The 4th IEEE International Conference on E-Health and Bioengineering - EHB 2013
Grigore T. Popa University of Medicine and Pharmacy, Iaşi, Romania, November 21-23, 2013

Artificial Prediction Markets for Lymph Node Detection

Adrian Barbu
Statistics Department
Florida State University
Tallahassee, FL, USA
e-mail: abarbu@stat.fsu.edu

Nathan Lay
Siemens Corporate Research
Princeton, NJ, USA
e-mail: enslay@gmail.com

***Abstract*- Prediction markets are forums aimed at predicting the outcome of future events of interest such as election results. People participate in a prediction market by buying contracts on the possible outcomes. They are rewarded after the outcome is known based on the number of contracts purchased for the correct outcome. The Artificial Prediction Market is a novel machine learning method that simulates a prediction market where the participants are trained classifiers instead of people. In this work we present an application of the Artificial Prediction Market to lymph node detection from CT images. An evaluation on 54 CT volumes shows that the detector trained with the Artificial Prediction Market has a detection rate of 81.2% at 3 false positives per volume, while an Adaboost classifier trained on the same features obtains a detection rate of 79.6% at the same false positive rate.**

Keywords: prediction markets, lymph node detection, medical imaging.

I. INTRODUCTION

The progress of many types of cancer including breast cancer as well as the success of radiation therapy and chemo-therapy can be monitored by measuring the size of lymph nodes near the cancer location. For this purpose, CT scans are taken are regular time intervals and the lymph nodes near the cancer location are found and measured by the radiologist. Finding the nearby lymph nodes is a tedious task because the lymph nodes are small, have irregular shape and low contrast relative to the surrounding tissue. Moreover, unlike the organs such as heart, liver, kidneys etc., the lymph nodes are not localized in any particular place, making them more difficult to detect. Because of these difficulties, an automatic method for lymph node detection would make this task more reliable and less time consuming.

There have been a number of methods for automatic lymph node detection, based on the Min-DD filter [1], [2], and on stable spring-mass models [3]. Recently machine learning approaches to lymph node detection [4] [5] have greatly reduced the number of false positives for the same detection rate. The learning based methods use a cascade of trained classifiers based on a number of features, starting with fast Haar features and ending with more computationally expensive and more discriminative segmentation-based features.

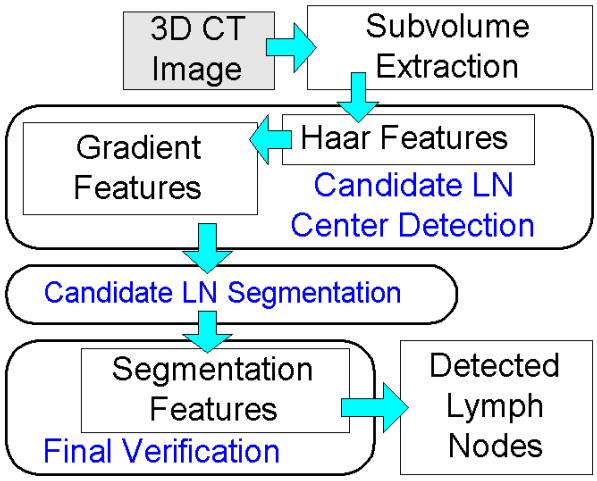

Fig. 1. Diagram of the lymph node detection system.

In this work we improve the lymph node detection performance by employing a recent machine learning method named Artificial Prediction Market [6] to construct a better model based on the same features as [4]. The procedure for lymph node detection and the features will be described in detail in Section II.

The Artificial Prediction Market is a simplified model of a real prediction market called the Iowa Electronic Market [7], which is a forum where contracts on the outcome of future events are bought and sold based on supply and demand. It has been observed that the contract prices of the possible outcomes of an event are a good approximation of the probability distribution of the outcomes. The Artificial Prediction Market replaces the human participants with small algorithms called classifiers and uses training examples to update their budgets (weights) using a zero-sum game that will be described in detail in Section III.

II. LYMPH NODE DETECTION OVERVIEW

The diagram of the lymph node detection system is illustrated in Figure 1. It contains the following stages:

1. **Subvolume Extraction.** A region of interest is extracted from the input 3D CT image by detecting one or more landmark points and extracting a subvolume of a predefined size (in mm) around the detected landmark.

