



Review

Artificial Intelligence Methods for Identifying and Localizing Abnormal Parathyroid Glands: A Review Study

Ioannis D. Apostolopoulos ^{1,*}, Nikolaos I. Papandrianos ², Elpiniki I. Papageorgiou ² and Dimitris J. Apostolopoulos ³

¹ Department of Medical Physics, School of Medicine, University of Patras, 26500 Patras, Greece

² Department of Energy Systems, Gaiopolis Campus, University of Thessaly, 41500 Larisa, Greece

³ Department of Nuclear Medicine, University General Hospital of Patras, School of Medicine, University of Patras, 26500 Patras, Greece

* Correspondence: ece7216@upnet.gr; Tel.: +30-697-386-6965

Abstract: Background: Recent advances in Artificial Intelligence (AI) algorithms, and specifically Deep Learning (DL) methods, demonstrate substantial performance in detecting and classifying medical images. Recent clinical studies have reported novel optical technologies which enhance the localization or assess the viability of Parathyroid Glands (PG) during surgery, or preoperatively. These technologies could become complementary to the surgeon's eyes and may improve surgical outcomes in thyroidectomy and parathyroidectomy. Methods: The study explores and reports the use of AI methods for identifying and localizing PGs, Primary Hyperparathyroidism (PHPT), Parathyroid Adenoma (PTA), and Multiglandular Disease (MGD). Results: The review identified 13 publications that employ Machine Learning and DL methods for preoperative and operative implementations. Conclusions: AI can aid in PG, PHPT, PTA, and MGD detection, as well as PG abnormality discrimination, both during surgery and non-invasively.



Citation: Apostolopoulos, I.D.; Papandrianos, N.I.; Papageorgiou, E.I.; Apostolopoulos, D.J. Artificial Intelligence Methods for Identifying and Localizing Abnormal Parathyroid Glands: A Review Study. *Mach. Learn. Knowl. Extr.* **2022**, *4*, 814–826. <https://doi.org/10.3390/make4040040>

Academic Editor: Andreas Holzinger

Received: 24 August 2022

Accepted: 20 September 2022

Published: 21 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: artificial intelligence; parathyroid glands; hyperparathyroidism; multigland disease; parathyroid adenoma; deep learning; machine learning

1. Introduction

A parathyroid adenoma (PTA) is a noncancerous (benign) tumor of the Parathyroid Glands (PGs). PGs are located in the neck, near or attached to the back side of the thyroid gland. PTA is part of a spectrum of parathyroid proliferative disorders that includes parathyroid hyperplasia, PTA, and parathyroid carcinoma [1].

Approximately eighty percent of primary hyperparathyroidism (PHPT) is caused by a PTA [2], followed by four-gland hyperplasia with ten to fifteen percent [2], and multiple adenomas with five percent [1].

Computer-Aided Diagnostic (CAD) assistance tools in PTA identification could significantly alleviate human tiredness and routine in everyday clinical practice, allowing the experts to put their efforts into nontrivial tasks. In addition, online surgical assisting tools that detect and localize important areas can aid in error prevention. Identification and preservation of the parathyroid glands (PGs) during thyroid surgery are very important. Damaging, devascularizing, autotransplanting, or inadvertent PGs can cause post-operative hypocalcemia. To this end, near infrared-induced autofluorescence (NIRAF) can deliver normal and pathologic PG localization in real time. Such tools are already embedded into modern image acquisition technologies and computer-enabled surgery frameworks. However, more modern solutions are worthy of examination; recent advances in Artificial Intelligence (AI) algorithms, specifically Deep Learning (DL) methods, demonstrate substantial performance in detecting and classifying medical images [2–4].

DL brought a revolution in feature-extraction from image data, enabling the computer-suggested capture of millions of potentially significant image features. DL algorithms can

learn to detect and distinguish important features that characterize an image according to a predefined label. For example, such methods have achieved remarkable success in various cancer-detection studies utilizing various imaging modalities [3–5]. DL implementations are also found in video processing and biomedical signal processing.

Recent clinical studies report novel optical technologies which enhance the localization or assess the viability of Parathyroid Glands (PG). These technologies could become complementary to the surgeon's eyes and may improve surgical outcomes in thyroidectomy and parathyroidectomy [6]. More importantly, combining such technologies with state-of-the-art image and video processing computational models can multiply the capabilities of these systems and greatly increase their necessity and utility in hospitals.

Non-invasive medical imaging acquisition modalities, such as SPECT, aid in the preoperative identification of hyperparathyroidism and abnormal PG localization. Again, AI methods can substantially contribute to the detection task and assist medical staff.

The present review investigates the implementation of AI for identifying and localizing abnormal PGs and PHPT. The Literature Review identifies 13 related papers from the year 2000 to July of 2022 and discusses their findings and methods. Current limitations and future suggestions are provided in the Discussion section.

2. Methods

2.1. Literature Review

The relevant publications were identified through extensive searches in approved publication-indexing websites and repositories. PubMed, Scopus, and Google Scholar were the major sources of information. Multiple keyword combinations were used to discover research papers and constitute the initial library, including:

- (Hyperparathyroidism OR Parathyroid Glands) AND (Deep Learning OR Artificial Intelligence)
- (Hyperparathyroidism OR Parathyroid Glands) AND (Convolutional Neural Networks OR Machine Learning)
- (Hyperparathyroidism or PHPT) AND (Deep Learning OR Artificial Intelligence)

The survey covered publications from January 2000 to July 2022. A total number of thirty-three publications constituted the initial library. Each publication's abstract and title were used to exclude irrelevant entries. Subsequently, a total of twelve research studies qualified for the review. The complete process is presented in Figure 1. This procedure identified 13 relevant papers which qualify for this review.

