

# O uso de Máquina Ensinável aplicada para reconhecimento Pessoas e Objetos em Lugares Controlados

1<sup>st</sup> Alexandre de Araújo

*Doutorado em Telecomunicação*

*Inatel - Instituto Nacional de Telecomunicações*

Santa Rita do Sapucaí, Brasil

alexandre.araujo@dtel.inatel.br

**Abstract**—The implementation of neural networks for recognition is challenging due to the difficulties in programming them. For this study, we used Google's Teachable Machine, a user-friendly tool that employs the Convolutional Neural Network (CNN) method, starting with data analysis. Two scenarios with different databases were presented: the first was trained with 2 classes and 900 images, while the second with 6 classes and 180 images. Accuracy, precision, and sensitivity were assessed, all of which showed satisfactory values. However, the precision may decrease due to the lighting conditions of the presented images. It should be noted that Google's Teachable Machine can generate learning models with an accuracy rate of 100%.

**Index Terms**—component, formatting, style, styling, insert

## I. INTRODUÇÃO

Atualmente, o reconhecimento de pessoa ou então a identificação de objetos por meio de dados de vídeo, tem se tornado essencial na visão computacional [20]. Onde a classificação e a identificação pode ser um problema, devido a percepção humana. Assim o desenvolvimento de novas tecnologia contribui para diversas aplicações. Tais como, estação de metrô, aeroportos, indústria e ambientes que necessita de monitoramento permanente ou então em setores de fluxo de mercadorias. Nestes ambientes, apresenta inúmeras câmeras de vídeo instaladas, cobrindo todos pontos de interesse que necessita de monitoramento constante ou que apresenta acontecimentos inesperáveis. Diante disto, é necessária reconhecer as ações e distinguir conforme padrão esperado.

Existe diversos sistemas que foi desenvolvido no passado, para reconhecimento de ações em vídeo. Onde inclui ações descomplicadas com (andar, correr e pular), entretanto, em as ações mais embaralhadas que podem envolver a interatividade de um grande número de pessoas ou de objetos [18]. Esses sistemas aplicam métodos de aprendizado de máquina, no qual são constituídos por redes neurais artificiais [11]. No entanto, para treinar estes sistemas a fim que consigam identificar as ações com qualidades e sem interferência, necessita de grandes bancos de dados de vídeo ou imagens.

O treinamento de um sistema constitui em obter um conjunto de frames de um vídeo específico, assim desenvolvendo uma sequencias de frames, o qual é retido resultados deles e enviando-os para entrada de um classificador. No decorrer

do treinamento, sabemos quais ações foram apresentadas ao classificador, de modo que conseguimos ajustar o classificar toda vez que estiver errado, assim aumentando o índice de acerto no futuro [7].

Apesar de existir diversos métodos, as ações de reconhecimento com boa precisão podem ser limitadas, onde tentando melhorar estas ações modernas de reconhecimento o qual veio fracassar devido a várias restrições. Onde à baixa qualidade do vídeo, ruídos, fundo confuso e similaridades nas ações [3]. Além disso, a verificação visual entre as imagens, poder ser tornar uma tarefa complicada, devido as semelhanças associado a imagem.

Em artigo publicado a relacionado ao reconhecimento de imagens, no qual, apresentaram método baseado em combinação de redes neurais convolucionais. Conforme apresentado por [12], o modelo proposto é baseado na CNN VGG19, onde foram criadas duas redes de treinamento a VGG19v e VGG19c, assim treinadas usando imagens similares. Ambas as redes foram avaliadas usando a classificação de media normalizada em um conjunto de teste. No qual, obtiveram a média normalizada 0,047 superando outras médias publicadas anteriormente. Estas combinações foram usadas para comprar imagens de marcas registras.

Em outro estudo, a técnica de aprendizado de maquinas com a implementação da inteligência artificial, no qual possibilita as maquinas analisar as informações, compreender padrões e executar o que foi aprendido [2]. Diante disto, o estudo [1] mostrou a implementação e aprendizagem de máquina para criação de aplicativo. Onde usando Teachable Machine da Google obteve resultados favoráveis para determinar vários parâmetros CNN.

