Discover the Expert: Context-Adaptive Expert Selection for Medical Diagnosis

Cem Tekin, *Member, IEEE*, Onur Atan, Mihaela van der Schaar, *Fellow, IEEE*

***Abstract*—In this paper we propose an expert selection system that learns online the best expert to assign to each patient depending on the context of the patient. In general, the context can include an enormous number and variety of information related to the patient's health condition, age, gender, previous drug doses, etc., but the most relevant information is embedded in only a few contexts. If these most relevant contexts were known in advance, learning would be relatively simple but they are not. Moreover, the relevant contexts may be different for different health conditions. To address these challenges, we develop a new class of algorithms aimed at discovering the most relevant contexts and the best clinic and expert to use to make a diagnosis given a patient's contexts. We prove that as the number of patients grows, the proposed context-adaptive algorithm will discover the optimal expert to select for patients with a specific context. Moreover, the algorithm also provides confidence bounds on the diagnostic accuracy of the expert it selects, which can be taken into account by the primary care physician before making the final decision. While our algorithm is general and can be applied in numerous medical scenarios, we illustrate its functionality and performance by applying it to a real-world breast cancer diagnosis dataset. Finally, while the application we consider in this paper is medical diagnosis, our proposed algorithm can be applied in other environments where expertise needs to be discovered.**

***Index Terms*—Semantic computing, context-adaptive learning, clinical decision support systems, healthcare informatics, distributed multi-user learning, contextual bandits.**

I. Introduction

One of the most important applications of semantic computing [1] is healthcare informatics [2]. The development of healthcare informatics tools and decision support systems is vital, since recent studies show that standard clinical practice often fails to fit the patient [3]. The landscape of healthcare is rapidly changing as the model for reimbursement shifts from fee-for-service, which emphasizes increasing volume, to pay-for-performance, which emphasizes improving quality of care and reducing costs [4]. Healthcare organizations are now tasked with developing metrics for measuring quality in terms of outcomes, patient experience, workflow efficiency, access, and organization. The widespread adoption of electronic health records (EHRs) to capture data routinely generated as part of standard of care is yielding new opportunities to leverage such information for quality improvement and evidence-based medicine. Nevertheless, an ongoing challenge is how to effectively apply this high-dimensional and unstructured dataset to support clinical decision making (e.g., determining the correct diagnosis) and improve resource management (e.g., matching a patient with the clinician who best handled "similar" cases, while also considering the workload of clinicians). This paper aims to optimize clinical workflows by personalizing the match of (new patient) cases with the appropriate diagnostic expertise whether a Clinical Decision Support Systems (CDSS), a domain expert who specializes in similar types of cases, or another institution. In current clinical practice, patients are referred to experts in an ad-hoc manner based on one or more of the following factors: signs and symptoms of the patient, patient's or primary care physician's preference, insurance plan, and availability of the physician. This paper develops a framework and associated methods and algorithms that uses semantic knowledge about the patient to assess and recommend expertise with the goal of optimizing the process for diagnosing a patient.

We assume that the diagnostic accuracy of an expert (either human or CDSS) depends on the *context* of the patient for which the decision is made. This context is all the information pertaining to the patient under consideration that can be utilized in the decision making process. For instance, in breast cancer diagnosis context includes patient profile, breast density, assessment history, characteristics of the opposite breast, modalities, etc., or in general, electronic medical records can be used as the context [5]. Since the context of a patient has many dimensions, learning the diagnostic accuracy of an expert suffers from the curse of dimensionality. The methodology we propose in this paper learns the *most relevant* context(s) pertinent to the current health condition of the patient and uses it/them to estimate the level of expertise exhibited by the expert. The level of expertise is defined based on the accuracy of their diagnostic. Moreover, different clinics have healthcare professionals with different expertise and some of these clinics may have access to CDSSs from different manufacturers and of different types while some others just rely on human experts. In our proposed system, these clinics can cooperate with each other to improve diagnostic accuracy by learning the contextual specializations of the other clinics (see Fig. 1). For instance, a rural clinic may have only a primary care physician (PCP), a registered nurse and equipment, but no specialist

The material is based upon work funded by the US Air Force Research Laboratory (AFRL) under the DDAS program. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not reflect the views of AFRL.

C. Tekin: Corresponding Author, Department of Electrical Engineering, UCLA, Los Angeles, CA, 90095. Email: cmtkn@ucla.edu. Phone: (310) 825-5843. Fax: (310) 206-4685.

O. Atan: Department of Electrical Engineering, UCLA, Los Angeles, CA, 90095. Email: oatan@ucla.edu. Phone: (310) 825-5843. Fax: (310) 206-4685.

M. van der Schaar: Professor, Department of Electrical Engineering, UCLA, Los Angeles, CA, 90095. Email: mihaela@ee.ucla.edu. Phone: (310) 825-5843. Fax: (310) 206-4685.

The authors would like to acknowledge prof. William Hsu (Radiology Department, UCLA) for numerous useful discussions and guidance and for providing us the data set used in the illustrative results section.

or CDSSs and may be able to request information from a more established hospital which implements multiple CDSSs and has several experienced human experts. Hence, based on the context of the patient, each clinic learns (i) whether it should rely on its own experts or request another clinic to make diagnostic decisions, and (ii) if it relies on its own experts, which expert it should assign the task of deciding the diagnostic, such that the reward (gain) obtained by selecting that particular expert is maximized. The reward for a particular diagnostic decision can be defined as the diagnostic accuracy minus the incurred cost (e.g. delay, money, etc.). Our proposed system learns online, meaning that its expert selection strategy is updated every time after the true health state of a patient is revealed. (Note that at times the true health state is not revealed immediately.) Based on this feedback, the expertise, i.e., the diagnostic accuracy, of the chosen expert is updated.

We model this problem as a distributed context-adaptive online learning problem. Each clinic decides on what diagnostic action to take based on the history of its own patient arrivals, patient arrivals to other clinics that requested a diagnostic action from the current clinic, and the success rate of each diagnostic action. In this way, each clinic is able to identify which experts make accurate decisions for patients with specific contexts.

The main contributions of the methods and algorithms presented in this paper are: (i) organizing the clinical data for each patient in terms of semantic contexts; (ii) discovering the most relevant context or set of patient contexts that are useful for assessing an expert; (iii) having the ability to provide confidence estimates on the accuracy of the selected diagnostic made by the experts; (iv) having the ability to select the "optimal" expert (a CDSS, a domain expert etc.) among one or multiple institutions, where "optimality" is defined by considering both the diagnosis accuracy and the experts' costs (workload, money charged, etc.).

The remainder of the paper is organized as follows. In Section II, we describe the related work. In Section III, we formalize the problem, and in Section IV, we present the proposed distributed, context-adaptive online algorithm which learns the best expert for diagnosing a patient based on his/her contextual information. In Section V, we discuss some extensions of our formalism. In Section VI, we illustrate the proposed system using a real-world breast cancer diagnostic data. Finally, the concluding remarks are given in Section VII.

II. Related Work

We categorize the related work into two key areas: work related to semantic computing, and work related to data mining and online learning.

A. Semantic computing

Semantic computing focuses on computing based on semantics ("context", "meaning", "intention") and it addresses all types of resources including data, document, tool, device, process and people [1]. Within the area of semantic computing, rule-based reasoning systems [6], [7] have emerged which deploy a database of the facts that are known about the problem currently being solved, and a decision engine which combines rules with the data to produce predictions. In these systems the decision rule is developed by a group of human experts, and rules are updated over time based on their effectiveness. Our proposed methodology fits within the class of semantic-based reasoning systems. However, in contrast to the existing work, we consider multiple experts, each adopting its own decision rule. Moreover, how well a specific decision rule (diagnostic) performs when applied to a patient, characterized by a specific context, is not known a priori. Hence, in this work we are interested in developing a rigorous and efficient methodology for learning how to select the expert adopting the best decision rule (diagnostic) for each patient.

B. Data mining and learning

Most of the prior work in online stream mining provides algorithms which are asymptotically converging to an optimal or locally-optimal solution without providing any rates of convergence. We do not only prove convergence results, but we are also able to explicitly characterize the performance loss incurred at each time slot (for each patient) with respect to the optimal solution.

Some of the existing solutions (including [8]–[10]) propose ensemble learning techniques which combine the diagnosis of multiple experts into a final diagnosis. In our work we only consider choosing the best expert (initially unknown), where the expert selection process is driven by the patient's context. This is especially important in resource constrained scenarios like healthcare informatics, where the human resources are limited either in terms of the number of experts that are making diagnostic decisions or the number of healthcare personnel that acts as an interface between the patient and the CDSS. We provide a detailed comparison to our work in Table I. As seen from Table I, our proposed system is context-adaptive, distributed, outputs confidence bounds, and provides an explicit rate of convergence to the optimal expert selection strategy as the number of patients grows.

In addition to the problems in data mining, our methods can be applied to any problem that can be formulated as a distributed contextual bandit problem. Contextual bandits have been studied before in [11]–[13] and other works in a single agent setting. Our work is very different from these because (i) we consider decentralized agents (clinics) who can learn to cooperate with each other, (ii) the set of available (diagnostic) actions and the context arrivals to the agents can be very different for each agent, (iii) instead of learning to take the best action considering the entire D-dimensional context vector, an agent learns to take the marginally best action by independently considering each D types of contexts, hence learning is much faster than existing learning algorithms whose convergence speed slows down exponentially with the dimension of the context space [14]. Due to its context-adaptive property, the order of the convergence speed of the algorithm we propose in this paper is independent of the dimension of the context space.

III. Problem Formulation

The system model is shown in Fig. 2 and Fig. 3. There are M clinics (learners) which are indexed by the set $\mathcal{M} :=$

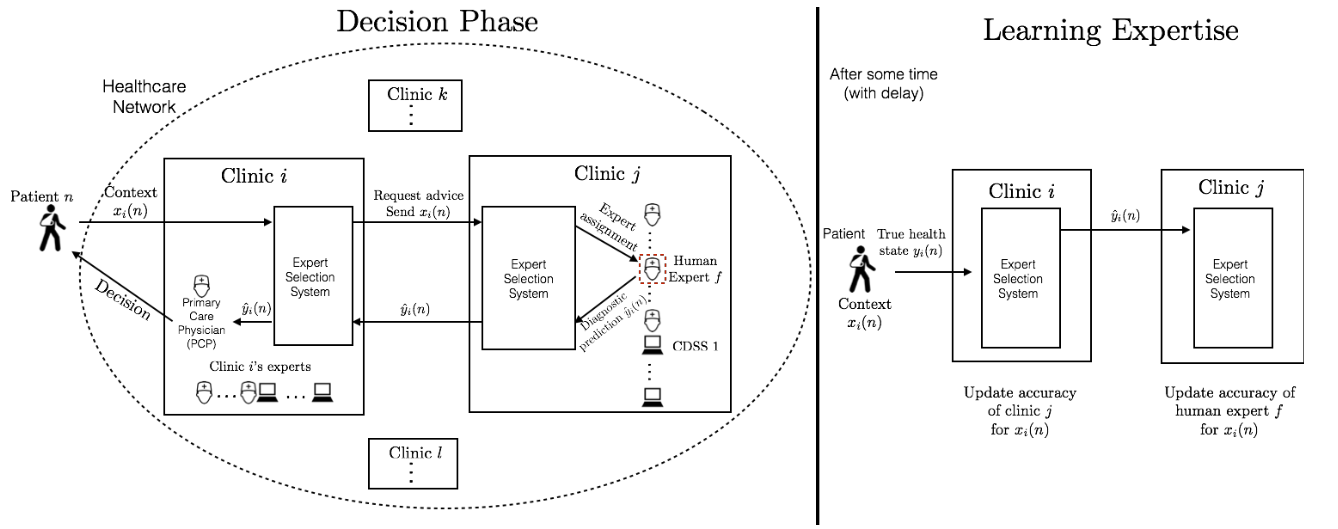

Fig. 1. Operation of the proposed system for clinic i. In this example, the diagnostic decision for the patient with context $x_i(n)$ is made by a human expert from clinic j. Then, after some time the true health state $y_i(n)$ of the patient is revealed. Based on this clinic i updates the diagnostic accuracy of clinic j for that context, while clinic j updates the diagnostic accuracy of its expert f.

