Comparing YOLO and Detectron2 models for automatic extracting patients information from leprosy assessment form

Anthony Militão[1], Hilson Gomes Vilar de Andrade[1,2], Kayo H. Monteiro[1], Elisson da Silva Rocha[1], Patricia Takako Endo[1]

[1]Programa de Pós-Graduação em Engenharia da Computação (PPGEC), Universidade de Pernambuco (UPE), Caruaru – PE – Brazil

[2]Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE) Recife – PE – Brazil

aamm2@ecomp.poli.br, hilsonvilar@recife.ifpe.edu.br,

{kayo.henrique, elisson.rocha, patricia.endo}@upe.br

Abstract. *Leprosy, caused by Mycobacterium leprae, remains a global challenge, requiring strategies to achieve disease elimination by 2030. In Brazil, the Simplified Neurological Assessment (from Portuguese Avaliação Neurológica Simplificada, ANS) is mandatory for suspected cases; however, the form is still manually fulfilled, which limits the use of data. This study evaluates computer vision models (YOLOv8x, YOLO11x, Faster R-CNN) for detecting hand and foot sensitivity regions from ANS forms. All models were evaluated based on precision, recall, mean average precision (mAP) and confusion matrix. YOLO variants achieved over 94% precision and 84% recall across all classes. Automating ANS data extraction can facilitate the creation of structured datasets, enhancing disease monitoring and enabling the train of predictive models.*

1. Introduction

The 2030 Agenda of the World Health Organization (WHO) prioritizes the fight against Neglected Tropical Diseases (NTDs) under the Sustainable Development Goal 3 (SDG 3), which aims to ensure healthy lives and promote well-being for all [WHO 2021]. Among these diseases, leprosy remains a critical world challenge, requiring strategies for early diagnosis and epidemiological surveillance, which are essential for the global elimination of the disease by 2030 [WHO 2024].

Caused by *Mycobacterium leprae*, leprosy affects the skin and peripheral nerves, potentially leading to loss of sensation, muscle weakness, paralysis, skin lesions, and, in advanced stages, deformities such as claw hand, foot drop, spontaneous amputations, and ocular complications that increase the risk of blindness [WHO 2017]. Considering the high potential of leprosy to cause irreversible physical disabilities [Brasil. Ministério da Saúde 2022], early detection and monitoring are essential to prevent severe impairments.

Only in 2023, 182,815 new cases of leprosy were reported globally, with Brazil accounting for 22,773 cases—representing nearly 92% of cases in the Americas [WHO 2024]. The monitoring the progression of leprosy and supporting its diagnosis still largely depends on clinical evaluations. In this context, in Brazil, the Simplified

Neurological Assessment (from Portuguese *Avaliação Neurológica Simplificada* - ANS) plays a crucial role in standardizing the assessment of sensation, muscle strength, and peripheral nerve thickening, enabling early identification of neurological impairments and the classification of physical disability grades[Nóbrega et al. 2024].

According to the guidelines of the Brazilian Ministry of Health, the ANS is a mandatory examination that must be performed at the time of diagnosis, every three months during Uniform Multidrug Therapy (U-MDT), and at the final supervised dose. In a 12-month treatment regimen, this recommendation implies conducting at least four assessments per year. Consequently, each completed form becomes a valuable source of data for monitoring disease progression and extracting crucial information for clinical practice [Ministério da Saúde 2022].

1.1. Motivation

The ANS form is a standardized neurological assessment tool used to evaluate basic neurological functions in patients. Currently, all ANS forms are completed manually, and there is no national digital system to store and manage the data collected. This lack of digitization hinders data analysis, decision-making, and the potential application of AI for deeper insights and automation.

Textual data can be extracted using conventional Optical Character Recognition (OCR) techniques. But, as this study explores non-textual elements in the form, especially the sensory evaluations performed with Semmes-Weinstein monofilaments [Leite et al. 2010] on the patients' hands and feet, we need other alternatives to cover image. In this case, developing methods to accurately detect Regions Of Interest (ROIs) related to sensitivity assessments from the ANS form is essential to complement existing techniques for extracting textual and tabular data.

1.2. Objectives

Although OCR-based methods have significantly advanced in extracting textual data from medical records [Hsu et al. 2021, Leite-Moreira et al. 2022], they often overlook non-textual elements—such as the color-coded monofilaments used in ANS forms for sensitivity assessments. This study aims to develop an automated pipeline for extracting image-based data from ANS forms to provide data for analysis on the monitoring of leprosy patients, focusing on the sensitivity assessment sections of the hands and feet. For that, we compare the performance of YOLOv8x [1], YOLO11x [2] and Faster R-CNN (implemented in Detectron2) [3]. These models are used for Detection of ROIs and the sensitivity markers (through color and symbol analysis).