## II. DESCRIÇÃO DO PROBLEMA

Portanto este artigo, o autor optou em usar o aplicativo Teachable Machine da Google, devido à facilidade de uso e rapidez no processamento, onde coletando imagens e a assim gerando um banco de dados satisfatório. Assim, gerando padrões de identificação e combinações para classificar imagens. Diante disto, as principais contribuições deste trabalho são apresentar ao usuário inexperiente as noções de básica de

implementação de maquina ensináveis utilizando ferramentas de código aberto, e aplicando o modelo para classificar imagens de pessoas e objetos em lugares controlados.

Será apresentado dois senários utilizando o modelo de aprendizagem por maquinas ensináveis, afim de classificar imagem conforme necessidade. No primeiro senário, estaremos identificando pessoas e classificando conforme padrão definido, ou seja, permitindo o acesso ou não de pessoas em um determinado ambiente, no qual, necessita o uso de máscara com equipamento de proteção.

Já para segundo senário, estaremos identificando objetos perigosos, ou seja, que possa causar algum maleficio as pessoas um determinado ambiente.

### III. MÉTODOS E CONCEITOS

O método baseado em CNNs profundas, no qual as redes neurais artificiais várias camadas, onde os pesos dos neurônios dentro de uma camada são compartilhados por vários mecanismos [17]. Essa arquitetura da rede permite que ela aprenda filtros convolucionais em diferentes escalas e níveis abstração.

No nosso trabalho, utilizamos o modelo Keras que é CNN aprendizado profunda, executada na plataforma TensorFlow, onde são estruturas os dados em camadas, oferecendo uma abstração essenciais e focado no aprendizado profundo [15].

- **Aprendizado de máquina**

O aprendizado de máquina é um ramo da inteligência artificial, no qual utiliza programa ou algoritmo, afim que possa aprender com dados apresentado, identificando padrões e tomar decisões, tornando o aprendizado instantâneo [10]. O aprendizado de máquina não é um modelo ligado de Inteligência Artificial que geralmente é mais amplo, por exemplo, pode pensar como o cérebro humano de forma independente [4].

Além disso, o aprendizado de máquina tem crescido no campo de ciência, devido a facilidade da interação de aprendizado tornando se útil para trabalho humano, pois sua capacidade de previsão é alta e com eficiência [16]. Assim apresentando habilidade muito útil onde o aprimoramento da infraestrutura de produção é um importante surgimento da indústria 4.0.

- **Rede Neural Convolucional**

Rede Neural Convolucional é um algoritmo de aprendizado profundo que é usado frequentemente para capturar imagens, som, vídeo e outros [14]. No entanto a CNN pode ser dividida em três itens principais, onde a camada principal recebe as características locais dos dados extraído e as camadas subsequentes são compostas por diferentes mapas de características fazendo com que seja possível lidar com variações e por final a camada de subamostragem que produz uma saída classificação. A Figura 1 representa a arquitetura simplificada de uma imagem nas camadas na rede neural convolucional [19].

- **Taxa de Aprendizado**

A taxa de aprendizado é parâmetro usado no processamento de treinamento da rede neural, indicando a que ritmo os pesos são atualizados. Assim, a decisão do valor não pode ser

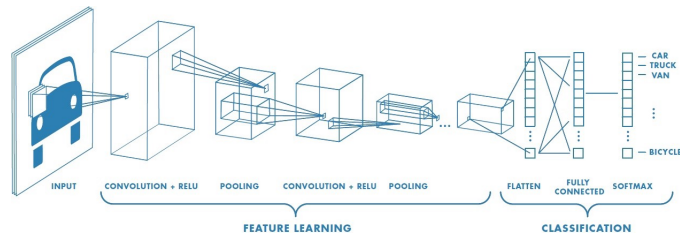


Fig. 1. Arquitetura Rede Neural Convolucional.

aleatória, pois valores muito altos podem provocar oscilações no treinamento e valores muito pequenos podem levar o aprendizado muito lento, o qual, pode interfere na convergência do processo de aprendizado e também influenciar na precisão do modelo [6].