978-1-4799-2373-1/13/$31.00 ©2013 IEEE

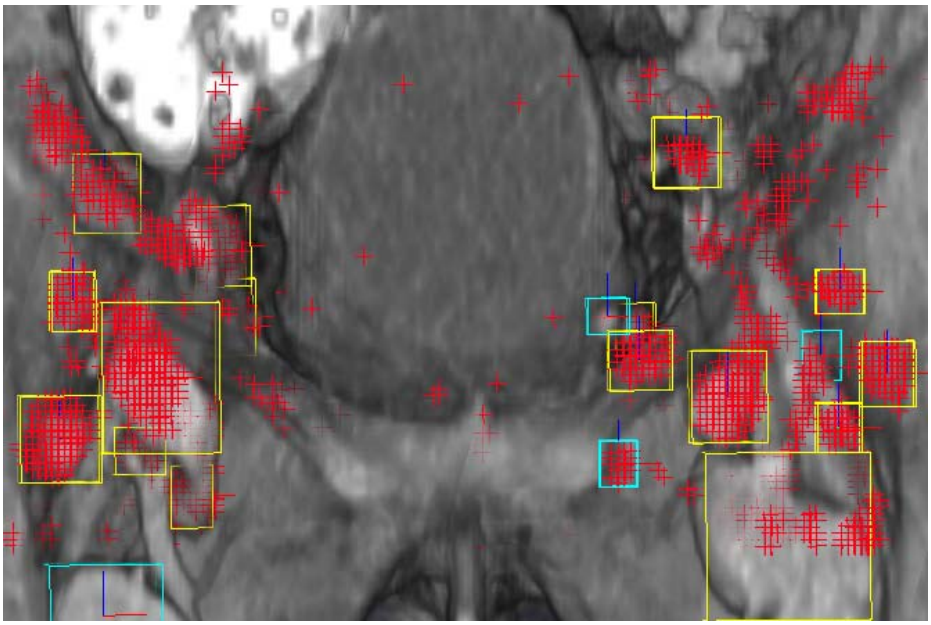

Fig. 3. Example of detected LN candidates in a pelvic region.

2. **Candidate LN Center Detection.** A first detector using features that can be computed quickly is used to prune out the locations in the 3D volume that most likely don't contain lymph nodes. This detector produces about 1000-3000 candidate lymph node centers in the 3D volume.
3. **Candidate LN Segmentation.** A segmentation is obtained at each of the detected candidate locations by gradient descent optimization of an energy function containing a Markov Random Field smoothness term.
4. **Final Verification.** The candidate LN segmentations are used to extract highly discriminative features that can better separate the true LN centers from other locations. A trained classifier uses only a few of these features to make the final decision of which lymph node candidates should be kept and which should be removed.
5. **Non-maximal suppression.** Since there will usually be many overlapping detections of the same lymph node, a non-maximal suppression step keeps only the best one and removes any other detection that overlaps with it.

This system is explained in more detail in [4]. The difference of this work from [4] is in the learning algorithm used to train the final verification stage. In [4] the final verification classifier was trained by Adaboost [8], while in this paper it is an Artificial Prediction Market. The detection steps are described in more detail in the remainder of this section.

A. Subvolume Extraction

The region of interest (e.g. axillary region for the axillary lymph nodes, pelvic region for the pelvic LN, etc) is obtained from the input 3D CT image by detecting one or more landmarks of interest (e.g. lung tips or the pubic symphysis tip) and extracting a subvolume of a predefined size (in mm) around the detected landmark. The landmarks are detected using specialized classifiers and 3D Haar features. The extracted region is resized so that the voxel size is 1.5mm.

B. Candidate Lymph Node Center Detection

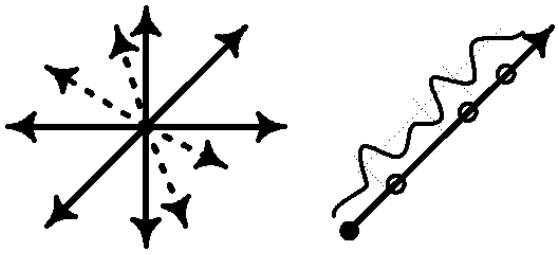

Fig. 2. Left: the gradient features are computed in 14 3D directions. Right: in each direction, the features are computed at the local gradient maxima above a threshold.

The first set of lymph node center candidates is constructed as the set of all voxels in the extracted subvolume that have the intensity in the range [-100,200] HU. These candidates are pruned with a detector trained as a four stage Adaboost [8] cascade with 3D Haar features for the first two stages and a set of gradient features for the last two stages.

The gradient features are computed based on 14 standard directions in 3D as illustrated in Figure 2, left. In each of the 14 directions, the features are computed based on the local maxima of the intensity gradient above one of 10 thresholds τ_j. Then the gradient features are:

- The 24 types of features from [9] computed at one of the first three local maxima in each of the 14 directions and above one of the 10 thresholds τ_j.
- The 24 types of features from [9] computed half way between the candidate center and one of the first three

local maxima obtained in one of the 14 directions which is above one of the 10 thresholds τ_j.

