A Study of Machine Learning-Driven Diagnosis of Laryngeal Cancer from Narrow-Band Imaging and CT Scans

Zahraa M. Abbas[1], Maalim A. Haasan[2]
[1,2]Department of Computer Science, Basra University, Basra, Iraq

Article Info

Article history:

Received Oct 18, 2025
Revised Dec 23, 2025
Accepted Feb 8, 2026

Keywords:

Cancer
Laryngeal Cancer
Narrow Band Imaging (NBI)
Machine Learning
Deep Learning

ABSTRACT

Laryngeal Carcinoma (LC) represents the most widely malignancy in the neck and head region. It is associated with risk factors such as alcohol consumption, smoking, and exposure to environmentally harmful substances. In its early stages, this disease is difficult to diagnose, leading to a poor prognosis for patients. This research attempts to discover the potential for improving the diagnostic accuracy of laryngeal cancer using some techniques of Machine Learning (ML) and deep learning (DL). This work demonstrates the use of advanced models and mechanisms, including convolutional neural networks (CNNs), support vector machines (SVMs), and random forests, to analyze laryngeal images taken from narrow-band imaging (NBI) computed tomography (CT). The study includes preprocessing methods such as normalization, denoising, segmentation, and data augmentation, which enhance the performance and effectiveness of the models. The project aims to evaluate the performance of models using metrics such as sensitivity, accuracy, specificity, and the F1 measure, so that it contributes to developing effective tools that help physicians make accurate data-driven diagnostic and treatment decisions. Ultimately, the research presents potential future applications of these technologies, based on integrated images and clinical data analysis, to predict disease progression and survival rates. These efforts will contribute to more effective tools for diagnosing and treating laryngeal cancer.

Corresponding Author:

Zahraa Mahmoud Abbas
Maalim A. Hassan
Department of Computer Science, Basra University College, Basra, Iraq
Email: zahraa.mahmoud@uobasrah.edu.iq

1. INTRODUCTION

Cancer is a severe disease, indicated by the development of unusual cells and their suspicious spread, damaging healthy tissue in the body. Various treatment strategies, including surgery, chemotherapy, and radiation therapy, are used to treat advanced cases of the disease, followed by adjuvant radiation therapy, either with or without concurrent chemotherapy [1].

Depending on where it originates, there are more than a hundred separate species that fall under each of the four groups, such as Leukemia, Lymphoma, Carcinoma, and Sarcoma. Throat cancer falls into a cancer category that has two basic types: adenocarcinoma (an uncommon type affecting organs) and squamous cell carcinoma (affecting the smooth throat cells). Throat cancer is separated into two categories: laryngeal cancer and pharyngeal cancer [2].

Laryngeal cancer is one of the most common neck and head tumors; it represents about a fourth to a third of neck and head tumors and forms 2% of total body cancers. Most of the laryngeal cancers originate in the vocal folds, and they mainly affect men [3]. It is often considered to be linked with factors such as alcohol consumption, smoking, air pollution, and human papillomavirus (HPV) infection [4]. Many researchers have measured the effect of the laryngeal cancer development diet and noticed that laryngeal cancer occurrence was increased in patients with an insufficiency in nutrients, as well as vitamins. Most laryngeal cancers are squamous cell carcinomas. One of the main symptoms is the hoarseness of voice in patients with the disease. If mishandled, it may lead to the development of disease and may be related to airway blockage, pain, and dysphagia [3].

1.1. The Laryngeal Cancers Diagnosis Methods

Doctors will examine the neck and throat for swelling, lumps, or other unusual tumefactions. Below, we highlight the current diagnostic methods, one or more of which may be done on the patient:

- **Indirect laryngoscopy:** or reflective laryngoscopy: Using a light source and a laryngoscope, the examination is performed, a small circular mirror with a changeable diameter that involves a thin metal handle.
- **Direct laryngoscopy:** A laryngoscope is used; a tool shaped like an extended tube with a camera and light at the end. The laryngoscope may be rigid (telelaryngoscopy) or flexible (fibroscopy) [5].
- **Biopsy:** It is the process of removing a small-scale piece of tissue to mine for cancer cells. A rigid laryngoscope is mostly used to remove this piece of tissue under general anesthesia. The tissue is then examined under a microscope by a pathologist to check whether cancer cells are present. The only certain method to know if the abnormal region is cancer or not is a biopsy.
- **Imaging tests:** Imaging checks are not dependent on determining laryngeal and hypopharyngeal cancers; instead, they may be conducted in many cases, both before and after a cancer diagnosis. Imaging checks make use of magnetic fields or radioactive substances, and X-rays to produce photographs of the inside of the body [6].

