DiagTree: Diagnostic Tree for Differential Diagnosis

Yejin Kim
POSTECH
Pohang, Korea
yejin89@postech.ac.kr

Jingyun Choi
POSTECH
Pohang, Korea
jingyunc@postech.ac.kr

Yosep Chong
St. Mary's Hospital
Seoul, Korea
ychong@catholic.ac.kr

Xiaoqian Jiang*
UC San Diego
La Jolla, California, USA
x1jiang@ucsd.edu

Hwanjo Yu*
POSTECH
Pohang, Korea
hwanjoyu@postech.ac.kr

ABSTRACT

Differential diagnosis is detection of one disease among similar diseases using evidence such as pathologic tests. A Partially Observed Markov Decision Process (POMDP) formulates the complex differential diagnosis process into a probabilistic decision-making model. However, differential diagnosis is not often fully formulated as POMDP because model construction does not consider the cost (or time) to finish the diagnosis process, or the practical convention on clinical tests. We propose a DIAGNOSTIC TREE (DIAGTREE), a new framework for diagnosing diseases, which combines several tests to reduce the diagnosis time and to incorporate real-world constraints into discrete optimization. DIAGTREE consists of multiple tests in internal nodes and posterior probabilities ("confidences") that the patient suffers the disease listed at each leaf node. The confidences are computed after a series of test results is applied in internal nodes. DIAGTREE is built to maximize the confidences at leaf nodes and to minimize the decision process time. We formulate this problem as integer programming and solve it by the Branch-and-Bound method and a greedy approach. We apply DIAGTREE to immunohistochemistry profiles to detect lymphoid neoplasms. We evaluate the accuracy and cost of the diagnosis rules from DIAGTREE compared to those obtained using rules that clinicians derived from their experience. DIAGTREE detected diseases with high accuracy and also reduced the diagnosis cost (or time) compared to the existing rules of clinicians. DIAGTREE can support clinicians by suggesting a simple diagnosis process with high accuracy and low cost among test candidates.

CCS CONCEPTS

•Applied computing →Health care information systems; Health informatics; •Mathematics of computing →Maximum likelihood estimation;

KEYWORDS

Decision process, Discrete optimization, Early diagnosis

ACM Reference format:
Yejin Kim, Jingyun Choi, Yosep Chong, Xiaoqian Jiang*, and Hwanjo Yu. 2017. DiagTree: Diagnostic Tree for Differential Diagnosis. In *Proceedings of CIKM'17, November 6–10, 2017, Singapore., ,* 10 pages.
DOI: https://doi.org/10.1145/3132847.3132924

1 INTRODUCTION

Differential diagnosis distinguishes a particular target disease from other candidates that show similar signs and symptoms. Clinicians use evidence such as pathologic tests and medical knowledge to establish that some candidate diagnoses are unlikely to be correct, and to increase the probability of correctly diagnosing the disease, but this probability inference becomes complex with increase in the numbers of associated tests and candidate diagnoses. Clinicians consider a patient's medical status, and would like to select the most effective and least expensive combination of tests to diagnose the disease. Most clinicians use *ad hoc* or heuristic strategies to select the combination of tests, but this selection process can be too complicated for intuition alone to be a sufficient guide [19, 25] because the results of individual tests are usually uncertain rather than deterministic, and because the cost of the tests may vary depending on the order in which they are done.

Partially Observable Markov Decision Process (POMDP) has shown great potential to provide computerized assistance to clinicians by successfully formulating the diagnosis process as a probabilistic decision-making model. POMDP consists of i) a set of hidden states, in which underlying diseases are observed as probabilities that diagnosis of a specific diseases is correct, ii) a set of actions, i.e., tests that clinicians can choose, and iii) a set of transition probabilities of the hidden states conditioned on hidden states and actions. POMDP finds a policy (i.e., tests that the clinician should perform at each stage). The policy provides an optimal sequence of actions with its consequences by maximizing a criterion called a *reward*. Several studies have used POMDP to support practitioners by suggesting computerized choices of best diagnosis [3, 10, 11].

However, POMDP has limited application to the diagnosis process, because the clinicians' reasoning process is sometimes different with POMDP in that it does not consider the following real-world concerns:

i) Time to finish the decision process: In practice, a series of tests must be completed within a certain time, because delayed

*Corresponding author

CIKM'17, November 6–10, 2017, Singapore.
 ISBN 978-1-4503-4918-5/17/11…$15.00
DOI: https://doi.org/10.1145/3132847.3132924

diagnosis could reduce the quality of treatment. Many emergency-department physicians believe that laboratory test time causes delayed emergency treatment more than 50% of the time [23]. For example, if a test requires one day to complete, then a one-by-one testing strategy might be impractical due to the time constraint, so to complete the diagnosis within a given time, several tests must be performed concurrently.

ii) Limited number of tests and no instant reward: Clinicians aim to maximize the confidence of a target disease after trying the certain number of tests, rather than to repeat the tests indefinitely as earnning an instant reward at every testing stage. The total number of tests that can be used is usually limited, because, in practice, the cost of tests is supported by insurance up to certain level (eg. 6). The test cost becomes abruptly expensive after that level. For example, Health Insurance Review & Assessment Service from Korea Ministry of Health and Welfare imposes an additional fee of about 1.8 times on seventh cell immunity test [21]. Accordingly, POMDP is not appropriate for differential diagnosis, because finding the optimal policy of POMDP assumes that the tests can be repeated indefinitely and requires an instant reward that depends on every state and action.

iii) Reused tests and not-covered diseases: The policy should avoid reusing tests, because such a medical test sometimes accompanies an irreversible chemical reaction that cannot be repeated on the same tissue again (eg. antibodies test to detect lymphoid neoplasms); thus POMDP's strategy that repeats the same test to reduce false positive is not applicable. The policy should also agree with a guideline that requires at least one test pertain directly to each disease [5, 24]. If a disease is not tested for, results can be ambiguous or erroneous. For example, among three candidate diseases, A, B, and C, we should not conclude that C is the target disease just because probabilities that A and B are the target disease are lower than the probability of C after only testing for A and B.

To the best of our knowledge, no previous study has satisfied those requirements. Decentralized POMDP (Dec-POMDP) [4] constructs a POMDP with multiple agents who independently take an action at each stage, and derives a joint policy that the agents perform together. A main assumption is that, while the Dec-POMDP is executing, each agent only receives its individual part of the joint policy and does not communicate with others to share the results. This method might help to satisfy the time constraint by distributing the actions among multiple clinicians, but use of multiple clinicians in the same diagnosis task is not desirable just for one patient. In addition, incorporating the real-world constraints above into Dec-POMDP is a challenging task. The state-of-the-art algorithm for optimal policy of Dec-POMDP is based on a joint equilibrium [16], which maximizes one agent's reward while fixing others' policies iteratively until the policies converge. This method requires an instant reward and does not ensure that all diseases will be tested at least once. Meanwhile, ρPOMDP devises the reward as a function of belief in the hidden state (i.e., entropy) to reduce uncertainty [1]. ρPOMDP expresses the reward as a piecewise linear and convex approximation function of entropy to seamlessly adapt normal POMDP frameworks. This method, however, cannot model our problem with the limited time and number of tests.