- **Épocas**

Época é um dos parâmetros da Rede Neural, no qual defini a quantidade de vezes que serão exibido os dados para rede neural. Assim, até concluir o processamento ou alcançar algum critério de parada e retornar para estágio inicial.

- **Tamanho Lote**

Tamanho do lote é um termo usado em aprendizado de máquina e refere-se ao número de exemplos de treinamento usados em uma iteração. No qual é dividido os dados em conjuntos de determinados tamanhos.

- **Teachable Machine Google**

Teachable Machine Google é um aplicativo Web que permite aprender de forma simples o conceito de inteligência artificial, onde utiliza a câmera do computador ou dados fornecidos pelo usuário, possibilitando reconhecer dados gerados, de tal modo classificando e obtendo respostas pré-determinadas [13].

O Teachable Machine é uma ferramenta que torna a criação de modelos de aprendizagem de máquina mais fácil e rápido, onde pessoas com pouco conhecimento em programação, podem criar modelos e treina-lo para classificar imagens, voz e poses. E tudo isso, sem a necessidade de criação de códigos [9].

- **Redimensionar Imagem**

O redimensionamento das imagens para treinamento foi um critério, na qual, ajudaria no processamento e também para obter um padrão. Para redimensionamento optamos pelo tamanho 160 pixels x 160 pixels para imagens de treinamento. Conforme Figura 2.

### IV. SOLUÇÃO PROPOSTA UTILIZANDO RNA

A análise de dados, é um processo de levantamento de dados realizado para obter informações que visa determinar as funções, objetivos e limitações a serem desenvolvidas. Diante disto, o uso de ferramentas que executa a combinação de dados é favorável. Para classificar nosso modelo foram usados dois tipos de conjuntos de dados.

Para o primeiro modelo de reconhecimento de faces, utilizaram 900 imagens, onde 450 apresentavam o uso de



Fig. 2. Redimensionamento das imagens.

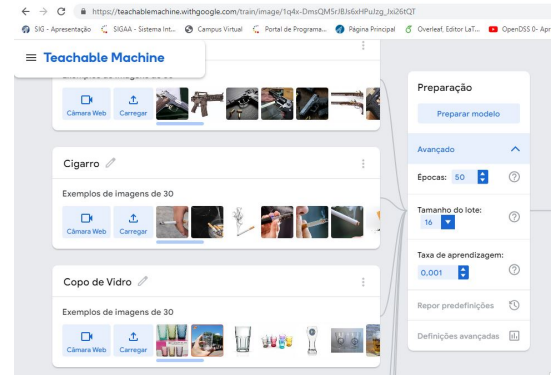


Fig. 4. Código do programa.

máscaras e as demais imagens não apresentava, todas as imagem foram coletados do banco de dados [8]. O conjunto de dados inclui imagens de homens e mulheres, onde apresenta diferentes enquadramentos, variação de iluminação, imagens sem foco, imagens tremidas e também vários tipos de máscaras.

No entanto, para segundo modelo de identificação de objetos, utilizaram 180 imagens, sendo divididas em seis classe diferentes, as imagens coletadas aleatórias na internet conforme parâmetros. Apresentado diversas característica com dimensões, cor, material e outros. Conforme Figura 3.