2. Material and methods

2.1. ANS form

The ANS form [Ministério da Saúde - Secretaria de Vigilância em Saúde 2021] exists in two versions. In the older version, the sections for inspection and sensitivity evaluation

[1] https://yolov8.com/

[2] https://yolo11.com/

[3] https://ai.meta.com/tools/detectron2/

include only three pairs of hands and feet, with larger outlines. In contrast, the newer version provides four pairs of each, offering an additional inspection. Despite these differences, both versions include textual records covering complaints related to the nose (dryness, wounds, septal perforation) and the eyes (reduced corneal sensitivity, decreased muscle strength of the upper eyelids, and corneal opacity). They also document complaints in the upper and lower limbs, as well as the palpation and strength evaluation of the radial, ulnar, and median nerves in the upper limbs, and the fibular and tibial nerves in the lower limbs.

The sensitivity assessment is related to hands and feet of patients. Each form consisted of two pages, with the first page featuring the hand sensitivity assessment at the bottom, while the second page displayed the foot sensitivity assessment at the top; and other information that are not the focus of this work. Figure 1 shows both hand and foot sensitivity images present in the ANS form.

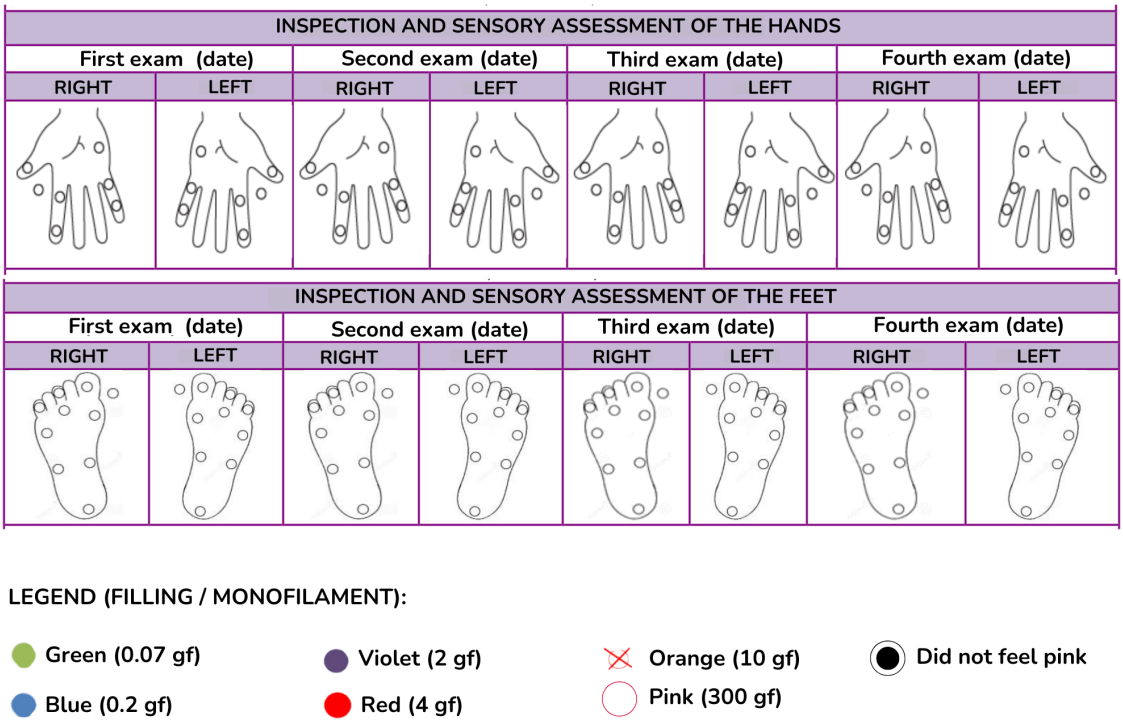


Figura 1. Inspection and sensory assessment of hands and feet present in the ANS form

