Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 2 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i02.49749

PAPER

Identification of Laryngeal Lesions Based on Narrowband Endoscopy Imaging Using Artificial Neural Networks and Visual Programming

Ana Luiza Moraes Fernandes da Costa[1](✉), Gerardo Antonio Idrobo Pizo[1], Luis Felipe Rivera Gómez[2]

[1]Graduate Program in Biomedical Engineering, UnB Gama College, University of Brasília, Brasília, Brazil

[2]Graduate Program in Aerospace Engineering, UnB Gama College, University of Brasília, Brasília, Brazil

analuizamoraesfc@gmail.com

ABSTRACT

Certain types of lesions on the laryngeal mucosa may indicate the early stages of laryngeal squamous cell carcinoma (LSCC), which constitutes 98% of malignant laryngeal tumors. This study aims to develop artificial intelligence-based methods to classify laryngeal lesions using digital images obtained through narrowband endoscopy imaging. A total of 1,320 digital images of laryngeal tissue, both healthy and lesioned, were classified into four categories. Five machine learning models were developed, utilizing conventional deep convolutional neural networks (CNNs) and capsule networks (CapsNet): VGG16, VGG19, Inception V3, CapsNet without data augmentation, and CapsNet with data augmentation. The latter used images synthetically generated by an adversarial generative network (GAN). These algorithms were implemented using the Orange visual programming software and the Colab computational platform. The inclusion of GAN-enhanced data augmentation significantly improved the performance of the CapsNet classifier across all lesion types. The CapsNet model with GAN data augmentation achieved an average recall, accuracy, and F1 score of 94.7%, marking it as the second-best performing model. The highest performance was achieved by the CNN Inception V3 model, with 97% recall, accuracy, and F1 score, facilitated through visual programming. The combination of CapsNet with GAN-based data augmentation presents a viable alternative for the classification of medical images. The use of the Orange visual programming tool enabled high classification performance—97% in both accuracy and sensitivity—at low computational costs, without the need for advanced programming skills from the user.

KEYWORDS

artificial intelligence, capsule networks (CapsNet), laryngeal lesions, narrowband imaging (NBI)

Costa, A.L.M.F., Pizo, G.A.I., Gómez, L.F.R. (2025). Identification of Laryngeal Lesions Based on Narrowband Endoscopy Imaging Using Artificial Neural Networks and Visual Programming. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(2), pp. 52–62. https://doi.org/10.3991/ijoe.v21i02.49749

Article submitted 2024-04-19. Revision uploaded 2024-07-11. Final acceptance 2024-07-11.

1 INTRODUCTION

Laryngeal cancer constitutes approximately 3% of all malignant neoplasms worldwide [1], with over 184,000 reported cases in 2020 alone. In Brazil, there were 4,450 deaths from this disease in the same year. For 2023, the estimated incidence rates are 6.21 new cases per 100,000 for men and 1.09 new cases per 100,000 for women [2] [3] [4]. Late diagnosis often necessitates aggressive treatments such as larynx and vocal fold resections, which significantly impact quality of life and increase mortality rates.

A primary challenge in early-stage diagnosis is detecting minor lesions in the laryngeal mucosa, which are often overlooked by specialists. To address this, machine learning methods, particularly artificial neural networks (ANNs), have been used to enhance the classification accuracy of various imaging modalities, significantly improving disease identification [5].

The scientific literature reveals that machine learning models, especially those using deep convolutional neural network (CNN) architectures, have achieved notable success in detecting laryngeal lesions, with accuracy rates reaching up to 98% [6]. Despite CNNs' promising results, the accessibility of these technologies is limited by the need for programming knowledge. Moreover, newer approaches, such as capsule neural networks, have not been fully explored in classifying images with laryngeal lesions. Capsule networks (CapsNet), a more recent architecture, aim to overcome some limitations of traditional CNNs by preserving positional information and hierarchical relationships within images.

Research has shown that various machine learning models and techniques can significantly improve the classification and diagnosis of medical images. For instance, studies have explored the application of ANNs in classifying learning styles in educational environments, which, although different in context, highlight the versatility and potential of ANNs in various fields [14] [31]. Comparisons between different deep learning models, such as YOLO and RetinaNet, have provided insights into optimizing performance in image classification tasks [15]. Additionally, the use of CNNs for detecting conditions like pterygium in ophthalmological centers demonstrates the applicability of these networks in medical diagnostics [16]. Further systematic reviews on the use of CNNs in vascular surgeries and multimodal medical image fusion underscore the broad utility and impact of these technologies [17] [18] [19] [20].