	[15]–[18]	[19], [20]	[21]	[12], [14]	This work
Message exchange	none	context	training residual	none	context (adaptively)
Learning approach	offline/online	offline	offline	Non-Bayesian online	Non-Bayesian online
Learning from other's contexts	N/A	no	no	no	yes
Using other's experts	no	all	all	no	sometimes-adaptively
Rate of convergence	no	no	no	yes - dimension dependent	yes - dimension independent
Context adaptive	no	no	no	no	yes
Confidence bounds	no	no	no	yes	yes

TABLE I
COMPARISON WITH RELATED WORK IN DATA MINING AND LEARNING.

$\{1, 2, \ldots, M\}$. The set of experts clinic i has is $\mathcal{F}_i$. As we discussed an expert can either be a human expert or a CDSS. The set of all experts is $\mathcal{F} = \cup_{i \in \mathcal{M}} \mathcal{F}_i$. Let $\mathcal{M}_{-i} := \mathcal{M} - \{i\}$ be the set of clinics clinic i can choose from to send its patient's context for diagnosis. The *diagnostic action set*[1] of clinic i is $\mathcal{K}_i := \mathcal{F}_i \cup \mathcal{M}_{-i}$. Throughout the paper we use index f to denote an element of $\mathcal{F}$, j to denote clinics in $\mathcal{M}_{-i}$, and k to denote an element of $\mathcal{K}_i$. Let $M_i := |\mathcal{M}_{-i}|$, $F_i := |\mathcal{F}_i|$ and $K_i := |\mathcal{K}_i|$, where $|\cdot|$ is the cardinality operator. A summary of notations is provided in Table II.

For each patient $n = 1, 2, \ldots, N$, the following events happen sequentially: (i) The nth patient with a D-dimensional context vector $\boldsymbol{x}_i(n) = (x_i^1(n), \ldots, x_i^D(n))$ arrives to clinic $i \in \mathcal{M}$, where $x_i^d(n) \in \mathcal{X}_d$ for $d \in \mathcal{D} := \{1, \ldots, D\}$ and $\mathcal{X}_d$ is the set of type-d contexts, and $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_D$ is the context space,[2] (ii) each clinic i assigns one of its own experts or another clinic to recommend a diagnosis $\hat{y}_i(n) \in \mathcal{Y}$ for the patient n, where $\mathcal{Y}$ is the set of possible diagnosis recommendations (in the application of breast cancer, this set includes breast tumor being malignant or benign), (iii) after some delay, the true health state of patient $y_i(n) \in \mathcal{Y}$ is revealed only to the clinic i where the patient has arrived,[3] (iv) if another clinic provided the diagnosis for that patient, then the clinic where the patient arrived passes the true health state of the patient to that clinic.

A. Context, diagnosis, diagnostic action accuracies

For each patient n, the context vector $\boldsymbol{x}_i(n)$ and true health state of the patient $y_i(n)$ are assumed to be drawn from an (unknown) joint distribution J over $\mathcal{X} \times \mathcal{Y}$ independently from the other patients. We do not require this draw to be independent among the clinics/learners. Given a context vector $\boldsymbol{x}$, there exists a conditional distribution $G_{\boldsymbol{x}}$ over $\mathcal{Y}$ (which depends on J). Similarly, depending on J, there is a marginal distribution H over $\mathcal{X}$ from which contexts are drawn. Given context vector $\boldsymbol{x}$, let $\pi_f(\boldsymbol{x}) := \int_{y \in \mathcal{Y}} \mathrm{I}(f(\boldsymbol{x}) = y) dG_{\boldsymbol{x}}(y)$ be the *joint accuracy* (or simply, accuracy) of expert $f \in \mathcal{F}$, where $f(\boldsymbol{x})$ is the diagnosis recommendation of expert f for context vector $\boldsymbol{x}$.[4] The diagnostic rule used by expert f, i.e., $f(\cdot)$ is allowed to be deterministic or random. $\mathrm{I}(\cdot)$ is the indicator function which is equal to 1 if the statement inside is true and 0 otherwise, and the expectation $\mathrm{E}[\cdot]$ is taken with respect to distribution $G_{\boldsymbol{x}}$. Let $\boldsymbol{x}^{-d} := (x^1, \ldots, x^{d-1}, x^{d+1}, \ldots, x^D)$ and $((\boldsymbol{x}')^{-d}, x^d) := (x'^1, \ldots, x'^{d-1}, x^d, x'^{d+1}, \ldots, x'^D)$. Then,

[1] In sequential online learning literature [22], [23], an action is also called an arm (or an alternative).

[2] Each dimension represents a different type of context. For example, first dimension may represent age, second dimension may represent weight, third dimension may represent gender, etc. In our analysis, we will assume that $\mathcal{X}_d = [0, 1]$ for all $d \in \mathcal{D}$. However, our algorithms will work and our results will hold even when the context space is discrete given that it is bounded. For instance, $\mathcal{X}_d$ can represent the set of normalized ages (formed for instance, by dividing the exact age of the patient with the maximum age 200). Then, every patient's normalized age will lie in $\mathcal{X}_d$.

[3] Our algorithm will also work when the true health state of some patients is never recovered by simply disregarding the history related to that patients.

[4] Although for simplicity of exposition we assumed that the decision only depends on the context vector, the radiological image can also be a part of the information sent to the expert which is denoted by $s_i(n)$. Then assuming that this data is i.i.d. given a context vector, the decision rule can be extended as $f(\boldsymbol{x}_i(n), s_i(n))$, and the expert accuracy can be defined analogously.

the *marginal accuracy* of expert f based on type-d context is defined as

$$\pi_f^d(x^d) := \int_{(\boldsymbol{x}')^{-d}} \pi_f((\boldsymbol{x}')^{-d}, x^d) dH((\boldsymbol{x}'^{-d}), x^d).$$

We say that the problem has *the similarity property* when each expert has similar marginal accuracies for similar contexts.

Definition 1: **Similarity Property.** If there exists $\alpha > 0$ and $L > 0$ such that for all $f, f' \in \mathcal{F}$ and $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$, we have

$$|\pi_f^d(x^d) - \pi_f^d((x')^d)| \leq L|x^d - (x')^d|^\alpha, \ \forall d \in \mathcal{D},$$

and

$$\pi_f(\boldsymbol{x}) - \pi_{f'}(\boldsymbol{x}) \leq Z\left(\max_{d\in\mathcal{D}} \pi_f^d(x^d) - \max_{d\in\mathcal{D}} \pi_{f'}^d(x^d)\right),$$

for some $Z > 0$, then we call J *a distribution with similarity constant* L *and similarity exponent* α.

Although, our model assumes a continuous context space, our algorithms will also work when the context space is discrete. Note that Definition 1 does not require the context space to be continuous. We assume that α is known by the clinics, while L does not need to be known. However, our algorithms can be combined with estimation methods for α. In reality, the knowledge of α is not required for our algorithms to run, however an estimate of α should be given as an input. If the estimate $\hat{\alpha}$ is chosen conservatively such that $\hat{\alpha} < \alpha$, the performance bounds we prove for our algorithm (Theorem 1 and Corollary 1) will hold with α replaced by $\hat{\alpha}$.

B. Unknowns, experts and diagnostic rewards

For each patient n, clinic i can either assign one of its experts or forward the patient's context to another clinic to have him/her diagnosed. We assume that for clinic i, assigning each expert $f \in \mathcal{F}_i$ incurs a cost $c_f^i \geq 0$. We assume that whenever the patient's context is sent to another clinic $j \in \mathcal{M}_{-i}$ a cost of c_j^i is incurred by clinic i.[5] This cost can be the delay and/or monetary costs associated with the forwarding action. When the diagnostic action $k \in \mathcal{K}_i$ is chosen for the nth patient of clinic i, and the diagnosis recommendation $\hat{y}_i(n)$ is made, the reward is equal to $r_i(n) := \mathrm{I}(\hat{y}_i(n) = y_i(n)) - c_k^i$. This reward is observed only after the true health state $y_i(n)$ is revealed. Since the costs are bounded, without loss of generality we assume that costs are normalized, i.e., $c_k^i \in [0, 1]$ for all $k \in \mathcal{K}_i$. The clinics are cooperative which implies that clinic $j \in \mathcal{M}_{-i}$ will return a diagnosis recommendation to i when called by i using its expert with the highest estimated diagnostic accuracy for i's context vector. Similarly, when called by $j \in \mathcal{M}_{-i}$, clinic i will return a diagnostic recommendation to j. In our theoretical analysis we do not consider the effect of this on i's learning rate; however, since our results hold for the case when other clinics are not forwarding their patient's context to i, they will also hold when other clinics forward the patient's context to i. Indeed, learning is faster for clinic i when other clinics ask clinic i for a diagnostic recommendation for their patients.

[5]The cost for clinic i does not depend on the cost of the expert chosen by clinic j. Since the clinics are cooperative, j will obey the rules of the proposed algorithm when assigning an expert to diagnose clinic i's patient. We assume that when called by clinic i, clinic j will select an expert from $\mathcal{F}_j$, but not forward i's patient's context to another clinic.

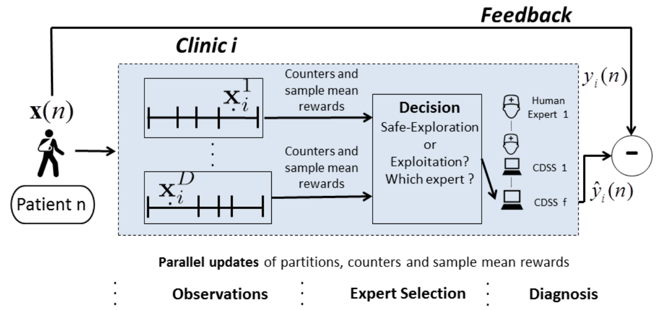

Fig. 2. Operation of clinic i for its nth patient when it chooses one of its own experts.

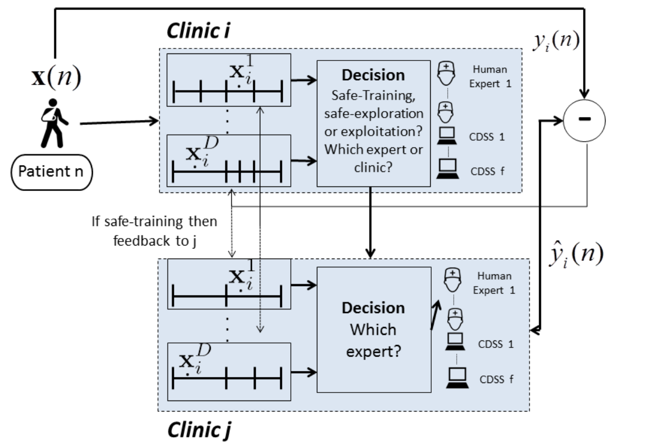

Fig. 3. Operation of clinic i for its nth patient when it chooses clinic j.

We assume that each expert produces a binary diagnostic recommendation,[6] thus $\mathcal{Y} = \{0, 1\}$. The best expert of a clinic $j \in \mathcal{M}$ for context vector $\boldsymbol{x}$ is $f_j^*(\boldsymbol{x}) = \arg\max_{f\in\mathcal{F}_j} \max_{d\in\mathcal{D}} \pi_f^d(x^d)$. Hence the accuracy of clinic j for context $\boldsymbol{x}$ is defined as $\pi_j(\boldsymbol{x}) := \pi_{f_j^*(\boldsymbol{x})}(\boldsymbol{x})$. Clinic j's marginal accuracy for a type-d context x^d is equal to the accuracy of its best expert, i.e.,

$$\pi_j^d(x^d) := \max_{f\in\mathcal{F}_j} \pi_f^d(x^d).$$

The goal of each clinic i is to maximize its total expected reward. This corresponds to minimizing the regret with respect to the benchmark solution which we will define in the next subsection.

C. Diagnosis with complete information

Our benchmark when evaluating the performance of the learning algorithms is the solution which selects the diagnostic action in $\mathcal{K}_i$ with the highest marginal accuracy minus cost (i.e., reward) given the context vector of the patient. For $k \in \mathcal{K}_i$, and $d \in \mathcal{D}$, let $\mu_k^d(x^d) := \pi_k^d(x^d) - c_k^i$. Specifically, the solution we compare against is given by

$$k_i^*(\boldsymbol{x}) := \arg\max_{k\in\mathcal{K}_i}\left(\max_{d\in\mathcal{D}} \mu_k^d(x^d)\right), \quad \forall \boldsymbol{x} \in \mathcal{X}.$$

[6]In general we can assume that diagnostic recommendations belong to $\mathbb{R}$ and define diagnostic error as some other metric. Our results can be adapted to this case as well.