Table 1: Example of probabilities P_{nm} that diagnosis test t_m is positive given disease y_n

	Test t_m				
Disease y_n	t_1	t_2	...	t_{M-1}	t_M
y_1	0	0.3		0.9	0.1
y_2	0.8	0.4		0.8	NaN
...			...		
y_{N-1}	0.5	NaN		0.1	0.5
y_N	0.2	0.3		0.2	0.2

Therefore, we propose Diagnostic Tree (DiagTree), a new framework based on POMDP. DiagTree is a tree-like structure that represents a policy and its results to address the above requirements. DiagTree minimizes the diagnosis time during model construction by allowing multiple tests to be performed concurrently. DiagTree consists of multiple tests in internal nodes and posterior probabilities (or confidence) of diseases at leaf nodes. The confidences of diseases are computed after a series of test results is applied in internal nodes. DiagTree is built to maximize the cofidence (at leaf nodes) and minimize the decision process time. We formulate this problem as integer programming that finds optimal binary decision variables while satisfying the real-world constraints (i.e., no reuse; full coverage), and solve it using the Branch- and-Bound method and a greedy approach.

We apply DiagTree to an immunohistochemistry (IHC) profile, which contains probabilities that antibody tests will detect lymphoid neoplasms. We evaluate DiagTree by measuring i) diagnosis accuracy and ii) time and the number of tests mandated by DiagTree and by manually-produced clinicians' rules. We use data from patients who have actually undergone the tests at UC San Diego Medical Center or Yeouido St. Mary hospital. DiagTree automatically produces equivalent or significantly better test-and-diagnosis procedures (with high confidence) than existing test-and-diagnosis procedures manually designed by clinicians. As a result, DiagTree can help clinicians by narrowing a complicated set of test candidates to a simple diagnosis process that has high likelihood of being correct, and thereby enabling early diagnosis.

2 PROBLEM FORMULATION

Clinicians diagnose a (target) disease among several possible (candidate) diseases using a set of diagnosis tests that react to the candidate diseases with certain probabilities. In practice, test results are uncertain, i.e., the probability p that a test reacts to a disease is < 1. For example, a test may react to one disease 7 out of 10 times (i.e., $p = 0.7$) and to another disease 2 out of 10 times (i.e., $p = 0.2$). Test results are useful to guide selection of the next test to reduce the set of candidate diseases and finally to diagnose the disease with high probability (or confidence); thus clinicians must identify which tests should be used and in which order. *An effective test procedure must both diagnose the target disease correctly with high probability and minimize the number of tests.* Sometimes multiple tests should be applied at the same time if the series of "one-by-one" tests takes too long to make the final diagnosis; i.e., the decision process consists of several stages that can involve multiple tests. Tests at the next stage are determined by the test results of the

current stage. Accordingly, diagnosis entails two types of cost: (1) the number of tests and (2) the number of stages (related to the time to make the final diagnosis).

By building the DIAGNOSTIC TREE, we aim to develop a decision process *that maximizes the probability (or confidence)* of identifying the disease while *minimizing the number of tests applied and the number of stages*. To solve the problem, we use the following information.

- Set of **diseases**, $y \in \{y_n\}_{n=1}^{N}$,
- Set of **tests**, $t \in \{t_m\}_{m=1}^{M}$: The test result t_m is either $t_m = 1$ (positive) or $t_m = 0$ (negative).
- Set of **probabilities** P_{nm} that test t_m is positive given disease y_n: $P_{nm} = p(t_m = 1|y_n)\ \forall n, m$.
- **Prior probability** $\pi(y)$ that a candidate disease is the target disease.

Based on the diagnosis process in practice, we have three restrictions in building the diagnostic tree as motivated in Section 1.

i) No reuse: A test is only applied once (due to irresversible chemical reaction).

ii) Full coverage: For each candidate disease, at least one test that can detect it must be applied.

iii) Maximum number of tests: The total number of tests that can be used is up to Q due to insurance policy. Accordingly, the maximum number q of tests applied at the same stage is also set to Q ($q = 1, \ldots, Q$).

3 DIAGNOSTIC TREE

We introduce DIAGTREE, a novel framework that propagates disease probabilities over a series of test results in a tree structure.

3.1 Structure

DIAGTREE is a tree-like structure that detects the target disease by iteratively applying tests and reducing the ambiguity of disease diagnoses. This tree consists of multiple tests in each internal node and a posterior probability vector $p(\mathbf{y})$ that candidate diseases at each leaf node are the one that the patient suffers (Fig. 1, 2, 3). Each internal node has $1 \le q \le Q$ tests; each is assigned to the jth position ($j = 1, \ldots, q$) of the node. The internal node is split in 2^q ways (positive or negative for each test) on the q tests. The 2^q-way child nodes are represented as the test results; i.e., a q-digit binary number. Each leaf node has a vector of posterior probabilities that the candidate diseases is the target disease conditioned on the set of tests results.

Each node in DIAGTREE has an integer address u (= 0, 1, ...). The address is assigned from parent to child node and from leftmost to rightmost child node. The root node is $u = 0$, and the 2^q child nodes of u have addresses of $2^q \cdot u + r, r = 1, \ldots, 2^q$, respectively. The node address u is self-explanatory because one can infer the parent node of u (i.e., $parent(u)$) and the parent node's test results (i.e., $\mathcal{R}(u)$) from which the child node u is generated. Because $\mathcal{R}(u)$ is the series of q results from each of the q tests, the jth value $\mathcal{R}(u)_j$ of $\mathcal{R}(u)$ is the test result of the jth test.

The stage of u, i.e., $stage(u)$, is the number of nodes from root to u. If u is the root, then $stage(u) = 0$. Depth K is the maximum stage in the tree plus one, i.e., the total number of stages (Fig. 1). The depth means the maximum number of stages needed to finish the diagnosis procedure. The depth matters because it measures the cost (i.e., maximum waiting time) to get the final diagnosis. When $q = 1$, each internal node has only one test, so the depth becomes $Q + 1$ (Fig. 1). When $1 < q < Q$, each internal node includes q tests, so the depth becomes $\left\lceil \frac{Q}{q} \right\rceil + 1$ (Fig. 2). When $q = Q$, DIAGTREE contains only root and leaf nodes (Fig. 3). The path of u, denoted as $path(u, \cdot)$, is an ordered set of nodes from root to u. $path(u, k)$ is the kth node of set $path(u, \cdot)$, i.e., the node at stage k on the path from root to u. $path(u, 0)$ represents the root node, and $path(u, stage(u))$ represents node u itself.

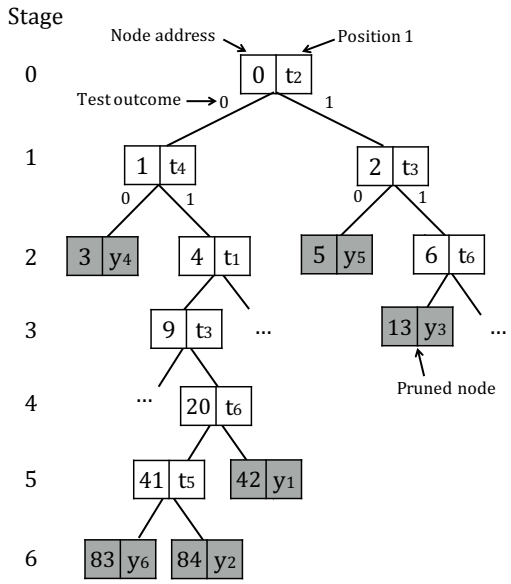


Figure 1: Example of DIAGTREE with (Q, q)=(6, 1). Each internal node is split in 2^1 ways where the result of q tests are 0 or 1. The depth $K = 7$. Leaf nodes with the target disease are marked as grey.