Laryngoscopy with narrowband imaging (NBI) uses narrowband light to enhance the visualization of pre-neoplastic and cancerous lesions in the mucosa of the upper aerodigestive tract. Early detection of laryngeal squamous cell carcinoma (LSCC) with NBI can significantly improve patient prognosis. However, manual identification of these lesions is subjective, highlighting the need for automated methods. This study aims to develop and evaluate neural network models, such as CapsNet and generative adversarial networks (GAN), to improve diagnostic accuracy and facilitate the early detection of laryngeal lesions.

This study explores several research questions: How does a supervised machine learning model based on CapsNet architecture perform in classifying laryngeal lesions in NBI compared to the main ANN architectures cited in the literature? Does the inclusion of synthetic medical images generated by GAN enhance the performance of a CapsNet-based classifier? How effective is an artificial intelligence solution implemented with visual programming in classifying lesions in laryngeal mucosa images?

Laryngeal cancer, which accounts for 98% of all laryngeal cancers, typically originates from abnormal squamous cell development in the laryngeal lining. Common symptoms include difficulty swallowing, persistent hoarseness, and the presence or absence of cervical nodules [7]. Treatment options can be invasive, involving total laryngectomies, chemotherapy, and radiotherapy, all of which can exacerbate symptoms and adversely affect the patient's quality of life [8]. Early detection of laryngeal cancer is vital for improving prognosis, reducing mortality rates, and preserving the anatomy of the larynx and vocal folds [9].

Emerging as a crucial tool for early squamous cell carcinoma of the larynx (LSCC) diagnosis, NBI utilizes light to enhance neoangiogenic patterns associated with preneoplastic and cancerous lesions in the upper aerodigestive tract mucosa. Key indicators for LSCC diagnosis through NBI include hypertrophic vessels with vesicular loops and leukoplakia in the vocal folds, with vocal fold leukoplakia characterized by flat or thick white epithelial plaques or spots [10] [11] [12].

The objective of this study is to develop and evaluate the performance of a classifier based on capsule neural networks for classifying laryngeal lesions in narrowband images obtained through laryngoscopy. It also involves augmenting the database with GAN and implementing CNN models within a visual programming software environment.

1.1 Research in context

Evidence before this study: A literature search was conducted in the Scopus, Science Direct, and Web of Science databases through November 16, 2023, to identify studies on the application of machine learning methods for diagnosing laryngeal cancer from images obtained via NBI. The search terms used were: ((“NBI” OR “narrowband imaging”) AND (“larynx” OR “laryngeal”) AND (“cancer” OR “carcinoma”) AND (“machine learning” OR “deep learning”)). This search yielded nine distinct studies, published between 2017 and 2022, which discussed the use of machine learning models to assist in diagnosing laryngeal carcinoma from NBI-acquired images. The highest reported sensitivity was 98% in studies related to multiclass image classification. However, there were no studies found that focused on the use of CapsNet combined with GAN data augmentation for cancer detection in laryngeal tissue images obtained by NBI. Furthermore, no results were found regarding image classification using visual programming software.

Added value of this study: This study explores the performance of machine learning models based on CapsNet enhanced with data augmentation using synthetic samples generated by GAN and models based on deep CNNs implemented via visual programming. The results demonstrate that these approaches enabled the classification of laryngeal lesions with up to 97% accuracy and sensitivity.

Implications of all the available evidence: When combined with existing evidence, it is clear that methods based on ANNs are significant tools in aiding the diagnosis of precursor lesions of laryngeal cancer. The approach developed using CapsNet and GAN presents a promising option for the classification of medical images, particularly when the dataset available for training is limited. Moreover, the implementation of CNNs through visual programming has proven to be an effective technique that is accessible to users without advanced programming skills, enhancing the capability to diagnose diseases based on image classification.

2 MATERIALS AND METHODS

The methodology of this study was structured into several key steps: Initially, a dataset consisting of NBI photos of the laryngeal mucosa was acquired. This was followed by the digital processing of these images. Classifiers based on ANNs, specifically focusing on CNN and CapsNet models, were then implemented. The dataset was further expanded using GAN to enhance the diversity and size of the data pool. The extended dataset underwent classification using the CapsNet-based classifier. Finally, the performance of each algorithm involved in the study was evaluated and compared.