Since calculating $k_i^*(\boldsymbol{x})$ requires knowledge of marginal expert accuracies only, we call $k_i^*(\boldsymbol{x})$ *the marginally best diagnostic action* given patient's context $\boldsymbol{x}$. Knowing this means that clinic i knows the expert in $\mathcal{F}$ that yields the highest diagnostic reward for each $x^d \in \mathcal{X}_d$, $d \in \mathcal{D}$. We call a policy that always acts according to this action an *optimal policy*.

D. The regret of learning

Let $a_i(n)$ be the diagnostic action of clinic i for its nth patient and $b_{i,j}(n)$ be the expert of clinic i that is assigned to the patient n of clinic j, when clinic j requests a diagnosis recommendation from clinic i. If there is no such request, then $b_{i,j}(n) = \emptyset$. Let $\boldsymbol{a}(n) := (a_1(n), \ldots, a_M(n))$, $\boldsymbol{b}_i(n) := \{b_{i,j}(n)\}_{j \in \mathcal{M}_{-i}}$ and $\boldsymbol{b}(n) = \{\boldsymbol{b}_i(n)\}_{i \in \mathcal{M}}$. Simply, the regret is the loss incurred due to the unknown expertise. Regret of a learning algorithm which assigns an expert $a_i(n) \in \mathcal{K}_i$ for patient n in the clinic i based on its context vector $\boldsymbol{x}_i(n)$ and the past observations is defined as

$$R_i(N) := \sum_{n=1}^{N} \left(\pi_{k_i^*(\boldsymbol{x}_i(n))}(\boldsymbol{x}_i(n)) - c^i_{k_i^*(\boldsymbol{x}_i(n))} \right) - \mathrm{E}\left[\sum_{n=1}^{N} (\mathrm{I}(\hat{y}_i(n) = y_i(n)) - c^i_{a_i(n)}) \right],$$

where $\hat{y}_i(n)$ denotes the diagnostic recommendation of the expert or other clinic $a_i(n)$ assigned by clinic i to the patient n, $y_i(n)$ denotes the true health state of the patient n that arrived to clinic i. For instance, when $a_i(n) = j$ and $b_{j,i}(n) = f \in \mathcal{F}_j$, then $\mathrm{E}[\mathrm{I}(\hat{y}_i(n) = y_i(n))] = \pi_f(\boldsymbol{x}_i(n))$. Regret gives the convergence rate of the total expected reward of the learning algorithm to the value of the benchmark solution $k_i^*(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{X}$. Any algorithm whose regret is sublinear, i.e., $R_i(N) = O(N^\gamma)$ such that $\gamma < 1$, will converge to the benchmark solution in terms of the average reward.

$\mathcal{M}$: Set of all clinics.
$\mathcal{F}_i$: Set of experts of clinic i. $F_i := \lvert\mathcal{F}_i\rvert$.
$\mathcal{M}_{-i}$: Set of clinics except clinic i. $M_i := \lvert\mathcal{M}_{-i}\rvert$.
$\mathcal{K}_i$: Set of diagnostic actions for clinic i. $K_i := \lvert\mathcal{K}_i\rvert$.
$\mathcal{F}$: Set of all experts.
$\mathcal{X} = [0,1]^D$: Context space.
$\mathcal{X}_d = [0,1]$: Context space for type-d context.
$\mathcal{Y}$: Set of all possible diagnosis recommendations.
$\pi_f(\boldsymbol{x})$: Joint accuracy of expert f given context vector $\boldsymbol{x}$.
$\pi_f^d(\boldsymbol{x}^d)$: Marginal accuracy of expert f for type-d context x^d.
c_k^i: Diagnostic cost of action k.
$k_i^*(\boldsymbol{x})$: Marginally best diagnostic action given context vector $\boldsymbol{x}$.
$R_i(N)$: Regret of clinic i up to patient N.

TABLE II
NOTATIONS USED IN PROBLEM FORMULATION.

IV. ADAPTIVELY LEARNING THE RELEVANT CONTEXTS

In this section we propose an online learning algorithm that achieves regret that is sublinear in the number of patients. We name our algorithm *Learn the EXpert* (LEX).

A. The LEX algorithm

The basic idea behind LEX is to learn the accuracies of different clinics and different experts by requesting diagnosis recommendations from them in a cost efficient way. Using LEX, a clinic can perform two tasks: (i) decide the diagnostic action to take for own patient; (ii) decide the expert to assign to the patients of other clinics which requested a diagnosis recommendation. Task (i) is performed by sub-algorithm *LEX for own patients* (LEX-OP), while task (ii) is performed by sub-algorithm *LEX for referred patients* (LEX-RP). The pseudocodes of LEX, LEX-OP and LEX-RP are given in Figures 4, 5 and 6, respectively. A summary of notations used in LEX is given in Table III.

There are three operation phases of LEX: exploitation, safe training and safe exploration. For any clinic and any patient LEX is only in one of these phases. In an exploitation phase, LEX is very confident about its expert selection decision. As we will show in Corollary 1, it is able provide confidence bounds on the probability that it selects the best expert among all possible experts and on the accuracy of the prediction made by the chosen expert. In the safe training phase, clinic i is not confident about how well some other clinic j knows its best expert for clinic i's patient. Hence, clinic i requests a diagnosis recommendation from clinic j which helps clinic j learn the accuracy of its own experts. In the safe exploration phase, clinic i is not confident about the accuracy of its diagnostic actions. It will choose a diagnostic action and receive a diagnosis recommendation to update the accuracy of the chosen diagnostic action (which is done after the true health state is revealed). Trainings and explorations are safe, which means that LEX alerts the clinician that is in charge of the patient that the diagnosis recommendation comes from an expert which may not be very reliable. Knowing this, the clinician may assign another expert or may choose to follow the recommendation based on his/her own expertise. This way the system learns, while the patient safety is not compromised. Whenever we refer to training and exploration, we mean safe training and safe exploration.

LEX adaptively divides the context space into finer and finer regions as more patients arrive such that the regions of the context space with large number of arrivals are trained and explored more accurately than regions of the context space with small number of arrivals, and then only uses the observations in those regions when estimating the rewards of diagnostic actions in $\mathcal{K}_i$ for contexts that lie in those regions. For each patient, LEX chooses a diagnostic action adaptively based on the estimated marginal accuracy of the diagnostic action given the context vector.

For each type-d context, LEX starts with a partition with a single element which is the entire context space $\mathcal{X}_d$, then divides the space into finer regions and explores them as more patients with those contexts arrive. In this way, LEX focuses on parts of the context space in which there are large number of patient arrivals, and does this independently for each type of context of the patients.

For each type-d context, we call an interval $(a2^{-l}, (a+1)2^{-l}] \subset [0,1]$ a level l hypercube for $a = 1, \ldots, 2^l - 1$,[7] where l is an integer. Let $\mathcal{P}_l^d$ be the partition of type-d context space $[0,1]$ generated by level l hypercubes. Clearly, $|\mathcal{P}_l^d| = 2^l$. Let $\mathcal{P}^d := \cup_{l=0}^{\infty} \mathcal{P}_l^d$ denote the set of all possible hypercubes. Note that $\mathcal{P}_0^d$ contains only a single hypercube

[7] The first level l hypercube is defined as $[0, 2^{-l}]$.

LEX for clinic i:

1: Input: $D_1(n)$, $D_2(n)$, $D_3(n)$, p, A
2: Initialization: $\mathcal{A}_i^d = \{[0,1]\}$, $d \in \mathcal{D}$. Run **Initialize**($\mathcal{A}_i^d$, d), $d \in \mathcal{D}$.
3: **while** $n \geq 1$ **do**
4: Run LEX-OP to get a_i, $\boldsymbol{C}_i$, d_i^*, C_i^*, train$_i$ and safe$_i$.
5: If $a_i \in \mathcal{M}_{-i}$, send $\boldsymbol{x}_i(t)$ to clinic a_i and request $\hat{y}_i(n)$.
6: Receive $\{\boldsymbol{x}_j(n)\}_{j \in \mathcal{M}_i(n)}$.
7: **if** $\mathcal{M}_i(n) \neq \emptyset$ **then**
8: Run LEX-RP to get $\boldsymbol{b}_i := \{b_{i,j}\}_{j \in \mathcal{M}_i(n)}$, $\{\boldsymbol{C}_j\}_{j \in \mathcal{M}_i(n)}$, $\{d_j^*\}_{j \in \mathcal{M}_i(n)}$, $\{C_j^*\}_{j \in \mathcal{M}_i(n)}$ and $\{\text{safe}_j\}_{j \in \mathcal{M}_i(n)}$.
9: **for** $j \in \mathcal{M}_i(n)$ **do**
10: Get $\hat{y}_j(n)$ from expert $b_{i,j}$. Send $\hat{y}_j(n)$, safe$_j$ to clinic j.
11: **if** not safe$_j$ **then**
12: Alert clinic j that $\hat{y}_j(n)$ may be inaccurate.
13: **end if**
14: **end for**
15: **end if**
16: **if** $a_i \in \mathcal{F}_i$ **then**
17: Pay $c^i_{a_i}$, get $\hat{y}_i(n)$ from expert a_i.
18: **else**
19: Pay $c^i_{a_i}$, get $\hat{y}_i(n)$ from clinic a_i.
20: **end if**
21: **if** not safe$_i$ **then**
22: Alert the PCP that $\hat{y}_i(n)$ may be inaccurate.
23: **end if**
24: When $y_i(n)$ is revealed, set $acc_i = \mathrm{I}(\hat{y}_i(n) = y_i(n))$, get reward $r_i(n) = acc_i - c^i_{a_i}$.
25: **if** train$_i$ **then**
26: $T^{\text{tr},i,d_i^*}_{a_i,C_i^*}++$.
27: **else if** not safe$_i$ **then**
28: $\hat{\pi}^{i,d_i^*}_{a_i,C_i^*} = \left(\hat{\pi}^{i,d_i^*}_{a_i,C_i^*} T^{i,d_i^*}_{a_i,C_i^*} + acc_i\right) / \left(T^{i,d_i^*}_{a_i,C_i^*} + 1\right)$.
29: $T^{i,d_i^*}_{a_i,C_i^*}++$.
30: **else**
31: Exploit but do not update.
32: **end if**
33: $T^{i,d}_{C_i^d}++$ for all $d \in \mathcal{D}$.
34: **for** $d \in \mathcal{D}$ **do**
35: **if** $T^{i,d}_{C_i^d} \geq A2^{pl_{C_i^d}}$ **then**
36: Create 2 level $l_{C_i^d}+1$ child hypercubes denoted by $\mathcal{A}_{C_i^d}$
37: Run **Initialize**($\mathcal{A}_{C_i^d}$, d)
38: $\mathcal{A}_i^d = \mathcal{A}_i^d \cup \mathcal{A}_{C_i^d} - C_i^d$
39: Send d and $\mathcal{A}_{C_i^d}$ to clinics $j \in \mathcal{M}_{-i}$ such that they can update clinic i's partition $\boldsymbol{\mathcal{A}}_i$.
40: **end if**
41: **end for**
42: **if** $\mathcal{M}_i(n) \neq \emptyset$ **then**
43: **for** $j \in \mathcal{M}_i(n)$ **do**
44: Obtain $y_j(n)$ from clinic j. Set $acc_j = \mathrm{I}(\hat{y}_j(n) = y_j(n))$.
45: **if** not safe$_j$ **then**
46: $\hat{\pi}^{i,d_j^*}_{b_{i,j},C_j^*} = \left(\hat{\pi}^{i,d_j^*}_{b_{i,j},C_j^*} T^{i,d_j^*}_{b_{i,j},C_j^*} + acc_j\right) / \left(T^{i,d_j^*}_{b_{i,j},C_j^*} + 1\right)$.
47: $T^{i,d_j^*}_{b_{i,j},C_j^*}++$.
48: **end if**
49: **end for**
50: **end if**
51: $n = n + 1$
52: **end while**

Initialize($\mathcal{A}$, d):

for $C \in \mathcal{A}$ **do**
Set $T^{i,d}_C = 0$, $T^{i,d}_{k,C} = 0$, $\hat{\pi}^{i,d}_{k,C} = 0$ for $k \in \mathcal{K}_i$, $T^{\text{tr},i,d}_{j,C} = 0$ for $j \in \mathcal{M}_{-i}$.
end for

Fig. 4. Pseudocode for LEX.