To measure the sensitivity of patients, the health professional uses calibrated monofilaments, named esthesiometer, in which each monofilament is identified by a color corresponding to a specific force. The nylon monofilaments are pressed against the patient's hands and feet, and when they bend, it confirms that the corresponding force, measured in gram-force (gf), has been applied. Under normal conditions, a 2.0 gf stimulus – represented by the violet or purple color, should be perceived, indicating the integrity of the nerve function. When the patient presents a nerve degeneration, i.e., they do not fell the monofilament in the specific point of hands or feet, the stimuli of increasing intensity

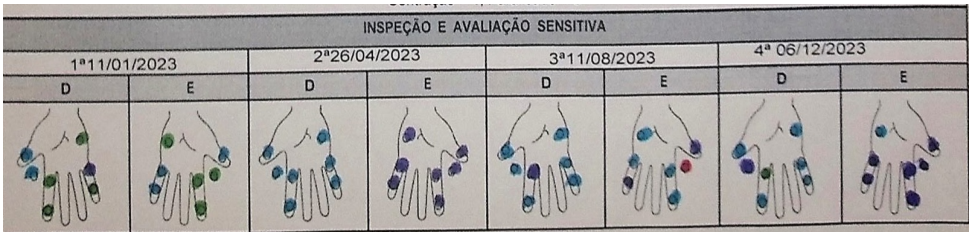
must be applied, such as 4.0 gf (red), 10.0 gf (orange, marked with an “X”), and 300 gf (pink or filled in with black if not perceived). This step-by-step assessment helps identify the extent of neurological impairment. The graphical representation of these results, through filled or marked circles, facilitates immediate clinical interpretation, allowing a rapid correlation between the sensory pattern and the stage of nerve injury.

2.2. Dataset preprocessing

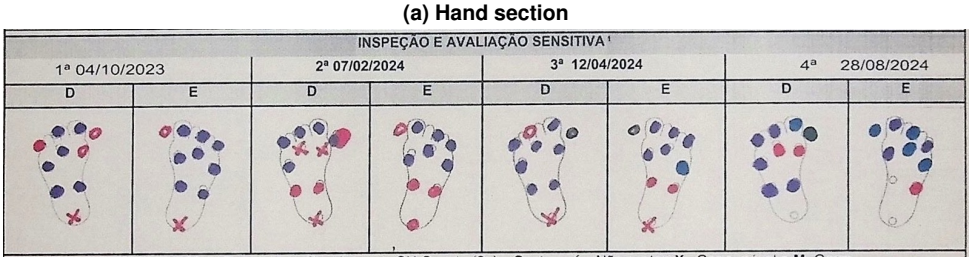
This study utilized 165 ANS clinical records collected at a Hospital Otávio de Freitas, Recife, Brazil. All records were fully digitized, resulting in images containing both pages, in JPG format. The digitization process maintained the integrity of the records, without loss of information or need to exclude low-quality images.

2.2.1. Image segmentation (cropping)

To extract the regions of interest from each form, a segmentation algorithm was developed that uses minimum and maximum coordinates to delimit the areas of interest. In this process, the limits for the hands are defined as (x_min=0, y_min=1360, x_max=1363, y_max=1920), corresponding to the bottom part of the first page, while the limits for the feet are defined as (x_min=1366, y_min=544, x_max=2676, y_max=1044), corresponding to the top part of the second page. Thus, for each ANS form, two images are generated: one containing the hand sensitivity assessment (Figure 2a) area and the other containing the foot sensitivity (Figure 2b) assessment area.



(a) Hand section



(b) Foot section

Figura 2. Example of a cropped image of the (a) hand and (b) foot sensitivity inspection section

During the digitization process, the original record images exhibited three distinct resolutions (2,724 x 1,926, 2,806 x 1,984, and 3,507 x 2,480 pixels). After the automated and manual cropping of the regions of interest, the resulting images were saved while maintaining the dimensions of the cropped areas corresponding to the hand and foot sensitivity regions. Additionally, images containing blank sensitivity assessments (i.e., forms where no assessment was recorded, either of the hands or feet) were removed from the dataset.