For the digital processing and initial model training, the Orange Data Mining software, version 2.27, was utilized. Orange is an open-access tool that is widely recognized in the academic community for its intuitiveness and user-friendliness [29]. It is freely available and features an interface that supports the use of predefined ANNs such as Inception V3, VGG16, and VGG19 through its Image Analytics add-on. The training of the neural networks within the orange software was carried out on a computer equipped with an Intel Core i7 processor running at 2.40 GHz with two cores and 8 gigabytes of RAM.

For tasks requiring more advanced computational capabilities, such as the generation of artificial images with GAN and the implementation of the CapsNet classifier, a Jupyter notebook hosted in the cloud was used. This notebook utilized a Tesla V100 GPU with 16 gigabytes of RAM, provided through the Colab platform. Colab is available via a payment-based model, which allows users to access enhanced computational resources for a predetermined duration.

2.1 Dataset acquisition

In this study, a publicly accessible dataset from the research of Moccia et al., 2017 [6], consisting of 1,320 images, was utilized. These images, each 100 × 100 pixels in size, represent healthy and early-stage cancerous laryngeal tissues. They were manually extracted from 33 narrow-band laryngoscopic images, each derived from a different patient diagnosed with squamous cell carcinoma of the larynx, confirmed through histopathological examination.

The dataset is methodically divided into four tissue classes, with each class comprising 330 images: HEA (healthy tissue), HV (tissue with hypertrophied vessels), LEU (tissue with leukoplakia), and IPCL (tissue with intrapapillary capillary lesions).

The "laryngeal dataset.tar" file contains three subfolders (FOLD 1, FOLD 2, FOLD 3), which are used for cross-validation purposes in the performance evaluation of tissue classification.

Each subfolder includes four subfolders corresponding to the four tissue classes, namely HEA, HV, LEU, and IPCL. This organization ensures that each tissue sample is a distinct and direct extraction from the laryngoscope images, capturing a diverse and representative array of tissue conditions essential for effective texture-based classification and early-stage diagnosis support.

2.2 Data augmentation

The performance of deep learning-based algorithms in image classification tasks can be significantly improved by employing data augmentation techniques. These techniques expand the training dataset, enhancing the model's ability to generalize.

Generative adversarial networks are a type of neural network designed to generate artificial samples based on the underlying data distribution. They comprise two main components: a "discriminator" that learns to classify a given example x into a label y based on a conditional probability distribution and a "generator" that models the joint probability distribution p (x, y), where x represents the input data and y represents the corresponding label [13].

The use of GANs to generate synthetic images has been shown to be an effective strategy to overcome the limitations imposed by the small size of medical image datasets [30]. By adding synthetic images generated by GANs, the variability of data is increased, which significantly enhances the training process of deep learning classifiers.

In this study, 440 images from the original dataset were used to generate artificial images, distributed as follows: 110 healthy tissue images, 110 with leukoplakia, 110 featuring hypertrophic vessels, and 110 containing intrapapillary capillary loops.

The use of a deep convolutional generative adversarial network (DCGAN) for data augmentation proved effective in creating an additional 200 artificial images for each of the four classes in the original dataset. This process involved the implementation of a DCGAN that adapted previously developed network models, specifically adjusting the network architecture and hyperparameters, such as a learning rate of 0.0002 and five layers, to meet the specific needs of our dataset and research objectives, as summarized in Table 1.

Table 1. Results of data augmentation with DCGAN

Classes	Images Generated	Training Epochs
HV	200	700
IPCL	200	100
LEU	200	700
HEA	200	350

Figure 1 showcases samples of artificial images generated by the DCGAN. These images represent the classes IPCL, HV, LEU, and HEA from the original set of images.