LEX-OP for clinic i:

Find $\boldsymbol{C}_i(n)$ in $\boldsymbol{\mathcal{A}}_i$ that contains $\boldsymbol{x}_i(n)$. Set $\boldsymbol{C}_i = \boldsymbol{C}_i(n)$.
if $\exists\, f \in \mathcal{F}_i$, $d \in \mathcal{D}$: $T^{i,d}_{f,C_i^d} \leq D_1(n)$ **then**
Set $a_i = f$, $d_i^* = d$, $C_i^* = C_i^d$, train$_i = false$, safe$_i = false$.
else if $\exists\, j \in \mathcal{M}_{-i}$, $d \in \mathcal{D}$: $T^{\text{tr},i,d}_{j,C_j^d} \leq D_2(n)$ **then**
Set $a_i = j$, $d_i^* = d$, $C_i^* = C_i^d$, train$_i = true$, safe$_i = false$.
else if $\exists\, j \in \mathcal{M}_{-i}$, $d \in \mathcal{D}$: $T^{i,d}_{j,C_j^d} \leq D_3(n)$ **then**
Set $a_i = j$, $d_i^* = d$, $C_i^* = C_i^d$, train$_i = false$, safe$_i = false$.
else
Set $a_i \in \arg\max_{k \in \mathcal{K}_i} \left(\max_{d \in \mathcal{D}} \hat{\pi}^{i,d}_{k,C_i^d} - c^i_k\right)$, $d_i^* = \arg\max_{d \in \mathcal{D}} \hat{\pi}^{i,d}_{a_i,C_i^d}$, $C_i^* = C_i^{d_i^*}$, train$_i = false$, safe$_i = true$.
end if

Fig. 5. Pseudocode for LEX-OP.

LEX-RP for clinic i.

for $j \in \mathcal{M}_i(n)$ **do**
Find $\boldsymbol{C}_j(n)$ in $\boldsymbol{\mathcal{A}}_j$ that contains $\boldsymbol{x}_j(n)$. Set $\boldsymbol{C}_j = \boldsymbol{C}_j(n)$.
if $\exists\, f \in \mathcal{F}_i$, $d \in \mathcal{D}$: $T^{i,d}_{f,C_j^d} \leq D_1(n)$ **then**
Set $b_{i,j} = f$, $d_j^* = d$, $C_j^* = C_i^d$, safe$_j = false$.
else
Set $b_{i,j} \in \arg\max_{f \in \mathcal{F}_i} \left(\max_{d \in \mathcal{D}} \hat{\pi}^{i,d}_{f,C_j^d}\right)$, $d_j^* = \arg\max_{d \in \mathcal{D}} \hat{\pi}^{i,d}_{b_{i,j},C_j^d}$, $C_j^* = C_j^{d_j^*}$, safe$_j = true$.
end if
end for

Fig. 6. Pseudocode for LEX-RP.

which is $\mathcal{X}_d$ itself. LEX keeps for the clinic i a set of mutually exclusive hypercubes that cover the context space of each type-d context. We call these hypercubes *active* hypercubes, and denote the set of active hypercubes for type-d context for patient n of clinic i by $\mathcal{A}_i^d(n)$. Let $\boldsymbol{\mathcal{A}}_i(n) := (\mathcal{A}_i^1(n), \ldots, \mathcal{A}_i^D(n))$ and $\boldsymbol{\mathcal{A}}(n) := \{\boldsymbol{\mathcal{A}}_i(n)\}_{i \in \mathcal{M}}$. All clinics know $\boldsymbol{\mathcal{A}}(n)$, so that they can form marginal sample mean diagnostic action reward estimates of their experts for the hypercubes of other clinics. This can be done via a simple message exchange between the clinics, where a clinic reports its new partition to the other clinics only when it updates its partition as shown in lines 34-41 of LEX. Clearly, we have $\cup_{C \in \mathcal{A}_i^d(n)} C = \mathcal{X}_d$. Denote the active hypercube that contains $x_i^d(n)$ by $C_i^d(n)$. Let $\boldsymbol{C}_i(n) := (C_i^1(n), \ldots, C_i^D(n))$ be the set of active hypercubes that contains $\boldsymbol{x}_i(n)$. The diagnostic action chosen by clinic i for patient n only depends on the diagnostic actions taken on previous context observations which are in $C_i^d(n)$ for some $d \in \mathcal{D}$. The number of such actions and observations can be much larger than the number of previous actions and observations in $\boldsymbol{C}_i(n)$. This is because in order for an observation to be in $\boldsymbol{C}_i(n)$, it should be in all $C_i^d(n)$, $d \in \mathcal{D}$. As we will explain below, for each patient n, LEX selects a diagnostic action based on a particular type of context that lies in $\mathcal{D}$. This type for patient n of clinic i is called the *main type* and is denoted by $d_i^*(n)$, and the hypercube that contains $x_i^{d_i^*(n)}(n)$ is called the *main hypercube* and is denoted by $C_i^*(n)$.

Before going into the details of the operation of LEX, we define several counters which count the number of occurances of specific events that will be used by LEX to decide whether to train, explore or exploit.

Definition 2: LEX uses the following counters to determine whether to train, explore or exploit for each patient n.

- $T^{i,d}_{C}(n)$: The number of patients of clinic i that arrived within the time window from the activation of C to the arrival of patient n whose type-d contexts are in C.
- $T^{\text{tr},i,d}_{j,C}(n)$: The number of patients of clinic i that arrived within the time window from the activation of C to the arrival of patient n whose type-d contexts are in C, for which LEX was in the training phase, the main type was d, and the diagnostic action was requested from clinic $j \in \mathcal{M}_{-i}$.
- $T^{i,d}_{j,C}(n)$, $j \in \mathcal{M}_{-i}$: The number of patients of clinic i that arrived in the time window from the activation of C to the arrival of patient n whose type-d contexts are in C, for which LEX was in exploration phase, the main type was d, and the diagnostic action was requested from clinic $j \in \mathcal{M}_{-i}$.
- $T^{i,d}_{f,C}(n)$, $f \in \mathcal{F}_i$: The number of patients that arrived to clinic i (own patients and patients from other clinics) within the time window from the activation of C to the arrival of patient n whose type-d contexts are in C, for which LEX was in exploration phase, the main type was d, and the diagnostic action was requested from expert $f \in \mathcal{F}_i$.

Once activated, a level l hypercube C will stay active until n such that $T^{i,d}_{C}(n) \geq A2^{pl}$, where $p > 0$ and $A > 0$ are parameters of LEX. After that, LEX will divide C into 2 level $l+1$ hypercubes. For each expert in $\mathcal{F}_i$, LEX have a single (deterministic) control function $D_1(n)$ which controls when to do safe exploration or exploitation, while for each clinic in $\mathcal{M}_{-i}$, LEX have two (deterministic) control functions $D_2(n)$ and $D_3(n)$, where $D_2(n)$ controls when to do safe training, $D_3(n)$ controls when to do safe exploration or exploitation when there are enough trainings.

For a type d hypercube C let

$$\mathcal{F}^{i,d}_{\text{ue},C}(n) := \{f \in \mathcal{F}_i : T^{i,d}_{f,C}(n) \leq D_1(t)\},$$

be the set of under-explored experts of clinic i,

$$\mathcal{M}^{i,d}_{\text{ue},C}(n) := \{j \in \mathcal{M}_{-i} : T^{i,d}_{j,C}(n) \leq D_3(t)\},$$

be the set of under-explored clinics in $\mathcal{M}_{-i}$, and

$$\mathcal{M}^{i,d}_{\text{ut},C}(n) := \{j \in \mathcal{M}_{-i} : T^{\text{tr},i,d}_{j,C}(n) \leq D_2(t)\},$$

be the set of under-trained clinics in $\mathcal{M}_{-i}$. For $\boldsymbol{C} = (C^1, \dots, C^D)$, let $\mathcal{F}^{i}_{\text{ue},\boldsymbol{C}}(n) := \bigcup_{d\in\mathcal{D}} \mathcal{F}^{i,d}_{\text{ue},C^d}(n)$, $\mathcal{M}^{i}_{\text{ue},\boldsymbol{C}}(n) := \bigcup_{d\in\mathcal{D}} \mathcal{M}^{i,d}_{\text{ue},C^d}(n)$ and $\mathcal{M}^{i}_{\text{ut},\boldsymbol{C}}(n) := \bigcup_{d\in\mathcal{D}} \mathcal{M}^{i,d}_{\text{ut},C^d}(n)$.

When patient n arrives to clinic i, LEX first finds $\boldsymbol{C}_i(n) \in \boldsymbol{\mathcal{A}}_i(n)$. Then, it enters training, exploration or exploitation phase based on the following order: (i) If $\mathcal{F}^{i}_{\text{ue},\boldsymbol{C}_i(n)}(n) \neq \emptyset$, then randomly select $a_i(n) \in \mathcal{F}^{i}_{\text{ue},\boldsymbol{C}_i(n)}(n)$ to explore it; (ii) Else if $\mathcal{M}^{i}_{\text{ut},\boldsymbol{C}_i(n)}(n) \neq \emptyset$, then randomly select $a_i(n) \in \mathcal{M}^{i}_{\text{ut},\boldsymbol{C}_i(n)}(n)$ to train it; (iii) Else if $\mathcal{M}^{i}_{\text{ue},\boldsymbol{C}_i(n)}(n) \neq \emptyset$, then randomly select $a_i(n) \in \mathcal{M}^{i}_{\text{ue},\boldsymbol{C}_i(n)}(n)$ to explore it; (iv) Else, if all the above sets are empty, then exploit the diagnostic action which have the highest marginal sample mean reward, i.e., $a_i(n) \in \arg\max_{k\in\mathcal{K}_i} \left(\max_{d\in\mathcal{D}} \bar{r}^{i,d}_{k,C^d_i(n)}(n)\right)$, where $\bar{r}^{i,d}_{k,C^d_i(n)}(n) = \hat{\pi}^{i,d}_{k,C^d_i(n)}(n) - c^i_k$.

Marginal sample mean diagnostic action accuracies, i.e., $\hat{\pi}^{i,d}_{k,C}(n)$, $k \in \mathcal{K}_i$, $C \in \mathcal{A}^d_i(n)$, $d \in \mathcal{D}$, are calculated only based on the diagnostic results $\mathrm{I}(\hat{y}_i(n') = y_i(n'))$ for the patients n' who arrived between the activation of C and patient n, for which LEX explored k, the main type was d and main hypercube was C. Note that the sample mean diagnostic action accuracies are not updated when LEX is in training or exploitation phase. The reason for not updating in training phase is that the clinic chosen by LEX may not know its best expert accurately, while the reason for not updating in exploitation phase is that the diagnostic action, and hence the main type, is chosen by looking at the values of all contexts, which will introduce bias to the sample mean estimate.

By this sample mean update mechanism, as the number of true health state observations increases, it is guaranteed that $\hat{\pi}^{i,d}_{k,C}(n)$ converges to a number very close to the true marginal expected accuracy of action k for contexts in C. This, together with the adaptive partitioning guarantees that the regret remains sublinear in the number of patients.

$\mathcal{A}^d_i(n)$: Set of type-d active hypercubes by patient n.
$\boldsymbol{\mathcal{A}}_i(n) := (\mathcal{A}^1_i(n), \dots, \mathcal{A}^D_i(n))$.
$C^d_i(n)$: Active hypercube that contains $x^d_i(n)$. $\boldsymbol{C}_i(n) := (C^1_i(n) \dots C^D_i(n))$.
$\mathcal{M}_i(n)$: Set of clinics that requested diagnosis recommendation from clinic i for their nth patient.
$\hat{\pi}^{i,d}_{k,C}(n)$: Marginal sample mean diagnostic accuracy of diagnostic action k of clinic i for type-d contexts in hypercube C.
$D_1(n), D_2(n), D_3(n)$: Control functions.

TABLE III
NOTATIONS USED IN LEX.

Next, we explain how clinic i assigns experts to other clinics' patients who request diagnosis recommendation from clinic i. Let $\mathcal{M}_i(n)$ be the set of clinics that request diagnosis recommendation from clinic i for their nth patient. For these clinics, clinic i selects its experts using LEX-RP. To do this, it first identifies the set of hypercubes $\boldsymbol{C}_j(n) \in \boldsymbol{\mathcal{A}}_j(n)$ that contains $\boldsymbol{x}_j(n)$. Then, it enters exploration or exploitation phase based on the following order: (i) If $\mathcal{F}^{i}_{\text{ue},\boldsymbol{C}_j(n)}(n) \neq \emptyset$, then randomly select $b_{i,j}(n) \in \mathcal{F}^{i}_{\text{ue},\boldsymbol{C}_j(n)}(n)$ to explore it; (ii) Else exploit the expert with the highest marginal sample mean accuracy, i.e., $b_{i,j}(n) \in \arg\max_{f\in\mathcal{F}_i} \left(\max_{d\in\mathcal{D}} \hat{\pi}^{i,d}_{f,C^d_j(n)}(n)\right)$. Again, the marginal sample mean accuracy of the chosen expert is updated only if the expert is explored, after the true label for the patient n of clinic j is sent to clinic i.