3.2 Probability inference in internal nodes

The probability that a candidate disease is the target disease is updated while going through tests in internal nodes. We infer the probability after tests as follows. In each internal node, prior probability from the parent node is used to compute the posterior probability that is conditioned on the test at the internal node. At first, the prior probability of disease $\pi(\mathbf{y})$ that has the maximum of ambiguity begins at the root node, and the probability is propagated along with the set of tests. When test $t_1, \ldots, t_q$ is applied, the posterior probability that is conditioned on the test results is $p(\mathbf{y}|t_1, \ldots, t_q)$, which is derived by Bayes theorem:

$$p(\mathbf{y}|t_1, \ldots, t_q) = \frac{\pi(\mathbf{y})p(t_1, \ldots, t_q|\mathbf{y})}{p(t_1, \ldots, t_q)} = \frac{\pi(\mathbf{y}) \prod_{m=1}^{q} p(t_m|\mathbf{y})}{\prod_{m=1}^{q} p(t_m)}$$

using the memoryless strategy in Markov decision process (i.e., actions are made solely based on current states regardless of past actions) [14]. As tests are applied repeatedly at the child nodes, posterior probabilities for some diseases become close to 1, and

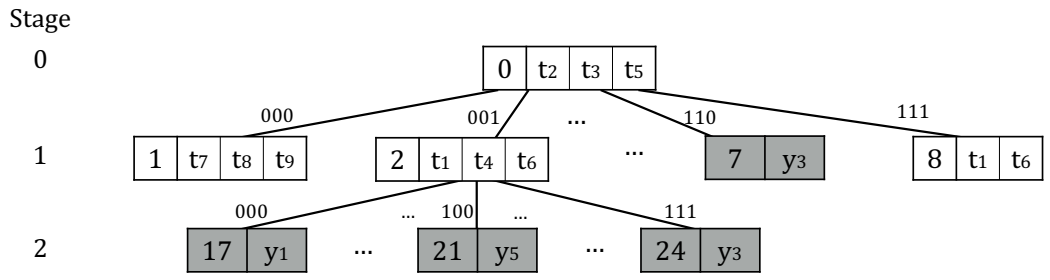


Figure 2: Example of DiagTree with (Q, q)=(6, 3). Each internal node is split in 2^3-way where the result of q tests are 000, 001, 010, 011, 100, 101, 110, and 111. The depth $K = 3$.

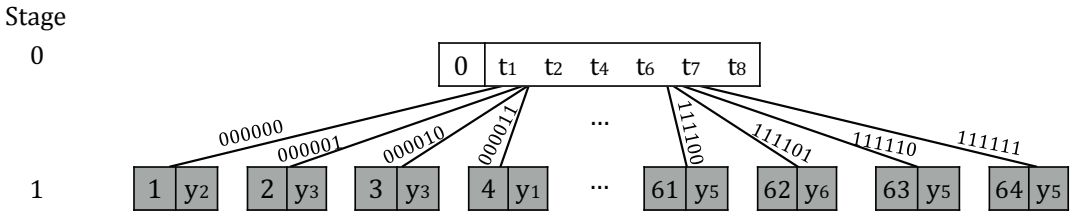


Figure 3: Example of DiagTree with (Q, q)=(6, 6). Each internal node is split in 2^6-way where the result of q tests are 000000, 000001, . . ., 111110, 111111. The depth $K = 2$.

others become close to 0; consequently, the ambiguity of candidate disease decreases.

3.3 Disease detection in leaf nodes

Among diseases at a leaf node u_l, one disease is selected as the target disease. For each leaf node u_l, a disease y^* for which the posterior probability of diagnosis is the largest, i.e.,

$$y^* = \text{argmax}_n \, p(y_n|u_l), \tag{1}$$

is regarded as the target disease, because clinicians select the disease for which the posterior probability is the largest among the candidate diseases. Thus, to maximize the confidence of diagnosis, we must build a DiagTree that maximizes the posterior probability of the target disease and minimizes the posterior probability of the other candidate diseases at each leaf node. Accordingly, the likelihood function to derive the optimal DiagTree is the ratio between the probabilities that the y^* is the target disease (correct diagnosis), and the geometric mean of the posterior probabilities that diagnoses of other candidates are the target disease:

$$l(u_l) = \frac{p(y^*|u_l)}{\sqrt[N-1]{\prod_{n, y_n \neq y^*} p(y_n|u_l)}}. \tag{2}$$

This likelihood is devised as the ratio form with geometric mean to make the objective function linear after logarithm in Section 5.1. When $p(y_n|u_l) \leq \epsilon$, where ϵ is an arbitrary small value, we set $p(y_n|u_l) = \epsilon$ so that the denominator cannot be less than ϵ. Note that minimizing probability that y^* is not the target disease (i.e., false negative) can be an alternative approach. For example, in Fig. 4, prior probabilites $\pi(\mathbf{y})$ is changed into posterior probability $p(\mathbf{y}|t_2, t_3, t_5)$ after test t_2, t_3, and t_5 are applied. In specific, $p(y_1|t_2 = 0, t_3 = 0, t_5 = 1)$ is computed as $\{\pi(y_1)(1 - p(t_2 = 1|y_1))(1 - p(t_3 = 1|y_1))p(t_5 = 1|y_1)\}/\{p(t_2 = 0)p(t_3 = 0)p(t_5 = 1)\}$. Here, $p(t = 0)$ is computed as $\sum_{n=1}^{5} \pi(y_n)p(t = 0|y_n)$. Thus, $p(y_1|t_2 = 0, t_3 = 0, t_5 = 1) = (0.13 * 0.9 * 1 * 0.72)/(0.79 * 0.96 * 0.85) \approx 0.13$. The likelihood at leaf node $u_l = 2$ is $l(u_l = 2) = \frac{0.97}{\sqrt[4]{0.13 * \epsilon * \epsilon * \epsilon}} = 1615$, and the likelihood at leaf node $u_l = 7$ is $l(u_l = 7) = \frac{1}{\sqrt[4]{\epsilon * \epsilon * \epsilon * \epsilon}} = 10000$, where $\epsilon = 10^{-4}$.

The choice of tests used in each internal node affects the likelihood of leaf nodes. Thus, we need a good strategy to appropriately assign tests to internal nodes to maximize the likelihood of leaf nodes.

3.4 Test assignments

To assign tests to nodes and formulate the structure, we introduce a set of binary indicator variables $X = \{x_{m,u,j}\}, \forall m = 1, \ldots, M$ (all tests), $\forall u = 0, \ldots, (2^q)^K - 2$ (all internal nodes), $\forall j = 1, \ldots, q$ to represent which test is assigned to each position of each node where

$$x_{m,u,j} = \begin{cases} 1 & \text{if test } t_m \text{ is assigned to position j on node u,} \\ 0 & \text{otherwise.} \end{cases}$$

Some positions of internal nodes can be pruned when no test decreases their ambiguity in posterior probabilities. In this case, for simplicity in objective function (Section 5) a dummy test ϕ is assigned to the position to indirectly represent pruning on DiagTree that has always full tree structure (Fig. 5). If a position j at node u is pruned, we denote this assignment as $x_{\phi,u,j} = 1$. For example, no more tests are needed at $u = 7$ in Fig. 2 when the probability that y_3

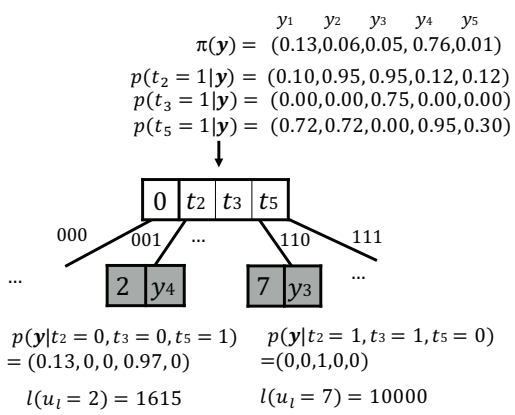


Figure 4: Example of probability inference and likelihood when (Q, q)=(6, 3)