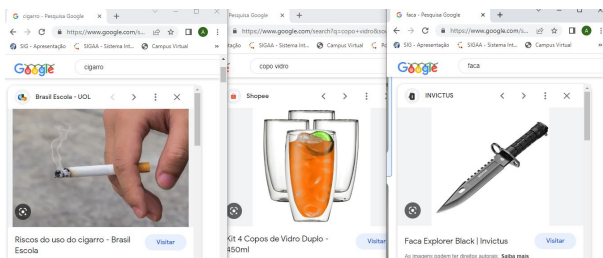


Fig. 3. Código do programa.

```
# Bibliotecas
from keras.models import load_model
from PIL import Image, ImageOps #install pillow instead of Pli
import numpy as np
from google.colab.patches import cv2_imshow
import matplotlib.pyplot as plt

# Carrega o modelo
model = load_model('keras_model.h5', compile=False)

class_names = open('labels.txt', 'r').readlines() # Carregar os rótulos

# Crie a matriz da forma certa para alimentar o modelo keras
# O 'comprimento' ou número de imagens que você pode colocar no array é
# Determinado pela primeira posição na tupla de forma, neste caso 1..
data = np.ndarray(shape=(1, 224, 224, 3), dtype=np.float32)

image = Image.open('01').convert('RGB') # Substitua isso pelo caminho para sua imagem

# Redimensionar a imagem para 224x224 com a mesma estratégia do TMI:
# Redimensionamos a imagem para pelo menos 224x224 e recortando do centro
size = (224, 224)
image = ImageOps.fit(image, size, Image.LANCZOS)

image_array = np.asarray(image) # Transforma a imagem em um array numpy
normalized_image_array = (image_array.astype(np.float32) / 127.5) - 1 # Normaliza a imagem
image_array = np.asarray(image) # Transforma a imagem em um array numpy
data[0] = normalized_image_array # Carrega a imagem no array

model.layers[0] # acessa a primeira camada:

# percorre as camadas uma a uma e imprime o nome de cada camada no console:
for layer in model.layers:
    print(layer.name)

# Executa a inferência
prediction = model.predict(data)
index = np.argmax(prediction)
class_name = class_names[index]
confidence_score = prediction[0][index]

print("Class:", class_name, end="")
print("Confidence score:", confidence_score)
img = cv2.imread("01")
cv2.imshow(img)
```

Fig. 5. Código do programa.

Mediante aos dados coletados, no qual permitirá cria os modelos de aprendizado de máquina para treinamento, assim utilizamos o Teachable Machine da Google para processar os dados e treinar a rede neural. Primeiramente acessamos a Web e criamos um modelo no Teachable Machine da Google, na qual permitirá a criação de um novo projeto, assim criando as classes e também inserindo as imagens coletadas anteriormente. Conforme Figura 4

Entretanto, Teachable Machine da Google apresenta algumas informações em gráficos que podem ser visualizada no site, como gráfico de precisão por época, perda por época e matriz de confusão. Abaixo estão os resultados obtidos com treinamento dos dois senários [5].

- Perda por época

Na 6 está apresentado o gráfico perda por época do Modelos A e a 7 está apresentando o gráfico perda por época do Modelo B.

Perda por época

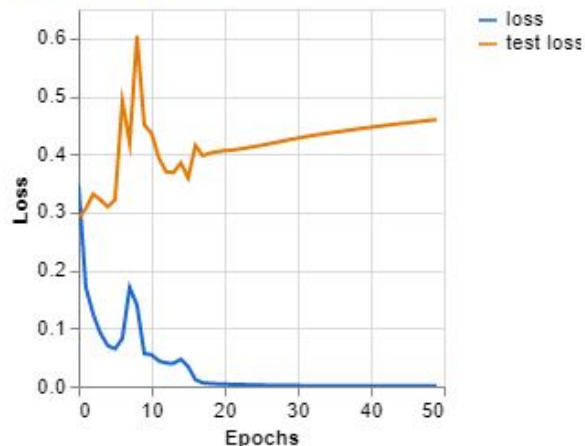


Fig. 6. Perda por época modelo A.

Perda por época

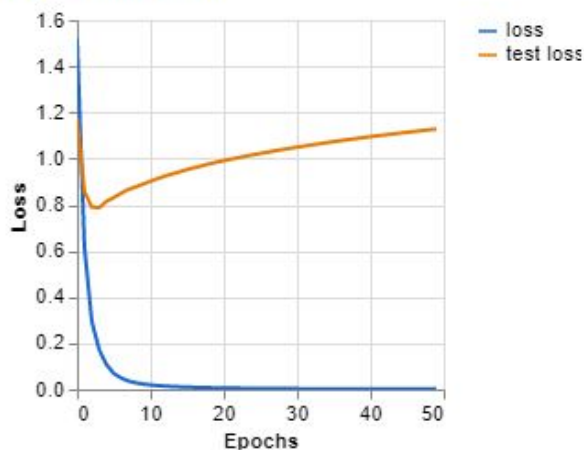


Fig. 7. Perda por época modelo B.