B. Analysis of the regret of LEX

In this subsection we analyze the regret of LEX and derive a sublinear upper bound on the regret, whose order of growth with N does not depend on D. We divide the regret $R_i(N)$ into three different terms. $R^e_i(N)$ is the regret of clinic i due to trainings and exploitations by patient N, $R^s_i(N)$ is the regret of clinic i due to selecting suboptimal diagnostic actions at exploitation steps by patient N, and $R_i^{ne}(N)$ is the regret of clinic i due to selecting near-optimal diagnostic actions in exploitation steps by patient N. Using the fact that LEX separate trainings, explorations and exploitations over time, and linearity of expectation operator, we get $R_i(n) = R_i^e(n) + R_i^s(n) + R_i^{ne}(n)$. In the following analysis, we will bound each part of the regret separately.

Lemma 1: **Regret due to safe trainings and safe explorations in a hypercube.** Consider clinics that use LEX with parameters $D_1(n) = D_3(n) = n^z \log n$ and $D_2(n) = F_{\max} n^z \log n$. For clinic i consider any level l hypercube for type-d contexts. The regret of clinic i in such a hypercube due to safe trainings and safe explorations up to its Nth patient is $O(MF_{\max}N^z \log N)$.

Proof: See Appendix A-B. ■

For clinic i let $\mu_k^d(x^d) := \pi_k^d(x^d) - c_k^i$, be the expected marginal reward of diagnostic action $k \in \mathcal{K}_i$ for a patient with type-d context $x^d \in \mathcal{X}_d$. For a type-d hypercube C, let $\dot{x}_C^d$ denote the context at the geometric center of C and let $\dot{\mu}_{k,C}^d := \mu_k^d(\dot{x}_C^d)$. Let $\mathcal{L}_{k,C}^{i,d}(n) := \{|\bar{r}_{k,C}^{i,d}(n) - \dot{\mu}_{k,C}^d| \geq BL2^{-l(C)\alpha}\}$, where $B := 4/(L2^{-\alpha}) + 2$, $\mathcal{L}_{k,\boldsymbol{C}}^i(n) := \cup_{d \in \mathcal{D}} \mathcal{L}_{k,C^d}^{i,d}(n)$ and $\mathcal{L}_{\boldsymbol{C}}^i(n) := \cup_{k \in \mathcal{K}_i} \mathcal{L}_{k,\boldsymbol{C}}^i(n)$. When $\mathcal{L}_{\boldsymbol{C}}^i(n)$ happens we say that the marginal expected reward estimate is *inaccurate* for at least one diagnostic action k.

For clinic i that exploits for its nth patient, the diagnostic action selection for patient n is called *suboptimal* if $\mathcal{L}_{\boldsymbol{C}_i(n)}^i(n)$ happens, otherwise, it is called *near optimal*. The next lemma bounds the regret due to *suboptimal* diagnostic action selections by bounding the expected number of times $\mathcal{L}_{\boldsymbol{C}_i(n)}^i(n)$ happens for $n = 1, \ldots, N$.

Lemma 2: **Regret due to suboptimal diagnostic action selections.** Consider clinics that use LEX with parameters $p > 0$, $2\alpha/p \leq z < 1$, $D_1(n) = D_3(n) = n^z \log n$ and $D_2(t) = F_{\max} n^z \log n$. The regret of clinic i due to suboptimal diagnostic action selections in its exploitation phases, i.e., $R_i^s(N)$, is $O(MF_{\max}N^{1-z/2})$.

Proof: See Appendix A-C. ■

Lemma 3: **Regret due to near-optimal clinics choosing suboptimal experts.** Consider clinics that use LEX with parameters $p > 0$, $2\alpha/p \leq z < 1$, $D_1(n) = D_3(n) = n^z \log n$ and $D_2(t) = F_{\max} n^z \log n$. Then, for any set of hypercubes $\boldsymbol{C}$ that has been active and contained $\boldsymbol{x}_i(n')$ for some patients $n' \in \{1, \ldots, N\}$ of clinic i, the regret due to a near optimal clinic choosing a suboptimal expert for these patients when called by clinic i is $O(1)$.

Proof: See Appendix A-D. ■

The next lemma bounds the regret due to clinic i choosing near optimal diagnostic actions for its patients up to the Nth patient.

Lemma 4: **Regret due to near-optimal diagnostic actions.** Consider clinics that use LEX with parameters $p > 0$, $2\alpha/p \leq z < 1$, $D_1(n) = D_3(n) = n^z \log n$ and $D_2(n) = F_{\max} n^z \log n$. Then, the regret of clinic i due to near optimal diagnostic action selections in its exploitation phases for its own patients $1, \ldots, N$ is $O\left(N^{\frac{1+p-\alpha}{1+p}}\right)$.

Proof: See Appendix A-E. ■

Next, we combine the results from Lemmas 1, 2, 3 and 4 to obtain the regret bound for LEX.

Theorem 1: **Convergence rate to the optimal diagnostic action.** Consider clinics that use LEX with parameters $p = \frac{3\alpha + \sqrt{9\alpha^2 + 8\alpha}}{2}$, $z = 2\alpha/p < 1$, $D_1(n) = D_3(n) = n^z \log n$ and $D_2(n) = F_{\max} n^z \log n$. Then, the regret of clinic i for its patients up to the Nth patient is

$$R_i(N) = O\left(F_{\max} M N^{f_1(\alpha)} \log N\right),$$

where $f_1(\alpha) = (2 + \alpha + \sqrt{9\alpha^2 + 8\alpha})/(2 + 3\alpha + \sqrt{9\alpha^2 + 8\alpha})$. Hence, $\lim_{N \to \infty} R_i(N)/N = 0$.

Proof: For clinic i, for each hypercube of each type-d context, the regret due to trainings and explorations is bounded by Lemma 1. It can be shown that for each type-d context there can be at most $4N^{1/(1+p)}$ hypercubes that are activated up to the Nth patient. Using this we get a $O(N^{z+1/(1+p)} \log N)$ upper bound on the regret due to explorations and trainings for a type-d context. Then we sum over all types of contexts $d \in \mathcal{D}$. Since this regret increases with z, we need to choose it as small as possible. From the results in Lemma 2, the smallest possible value of z is $2\alpha/p$. For this z value, the regret due to suboptimal action selections in exploitation phases (given in Lemma 2) is $O(N^{1-\alpha/p})$. The regret due to near optimal action selection in exploitation phases (given in Lemma 4) is $O(N^{1-\alpha/(p+1)})$. Hence the highest orders of the regret come from trainings and explorations and near-optimal action selections. The value of p that makes these two terms equal is $\frac{3\alpha + \sqrt{9\alpha^2 + 8\alpha}}{2}$. ■

From the result of Theorem 1, it is observed that the regret increases linearly with the number of clinics in the system and their number of experts (which $F_{\max}$ is an upper bound on). We note that the regret is the gap between the total expected reward of the optimal distributed policy that can be computed by a genie which knows the marginal accuracies of every expert, and the total expected diagnostic reward of LEX. Since the performance of optimal distributed policy never gets worse as more clinics are added to the system or as more experts are introduced, the benchmark we compare our algorithm against with may improve. Therefore, the total reward of LEX may improve even if the regret increases with M, F_i and $F_{\max}$.

Theorem 1 gives a bound on the long-term performance of LEX. In a clinical setting, for interpreting the diagnosis recommendation provided by LEX, clinicians may want to know the confidence about the proposed diagnosis recommendation for the patient under consideration. LEX can provide the clinicians sharp confidence bounds on the diagnostic accuracy of the expert it selects. These bounds reveal the context-specific expertise level of the human experts or CDSSs.

Corollary 1: **Confidence bounds on the diagnosis recommendation.** Consider clinics that use LEX with parameters $p = \frac{3\alpha + \sqrt{9\alpha^2 + 8\alpha}}{2}$, $z = 2\alpha/p < 1$, $D_1(n) = D_3(n) = n^z \log n$ and $D_2(n) = F_{\max} n^z \log n$. Then, we have the following confidence bounds on the diagnostic recommendation $\hat{y}_i(n)$: (i) If the prediction is made in an exploitation phase, then with probability at least

$$P_{\text{opt}} := \min\left\{0, 1 - \frac{2K_i D}{n^2} - \frac{2MF_{\max} D \beta_2}{n^{z/2}}\right\},$$

the recommendation comes from a near-optimal diagnostic action (the complement of event $\mathcal{L}^i_{\boldsymbol{C}_i(n)}(n)$ happens), where $\beta_2 = \sum_{n=1}^{\infty} 1/n^2$; (ii) With probability at least P_{opt}, we have

$$\pi^d_{a_i(n)}(x^d) \geq \hat{\pi}^{i,d}_{a_i(n),C^d_i(n)}(n) - 2(B+1)L2^{l(C^d_i(n))\alpha},$$

for all $d \in \mathcal{D}$, where $a_i(n)$ is the diagnostic action taken by clinic i for its nth patient.

Proof: The proof is contained within the proofs of Lemma 2 and Lemma 4. ■

Corollary 1 implies that when LEX exploits for a patient n, it can tell the clinician the probability that the chosen expert is one of the best (near-optimal) experts for the context of patient n. Moreover, it can also tell the clinician a bound on the accuracy of the current diagnostic recommendation. This bound given in Corollary 1 says that for the patient population with type-d context x^d, the true accuracy of the best expert corresponding to the diagnostic action $a_i(n)$ will almost be as high as its estimated accuracy minus some uncertainty term related to the length of the hypercube that the accuracy estimates are formed over. Using this information, the clinician can arrive at a decision: it can follow the recommendation of LEX, or it can find and assign another expert to the patient.

Remark 1: From Corollary 1 when $\alpha = 1$, the probability that a suboptimal expert is chosen for the nth patient when LEX exploits for the nth patient is $O(n^{-0.28})$. Although this goes to zero quickly, the recommendation for the initial set of patients may not be very accurate. This is not a problem since prior knowledge can be incorporated to LEX. Assume that LEX is a priori trained with N_0 patients for each clinic. All the recommendations in this training set is done for the purpose of training LEX and does not affect a clinician's final decision on the patient. Then, for the nth patient that arrives to clinic i after the initial training, the probability that LEX chooses a suboptimal expert will be $O((n+N_0)^{-0.28})$, which can be made arbitrarily small by adjusting the initial training population.

V. Discussion and Extensions of LEX

A. Context of the experts

Our current algorithm LEX is learning independently for each expert. When the number of experts is large, learning for groups of experts using the similarity between their contexts can speed up the learning process. Moreover, depending on the context of expert, the set of patients that can be assigned to that expert can be different.

Below we describe the modification to LEX by which expert context can be taken into account: Consider the set of experts $\mathcal{F}$. For each $f \in \mathcal{F}$, let $\boldsymbol{z}(f) \in \mathcal{Z}$ be the context of expert f, which is a vector in some d_e-dimensional Euclidian space $\mathcal{Z}$. Consider a patient with context vector $\boldsymbol{x}$ and two experts f and f'. Then, the diagnosis accuracy of these experts for the patient will depend on how similar these experts are. This can be mathematically modeled as a *similarity property* between the experts which is stated as follows:

Definition 3: **Similarity of the experts.** There exists $L_e > 0$ and $\alpha_e > 0$ such that for all $f, f' \in \mathcal{F}$, $\boldsymbol{x} \in \mathcal{X}$ and $d \in \mathcal{D}$, we have $|\pi^d_f(x^d) - \pi^d_{f'}(x^d)| \leq L_e||\boldsymbol{z}(f) - \boldsymbol{z}(f')||^{\alpha_e}$.

Now consider the following modification to LEX: Instead of forming marginal sample mean accuracy estimates for experts $f \in \mathcal{F}_i$, each clinic i creates a partition of the space of contexts $\mathcal{Z}$ of the experts denoted by $\mathcal{E}_i$. For each $e \in \mathcal{E}_i$, let $\mathcal{F}(e) \subset \mathcal{F}_i$ be the subset of experts whose contexts lie in $e \subset \mathcal{Z}$. For each $e \in \mathcal{E}_i$, we define a new diagnostic action g_e. Let $\mathcal{G}_i = \{g_e\}_{e \in \mathcal{E}_i}$. When clinic i takes diagnostic action g_e, it randomly selects an expert from $\mathcal{F}(e)$ to assign to the patient.