2.2.2. Image standardization (resizing)

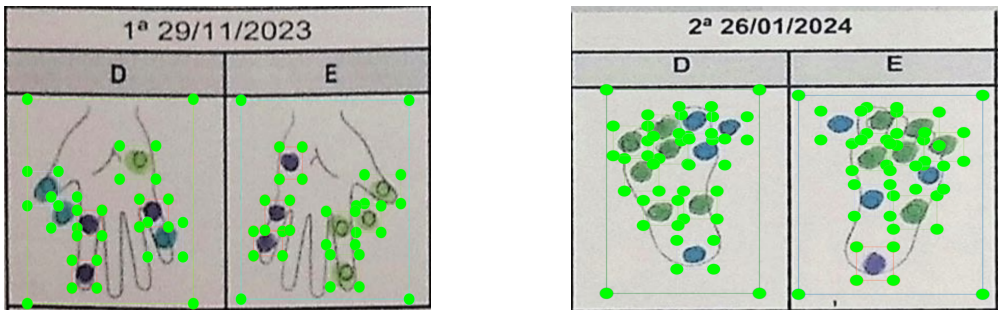
For models training, the images were resized to 640 pixels by both architectures, ensuring uniformity in the input. This resizing was performed while preserving the proportions of the images and without distortions, which ensured consistency in the data for the training and inference stages. In this way, the differences in the original resolutions of the digitized images did not impact the model, as standardization occurred during the internal processing by the detection networks.

After this step, the resulting images were stored individually, each image has its respective section for sensitivity assessment of the hands and feet. At the end of the preprocessing stage, a total of 262 images were generated, comprising 165 hand images and 97 foot images, all resized to 640x640 pixels. These images were then splitted into three subsets: training, validation, and test, with the following proportions 70%, 15% e 15%, respectively.

2.2.3. Image annotation (labeling)

Using labelImg [4], an open-source tool for image annotation, each image was manually annotated following a unified labeling scheme employing bounding boxes. Each object is represented by a line in an annotation file. Each line contains the class identifier and the bounding box's normalized coordinates—specifically, the center (x, y), width, and height relative to the image dimensions. The image is divided into an S×S grid, with each cell responsible for predicting B bounding boxes along with their confidence scores. During training, a loss function is optimized to minimize localization and classification errors [Redmon et al. 2016].

Using a Python script, the bounding box annotations from each image were automatically converted to the COCO JSON format, the annotation model used by Detectron2. This process involves extracting, from each annotation file in the YOLO format, the class identifier and the normalized coordinates of the bounding boxes, restructuring this information into a single JSON file containing keys such as "images", "annotations", and "categories". After the conversion, all the boxes were displayed on each image to verify the integrity of the annotations, as shown in Figure 3.



(a) Labeled hands **(b) Labeled feet**

Figura 3. Example of labeled (a) hands and (b) feet

[4] https://github.com/tzutalin/labelImg

After preprocessing, the final dataset comprised 10,220 instances distributed across 11 distinct classes—including categories such as colors, hand, and foot, as detailed in Table 1. The dataset was then partitioned into training, validation and test sets.

Classes	Train	Validation	Test
All	7448	1520	1252
Green	1236	341	205
Blue	1146	149	169
Purple	1179	175	250
Red	387	68	111
Orange	362	98	59
Pink	656	224	126
Did not feel pink	1658	299	194
Right Hand	261	49	44
Left Hand	262	49	44
Right Foot	152	34	25
Left Foot	149	34	25

Tabela 1. Instances per class of the dataset

2.3. Models and Parameters

The YOLO family is based on a unified architecture that performs object detection in a single pass through the network, dividing the process into three main stages: a backbone that extracts relevant features from the image, a neck that aggregates information at multiple scales, and a decoupled head that separates regression (localization) from object classification. This single-shot structure allows the model to process the entire image at once, optimizing inference time and facilitating real-time applications [Redmon et al. 2016].

In the case of the YOLOv8x and YOLO11x variants, both maintain this fundamental structure; however, YOLO11x presents additional refinements. While YOLOv8x already uses C2f-type modules in the backbone to improve gradient flow and employs multi-scale integration through the neck, YOLO11x enhances this feature extraction with adjustments in the backbone architecture and the fusion of information from different resolutions. Furthermore, the decoupled head of YOLOv8x, which uses loss functions such as CIoU and focal techniques to align localization and classification predictions—is refined in YOLO11x through a more robust task alignment score[5].

Both models were trained with the same configuration, but using different GPUs: YOLOv8x was trained on an NVIDIA Tesla T4 [6], while YOLO11x was trained on an NVIDIA L4 [7]. Was used Ultralytics 8.3.75, Python 3.11.11, and Torch 2.5.1+cu124, with images resized to 640×640 pixels, 100 training epochs, and a batch size of 16. The loss function weights were set to 0.05 for the localization loss (box loss) and 0.5 for the classification loss (cls loss), allowing that the observed differences stem from architectural variations rather than training parameters.