- The distances to the local gradient maxima.
- The difference between distances of the corresponding local gradient maxima for any two directions.

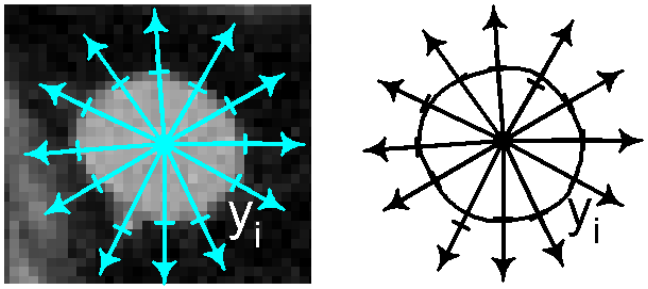

Fig. 4. Left: measurements y_i are found based on intensity difference from the center. Right: the segmentation is found by energy minimization to fit through most of the y_i.

These produce about 64000 gradient features. The gradient features are more computationally expensive than the Haar features, which is why they were used only in the last two stages of the cascade, after many easy negatives were pruned out by the first two levels using the Haar features. An example of detected LN center candidates is shown in Figure 3, with the detected candidates shown as red crosses.

C. Candidate Lymph Node Segmentation

At each of the obtained candidate lymph node centers a segmentation is obtained using energy minimization and a spherical representation.

The spherical representation on the segmentation is obtained by measuring the distance from the lymph node candidate center to the lymph node boundary in 162 directions. These directions are obtained using a triangulation of the sphere with 162 vertices and 320 triangles. This way the segmentation is represented as a 162 dimensional vector of radii $\boldsymbol{r} = (r_1, \dots, r_{162})$.

To obtain the segmentation from a lymph node candidate center, first the most probable boundary distances y_i are found in each of the 162 directions, as illustrated in Figure 4, left. The segmentation $\boldsymbol{r}$ is then obtained by minimizing the following energy by gradient descent.

$$E(r) = \alpha \sum_{i=1}^{162} \rho(r_i - y_i) + \sum_{i=1}^{162} \frac{1}{2|\partial i|} \sum_{j \in \partial i} (r_i - r_j)^2$$

where $\rho(x) = ln\,(1 + \frac{x^2}{2})$. Because the gradient can be computed analytically, the minimization is very fast. Because of the robust function $\rho(x)$, the segmentation can ignore some noisy measurements y_i, as seen in Figure 4, right. Other segmentation methods, e.g. based on fuzzy logic [10], could be used instead of the method presented in this section.

D. Final Verification

The segmentation is used to obtain informative features that can better discriminate the lymph nodes from other structures such as blood vessels.

First, the bounding box of the segmentation is computed and the candidates for which two of the three bounding box dimensions are less than 9mm are automatically rejected.

For the remaining candidates, the following features are computed from the segmentation:

- The 24 feature types from [9] are computed at the 162 vertices of the segmentation triangulation. The 162 values are sorted in decreasing order.
- For each of the 24 feature types, the 81 sums of values at the opposite vertices are computed and sorted in decreasing order.
- The 81 diameters (distances between opposite vertices) are sorted in decreasing order. Then for each diameter are computed the following features:
    - The value of the diameter
    - The ratio of the larger over the smaller radius
    - The ratio of the i-th sorted diameter over the j-th sorted diameter for $1 \le i < j \le 81$.
    - For each of the 24 feature types, the maximum or minimum of the values at the two diameter endpoints
    - For each of the 24 feature types, the maximum or minimum of the values half way to two diameter endpoints

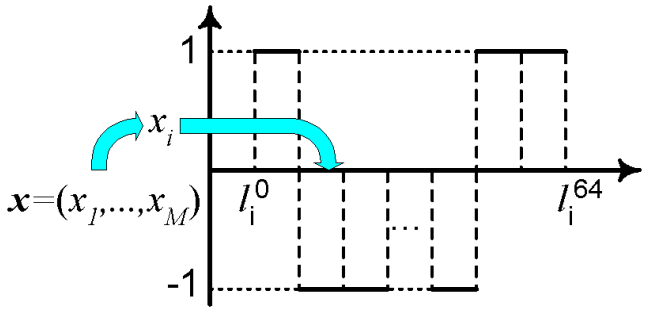

Fig. 5. Diagram of an Adaboost weak classifier for feature *i*.

