# Uso de aprendizado de máquina para detecção de faces falsas geradas por inteligência artificial

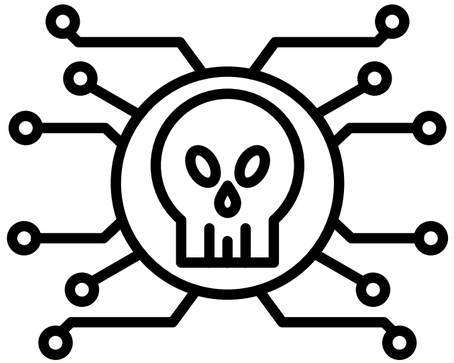
João Pedro Vieira Rodrigues, 201022613
Orientador: Prof. Dr. Kelton Augusto Pontara da Costa
Bacharelado em Ciência da Computação - UNESP / Bauru

# INTRODUÇÃO

# PROBLEMA

**O avanço da inteligência artificial permitiu a criação de rostos falsos através de algoritmos, acarretando problemas como:**


CRIAÇÃO DE NOTÍCIAS FALSAS



ATAQUES CIBERNÉTICOS

CRIAÇÃO DE PERFIS FALSOS

# JUSTIFICATIVA

**Principais motivações do trabalho:**

- Área da segurança digital em alta
- O aprendizado de máquina é uma alternativa para resolver o problema

# OBJETIVOS

- **Objetivo Geral:**

Avaliar o desempenho da Capsule Neural Network na classificação de faces falsas.

- **Objetivos Específicos**

1. Estudar e implementar a Capsule Neural Network
2. Treinar essa rede para reconhecer faces falsas e verdadeiras
3. Interpretar os dados gerados pelo treinamento
4. Comparação do seu desempenho com outras técnicas

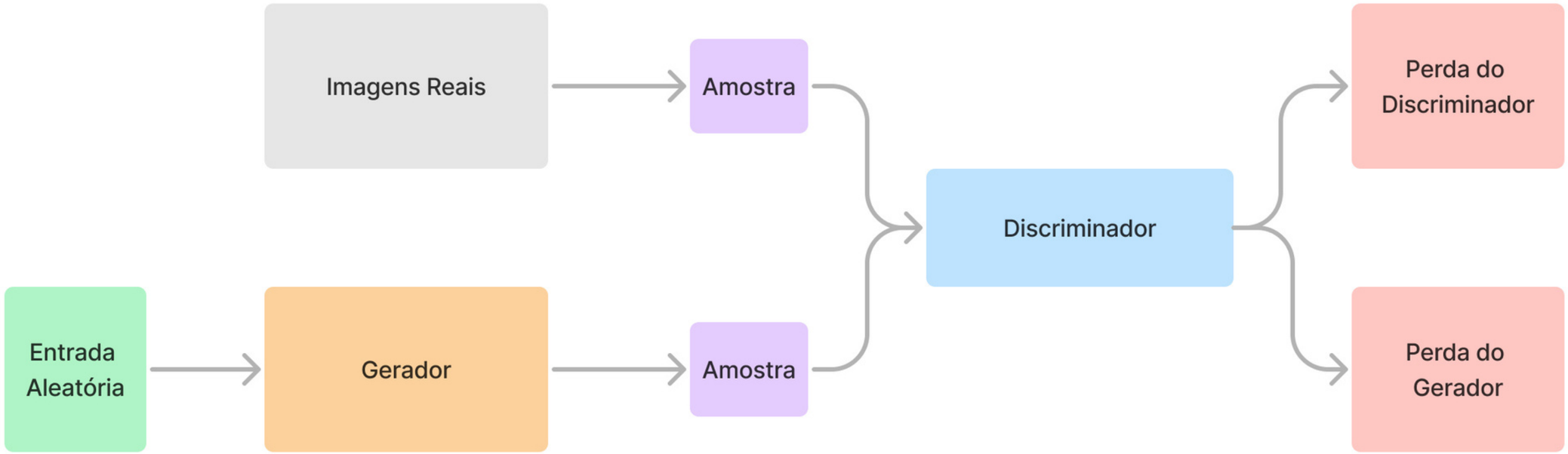
# FUNDAMENTAÇÃO TEÓRICA

# GENERATIVE ADVERSARIAL NETWORK

- Publicado em 2014 por Ian Goodfellow no artigo "*Generative Adversarial Network*"
- Algoritmo utilizado para criação das imagens falsas do banco de dados