1.2. Narrow-band Imaging (NBI)

Narrow-band imaging (NBI) endoscopy is commonly used to diagnose different laryngeal diseases. NBI depends on the depth of light breakthrough, which is based on the wavelength; the breakthrough depth increases with increasing wavelength. The NBI Technique uses a filter to adjust the wide-band white light released by the endoscope xenon light source [7]. NBI filters produce only blue (415 nm of wavelength) and green (540 nm of wavelength) light. A narrow band of blue light (415 nm) will show the superficial capillary vessel network, and a narrow band of green light (540 nm) will show the subcutaneous vessel. This technique allows the physician to evaluate the blood vessels in mucosal lesions, increasing the diagnostic accuracy of conventional laryngoscopy using high-resolution video [8]. Ni et al. compared the sensitivity and specificity of NBI and white-light imaging (sensitivity: 88.9% vs. 68.9% and specificity: 93.2% vs. 89.8%). NBI has evident advantages in guiding biopsy placement and the early diagnosis of laryngeal malignant diseases; however, it requires acquiring the skill necessary to review NBI checks, lengthy training, and extensive experience [7].

Standard diagnostic methods above are effective but often time-consuming, which can negatively impact the speed of treatment for patients. With the increasing need for faster and more accurate results, AI becomes a suitable option. AI technologies can analyze data quickly and efficiently, which contributes to accelerating the diagnosis process and improving its accuracy.

1.3. Overview of Machine Learning (ML) and Deep Learning (DL)

Machine Learning (ML) is not a modern term in computer science. However, it is an increasingly and highly developed field. ML teaches how to employ machines to simulate human behavior and ability. It takes into account a technique used to extract meanings from data and answer questions. It brings together computer science and statistics to allow machines or computers to perform a given task [9]. It is a branch of artificial intelligence that enables computers to learn new tasks and information independently without human intervention. This involves developing models and algorithms that enable computers to draw conclusions or predict patterns in data. Machine Learning has recently gained popularity due to its ability to handle data sets and make faster and more accurate decisions compared to traditional methods [10]. It has shown promising potential as an adjunctive tool to help physicians in stratifying the overall survival (OS) risk of different cancers, including laryngeal cancer [11]. Currently, commonly used ML techniques related to cancer prediction, such as RF, SVM, DT, LASSO, ANN, etc... Among them, the ML technique has been commonly used in cancer prognosis research, such as kidney cancer, lung cancer, laryngeal cancer, breast cancer, etc. [12].

Deep Learning (DL) is a well-known and commonly used part of ML. DL shares comparable properties with ML. Both notions are based on data classification and training models. DL is a part of ML, but very complex. DL enables computers with the capability to recognize and learn examples. DL needs to use the performance of high computing for data transactions and large datasets. DL uses various methods for training data based on data type [9]. Lately, many researchers have reported the application of deep learning models, as it is used in nuclear to solve classification problems [13]. Machine Learning algorithms are often classified into four principal categories, as outlined below:

- **Supervised Learning (SL):** The basic concept of this type is that the model is trained on categorized data and used to predict new data. It requires dividing data for two sets, including the training and testing sets.

In the beginning, the model has to be trained on the training set and then test its performance on a testing set [14]. This means in SL, each example includes a pair of input items and the desired output value. The system will be able to reliably identify the class labels for unseen instances in the best-case scenario. There are many SL algorithms to choose from, each with its own set of drawbacks and benefits. For all SL methods, no one learning algorithm works well [15].

- **Unsupervised Learning (UL):** It also involves training data, except that the target or categorized values are unknown. Thus, by finding the hidden pattern, machines try to cluster comparable types of data. Instead of prediction, the main goal of UL is to discover patterns. The efficiency of any model within UL cannot be evaluated if unknown or absent values of the labels are found. The algorithms involved in UL are Association Rule Mining, Dimensionality Reduction Techniques, and K-means clustering [14].
- **Semisupervised Learning (SSL):** This compelling technology consolidates unsupervised as well as supervised learning to develop the learning accuracy. By benefiting from plenty of unlabeled data, this technique handles a massive amount of bounded labeled data. SSL uses applications for various fields such as web content, speech recognition and analysis, classification of text documents, and protein sequence classification [16].
- **Reinforcement Learning (RL):** This kind uses previous data without a previously defined label to develop the upcoming processing of the active system. It refers to a learning method in which the agent reacts with the environment by discovering errors or rewards and producing actions. The operator is placed within a new environment to perform and learn via receiving some signals. These signals are able to reward or punish based on the operator's action. This allows the operator to modify its actions and learns best when upcoming actions are considered. The RL aims to maximize the rewards instead of pattern discovery or labeled data, just like the case of Supervised and Unsupervised Learning. The RL connections in a formal scope between the operator and its environment are stated in terms of signals, action, and state (punishment or reward) [9].