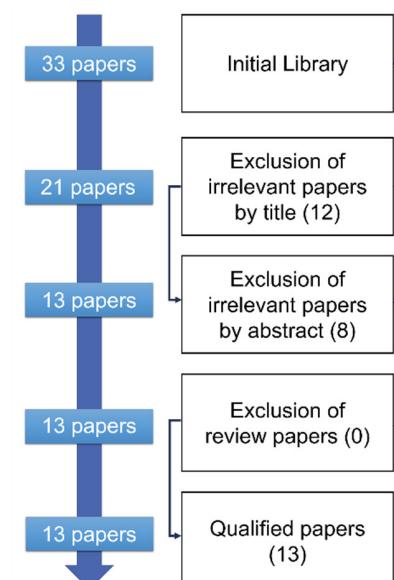


Figure 1. Literature Review methodology.

2.2. Machine Learning and Deep Learning in a Nutshell

This section describes the AI methods and algorithms reported in the literature review.

2.2.1. Machine Learning

ML is a part of AI [7]. It uses structured or unstructured data to learn patterns, forecast future values, or discover underlying knowledge [8]. The general idea of a machine that learns through a set of past observations is not an idea of our time [9]. The large amounts of data of any kind which are at the disposal of medical research centres and hospitals do not guarantee the successful development of an ML model. One of the most difficult challenges for engineers and programmers is labelling [10]. An ML model is commonly built upon a specific question or hypothesis to be investigated. For example, the malignancy suspiciousness rating of nodules inside specific organs and tissues in our body could be the focal point of ML methods. In essence, the medical dilemma of whether an observed nodule is malignant or benign is a likely domain for applying an ML model. We can assign varying levels of discretion to the methods and algorithms of any ML implementation concerning the focal point. Hence, we discuss supervised ML, unsupervised ML, and semi-supervised ML [10].

Supervised learning involves working with labelled datasets for training and testing. Every instance in the training data is accompanied by a specific and desired output/target, which is utilized by the algorithm in order to learn [11]. Examples of data where the desired output is known and their values are predefined are called labelled examples. In the case of parathyroid gland detection, the actual location of an image finding is considered to be the label of each instance. Based on this label, the ML model learns to identify patterns in the image related to this location. In a similar example, research may focus on distinguishing between normal and abnormal parathyroid images of various modalities (e.g., SPECT or histopathological images). In that case, the actual label of the image (normal or abnormal) is considered the ground truth.

Contrary to supervised learning, unsupervised learning utilizes unlabeled data, aiming to discover hidden patterns that group the data into clear and sufficiently demarcated sets [12]. Unsupervised ML can reveal new knowledge from data by analyzing the suggested patterns and performing cross-examination [13]. Unsupervised learning is often identical to data mining, a broader field aiming to discover patterns for data, deploying both ML and statistical or mathematical tools.

Dealing with labelled and unlabeled data is the objective of semi-supervised learning [14]. Not necessarily reliant upon discovering underlying patterns within the unlabeled data, this method instead focuses on discovering basic patterns from a set of labelled data and matching them with similar patterns of a set of unlabeled data [15]. Based on the confidence of the prediction, a certain amount of unlabeled data is incrementally incorporated into the labelled data to increase their size.

The most popular methods for medical tasks include Bayesian Networks [16], Decision Trees [17], Support Vector Machines [18], Regression models, Artificial Neural Networks (ANNs) [19], Genetic Algorithms [20], and Convolutional Neural Networks (CNNs) [21].

2.2.2. Deep Learning

DL alludes to various ML approaches utilizing many nonlinear processing units grouped by layers to process the input information by gradually applying specific transformations [22]. In the basic approach, the layers are usually sequentially connected. In essence, each layer processes the previous layer's output [23]. In this way, different levels of abstraction can acquire hierarchical representations of the input data. Special neural networks are utilized in DL's applications, which are related to image feature extraction.

Those networks are known as CNN, and their name comes from the convolution operation, which is the cornerstone of such methods. CNNs were introduced by LeCun [24]. CNN is a deep neural network that mainly uses convolution layers to extract useful information from the input data, usually feeding a final Fully Connected (FC) layer [25].

They exhibit impressive performance in a variety of tasks. A detailed explanation of the convolution process is presented in the next section. A convolution operation is performed as a filter, which is a table of weights slides throughout the input image. An output pixel produced at every position is a weighted sum of the input pixels (the pixels that the filter has passed from). The weights of the filter, as well as the size of the table (kernel), are constant for the duration of the scan. Therefore, convolutional layers can seize the shift-invariance of visible patterns and depict robust features.

After several convolutional and pooling layers, one or more FC layers may aim to perform high-level reasoning. FC layers connect all previous layers' neurons with every neuron of the FC layer.

The last layer of CNN is an output layer. The Softmax [26] operator is a common classifier for CNNs. A Support Vector Machine (SVM), usually combined with CNN features, is used. CNNs have been widely used for medical image classification [27–32].

3. Results

The review identified two major categories, namely, thyroidectomy-assisting methods for localizing PGs and preoperative PG detection and abnormality identification. Tables 1 and 2 summarize the type and the results of the reported 13 studies, respectively.

3.1. Thyroidectomy Assisting Methods for Localizing Parathyroid Glands

Early and precise detection of PGs is a challenging problem in thyroidectomy due to their small size and an appearance similar to that of surrounding tissues. Several AI methods have been designed and proposed to assist surgeons in localizing and identifying PGs. Recent literature fully uses emerging ML and DL algorithms to achieve high detection rates.

Kim et al. [33] introduced a prototype solution for the reduction of false-positive PGs localized using near-infrared autofluorescence (NIRAF) [34] methods. Their appliance is equipped with a coaxial excitation light (785 nm) and a dual-sensor. Under this setup, the authors employed the YOLO v5 [35] network, a real-time object detection DL model, to identify and localize PGs. The authors evaluated their solution's clinical feasibility in situ and ex vivo using sterile drapes on ten human subjects. Video data of 1287 images of well-visualized and localized PGs from six human subjects were utilized. This method yielded a mean average precision of 94.7% and a 19.5-millisecond processing time/detection. It is a matter for future research whether the proposed method remains at a top performance after the inclusion of more human participants.