# GENERATIVE ADVERSARIAL NETWORK

- Representação de uma GAN

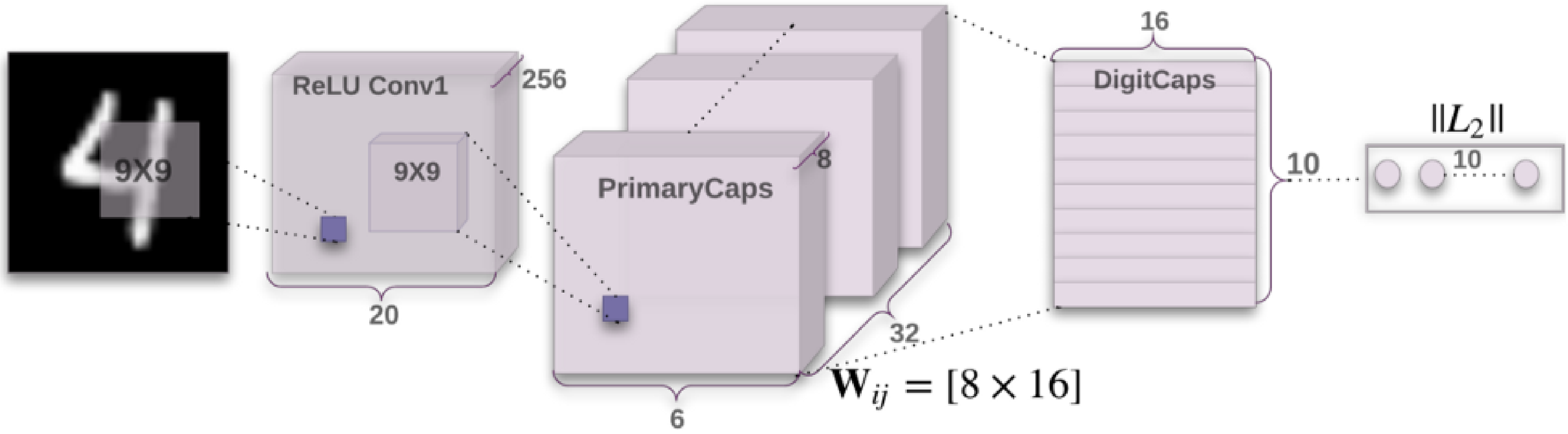


# CAPSULE NEURAL NETWORK

- Proposta em 2017 no artigo *"Dynamic routing between capsules"*
- Utiliza o conceito de cápsulas para extração de características de uma imagem