Clinically, AI is useful for the diagnosis of diseases, prognosis prediction, and determination of treatment direction. However, when building a highly accurate classifier, a large amount of data is initially needed. When data is preprocessed (such as data curation), securing and verification are important to build a model, and a generalization that achieves consistent produces across various applications must be guaranteed [17]. This is achieved using ML algorithms learning, including supervised algorithms such as Support Vector Machines (SVM), Decision Trees (DT), and Random Forests (RF). Additionally, DL techniques such as Convolutional Neural Networks (CNN) play a significant role in enhancing model performance and accuracy. Each of these algorithms is explained below:

- **Decision Tree (DT)** or Classification trees are a well-known and effective prediction approach and one of the oldest and most outstanding algorithms, ideal for classification tasks used in statistics and data mining. The approach of building a decision tree includes dividing the dataset used into datasets smaller than the previous one until a tree finds the decision leaf nodes that represent a decision or classification [16].
- **Random Forests (RF)** are a technique of ensemble learning and a supervised Machine Learning that is commonly applied for regression and classification tasks [16]. RF works as an ensemble and consists of multiple decision trees. For each decision tree within this classifier, a class prediction is fetched, so the class that gets the highest votes becomes the model’s prediction [18]. RF algorithm has various benefits, as well as its capability to address larger datasets, robustness, and stability against outliers and noise. However, it has some drawbacks, including high-level complexities that require supplemental resources, more execution time, a longer training period than other Machine Learning algorithms, and a delay in real-time prediction [16].
- **Support Vector Machine (SVM)** is a type of supervised learning algorithm that analyzes and arranges the data into one of the given classes. It is a technique of Machine Learning that deals with processing data for classification and regression analysis. SVM is preferred because it overcomes the dimensionality problem that arises during investigating our high-dimensional feature vector [18].
- **K-Nearest Neighbors (KNN)** is a supervised Machine Learning algorithm that classifies new data points with existing data points by comparing similarity measures (such as Euclidean distance). KNN is employed by getting the nearest neighbors to each data point to classify it using plurality voting. It is robust against

noise, and its accuracy is based on the quality of data. KNN is commonly used in pattern recognition and statistical estimation. It does not suppose any distribution of underlying data, since it is a non-parametric technique [16].

- **Convolutional Neural Networks (CNNs)**: In image preprocessing, CNNs represent a subset of the commonly used artificial neural networks (ANNs) [19]. The architecture of CNN represents four layers. The input layer possesses the image data. The second layer represents the convolution layer and is considered a core part of CNN. Besides the Pooling layer, which strives to reduce the sample sizes to decrease computation as well as memory troubles. The last layer, also called the Fully Connected layer, strives to find and determine the definitive output class [9].

2. Related Works

The following is an overview of earlier research that used ML and DL approaches to diagnose laryngeal cancer problems. **Table 1** summarizes these studies and includes the source number, publication year, topic, number of patients, techniques employed, and best accuracy attained.

- Tamashiro et al. [20] developed an AI-based diagnostic system using a deep learning approach for the early detection of laryngeal cancer using endoscopic images. Convolutional neural networks (CNNs) were used with a dataset from a hospital in Tokyo, 5430 images from 247 patients taken using two different techniques: white-light and narrow-band imaging. The model performed better with narrow-band images, achieving 69% **accuracy** and 85.6% **sensitivity**, compared to white-light images of the same patient, which achieved 61.8% **accuracy** and 70.1% **sensitivity**. A limitation of this research is the ratio of superficial to advanced tumors (202 vs. 45). This uneven distribution between the categories may bias the model towards superficial tumors more than advanced tumors, thus reducing its reliability due to poor generalizability.

- Esmaeili et al. [21] developed deep learning model developed using a cut-off-layer technique, merged with a fine-tuned ResNet50 architecture, was chosen as the DCNN design to achieve optimal performance to achieve optimal performance in classifying laryngeal lesions from 8181 narrow-band contact endoscopy (CE-NBI) images from 146 patients. The best-performing model achieved a test accuracy of 83.5%, with sensitivity and specificity of 87.3% and 76.1%, respectively. One limitation of this study, although it did not explicitly state the number of training and validation data sets, was the distribution between categories. This distribution revealed an imbalance between benign and malignant categories (5313 vs. 2868). This imbalance leads to a bias in the model towards the larger (benign) category, affecting its accuracy in detecting cancerous cases and thus weakening the model's performance and reliability.