[5] https://docs.ultralytics/
[6] https://www.nvidia.com/en-us/data-center/tesla-t4/
[7] https://www.nvidia.com/en-us/data-center/l4/

Regarding the structure, YOLOv8x initially features 365 layers, 68 million parameters, and a computational cost of 258.2 GFLOPs, which is reduced to 257.4 GFLOPs and 112 layers after fusion. In contrast, YOLO11x has a deeper architecture, with 631 layers in its initial training structure, showing 195.5 GFLOPs before fusion; after fusion, the model is optimized to 190 layers and 194.5 GFLOPs. These metrics indicate that, although YOLO11x starts with a more complex structure, the implemented optimizations significantly reduce the computational complexity, as reflected by the lower number of GFLOPs compared to YOLOv8x.

The Faster R-CNN model is part of the family R-CNN which differ when defining Regions Proposal Network, initially for object detection in Convolutional Neural Networks(CNN) [Li et al. 2022]. Such as YOLO models, the R-CNN family also performs classification, object detection and segmentation. The Faster R-CNN training was enhanced with ResNet-50 [He et al. 2016] and Feature Pyramid Network (FPN) [Lin et al. 2017] within the Detectron2 framework. ResNet-50 is a deep convolutional neural network with 50 trainable layers, designed to mitigate the vanishing gradient problem through residual shortcuts (skip connections), allowing for more efficient training of deep networks. The FPN improves object detection across different scales by combining information from multiple levels of the convolutional network, ensuring higher accuracy in detecting both small and large objects.

The same image dataset split defined for training, validation, and testing in YOLO was applied to this training, with 8 images per batch and 2 workers to optimize data loading. The initial learning rate was set to 0.0001, with no decay throughout the training process. The total number of iterations was 10,000, with checkpoints saved every 250 iterations. The batch size for the ROI head was set to 128. Training was also conducted on an NVIDIA Tesla T4 and included automatic periodic evaluation every 250 iterations.

2.4. Metrics

To evaluate the performance of object detection models, we adopted a set of well-established metrics in computer vision, encompassing precision, recall, mean average precision (mAP), and confusion matrix.

The precision indicates the fraction of detections that are correctly classified relative to the total number of positive predictions, illustrating how many false positives occur. The recall represents the fraction of objects that were effectively detected out of all objects actually present in the images, reflecting the number of false negatives committed. In addition to reporting overall precision and recall values, the models also provide mAP at different Intersection over Union (IoU) [Girshick et al. 2014] thresholds. The mAP computed at IoU = 0.5 (mAP50) assesses the model's ability to correctly detect an object with at least 50% overlap between the predicted and ground-truth bounding boxes, while the mAP calculated over an IoU range from 0.5 to 0.95 (mAP50-95) investigates the model's robustness under increasingly stringent overlap requirements. The confusion matrix lists, for each ground-truth class (columns), which predicted class (rows) was assigned by the model, thus making it clear when recurring misclassifications occur between specific categories. The multiple threshold approach gives a more complete picture of the model's performance by requiring different levels of location accuracy. In addition, making the results available both in aggregate form and by class reveals possible specific weaknesses, guiding future improvements to the model.

3. Results

We first compare the values of these metrics to provide an overview of the effectiveness of each model, and finally, we analyze the confusion matrices of the models on the test set, allowing a understanding of the classification errors and prediction patterns observed.

Table 2 shows that, on average (row "All"), YOLOv8x achieves a precision and a recall of 96.81%, with an mAP50–95 of 68.1%. YOLO11x exhibits slightly higher precision and recall at 97.45% and 97.84%, respectively, and a mAP50–95 of 68.8%. In contrast, Faster R-CNN, while maintaining a precision of 96.33%, records a lower recall of 94.22% and a significantly lower mAP50–95 of 60.20% compared to both versions of YOLO.