# CAPSULE NEURAL NETWORK

- Representação da CapsNet

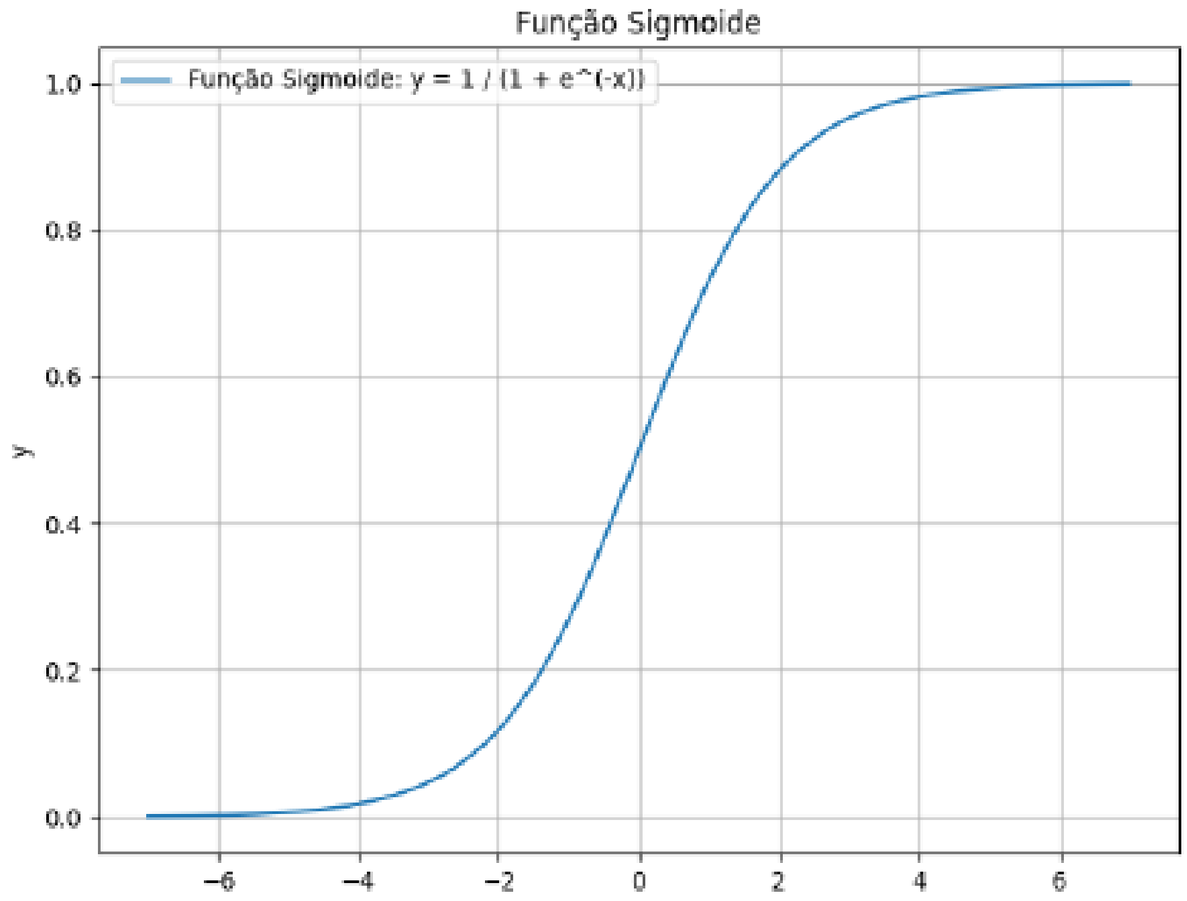


# FUNÇÃO DE ATIVAÇÃO

## Sigmoid:

Fórmula:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



# MÉTRICAS DE AVALIAÇÃO

**As métricas utilizadas no trabalho foram a acurácia e precisão**

## Acurácia

$$\frac{VP + VN}{VP + VN + FN + FP}$$

## Precisão

$$\frac{VP}{VP + FP}$$

# METODOLOGIA

# BANCO DE DADOS

## Banco de dados de 140 mil imagens de rostos falsos e verdadeiros do site Kaggle

- 70 mil rostos falsos criados pela Style-GAN
- 70 mil rostos verdadeiros retirados do site Flickr

# BANCO DE DADOS

Imagem falsa

Imagem real

# TREINAMENTO, VALIDAÇÃO E TESTE

**As imagens foram divididas de acordo com a tabela:**

| | **Rostos Reais** | **Rostos Falsos** |
|---|---|---|
| **Treinamento** | 50 mil imagens | 50 mil imagens |
| **Validação** | 10 mil imagens | 10 mil imagens |
| **Teste** | 10 mil imagens | 10 mil imagens |

