WILEY **Statistics** in Medicine

Discovering anomalous patterns in large digital pathology images

Sriram Somanchi[1] (iD) | Daniel B. Neill[2] | Anil V. Parwani[3]

[1] IT, Analytics, and Operations, Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556, U.S.A.

[2] Event and Pattern Detection Laboratory, H.J. Heinz III College, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213, U.S.A.

[3] Department of Pathology, Ohio State University, Columbus OH 43210, U.S.A.

Correspondence
Sriram Somanchi, 344 Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556, U.S.A.
Email: somanchi.1@nd.edu

Advances in medical imaging technology have created opportunities for computer-aided diagnostic tools to assist human practitioners in identifying relevant patterns in massive, multiscale digital pathology slides. This work presents Hierarchical Linear Time Subset Scanning, a novel statistical method for pattern detection. Hierarchical Linear Time Subset Scanning exploits the hierarchical structure inherent in data produced through virtual microscopy in order to accurately and quickly identify regions of interest for pathologists to review. We take a digital image at various resolution levels, identify the most anomalous regions at a coarse level, and continue to analyze the data at increasingly granular resolutions until we accurately identify its most anomalous subregions. We demonstrate the performance of our novel method in identifying cancerous locations on digital slides of prostate biopsy samples and show that our methods detect regions of cancer in minutes with high accuracy, both as measured by the ROC curve (measuring ability to distinguish between benign and cancerous slides) and by the spatial precision-recall curve (measuring ability to pick out the malignant areas on a slide which contains cancer). Existing methods need small scale images (small areas of a slide preselected by the pathologist for analysis, eg, 32×32 pixels) and may not work effectively on large, raw digitized images of size $100K \times 100K$ pixels. In this work, we provide a methodology to fill this significant gap by analyzing large digitized images and identifying regions of interest that may be indicative of cancer.

KEYWORDS
anomalous pattern detection, pathology informatics, prostate cancer, subset scanning

1 | INTRODUCTION

1.1 | Background

Anatomic pathology is a medical specialty, which includes diagnosing a disease from biopsy samples of an organ. For decades, pathology workflows have been highly manual: thin slices of biopsy specimens are histochemically stained on a glass slide and analyzed by a pathologist under an optical microscope. More recently, advances in computer-aided medical diagnostics have introduced a digital workflow, and use of digital pathology has grown dramatically in the last 10 years.[1] Many pathology laboratories are on the path towards modernizing and updating their work flows using these advanced techniques. Digital pathology offers many advantages over traditional anatomic pathology, including secure

and easy distribution of digital images and the ability to use informatics tools to analyze them. It is typically believed that pathologists are often under time pressure to analyze these images,[2] which may lead to misdiagnoses. Thus, the use of automated tools to assist in diagnosis and focus attention on the most relevant areas of the slide has great potential to improve the digital pathology workflow, enable pathologists to better manage their workload, and improve overall accuracy of diagnosis. Automatically differentiating between cancerous and benign images can help prioritize cases that need urgent attention, while identification of malignant areas on a slide can provide specific evidence in support of the cancer diagnosis. In this paper, we introduce novel detection methods that can aid pathologists by quickly identifying potential regions of interest in a high-resolution digital image. These regions are automatically detected and highlighted in an image viewer for further examination. This process helps in timely diagnosis and can provide input for secondary opinions on regions of interest that the pathologist might have missed.

1.2 | Approach based on subset scanning

Typically, the regions of interest for a pathologist in a tissue biopsy contain patterns that are abnormal as compared with the rest of the tissue structure. Hence, our approach for aiding medical diagnostics is to detect regions that contain anomalous patterns in these large-scale multiresolution images. Specifically, we approach pattern detection as a "subset scan" problem,[3,4] where we search over subsets of data with the goal of finding the most anomalous subsets. There are 2 main challenges in subset scan approaches: statistically quantifying the anomalousness of a subset and efficiently searching over the exponential number of subsets to identify the most anomalous subsets. Anomalousness of a subset can be typically quantified by a log-likelihood ratio statistic like the Expectation Based Poisson[4,5] or Expectation Based Gaussian[3] statistics. A major computational hurdle of the "subset scan" approach to pattern detection is that we need to optimize the score function over an exponential number of subsets. Linear Time Subset Scanning (LTSS)[3,4] is a novel approach to anomalous pattern detection that enables efficient computation by identifying the most anomalous subset without exhaustive search.