Tabela 2. Comparison of YOLOv8x, YOLO11x and Faster R-CNN

	YOLOv8x			YOLO11x			Faster R-CNN		
Class	**Precision**	**Recall**	**mAP50-95**	**Precision**	**Recall**	**mAP50-95**	**Precision**	**Recall**	**mAP50-95**
All	96.81%	96.81%	68.1%	97.45%	97.84%	68.8%	96.33%	94.22%	60.20%
Green	97.10%	98.05%	56.3%	98.02%	96.59%	56.0%	96.59%	99.50%	53.1%
Blue	98.69%	89.35%	58.7%	96.45%	96.45%	60.3%	91.12%	96.86%	51.8%
Purple	95.00%	98.80%	63.0%	98.42%	99.60%	65.6%	98.80%	92.51%	57.2%
Red	95.28%	90.99%	55.2%	94.44%	91.89%	57.5%	92.79%	92.79%	49.0%
Orange	96.67%	98.31%	66.4%	95.00%	96.61%	63.1%	93.22%	79.71%	46.4%
Pink	95.38%	98.41%	56.9%	96.88%	98.41%	59.6%	96.03%	95.28%	49.2%
Did not feel (pink)	97.47%	99.48%	55.5%	97.98%	100.0%	57.7%	98.45%	98.96%	54.8%
Right Hand	100.0%	100.0%	83.5%	100.0%	100.0%	84.3%	97.73%	95.56%	71.2%
Left Hand	100.0%	100.0%	84.4%	97.78%	100.0%	84.3%	100.0%	93.62%	79.0%
Righ Foot	100.0%	100.0%	84.5%	100.0%	100.0%	82.9%	100.0%	75.76%	73.4%
Left Foot	100.0%	100.0%	84.1%	100.0%	100.0%	85.7%	100.0%	83.33%	77.0%

For the color classes, both versions of YOLO models maintain precision and recall above 90%. For instance, for the Green class, YOLOv8x achieves 97.10% of precision and 98.05% of recall, while YOLO11x reaches 98.02% of precision and 96.59% of recall.

Across the color classes, the YOLO models exhibit mAP50–95 values in a range from 55% to 66%. In comparison, Faster R-CNN, although achieving recall such as 99.50% in Green, shows a wider variability in mAP50–95 (ranging from 46.4% to 57.2%).

In the limb categories, all models achieve near-perfect recall, approaching 100%. However, while YOLOv8x and YOLO11x obtain mAP50–95 scores above 80% (approximately 83.5%–84.5%), Faster R-CNN exhibits lower performance, with mAP50–95 values ranging from 71.2% to 77.0%, representing a decrease of approximately 5–10 percentage points.

Figure 4 displays the precision per each of the 11 classes, highlighting the best and worst performances across all models. For Faster R-CNN, the highest precision is found in the limb categories, with Left Hand, Right Foot, and Left Foot achieving 100%, while the lowest precisions occur in the Blue (91.12%) and Red (92.79%) classes. In YOLO11x, among the color classes, the highest precision is seen in Purple (98.42%) and Green (98.02%), with Red showing the lowest value (94.44%). Meanwhile, YOLOv8x achieves its highest precision in Blue (98.69%) and its lowest in Purple (95.00%).

Figure 5 presents the confusion matrices, with rows representing the predicted classes and columns representing the actual classes. For both YOLOv8x and YOLO11x, a strong concentration of values is observed along the diagonal, reflecting accurate predictions, though minor off-diagonal errors appear in some color classes (e.g., Green, Blue, and Purple).

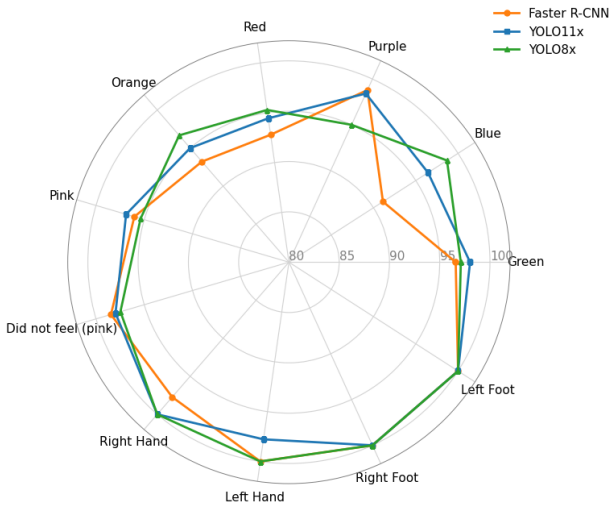
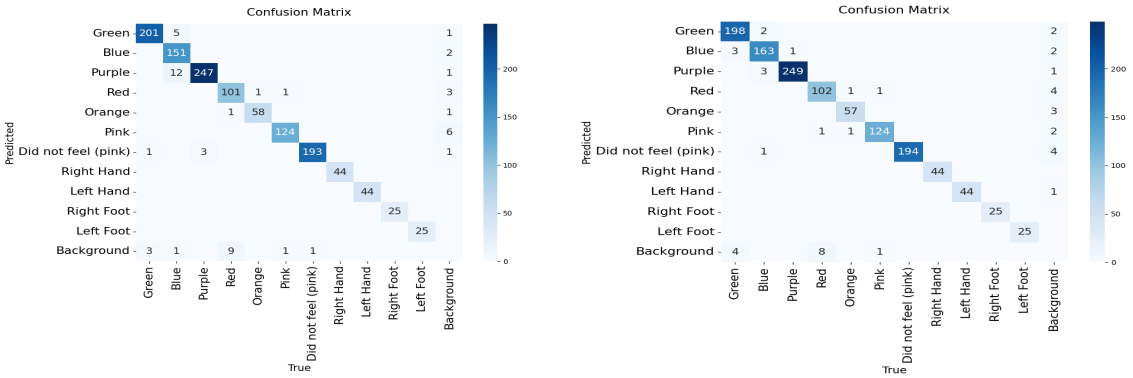
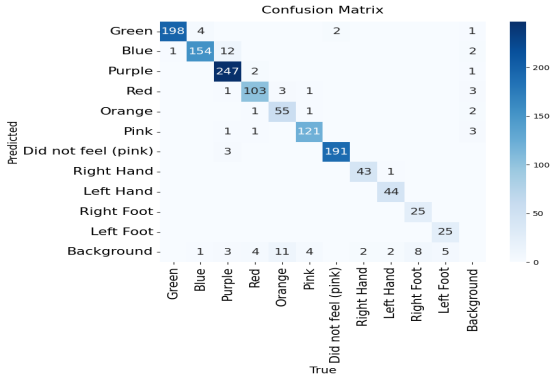