- Wellenstein et al. [22] developed a deep learning algorithm specifically designed for detecting and locating laryngeal cancer during endoscopy. The YOLOv5 algorithm, in its various versions (YOLOv5s, YOLOv5m, and YOLOv5l), along with an ensemble model combining the YOLOv5s and YOLOv5m, was used as the primary algorithm for detecting and locating laryngeal cancer. The model was trained and validated using a dataset of 4,488 images, comprising a private hospital dataset in China and the open-source dataset Laryngoscope8. The ensemble model outperformed the other models, achieving a better balance between precision and recall (70% and 73%) compared to the YOLOv5s, YOLOv5m, and YOLOv5l models (0.77, 0.67), (0.82, 0.67), and (0.74,0.72), respectively. A limitation of this study is the lack of performance metrics for the model, such as accuracy, AUC, sensitivity, and specificity. Although the combined models performed better due to their combination of speed and accuracy, their overall performance remains average and requires further improvement.

- Yumii et al. [23] constructed a predictive model based on the use of radiomics analysis for the depth diagnosis of laryngopharyngeal cancer from NBI endoscopic images, where the Machine Learning algorithm, Least Absolute Shrinkage and Selection Operator (LASSO), was applied with a neural network containing ten hidden layers. A dataset from Hiroshima University in Japan was used, divided into a training set of 71 images and a validation set of 24 images. Their AI model demonstrated equivalent performance to that of skilled endoscopists, achieving an average AUC of 0.868 and an **accuracy** of 83.3% in predicting subepithelial invasion, with 87.3% **sensitivity** and 76.1% **specificity**. A limitation of this study is the small size of the dataset (only 95 images). In addition, there is a possibility of overfitting because the model may retain features specific to the sample instead of generalizing the public relations, thus achieving high performance on training data but poor performance on unseen data

- Li et al. [24] developed a Machine Learning model using Support Vector Machines (SVM) to predict the 5-year survival status of laryngeal squamous cell carcinoma (LSCC) patients. A clinical dataset of 150 cases (117 from the Zibo Center and 33 from the Beijing Center) was used, divided into a 60% training (90 cases) and a 40% testing (60 cases). The model accomplished 85.0% **accuracy**, 87.5% **sensitivity**, and 81.2% **AUC**, and was implemented as a practical online prediction platform for clinical use. A limitation of this study is the small dataset size (150 cases), which may weaken the model's generalizability. Due to the small sample size, the model learns specific characteristics unique to this group and may fail when applied to different or larger populations. In addition, there is a possibility of overfitting because the model may retain features specific to the sample instead of generalizing the public relations, thus achieving high performance on training data but poor performance on unseen data.

Table 1: An Overview of the previously listed studies.

Ind.	Reference	Year	Subject	Dataset size	Method of Classification	Outcomes
1.	Tamashiro et al.	2020	Pharyngeal Cancer (NBI)	5430 Images	DCNN	Sensitivity: 85.6%
2.	Esmaeili et al.	2021	Laryngeal Cancer	8181 Images	ResNet50	Accuracy: 83.5%
3.	Wellenstein et al.	2023	Laryngeal Carcinoma	4488 Images	YOLOv5(Ensemble YOLOs/YOLOm	Recall: 73.0%
4.	Yumii et al.	2024	Depth of Laryngopharyngeal Cancer	95 Images	LASSO	AUC: 0.868%
5.	Li et al.	2024	Laryngeal SCC Survival Prediction	150 Clinical data	SVM	Accuracy: 85%

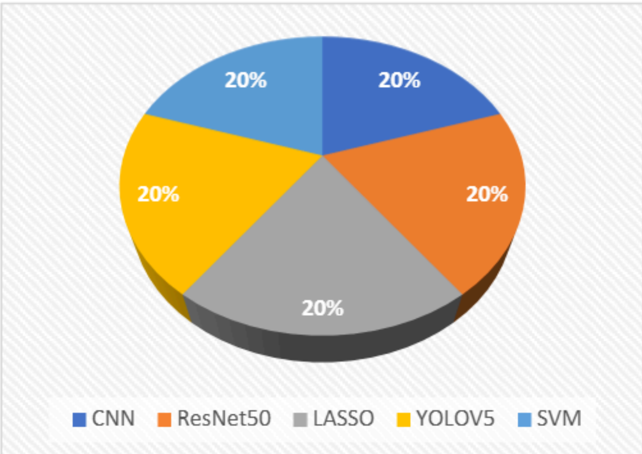

Fig. 1: Number of research papers published using ML and DL models for predicting laryngeal cancer.