Akbulut et al. [36] proposed a decision tree for intraoperative autofluorescence assessment of PGs in PHPT. The study involved 102 patients and 333 confirmed PGs. The authors extracted predictors from each PG, and the developed decision tree used normalized autofluorescence intensity, heterogeneity index, and gland volume to predict normal versus abnormal glands and subclasses of parathyroid pathologies. The algorithm achieved 95% accuracy in distinguishing between normal and abnormal PGs and 84% in predicting parathyroid pathologies' subclasses. However, the authors do not report the training and evaluation samples.

Wang et al. [37] benchmarked the YOLO V3, Faster R-CNN, and Cascade algorithms for identifying PGs during endoscopic approaches. The study involved 166 endoscopic thyroidectomy videos, of which 1700 images were employed (frames). The experiments revealed the superiority of Faster R-CNN in this task, which achieved precision, recall rate, and F1 scores of 88.7%, 92.3%, and 90.5%, respectively. The authors evaluated this network further using an independent external cohort of 20 videos. Senior and junior surgeons' visual estimation was used for comparisons. In this test set, the parathyroid identification rate of their method was 96.9%, while senior surgeons and junior surgeons achieved 87.5% and 71.9%, respectively.

Avci et al. [38] used the Google AutoML platform to identify an optimal DL model to localize parathyroid-specific autofluorescence on near-Infrared imaging. The study involved

466 intraoperative near-infrared images of 197 participants undergoing thyroidectomy or parathyroidectomy procedures. The study was split into three sets, training, validation, and test. 527 PG AF signals from the near-infrared images obtained intraoperatively from these procedures were used to develop the model's training set. The method yielded a recall of 90.5% and a precision of 95.7%, respectively. Those scores correspond to a 91.9% accuracy in detecting PGs.

Avci et al. [39] repeated the above study using a total of 906 intraoperative parathyroid autofluorescence images of 303 patients undergoing parathyroidectomy/thyroidectomy. The dataset was split, and 20% was kept for evaluation. The authors evaluated their models based on AUROC and AUPRC, which were found to be 0.9 and 0.93, respectively. Precision and recall were reported at 89% each.

Wang et al. [40] proposed an innovative method for identifying PGs based on laser-induced breakdown spectroscopy (LIBS). The study involved 1525 original spectra (773 PG spectra and 752 NPG spectra) from 20 smear samples of three rabbits. The authors extracted the emission lines related to K, Na, Ca, N, O, CN, and C2 and built several ML algorithms to distinguish between PGs and nPGs. The predictive attributes were ranked based on the importance weight calculated by Random Forest. The Artificial Neural Network model and the Random-Forest-based feature selection achieved a 92% accuracy.

3.2. Preoperative Parathyroid Gland Detection and Abnormality Identification

Sandqvist et al. [41] proposed an ensemble of decision trees with Bayesian hyperparameter optimization for predicting the presence of overlooked PTAs at a preoperative level using ^{99m}Tc -Sestamibi-SPECT/CT technology in Multiglandular Disease (MGD) patients. The authors used six predictors, namely, the preoperative plasma concentrations of parathyroid hormone, total calcium, and thyroid stimulating hormone, the serum concentration of ionized calcium, the 24-h urine calcium, and the histopathological weight of the localized PTA at imaging. The retrospective study involved 349 patients, whilst the dataset was split into 70% for training and 30% for testing. The authors designed their framework utilizing two response classes; patients with Single-Gland Disease (SGD) correctly localized at imaging and MGD patients in whom only one PTA was localized on imaging. Their algorithm achieved a 72% true positive prediction rate for MGD patients and a misclassification rate of 6% for SGD patients. This study confirmed that AI could aid in identifying patients with MGD for whom ^{99m}Tc -Sestamibi-SPECT/CT failed to visualize all PTAs.

Stefaniak [42] et al. developed an ANN to detect and locate pathological parathyroid tissue in the planar neck scintigrams. This study involved 35 participants. The detailed data consisted of sets of three single pixels, each belonging to one of the three consecutive neck scintigrams generated 20 min after $(^{99m}\text{TcO}_4)$ -administration, 10 min after $(^{99m}\text{Tc})\text{-MIBI}$ injection, and 120 min after $(^{99m}\text{Tc})\text{-MIBI}$ injection, respectively. The results of the ANN were compared to the conventional assessment of two radionuclide parathyroid examinations, namely, the subtraction method and $(^{99m}\text{Tc})\text{-MIBI}$ double-phase imaging. The ANN yielded a close relationship with the visual assessment of original neck scintigrams, with R square coefficient R^2 of 0.717 and standard error equal to 0.243 during its training. Multidimensional regression analysis yielded a weaker relationship, with an R^2 of 0.543 and a standard error of 0.567.

Yoshida et al. [43] employed RetinaNet [44], a DL network for the detection of PTA by parathyroid scintigraphy with $^{99m}\text{technetium sestamibi}$ ($^{99m}\text{Tc-MIBI}$) before surgery. The study enrolled 237 patients who took parathyroid scintigrams using $^{99m}\text{Tc-MIBI}$ and each of whom were determined to be a positive or negative case. Those patients' scans included 948 scintigraphy with 660 annotations, which were used for training and validation purposes. The test set included 44 patients (176 scintigrams and 120 annotations). The models' lesion-based sensitivity and mean false positive indications per image (mFPI) were assessed with the test dataset. The model yielded a sensitivity of 82%, with an mFPI

of 0.44 for the scintigrams of the early-phase model. For the delayed-phase model, the results reported 83% sensitivity and 0.31 mFPI.

Somnay et al. [45] employed several ML models for recognizing PHPT using clinical predictors, such as age, sex, and serum levels of preoperative calcium, phosphate, parathyroid hormone, vitamin D, and creatinine. The study enrolled 11,830 patients managed operatively at three high-volume endocrine surgery. Under a 10-fold cross-validation procedure, the Bayesian network was found superior to the rest of the ML models, achieving 95.2% accuracy and an AUC score of 0.989. This performance by the Bayesian network is interesting because, in general, such networks tend to overfit and their generalization capabilities are very limited.