There are about 17000 features obtained this way. A small number of these features are selected by Adaboost and combined into the verification classifier. The weak classifiers depend on one feature each, say $h_i(\boldsymbol{x}) = h_i(x_i)$. The range of the feature values is divided into 64 intervals with endpoints $(l_i^0, \dots, l_i^{64})$ and the weak classifier returns $h_i(x) = b_{ik}$ if $x_i \in (l_i^{k-1}, l_i^k]$, where $b_{ik} = \pm 1$, as illustrated in Figure 5.

E. Non-Maximal Suppression

The non-maximal suppression algorithm removes overlapping detections, which most likely belong to the same lymph node. It starts with a set D of detections with the score (obtained by the verification classifier) above a threshold and proceeds as follows. It finds the detection from D with the highest score and adds it to the list L of detected lymph

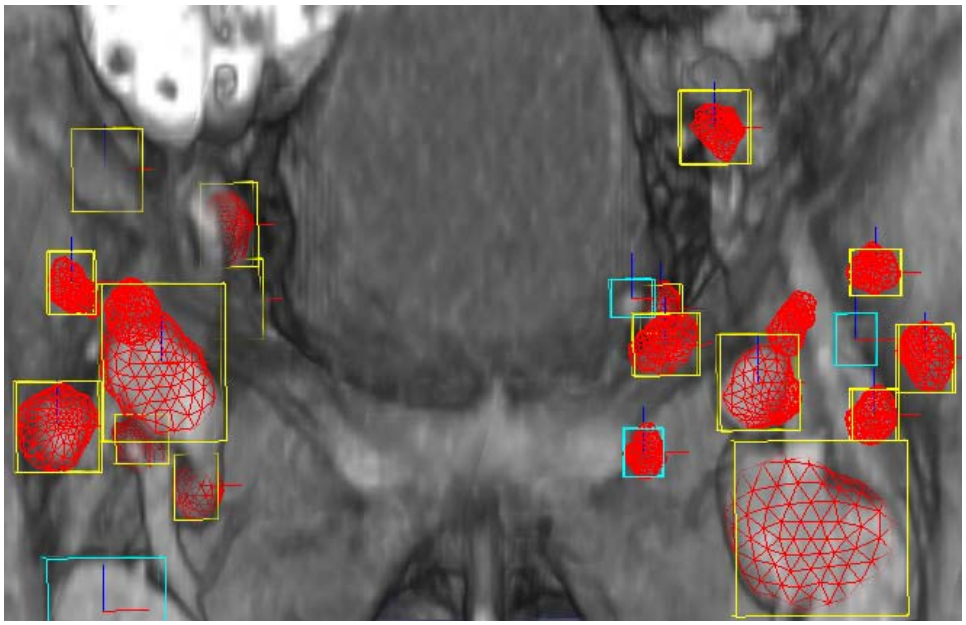

Fig. 6. Detected and segmented lymph nodes.

nodes. It then removes all detections from D that have the segmentation center inside the bounding box of any detection from L. Then it adds to the list L the detection from D with the highest score from the remaining ones and again removes from D all other detections with the center inside its bounding box. The procedure of adding the best remaining detection from D and removing all other detection with the center inside its bounding box is repeated until there are no more elements left in D.

An example of detected and segmented LN obtained after the non-maximal suppression step is shown in Figure 6.

III. The Artificial Prediction Market

The artificial prediction market is a machine learning tool designed to simulate a real prediction market. A prediction market is a forum where people can buy and sell contracts for the outcomes of future events of interest such as presidential elections or international conflicts. The contract prices fluctuate based on the market phenomena of supply and demand. After the outcome is known, the owners of contracts for the correct outcome can redeem their face value for cash.

The artificial prediction market simulates the Iowa Electronic Market [7], which is a prediction market where each contract pays $1 if the outcome is realized. It has been observed that the price of the contracts for an outcome is a good approximation of the probability that the outcome is realized.

As opposed to the Iowa Electronic Market, the participants in the artificial prediction market are computer algorithms, so we can assume to know the number of contracts they will buy for any possible market price. The participants will try to predict the outcome for a number of instances which are vectors $\boldsymbol{x}_i \in R^F$.

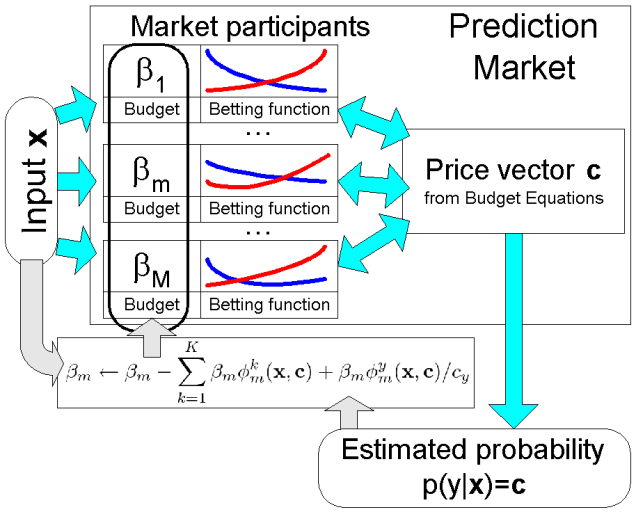

Fig. 7. Diagram of the market training using an instance $\boldsymbol{x}$.

