Published in final edited form as:

J Surg Res. 2017 November ; 219: 173–179. doi:10.1016/j.jss.2017.05.117.

Machine Learning to Identify Multigland Disease in Primary Hyperparathyroidism

J.R. Imbus¹, R.W. Randle¹, S.C. Pitt¹, R.S. Sippel¹, and D.F. Schneider¹

¹University of Wisconsin, Department of Surgery, Madison, WI, USA

Abstract

Background—20–25% of patients with primary hyperparathyroidism (PHPT) will have multigland disease (MGD). Preoperative imaging can be inaccurate or unnecessary in MGD. Identification of MGD could direct the need for imaging and inform operative approach. The purpose of this study is to use machine learning (ML) methods to predict MGD.

Methods—Retrospective review of a prospective database. The ML platform, WEKA, was used and we selected models for 1. overall accuracy and 2. preferential identification of MGD. A review of imaging studies was performed on a cohort predicted to have MGD.

Results—2010 patients met inclusion criteria: 1532 patients had single adenoma (SA) (76%) and 478 had MGD (24%). After testing many algorithms, we selected two different models for potential integration as clinical decision-support tools. The best overall accuracy was achieved using a boosted tree classifier, *RandomTree*: 94.1% accuracy; 94.1% sensitivity, 83.8% specificity, 94.1% positive predictive value, 0.984 Area Under the Receiver Operating Characteristics curve (AUC). To maximize positive predictive value of MGD prediction, a rule-based classifier, *JRip*, with cost-sensitive learning was used and achieved 100% PPV for MGD. Imaging reviewed from the cohort of 34 patients predicted to have MGD by the cost-sensitive model revealed 39 total studies performed: 28 sestamibi scans and 11 ultrasounds. Only 8 (29%) sestamibi scans and 4 (36%) ultrasounds were correct.

Conclusion—ML methods can help distinguish MGD early in the clinical evaluation of PHPT, guiding further workup and surgical planning.

Keywords

parathyroid glands; primary hyperparathyroidism; machine learning; clinical decision support; diagnosis; multigland disease

Corresponding author: Joseph Imbus, MD, imbus@wisc.edu, Phone: 513-675-1182, Address: 2901 Stevens Street, Madison, WI 53705.

Department of Surgery, University of Wisconsin, 600 Highland Avenue, K4/728 CSC, Madison, WI 53792

Author contributions: Joseph Imbus is the first author with the following contributions: drafting of the manuscript, data collection, data analysis. Reese Randle contributed to editing and critical revision of the manuscript. Susan Pitt contributed to methodology and manuscript editing. Rebecca Sippel contributed to methodology and manuscript editing. David Schneider is the senior author and responsible for study design and conception, drafting of the manuscript, and critical revision with editing.

Disclosure: The authors report no proprietary or commercial interest in any product mentioned or concept discussed in this article. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Introduction

Primary hyperparathyroidism (PHPT) is a complicated disease process with a diverse array of clinical presentations which affects over 100,000 new patients in the United States every year(1). Collectively, PHPT is attributed to abnormal autogenous parathyroid hormone (PTH) secretion. Although there are four parathyroid glands, the disease manifests with a single adenoma (SA) in 80% of cases. The remaining 20% of cases occur as double adenoma (DA), or four-gland hyperplasia(2–5). The term multigland disease (MGD) describes PHPT from DA or hyperplasia.

Surgical resection of the abnormal parathyroid gland(s) is the only cure for this increasingly diagnosed condition(6, 7). The last 20 years has seen a change in operative approach for PHPT. Minimally invasive parathyroidectomy has been popularized as an effective alternative to traditional four gland exploration for PHPT, preoperatively identifying a SA with imaging and ruling out MGD with intraoperative PTH (IoPTH) measurement. However, when MGD is known or suspected preoperatively, planned bilateral neck exploration with identification of all four parathyroid glands by an experienced surgeon remains a safe and extremely successful approach (8–10).