Imbus et al. [46] benchmarked ML classifiers for predicting MGD in PHPT patients. The study involved 2010 participants (1532 patients with SGD and 478 with MGD). The fourteen predictor variables included patient demographic, clinical, and laboratory attributes. The boosted tree classifier was found superior to the rest ML modes, reaching an accuracy of 94.1%, a sensitivity of 94.1%, a specificity of 83.8%, a PPV of 94.1%, and an AUC score of 0.984.

Chen et al. [47] applied transfer learning for the automatic detection of PHPT from ultrasound images annotated by senior radiologists. The study involved 1000 ultrasound images containing PHPTs, of which 200 images were used to evaluate the developed model. For this purpose, they employed three well-established Convolutional Neural Networks to analyze the PHPT ultrasound and suggest potential features underlying the presence of PHPT. This study achieved the best recall, at 0.956.

In a recent work by Apostolopoulos et al. [48], the authors developed a three-path VGG19-based network to identify abnormal PGs in the early MIBI, late MIBI and TcO₄ thyroid scan images. The study includes 632 parathyroid scans (414 PG, 168 nPG). The proposed model, which is called ParaNet, exhibits top performance, reaching an accuracy of 96.56% in distinguishing between abnormal PGs and normal PGs scans. Its sensitivity and specificity are 96.38% and 97.02%, respectively. PPV and NPV values are 98.76% and 91.57%, respectively.

Table 1. Overview of the reviewed studies and their major outcomes. N.B.: PG: Parathyroid Gland, PTA: Parathyroid Adenoma, MGD: Multiglandular Disease, SGD: Single Gland Disease, PT: Parathyroid Tissue, PHPT: Primary Hyperparathyroidism, mFPI: mean false positive indications per image, mAP: mean average precision, PPV: positive predictive value, MR: misclassification rate.

Study	First Author	Year	Category	Aim	Major Findings
[33]	Kim	2022	Operative	PG detection	mAP: 94.7%
[41]	Sandqvist	2022	Preoperative	PTA detection	MGD-patients PPV: 72% SGD-patients MR: 6%
[42]	Stefaniak	2003	Preoperative	PTA detection	R ² of 0.543 and standard error of 0.567
[36]	Akbulut	2021	Operative	PG normal-abnormal classification and parathyroid pathology discrimination	PG normal-abnormal Accuracy: 95% Parathyroid pathology discrimination Accuracy: 84%
[37]	Wang	2022	Operative	PG identification	Precision: 88.7% Recall: 92.3% F1: 90.5%
[38]	Avci	2022	Operative	PG identification	Precision: 95.7% Recall: 90.5%

Table 1. Cont.

Study	First Author	Year	Category	Aim	Major Findings
[39]	Avci	2022	Operative	PG identification	Precision: 89% Recall: 89% AUC: 0.9
[43]	Yoshida	2022	Preoperative	PTA identification	Early Phase Sensitivity: 82% mFPI: 0.44 Delayed Phase Sensitivity: 83% mFPI: 0.31
[45]	Somnay	2017	Preoperative	PHPT recognition	Accuracy: 95.2% AUC: 0.989
[40]	Wang	2021	Operative	PG identification	Accuracy: 92%
[46]	Imbus	2017	Preoperative	MGD detection	Accuracy: 94.1% Sensitivity: 94.1% Specificity: 83.8% PPV: 94.1% AUC: 0.984
[47]	Chen	2020	Preoperative	PHPT detection	Recall: 96%
[48]	Apostolopoulos	2022	Preoperative	PG identification	Accuracy: 96.56% Sensitivity: 96.38% Specificity: 97.02% PPV: 98.76% NPV: 91.57%

Table 2. Key findings and experiment information of the presented literature.

Study	Method	Data Information	Major Findings
[33]	Deep Learning (YOLO v5)	Participants: 6 human subjects Classes: Not applicable Validation: 4 for training, 2 for testing Data Type: Video data (1287 images)	mAP: 94.7%
[41]	Machine Learning (Ensemble of Decision Trees)	Participants: 349 patients Classes: Patients with Single-Gland Disease (SGD) correctly localized at imaging and MGD patients in whom only one PTA was localized on imaging. Distribution between the two classes is not mentioned Validation: 70% for training and 30% for testing Data Type: Tabular—Six predictor variables	MGD-patients PPV: 72% SGD-patients MR: 6%
[42]	Machine Learning (ANN)	Participants: 35 patients Classes: Visually detectable Parathyroid Adenoma, probable Parathyroid Adenoma, background and/or outside body area, and thyroid gland. Distribution between the classes is not mentioned Validation: 25 for training, 10 for testing Data Type: Planar neck scintigrams	R ² of 0.543 and standard error of 0.567
[36]	Machine Learning (Ensemble of Decision Trees)	Participants: 333 PGs Classes: abnormal (n = 149) versus normal PGs (n = 184) Data Type: Tabular—Three predictor variables	PG normal-abnormal Accuracy: 95% Parathyroid pathology discrimination Accuracy: 84%

Table 2. Cont.