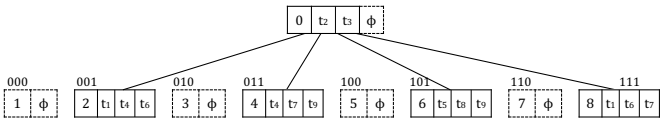


Figure 5: Example of pruning when (Q, q)=(6, 3). Pruned position is marked as ϕ in root node. Child nodes whose $\mathcal{R}(u)$ contains negative (0) result are marked as dotted line.

is the correct diagnosis is high enough compared to the probabilities of other diagnoses, and applying more tests just reduces the certainty that y_3 is correct. In this case, dummy tests are assigned at all positions 1, 2, and 3 (i.e., $x_{\phi,7,1} = 1, x_{\phi,7,2} = 1, x_{\phi,7,3} = 1$). Two tests are also enough at $u = 8$ in Fig. 2 when extra tests increase the ambiguity, so dummy test is assigned to one of positions (say position 3) $x_{\phi,8,3} = 1$. Accordingly, the number of tests executed concurrently can be less than q when dummy tests are used. To seamlessly incorporate dummy tests in the probability inference process, we assume $p(\phi = 1|\mathbf{y}) = (1, \ldots, 1)$ and let the child node with positive result inherit the posterior probability of its parent node as a prior probability. For example, in Fig. 5, node $u = 6$ inherits posterior probability of $p(\mathbf{y}|t_2 = 1, t_3 = 0, \phi = 1) = p(\mathbf{y}|t_2 = 1, t_3 = 0)$ as prior probability because $p(\phi = 1|\mathbf{y}) = (1, \ldots, 1)$. Consequently $u = 6$ is a child node of $u = 0$ with $t_2 = 1$ and $t_3 = 0$ regardless of the dummy test.

Child nodes from negative results of the dummy test always have dummy tests to indirectly express pruned nodes. That is, if $x_{\phi,u,j} = 1$, the dummy test's child nodes $2^q \cdot u + r$, $r \in \{r|\mathcal{R}(2^q \cdot u + r)_j = 0\}$ for all j are always assigned to dummy tests; this means that node u and negative child nodes $2^q \cdot u + r$ have inequality:

$$\begin{aligned} x_{\phi,2^q \cdot u+r,1} &\geq x_{\phi,u,j} \\ &\ldots \\ x_{\phi,2^q \cdot u+r,q} &\geq x_{\phi,u,j} \forall j,\ u \end{aligned} \tag{3}$$

where $r \in R = \{r|\mathcal{R}(2^q \cdot u + r)_j = 0\}$ for all j. Especially, when dummy tests are assigned at all positions within node u, all the child nodes follow this inequality (i.e., $r \in \{1, \ldots, 2^q\}$).

4 HEURISTIC APPROACH

An intuitive approach to find DiagTree is a greedy heuristic that selects the most "promising" tests greedily at each stage without considering the previous or next stages. Let us denote the entropy of internal node u without selecting tests yet as

$$H(u) = -\sum_n p(y_n|u) \log_2 p(y_n|u) \tag{4}$$

where $p(y|u)$ is the posterior probability of y that is conditioned on previous tests at parent nodes. The entropy after splitting the node u with test t_m becomes

$$\begin{aligned} &H(u|t_m) = \\ &p(t_m = 1)\left\{-\sum_{n=1}^{N} p(y_n|u, t_m = 1) \log_2 p(y_n|u, t_m = 1)\right\} \\ &+p(t_m = 0)\left\{-\sum_{n=1}^{N} p(y_n|u, t_m = 0) \log_2 p(y_n|u, t_m = 0)\right\}. \end{aligned}$$

The information gain at node u with t_m is given by

$$IG(u, t_m) = H(u) - H(u|t_m). \tag{5}$$

Then, the greedy heuristic selects top q tests t_m ($m = 1, \ldots, q$) that are not yet used and maximize information gain $IG(u, t_m)$. This selecting process stops when Q tests are used in total. As a variation we can select the next q tests that maximize likelihood gain: $LG(u, t_m) = -\log(l(u)) + \log(l(u|t_m))$ using $-\log(l(u))$ instead of $H(u)$. These intuitive methods, information gain (IG) or likelihood gain (LG) are simple and straightforward, but not fully optimized. Thus, we propose an integer programming approach to build DiagTree.

5 BUILDING DIAGNOSTIC TREE

We formulate the problem of building DiagTree (or assigning tests and pruning on positions) as integer programming. We first formulate the objective function as a linear function, derive constraints, and use the Branch-and-Bound algorithm to find the solution [8].

5.1 Objective function

The optimal diagnosis approach is to find the tree X that maximizes the posterior probability ratio in Eq.(2) at each leaf node, and that minimizes the number of tests used. For simplicity, we take the logarithm of $l(u_{li})$, which is the posterior probability ratio at ith leaf node u_{li}. The leaf nodes u_{li} are all at the deepest stage, (i.e., $stage(u_{li}) = K - 1$) of full tree containing pruned nodes. Thus, the objective function is

$$\text{argmax}_X \sum_i \log(l(u_{li})) - \mu \sum_{m, m \neq \phi} \sum_u \sum_j x_{m,u,j}, \tag{6}$$

where μ is a weighting constant for the number of positions used (i.e., not pruned). We do not impose weight $p(u_{li})$ on the $\log(l(u_{li}))$ to make each leaf node contribute to the objective function equally.

The $\log(l(u_{li}))$ is expressed as

$$\begin{aligned}\log(l(u_{li})) &= \log p(y^*|u_{li}) - \frac{1}{N-1}\left(\sum_{n, y_n \neq y^*} \log p(y_n|u_{li})\right) \\ &\propto N \log p(y^*|u_{li}) - \left(\sum_n \log p(y_n|u_{li})\right).\end{aligned} \tag{7}$$

Using Bayes theorem, we can transform $\log p(y_n|u_{li})$ of the second term to $\log p(y_n|u_{li}) = \log p(u_{li}|y_n) + \log \pi(y_n) - \log p(u_{li})$. After removing constant terms $\log \pi(y_n)$ and $\log p(u_{li})$, we derive the first term $\log p(u_{li}|y_n)$ explicitly. $p(u_{li}|y_n)$ (i.e., probability to reach u_{li} given tests) is a product of $p(t_m|y_n)$ (i.e., the probabilities of the test results) in nodes of $path(u_{li}, \cdot)$. That is, for y_n, $p(t_m|y_n)$ is given as

$$p(t_m|y_n) = \begin{cases} P_{nm} & \text{if } t_m = 1 \text{ (positive)} \\ 1 - P_{nm} & \text{if } t_m = 0 \text{ (negative).} \end{cases} \tag{8}$$

Then $p(u_{li}|y_n)$ is

$$p(u_{li}|y_n) = \prod_{k=0}^{K-2} \prod_{j=1}^{q} \prod_{m=1}^{M} p(t_m|y_n)^{x_{m,path(u_{li},k),j}} \tag{9}$$

where $path(u_{li}, k) = parent(path(u_{li}, k+1))$. Thus, the objective function is derived using X,

$$\begin{aligned}&\text{argmax}_X \sum_i \Bigg[N \log p(y^*|u_{li}) \\ &- \left(\sum_{n=1}^{N} \sum_{k=0}^{K-2} \sum_{j=1}^{q} \sum_{m=1}^{M} \log p(t_m|y_n) x_{m,path(u_{li},k),j}\right)\Bigg] \\ &- \mu \sum_{m, m\neq\phi} \sum_u \sum_j x_{m,u,j}.\end{aligned} \tag{10}$$

5.2 Constraints

We have constraints on target disease posterior probability, structure, no-reuse, and full coverage as follows.

