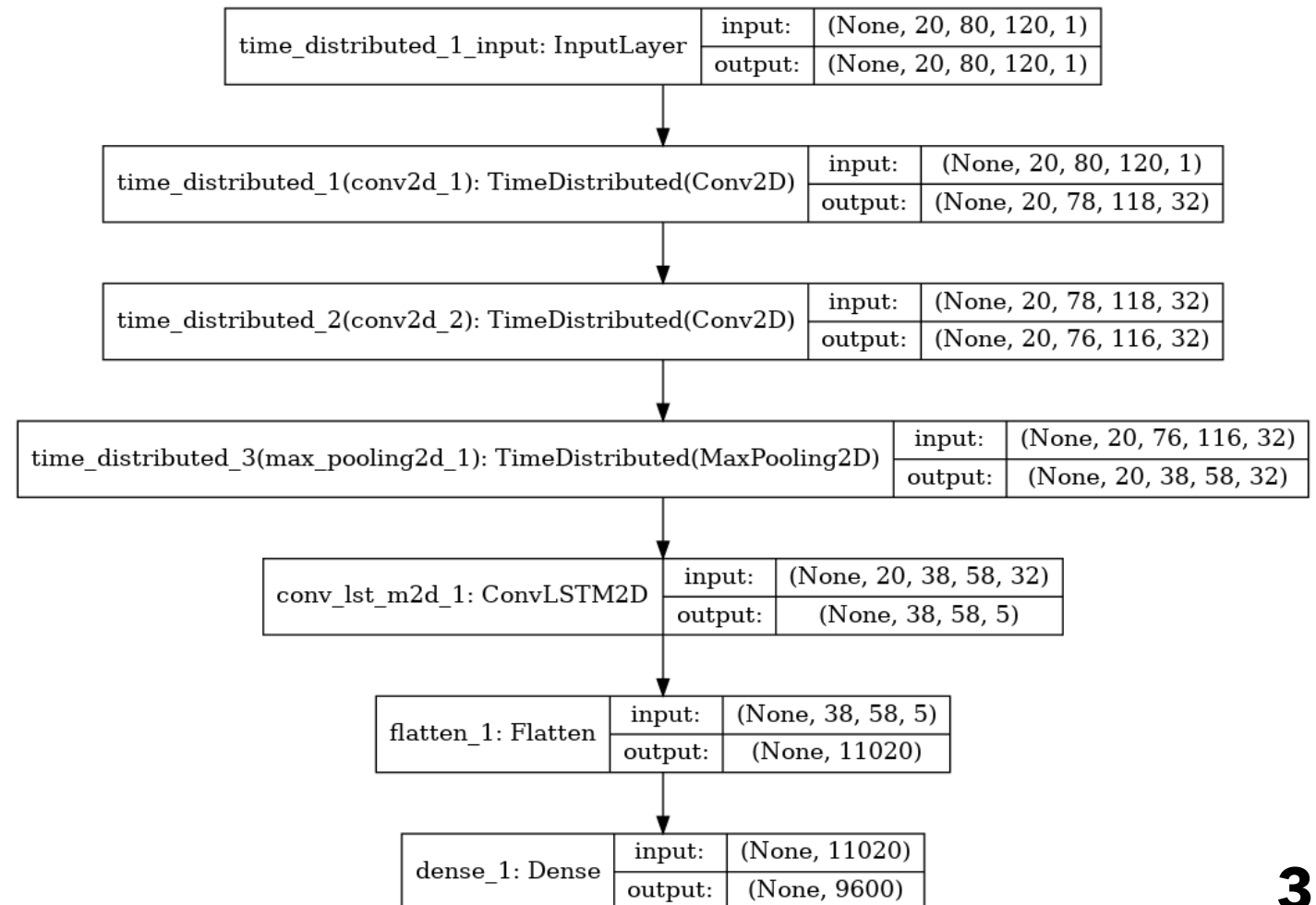
- Dinámica lineal:
  - Pendiente nula.
  - Altura inicial del píxel fija.
  - 800 entrenamiento; 100 test.
- No es una estrategia adecuada.
- La expansión del píxel produce mejora.

| DINÁMICA |           | CNN   | CNN + LSTM |
|----------|-----------|-------|------------|
| Lineal   | Discreto  | 0.0 % | 29.6 %     |
|          | Expandido |       | 21.5 %     |

# 6. Predicción con imágenes crudas

## Estructuras propuestas - ConvLSTM-1

- 2 capas convolucionales con 32 neuronas.
- 1 capa de MaxPooling 2x2.
- 1 capa ConvLSTM con 5 neuronas.