Assume there are K possible outcomes (or classes). The market price is a vector $\boldsymbol{c}$=(c_1,…,c_K) where each $c_i \geq 0$. To obtain probabilities, we assume that $c_1+\ldots+c_K=1$. Thus the market price $\mathbf{c}$ belongs to the simplex $\Delta \subset R^K$.

A participant in the artificial prediction market is a pair (β, ϕ) where β is its budget, representing the amount of virtual money the participant currently possesses. The betting function $\phi(\boldsymbol{x}, \boldsymbol{c})$: $R^F \times \Delta \rightarrow R^K$ represents percentage of the budget β the participant will spend for the current market price $\boldsymbol{c}$ and current instance $\boldsymbol{x} \in R^F$.

A. The Market Update Equation

Given and instance $\boldsymbol{x} \in R^F$, assume the current contract prices form the vector $\boldsymbol{c} \in \Delta$ and the correct outcome is y. Then after the participant (β_m, ϕ_m) purchases contracts at the

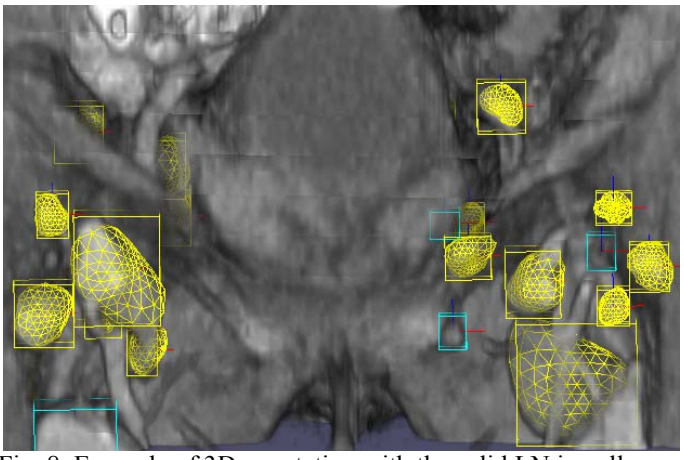

Fig. 8. Example of 3D annotation with the solid LN in yellow and non-solid in light blue.

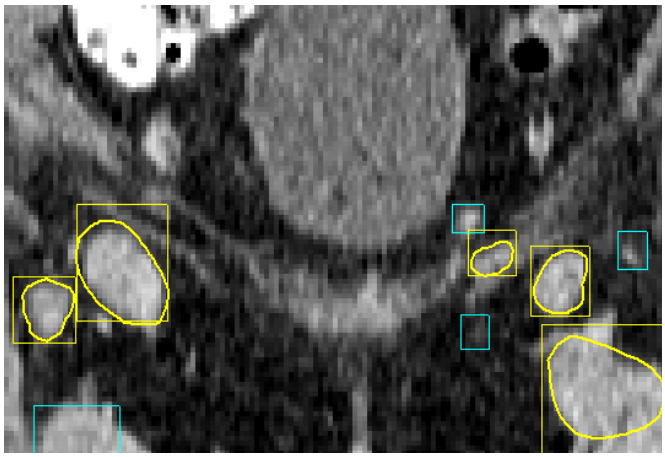

Fig. 9. Cross-section on the example from Figure 8.

price $\boldsymbol{c} \in \Delta$ and is rewarded based on the number of contract purchased for the class y, its budget is updated as follows:

$$\beta_m \leftarrow \beta_m - \beta_m \sum_{k=1}^{K} \phi_m^k(\boldsymbol{x},\boldsymbol{c}) + \frac{\beta_m}{c_y} \phi_m^y(\boldsymbol{x},\boldsymbol{c})$$

B. *The Budget Equations*

Assuming that the sum of the budgets of all participants is the same after presenting an instance to the market participants, independent of the possible outcome, the contract prices $\boldsymbol{c}$ must satisfy the following equations:

$$\sum_{m=1}^{M} \beta_m \phi_m^k(\boldsymbol{x},\boldsymbol{c}) = c_k \sum_{m=1}^{M} \sum_{i=1}^{K} \beta_m \phi_m^i(\boldsymbol{x},\boldsymbol{c})$$

In many cases, these equations are sufficient to determine a unique price vector $\boldsymbol{c}$.