To date, researchers have been unsuccessful in attempting to definitively distinguish SA from MGD based on biochemical and clinical factors alone. Hereditary factors, prior lithium therapy, and a history of head and neck irradiation are the only known risk factors for MGD, but these factors only account for a small minority of patients with PHPT(11, 12). Surgeons typically use sestamibi scans, neck ultrasonography, and four-dimensional computed tomography as the primary imaging modalities to attempt localization of the abnormal gland(s). However, these imaging studies are significantly less accurate in the case of MGD(13, 14). Furthermore, these studies may falsely indicate SA, misleading the surgeon to attempt a minimally invasive surgery. The prevalence of unrecognized MGD in patients with positively localized imaging has been estimated to be as high as 22%(15). These cases often result in conversion to a bilateral exploration after an unnecessary delay waiting for IoPTH results. The literature also suggests that despite appropriate use of adjunctive imaging and IoPTH, some occult but pathologic glands go unrecognized at surgery(16).

Unfortunately, imaging tests are often ordered to inform or confirm diagnosis of PHPT rather than for operative planning(17). Experts agree that the diagnosis of PHPT is based on biochemical testing whereas imaging assists with operative planning(18). The overutilization of medical imaging in the United States is a recognized problem at the national level, contributing to unnecessary healthcare costs(19). Developing decision support tools for physicians to use at the point of care may help curb superfluous imaging(20). In PHPT, the challenge is to identify the subset of patients who may have MGD where imaging will not be as helpful. In addition, preoperative identification of MGD may reduce operative time spent waiting for IoPTH levels since surgeons could immediately perform a bilateral exploration.

Machine learning (ML) is a growing field with many different applications. Broadly, ML is a collection of methods that allow a computer to learn rules from known/existing datasets to make predictions. In clinical medicine, labeled examples of an outcome (i.e. cancer vs.

benign) can be used by ML algorithms to predict or classify future outcomes when given new data.(21) A number of different ML algorithms now exist, each with different underlying mathematical methods, ranging from decision rules, tree-based learning, and Bayesian calculations.(22) Human clinical decision making is prone to error and bias(23). The strength of ML is the ability to synthesize complex combinations of numerous variables in a short period of time, often in a non-linear fashion(24). For this study, we hypothesized that ML methods could be used to distinguish MGD from SA preoperatively, thereby streamlining workup and operative planning.

Methods

Subjects and Study Design

We performed a retrospective review of our institution's IRB-approved prospectively collected database containing patients with surgically treated hyperparathyroidism. This database was queried for patients eighteen years and older who underwent an initial parathyroidectomy for PHPT between 2001 and 2016. Clinical data was reviewed and patients with a history of lithium use, prior neck radiation exposure, prior neck surgery, secondary/tertiary hyperparathyroidism, concomitant thyroid disease, and hereditary disorders were excluded. Cases which involved concomitant thyroid resection and cases which did not result in cure were excluded. Cure was defined as normal serum calcium level at six months post-operatively. In addition, patients with a final pathology result of parathyroid carcinoma were excluded.

Data Selection and Machine Learning Approach

For our study, we labeled the final cohort as SA (one gland removed) and MGD (> 1 gland removed). Predictor variables included patient demographic, clinical, and laboratory attributes. We utilized the ML platform Waikato Environment for Knowledge Analysis (version 3.8.0, Weka).

Model Selection and Cost-Sensitive Learning

We tested several different categories of supervised ML classifiers including Bayesian network, rule-based, tree, and regression-based classifiers. An automated method was used to determine optimal parameter tuning(25). We decided to create two separate classifier configurations: 1. A ML model to achieve the highest overall accuracy in prediction of SA vs MGD and 2. A ML model to achieve the highest positive predictive value (PPV) for MGD.

We chose to optimize MGD PPV because this measure will be most useful in the clinical setting. A high PPV minimizes false positives, and a MGD prediction under these conditions truly represents patients who need minimal imaging and exploration of all four glands. Attempting to optimize the prediction of a minority class is not innate to ML methods, but adjustments to the training process can take the cost of misclassification into account. Cost-sensitive learning uses weight adjustments to make the decision structure biased towards avoiding one type of error(26). We used a cost-sensitive learning scheme which effectively

penalized false positive predictions of MGD three times more heavily than false negative predictions.