3. Methodology

This section presents the methodologies previously used within the scope of our study:

- **Wu et al.** [25], relied on deep learning using Convolution Neural Networks (CNNs) as the primary technique for classifying laryngeal images into cancerous and normal. The study used a dataset of 10,760 images. The data was divided into an 80% training set (8,608 images) and a 20% test set (2,152 images). The following preprocessing was performed: resizing the images to a scaling size (224×224 pixels) to meet the requirements of the models used later; and employing data enhancement techniques, such as image translation and rotation, to avoid overfitting, generate additional images, and increase their usefulness. This study utilized several models: AlexNet, VGG16, ResNet18, and Inceptionv3, in addition to an improved model combining ResNet18 with Inceptionv3 to extract deep features from the ResNet18 model and then using Inceptionv3 to achieve more accurate classification.

- **Guo et al.** [26], conducted a retrospective study to develop a predictive model based on CT scans of 256 patients with laryngeal and hypopharyngeal cancer. Preprocessing steps included resampling the size of all images to ensure that extracted features were not affected by varying image resolutions. Several signal intensity transformation filters were used: exponential, square, logarithmic, and a Wavelet filter for extracting multiple features from images. The Least Absolute Shrinkage and Selection Operator (LASSO) was used as a feature selection technique to reduce dimensionality, select the most relevant features, avoid overfitting, and improve model efficiency. The study employed the following predictive models: a logistic regression model, and a logistic regression model combined with the Synthetic Minority Oversampling Technique (SMOTE) and Support Vector Machine (SVM) algorithms: LR-SVMSMOTE, to address the problem of data imbalance between categories.
- **Singh et al**. [27], aimed to classify laryngeal cancer stages with high accuracy using a set of CT scan data (560 images) combined with clinical data. The data was divided into 70% training, 15% validation, and 15% testing to prepare it for the preprocessing step, which included resizing the images to 224×224 pixels and augmentation the data through zoom in/out, shifts, random rotations, and horizontal and vertical flipping. For the clinical data, records were checked for null or missing values, and missing values were filled using either the mean or column median of the most frequently occurring value for the categorical data. The KNN algorithm was then used to classify the medical text after converting it to vectors. Finally, pre-trained deep learning models VGG19 and Resnet50 were used to extract features from the preprocessed medical images.
- **Xu et al.** [28], used deep convolutional neural networks, specifically the DenseNet201 architecture, to diagnose laryngeal cancer using a dataset of endoscopic laryngeal images of both benign and malignant cases. The dataset, consisting of 1029 images, was divided into an 80% training set (823 images) and a 20% test set (206 images) to prepare them for preprocessing before training. Preprocessing included resizing the images and converting them to a standardized size of 512×512 pixels and normalizing the pixel values to a defined range. The study employed several deep learning models: VGG19, Squeeznet1, ResNet152, MobileNet v3, MnasNet, InceptionV3, Alexnet, and DenseNet201.

4. Suggested Dataset

4.1. Laryngeal Cancer Dataset (LC)

The LC dataset was acquired from the Kaggle website [29]. We will utilize the larynx dataset, containing 1320 images of early-stage cancerous and healthy laryngeal tissues. The images (100 * 100) pixels were produced by hand from 33 narrow-band laryngoscopy data images collected from 33 patients affected with laryngeal spinocellular carcinoma (diagnosed by a histopathological laboratory test). The folder laryngeal dataset.rar contains three subfolders, each of which includes four folders related to the four tissue classes: He, Hbv, Le, and IPCL. Table 2 provides detailed information about the suggested dataset.

Table 2: Details of the Laryngeal Cancer Dataset.

Dataset	No.	Class	Resource
Laryngeal Cancer	1320	4	Kaggle

4.2. Laryngoscope8 Datasets

The dataset was acquired from the GitHub website [30]. The Laryngoscope8 dataset contains 3057 images of laryngeal specimens. These data were collected from the head and neck surgery operations in the Otorhinolaryngology Department/Sixth Medical Center/PLA General Hospital. 1950 patients are collected in this dataset, from which 1297 patients produced only one image of the larynx, while the residual of 653 patients produced more than two laryngeal images. All images within this dataset were classified into 8 categories given by hospital otolaryngologists. These categories include: Glottic Cancer, Granuloma, Reinke's Edema, Vocal Cord Nodules, Vocal Cord Leukoplakia, Vocal Cord Polyps, Vocal Cord Cyst, and Normal. Table 3 provides detailed information about this dataset.

Table 3: Details of the Laryngoscope8 Dataset.