Then, the set of diagnostic actions of clinic i will be $\mathcal{K}'_i = \mathcal{M}_{-i} \cup \mathcal{G}_i$. Now, instead of keeping marginal sample mean accuracy estimates $\hat{\pi}^{i,d}_{f,C}$ for hypercubes C of patients' contexts for expert $f \in \mathcal{F}_i$, LEX will keep marginal sample mean accuracy estimates $\hat{\pi}^{i,d}_{g_e,C}$ for hypercubes C of patients' contexts for set of experts $\mathcal{F}(e)$ for all $g_e \in \mathcal{G}_i$. With this modification, the performance of LEX will depend on the cardinality of $\mathcal{G}_i$ and similarity constants L_e and α_e. The regret due to explorations will increase with $|\mathcal{G}_i|$ since more actions need to be explored, while the regret due to *grouping* the experts decreases since experts' contexts within the same group will be more similar to each other as the size of the sets $e \in \mathcal{E}_i$ decreases.

B. Cooperation among the experts

In our current setting the clinics cooperate with each other by making diagnosis recommendations for each other's patients when requested. Such cooperation is very beneficial when the expertise of the experts vary among the clinics [24], [25]. However, LEX assigns a single expert to each patient (whether an own expert of the clinic or another clinic's expert). In some clinical applications, such as some complex diseases, multiple experts and clinics can work simultaneosly to diagnose a patient, which can significantly improve the diagnosis accuracy.

LEX can be easily adapted to learn the best group of experts and clinics to assign to a patient given its context vector. We describe how LEX should be modified for this below.

Consider a new diagnostic action set $\mathcal{K}^{\text{coop}}_i$ for clinic i, whose elements are the group of experts/clinics that can be assigned to a patient. Now, instead of keeping marginal sample mean accuracy estimates $\hat{\pi}^{i,d}_{k,C}$ for hypercubes C of patients' contexts for a diagnostic action in $k \in \mathcal{K}_i$, LEX will keep sample mean accuracy estimates $\hat{\pi}^{i,d}_{k,C}$ for hypercubes C of patients' contexts for groups of experts/clinics $k \in \mathcal{K}^{\text{coop}}_i$ such that $k \subset \mathcal{K}_i$. Since LEX will learn independently for each group $k \in \mathcal{K}^{\text{coop}}_i$, the regret of LEX will be linear in the cardinality of $\mathcal{K}^{\text{coop}}_i$. This can be large if $\mathcal{K}^{\text{coop}}_i$ is set to contain all possible groups of experts and clinics. However, clinic i can intelligently adapt the set $\mathcal{K}^{\text{coop}}_i$ in a way that certain combinations of experts and clinics are discarded based on how these individual experts and clinics performed in the past. For instance, LEX can set a threshold τ and discard experts/clinics whose sample mean accuracy falls below τ. Moreover, LEX can also use the contextual information of experts when creating the groups in $\mathcal{K}^{\text{coop}}_i$. For instance, similarity of the experts described in the previous subsection can be be used when creating the groups of experts. While forming groups with experts that have contexts that are very similar to each other may not improve the diagnosis accuracy a lot, adding experts with different contexts (background, education, etc.) may significantly improve the diagnostic accuracy because one expert may look to the patient data from a different perspective from the other experts within the group.

C. Congestion cost for experts

In our framework, we have not considered the congestion of the experts. In reality, making a diagnosis for a new patient requires a certain amount of time devoted by the expert. Hence, an expert gets congested when too many new patients are recommended, which results in delay in making a diagnosis. We model this as the congestion cost for $f \in \mathcal{F}_i$ as $c_f(n)$, which is given as $c_f(n) := \min\{(c_f(n-1) + u_{\boldsymbol{x}_i(n),f}\mathrm{I}(a_i(n) = f) + \sum_{j\in\mathcal{M}_{-i}} u_{\boldsymbol{x}_j(n),f}\mathrm{I}(b_{i,j}(n) = f) - p_f\delta(n-1,n))^+, 1\}$, where $p_f \in [0,1]$ is the amount of work that expert f can do in one unit of time, $\delta(n-1,n)$ is the time between the arrival of patient $n-1$ and n, and $u_{\boldsymbol{x},f} \in [0,1]$ is the amount of work to diagnose the patient with context $\boldsymbol{x}$. We assume that $c_f(n)$ is known (as $u_{\boldsymbol{x},f}$ and p_f can be calculated from the historical patient data). LEX can take this congestion cost into account by setting the marginal sample mean reward of expert f for clinic i's nth patient to $\bar{r}^{i,d}_{f,C^d_i(n)}(n) = \hat{\pi}^{i,d}_{f,C^d_i(n)}(n) - c_f(n) - c^i_f$. We present numerical results considering congestion cost in Section VI.

D. Generality of LEX

Our proposed algorithm is general, can be applied to discover the expertise in many applications such as personalized education, recommender systems and task assignments. Note that although we made a case study in breast cancer dataset, there is nothing specific about breast cancer or medical applications in our proposed algorithm. For example, in a personalized education system, the context can be the GPA, quiz and/or homework scores of the student, while experts can be instructors, teaching assistants and/or online courses.

VI. Illustrative Results

In this section, we illustrate the functioning and performance of our algorithm by comparing it against several state-of-art online ensemble learning algorithms and other multi-armed bandit algorithms. While many of these algorithms are centralized and they cannot easily or at all be deployed in the envisioned distributed clinic setting, we compare against these various methods to highlight the merits of our proposed scheme, including the importance of using contextual (semantic) information in making decisions.

A. Dataset

Breast Cancer Dataset We consider a breast cancer data set provided by UCLA radiology department. The data set includes 45450 patients. A radiologist interprets the breast image of the patients and assigns a BI-RADS score. Score '1' is negative, '2' and '3' are associated with benign, '4' is suspicious, '5' is highly probable malignancy, and '6' is known malignant. The score '4' is further divided into three subcategories with 4A indicating low suspicion of malignancy, 4B indicating intermediate suspicion and 4C indicating moderate concern. We focus only on the BI-RADS 4, 4A, 4B and 4C patients who need to be further monitored and/or screened to decide their cancer status (i.e. benign or malignant tumor). These patients are assigned to an expert, which decides whether or not to undergo a biopsy based on the patients' context, which includes their age, imaging modality, and breast density. Note that some instances of the context vector are missing for some patients. For instance, the breast density information is available for only 45% of the patients.

B. Algorithms that we compare against

-LinUCB [26] is a contextual bandit algorithm which assumes that the expected reward of a diagnostic action is a linear combination of the components of the context vector. However, the coefficients in the linear combination is different for each diagnostic action and is unknown.

-Hybrid-ϵ (Hybrid) [11] combines the context (side) information with an ϵ-greedy algorithm, by extracting the history of context arrivals within a small region of the context space and running an ϵ-greedy algorithm within that space.

-Weighted Majority (WM) [27] is an offline algorithm that assigns and updates weights for the experts based on training data, and produces a final diagnosis recommendation by weighting the diagnosis recommendations of all the experts.

-Sliding Window Adaboost (AdaSliding) [28] is an online version of Adaboost [29], which aims to find the optimal weighting among the experts by an exponential weight update mechanism, where the weights are updated using a window of recent past observations and decisions.

All the algorithms above are centralized, i.e., they require a central clinic which has direct access to all the experts of all clinics. From the above algorithms, we modified LinUCB and Hybrid-ϵ such that they can run on the distributed setting we consider. For Weighted Majority and Sliding Window Adaboost we assume that there is a central clinic which has direct access to all the experts of all clinics.

C. Performance of LEX in breast cancer dataset

Unless stated otherwise, we assume that there are $M = 4$ clinics and $F_i = 2$ experts for each clinic $i \in \mathcal{M}$. A clinic can select one of its own experts without incurring any cost, while the cost of selecting another clinic is set to 0.01.

While LinUCB and Hybrid-ϵ are bandit algorithms (i.e., they require a diagnosis only from the expert they select), AdaSliding and AdaWeighted require the diagnostic recommendations of all the experts in the system for each patient. Hence, these algorithms are run on a centralized system in which the clinic has access to all experts.

Results on the diagnostic accuracy: The comparison between LEX and the above algorithms is given in Table IV. As the performance metric, we use the diagnostic accuracy (i.e., the percentage of patients that are correctly diagnosed). LEX outperforms other learning algorithms by achieving a diagnostic accuracy that is at least 13% higher than the best of the other methods. Moreover, as the number of patients N increases, the diagnostic accuracy increases because LEX learns the expertise of the experts with a higher accuracy as more patients arrive. The poor performance of LinUCB and Hybrid-ϵ algorithms is due to the fact that they don't have the training phase that LEX have, and they learn considering all the contexts in the context vector, rather than learning for the most relevant context as LEX does.

Results on cooperation among the clinics: In order to assess the effect of cooperation between clinics, we simulate the performance of LEX for different numbers of clinics that clinic i can forward its patient's context for diagnosis recommendation. As shown in Table V diagnosis accuracy of LEX increases with the number of clinics that clinic i is connected to. This is due to the diversity of the expertise among different clinics. While a clinic can be good at making diagnosis recommendation to patients with a specific type of context, another clinic may be better specialized for other types of contexts.

TABLE V
DIAGNOSTIC ACCURACY OF LEX AS A FUNCTION OF THE NUMBER OF CLINICS

Number of clinics	M=1	M=2	M=4
LEX	75.96%	81.09%	83.32%

Results on costs associated with cooperation: How and when the clinics cooperate with each other depends on several factors including the expertise of the clinics, contexts of the patients and costs of cooperation (delay, money, etc.). Here, we evaluate the percentage of times a clinic cooperates with the other 4 clinics in exploitations using LEX, as a function of the cost of choosing another clinic for diagnosis recommendation. We assume this cost is the same for all clinics and denote it by c. As seen in Table VI, the percentage of cooperation decreases as the cost increases, and reaches zero when the cost exceeds some threshold. When the cost is too high, asking for expertise of another clinic is not advantageous, even when it improves the diagnostic accuracy.

TABLE VI
COOPERATION % VS. COOPERATION COST. COOPERATION % IS THE PERCENTAGE OF TIMES ANOTHER CLINIC IS CALLED IN EXPLOITATIONS.

Cost c	0.01	0.05	0.1	0.2	0.5
Cooperation	12.67%	7.68%	5.53%	0%	0%

Results on delayed health state observations: For most cases, it may not be possible observe the true health state of the patient just after the diagnostic decision is made. When the decision is malignant, usually a biopsy is performed in a short amount of time, and this reveals the true health state of the patient. However, when the decision is benign, the patient usually waits until the next screening, without any immediate action.

We simulate this by introducing a delay d_n, in terms of the number of patients that have arrived after the nth patient before the true health state of the nth patient is revealed. We assume that $d_n \sim \text{geometric}(\lambda)$, for some parameter $\lambda > 0$. In Table VII, the performance of LEX as a function of the delay is shown in terms of parameter λ. Smaller values of λ imply larger delay, but as seen from the table, the performance of LEX is only slightly affected by the delay.

TABLE VII
TRADEOFF BETWEEN DIAGNOSTIC ACCURACY AND DELAY

λ	0.02	0.01	0.004	0.002
LEX	81.48%	80.43%	80.22%	79.18%

Results on workload balance among the experts: Tables VIII and IX provide a comparison of the diagnostic accuracy and the workloads of the experts for a single clinic $M = 1$ with 6 experts under two different models (i) a model that assumes congestion cost for each expert which depends on the number of patients that expert is assigned to, and (ii) a model that does not consider congestion. Table VIII shows the fraction of times the experts 1, 2, 3 (the most congested experts) and the rest of the experts (expert 4,5,6) are recommended to the patients. We simulate this by taking the congestion cost $c_f(n)$, $f \in \mathcal{F}$ defined in Section V-C, and setting its parameters to $u_{\boldsymbol{x},f} = 0.4$, $p_f = 1$, $\lambda = 1$ for all $\boldsymbol{x} \in \mathcal{X}$ and $f \in \mathcal{F}$ and arrival times $\delta(n-1, n) \sim \exp(\beta)$. In this setting smaller values of β imply more frequent patient arrivals.