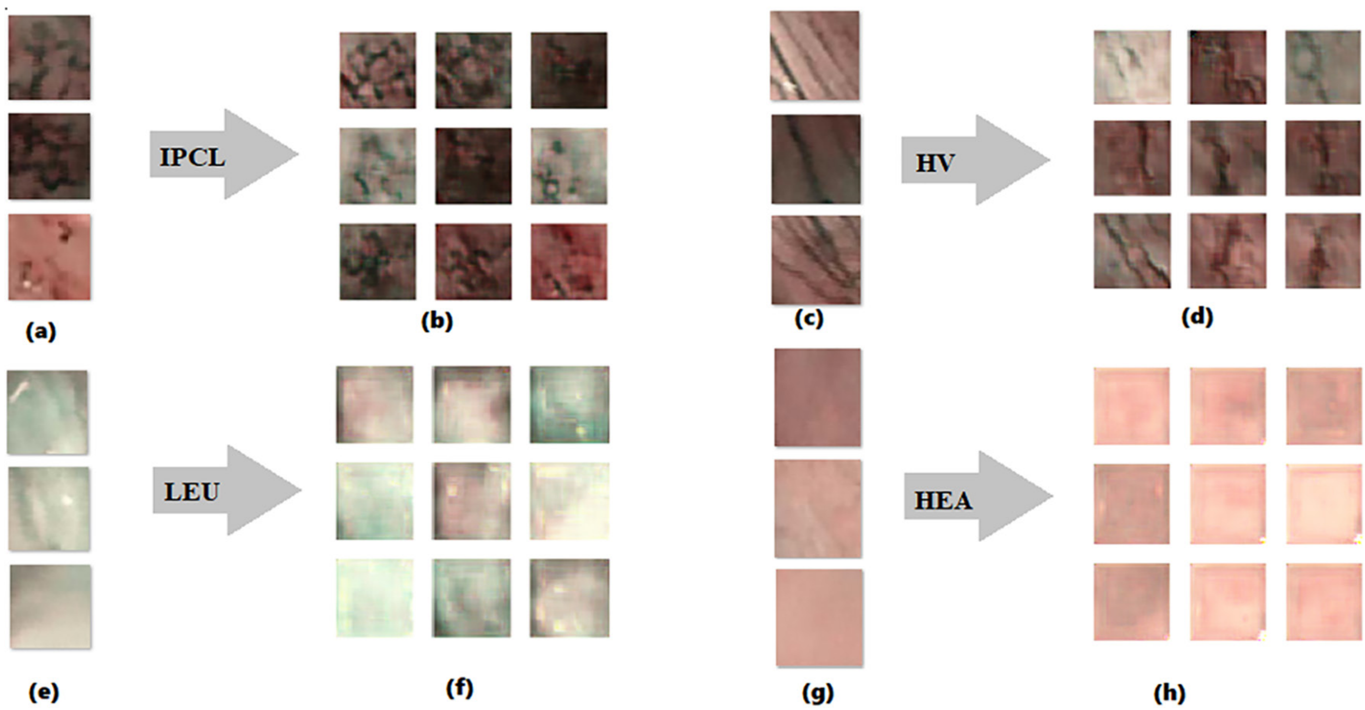

Fig. 1. Sample of images generated by data augmentation with DCGAN

Notes: (a) Original images of the IPCL (Intrapapillary Capillary Loops) class, (b) Generated images for the IPCL class after data augmentation with DCGAN, (c) Original images of the HV (Hypertrophic Vessels) class, (d) Generated images for the HV class after data augmentation with DCGAN, (e) Original images of the LEU (Leukoplakia) class, (f) Generated images for the LEU class after data augmentation with DCGAN, (g) Original images of the HEA (Healthy Tissue) class, (h) Generated images for the HEA class after data augmentation with DCGAN.

3 RESULTS

This study presents a proposal for employing deep ANNs to automate the classification of laryngeal lesions from images acquired through laryngoscopy using NBI techniques. The results achieved by each implemented ANN classifier are depicted in Table 2.

Table 2. Results of data augmentation with DCGAN for CapsNet

Architecture	F1-Score	Precision	Recall	Image used on Training Step
CNN VGG 16	0.947	0.947	0.947	1188
CNN VGG 19	0.931	0.936	0.932	1188
CNN Inception V3	0.970	0.970	0.970	1188
CapsNet	0.931	0.934	0.932	1056 + 132 images in validation set
CNN VGG 16	0.947	0.947	0.947	1856 + 132 images in validation set

Among all models implemented, the CNN Inception V3 applied to the original dataset performed the best, achieving values greater than 97% across all three evaluated metrics. The CapsNet classifier, enhanced with data augmentation using artificial images generated by GAN, showed the second-best performance, achieving a recall, f1-score, and overall accuracy of 94.7%, which matched the performance of the CNN VGG16 model.

The integration of synthetic medical images generated by DCGAN positively impacted the performance of the CapsNet classifier across all evaluated metrics, improving the f1-score by 1.6 percentage points, recall by 1.5 percentage points, and accuracy by 1.32 percentage points.

Overfitting, a common issue where a model fits the training data well but struggles to generalize to new, unseen data, was observed. This phenomenon can occur in situations where the input data sample is too small or lacks variability. Despite efforts to expand the image dataset used for training, it was not possible to completely eliminate the occurrence of overfitting in the CapsNet model. This issue was evident from the divergence of curves representing the accuracy values achieved by the model during the training and validation stages, as shown in Figure 2. Overfitting indicates that, although the models achieved high performance metrics on the training dataset, their generalization ability may be limited when applied to new datasets. Therefore, it is crucial to validate the model with additional independent data to ensure its robustness and reliability.