In general, these equations can be solved by the double bisection method or by the Mann Iteration (a fixed point algorithm). For more details, see [6].

In the case when all betting functions are *constant* with respect to the price $\boldsymbol{c}$, i.e. $\phi_m(\boldsymbol{x}, \boldsymbol{c}) = \phi_m(\boldsymbol{x})$, the contract price has a simple formula

$$\boldsymbol{c} = \frac{\sum_{m=1}^{M} \beta_m \phi_m(\boldsymbol{x})}{\sum_{m=1}^{M} \sum_{k=1}^{K} \beta_m \phi_m^k(\boldsymbol{x})}$$

C. *Training the Artificial Prediction Market*

Given a number of training instances $(x_i, y_i) \in R^F \times \{1, \dots, K\}, i \in \{1, \dots, N\}$, the artificial prediction market is trained by presenting the training instances to the market, one at a time, computing the market price $\boldsymbol{c}$ for each presentation using the budget equations and updating the participant budgets accordingly. The procedure is illustrated in Figure 7.

The market is trained for a number of epochs, where one epoch consists of presenting all training instances once to the market. After training, the participants have budgets reflecting their capability of predicting the current outcomes compared to the other participants.

It has been observed experimentally in [6] that a trained market is better at predicting the class outcome using the price vector $\boldsymbol{c}$ than an untrained market where all participants have the same budget. In fact, it has also been proved in [6] that for the prediction market with constant betting functions, the market training actually performs maximum likelihood optimization by a version of stochastic gradient ascent.

D. *Specialized Market Participants*

It is not necessary that all market participants be involved in betting for each possible instance $\boldsymbol{x} \in R^F$. There could exist a market participant (β_m, ϕ_m) that has a subdomain $D_m \subset R^F$ for which it participates, since it has been trained to predict well only for instances in that subdomain. In this case the function ϕ_m is zero outside of the subdomain D_m. These are the specialized participants.

Examples of specialized participants are the leaves of the trees in a random forest. If each leaf is given a budget, they can be combined into a prediction market, for example using constant betting. These participants are specialized since they bet 0 outside the domain specified by the attribute values that define each leaf.

E. *Prediction Market for Lymph Node Verification*

As already mentioned in Section II.D, from the 17000 segmentation based features are selected a small number F of features by Adaboost. The number of selected features can be decided by cross-validation. For the pelvic lymph node detection application have been selected F=32 features.

For each feature i Adaboost found a classifier that divides the range of the feature values in B=64 bins. For each observation $\boldsymbol{x}$ the classifier finds the bin k corresponding to the feature value x_i and returns the value $b_{ik} \in \{-1,1\}$.

The market used for lymph node verification is a constant market with participants the $M = F \times B = 2048$ bins of the $F = 32$ selected features. Thus the participants can be indexed as (i, k) where i is the feature index and k is the bin

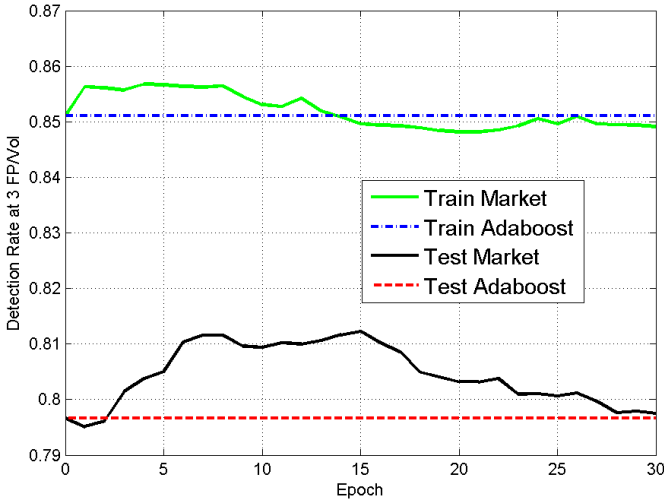

Fig. 10. Detection rate at 3FP/vol vs. number of training epochs of the market.