Study	Method	Data Information	Major Findings
[37]	Deep Learning (Faster R-CNN)	Participants: 166 endoscopic thyroidectomy videos 1700 images were employed (frames) Classes: Not applicable Validation: Training-validation ratio 15:2 20 full length videos were used as controls Data Type: Thyroidectomy videos	Precision: 88.7% Recall: 92.3% F1: 90.5%
[38]	Deep Learning (Google AutoML)	Participants: 466 intraoperative near-infrared images of 197 participants Classes: Not applicable Validation: 80% for training, 10% for validation, 10% for testing Data Type: Near-infrared images	Precision: 95.7% Recall: 90.5%
[39]	Deep Learning (Google AutoML)	Participants: 906 intraoperative parathyroid autofluorescence images of 303 participants Classes: 78 abnormal and 628 normal PG images Validation: 80% for training, 10% for validation, 10% for testing Data Type: Near-infrared images	Precision: 89% Recall: 89% AUC: 0.9
[43]	Deep Learning (Retina Net)	Participants: 281 patients Classes: Not applicable Validation: 192 for training, 45 for validation, 44 for testing Data Type: Early- and late-phase parathyroid scintigrams	Early Phase Sensitivity: 82% mFPI: 0.44 Delayed Phase Sensitivity: 83% mFPI: 0.31
[45]	Machine Learning (Bayesian Networks)	Participants: 11830 patients Classes: 6777 patients (study) with biochemical PHPT, 5053 patients without Validation: 10-fold cross-validation Data Type: Tabular—Clinical predictors	Accuracy: 95.2% AUC: 0.989
[40]	Machine Learning (ANN)	Participants: 1525 original spectra from 20 smear samples of three rabbits Classes: 773 PG spectra and 752 NPG spectra Validation: 3-fold cross-validation Data Type: Tabular—Clinical predictors	Accuracy: 92%
[46]	Machine Learning (Boosted Tree)	Participants: 2010 participants Classes: 1532 patients with Single Adenoma SGD and 478 with MGD Validation: 10-fold cross-validation Data Type: Tabular—14 predictor variables	Accuracy: 94.1% Sensitivity: 94.1% Specificity: 83.8% PPV: 94.1% AUC: 0.984
[47]	Deep Learning (CNN)	Participants: 1000 ultrasound images containing PHPTs Classes: Not mentioned Validation: 200 images (of the initial 1000) Data Type: Ultrasound images	Recall: 96%
[48]	Deep Learning (CNN)	Participants: 632 parathyroid scans Classes: PG (414 samples), nPG (168 samples) Validation: 10-fold cross-validation Data Type: Parathyroid scans	Accuracy: 96.56% Sensitivity: 96.38% Specificity: 97.02% PPV: 98.76% NPV: 91.57%

4. Discussion

The research study identified and described 13 studies addressing the issue of PG identification and localization, PHPT, PTA, and MGD detection. Most studies focus on PG detection (42%), while PG localization is addressed in 33% of the total studies.

There has been a significant amount of research conducted for preoperative delivery using ultrasound and scintigraphy image sources. Preoperative detection of abnormalities is also addressed using ML approaches without deploying any imaging modality. Significant clinical and demographical predictors are revealed in the literature, contributing to the diagnosis of PHPT and MGD. Overall, the preoperative delivery methods are introduced in 54% of the reviewed publications (Figure 2). The studies report very promising results in preoperative classification tasks, such as normal-abnormal image discrimination or MGD prediction using clinical factors. The observed sensitivity varies between 82 and 96 per cent. The majority of studies report an accuracy that ranges between 91 and 96 per cent. However, PG localization is not yet explored. It is expected that localizing each abnormal PG in thyroid scans would yield a number of false positive findings, thereby making this task very challenging.

The research community is also making efforts to provide novel appliances and topologies to improve the detection of findings during surgery. Most relevant publications accompany their technological solutions with traditional ML and DL approaches to enhance detection accuracy or to provide assisting computational tools. Studies presenting technological and AI solutions that deliver during surgery report better results regarding PG localization. It is observed that none of the reviewed research works integrates clinical factors and imaging data. It is expected that combining any available demographic, clinical, and biological data, where existent, would improve the diagnostic accuracy of image-based approaches and reduce the many reported false positive cases.

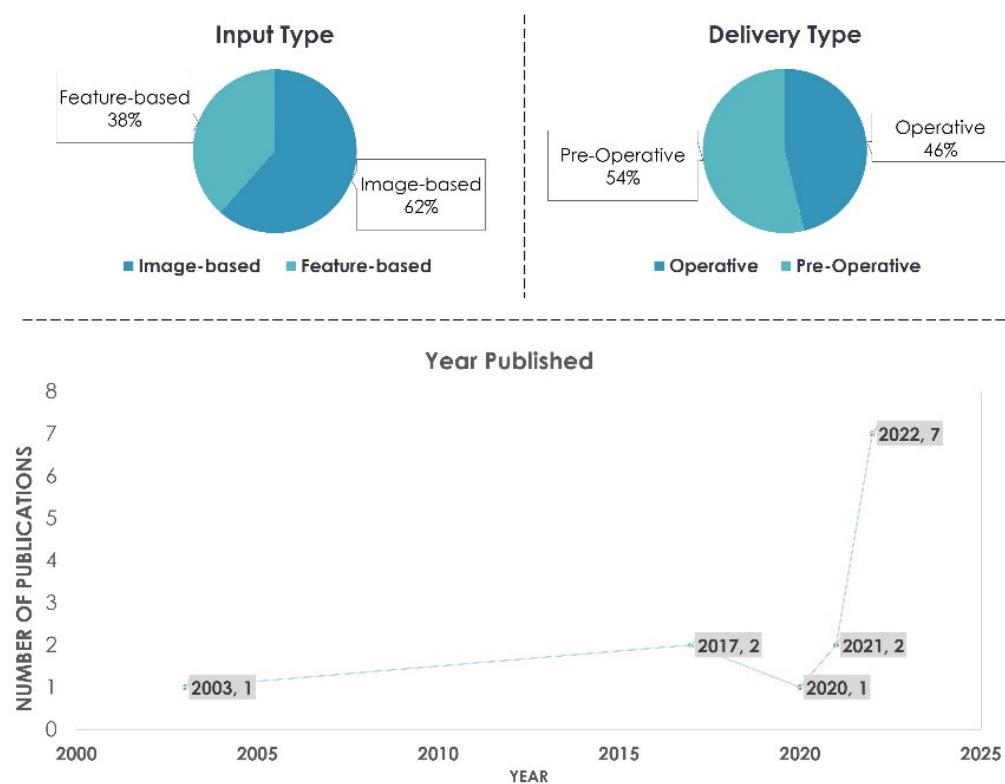


Figure 2. Analysis of the identified literature.