Imaging Studies

The number and correctness of sestamibi scans and ultrasounds performed for parathyroid localization were evaluated. We considered a sestamibi scan correct if the gland removed was on the same side as the gland identified by sestamibi. We considered an ultrasound correct if the gland removed was the same gland (laterality and upper vs. lower) identified on ultrasound. A negative ultrasound or sestamibi scan was correct only if the patient was found to have three or more enlarged glands. Since the purpose of this study was to evaluate an algorithm that distinguishes SA from MGD using biochemical and clinical data alone, we did not evaluate the concordance of imaging for this cohort.

Statistical Analysis

Classifiers were evaluated for generalization performance by 10-fold cross validation. Other measures included overall accuracy, sensitivity (recall), specificity (1 – false positive rate), PPV (recall), and area under the receiver operating characteristic (ROC) curve. Aside from the ML analysis stated above, descriptive statistics were generated using SAS software, Version 9.4 (Copyright © 2013 SAS Institute Inc. Cary, NC, USA). Continuous variables are represented by their median $\pm$ standard deviation, and comparisons were made using Student *t* test. Categorical variables are represented as their number (percentage), and the Pearson χ^2 test was used for statistical comparisons. A P value of 0.05 or less defined statistical significance.

Results

Patient Characteristics

A total of 2010 patients with PHPT met inclusion criteria - 1532 (76%) had SA and 478 (24%) had MGD. Characteristics of each of these groups are shown in Table I. Compared to the SA group, patients with MGD had lower pre-operative calcium levels (10.4 ± 0.86 vs. 10.9 ± 0.83 , $p < 0.001$), lower pre-operative urinary calcium levels (243 ± 157 vs 300 ± 209), lower pre-operative PTH levels (89 ± 78 vs. 105 ± 103 , $p < 0.001$), and lower pre-operative vitamin D levels (33 ± 14.7 vs. 30 ± 12.9 , $p < 0.001$). 121 (6%) of all patients had a pre-operative vitamin D level less than 15 ng/mL. The MGD cohort was also more likely to have a history of nephrolithiasis (21% vs 16.5%, 0.016). Median patient age and gender distribution were similar between the two groups (Table I).

Model for Overall Accuracy

After testing multiple algorithms, we found a boosted model using the tree-based classifier, RandomTree, was the most successful at overall discrimination between SA and MGD. RandomTree constructs a decision tree by considering a random subset of all the attributes at each branch point. The random selection process reduces overfitting of the model to the training data, and there is less likelihood that the learning will incorporate random data “noise” instead of maintaining generalizability for prediction of the intended outcome. In addition, we used an ensemble, combining RandomTree with AdaBoost. AdaBoost is a

meta-algorithm that creates iterative models, placing more weight on previously misclassified instances to “boost” the accuracy of subsequent predictions. For this dataset, the model achieved 94% overall accuracy, and the area under the ROC curve was 0.984. Model sensitivity was 94%, specificity was 83.8%, and PPV was 94% (Table II).

Optimizing MGD Prediction

Since our goal was to develop a classifier that maximized PPV of MGD, we attempted to improve upon the RandomTree model by testing several additional classifiers in cost-sensitive schemes. With MGD being the minority class, overall classification accuracy does not necessarily translate to meaningful prediction of MGD. With respect to classifier selection in this scheme, we had the most success using a rule-based algorithm, JRip, which utilizes repeated incremental pruning. JRip first generates a set of rules consisting of different combinations each attribute’s values and then “prunes” (i.e. deletes) certain rules based on their performance to reduce error and form a final rule set. Of the various cost-sensitive weight assignments trialed, we found 3:1 to be the optimal ratio of cost assigned to misclassification of MGD (i.e. 3 times costlier to predict MGD for actual SA cases (false positives) than to predict SA for actual MGD cases.). This strategy effectively reduced the number of false positives to 0 at the expense of 444 false negatives, achieving 100% PPV for MGD (Table II).