Dataset	No.	Class	Resource
Laryngscope8	3057	8	GitHub

4.3. Preprocessing

Preprocessing is an important stage in data preparation for Machine Learning, focused on improving the quality of data and model performance [31]. For medical images, the image data size is far less than that of natural images. However, the quantity and quality of training data affect the final model of Machine Learning to a certain extent. So, before large-scale training, image pretreatment is necessary [32]. The most common preprocessing technique involves normalization, noise reduction, segmentation, and Augmentation (Rotation, flipping, and contrast adjustment).

5. Performance Metrics

Researchers evaluating the efficacy of prediction algorithms for laryngeal cancer have employed some measurements, such as Area Under the Curve (AUC), accuracy, F1-score, sensitivity, and specificity. Among these measures, AUC and accuracy are commonly used. Each of the above-mentioned performance measures, along with the corresponding mathematical formula, is explained as follows.

- **Accuracy:** Defined as the ratio of true cases obtained, both negative and positive, to all cases obtained. Accuracy is a weighted arithmetic average of precision and opposite precision; it is computed by the following equation (1) [33].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

  - True Positive (TP): indicates that an actual output is positive and the predicted result of the expected value is also positive.
  - True Negative (TN): indicates that the actual output is negative and the predicted result of the expected value is also negative.
  - False Positive (FP): indicates that the actual output is positive and the predicted result of the expected value is negative.
  - False Negative (FN): indicates that the actual output is negative and the predicted result of the expected value is positive [19].

- **Precision metric:** is calculated by dividing the Correct Positive Predictions (TP) for the class by the overall positive predictions (TP + FP) of the class. Knowing that the best accuracy is 1.0, while the worst is 0.0 [34]. It is computed by the following equation:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- **Specificity:** It is the number of correctly retrieved Negative cases divided by all correct cases [33]. The True Negative Rate (TNR), another name for this measure. It’s explained by the following equation:

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

- **Sensitivity (Recall):** It is calculated by dividing the number of accurate positive predictions (TP)for the class by the overall number of positive predictions (P) of the class. Also called the True Positive Rate (TPR). The best TP Rate is 1.0, and the worst is 0.0 [34]. It’s computed by the following equation:

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

- **F1-Score:** Is known as a weighted average for both precision and recall based on the β (weight function), as shown in equation 5.1. The F1-score refers to the harmonic meaning of precision and recall, as shown in equation 5.2. The F-score, which is also called as F-measure. The F1-score has various indices, which can give various weights of precision and recall as well [33].

$$F1 - Score: F\beta = (1+\beta^2) * \frac{(Precision*Recall)}{(Precision+Recall)} \quad (5.1)$$

With $\beta = 1$the standard F-score is obtained, as shown in Formula 5.2.

$$F1 - Score = 2 * \frac{(Precision*Recall)}{(Precision+Recall)} \quad (5.2)$$

- **Area Under the Receiver Operating Characteristic curve (AUC-ROC):** The Receiver Operating Characteristic Curve (ROC) evaluates the performance of a classification model by comparing the True Positive (TP) rate vs. the False Positive (FP) rate at various classification thresholds. The area under the ROC curve (AUC) refers to the total area under the curve, providing a measure of overall performance over all possible classification thresholds [19].

6. Discussion

5.1. Limitations

While Machine Learning has shown great promise in cancer prediction, several challenges remain. The availability and quality are one major issue with labeled datasets. Datasets of medical imaging often require expert annotation, which is time-consuming and costly. For laryngeal cancer, in particular, the availability of large, annotated datasets remains a limitation, which can impact the performance and generalizability of ML models.

Another challenge is model interpretability. Although deep learning models, such as CNNs, have proven highly accurate, they are often considered 'black boxes' due to the loss of transparency in their decision-making process. This can hinder clinical adoption, as medical professionals may be reluctant to rely on models whose rationale is unclear. Recent studies, such as those by Lundervold, A.S., and Lundervold, A. (2019), have explored ways to improve the understanding of Machine Learning (ML) models in healthcare settings, which will be critical for gaining trust in ML-based diagnostic tools.

Lastly, ensuring that ML models are generalized across different populations and medical facilities is another critical concern. When models are trained using data limited to a single organization or demographic group, they may not perform as well when handling data from different sources. This raises the need for multi-institutional collaborations and more diverse datasets to ensure robustness in real-world clinical applications.

5.2. Future Work

Future work aimed at predicting laryngeal cancer using AI models:

1. We plan to leverage a large, previously unused dataset.
2. Develop a system to predict a patient's survival after a specified period to help doctors understand their risk of developing the disease. This will rely on the combined use of imaging and clinical data.
3. Develop a system to predict disease progression before treatment to help doctors make appropriate decisions regarding a patient's health.