Figura 4. Radar graph related to the precision metric of the evaluated models



(a) YOLOv8x **(b) YOLO11x**



(c) Faster R-CNN

Figura 5. Confusion matrix of the three models: (a) YOLOv8x, (b) YOLO11x and (c) Faster R-CNN

In contrast, Faster R-CNN exhibits a less concentrated diagonal, particularly in classes such as Blue, Red, and Orange. For example, Faster R-CNN achieves a precision of 91.12% for Blue, 92.79% for Red, and 93.22% for Orange, with corresponding recall values of 96.86%, 92.79%, and 79.71%.

4. Discussion

The models achieved performance metrics such as precision, recall, and mAP on the test set. YOLO11x stood out by achieving the highest mAP and a good balance between precision and recall, followed closely by YOLOv8x, whereas Faster R-CNN obtained lower values principally in mAP50-95, harsher scenario.

It was also observed that larger objects (such as the depictions of feet and hands on the form) were detected more easily by the models, whereas smaller or subtler elements had lower accuracy. This is consistent with findings in computer vision, which highlight a greater difficulty in detecting small objects due to limited visual information [Mirzaei et al. 2023].

The study involved images from two versions of the ANS form, which differed in layout and printing resolution. These discrepancies affected performance evidencing the challenges arising from heterogeneous input data. And it is known that image resolution directly impacts the accuracy of object detection [Hao et al. 2023]; thus, variations in the size and clarity of form elements (between the older and the newer versions) require the model to generalize beyond a specific training pattern.

The models were trained without hyperparameters tuning methods and specific fine-tuning for the ANS form. The deep learning literature indicates that hyperparameter optimization can significantly influence the performance of neural models [Ilemobayo et al. 2024]. Without such tuning, the results may not reflect the maximum potential of the tested architectures, especially the Faster R-CNN, which aims to be more robust than the single shot models. In other words, additional improvements in accuracy and generalization could be achieved by adjusting parameters such as learning rate, batch size, or regularization strategies, as evidenced in studies on model optimization.

5. Conclusions and Future Works

Our main goal with the digitization of the ANS forms was to enhance neurological monitoring of leprosy patients by providing structured data for detailed analyses for decision making process. In line with the WHO's global "Zero Leprosy" strategy for 2030 [Sasakawa 2024], digital health technologies have demonstrated practical efficacy in monitoring patients with leprosy [Dahoklory et al. 2023].

This study represents an initial step toward the digitization and automation of neurological analysis in leprosy, enabling more efficient monitoring. In future work, the extracted data can be used to train predictive models of disease progression, OCR techniques should be integrated to supplement textual information extraction, and the system could be validated across different centers to ensure its robustness and applicability.

Acknowledgments

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Secretaria de Ciência, Tecnologia e Inovação e do Complexo Econômico-Industrial da Saúde (SECTICS), Ministério da Saúde (MS). We would like to thank Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE) and Universidade de Pernambuco (UPE).