Imaging Review

To indicate how the cost-sensitive classifier could impact clinical care for MGD patients in a system adopting our proposed framework, we reviewed all pre-operative imaging studies performed on the cohort of patients predicted to have MGD with 100% PPV. This included 34/478 patients with MGD and 0/1532 patients with SA. There were 39 total imaging tests performed for the 34 patients: 28 sestamibi scans and 11 ultrasounds. Of the sestamibi scans performed, 71% (n=20) were incorrect, and 64% (n=7) of ultrasounds were incorrect (Figure 1). All patients ultimately underwent a curative bilateral neck exploration as expected, but 58% (n = 20) of these patients had unrecognized MGD as evidenced by mistakenly positive localization of a single adenoma on imaging and need for conversion from failed minimally invasive parathyroidectomy.

Discussion

We report on the novel application of two promising ML methods that accurately distinguish MGD from SA in patients with PHPT early in their workup. The end result is a predictive algorithm that will streamline the workup and operative approach for patients with PHPT. The boosted method performed remarkably well, correctly classifying the morphologic etiology of PHPT with greater than 94% accuracy using patient characteristics and relevant biochemical data. Overall, we believe that these methods can provide utility to both referring providers and surgeons at multiple time points along the trajectory of care for patients with PHPT. By identifying patients who will need a bilateral neck exploration, there would be an opportunity to refer patients to centers that are experienced with this operative approach. Instead of attempting to treat all PHPT patients as if they had SA, our proposed clinical

pathway (Figure 2) provides a more efficient, cost-saving, individualized approach to the workup and surgical care.

As early as the time of diagnosis of PHPT, there is benefit to recognizing the possibility or suspicion of MGD. The provider could be alerted by a best-practice advisory within the electronic health record to forgo ordering a sestamibi scan, neck ultrasound, or four-dimensional CT. The patient would then be referred to a surgeon comfortable with bilateral exploration for further evaluation. The use of ultrasound by surgeons for evaluation of patients with PHPT has been shown to be beneficial, and ultrasound is also advisable to detect potential thyroid pathology(18, 27). It would be naïve to assume that an implemented and widely validated version of this model would be completely immune to false positives, especially as PPV varies with disease prevalence; thus, this surgeon performed ultrasound would also serve as a safety net to catch potentially misclassified SA cases deserving of a minimally invasive approach. In addition, the use of ultrasound would allow the surgeon to evaluate imaging findings in light of any concomitant thyroid nodular disease. The remaining patients would be able to undergo a bilateral neck exploration without delay. This represents just one possible clinical framework to integrate our ML algorithm. In fact, the inevitable variability between institutions and surgeon-specific practices opens up the possibility of different adaptations of our proposed pathway. For example, for surgeons who already received PHPT referrals without prior imaging studies ordered, the ML algorithm would aid the surgeon in the selection of imaging and operative planning.

Although it is difficult to truly predict the benefits to work flow and patient triage, there are certainly potential financial benefits at the individual level for patients and on a population level as well. For patients who have actual MGD, the sequential ordering of multiple imaging studies in the pre-operative period to search for and confirm the location of a SA can be costly; to that end, our imaging review was intended to demonstrate the benefits of streamlined care on a population level. The presence of underlying thyroid pathology (i.e. multinodular goiter) or prior neck surgery can confound the interpretation of imaging studies, and we attempted to control for this by excluding such patients. Future studies are needed to evaluate the potential for cost savings from foregone imaging and operative approach. As was the case in this cohort, the ideal real-life scenario would be prediction of a pure cohort of actual MGD with no down side to MGD prediction (i.e. no patients with SA mislabeled as MGD who consequentially would be referred for a bilateral neck exploration unnecessarily). The cost of patient time associated with having unnecessary preoperative imaging performed should also be considered. For example, 25 minutes is the average duration spent in the operating room waiting for IoPTH results.(15) At an approximate cost of fifteen dollars per minute OR time(28), this represents a potentially avoidable cost of \$375 per patient.