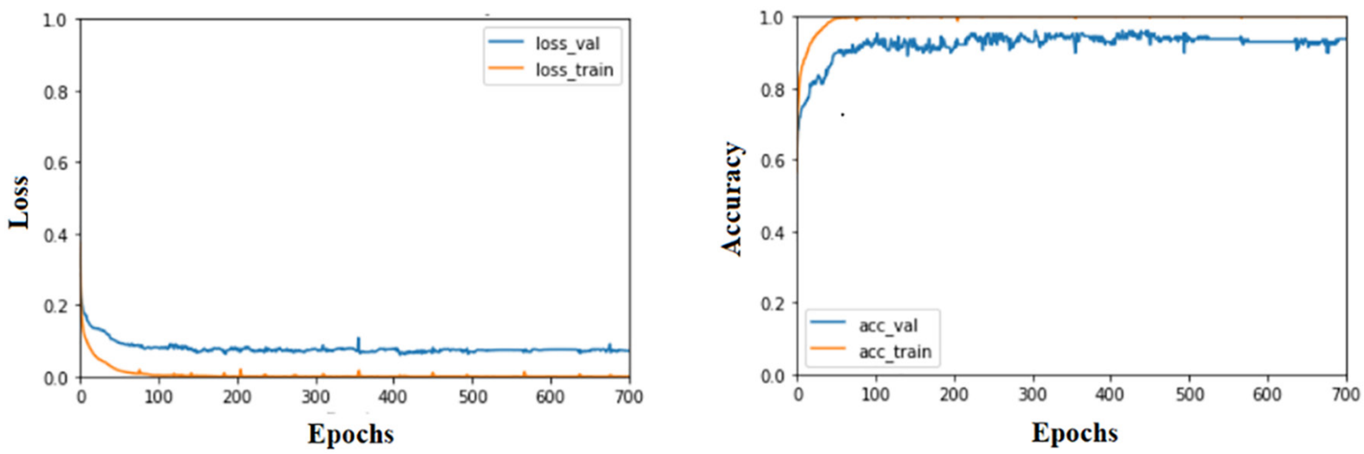

Fig. 2. Loss and accuracy obtained by the CapsNet classifier with data augmentation

Additionally, classifiers based on conventional CNN were also implemented using the Orange Data Mining visual programming and data mining tool. Figure 3 displays the structure of the visual programming workflow implemented in the Orange Data Mining Software for CNN classification. The model based on Inception V3, implemented using the visual programming workflow in the free Orange Image Analytics software (see Figure 3), achieved the best overall performance, even surpassing the CapsNet model in image classification with an average sensitivity, accuracy, and f1-score equivalent to 97%, which is close to the state of the art (98%) (refer to Table 3).

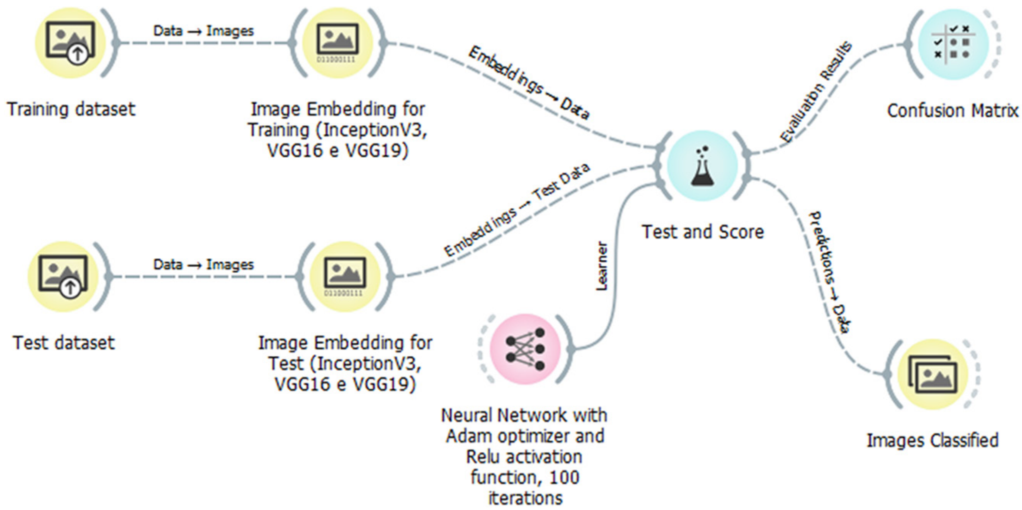

Fig. 3. Visual programming workflow implemented in Orange Datamining Software for CNN classification