Despite their promising results, most studies use very few participants to train and evaluate their models. Most studies address this issue by extracting many video frames and slices from each patient. Therefore, the amount of samples is adequate for model training.

However, the datasets remain biased because the utilized frames/slices share the same origin. As a result, the study's results might be misleading. Still, there are studies that use more participants and report acceptable results and meaningful conclusions [41,46,48]. Most studies are validated on cohorts that do not exceed 500. As a result, the reported results, though undeniably encouraging, are not yet well-grounded. While the number of studies published peaks after 2021, the research on PG identification and localization, PHPT, PTA, and MGD detection is still constrained. The absence of publicly available data repositories covering relevant tasks impedes Biomedical Engineering experts from exploring the full potential of Artificial Intelligence in this domain. Nevertheless, the significant results reported in the literature undeniably open the horizons. Specifically, in PG detection and localization, the emergence of large-scale image datasets could accelerate the exploration of novel and state-of-the-art DL approaches and provide trustworthy solutions for medical assisting tools.

The emerging field of eXplainable Artificial Intelligence (XAI) as a set of algorithms and methods providing explanations can increase the medical importance and usefulness of AI methods in PHPT detection and PG abnormality discrimination. Most studies do not use explainable algorithms that inform the user of their decisions. As an example of explainable AI, the study of Imbus et al. [46] uses a decision tree for discriminating MGD from SGD. Decision trees are inherently self-explanatory. However, in studies where an ensemble of decision trees is employed (e.g., [36]), it is difficult to provide explanations. In studies where DL is employed (e.g., [43,48]), post-hoc explainability methods, such as the Grad-CAM algorithm [49], are not considered. Future studies could consider adopting explainable strategies to enhance their results and provide frameworks that are meaningful in everyday practice.

It was observed that many studies do not extensively report their methodology in terms of the employed ML and DL algorithms. Moreover, the majority of studies employ basic AI methods without mentioning any parameter tuning. For example, in studies where the decision trees are designed, the maximum number of leaf nodes and the maximum depth are not documented.

It is concluded that more effort should be put into designing and furnishing problem-specific models with well-grounded parameter selection. As an example of such methodology, in [41], the authors performed a Bayesian hyperparameter tuning, one which improved their results.

Finally, there is no established and documented method for validating the results. Some studies consider a train-test split solely, without any cross-validation method. This method is only suitable when large amounts of data are involved. In studies with few samples, partitioning the dataset at random may introduce biases. Other studies perform a cross-validation method (e.g., 3-fold, 10-fold) but do not consider control groups and external test sets. As a result, comparisons between studies are difficult.

Moreover, the robustness of the proposed pipelines regarding acquisition device variation is not explored. It is usual that different devices yield different image characteristics, e.g., resolution, pixel intensities, and video frames. Some variations regarding the models' effectiveness are expected and should be investigated.

5. Conclusions

This review study presented twelve works addressing the issue of PG identification and localization, HPT, PTA, and MGD detection. The reviewed studies were focused on both preoperative and operative solutions. Significant clinical and demographical predictors are revealed in the literature, contributing to the effective diagnosis of PHPT and MGD. Most relevant publications accompany their technological solutions with traditional ML and DL approaches to enhance the detection accuracy or to provide assisting computational tools. In the task of PG detection and localization, the emergence of large-scale image datasets could accelerate the exploration of novel and state-of-the-art DL approaches and provide trustworthy solutions for medical assisting tools. Moreover, explainable al-

gorithms must be introduced to enhance the results and increase the significance of the proposed methods.