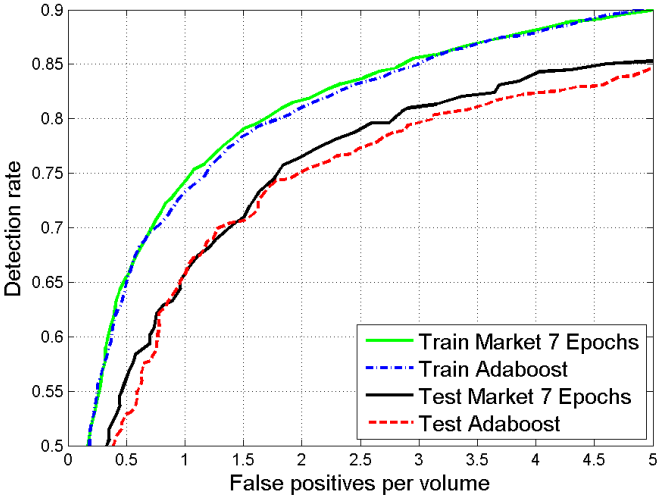

Fig. 11. ROC curve for the market trained 7 epochs.

index. Following the notation from Section II.D, participant (i,k) only bets on observations $\boldsymbol{x}$ that have the value of feature x_i in the range $(l_i^{k-1}, l_i^k]$. Therefore this market contains specialized participants. The participant (i,k) only bets for the class b_{ik}, and its betting function is

$$\phi_{ik}(\boldsymbol{x},\boldsymbol{c}) = \begin{cases} b_{ik} \text{ if } x_i \in (l_i^{k-1}, l_i^k] \\ 0 \quad \text{else} \end{cases}$$

Observe that for a fixed feature i only one of the participants $(i,k), k = 1 \ldots B$ will bid for any observation. Denote by $k(\boldsymbol{x},i)$ the index k such that $x_i \in (l_i^{k-1}, l_i^k]$.

Denote by β_{ik} the budget of participant (i,k). Then the contract price $\boldsymbol{c} = (c_{-1}, c_1)$ for observation $\boldsymbol{x} = (x_1, \ldots, x_M)$ is

$$c_1(\boldsymbol{x}) = \frac{\sum_{i,b_{ik(\boldsymbol{x},i)}=1} \beta_{ik(\boldsymbol{x},i)}}{\sum_{i=1}^{M} \beta_{ik(\boldsymbol{x},i)}}, c_{-1}(\boldsymbol{x}) = 1 - c_1(\boldsymbol{x})$$

The market is initialized with $\beta_{ik} = \beta_i > 0$, where β_i is the Adaboost coefficient of the weak classifier corresponding to feature i. Thus the initial market price is in a linear relationship with to the normalized Adaboost score $h(\boldsymbol{x})$

$$c_1(\boldsymbol{x}) = \frac{h(\boldsymbol{x}) + 1}{2}, h(\boldsymbol{x}) = \frac{\sum_{i=1}^{M} \beta_i h_i(x_i)}{\sum_{i=1}^{M} \beta_i}$$

The market is updated with the budget update equations from Section III.A by presenting the training examples one at a time for a number of epochs.

IV. Experiments

The Artificial Prediction Market from Section III.E was used for lymph node verification. First the market is trained with the same training examples that were used to train the Adaboost verification classifier. The positive examples are the feature vectors obtained from positions close to the annotated lymph node centers. The negatives are the feature vectors from candidates detected as described in Section II.B, about 3000 per volume, that are not inside any lymph node.

Since the number of positive examples is much smaller than the number of negatives, we modify the prediction market formulation by adding weights to the training examples, with the positives having weights $w_i = 1/N_{pos}$ and the negatives $w_i = 1/N_{neg}$, where N_{pos}, N_{neg} are the number of positives and negatives respectively.

The market price $\boldsymbol{c}$ is computed the same way, but when presented with observation i with weight w_i the budgets are updated by the modified equation:

$$\beta_m \leftarrow \beta_m - \beta_m w_i \sum_{k=1}^{K} \phi_m^k(\boldsymbol{x},\boldsymbol{c}) + \frac{\beta_m}{c_y} w_i \phi_m^y(\boldsymbol{x},\boldsymbol{c})$$

The system is evaluated on a dataset containing 54 volumes and 569 solid lymph nodes. The volumes also contain non-solid lymph nodes, which were annotated with a different color. An example of annotation is shown in Figure 8 with a cross-section shown in Figure 9.

The training examples are obtained from the LN candidates detected by the LN candidate detector described in Section II.C. The positive examples are obtained from the centers of the solid lymph node annotations. The negatives were all detected candidates outside the solid or non-solid lymph nodes.

Similar to [4], the detection criteria are as follows: a lymph node detection is considered correct if its center is inside a solid lymph node mesh. A detection is considered a false positive if it is outside all solid lymph node meshes and non-solid lymph node bounding boxes. Thus a detection inside a non-solid lymph node bounding box is neither correct nor it is a false positive.

The market is trained using the budget update equation shown above for up to 30 epochs. For each observation, the prediction from the market is obtained as the market price vector $\boldsymbol{c} = (c_{-1}, c_1)$ with c_1 being the estimated probability of class 1, i.e. the lymph nodes. The performance of the market classifier after each epoch is calculated on the training set and the test set.