Table 3. Comparison of obtained results with correlated works

Author	Classification	Machine Learning Architecture	Average Recall
Araújo et al., 2019 [21]	Multiclass	SVM and CNN	98%
Moccia et al., 2017 [6]	Multiclass	SVM	98%
Present work	Multiclass	CNN Inception V3 using visual programming with Orange Datamining software	97%
Kwon et al., 2022 [22]	Binary	Decision Trees	97%
Inaba et al., 2020 [23]	Binary	Retina Net	95.50%
Azam et al., 2022a [24]	Multiclass	SegMENT	95%
Present work	Multiclass	CapsNet and DCGAN	94.70%
Xiong et al., 2019 [25]	Multiclass	CNN	92.20%
He et al., 2021 [26]	Binary	CNN	90.10%
Esmaeili et al., 2020 [27]	Binary	RF (Random Forest)	84.60%
Azam et al., 2022b [28]	Binary	CNN YOLO	62%

4 DISCUSSION AND CONCLUSION

In comparison to related studies on the multiclass classification of laryngeal lesions in images obtained by NBI, the results obtained using the Inception V3 model were closest to the state of the art, achieving 98% accuracy. The CapsNet-based method with data augmentation via GAN achieved results close to 95%. Table 3 below displays the arrangement of the best results obtained for the recall metric, comparing the present study with related works.

It is crucial to note that this study focused exclusively on the application of database augmentation techniques with GAN, using only a CapsNet based model for the automated classification of laryngeal lesions. Further research is necessary to analyze the impact of this data augmentation method when combined with other classifier models.

A limitation identified in this study was the dataset division used by the CNN classifiers implemented with visual programming, as images were not reserved for the validation subset. Thus, there is an opportunity to enhance CNN classifiers by allocating part of the images for model validation.

The DCGAN model used in this study exhibited instability during training due to the reduced sample of input data. Despite this, the technique was able to improve the performance of the CapsNet classifier. Recommendations for future work include fine-tuning to increase the stability of the DCGAN model during training and adopting a mechanism to automate the validation of samples generated by the DCGAN network, with the aid of reliability assessment metrics such as Frechet Inception Distance (FID).

The noticeable mismatch between the validation and training accuracy curves in the classification performed with the CapsNet models suggested the presence of overfitting. Therefore, using a larger database for training and implementing automated overfitting control mechanisms in the CapsNet models are recommended enhancements for future works implemented on the Colab platform.

Laryngeal cancer is a pathology that can severely compromise the quality of life and often leads to death when identified in an advanced stage. Early identification of specific lesions in the larynx can signal the development of precancerous tissue, emphasizing the importance of early detection for effective treatment.

Methods based on ANNs are critical tools for aiding the diagnosis of diseases in imaging exams. The solution developed in this study, which combines CapsNet and GANs, may represent a promising alternative for the automated classification of medical images, particularly in cases involving rare diseases where training data may be scarce. However, improvements are needed to prevent overfitting and reduce the instability of the generative model.

This study contributes to the advancement of health-related technologies by exploring new methods to aid in the diagnosis of precursor lesions of laryngeal cancer. The primary contribution of this study was the presentation of an artificial intelligence solution based on the implementation of CNNs through a visual programming tool. This approach proved to be an efficient technique for the classification of medical images, achieving 97% accuracy and sensitivity, providing a user-friendly interface, low operating costs, and requiring no advanced programming skills from the user.

5 REFERENCES

[1] S. Dogantemur *et al.*, "Assessment of HPV 16, HPV 18, p16 expression in advanced stage laryngeal cancer patients and prognostic significance: Assessment of HPV 16, HPV 18, p16 expression in laryngeal cancer," *Braz. J. Otorhinolaryngol.*, vol. 86, no. 3, pp. 351–357, 2020. https://doi.org/10.1016/j.bjorl.2019.11.005

[2] World Health Organization, "The Impact of Cancer [2020]," 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer. [Accessed: Apr. 8, 2021].

[3] Instituto Nacional do Câncer – INCA, "Câncer de laringe," Dec. 2020. [Online]. Available: https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/laringe/versao-para-profissionais-de-saude. [Accessed: Dec. 27, 2020].