For the many scoring functions that satisfy the LTSS property, we can find the subset that maximizes the scoring function by evaluating only a linear number of subsets: for N data elements, the number of subsets evaluated is proportional to N rather than 2^N. However, when the data is huge (ie., number of data elements is in billions or trillions) even evaluating a linear number of subsets might not be feasible. Also, LTSS only finds the best subset of pixels, without enforcing any constraints on the detected subsets, and thus might include many "noise" pixels which are individually anomalous simply by chance. In this paper, we introduce a new framework called Hierarchical Linear Time Subset Scanning (HLTSS) to address these issues. HLTSS assumes that data elements are aggregated at multiple levels of hierarchy. For example, in the case of digital pathology, the image of a biopsy sample is stored at multiple resolutions in order to help a pathologist analyze it through virtual microscopy. HLTSS takes advantage of this hierarchical structure in the data to achieve high accuracy and computational efficiency for detecting anomalous subsets. We show that any scoring function that satisfies the LTSS property can be incorporated into our new HLTSS framework.

1.3 | Application

We develop the HLTSS framework for identifying regions of interest in digital pathology images. Digital pathology images are stored at multiple resolutions, as "Whole Slide Images" (WSI), with the lowest (coarsest) resolution image providing the overall big picture and the highest (finest) resolution image providing very detailed structure of the biopsy sample. There is a pyramidal structure to WSI where each level in the pyramid is an image at some resolution. A special file format (Tagged Image File Format (TIFF), ScanScope Virtual Slide (SVS)) is necessary to store these multiple resolution images in a single WSI file.[6] Figure 1 provides an example image with multiple resolution images shown in the form of a pyramid. Typically, biopsy samples are placed on a glass slide and histologically stained for physical examination. These glass slides are converted to WSI by various high resolution scanners (Aperio, Mirax, etc).

At the highest resolution, these images are very large, typically in the range of $100K \times 100K$ pixels. A pathologist typically starts from low resolution and zooms in to high resolution to pinpoint abnormal regions on a slide. Our method is motivated by this procedure: we take a multiresolution digital image, identify the most anomalous regions at a coarse level, and continue to analyze the data at increasingly granular resolutions until we accurately identify the most anomalous subregions. Our objectives are to perform this task with high detection power for distinguishing cancerous from benign images, high accuracy for pinpointing the affected areas of a cancerous image, and computational efficiency.
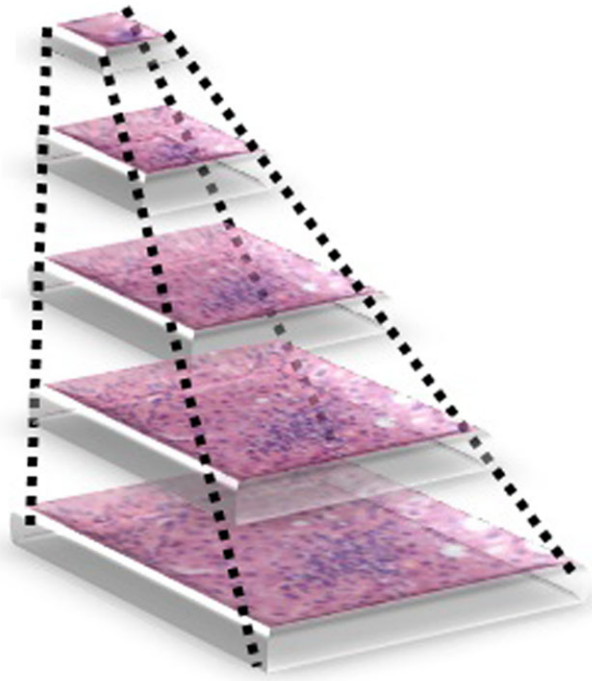
FIGURE 1 An example of a Hematoxylin and Eosin (HE) stained Whole Slide Image, where images at multiple resolutions are stored in a pyramidal structure. This structure helps a pathologist to efficiently view the slide through virtual microscopy [Colour figure can be viewed at wileyonlinelibrary.com]