5.2.1 Max target disease posterior probability. To make the objective function linear, we transform the max function on target disease (i.e., $p(y^*|u_{li})$) to a linear function by deriving new continuous variables and indicator variables. To represent the maximum posterior probability $p(y^*|u_{li})$ of y^*, as a linear function of X, we define a new continuous variable $z_{u_{li}}$ that replaces $\log(p(y^*|u_{li}))$,

$$z_{u_{li}} = \log(p(y^*|u_{li})) \; \forall u_{li} \tag{11}$$

and define a binary variable $s_{n,u_{li}}$ which indicates that $y_n \neq y^*$ as

$$s_{n,u_{li}} = \begin{cases} 1 & \text{if } y_n \neq y^* \\ 0 & \text{if } y_n = y^* \end{cases} \quad \forall n, u_{li} \tag{12}$$

Then we have the constraints that if $s_{n,u_{li}} = 0$, then $z_{u_{li}} = \log(p(y_n|u_{li}))$, which is equivalent to $z_{u_{li}} \leq \log(p(y_n|u_{li}))$ to maximize $z_{u_{li}}$. If $s_{n,u_{li}} = 1$, $z_{u_{li}}$ does not have any constraints with $\log(p(y_n|u_{li}))$; this condition is equivalent to $z_{u_{li}} \leq \infty$. Accordingly, we have

Constraint 1. *For all i,*

$$z_{u_{li}} \leq \log p(y_n|u_{li}) + B s_{n,u_{li}} \tag{13}$$

where B is an arbitrary big number to represent ∞.

Also, there exists only one target disease y^* with largest value of $p(y^*|u_{li})$ in each u_{li}. That is, there exist $N-1$ candidate diseases that are not y^* at each leaf node u_l:

Constraint 2. *For all i,*

$$\sum_{n=1}^{N} s_{n,u_{li}} = N - 1. \tag{14}$$

5.2.2 No reuse and full coverage. Because a test is only applied once, the number of test t_m used in $path(u_{li}, \cdot)$ is less than or equal to one:

Constraint 3. *For all m and i*

$$\sum_{k=0}^{K-2} \sum_{j=1}^{q} x_{m,path(u_{li},k),j} \leq 1. \tag{15}$$

DiagTree should contain at least one test t_m that can detect the disease y_n to be positive. We define binary indicator variables a_{nm} and a tolerance τ to ensure tests to detect diseases with probability that is significantly > 0 such that

$$a_{nm} = \begin{cases} 1 & \text{if } P_{nm} > \tau \\ 0 & \text{if } P_{nm} \leq \tau \end{cases} \quad \forall n, m. \tag{16}$$

Then, test t_m with $a_{nm} = 1$ is used throughout the DiagTree $a_{nm}\left(\sum_u \sum_j x_{m,u,j}\right)$ times. The total number of tests used with $a_{nm} = 1$ is $\sum_m a_{nm}\left(\sum_u \sum_j x_{m,u,j}\right)$:

Constraint 4. *For all n,*

$$\sum_m a_{nm}\left(\sum_u \sum_j x_{m,u,j}\right) \geq 1. \tag{17}$$

5.2.3 Full structure. DiagTree has a structure of a full 2^q-way tree, and each node has q tests (including dummy test):

Constraint 5. *For all u,*

$$\sum_m \sum_j x_{m,u,j} + \sum_j x_{\phi,u,j} = q, \tag{18}$$

and only one test (including dummy test) can be assigned on each position:

Constraint 6. *For all u, j*

$$x_{\phi,u,j} + \sum_{m=1}^{M} x_{m,u,j} = 1. \tag{19}$$

Also, we have constraints on dummy tests that child nodes of negative results are all assigned to dummy tests in Eq. (3):

Constraint 7. *Eq. (3).*

To sum up, our problem is formulated as finding the optimal $X = \{x_{m,u,j}\}$ such that

$$\begin{aligned}&\text{argmax}_X \sum_i \Bigg[N z_{u_{li}} \\ &- \left(\sum_{n=1}^{N} \sum_{k=0}^{K-2} \sum_{j=1}^{q} \sum_{m=1}^{M} \log p(t_m|y_n) x_{m,path(u_{li},k),j}\right)\Bigg] \\ &- \mu \cdot \sum_{m, m\neq\phi} \sum_u \sum_j x_{m,u,j}\end{aligned}$$

subject to

$$z_{u_{li}} \leq \log p(y_n|u_{li}) + Bs_{n,u_{li}} \quad \forall i$$

$$\sum_{n=1}^{N} s_{n,u_{li}} = N-1 \quad \forall i$$

$$\sum_{k=0}^{K-2}\sum_{j=1}^{q} x_{m,path(u_{li},k),j} \leq 1 \quad \forall m,i$$

$$\sum_{m} a_{nm}\left(\sum_{u}\sum_{j} x_{m,u,j}\right) \geq 1 \quad \forall n$$

$$\sum_{m}\sum_{j} x_{m,u,j} + \sum_{j} x_{\phi,u,j} = q \quad \forall u$$

$$x_{\phi,u,j} + \sum_{m=1}^{M} x_{m,u,j} = 1 \quad \forall u,j.$$

5.3 Branch-and-Bound Algorithm

This optimization task is a non-convex *mixed-integer programming* problem [12], so we use a Branch-and-Bound (BNB) algorithm, which is a non-heuristic method to find the local optimal solution from integer programming problem. The task of finding the global optimal solution in POMDP models is a difficult open challenge, and is beyond the scope of this paper. The time complexity of POMDP and Dec-POMDP are exponential and non-deterministic exponential [4], respectively. Particularly, Dec-POMDP requires that a guess about the solution be generated in a non-deterministic way, and verifying whether the guess is a solution takes exponential time. Finding an approximate local optimum solution for Dec-POMDP is also non-deterministic exponential time [17].

The BNB algorithm gradually finds local optima while maintaining provable lower and upper bounds on the global optimal solution [2]. Specifically, we first transform our problem to a linear programming problem by relaxing binary variables to be continuous variables. This linear programming problem can be solved in polynomial time by using the simplex method [20]. We restore the relaxed continuous variables to binary variables by iteratively selecting the most 'promising' one. That is, we select the relaxed variable that least decreases the objective value when we set the variable to 0 or 1.

6 TIME COMPLEXITY

Because usually $Q \leq 6$ in practice due to insurance policy, the number of DiagTree's internal nodes is $S = \frac{(2^q)^{Q/q}-1}{2^q-1} \approx 2^{(Q-q)} \leq 2^5$. Time complexity of BNB is exponential to obtain the global optimum. The number of binary variables used in BNB is $\#x_{m,u,j} + \#s_{n,u_{li}} = MSq + NS = (Mq+N)S$; so BNB have the worst-case complexity of $O(2^{(M+N)S})$ in the case of exhaustive search for if all variables are binary, and the best-case complexity of $O((M+N)S)$ when always a relaxed binary variable exists that does not decrease the objective value at every iteration [15]. In contrast, greedy approaches have linear time complexity on M, N, and S. For all DiagTree's internal nodes, linear search to select q tests takes $(2^q)^0NM + (2^q)^1N(M-q) + (2^q)^2N(M-2q) + \ldots + (2^q)^{Q/q-1}N(M-(Q/q-1)q)$, or $O(MNS)$. Thus, for large M and N, the greedy approaches work efficiently to find DiagTree; for relatively small M and N, the BNB approach can also work. In practice, the number M of tests that are considered to use together does not exceed 30 because functions of the tests are clearly separated among them. The number N of diseases that need careful discrimination by tests also does not exceed 10 because in differential diagnosis, clinicians use the tests to discriminate only similar diseases.