# RESULTADOS

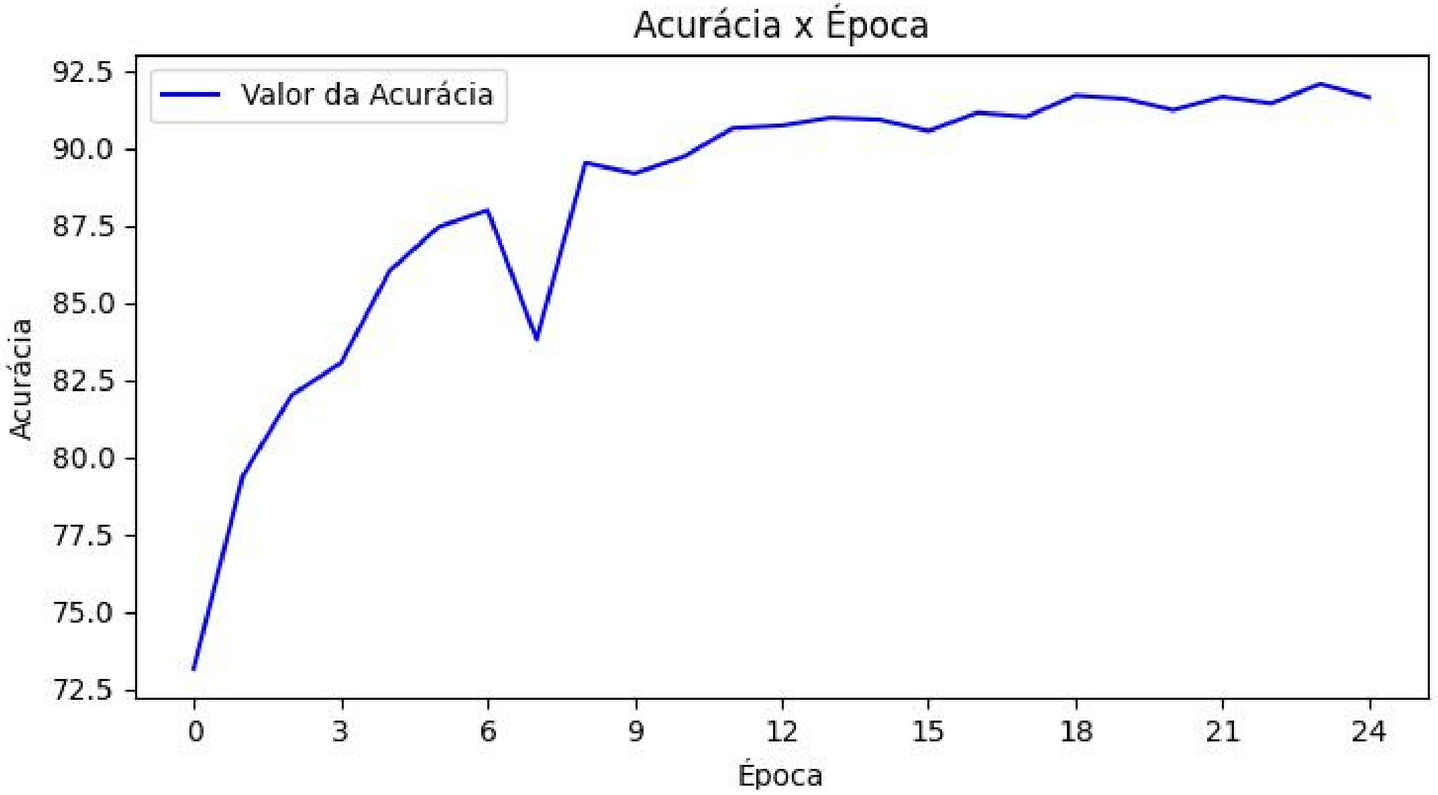
# TREINAMENTO E VALIDAÇÃO

| Época | Perda de Treinamento | Perda de Validação | Acurácia |
|---|---|---|---|
| 1 | 0.0348 | 0.0257 | 0.7316 |
| 2 | 0.0236 | 0.0215 | 0.7939 |
| 3 | 0.0202 | 0.0196 | 0.8201 |
| 4 | 0.0182 | 0.0189 | 0.8304 |
| 5 | 0.0166 | 0.0164 | 0.8602 |
| 6 | 0.0155 | 0.0154 | 0.8745 |
| 7 | 0.0145 | 0.0150 | 0.8798 |
| 8 | 0.0137 | 0.0174 | 0.8380 |
| 9 | 0.0130 | 0.0138 | 0.8951 |
| 10 | 0.0124 | 0.0139 | 0.8918 |
| 11 | 0.0119 | 0.0136 | 0.8972 |
| 12 | 0.0114 | 0.0127 | 0.9063 |
| 13 | 0.0110 | 0.0128 | 0.9072 |
| 14 | 0.0106 | 0.0124 | 0.9098 |
| 15 | 0.0102 | 0.0127 | 0.9091 |
| 16 | 0.0099 | 0.0129 | 0.9055 |
| 17 | 0.0097 | 0.0126 | 0.9114 |
| 18 | 0.0094 | 0.0120 | 0.9101 |
| 19 | 0.0092 | 0.0117 | 0.9170 |
| 20 | 0.0090 | 0.0117 | 0.9159 |
| 21 | 0.0087 | 0.0120 | 0.9123 |
| 22 | 0.0085 | 0.0117 | 0.9166 |
| 23 | 0.0083 | 0.0116 | 0.9144 |
| 24 | 0.0081 | 0.0114 | 0.9207 |
| 25 | 0.0080 | 0.0116 | 0.9164 |

- Época com menor perda de treinamento: 25
- Época com menor perda de validação: 24
- Época com a melhor acurácia: 24