7. Conclusion

This survey highlights the prospects of Machine Learning and deep learning for improving laryngeal cancer diagnosis. This study also explores recent research related to Machine Learning and deep learning methods, including SVM as well as CNN techniques, which have demonstrated high-level accuracy for predicting and classifying some of the disease stages. The metrics that were considered for evaluating model performance for this research are accuracy, specificity, sensitivity, F1-score, and AUC. We hope that our future application will contribute to providing more accurate and effective solutions to the challenges associated with diagnosis and treatment, thereby improving patients' quality of life and reducing the burden of treatment procedures. The study examines the significant challenges of early detection with the aid of artificial intelligence, and our future efforts aim to expand the dataset size, utilize additional machine learning models, and employ feature selection techniques to further enhance the results.

REFERENCES

[1] A. Sahu *et al.*, "Imaging Recommendations for Diagnosis, Staging and Management of Larynx and Hypopharynx Cancer," Feb. 01, 2023, *Georg Thieme Verlag*. doi: 10.1055/s-0042-1759504.

[2] R. Akshara and T. P. Latchoumi, "A Survey: Identification of Throat Cancer byMachine Learning," 2021. [Online]. Available: http://annalsofrscb.ro6616

[3] M. Q. Ali, A. Raji, and M. S. Abdulameer, "Patterns of laryngeal cancer presentation of Iraqi patients," *Revista Latinoamericana de Hipertension*, vol. 16, no. 4, pp. 298–303, 2021, doi: 10.5281/zenodo.5812217.

[4] W. Wang *et al.*, "Creation of a machine learning-based prognostic prediction model for various subtypes of laryngeal cancer," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-56687-x.

[5] "Laryngoscopy | Quirónsalud." Accessed: Jun. 21, 2025. [Online]. Available: https://www.quironsalud.com/en/diagnostic-tests/laryngoscopy

[6] "Researched and Authored by Prof Michael C Herbst [D Litt et Phil (Health Studies); D N Ed; M Art et Scien; B A Cur; Dip Occupational Health; Dip Genetic Counselling; Diagnostic Radiographer; Dip Audiometry and Noise Measurement] Approved by Ms Elize Joubert, Chief Executive Officer [BA Social Work (cum laude); MA Social Work] Cancer Association of South Africa (CANSA) Fact Sheet on Cancer of the Larynx," 2021.

[7] Y. He *et al.*, "A deep convolutional neural network-based method for laryngeal squamous cell carcinoma diagnosis," *Ann Transl Med*, vol. 9, no. 24, pp. 1797–1797, Dec. 2021, doi: 10.21037/atm-21-6458.

[8] J. Galli *et al.*, "Role of narrow band imaging technology in the diagnosis and follow-up of laryngeal lesions: Assessment of diagnostic accuracy and reliability in a large patient cohort," *J Clin Med*, vol. 10, no. 6, pp. 1–10, Mar. 2021, doi: 10.3390/jcm10061224.

[9] H. Alaskar and T. Saba, "Machine Learning and Deep Learning: A Comparative Review," 2021, pp. 143–150. doi: 10.1007/978-981-33-6307-6_15.

[10] L. Deribe Abera, Y. Li, and S. Li, "OVERVIEW OF MACHINE LEARNING AND DEEP LEARNING," *www.irjmets.com @International Research Journal of Modernization in Engineering*, 1793, [Online]. Available: www.irjmets.com

[11] R. O. Alabi, A. Almangush, M. Elmusrati, I. Leivo, and A. A. Mäkitie, "Interpretable machine learning model for prediction of overall survival in laryngeal cancer," *Acta Otolaryngol*, 2024, doi: 10.1080/00016489.2023.2301648.

[12] W. Zhao *et al.*, "Construction of Prediction Model of Early Glottic Cancer Based on machine learning," Jan. 02, 2024. doi: 10.21203/rs.3.rs-3736160/v1.

[13] M. Nakajo *et al.*, "The usefulness of machine-learning-based evaluation of clinical and pretreatment 18F-FDG-PET/CT radiomic features for predicting prognosis in patients with laryngeal cancer," *British Journal of Radiology*, vol. 96, no. 1149, 2023, doi: 10.1259/bjr.20220772.

[14] A. Dhillon and A. Singh, "Biology and Today's World Machine Learning in Healthcare Data Analysis: A Survey," *J. Biol. Today's World*, vol. 8, no. 2, pp. 1–10, 2018, doi: 10.15412/J.JBTW.01070206.

[15] V. C. Gandhi and P. P. Gandhi, "A Survey - Insights of ML and DL in Health Domain," in *International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 239–246. doi: 10.1109/ICSCDS53736.2022.9760981.