#### • Precisão por época

Na 8 está apresentado o gráfico precisão por época do Modelos A e a 9 está apresentando o gráfico precisão por época do Modelo B.

Precisão por época

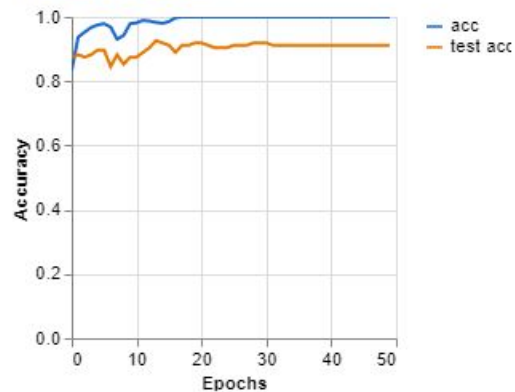


Fig. 8. Precisão por época modelo A.

Precisão por época

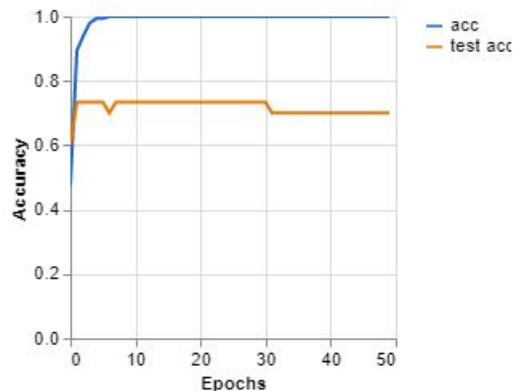


Fig. 9. Precisão por época modelo B.

#### • Matriz de confusão

Na Figura 10 está apresentado a matriz de confusão do Modelos A e a Figura 11 está apresentando a matriz de confusão do Modelo B.

Matriz de confusão

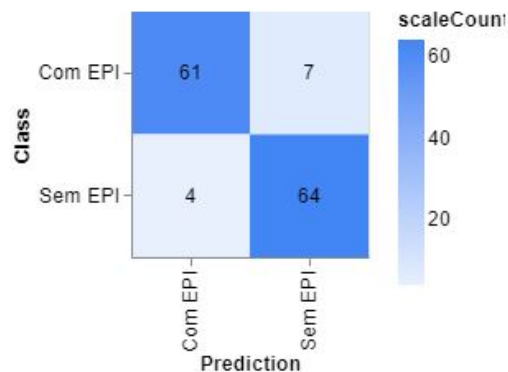


Fig. 10. Matriz de confusão modelo A.

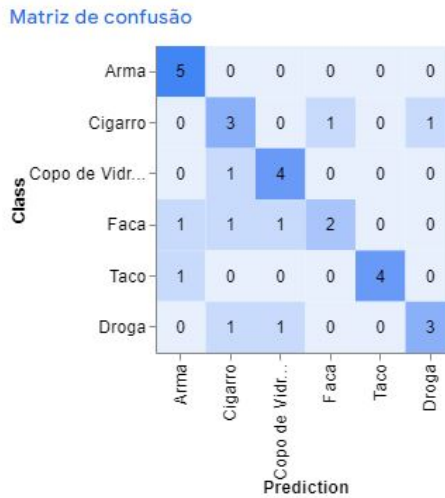


Fig. 11. Matriz de confusão modelo B.

## V. RESULTADOS

Na avaliação dos resultados após o treinamento da máquina ensinável, na qual apresentamos ao sistema outras imagens distintas e aleatória. Assim o sistema identificou e classificou a imagem com forme aprendizagem. Para modelo A, foi apresentado quatro imagens, sendo duas com máscaras e duas sem máscara, assim apresentou 7 camadas sendo 5 profundas e resultado favoráveis conforme apresentado na tabela I.