TABLE VIII
ALLOCATION OF THE EXPERTS UNDER CONGESTION COST

Allocation	No congestion	$\beta = 0.2$	$\beta = 0.1$
Expert 1	67.39%	28.95%	21.87%
Expert 2	12.80%	27.19%	20.90%
Expert 3	7.25%	20.36%	18.15%
Other Experts	12.56%	23.5%	39.08%

As seen from Table VIII when the congestion cost is introduced, based on the level of congestion the workload becomes more uniform among the experts. The effect of this on the diagnostic accuracy is given in Table IX . From this table, it is observed that the diagnostic accuracy is only slightly affected due to the congestion, since LEX trade-offs between the accuracy and the congestion cost by shifting the workloads to less accurate but less congested experts when the rate of patient arrivals increases.

VII. CONCLUSION

In this paper we proposed a context-adaptive medical diagnosis system that selects from a pool of human experts and CDSSs to make diagnosis recommendations. The system learns online, which context of the patient to use, and which expert to rely on when making diagnosis recommendations. We prove that the diagnostic accuracy of the proposed system converges to the accuracy of the best context-adaptive expert, which means that the best diagnosis mechanism (whether a human expert or a CDSS) for each context is perfectly learned. Moreover, the proposed algorithm LEX learns the best expert for treating a patient with a specific context not only within a clinic but also across all clinics; hence, its performance is better than the performance of the best expert within any given clinic. In a clinical deployment of

TABLE IX
RESULTING PERFORMANCE OF LEX UNDER CONGESTION COST

Congestion	No congestion	$\beta = 0.2$	$\beta = 0.1$
Diagnostic Accuracy	81.17%	79.90%	77.93%

TABLE IV
COMPARISION OF LEX WITH STATE-OF-THE-ART LEARNING ALGORITHMS IN TERMS OF DIAGNOSTIC ACCURACY

Patients	Type	Type	Type	$N = 1000$	$N = 3000$	$N = 5439$
LEX	Distributed	Contextual	Online	80.03%	82.49%	83.32%
LinUCB	Distributed	Contextual	Online	63.03%	65.93%	66.43%
Hybrid	Distributed	Contextual	Online	63.15%	65.53%	67.91%
AdaSliding	Centralized	-	Online	66.10%	71.20%	73.16%
WM	Centralized	-	Offline	60.50%	59.93%	59.48%

the proposed system, diagnosis recommendations made by the system will be examined by a clinician before the final prediction is made. This will provide an additional layer of safety. For any patient, it is the clinician's discretion whether to rely on or to disregard the recommendation of the LEX algorithm. Future work includes designing algorithms that can track the changes in a clinician's performance by exploiting a recent time window of patient histories, and adapting LEX to discover expertise in other settings.

APPENDIX A
PROOF OF THE LEMMAS

A. Preliminaries

Let $\beta_2 = \sum_{n=1}^{\infty} 1/n^2$. For a set A let A^C denote the complement of that set. For a set of hypercubes $\boldsymbol{C}$, let $l_{\min}(\boldsymbol{C})$ be the level of the hypercube in $\boldsymbol{C}$ with the minimum level. We start with a simple lemma which gives an upper bound on the highest level hypercube that is active for any patient n.

Lemma 5: **A bound on the level of active hypercubes.** All the active hypercubes $\mathcal{A}_i^d(n)$ for type-d contexts for patient n have at most a level of $(\log_2 n)/p + 1$.

Proof: Let $l+1$ be the level of the highest level active hypercube. We must have $A\sum_{j=0}^{l} 2^{pj} < n$, otherwise the highest level active hypercube will be less than $l+1$. We have for $n/A > 1$, $A\frac{2^{p(l+1)}-1}{2^p-1} < n \Rightarrow 2^{pl} < \frac{n}{A} \Rightarrow l < \frac{\log_2 n}{p}$. ■

B. Proof of Lemma 1

This directly follows from the number of trainings and explorations that are required before any diagnostic action can be exploited which is determined by $D_1(n)$, $D_2(n)$ and $D_3(n)$.

C. Proof of Lemma 2

Let Ω denote the space of all possible outcomes, and w be a sample path. The event that the LEX exploits when $\boldsymbol{x}_i(n) \in \boldsymbol{C}$ is given by $\mathcal{W}_{\boldsymbol{C}}^i(n) := \{w : \mathcal{F}_{\text{ue},\boldsymbol{C}}^i(n) \cup \mathcal{M}_{\text{ue},\boldsymbol{C}}^i(n) \cup \mathcal{M}_{\text{ut},\boldsymbol{C}}^i(n) = \emptyset, \boldsymbol{x}_i(n) \in \boldsymbol{C}, \boldsymbol{C} \in \boldsymbol{\mathcal{A}}_i(n)\}$. We will bound the probability that LEX chooses a suboptimal action for clinic i in an exploitation phase when i's context vector is in the set of active hypercubes $\boldsymbol{C}$ for any $\boldsymbol{C}$, and then use this to bound the expected number of times a suboptimal action is chosen by clinic i for its patients in exploitation steps using LEX. Recall that reward loss in every step in which a suboptimal action is chosen can be at most 2, hence the regret due to suboptimal action selections is directly proportional to the number of times event $\mathcal{L}_{\boldsymbol{C}}^i(n) \cap \mathcal{W}_{\boldsymbol{C}}^i(n)$ occurs.

Let $\mathcal{B}_{j,C}^{i,d}(n)$ be the event that clinic j made a suboptimal expert assignment for clinic i for at most n^ϕ of the patients of clinic i whose type-d contexts are in C for which clinic i explored clinic j up to its nth patient according to the type-d context. For $f \in \mathcal{F}_i$ let $\mathcal{B}_{f,\boldsymbol{C}}^{i,d}(n) := \Omega$ for all $d \in \mathcal{D}$. We have

$$\begin{aligned} \mathrm{P}\left(\mathcal{L}_{\boldsymbol{C}}^i(n), \mathcal{W}_{\boldsymbol{C}}^i(n)\right) &\le \sum_{k\in\mathcal{K}_i} \mathrm{P}\left(\mathcal{L}_{k,\boldsymbol{C}}^i(n), \mathcal{W}_{\boldsymbol{C}}^i(n)\right) \\ &\le \sum_{k\in\mathcal{K}_i}\sum_{d\in\mathcal{D}} \Big(\mathrm{P}\left(\mathcal{L}_{k,C^d}^{i,d}(n), \mathcal{B}_{k,C}^{i,d}(n), \mathcal{W}_{\boldsymbol{C}}^i(n)\right) \\ &+ \mathrm{P}\left(\mathcal{B}_{k,C}^{i,d}(n)^C\right)\Big). \end{aligned} \tag{A.1}$$

We generate an artificial i.i.d. processes for each type-d hypercube C^d to bound the probabilities related to deviation of sample mean reward estimates $\bar{r}_{k,C^d}^{i,d}(n) := \hat{\pi}_{k,C^d}^{i,d}(n) - c_k^i$, $k \in \mathcal{K}_i$, $d \in \mathcal{D}$ from the expected rewards, which will be used to bound the probability of choosing a suboptimal action. We call this the *center* process, in which rewards are generated according to a bounded i.i.d. process with expected reward $\dot{\mu}_{k,C^d}^d$. Let $\dot{r}_{k,C^d}^{i,d}(n)$ denote the sample mean of the n samples from the center process for type-d hypercube C^d and action k. Let $\tilde{T}_k^d := T_{k,C^d}^{i,d}(n)$. We have

$$\begin{aligned} &\mathrm{P}\left(\mathcal{L}_{k,C^d}^{i,d}(n), \mathcal{W}_{\boldsymbol{C}}^i(n), \mathcal{B}_{k,C}^{i,d}(n)\right) \\ &= \mathrm{P}\left(|\bar{r}_{k,C^d}^{i,d}(n) - \dot{\mu}_{k,C^d}^d| \ge BL2^{-l(C^d)\alpha}, \mathcal{B}_{k,C}^{i,d}(n), \mathcal{W}_{\boldsymbol{C}}^i(n)\right) \\ &\le \mathrm{P}\left(|\dot{r}_{k,C^d}^{i,d}(\tilde{T}_k^d) - \dot{\mu}_{k,C^d}^d| \ge (B-2)L2^{-l(C^d)\alpha} - 2\frac{n^\phi}{\tilde{T}_k^d},\right. \\ &\left.\mathcal{W}_{\boldsymbol{C}}^i(n)\right) \le 2/n^2, \end{aligned} \tag{A.2}$$

by using a Chernoff-Hoeffding bound. This can be verified by checking that the last inequality holds when the condition

$$(B-2)L2^{-l(C^d)\alpha} - 2\frac{n^\phi}{\tilde{T}_k^d} \ge 2n^{-z/2}$$

holds, which holds when

$$(B-2)L2^{-l(C^d)\alpha} - 2n^{\phi - z} \ge 2n^{-z/2} \tag{A.3}$$

holds. By Lemma 5, (A.3) holds when

$$(B-2)L2^{-\alpha} n^{-\alpha/p} \ge 4n^{-z/2}. \tag{A.4}$$

Since the parameters in the statement of the lemma are $\phi = z/2$ and $z \ge 2\alpha/p$, and since $B = 4/(L2^{-\alpha}) + 2$, (A.4) holds.

For $j \in \mathcal{M}_{-i}$, let $X_{j,\boldsymbol{C}}^{i,d}(n)$ be the number of times a suboptimal expert (suboptimal diagnostic action) of clinic j is selected when clinic i calls clinic j in exploration phases of clinic i with d as the main type of context, and when the context vector is in the set of hypercubes $\boldsymbol{C}$ of the nth patient. Since $\mathrm{P}(\mathcal{B}_{k,\boldsymbol{C}}^{j,d}(n)^C) = 0$ for $k \in \mathcal{F}_j$, $d \in \mathcal{D}$, using the result in (A.2) for clinic j instead of i, we have

$$\begin{aligned} E[X_{j,\boldsymbol{C}}^{i,d}(n)] &\le \sum_{n=1}^{N}\sum_{k\in\mathcal{F}_j} \mathrm{P}\left(\mathcal{L}_{k,\boldsymbol{C}}^j(n), \mathcal{W}_{\boldsymbol{C}}^j(n)\right) \\ &\le 2F_{\max} D\beta_2, \end{aligned}$$

where $\beta_2 = \sum_{n=1}^{\infty} 1/n^2$. Hence, $\mathrm{P}(\mathcal{B}^{i,d}_{j,\boldsymbol{C}}(n)^c) \leq \mathrm{E}[X^{i,d}_{j,\boldsymbol{C}}(n)]/n^{\phi} \leq 2F_{\max} D\beta_2 n^{-z/2}$, for $j \in \mathcal{M}_{-i}$.

Combining all of these we get

$$\mathrm{P}\left(\mathcal{L}^i_{\boldsymbol{C}}(n), \mathcal{W}^i_{\boldsymbol{C}}(n)\right) \leq \frac{2K_i D}{n^2} + \frac{2MF_{\max} D\beta_2}{n^{z/2}}.$$

Hence

$$R^s_i(N) \leq \sum_{n=1}^{N} \left(\frac{2K_i D}{n^2} + \frac{2MF_{\max} D\beta_2}{n^{z/2}}\right) = O(MF_{\max} N^{1-z/2}).$$

D. Proof of Lemma 3

Let $X^i_{j,\boldsymbol{C}}(N)$ denote the random variable which is the number of times a suboptimal expert of clinic $j \in \mathcal{M}_{-i}$ is chosen in exploitation phases of clinic i when $\boldsymbol{x}_i(n')$ is in set $\boldsymbol{C} \in \boldsymbol{\mathcal{A}}_i(n')$ for $n' \in \{1, \ldots, N\}$. Similar to the proof of Lemma 2, it can be shown that $E[X^i_{j,\boldsymbol{C}}(N)] \leq 2F_{\max} D^2 \beta_2$. Thus, the contribution to the regret from suboptimal actions of clinic j is bounded by $4F_{\max} D^2 \beta_2$. We get the final result by considering the regret from all $M-1$ other clinics.

E. Proof of Lemma 4

The following lemma bounds the per-patient (one-step) regret to clinic i from choosing near optimal actions. This lemma is used later to bound the total regret from near optimal actions.

Lemma 6: **One-step regret due to near-optimal actions for a set of hypercubes.** Consider clinics using LEX with parameters $p > 0$, $2\alpha/p \leq z < 1$, $D_1(n) = D_3(n) = n^z \log n$ and $D_2(n) = F_{\max} n^z \log n$. Then, for any set of hypercubes $\boldsymbol{C}_i(n) = \boldsymbol{C}$, the one-step regret of clinic i in an exploitation phase n from choosing one of its near optimal diagnostic actions (i.e., when $\mathcal{L}^i_{\boldsymbol{C}}(n)^C$ happens) is bounded above by $4Z(B+1)L2^{-l_{\min}(\boldsymbol{C})\alpha}$.