[16] A. F. A. H. Alnuaimi and T. H. K. Albaldawi, "An overview of machine learning classification techniques," in *BIO Web of Conferences*, EDP Sciences, Apr. 2024. doi: 10.1051/bioconf/20249700133.

[17] 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*, Jan. 2022, doi: 10.1016/j.jvoice.2022.07.007.

[18] V. P. Singh and A. K. Maurya, "Role of Machine Learning and Texture Features for the Diagnosis of Laryngeal Cancer," in *Machine Learning for Healthcare Applications*, Wiley, 2021, pp. 353–367. doi: 10.1002/9781119792611.ch23.

[19] D. Sabeeh and Maalim A. Aljabery, "A Study on Cirrhosis Prediction Based on Machine Learning Techniques," *Journal of Education for Pure Science*, vol. 14, no. 4, Dec. 2024, doi: 10.32792/jeps.v14i4.461.

[20] A. Tamashiro *et al.*, "Artificial intelligence-based detection of pharyngeal cancer using convolutional neural networks," *Digestive Endoscopy*, vol. 32, no. 7, pp. 1057–1065, Nov. 2020, doi: 10.1111/den.13653.

[21] N. Esmaeili *et al.*, "Deep convolution neural network for laryngeal cancer classification on contact endoscopy-narrow band imaging," *Sensors*, vol. 21, no. 23, Dec. 2021, doi: 10.3390/s21238157.

[22] D. J. Wellenstein, J. Woodburn, H. A. M. Marres, and G. B. van den Broek, "Detection of laryngeal carcinoma during endoscopy using artificial intelligence," *Head Neck*, vol. 45, no. 9, pp. 2217–2226, Sep. 2023, doi: 10.1002/hed.27441.

[23] K. Yumii *et al.*, "Artificial intelligence-based diagnosis of the depth of laryngopharyngeal cancer," *Auris Nasus Larynx*, vol. 51, no. 2, pp. 417–424, Apr. 2024, doi: 10.1016/j.anl.2023.09.001.

[24] Z. Li, T. Li, P. Zhang, and X. Wang, "A practical online prediction platform to predict the survival status of laryngeal squamous cell carcinoma after 5 years," *American Journal of Otolaryngology - Head and Neck Medicine and Surgery*, vol. 45, no. 3, May 2024, doi: 10.1016/j.amjoto.2023.104209.

[25] F. Wu, P. Wu, Y. Hou, and H. Shang, "Neural Network for Image Classification of Laryngeal Cancer," in *Proceedings - 2021 International Conference on Networking Systems of AI, INSAI 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 239–243. doi: 10.1109/INSAI54028.2021.00051.

[26] R. Guo *et al.*, "CT-based radiomics features in the prediction of thyroid cartilage invasion from laryngeal and hypopharyngeal squamous cell carcinoma," *Cancer Imaging*, vol. 20, no. 1, Dec. 2020, doi: 10.1186/s40644-020-00359-2.

[27] J. Singh *et al.*, "Deep Learning-based Staging of Throat Cancer for Enhancing Diagnostic Accuracy Through Multimodal Data Integration," *Open Bioinforma J*, vol. 18, no. 1, Apr. 2025, doi: 10.2174/0118750362355440250203115520.

[28] Z. H. Xu, D. G. Fan, J. Q. Huang, J. W. Wang, Y. Wang, and Y. Z. Li, "Computer-Aided Diagnosis of Laryngeal Cancer Based on Deep Learning with Laryngoscopic Images," *Diagnostics*, vol. 13, no. 24, Dec. 2023, doi: 10.3390/diagnostics13243669.

[29] "Laryngeal dataset." Accessed: Apr. 23, 2025. [Online]. Available: https://www.kaggle.com/datasets/mahdiehhajian/laryngeal-dataset?resource=download

[30] "GitHub - greenyin/Laryngoscope8." Accessed: Apr. 23, 2025. [Online]. Available: https://github.com/greenyin/Laryngoscope8

[31] M. Paterson, J. Moor, and L. Cutillo, "Detecting Throat Cancer from Speech Signals using Machine Learning: A Scoping Literature Review," Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.09230

[32] H. Shang and P. Wu, "Semi-supervised deep learning for automatic classification of medical images," SPIE-Intl Soc Optical Eng, Jul. 2024, p. 12. doi: 10.1117/12.3032744.

[33] H. Dalianis, "Evaluation Metrics and Evaluation," in *Clinical Text Mining*, Springer International Publishing, 2018, pp. 45–53. doi: 10.1007/978-3-319-78503-5_6.

[34] Ž. Vujović, "Classification Model Evaluation Metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.