The literature contains numerous examples of studies that attempt to make predictions about the existence of SA vs MGD, and most predictive efforts to date have focused on the confirmation of single gland disease, suitability, and execution of a successful targeted parathyroidectomy(29). Clearly, higher Ca and PTH levels are associated with SA(30). The CaPTHUS scoring model proposed by Kebebew et al(31) was created for SA prediction and incorporated pre-operative serum calcium and PTH levels, but was contingent on the results

of imaging studies. The “Wisconsin Index” also incorporated calcium and PTH but was developed as an intraoperative tool to predict the likelihood of additional hyperfunctioning parathyroid glands(32). Finally, numerous studies have reported gland weight/size as a predictive factor with smaller glands favoring MGD (33–35). We hope to predict MGD far prior to being able to make that assessment. Unlike the results of Smith et al who failed to find a difference in SA vs MGD based on clinical factors alone, we report a highly accurate capability to distinguish MGD from SA(36, 37). In contrast to the existing literature detailed above, our study omits imaging or intra-operative data and enables accurate prediction before money and time is wasted on imaging or operative time. Future studies are needed to continue to explore the predictive possibilities of ML in primary hyperparathyroidism, including the identification and localization of ectopic glands.

The capabilities of ML are increasingly being applied to advance clinical decision making. More and more researchers and physicians are embracing these techniques to tackle challenging aspects of patient care - from recognition of breast cancer on mammography(38), to prediction of transplantation outcomes and donor selection(39), to the determination of burn healing time(40). We have previously used ML to accurately make the diagnosis of PHPT, even in mild disease(41). Our diagnostic model, in combination with the results of this study, exemplify how ML can and will likely assist in decision making along the entire continuum of surgical care – from disease recognition(41) and diagnosis, to workup, and even operative approach (Figure 2).

There are several limitations with our study. First, this was a retrospective study design and only included data from patients at a single institution. Our relatively large sample size was beneficial, however geographic variation in MGD prevalence and patient demographics cannot be accounted for. To extend upon this, our models have yet to be validated with other patient populations. This includes patients with multinodular goiter/thyroid nodules as this patient population was not accounted for in our machine learning methods. We are hopeful that the cross-validation procedures during the training process will allow generalizability of the models, but there is certainly a potential for overfitting. The lack of 4D-CT use at our institution during the study period also is a source of institutional bias when considering imaging performed. Lastly, as the end-goal of many predictive ML models such as ours is use as a clinical decision support tool, there remain barriers to implementation and adoption, which are beyond the realm of the current study. Much work lies ahead regarding implementation, usability, and perfecting human-computer interactions. To address the potential issue of non-universal electronic health records for integration, our machine learning model could be easily implemented as a web-based “calculator”, distributed as a stand-alone program, or offered as a mobile application.

In conclusion, these models demonstrate that MGD can be identified using ML methods early in the management of PHPT, providing an opportunity to reduce the costs associated with pre-operative workup and operation.