Proof: Consider the event $\mathcal{L}^i_{\boldsymbol{C}}(n)^C$ in which clinic i exploits and chooses diagnostic action k. Then, the one step regret for any context vector $\boldsymbol{x} \in \boldsymbol{C}$ is equal to

$$\begin{aligned}\Delta^i(n) &:= (\pi_{k^*_i(\boldsymbol{x})}(\boldsymbol{x}) - c^i_{k^*_i(\boldsymbol{x})}) - (\pi_k(\boldsymbol{x}) - c^i_k) \\ &\leq Z\left(\max_{d\in\mathcal{D}} \mu^d_{k^*_i(\boldsymbol{x})}(x^d) - \max_{d\in\mathcal{D}} \mu^d_k(x^d)\right), \end{aligned} \quad \text{(A.5)}$$

by the *Similarity Property*. Again by the *Similarity Property*, we have

$$\mu^d_{k^*_i(\boldsymbol{x})}(x^d) \leq \dot{\mu}^d_{k^*_i(\boldsymbol{x}),C^d} + L2^{-l(C^d)\alpha},$$
$$\mu^d_k(x^d) \geq \dot{\mu}^d_{k,C^d} - L2^{-l(C^d)\alpha}.$$

Since $|\bar{r}^{i,d}_{k,C^d}(n) - \dot{\mu}^d_{k,C^d}| \leq BL2^{-l(C^d)\alpha}$ for all $k \in \mathcal{K}_i$ and $d \in \mathcal{D}$ on event $\mathcal{L}^i_{\boldsymbol{C}}(n)^C$, we have

$$\mu^d_{k^*_i(\boldsymbol{x})}(x^d) \leq \bar{r}^{i,d}_{k^*_i(\boldsymbol{x}),C^d}(n) + 2(B+1)L2^{-l(C^d)\alpha},$$
$$\mu^d_k(x^d) \geq \bar{r}^{i,d}_{k,C^d}(n) - 2(B+1)L2^{-l(C^d)\alpha},$$

where the factor 2 on the right hand side of the equations accounts for a near optimal clinic $k \in \mathcal{M}_{-i}$ choosing one of its near optimal actions in $\mathcal{F}_k$.

The equations above imply that

$$\max_{d\in\mathcal{D}} \mu^d_{k^*_i(\boldsymbol{x})}(x^d) \leq \max_{d\in\mathcal{D}} \bar{r}^{i,d}_{k^*_i(\boldsymbol{x}),C^d}(n) + 2(B+1)L2^{-l_{\min}(\boldsymbol{C})\alpha},$$
$$\max_{d\in\mathcal{D}} \mu^d_k(x^d) \geq \max_{d\in\mathcal{D}} \bar{r}^{i,d}_{k,C^d}(n) - 2(B+1)L2^{-l_{\min}(\boldsymbol{C})\alpha}.$$

Since k is chosen by clinic i in its exploitation phase, we must have $\max_{d\in\mathcal{D}} \bar{r}^{i,d}_{k,C^d}(n) \geq \max_{d\in\mathcal{D}} \bar{r}^{i,d}_{k^*_i(\boldsymbol{x}),C^d}(n)$. Combining the above equations with (A.5), we get

$$\Delta^i(n) \leq 4Z(B+1)L2^{-l_{\min}(\boldsymbol{C})\alpha}.$$

■

Let $\tau_i(N)$ be the set of patients up to the Nth patient of clinic i for which clinic i exploits. Using the results of the above lemma, the regret of clinic i due to near optimal actions up to its Nth patient is bounded by

$$\begin{aligned}\sum_{n\in\tau_i(N)} \Delta^i(n) &\leq 4Z(B+1)L \sum_{n\in\tau_i(N)} 2^{-l_{\min}(\boldsymbol{C}_i(n))\alpha} \\ &\leq 4Z(B+1)L \sum_{n\in\tau_i(N)} \sum_{d\in\mathcal{D}} 2^{-l(C^d_i(n))\alpha} \\ &\leq 4Z(B+1)LD \max_{d\in\mathcal{D}} \sum_{n\in\tau_i(N)} 2^{-l(C^d_i(n))\alpha}.\end{aligned} \quad \text{(A.6)}$$

We know that the length of the hypercubes used by LEX decreases over time due to its accounting for the tradeoff between patient arrivals and reward variations within a hypercube. In order to bound (A.6), we assume a worst case scenario, where context vectors arrive such that at each n, the active hypercube that contains the context of each type has the maximum possible length. This happens when type d contexts arrive in a way that all level l hypercubes are split to level $l+1$ hypercubes, before any arrivals to these level $l+1$ hypercubes happen, for all $l = 0, 1, 2, \ldots$. This way it is guaranteed that the length of the hypercube that contains the context for each $n \in \tau_i(N)$ is maximized. Let $l_{\max}$ be the level of the maximum level hypercube in $\boldsymbol{\mathcal{A}}_i(N)$. For the worst case context arrivals we must have

$$\sum_{l=0}^{l_{\max}-1} 2^l 2^{pl} < N \Rightarrow l_{\max} < 1 + \log_2 N/(1+p),$$

since otherwise maximum level hypercube will have level larger than $l_{\max}$. Hence, continuing from (A.6), we have

$$\begin{aligned}\sum_{n\in\tau_i(N)} \Delta^i(n) &\leq 4Z(B+1)LD \max_{d\in\mathcal{D}} \sum_{n\in\tau_i(N)} 2^{-l(C^d_i(n))\alpha} \\ &\leq 4Z(B+1)LD \sum_{l=0}^{1+\log_2 N/(1+p)} 2^l 2^{pl} 2^{-l\alpha} \\ &\leq 4Z(B+1)LD2^{2(p+1-\alpha)} N^{\frac{p+1-\alpha}{p+1}}.\end{aligned}$$

References

[1] P. Sheu, H. Yu, C. Ramamoorthy, A. K. Joshi, and L. A. Zadeh, *Semantic Computing*. John Wiley & Sons, 2011.

[2] C. C. Wang, D. A. Hecht, P. C.-Y. Sheu, and J. J. Tsai, "Semantic computing and drug discovery - a preliminary report," *Sixth IEEE International Conference on Semantic Computing*, pp. 453–458, 2013.

[3] C. C. Bennett and K. Hauser, "Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach," *Artificial Intelligence in Medicine*, vol. 57, no. 1, pp. 9–19, 2013.
[4] M. B. Rosenthal and R. A. Dudley, "Pay-for-performance: will the latest payment trend improve care?" *Jama*, vol. 297, no. 7, pp. 740–744, 2007.
[5] W. Hsu, R. K. Taira, S. El-Saden, H. Kangarloo, and A. A. Bui, "Context-based electronic health record: toward patient specific healthcare," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 2, pp. 228–234, 2012.
[6] M. Barla, M. Tvarožek, and M. Bieliková, "Rule-based user characteristics acquisition from logs with semantics for personalized web-based systems," *Computing and Informatics*, vol. 28, no. 4, pp. 399–428, 2012.
[7] O. Cordón, "A historical review of evolutionary learning methods for mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems," *International Journal of Approximate Reasoning*, vol. 52, no. 6, pp. 894–913, 2011.
[8] S. Oh, M. S. Lee, and B.-T. Zhang, "Ensemble learning with active example selection for imbalanced biomedical data classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 8, no. 2, pp. 316–325, 2011.
[9] E. Alpaydin, *Introduction to machine learning*. The MIT Press, 2004.
[10] P. Zhang, X. Zhu, Y. Shi, L. Guo, and X. Wu, "Robust ensemble learning for mining noisy data streams," *Decision Support Systems*, vol. 50, no. 2, pp. 469–479, 2011.
[11] D. Bouneffouf, A. Bouzeghoub, and A. L. Gançarski, "Hybrid-ε-greedy for mobile context-aware recommender system," in *Advances in Knowledge Discovery and Data Mining*. Springer, 2012, pp. 468–479.
[12] T. Lu, D. Pál, and M. Pál, "Contextual multi-armed bandits," in *Proc. of International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010, pp. 485–492.
[13] W. Chu, L. Li, L. Reyzin, and R. E. Schapire, "Contextual bandits with linear payoff functions," in *Proc. of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.
[14] A. Slivkins, "Contextual bandits with similarity information," in *24th Annual Conference on Learning Theory (COLT)*, 2011.
[15] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.
[16] P. Bühlmann and B. Yu, "Boosting with the l 2 loss: regression and classification," *Journal of the American Statistical Association*, vol. 98, no. 462, pp. 324–339, 2003.
[17] A. Lazarevic and Z. Obradovic, "The distributed boosting algorithm," in *Proc. of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2001, pp. 311–316.
[18] B. Chen, R. Jiang, T. Kasetkasem, and P. K. Varshney, "Channel aware decision fusion in wireless sensor networks," *Signal Processing, IEEE Transactions on*, vol. 52, no. 12, pp. 3454–3458, 2004.
[19] G. Mateos, J. A. Bazerque, and G. B. Giannakis, "Distributed sparse linear regression," *Signal Processing, IEEE Transactions on*, vol. 58, no. 10, pp. 5262–5276, 2010.
[20] H. Kargupta, B. Park, D. Hershberger, and E. Johnson, "Collective data mining: A new perspective toward distributed data mining," *Advances in Distributed and Parallel Knowledge Discovery*, no. part II, pp. 131–174, 1999.
[21] H. Zheng, S. R. Kulkarni, and H. Poor, "Attribute-distributed learning: models, limits, and algorithms," *Signal Processing, IEEE Transactions on*, vol. 59, no. 1, pp. 386–398, 2011.
[22] T. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol. 6, pp. 4–22, 1985.
[23] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, 2002.
[24] Y. Jiang, Y. Zhou, and W. Wang, "Task allocation for undependable multiagent systems in social networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 8, pp. 1671–1681, 2013.
[25] Y. Jiang and J. Jiang, "Contextual resource negotiation-based task allocation and load balancing in complex software systems," *Parallel and Distributed Systems, IEEE Transactions on*, no. 5, pp. 641–653, 2009.
[26] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proc. of the 19th International Conference on World Wide Web*. ACM, 2010, pp. 661–670.
[27] N. Littlestone and M. K. Warmuth, "The weighted majority algorithm," in *Foundations of Computer Science, 30th Annual Symposium on*. IEEE, 1989, pp. 256–261.
[28] H. Wang, W. Fan, P. S. Yu, and J. Han, "Mining concept-drifting data streams using ensemble classifiers," in *Proc. of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2003, pp. 226–235.
[29] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Computational Learning Theory*. Springer, 1995, pp. 23–37.

Cem Tekin Cem Tekin is a Postdoctoral Scholar at University of California, Los Angeles. He received the B.Sc. degree in electrical and electronics engineering from the Middle East Technical University, Ankara, Turkey, in 2008, the M.S.E. degree in electrical engineering: systems, M.S. degree in mathematics, Ph.D. degree in electrical engineering: systems from the University of Michigan, Ann Arbor, in 2010, 2011 and 2013, respectively. His research interests include machine learning, multi-armed bandit problems, data mining, multi-agent systems and game theory. He received the University of Michigan Electrical Engineering Departmental Fellowship in 2008, and the Fred W. Ellersick award for the best paper in MILCOM 2009.

Onur Atan Onur Atan received B.Sc degree in Electrical Engineering from Bilkent University, Ankara, Turkey in 2013. He is currently pursuing the Ph.D. degree in Electrical Engineering at University of California, Los Angeles. His research interests include online learning and multi-armed bandit problems.

Mihaela van der Schaar Mihaela van der Schaar is Chancellor Professor of Electrical Engineering at University of California, Los Angeles. Her research interests include network economics and game theory, online learning, dynamic multi-user networking and communication, multimedia processing and systems, real-time stream mining. She is an IEEE Fellow, a Distinguished Lecturer of the Communications Society for 2011-2012, the Editor in Chief of IEEE Transactions on Multimedia and a member of the Editorial Board of the IEEE Journal on Selected Topics in Signal Processing. She received an NSF CAREER Award (2004), the Best Paper Award from IEEE Transactions on Circuits and Systems for Video Technology (2005), the Okawa Foundation Award (2006), the IBM Faculty Award (2005, 2007, 2008), the Most Cited Paper Award from EURASIP: Image Communications Journal (2006), the Gamenets Conference Best Paper Award (2011) and the 2011 IEEE Circuits and Systems Society Darlington Award Best Paper Award. She received three ISO awards for her contributions to the MPEG video compression and streaming international standardization activities, and holds 33 granted US patents.