## 6. Predicción con imágenes crudas

### Resultados - ConvLSTM-1

- 10000 muestras de test.
- Valores altos de máximo.
- Muchos *outliers*.

| DINÁMICA   |       | MEDIA ERROR RELATIVO |
|------------|-------|----------------------|
| Lineal     | 1 DOF | 0.06 %               |
|            | 2 DOF | 0.29 %               |
| Parabólica | 1 DOF | 0.01 %               |
|            | 2 DOF | 0.03 %               |
|            | 3 DOF | 3.76%                |
| Sinusoidal | 1 DOF | 0.01 %               |
|            | 2 DOF | 1.12 %               |
|            | 3 DOF | 3.44 %               |
|            | 4 DOF | 13 %                 |

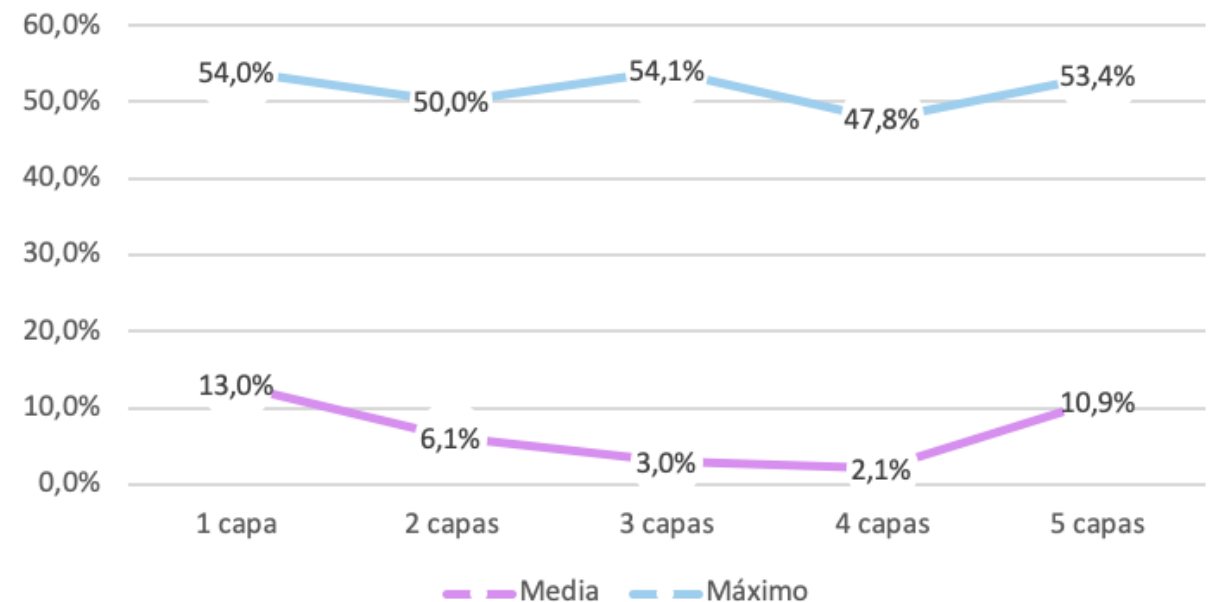
# 6. Predicción con imágenes crudas

## ConvLSTM-4 - Aumento de capas

- Mejora la media.
- Límite en 4 capas.