Each pixel in an image is typically represented in Alpha, Red, Green, Blue (ARGB) format, with 8 bits each for Red, Green, and Blue (RGB) components of pixel color. Pyramidal images are typically representative of the hierarchical aggregation of each component.[6] That is, if a given pixel at a higher (coarser) level is representative of a grid of 2×2 pixels at a lower level in the hierarchy, then each component (R,G,B) of the higher-level pixel is approximately the average of the corresponding component of the 4 pixels below it.

Digital pathology images are typically Hematoxylin and Eosin (HE) stained images. The resulting image is composed of hematoxylin (violet color), eosin (pink color), and white (background color). Hematoxylin is typically indicative of cell nuclei, and pink is indicative of other tissue regions like cytoplasm. In this work, we focus on identifying regions that contain a higher than expected concentration of violet pixels (hematoxylin dye).

There are various applications where identifying this pattern is potentially useful: identifying regions of inflammation in gastrointestinal tracts for Crohn's disease; finding regions of inflammation (gastritis) in the stomach, which may be indicative of colonization by *Helicobacter pylori*; and diagnosis of prostatic intraepithelial neoplasia, which may lead to prostate cancer. Specifically, we apply this methodology for identifying cancerous regions in a prostate biopsy sample, as described below.

1.3.1 | Detection of prostate cancer

Prostate cancer is the most prevalent form of cancer and second most common cause of cancer deaths among men in the U.S.[7] About 1 in 6 men will be diagnosed with prostate cancer during their lifetime.[7] Pathologists rely on examination of biopsy samples under a microscope for detecting cancerous cases. It is extremely important for pathologists to accurately identify cases of cancer so that early intervention can improve the prognosis. For diagnosis of prostate cancer, the single most important feature is the presence of a prominent and enlarged nucleolus,[1] as well as darker coloration of the cytoplasm,[8] and thus areas affected by prostate cancer are generally more violet in color than benign locations.[8] Pathologists also use luminal and architectural characteristics, which are based on cell shape, and our method does not take these into account. Nevertheless, because of the importance of the nucleic and cytoplasmic color characteristics, we hypothesize that our anomaly detection algorithm (which finds regions with higher than expected concentrations of violet pixels) can be used for identifying prostate cancer.

1.4 | Related work

Previous work applies machine learning approaches to automatically classify pathology images as cancerous or benign.[9-13] These methods extract image features from labeled training examples and use standard techniques like Support Vector Machines or penalized logistic regression to classify images as cancerous or noncancerous. Recent work by Janowczyk and Madabhushi,[14] described in more detail in Section 3.2, uses deep learning for subtype classification of digital pathology images. However, this method uses small-scale images of size 32×32 pixels (small areas of a slide preselected by the pathologist for analysis) and may not work effectively on large, raw digitized images of size $100K \times 100K$ pixels. In this paper, we provide a methodology to fill this significant gap by analyzing large digitized images and identifying regions of interest that may be indicative of cancer. An additional, important distinction is that the above methods are based on supervised learning and require large numbers of manually labeled training examples, whereas our method is unsupervised and does not require training data.

1.5 | Contributions

We quantify the anomalousness of a region using the Expectation Based Binomial (EBB) scoring function, where we map each pixel to a continuum of violet (hematoxylin), pink (eosin), and white (absence of tissue) and pinpoint regions containing a higher than expected proportion of violet pixels. We show that EBB satisfies the LTSS property and hence can be efficiently optimized over subsets of pixels. Finally, we show how to incorporate the hierarchical, multiresolution structure of digital pathology images and present the HLTSS framework for efficiently identifying regions of interest. We apply our methodology to digital images of prostate biopsy samples from the Department of Pathology at University of Pittsburgh Medical Center (UPMC) to identify cancerous locations in these images. We show that our methodology helps in differentiating between cancerous and benign images. Further, we show that our methods have good accuracy in picking out the malignant areas on a slide which contains cancer.