7 EXPERIMENTS

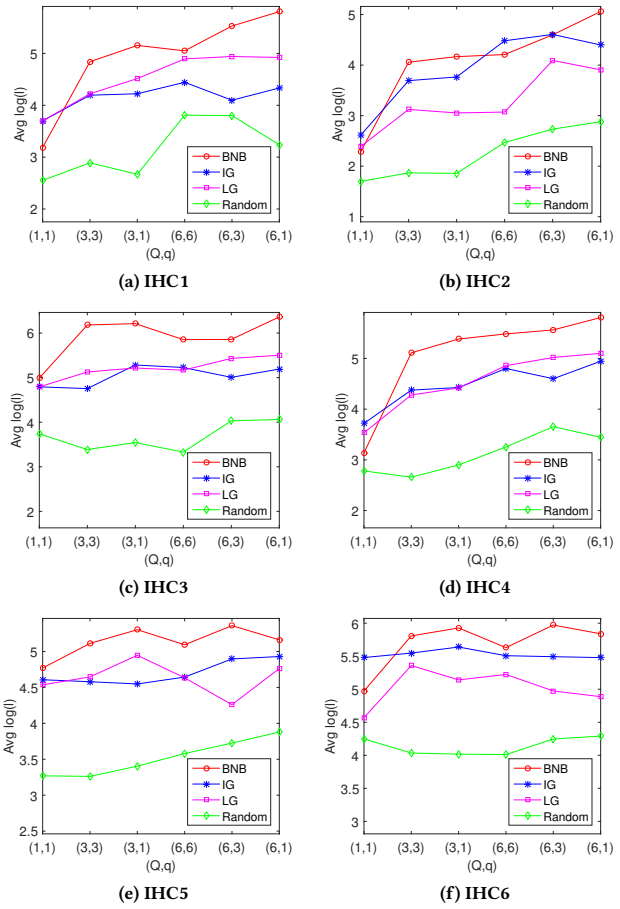
We apply DiagTree to immunohistochemistry (IHC) profiles (Sectioin 7.1) and evaluate DiagTree in two ways: i) evaluate the overall confidence (or likelihood of leaf nodes) and runtime of the algorithms at various Q and q (Section 7.2), and ii) compare the accuracy and cost of the diagnosis rules from DiagTree with those obtained using clinicians' knowledge (Section 7.3). We demonstrate that DiagTree actually produces diagnosis rules that cost the same or less than clinicians' rules. We also apply DiagTree to a synthetic dataset and perform threshold analysis of μ , but due to the space limit, evaluation on synthetic data and the thresholds are omitted and can be found in [13].

7.1 IHC Profiles

We apply DiagTree to the IHC profiles that pathologists use for immunohistochemical staining. IHC staining is the process of using specific antibodies to detect antigens within tissue sections, and is used to diagnose abnormal cells found in cancer tumors [7]. Recently, the growing number of the antibodies across multiple diseases makes pathologists difficult to use intuition alone to design the diagnosis process [22]; thus DiagTree can reduce burden on pathologists by suggesting an optimized diagnosis process. IHC profiles contain probabilities that 115 antibody tests will be positive given 104 lymphoid neoplasm diseases ($p(t|y)$). The diseases are categorized using the World Health Organization's classification [18]. Among $115 \times 104 = 11,960$ pairs of test and disease, only 4,245 have known probability values, which were collected from [6] and other sources by clinicians. The other 7,715 pairs have no clinical proofs of interactions; for those pairs, we assume independence between test and disease, and skip the test to detect the disease, i.e., $p(y|t_1,t_2,t_3) = p(y|t_2,t_3)$ when $p(t_1|y)$ is unknown and replaced by $p(t_1)$. We divide the full profile into six sub-profiles according to the disease sets of interest that need sophisticated differential diagnosis. The six sub-profiles can be found in [13]. Prior probabilities of diseases are computed by normalizing the number of disease instances reported in [26]. Note that DiagTree can be also applied to various diagnosis process other than IHC profiles such as detecting drug infusion, kidney transplantation [19], and ischemic heart disease [10].

7.2 Algorithm Evaluation

On the six sub-profiles, we evaluate the overall confidence of DiagTree from BNB with three baseline methods – Greedy Information Gain (IG), Greedy Likelihood Gain (LG), and random (i.e., selects tests randomly). DiagTree's confidence is measured by averaging the log posterior probability ratio ($log(l)$) in Eq.(2). After several trials, we set $\epsilon = 10^{-4}$ and maximum runtime as 7,200 s. We set the thresholds $\tau = 0.05$, and set the threshold μ to 0 (Fig. 7) for

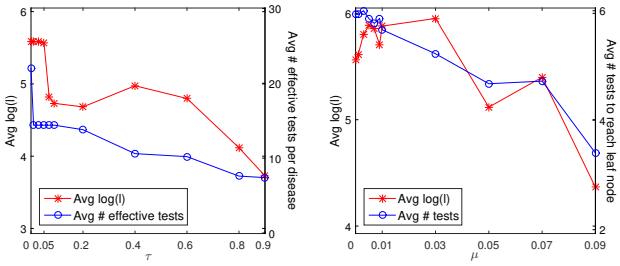


(a) IHC1 (b) IHC2 (c) IHC3 (d) IHC4 (e) IHC5 (f) IHC6

Figure 6: Average log likelihood over various Q and q on IHC profiles.

Table 2: Runtime (sec.) of BNB, IG, and LG

		(Q, q)					
		(1,1)	(3,3)	(3,1)	(6,6)	(6,3)	(6,1)
IHC1	BNB	0.038	5.74	0.307	3421	1968	3218
	IG	0.024	0.008	0.027	0.011	0.032	0.063
	LG	0.016	0.007	0.023	0.013	0.031	0.082
IHC2	BNB	0.037	0.051	0.067	1296	1927	3752
	IG	0.001	0.002	0.005	0.01	0.018	0.044
	LG	0.002	0.002	0.006	0.012	0.021	0.058
IHC3	BNB	0.039	0.301	0.747	1495	2343	3678
	IG	0.001	0.002	0.006	0.011	0.017	0.053
	LG	0.001	0.002	0.007	0.012	0.02	0.067
IHC4	BNB	0.037	5.79	0.478	3772	2768	2950
	IG	0.001	0.002	0.006	0.011	0.02	0.054
	LG	0.001	0.002	0.008	0.012	0.02	0.068
IHC5	BNB	0.01	0.044	5.815	5851	1466	2355
	IG	0.001	0.002	0.005	0.011	0.015	0.046
	LG	0.001	0.002	0.006	0.012	0.018	0.056
IHC6	BNB	0.014	0.147	0.693	3084	6809	13.8
	IG	0.001	0.002	0.006	0.011	0.017	0.056
	LG	0.001	0.002	0.008	0.012	0.021	0.072



(a) τ (b) μ

Figure 7: Sensitivity analysis on thresholds τ and μ.

fair comparison with baselines. We measure $\log(l)$ of DiagTree with various combinations of q and Q, i.e., (1, 1), (3, 3), (3, 1), (6, 6), (6, 3), and (6, 1).

As a result, BNB produces the highest $\log(l)$ on most IHC subprofiles (Fig. 6). IG produces higher $\log(l)$ than LG on IHC2 and IHC6, LG produces higher $\log(l)$ than IG on IHC1, and IG and LG produce comparable $\log(l)$ on IHC3, IHC4, and IHC5. All three methods produce significantly higher $\log(l)$ than the random baseline. $\log(l)$ increased as q decreased with fixed Q because tests are selected after observing all the previous test results; and $\log(l)$ increased as Q increases with fixed q because more tests are used.