References

1. Heath H 3rd, Hodgson SF, Kennedy MA. Primary hyperparathyroidism. Incidence, morbidity, and potential economic impact in a community. *The New England journal of medicine*. 1980; 302:189–193. [PubMed: 7350459]
2. Arnold A, Staunton CE, Kim HG, Gaz RD, Kronenberg HM. Monoclonality and abnormal parathyroid hormone genes in parathyroid adenomas. *The New England journal of medicine*. 1988; 318:658–662. [PubMed: 3344017]
3. Backdahl M, Howe JR, Lairmore TC, Wells SA Jr. The molecular biology of parathyroid disease. *World journal of surgery*. 1991; 15:756–762. [PubMed: 1685045]
4. Westin G, Bjorklund P, Akerstrom G. Molecular genetics of parathyroid disease. *World journal of surgery*. 2009; 33:2224–2233. [PubMed: 19373510]
5. Vandenbulcke O, Delaere P, Vander Poorten V, Debruyne F. Incidence of multiglandular disease in sporadic primary hyperparathyroidism. *B-ent*. 2014; 10:1–6. [PubMed: 24765822]
6. Silverberg SJ, Shane E, Jacobs TP, Siris E, Bilezikian JP. A 10-Year Prospective Study of Primary Hyperparathyroidism with or without Parathyroid Surgery. *New England Journal of Medicine*. 1999; 341:1249–1255. [PubMed: 10528034]
7. Yeh MW, Ituarte PH, Zhou HC, Nishimoto S, Liu IL, et al. Incidence and prevalence of primary hyperparathyroidism in a racially mixed population. *J Clin Endocrinol Metab*. 2013; 98:1122–1129. [PubMed: 23418315]
8. Allendorf J, DiGorgi M, Spanknebel K, Inabnet W, Chabot J, et al. 1112 consecutive bilateral neck explorations for primary hyperparathyroidism. *World J Surg*. 2007; 31:2075–2080. [PubMed: 17768656]
9. Abdulla AG, Ituarte PH, Harari A, Wu JX, Yeh MW. Trends in the frequency and quality of parathyroid surgery: analysis of 17,082 cases over 10 years. *Annals of surgery*. 2015; 261:746–750. [PubMed: 24950283]
10. Chen H, Wang TS, Yen TW, Doffek K, Krzywda E, et al. Operative failures after parathyroidectomy for hyperparathyroidism: the influence of surgical volume. *Annals of surgery*. 2010; 252:691–695. [PubMed: 20881776]
11. Yip L, Ogilvie JB, Challinor SM, Salata RA, Thull DL, et al. Identification of multiple endocrine neoplasia type 1 in patients with apparent sporadic primary hyperparathyroidism. *Surgery*. 2008; 144:1002–1006. discussion 1006–1007. [PubMed: 19041010]
12. Awad SS, Miskulin J, Thompson N. Parathyroid adenomas versus four-gland hyperplasia as the cause of primary hyperparathyroidism in patients with prolonged lithium therapy. *World journal of surgery*. 2003; 27:486–488. [PubMed: 12658498]
13. Bergenfelz A, Lindblom P, Tibblin S, Westerdaal J. Unilateral versus bilateral neck exploration for primary hyperparathyroidism: a prospective randomized controlled trial. *Annals of surgery*. 2002; 236:543–551. [PubMed: 12409657]
14. Solorzano CC, Carneiro-Pla D. Minimizing cost and maximizing success in the preoperative localization strategy for primary hyperparathyroidism. *The Surgical clinics of North America*. 2014; 94:587–605. [PubMed: 24857578]
15. Morris LF, Zanolco K, Ituarte PH, Ro K, Duh QY, et al. The value of intraoperative parathyroid hormone monitoring in localized primary hyperparathyroidism: a cost analysis. *Ann Surg Oncol*. 2010; 17:679–685. [PubMed: 19885701]
16. Siperstein A, Berber E, Barbosa GF, Tsinberg M, Greene AB, et al. Predicting the success of limited exploration for primary hyperparathyroidism using ultrasound, sestamibi, and intraoperative parathyroid hormone: analysis of 1158 cases. *Annals of surgery*. 2008; 248:420–428. [PubMed: 18791362]
17. Press DM, Siperstein AE, Berber E, Shin JJ, Metzger R, et al. The prevalence of undiagnosed and unrecognized primary hyperparathyroidism: a population-based analysis from the electronic medical record. *Surgery*. 2013; 154:1232–1237. discussion 1237–1238. [PubMed: 24383100]
18. Wilhelm SM, Wang TS, Ruan DT, Lee JA, Asa SL, et al. The American Association of Endocrine Surgeons (AAES) Guidelines for Definitive Management of Primary Hyperparathyroidism. *JAMA surgery*. 2016 in press.