TABLE I  
RESULTADOS POR CLASSE DE PONTUAÇÃO DE CONFIANÇA E A PRECISÃO DO MODELO A

Classe	Pontuação de Confiança	Precisão
Com Máscara	100%	0.91
Sem Máscara	99%	0.91

No entanto, para o segundo modelo foi apresentado doze imagens distintas e aleatórias, no qual apresentou 3 camadas sendo 1 profunda e uma boa pontuação de confiança, porem em algumas classes o nível de precisão foi baixo, conforme tabela II.

TABLE II  
RESULTADOS POR CLASSE DE PONTUAÇÃO DE CONFIANÇA E A PRECISÃO DO MODELO B

Classe	Pontuação de Confiança	Precisão
Arma	100%	1.00
Cigarro	99%	0.60
C. Vidro	95%	0.80
Droga	86%	0.60
Faca	98%	0.40
Taco	82%	0.80

## VI. CONCLUSÃO

Neste trabalho abordamos a dificuldade de programação de rede neural devido complexidade. Assim propusemos usar uma plataforma livre de máquina de ensinável. Além disto

apresentamos o modo de utilizar a ferramenta Teachable Machine da Google, onde criamos dois modelos distintos para treinamento, com isto as pessoas poderão implementar o aprendizado de máquina e determinar parâmetros CNN para máxima exatidão, precisão e resultado de sensibilidade sem altera todo modelo.

De acordo com teste realizado, a Teachable Machine da Google apresenta uma capacidade alta e confiável em seus resultados, assim sendo possível apresentar valores de 100% de exatidão, sensibilidade e precisão. Além disto a interface Teachable Machine possibilita interagir com outros dispositivos.

## VII. TRABALHOS FUTUROS

Baseado nos desafios e avanços recentes no campo do reconhecimento de ações e objetos em vídeos, várias direções de trabalho futuros podem ser exploradas. Primeiramente, o desenvolvimento de modelos híbridos e multimodais (Sugestão 1) se mostra promissor. A combinação de diferentes arquiteturas de redes neurais, como CNNs, RNNs e Transformers, pode melhorar a precisão no reconhecimento de ações complexas, especialmente em cenários que envolvem interações humanas e objetos.

Além disso, a melhoria da qualidade de vídeos e a robustez dos modelos (Sugestão 2) são essenciais para aumentar a eficácia do reconhecimento. Métodos de pré-processamento de vídeos para melhorar a qualidade da imagem e reduzir ruídos, assim como o uso de Redes Adversárias Generativas (GANs) para criar dados de treinamento mais variados e realistas, podem significar um avanço significativo na precisão dos modelos.

Por fim, focar em aplicações específicas em ambientes críticos (Sugestão 3) como estações de metrô, aeroportos e indústrias, onde a precisão e rapidez no reconhecimento de ações são essenciais, pode trazer benefícios diretos para a segurança e eficiência desses locais. Desenvolver soluções customizadas para esses ambientes críticos garantirá que os sistemas de monitoramento sejam adaptados às necessidades particulares de cada contexto, melhorando assim a segurança e a gestão operacional.

## REFERENCES

- [1] Agustian, D., Pertama, P. P. G. P., Crisnapati, P. N., Novayanti, P. D., 2021. Implementation of machine learning using google;s teachable machine based on android. In: 2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS). pp. 1–7.
- [2] Chang, M.-C. L., Huang, Y.-H., Lin, W.-C., Sun, S.-W., 2021. Digital fabrication: Machine learning-based immersive exper- encing for the virtual space in a future museum. In: 2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). pp. 102–105.
- [3] Fan, D., He, W., Li, H., Liu, W., Shi, W., Jiang, Z., 2020. Facial expression recognition based on ensemble convolutional neural network. In: 2020 International Conference on Virtual Reality and Visualization (ICVRV). pp. 145–148.
- [4] Fischer, K., 2012. Human tutors intuitively reduce complexity in socially guided embodied grammar learning. In: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. pp. 871–877.