Author Contributions: Conceptualization, I.D.A., N.I.P. and D.J.A.; methodology, I.D.A. and E.I.P.; formal analysis, D.J.A. and N.I.P.; investigation, D.J.A.; resources, D.J.A.; data curation, I.D.A., N.I.P. and E.I.P.; writing—original draft preparation, I.D.A.; writing—review and editing, I.D.A., N.I.P. and E.I.P.; supervision, D.J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Wieneke, J.A.; Smith, A. Parathyroid Adenoma. *Head Neck Pathol.* **2008**, *2*, 305–308. [[CrossRef](#)]
- Walker, M.D.; Silverberg, S.J. Primary Hyperparathyroidism. *Nat. Rev. Endocrinol.* **2018**, *14*, 115–125. [[CrossRef](#)] [[PubMed](#)]
- Astaraki, M.; Zakkou, Y.; Toma Dasu, I.; Smedby, Ö.; Wang, C. Benign-Malignant Pulmonary Nodule Classification in Low-Dose CT with Convolutional Features. *Phys. Med.* **2021**, *83*, 146–153. [[CrossRef](#)] [[PubMed](#)]
- Haggenmüller, S.; Maron, R.C.; Hekler, A.; Utikal, J.S.; Barata, C.; Barnhill, R.L.; Beltraminelli, H.; Berking, C.; Betz-Stablein, B.; Blum, A.; et al. Skin Cancer Classification via Convolutional Neural Networks: Systematic Review of Studies Involving Human Experts. *Eur. J. Cancer* **2021**, *156*, 202–216. [[CrossRef](#)]
- Lee, S.-Y.; Kang, H.; Jeong, J.-H.; Kang, D. Performance Evaluation in [18F]Florbetaben Brain PET Images Classification Using 3D Convolutional Neural Network. *PLoS ONE* **2021**, *16*, e0258214. [[CrossRef](#)] [[PubMed](#)]
- Abbaci, M.; De Leeuw, F.; Breuskin, I.; Casiraghi, O.; Lakhdar, A.B.; Ghanem, W.; Laplace-Buillé, C.; Hartl, D. Parathyroid Gland Management Using Optical Technologies during Thyroidectomy or Parathyroidectomy: A Systematic Review. *Oral. Oncol.* **2018**, *87*, 186–196. [[CrossRef](#)]
- Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. Machine Learning Basics. *Deep. Learn.* **2016**, 98–164. Available online: http://whdeng.cn/Teaching/PPT_01_Machine%20learning%20Basics.pdf (accessed on 23 August 2022).
- Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 160. [[CrossRef](#)]
- Denton, E.; Hanna, A.; Amironesei, R.; Smart, A.; Nicole, H. On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet. *Big Data Soc.* **2021**, *8*, 205395172110359. [[CrossRef](#)]
- Jiang, T.; Gradus, J.L.; Rosellini, A.J. Supervised Machine Learning: A Brief Primer. *Behav. Ther.* **2020**, *51*, 675–687. [[CrossRef](#)]
- Sen, P.C.; Hajra, M.; Ghosh, M. Supervised Classification Algorithms in Machine Learning: A Survey and Review. In *Emerging Technology in Modelling and Graphics*; Mandal, J.K., Bhattacharya, D., Eds.; Advances in Intelligent Systems and Computing; Springer: Singapore, 2020; Volume 937, pp. 99–111. ISBN 9789811374029.
- Alloghani, M.; Al-Jumeily, D.; Mustafina, J.; Hussain, A.; Aljaaf, A.J. A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In *Supervised and Unsupervised Learning for Data Science*; Berry, M.W., Mohamed, A., Yap, B.W., Eds.; Unsupervised and Semi-Supervised Learning; Springer International Publishing: Cham, Switzerland, 2020; pp. 3–21. ISBN 978-3-030-22474-5.
- Berry, M.W.; Mohamed, A.; Yap, B.W. (Eds.) *Supervised and Unsupervised Learning for Data Science*; Unsupervised and Semi-Supervised Learning; Springer International Publishing: Cham, Switzerland, 2020; ISBN 978-3-030-22474-5.
- Hadji, M.F.A.; Schwenker, F. Semi-Supervised Learning. In *Handbook on Neural Information Processing*; Bianchini, M., Maggini, M., Jain, L.C., Eds.; Intelligent Systems Reference Library; Springer: Berlin/Heidelberg, Germany, 2013; Volume 49, pp. 215–239. ISBN 978-3-642-36656-7.
- Van Engelen, J.E.; Hoos, H.H. A Survey on Semi-Supervised Learning. *Mach. Learn.* **2020**, *109*, 373–440. [[CrossRef](#)]
- Marcot, B.G.; Penman, T.D. Advances in Bayesian Network Modelling: Integration of Modelling Technologies. *Environ. Model. Softw.* **2019**, *111*, 386–393. [[CrossRef](#)]
- Kotsiantis, S.B. Decision Trees: A Recent Overview. *Artif. Intell. Rev.* **2013**, *39*, 261–283. [[CrossRef](#)]
- Suthaharan, S. Support Vector Machine. In *Machine Learning Models and Algorithms for Big Data Classification*; Integrated Series in Information Systems; Springer: Boston, MA, USA, 2016; Volume 36, pp. 207–235. ISBN 978-1-4899-7640-6.
- Li, H.; Zhang, Z.; Liu, Z. Application of Artificial Neural Networks for Catalysis: A Review. *Catalysts* **2017**, *7*, 306. [[CrossRef](#)]
- Kramer, O. Genetic Algorithms. In *Genetic Algorithm Essentials*; Studies in Computational Intelligence; Springer International Publishing: Cham, Switzerland, 2017; Volume 679, pp. 11–19. ISBN 978-3-319-52155-8.
- LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
- Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
- LeCun, Y.; Bengio, Y. Others Convolutional Networks for Images, Speech, and Time Series. *Handb. Brain Theory Neural Netw.* **1995**, *3361*, 1995.