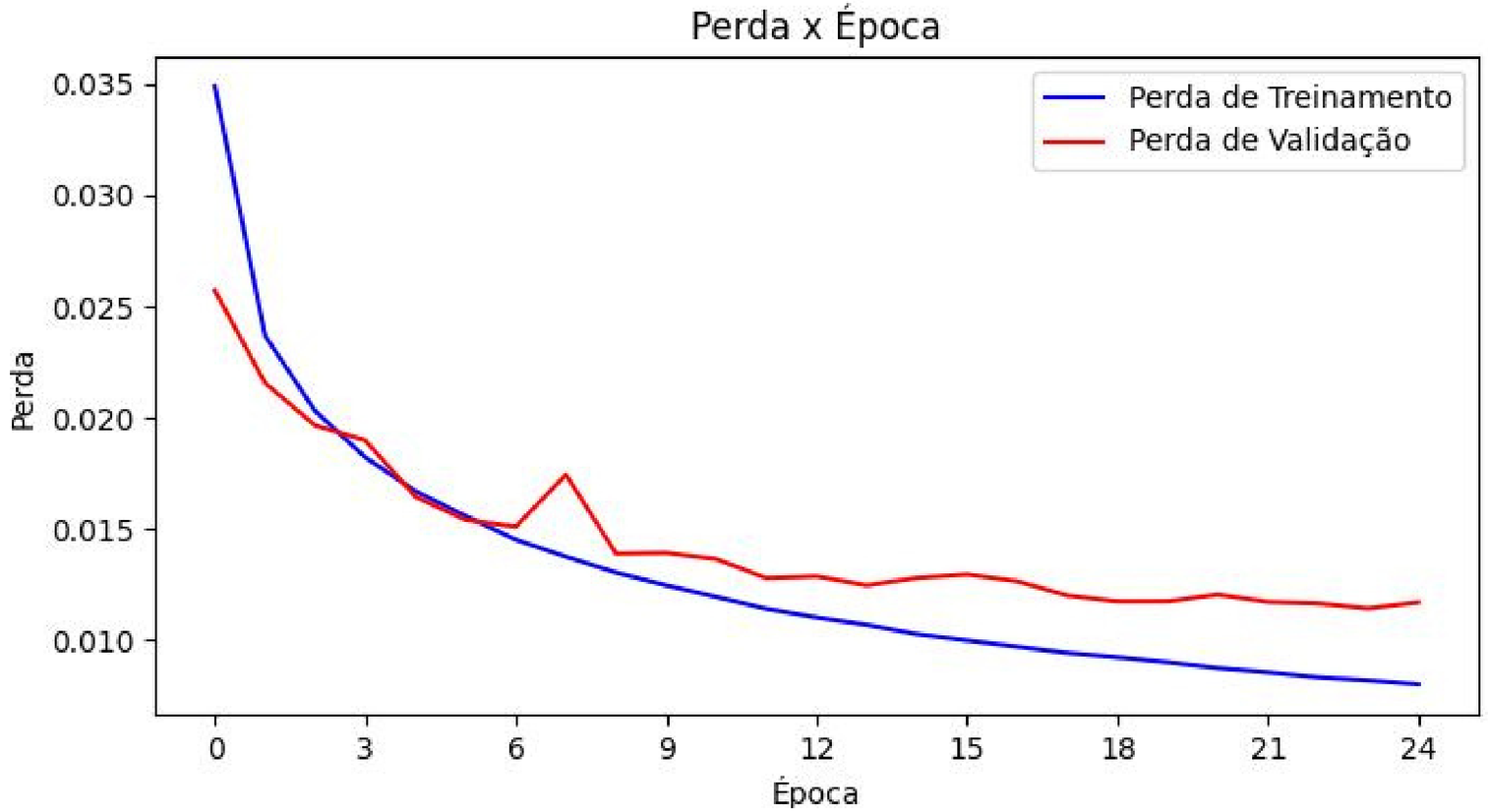
# TREINAMENTO E VALIDAÇÃO

Gráfico das acurácias durante as épocas



# TREINAMENTO E VALIDAÇÃO

Gráfico das perdas durante as épocas



# TESTE

## Matriz de confusão

| | | Valor Predito | Valor Predito |
|---|---|---|---|
| | | Positivo | Negativo |
| Valor Real | Positivo | 8649 | 976 |
| Valor Real | Negativo | 650 | 9725 |

## Métricas

$$Precisão = 93.01\%$$

$$Acurácia = 91.87\%$$

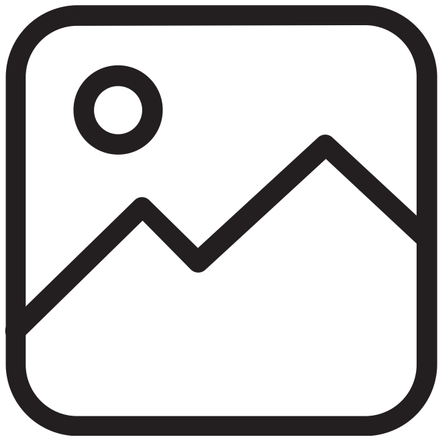
# COMPARAÇÃO

Comparação da CapsNet com as técnicas do artigo (WANG; ZARGHAMI; CUI, 2021)

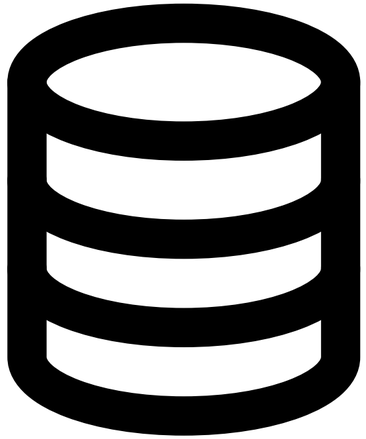
| Modelos | Acurácia | Precisão |
|---|---|---|
| CapsNet | 91.87 | 93.01 |
| Res-Net | 98.67 | 99.12 |
| Gram-Net | 98.71 | 98.98 |
| LBP-Net | 98.58 | 98.96 |

# CONCLUSÃO

# TRABALHOS FUTUROS



**Qualidade da imagem**



**Banco de dados diferente**

# REFERÊNCIAS


GOODFELLOW, I. J.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. GENERATIVE ADVERSARIAL NETWORKS. 2014.

WANG, Y.; ZARGHAMI, V.; CUI, S. FAKE FACE DETECTION USING LOCAL BINARY PATTERN AND ENSEMBLE MODELING. IN: 2021 IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP). [S.L.: S.N.], 2021. P. 3917–3921

SABOUR, S.; FROSST, N.; HINTON, G. E. DYNAMIC ROUTING BETWEEN CAPSULES. 2017.

# OBRIGADO