[4] Instituto Nacional do Câncer – INCA, "Câncer de laringe," Dec. 2023. [Online]. Available: https://www.inca.gov.br/publicacoes/livros/estimativa-2023-incidencia-de-cancer-no-brasil. [Accessed: Jan. 20, 2023].

[5] M. K. Santos, J. R. Ferreira Júnior, D. T. Wada, A. P. M. Tenório, M. H. Nogueira-Barbosa, and P. M. de A. Marques, "Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: Advances in imaging towards precision medicine," *Radiol. Bras.*, vol. 52, no. 6, pp. 387–396, 2019. https://doi.org/10.1590/0100-3984.2019.0049

[6] S. Moccia *et al.*, "Confident texture-based laryngeal tissue classification for early stage diagnosis support," *Journal of Medical Imaging*, vol. 4, no. 3, 2017. https://doi.org/10.1117/1.JMI.4.3.034502

[7] R. Hermans, "Staging of laryngeal and hypopharyngeal cancer: Value of imaging studies," *European Radiology*, vol. 16, pp. 2386–2400, 2006. https://doi.org/10.1007/s00330-006-0301-7

[8] C. T. V. Maciel, I. C. G. Leite, and V. T. Soares, "Câncer de laringe: um olhar sobre a qualidade de vida," *Revista Interdisciplinar de Estudos Experimentais – Animais e Humanos (RIEE)*, vol. 2, no. 4, pp. 126–134, 2010.

[9] J. Unger, J. Lohscheller, M. Reiter, K. Eder, C. S. Betz, and M. Schuster, "A noninvasive procedure for early-stage discrimination of malignant and precancerous vocal fold lesions based on laryngeal dynamics analysis," *Cancer Res.*, vol. 75, no. 1, pp. 31–39, 2015. https://doi.org/10.1158/0008-5472.CAN-14-1458

[10] C. Piazza, F. Del Bon, G. Peretti, and P. Nicolai, "Narrow band imaging in endoscopic evaluation of the larynx," *Current Opinion in Otolaryngology, Head and Neck Surgery*, vol. 20, no. 6, pp. 472–476, 2012. https://doi.org/10.1097/MOO.0b013e32835908ac

[11] A. Panwar, R. Lindau, and A. Wieland, "Management of premalignant lesions of the larynx," *Expert Review of Anticancer Therapy*, vol. 13, no. 9, pp. 1045–1051, 2013. https://doi.org/10.1586/14737140.2013.829643

[12] H. Klimza, J. Jackowska, C. Piazza, J. Banaszewski, and M. Wierzbicka, "The role of intraoperative narrow-band imaging in transoral laser microsurgery for early and moderately advanced glottic cancer," *Braz. J. Otorhinolaryngol.*, vol. 85, no. 2, pp. 228–236, 2019. https://doi.org/10.1016/j.bjorl.2018.01.004

[13] A. T. Oliveira, "Classificação automática de arritmias cardíacas usando uma combinação de redes neurais," *Teste (Doutorado em Engenharia Mecânica) – Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas*, p. 81, 2019.

[14] Y. Gambo and M. Z. Shakir, "An Artificial Neural Network (ANN)-based learning agent for classifying learning styles in self-regulated smart learning environment," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 18, pp. 185–199, 2021. https://doi.org/10.3991/ijet.v16i18.24251

[15] N. I. Nife and M. Chtourou, "A comprehensive study of deep learning and performance comparison of deep neural network models (YOLO, RetinaNet)," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 12, pp. 62–77, 2023. https://doi.org/10.3991/ijoe.v19i12.42607

[16] E. J. Ticlavilca-Inche, M. I. Moreno-Lozano, P. Castañeda, S. Wong-Durand, and A. Oñate-Andino, "Mobile application based on convolutional neural networks for pterygium detection in anterior segment eye images at ophthalmological medical centers," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 20, no. 8, pp. 115–138, 2024. https://doi.org/10.3991/ijoe.v20i08.48421