19. Hendee WR, Becker GJ, Borgstede JP, Bosma J, Casarella WJ, et al. Addressing overutilization in medical imaging. *Radiology*. 2010; 257:240–245. [PubMed: 20736333]
20. Siström CL. The appropriateness of imaging: a comprehensive conceptual framework. *Radiology*. 2009; 251:637–649. [PubMed: 19474372]
21. Liu, YI., Kamaya, A., Desser, TS., Rubin, DL. A Bayesian classifier for differentiating benign versus malignant thyroid nodules using sonographic features. *AMIA ... Annual Symposium proceedings / AMIA Symposium*. AMIA Symposium; 2008; p. 419-423.
22. Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, et al. The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine*. 2009; 46:5–17. [PubMed: 18790621]
23. Cook C. Is Clinical Gestalt Good Enough? *The Journal of Manual & Manipulative Therapy*. 2009; 17:6–7. [PubMed: 20046560]
24. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *The New England journal of medicine*. 2016; 375:1216–1219. [PubMed: 27682033]
25. Thornton, C., Hutter, F., Hoos, HH., Leyton-Brown, K. Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*; Chicago, Illinois, USA: ACM; 2013. p. 847-855.
26. Weng, CG., Poon, J. A new evaluation measure for imbalanced datasets. *Proceedings of the 7th Australasian Data Mining Conference - Volume 87*; Glenelg, Australia: Australian Computer Society, Inc; 2008. p. 27-32.
27. Arora S, Balash PR, Yoo J, Smith GS, Prinz RA. Benefits of surgeon-performed ultrasound for primary hyperparathyroidism. *Langenbecks Arch Surg*. 2009; 394:861–867. [PubMed: 19547997]
28. Stahl JE, Sandberg WS, Daily B, Wiklund R, Egan MT, et al. Reorganizing patient care and workflow in the operating room: a cost-effectiveness study. *Surgery*. 2006; 139:717–728. [PubMed: 16782425]
29. Kavanagh DO, Fitzpatrick P, Myers E, Kennelly R, Skehan SJ, et al. A predictive model of suitability for minimally invasive parathyroid surgery in the treatment of primary hyperparathyroidism [corrected]. *World J Surg*. 2012; 36:1175–1181. [PubMed: 22170475]
30. Stawicki SP, El Chaar M, Baillie DR, Jaik NP, Estrada FP. Correlations between biochemical testing, pathology findings and preoperative sestamibi scans: a retrospective study of the minimally invasive radioguided parathyroidectomy (MIRP) approach. *Nuclear medicine review. Central & Eastern Europe*. 2007; 10:82–86. [PubMed: 18228211]
31. Kebebew E, Hwang J, Reiff E, Duh QY, Clark OH. Predictors of single-gland vs multigland parathyroid disease in primary hyperparathyroidism: a simple and accurate scoring model. *Archives of surgery (Chicago, Ill : 1960)*. 2006; 2006:777–782.
32. Mazeh H, Chen H, Levenson G, Sippel RS. Creation of a “Wisconsin Index” Nomogram to Predict the Likelihood of Additional Hyperfunctioning Parathyroid Glands During Parathyroidectomy. *Annals of surgery*. 2012
33. Norlén O, Glover A, Zaidi N, Aniss A, Sywak M, et al. The Weight of the Resected Gland Predicts Rate of Success After Image-Guided Focused Parathyroidectomy. *World Journal of Surgery*. 2015; 39:1922–1927. [PubMed: 25677011]
34. McCoy KL, Chen NH, Armstrong MJ, Howell GM, Stang MT, et al. The small abnormal parathyroid gland is increasingly common and heralds operative complexity. *World journal of surgery*. 2014; 38:1274–1281. [PubMed: 24510243]
35. Sepahdari AR, Bahl M, Harari A, Kim HJ, Yeh MW, et al. Predictors of Multigland Disease in Primary Hyperparathyroidism: A Scoring System with 4D-CT Imaging and Biochemical Markers. *AJNR Am J Neuroradiol*. 2015; 36:987–992. [PubMed: 25556203]
36. Moalem J, Guerrero M, Kebebew E. Bilateral neck exploration in primary hyperparathyroidism--when is it selected and how is it performed? *World J Surg*. 2009; 33:2282–2291. [PubMed: 19234738]
37. Mozes G, Curlee KJ, Rowland CM, van Heerden JA, Thompson GB, et al. The predictive value of laboratory findings in patients with primary hyperparathyroidism. *Journal of the American College of Surgeons*. 2002; 194:126–130. [PubMed: 11848628]

38. Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney GA, et al. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology*. 2006; 240:666–673. [PubMed: 16926323]
39. Shouval R, Bondi O, Mishan H, Shimoni A, Unger R, et al. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. *Bone Marrow Transplant*. 2014; 49:332–337. [PubMed: 24096823]
40. Yeong EK, Hsiao TC, Chiang HK, Lin CW. Prediction of burn healing time using artificial neural networks and reflectance spectrometer. *Burns*. 2005; 31:415–420. [PubMed: 15896502]
41. Somnay YR, Craven M, McCoy KL, Carty SE, Wang TS, et al. Improving diagnostic recognition of primary hyperparathyroidism with machine learning. *Surgery*. 2016

Figure 1. Flow chart of imaging studies performed on the cohort of patients predicted to have MGD by the cost-sensitive model. Shaded cells represent misleading imaging studies (i.e. incorrect for MGD)

Figure 2.

Conceptual framework outlining the utility of a clinical decision support (CDS) tool optimized to predict MGD. Shaded cells indicate efficient care for patients with actual MGD. Patients predicted to have MGD could avoid imaging studies and be referred to a surgeon early after diagnosis. Ultrasound is proposed as a means to detect thyroid pathology and detect misclassified patients.

Table I

Characteristics of Patients with Primary Hyperparathyroidism by Etiology

	Single Adenoma (n=1532)	Multigland Disease (n=478)	P value
Age	61 ± 13.8	61 ± 12.8	0.82
Body Mass Index	30 ± 7.69	29 ± 8.2	0.17
Calcium (mg/dL)	10.9 ± 0.83	10.4 ± 0.86	<.001
Phosphate	2.8 ± 3.7	3.1 ± 3.1	0.59
PTH (pg/mL)	105 ± 103	89 ± 78	<.001
Urine Calcium (mg/24 hr)	300 ± 209	243 ± 157	<.001
ALP (U/L)	88 ± 35.6	86 ± 41.5	0.56
Vitamin D (ng/mL)	30 ± 12.9	33 ± 14.7	<.001
Creatinine (mg/dL)	0.9 ± 0.48	0.9 ± 0.52	0.75
Female	1180 (77)	383 (80)	0.15
Bone Disease	616 (40.2)	223 (46.6)	0.01
Prior Fracture	101 (6.6)	48 (10.1)	0.01
GERD	290 (18.9)	116 (24.3)	0.01
Nephrolithiasis	253 (16.5)	102 (21)	0.016

Continuous variables represented by their median ± standard deviation; Categorical variables represented by their number (%). PTH = parathyroid hormone; ALP = alkaline phosphatase; GERD = gastroesophageal reflux

TABLE II

Classification Results using WEKA

Classifier	Accuracy	Recall (sensitivity)	1 – FP Rate (specificity)	Precision (PPV)	ROC Area
I. Maximizing overall accuracy for prediction of single adenoma vs. multigland disease					
<i>RandomTree + AdaBoost</i> ^a	94.1%	94.1%	83.8%	94.1%	0.984
II. Maximizing PPV for detection of multigland disease					
<i>JRip + Cost-Sensitive Learning</i> ^b	77.9%	7.1%	100%	100%	0.536

^a*RandomTree* is a tree classifier; with boosting, this model achieved the best overall accuracy

^b*JRip* classifier is a rule-based classifier; cost-sensitive learning weights of 3:1 were assigned to reduce MGD prediction false positives and maximize PPV

FP rate = false positive rate; PPV = positive predictive value; ROC = receiver operating characteristic curve