Method	Target Area	# Cases	LN Size	#LN	#TP	#FP	TPR	FP/Vol	PPV	Time/Vol
Prediction Market trained 7 epochs	Pelvic+Abdominal	54	>10mm	569	463	170	81.4%	3.1	73.1%	15-40sec
Barbu [4]	Pelvic+Abdominal	54	>10mm	569	455	172	80.0%	3.2	72.6%	15-40sec
Feulner [5]	Mediastinum	54	>10mm	289	153	167	52.9%	3.1	47.8%	106sec
Feuerstein [1] implementation	Pelvic+Abdominal	54	>10mm	569	424	5741	74.5%	106.3	0.07%	5-10min
Feuerstein [1]	Mediastinum	5	>1.5mm	106	87	567	82.1%	113.4	13.3%	1-6min
Kitasaka [2]	Abdomen	5	>5mm	221	126	290	57.0%	58	30.3%	2-3h
Dornheim [3]	Neck	1	>8mm	29	29	9	100%	9	76.3	17min

TABLE I. LYMPH NODE DETECTION RESULTS

In Figure 10 is shown the detection rate at 3 FP/vol (false positives per volume), which is a clinically acceptable false positive rate. One could see that the detection rate on the test set increases from about 79.6% to about 81.2% at epoch 15 and then it starts decreasing.

In Figure 11 are shown the ROC curves on the training and test sets for the Adaboost classifier and the prediction market trained for 7 epochs. One could see that the performance on the test set improves for a large range of false positive rate, except from 0.7 to 1.7 FP/vol.

In Table I is shown one point from the ROC curve from Figure 11, where the false positive rate is about 3FP/volume. The results from [1], [2], [3], [4], [5] are also shown for comparison. One could see that compared to [4], which is the same as the "Test Adaboost" from Figures 10 and 11, the detection rate increases from 80.0% to 81.4%.

V. CONCLUSION AND FUTURE WORK

In this paper we presented an application of the Artificial Prediction Market to improve the lymph node detection performance from CT images. By training the final verification classifier with the Artificial Prediction Market the detection performance improved from 80.0% with 3.2 FP/volume to 81.4% with 3.1 FP/volume.

This improvement is probably due to the fact that the bins of the Adaboost classifier don't have the best possible weights, and the Prediction Market found better weights by a constraint version of Maximum Likelihood optimization.

In the future we plan to study methods for feature selection for the prediction market, in which the weakest participants are gradually removed. Another direction of study is finding other types of specialized participants besides the leaves of decision trees.

REFERENCES

[1] M. Feuerstein, D. Deguchi, T. Kitasaka, S. Iwano, K. Imaizumi, Y. Hasegawa, Y. Suenaga and K. Mori, "Automatic mediastinal lymph node detection in chest CT," in *SPIE*, 2009.

[2] T. Kitasaka, Y. Tsujimura, Y. Nakamura, K. Mori, Y. Suenaga, M. Ito and S. Nawano, "Automated extraction of lymph nodes from 3-d abdominal ct images using 3-d minimum directional difference filter," *LNCS,* vol. 4792, p. 336, 2007.

[3] L. Dornheim and J. Dornheim, "Automatische Detektion von Lymphknoten in CT-Datensatzen des Halses," in *BMV*, 2008.

[4] A. Barbu, M. Suehling, X. Xu, D. Liu, S. K. Zhou and D. Comaniciu, "Automatic Detection and Segmentation of Lymph Nodes from CT Data," *IEEE Trans Medical Imaging,* vol. 31, no. 2, p. 240–250, 2012.

[5] J. Feulner, S. Zhou, M. Hammon, J. Hornegger and D. Comaniciu, "Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior," *Medical Image Analysis,* 2012.

[6] A. Barbu and N. Lay, "An Introduction to Artificial Prediction Markets for Classification," *J. of Machine Learning Research,* vol. 13, p. 2177–2204, 2012.

[7] J. Wolfers and E. Zitzewitz, "Prediction markets," *Journal of Economic Perspectives,* pp. 107-126, 2004.

[8] R. Schapire, "The strength of weak learnability," *Machine Learning,* vol. 5, no. 2, p. 197–227, 1990.

[9] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering and D. Comaniciu, "Four-Chamber Heart Modeling and Automatic Segmentation for 3D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features," *IEEE Trans. Medical Imaging,* no. 11, 2008.

[10] H. Costin, "A Fuzzy Rules-Based Segmentation Method for Medical Images Analysis," *Int. J. of Comp. Comm. and Control,* vol. 8, no. 2, pp. 196-205, 2013.