24. LeCun, Y.; Kavukcuoglu, K.; Farabet, C. Convolutional networks and applications in vision. In Proceedings of the 2010 IEEE International Symposium on Circuits and Systems, Paris, France, 30 May–2 June 2010; IEEE: Paris, France, 2010; pp. 253–256.
25. Affonso, C.; Rossi, A.L.D.; Vieira, F.H.A.; de Leon Ferreira, A.C.P. Others Deep Learning for Biological Image Classification. *Expert Syst. Appl.* **2017**, *85*, 114–122. [[CrossRef](#)]
26. Liu, W.; Wen, Y.; Yu, Z.; Yang, M. Large-Margin Softmax Loss for Convolutional Neural Networks. In Proceedings of the ICML, New York, NY, USA, 19–24 June 2016; Volume 2, p. 7.
27. Apostolopoulos, I.D.; Aznaouridis, S.I.; Tzani, M.A. Extracting Possibly Representative COVID-19 Biomarkers from X-ray Images with Deep Learning Approach and Image Data Related to Pulmonary Diseases. *J. Med. Biol. Eng.* **2020**, *40*, 462–469. [[CrossRef](#)]
28. Apostolopoulos, I.D.; Papathanasiou, N.D.; Spyridonidis, T.; Apostolopoulos, D.J. Automatic characterization of myocardial perfusion imaging polar maps employing deep learning and data augmentation. *Hell. J. Nucl. Med.* **2020**, *23*, 125–132.
29. Apostolopoulos, I.D.; Apostolopoulos, D.I.; Spyridonidis, T.I.; Papathanasiou, N.D.; Panayiotakis, G.S. Multi-input deep learning approach for Cardiovascular Disease diagnosis using Myocardial Perfusion Imaging and clinical data. *Phys. Med.* **2021**, *84*, 168–177. [[CrossRef](#)] [[PubMed](#)]
30. Apostolopoulos, I.D.; Pintelas, E.G.; Livieris, I.E.; Apostolopoulos, D.J.; Papathanasiou, N.D.; Pintelas, P.E.; Panayiotakis, G.S. Automatic classification of solitary pulmonary nodules in PET/CT imaging employing transfer learning techniques. *Med. Biol. Eng. Comput.* **2021**, *59*, 1299–1310. [[CrossRef](#)]
31. Papandrianos, N.I.; Feleki, A.; Moustakidis, S.; Papageorgiou, E.I.; Apostolopoulos, I.D.; Apostolopoulos, D.J. An Explainable Classification Method of SPECT Myocardial Perfusion Images in Nuclear Cardiology Using Deep Learning and Grad-CAM. *Appl. Sci.* **2022**, *12*, 7592. [[CrossRef](#)]
32. Papandrianos, N.I.; Apostolopoulos, I.D.; Feleki, A.; Apostolopoulos, D.J.; Papageorgiou, E.I. Deep Learning Exploration for SPECT MPI Polar Map Images Classification in Coronary Artery Disease. *Ann. Nucl. Med.* **2022**, *36*, 823–833. [[CrossRef](#)] [[PubMed](#)]
33. Kim, Y.; Lee, H.C.; Kim, J.; Oh, E.; Yoo, J.; Ning, B.; Lee, S.Y.; Ali, K.M.; Tufano, R.P.; Russell, J.O.; et al. A coaxial excitation, dual-red-green-blue/near-infrared paired imaging system toward computer-aided detection of parathyroid glands in situ and ex vivo. *J. Biophotonics* **2022**, *15*, e202200008. [[CrossRef](#)]
34. Solórzano, C.C.; Thomas, G.; Baregamian, N.; Mahadevan-Jansen, A. Detecting the Near Infrared Autofluorescence of the Human Parathyroid: Hype or Opportunity? *Ann. Surg.* **2020**, *272*, 973–985. [[CrossRef](#)] [[PubMed](#)]
35. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo Algorithm Developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [[CrossRef](#)]
36. Akbulut, S.; Erten, O.; Kim, Y.S.; Gokceimam, M.; Berber, E. Development of an Algorithm for Intraoperative Autofluorescence Assessment of Parathyroid Glands in Primary Hyperparathyroidism Using Artificial Intelligence. *Surgery* **2021**, *170*, 454–461. [[CrossRef](#)] [[PubMed](#)]
37. Wang, B.; Zheng, J.; Yu, J.; Lin, S.; Yan, S.; Zhang, L.; Wang, S.; Cai, S.; Abdelhamid Ahmed, A.H.; Lin, L.; et al. Development of Artificial Intelligence for Parathyroid Recognition During Endoscopic Thyroid Surgery. *Laryngoscope* **2022**. [[CrossRef](#)]
38. Avci, S.N.; Isiktas, G.; Berber, E. A Visual Deep Learning Model to Localize Parathyroid-Specific Autofluorescence on Near-Infrared Imaging: Localization of Parathyroid Autofluorescence with Deep Learning. *Ann. Surg. Oncol.* **2022**, *29*, 4248–4252. [[CrossRef](#)]
39. Avci, S.N.; Isiktas, G.; Ergun, O.; Berber, E. A Visual Deep Learning Model to Predict Abnormal versus Normal Parathyroid Glands Using Intraoperative Autofluorescence Signals. *J. Surg. Oncol.* **2022**, *126*, 263–267. [[CrossRef](#)]
40. Wang, Q.; Xiangli, W.; Chen, X.; Zhang, J.; Teng, G.; Cui, X.; Idrees, B.S.; Wei, K. Primary Study of Identification of Parathyroid Gland Based on Laser-Induced Breakdown Spectroscopy. *Biomed. Opt. Express* **2021**, *12*, 1999. [[CrossRef](#)]
41. Sandqvist, P.; Sundin, A.; Nilsson, I.-L.; Grybäck, P.; Sanchez-Crespo, A. Primary Hyperparathyroidism, a Machine Learning Approach to Identify Multiglandular Disease in Patients with a Single Adenoma Found at Preoperative Sestamibi-SPECT/CT. *Eur. J. Endocrinol.* **2022**, *187*, 257–263. [[CrossRef](#)]
42. Stefaniak, B.; Cholewiński, W.; Tarkowska, A. Application of Artificial Neural Network Algorithm to Detection of Parathyroid Adenoma. *Nucl. Med. Rev.* **2003**, *6*, 111–117.
43. Yoshida, A.; Ueda, D.; Higashiyama, S.; Katayama, Y.; Matsumoto, T.; Yamanaga, T.; Miki, Y.; Kawabe, J. Deep Learning-Based Detection of Parathyroid Adenoma by ^{99m}Tc -MIBI Scintigraphy in Patients with Primary Hyperparathyroidism. *Ann. Nucl. Med.* **2022**, *36*, 468–478. [[CrossRef](#)]
44. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
45. Somnay, Y.R.; Craven, M.; McCoy, K.L.; Carty, S.E.; Wang, T.S.; Greenberg, C.C.; Schneider, D.F. Improving Diagnostic Recognition of Primary Hyperparathyroidism with Machine Learning. *Surgery* **2017**, *161*, 1113–1121. [[CrossRef](#)]
46. Imbus, J.R.; Randle, R.W.; Pitt, S.C.; Sippel, R.S.; Schneider, D.F. Machine Learning to Identify Multigland Disease in Primary Hyperparathyroidism. *J. Surg. Res.* **2017**, *219*, 173–179. [[CrossRef](#)] [[PubMed](#)]
47. Chen, J.; Guo, Q.; Jiang, Z.; Wang, H.; Yu, M.; Wei, Y. Recognition of Hyperparathyroidism Based on Transfer Learning. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea, 16–19 December 2020; IEEE: Piscataway, NJ, USA; pp. 2959–2961.

48. Apostolopoulos, I.D.; Papathanasiou, N.D.; Apostolopoulos, D.J. A Deep Learning Methodology for the Detection of Abnormal Parathyroid Glands via Scintigraphy with ^{99m}Tc -Sestamibi. *Diseases* **2022**, *10*, 56. [[CrossRef](#)]
49. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 618–626.