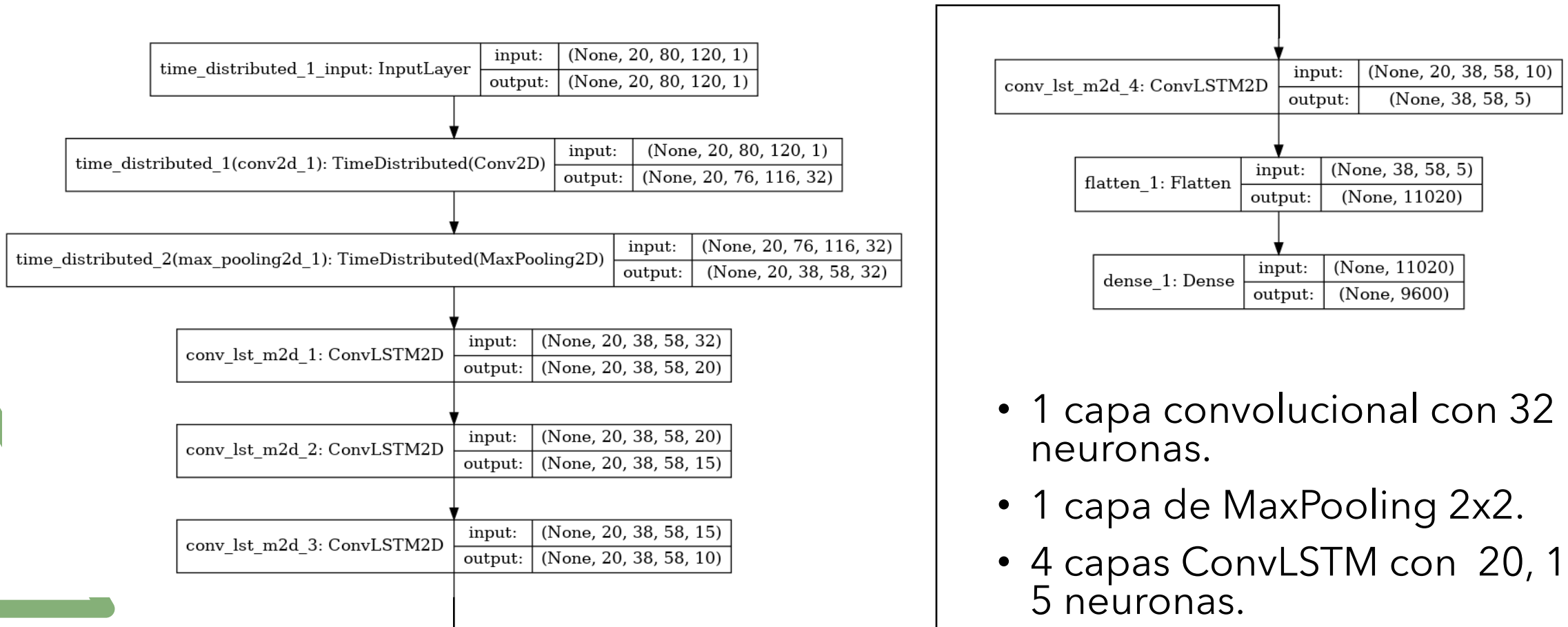
- Dinámica sinusoidal 4 DOF.
- 80000 muestras entrenamiento.
- 10000 muestras test.

Error relativo - Aumento de capas



# 6. Predicción con imágenes crudas

## Estructuras propuestas - ConvLSTM-4





## 6. Predicción con imágenes crudas

### Resultados - ConvLSTM-4

- 10000 muestras de test.
- Mejores resultados.
- Valores altos de máximo.
- Muchos *outliers*.

| DINÁMICA   |       | MEDIA ERROR RELATIVO |
|------------|-------|----------------------|
| Lineal     | 2 DOF | 0.07 %               |
| Parabólica | 3 DOF | 0.87 %               |
| Sinusoidal | 2 DOF | 0.14 %               |
|            | 3 DOF | 0.51 %               |
|            | 4 DOF | 2.06%                |
| Combinada  |       | 2.01 %               |

The background features abstract geometric shapes. A large blue semi-circle occupies the right side. A green circle is in the upper left. A green square outline is on the left. Various green dashed lines and a solid green line are scattered across the white background.

# CONCLUSIONES

# 7. Conclusiones

Se puede predecir

- Las **imágenes modeladas** son **más sencillas** para las redes que las crudas.
- El número de muestras de entrenamiento afecta.
  - **↑ Complejidad ↑ Muestras**
- La **recurrencia mejora** los resultados.
- **↑ N° capas → ↑ Prestaciones**
- Conviene el uso de redes que capten las **correlaciones espacio-temporales simultáneamente** (ConvLSTM).
- **↑ Gap → ↓ Prestaciones**

# 7. Conclusiones

## Líneas futuras

- **Predicción de objetos reales** en movimiento:
  - Mayor tamaño de imagen.
  - Distintas formas y tamaños de objeto.
  - Dinámicas ruidosas.
  - Presencia de aceleración.
  - Pérdida de muestras.
- Integración en **aplicación real** → Seguimiento.



# Enlaces

- Repositorio:  
<https://github.com/RoboticsLabURJC/2017-tfm-nuria-oyaga>
- Bitácora:  
<https://roboticslaburjc.github.io/2017-tfm-nuria-oyaga/>