[17] J. Gamboa-Cruzado, M. Rojas-Morales, J. López-Goycochea, E. Condor Tinoco, G. Paucar-Carlos, and A. Sifuentes Damián, "Systematic literature review on convolutional neural networks for vascular surgeries," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 12, pp. 138–158, 2022. https://doi.org/10.3991/ijoe.v18i12.31535
[18] N. Alseelawi, H. T. Hazim, and H. T. S. Alrikabi, "A novel method of multimodal medical image fusion based on hybrid approach of NSCT and DTCWT," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 3, pp. 114–133, 2022. https://doi.org/10.3991/ijoe.v18i03.28011
[19] P. Afshar *et al*., "3D-MCN: A 3D multiscale capsule network for lung nodule malignancy prediction," *Sci. Rep.*, vol. 10, p. 7948, 2020. https://doi.org/10.1038/s41598-020-64824-5
[20] S. Yadav and S. M. Jadhav, "Deep convolutional neural network-based medical image classification for disease diagnosis," *Journal of Big Data*, vol. 6, 2019. https://doi.org/10.1186/s40537-019-0276-2
[21] T. Araújo *et al*., "Learned and handcrafted features for early-stage laryngeal SCC diagnosis," *Med/ Biol/ Eng/ Comput/*, vol. 57, pp. 2683–2692, 2019. https://doi.org/10.1007/s11517-019-02051-5
[22] I. Kwon *et al*., "Diagnosis of early glottic cancer using laryngeal image and voice based on ensemble learning of convolutional neural network classifiers," *Journal of Voice*, vol. 39, no. 1, pp. 245–257, 2022. https://doi.org/10.1016/j.jvoice.2022.07.007
[23] A. Inaba *et al*., "Artificial intelligence system for detecting superficial laryngopharyngeal cancer with the high efficiency of deep learning," *Head & Neck*, vol. 42, no. 9, pp. 2581–2592, 2020. https://doi.org/10.1002/hed.26313
[24] M. A. Azam *et al*., "Videomics of the upper aero-digestive tract cancer: Deep learning applied to white light and narrow band imaging for automatic segmentation of Endoscopic Images," *Frontiers in Oncology*, vol. 12, 2022. https://doi.org/10.3389/fonc.2022.900451
[25] H. Xiong *et al*., "Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images," *EBioMedicine*, vol. 48, pp. 92–99, 2019. https://doi.org/10.1016/j.ebiom.2019.08.075
[26] Y. He *et al*., "A deep convolutional neural network-based method for laryngeal squamous cell carcinoma diagnosis," *Annals of Translational Medicine*, vol. 9, no. 24, 2021. https://doi.org/10.21037/atm-21-6458
[27] N. Esmaeili *et al*., "Laryngeal lesion classification based on vascular patterns in contact endoscopy and narrow band imaging: Manual versus automatic approach," *Sensors*, vol. 20, no. 14, 2020. https://doi.org/10.3390/s20144018
[28] M. A. Azam *et al*., "Deep learning applied to white light and narrow band imaging videolaryngoscopy: Toward real-time laryngeal cancer detection," *The Laryngoscope*, vol. 132, no. 9, pp. 1798–1806, 2021. https://doi.org/10.1002/lary.29960
[29] J. Demšar *et al*., "Orange: Data mining toolbox in Python," *J. Mach. Learn. Res.*, vol. 14, pp. 2349–2353, 2013.
[30] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review." *Medical Image Analysis*, vol. 58, p. 101552, 2019. https://doi.org/10.1016/j.media.2019.101552
[31] L. R. Al-Khazraji, A. R. Abbas, and A. S. Jamil, "The effect of changing targeted layers of the deep dream technique using VGG-16 model," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 3, pp. 34–47, 2023. https://doi.org/10.3991/ijoe.v19i03.37235

6 AUTHORS

Ana Luiza Moraes Fernandes da Costa holds a degree in Systems Analysis and Development from the Instituto de Educação Superior de Brasília (2017). She is a specialist in Project Management from the Instituto de Educação Superior de Brasília (2018) and holds a master's degree in Biomedical Engineering from the University of Brasília (2023). She has experience in the field of Computer Science, with an emphasis on Information Systems. Her interests include Artificial Intelligence, Machine Learning, and Computer Vision, applied to biomedical systems (E-mail: analuizamoraesfc@gmail.com; CV: http://lattes.cnpq.br/3778077298891929).

Gerardo Antonio Idrobo Pizo holds a degree in Physical Engineering from the University of Cauca, Colombia, and a master's and doctorate in Mechatronic Systems from the University of Brasília, Brazil. He is currently an Associate Professor at the University of Brasília (UnB), working in the postgraduate program in Biomedical Engineering at the UnB-Gama Faculty (FGA). His main research interests include Artificial Intelligence, Optical Spectroscopy, Biomedical Systems Development, and Computer Vision (E-mail: gerardo_idrobo@unb.br; CV: http://lattes.cnpq.br/6015706048119134).

Luis Felipe Rivera Gómez is currently pursuing a degree in Aerospace Engineering at the University of Brasília. His research interests include Computer Vision and Artificial Intelligence (CV: http://lattes.cnpq.br/7620342465370630).