Because BNB has exponential time complexity in the worst case, it is slower than IG and LG. Most DiagTrees except (6, 1) on IHC5 and IHC6 reached adequate local optimal solutions within the time limit (Table 2). DiagTree with (6, 1) on IHC5 and IHC6, which is more information-rich than (6, 6) or (6, 3), reached a local optimal solution, but a local optimum that had higher $\log(l)$ than that of (6, 6) or (6, 3) was not reached within the time limit. In contrast, IG and LG took < 0.03 s for all cases.

DiagTree uses τ to determine whether the test m is effective to detect the disease n (i.e., $P_{nm} > \tau$). We evaluate the sensitivity of the BNB's $\log(l)$ and the average number of effective tests per disease (i.e., $(1/n)\sum_n \sum_m a_{nm}$) with respect to different τ on IHC4 (Fig. 7a). We select $\tau = 0.05$ for optimal threshold, because the number of effective tests abruptly decreases as τ increases from 0 to 0.05, and the BNB's $\log(l)$ then abruptly decreases after $\tau = 0.05$.

We also evaluate the sensitivity of the BNB's $\log(l)$ and the average number of tests to reach leaf node in DiagTree with respect to different μ, the weighting constant to miminize the number of tests used in Eq. (6) (Fig. 7b). We found that the $\log(l)$ and the number of tests decrease as μ increases, and $\mu = 0.03$ is the optimal threshold.

7.3 Comparison with Clinicians' Rules

We compare the accuracy and cost of the diagnosis rules from DiagTree and clinicians' practice.

7.3.1 Diagnosis Rules from DiagTree. We present DiagTree with (6, 3) on IHC4 (Fig. 8); other DiagTrees can be found in [13].

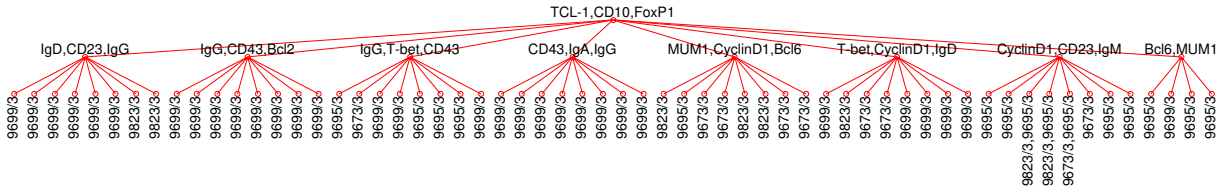


Figure 8: Example of DiagTree on (6, 3) from IHC4 built by BNB.

Antibodies of TCL-1, CD10, and FoxP1 are first applied to examine which disease is the target disease among 9823/3, 9673/3, 9680/3 CD5+, 9699/3, and 9695/3, If the test results are all positive, Bcl6 and MUM1 are then applied. If Bcl6 is negative and MUM1 is positive, the target disease is 9699/3: Extranodal marginal zone lymphoma of mucos-associated lymphoid tissue (MALT lymphoma).

7.3.2 Accuracy. We evaluate the accuracy of DiagTree when clinicians actually applied it for diagnosis. We simulate application of DiagTree using the past test results from pathologic reports of 131 patients at Yeouido St. Mary's hospital, Korea obtained from 2009 to 2015. The pathologic reports contain patients' antibody test names, results of the test, and disease codes. We discard patients who had diseases that are not in IHC profiles or that occurred less than five times. The pathologic reports contain 24 diseases in total; nine diseases are common in both IHC profiles and the pathologic reports, so we have 102 patients with the nine frequent diseases. Some test results are unavailable when the tests are not performed. In that case, we impute the test results according to probability in IHC profiles. Finally, we diagnose the diseases as our diagnosis rules from DiagTree with the expected test results from the pathologic reports. We then compare the detected disease and the actual disease to derive the accuracy, true positive rate (TPR), and false positive rate (FPR). DiagTree detects the nine diseases with high accuracy, high TPR and low FPR (Table 3). Among the three DiagTrees with setting of (Q, q)= (6, 1), (6, 3), and (6, 6), the DiagTree with = (6, 1) has the highest accuracy and TPR.

7.3.3 Cost. We compare the cost (i.e., the number of tests and stages to diagnose diseases) of diagnosis rules from DiagTree with those of the rules manually produced by clinicians. We extract manually-produced diagnosis rules from the pathologic reports of Yeouido St. Mary's hospital and UC San Diego Medical Center (UCSD). The UCSD pathologic reports contain 2,029 patients with 49 tests and 89 disease codes from 2009 to 2014, but do not include test results. In total, 17 diseases occur in both hospitals and IHC profiles. For these 17 diseases, we compute the average number of tests used and the estimated number of stages (i.e., maximum waiting time) by assuming that up to three tests can be performed simultaneously. We then compute the cost from DiagTree for the 17 diseases. To measure the confidence of the results from DiagTree, we compute the average posterior probabilities at leaf nodes for each disease. Compared to clinician's rules, DiagTree produces lower costs from most diseases (Table 4). For example, clinicians usually need 8.9 tests and three stages to diagnose mantle cell lymphoma. When clinicians use DiagTree, they detects it using six tests and one stage with 0.91 confidence. DiagTree produces relatively low posterior probabilities on the last four diseases, i.e., (1) Adult T-cell leukemia/lymphoma, (2) Diffuse large B-cell lymphoma (DLBCL) CD5+, (3) B-cell lymphoma between DLBCL and Burkitt lymphoma, and (4) Anaplastic large cell lymphoma, ALK negative. DiagTree detects these diseases with posterior probabilities < 0.23. This poor result occurs because these diseases are very rare and lack clinical test data [26]. The prior probabilities of the four diseases are 0.0019 (in IHC5), 0.0009 (in IHC4), 0.0003 (in IHC6), 0.0060 (in IHC1), respectively. Also DiagTree does not reduce the number of tests for Plasma cell neoplasms; clinicians' rules detect plasma cell neoplasms with four tests in three stages.

8 CONCLUSION AND FUTURE WORK

This paper presents DiagTree, a new decision process framework for diagnosing diseases that minimizes the cost to finish the decision process and satisfies the actual setting of differential diagnosis. DiagTree consists of multiple tests in each internal node, and posterior probabilities (or confidences) of diagnoses at each leaf node. A disease with maximum confidence is detected after a series of test results are applied in internal nodes of DiagTree. Experimental results on synthetic and IHC profile datasets demonstrate that DiagTree with BNB algorithm produces a decision process that has high confidence, and also reduces the cost to finish the tests with high accuracy compared to existing diagnosis procedures designed by clinicians. As a result, DiagTree can support clinicians by suggesting simple diagnosis processes of high accuracy and low cost from the test candidates. Future work will focus on incorporating importance (such as reliability or price) of each test in our formulation and on reducing the runtime of BNB approach by using an isomorph-free BNB search [9].

Acknowledgements: This research was partly supported by Brain Korea 21 PLUS project for POSTECH Computer Science and Engineering Institute, Kyung-hee University (2014-0-00147), NRF (2016R1E1A1A01942642, 2016R1D1A1A02937427), NIH (R01GM118574, R01GM118609, and U01EB023685).