Referências

Brasil. Ministério da Saúde (2022). Protocolo clínico e diretrizes terapêuticas da hanseníase. https://www.gov.br/conitec/pt-br/midias/protocolos/publicacoes_ms/copy_of_20230131_PCDT_Hanseniase_2022_eletronica_ISBN.pdf.

Dahoklory, D. F., Haryanto, J., and Indarwati, R. (2023). The application of digital health as a nursing solution for leprosy patients during the covid-19 pandemic: A systematic review. *JPMA. The Journal of the Pakistan Medical Association*, 73(Suppl 2)(2):S170–S174. Department of Nursing, Airlangga University, Surabaya, Indonesia.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587.

Hao, Y., Pei, H., Lyu, Y., Yuan, Z., Rizzo, J.-R., Wang, Y., and Fang, Y. (2023). Understanding the impact of image quality and distance of objects to object detection performance. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11436–11442.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Hsu, E., Malagaris, I., Kuo, Y., Sultana, R., and Roberts, K. (2021). Deep learning-based NLP data pipeline for EHR scanned document information extraction. *CoRR*, abs/2110.11864.

Ilemobayo, J., Durodola, O., Alade, O., Awotunde, O., Adewumi, T., Falana, O., Ogungbire, A., Osinuga, A., Ogunbiyi, D., Odezuligbo, I., Edu, O., and Ifeanyi, A. (2024). Hyperparameter tuning in machine learning: A comprehensive review. *Journal of Engineering Research and Reports*, 26:388–395.

Leite, S., Barros, A., Fonseca, M., Andrade, T., Foss, N., and Frade, M. (2010). Avaliação sensitiva de hansenianos pelos monofilamentos semmes-weinstein em serviço terciário de fisioterapia. *Hansenologia Internationalis: hanseníase e outras doenças infecciosas*, 35:9–16.

Leite-Moreira, A., Mendes, A., Pedrosa, A., Rocha-Sousa, A., Azevedo, A., Amaral-Gomes, A., Pinto, C., Figueira, H., Pereira, N. R., Mendes, P., and Pimenta, T. (2022). An nlp solution to foster the use of information in electronic health records for efficiency in decision-making in hospital care.

Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2022). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):6999–7019.

Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ministério da Saúde - Secretaria de Vigilância em Saúde (2021). *Formulário para Avaliação Neurológica Simplificada e Classificação do Grau de Incapacidade Física em Hanseníase*. Ministério da Saúde, Brasil.

Ministério da Saúde, Secretaria de Ciência, T. I. e. I. E. e. S. (2022). Protocolo clÍnico e diretrizes terapÊuticas da hansenÍase. https://www.gov.br/conitec/pt-br/midias/protocolos/20220818_pcdt_hanseniase.pdf. Portaria SCTIE/MS Nº 67, de 7 de julho de 2022.

Mirzaei, B., Nezamabadi-Pour, H., Raoof, A., and Derakhshani, R. (2023). Small object detection and tracking: A comprehensive review. *Sensors (Basel)*, 23(15):6887. Intelligent Data Processing Laboratory (IDPL), Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran. Department of Earth Sciences, Utrecht University, Utrecht, Netherlands.

Nóbrega, M. d. M., Santana, E. M. F., Brito, K. K. G., Antas, E. M. V., Pacheco, F. C. d. S., Oliveira, S. H. d. S., and Soares, M. J. G. O. (2024). PercepÇÕes de profissionais sobre intervenÇÃo educativa acerca da avaliaÇÃo neurolÓgica simplificada da hansenÍase. *Revista Baiana de Enfermagem*, 38(.).

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788.

Sasakawa, Y. (2024). What's needed to achieve zero leprosy. *Bulletin of the World Health Organization*, 102(8):554–554A. The Nippon Foundation, The Nippon Zaidan Building 1-2-2 Akasaka, Minato-ku Tokyo 107-8404, Japan.

WHO (2017). *Guidelines for the Diagnosis, Treatment and Prevention of Leprosy*. World Health Organization, Regional Office for South-East Asia, New Delhi, India. Licence: CC BY-NC-SA 3.0 IGO.

WHO (2021). Towards zero leprosy. global leprosy (hansen's disease) strategy 2021–2030. Advocacy Brief.

WHO (2024). Global leprosy (hansen disease) update, 2023: Elimination of leprosy disease is possible – time to act! WHO Weekly Epidemiological Record, WER No 37, 2024, 99, 501–521.