- [5] Khalid, F., Abdullah, A. H., Abdullah, L. N., 2022. Smartflora mobile flower recognition application using machine learning tools. In: 2022 IEEE 18th International Colloquium on Signal Processing e Applications (CSPA). pp. 204–209.
- [6] Lakshmanarao, A., Ushanag, S., Leela, B. S., 2021. Ad pre- diction using click through rate and machine learning with reinforcement learning. In: 2021 Fourth International Confer- ence on Electrical, Computer and Communication Technologies (ICECCT). pp. 1–5.
- [7] Liu, W., Li, W., Sun, L., Zhang, L., Chen, P., 2017. Finger vein recognition based on deep learning. In: 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA). pp. 205–210.
- [8] MACEDO, S., 2020. Deep learning serie: de- tectando faces em video com deep learn- ing. In: <https://github.com/sandeco/CanalSandeco/tree/master/Deep%20Learning%20s%C3%A9rie/%2316%20%20Detectando%20Faces%20em%20v%C3%ADdeo%20com%20Deep%20Learning>.
- [9] Miyamoto, Y., 2022. An approach to classify thumbnail images on video sites by the number of accesses. In: 2022 37th Inter- national Technical Conference on Circuits/Systems, Computers and Communications (ITC- CSCC). pp. 194–197.
- [10] P. B. K., Sinha, P., Maharana, M. K., Jena, C., Pavan Kumar, A. V., Akkenaguntla, K., 2021. Power system fault detection using image processing and pattern recognition. In: 2021 IEEE 2nd International Conference on Applied Electromagnetics, Sig- nal Processing, & Com- munication (AESPC). pp. 1–5
- [11] Peres, A. A., Vieira, S. M., Pinto, J. R. C., 2018. Hybrid neural models for automatic handwritten digits recognition. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–8.
- [12] Perez, C. A., Estevez, P. A., Galdames, F. J., Schulz, D. A., Perez, J. P., Bastias, D., Vilar, D. R., 2018. Trademark image retrieval using a combination of deep convolutional neural networks. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–7.
- [13] Salim, S., Jamil, M. M. A., Ambar, R., Roslan, R., Kamar- dan, M. G., 2022. Sign language digit detection with mediapipe and machine learning algorithm. In: 2022 IEEE 12th Interna- tional Conference on Control System, Computing and Engineer- ing (ICCSCE). pp. 180–184.
- [14] Samudre, P., Shende, P., Jaiswal, V., 2019. Optimizing perfor- mance of convolutional neural network using computing tech- nique. In: 2019 IEEE 5th International Conference for Conver- gence in Technology (I2CT). pp. 1–4.
- [15] Shivaanivarsha, N., Lakshmidhevi, P. B., Josy, J. T., 2022. A convnet based real-time detection and interpretation of bovine disorders. In: 2022 International Conference on Communica- tion, Computing and Internet of Things (IC3IoT). pp. 1–6.
- [16] Sistu, S. K. H., Chadaram, P. S. S., Varma, A. K., Gurala, C., 2022. A framework for iot-enabled secure access using face- mask detection. In: 2022 International Conference on Break- through in Heuristics And Reciprocity of Advanced Technolo- gies (BHARAT). pp. 18–20.
- [17] Subramanian, M., Shanmugavadeivel, K., Muthusamy, S., Bajaj, M., Rubanenko, O., Danylchenko, D., 2022. Development of a surveillance system to detect forest fire and smoke using deep neural networks. In: 2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek). pp. 1–6.
- [18] T. S., S. B., M. H. S., S. V. V., M. P., 2022. Suspicious activity de- tecting camera. In: 2022 International Conference on Com- munication, Computing and Internet of Things (IC3IoT). pp. 1–4.
- [19] Utama, I. B. K. Y., Faqih, A., Kusumoputro, B., 2019. Three mixture of odor classification using convolutional neural net- work. In: 2019 16th International Conference on Quality in Re- search (QIR): International Symposium on Electrical and Com- puter Engineering. pp. 1–4.
- [20] Zhdanov, P., Khan, A., Ramirez Rivera, A., Khattak, A. M., 2018. Improving human action recognition through hierarchical neural network classifiers. In: 2018 International Joint Confer- ence on Neural Networks (IJCNN). pp. 1–7.6