REFERENCES

[1] M. Araya, O. Buffet, V. Thomas, and F. Charpillet. A pomdp extension with belief-dependent rewards. In *Advances in Neural Information Processing Systems*, pages 64–72, 2010.
[2] V. Balakrishnan, S. Boyd, and S. Balemi. Branch and bound algorithm for computing the minimum stability degree of parameter-dependent linear systems. *International Journal of Robust and Nonlinear Control*, 1(4):295–317, 1991.
[3] C. C. Bennett and K. Hauser. Artificial intelligence framework for simulating clinical decision-making: A markov decision process approach. *Artificial intelligence in medicine*, 57(1):9–19, 2013.
[4] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein. The complexity of decentralized control of markov decision processes. *Mathematics of operations research*, 27(4):819–840, 2002.
[5] E. Campo, S. H. Swerdlow, N. L. Harris, S. Pileri, H. Stein, and E. S. Jaffe. The 2008 who classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood*, 117(19):5019–5032, 2011.
[6] C. V. Cotta. Diagnostic immunohistochemistry theranostic and genomic applications by david j. dabbs, 2010.

Table 3: Accuracy of DiagTree for diagnosing the nine diseases

ICD-O-3	Disease	(Q, q) = (6, 1)			(Q, q) = (6, 3)			(Q, q) = (6, 6)		
		ACC	TPR	FPR	ACC	TPR	FPR	ACC	TPR	FPR
9699/3	MALT lymphoma	0.96	1	0.04	0.91	1	0.11	0.94	0.93	0.06
9695/3	Follicular lymphoma	0.97	1	0.03	0.94	0.67	0.05	0.93	0.33	0.04
9673/3	Mantle cell lymphoma	0.97	0.88	0.02	0.94	0.75	0.05	0.94	1	0.07
9680/3	DLBCL, NOS, (with c-myc rearrangement)	0.97	1	0.06	0.94	1	0.13	0.97	1	0.06
9680/3 CD5+	Diffuse large B-cell lymphoma (DLBCL) CD5+	0.96	0.67	0.03	0.91	0	0.07	0.95	0	0.03
9719/3	Extranodal NK/T-cell lymphoma. nasal type	0.96	1	0.04	0.94	1	0.07	0.97	1	0.03
9702/3	Peripheral T-cell lymphoma. NOS	0.96	0.5	0.03	0.94	1	0.06	0.97	1	0.03
9714/3 negative	Anaplastic large cell lymphoma ALK negative	0.96	0	0.01	0.94	0	0.04	0.97	0	0
9650/3	Classical Hodgkin lymphoma introduction	0.96	1	0.04	0.94	1	0.07	0.97	1	0.03

Table 4: Costs of DiagTree compared to clinician's rules.

ICD-O-3	Disease	Clinician's rule		DiagTree		
		# tests	# stages	#tests	# stages	Prob.
9687/3	Burkitt lymphoma	11	7	6	1	> 0.95
9766/1	Lymphomatoid granulomatosis	5	2	6	2	> 0.95
9811/3	B lymphoblastic leukaemia/lymphoma NOS	10.5	4	6	2	> 0.95
9673/3	Mantle cell lymphoma	8.9	3	6	1	0.91
9695/3	Follicular lymphoma	6	2	6	1	0.87
9823/3	Chronic lymphocytic leukaemia /small lymphocytic lymphoma	8	3	6	3	0.83
9650/3	Classical Hodgkin lymphoma introduction	6	3	6	3	0.76
9659/3	Nodular lymphocyte predominant Hodgkin lymphoma	6.1	4	6	6	0.77
9702/3	Peripheral T-cell lymphoma. NOS	10.5	4	6	1	0.73
9719/3	Extranodal NK/T-cell lymphoma nasal type	7.1	3	6	1	0.70
9714/3	Anaplastic large cell lymphoma ALK positive	7	5	6	6	0.66
9731/3	Plasma cell neoplasms	4	3	6	3	0.61
9699/3	MALT lymphoma	6	2	6	2	0.54
9827/3	Adult T-cell leukemia/lymphoma	18	5	6	6	0.23
9680/3 CD+	DLBCL CD5+	9.3	4	6	6	<0.01
9680/3, 9687/3	B-cell lymphoma between DLBCL and Burkitt lymphoma	9	3	6	6	<0.01
9714/3 negative	Anaplastic large cell lymphoma ALK negative	11	4	6	6	<0.01

[7] D. J. Dabbs. *Diagnostic immunohistochemistry.* Elsevier Health Sciences, 2013.

[8] F. Glover. Future paths for integer programming and links to artificial intelligence. *Computers & operations research*, 13(5):533–549, 1986.

[9] M. Grzes, P. Poupart, and J. Hoey. Isomorph-free branch and bound search for finite state controllers. In *IJCAI*. Citeseer, 2013.

[10] M. Hauskrecht and H. Fraser. Planning treatment of ischemic heart disease with partially observable markov decision processes. *Artificial Intelligence in Medicine*, 18(3):221–244, 2000.

[11] C. Hu, W. S. Lovejoy, and S. L. Shafer. Comparison of some suboptimal control policies in medical drug therapy. *Operations Research*, 44(5):696–709, 1996.

[12] L. Hyafil and R. L. Rivest. Constructing optimal binary decision trees is np-complete. *Information Processing Letters*, 5(1):15–17, 1976.

[13] Y. Kim, J. Choi, Y. Chong, X. Jiang, and H. Yu. *DiagTree : Diagnostic Tree for Differential Diagnosis.* http://dm.postech.ac.kr/DiagTree, 2017.

[14] M. L. Littman. Memoryless policies: Theoretical limitations and practical results. In *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, volume 3, page 238. MIT Press, 1994.

[15] C. J. McDiarmid and G. M. Provan. An expected-cost analysis of backtracking and non-backtracking algorithms. In *IJCAI*, pages 172–177, 1991.

[16] R. Nair, M. Tambe, M. Yokoo, D. Pynadath, and S. Marsella. Taming decentralized pomdps: Towards efficient policy computation for multiagent settings. In *IJCAI*, 2003.

[17] Z. Rabinovich, C. V. Goldman, and J. S. Rosenschein. The complexity of multiagent systems: the price of silence. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 1102–1103. ACM, 2003.

[18] E. Sabattini, F. Bacci, C. Sagramoso, and S. Pileri. Who classification of tumours of haematopoietic and lymphoid tissues in 2008: an overview. *Pathologica*, 102, 2010.

[19] A. J. Schaefer, M. D. Bailey, S. M. Shechter, and M. S. Roberts. Modeling medical treatment using markov decision processes. In *Operations research and health care*, pages 593–612. Springer, 2005.

[20] A. Schrijver. *Theory of linear and integer programming.* John Wiley & Sons, 1998.

[21] H. I. R. . A. Service. Health insurance coverage, 2017.

[22] D. Shin, G. Arthur, C. Caldwell, M. Popescu, M. Petruc, A. Diaz-Arias, and C.-R. Shyu. A pathologist-in-the-loop ihc antibody test selection using the entropy-based probabilistic method. *Journal of pathology informatics*, 3, 2012.

[23] S. J. Steindel and P. J. Howanitz. Physician satisfaction and emergency department laboratory test turnaround time: observations based on college of american pathologists q-probes studies. *Archives of pathology & laboratory medicine*, 125(7):863–871, 2001.

[24] C. R. Taylor. Ihc and the who classification of lymphomas: cost effective immunohistochemistry using a deductive reasoning "decision tree" approach. *Applied Immunohistochemistry & Molecular Morphology*, 17(5):366–374, 2009.

[25] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making*, pages 141–162. Springer, 1975.

[26] S. O. Yoon, C. Suh, D. H. Lee, H.-S. Chi, C. J. Park, S.-S. Jang, H.-R. Shin, B.-H. Park, and J. Huh. Distribution of lymphoid neoplasms in the republic of korea: analysis of 5318 cases according to the world health organization classification. *American journal of hematology*, 2010.