In the remainder of this paper, we discuss the application of EBB for scoring areas of anomalous coloration and propose our HLTSS framework in Section 2. Evaluation results on simulated data are shown in Section 3. We show the effectiveness of our methods in identifying cancerous locations in prostate biopsy samples and provide case studies and discussion of accuracy results in Section 4. We conclude and provide future directions of this work in Section 5.

2 | METHODOLOGY

We first describe the EBB scoring function used for statistically quantifying a region as anomalous, and then present the HLTSS framework for efficiently identifying the regions of interest.

2.1 | Expectation Based Binomial (EBB) for scoring discoloration

To score a subset S, which may consist of pixels s_i at multiple levels of hierarchy, we first compute its aggregate color (R, G, B) and number of base-level pixels N: $R = \frac{\sum_{s_i \in S} R_i * N_i}{\sum_{s_i \in S} N_i}$, $G = \frac{\sum_{s_i \in S} G_i * N_i}{\sum_{s_i \in S} N_i}$, $B = \frac{\sum_{s_i \in S} B_i * N_i}{\sum_{s_i \in S} N_i}$, and $N = \sum_{(s_i \in S)} N_i$, where each pixel s_i has color (R_i, G_i, B_i) and corresponds to N_i base-level pixels. (For example, if pixel s_i is in the level with downsampling of 4, then it represents $N_i = 4 * 4 = 16$ base-level pixels.) Our generative model assumes that, at the base level, each pixel is either white, pink, or violet, where $White = (R_w, G_w, B_w)$, $Pink = (R_p, G_p, B_p)$, and $Violet = (R_v, G_v, B_v)$ can be represented as 3-dimensional points in RGB space. Given the aggregated color (R, G, B) for subset S, we can compute the corresponding proportions of white, pink, and violet pixels by projecting the point onto the plane formed by $White$, $Pink$, and $Violet$ and computing W, P, V such that $(R, G, B) = (W * White + P * Pink + V * Violet)$, where $W, P, V \geq 0$ and $W + P + V = 1$. See Section 2.5 for more details. The 3 base colors $White$, $Pink$, and $Violet$ can be estimated from a small number of slides from a given laboratory; we do this with an algorithm presented by Ruifrok and Johnston[15] using ImageJ software.[16]

Given the aggregated white, pink, and violet values (W, P, V) of subset S and its corresponding number of base-level pixels N, we compute the score of subset S using the log-likelihood ratio statistic $F(S) = \log \frac{P(Data|H_1(S))}{P(Data|H_0)}$, where the null hypothesis H_0 assumes that no subset of pixels is affected, and the alternative hypothesis $H_1(S)$ assumes that subset

S is affected because of cancer in the corresponding cells. Our generative model assumes that each nonwhite pixel at the base level is either violet with probability p_i or pink with probability $1-p_i$; note that white is the background color and is eliminated from the analysis as it may not be pathologically relevant. For benign images, we assume $p_i = p_0$ everywhere for some constant p_0, which can be estimated from a small sample of benign images. Further, we assume that if a region is affected by cancer, then the probability of violet pixels in the region increases by a multiplicative factor $q > 1$. Let $c_i \in \{0,1\}$ represent the state of a nonwhite pixel s_i, where $c_i = 1$ and $c_i = 0$ represent violet and pink, respectively. We then test the null hypothesis H_0 against the set of alternative hypothesis $H_1(S,q)$, where $H_0: c_i \sim Bernoulli(p_0)$ for all pixels s_i, and $H_1(S,q): c_i \sim Bernoulli(q*p_0)$ for pixels $s_i \in S$ and $c_i \sim Bernoulli(p_0)$ for $s_i \notin S$. The quantity q represents the relative risk, ie, the multiplicative increase in the probability of generating violet pixels. As is typical in the scan statistic framework, we use a Generalized Likelihood Ratio Test, which assumes the maximum likelihood estimate of the free parameter q. The log-likelihood ratio statistic $F(S)$ can then be written as follows:

$$F(S) = \log\frac{\max_{1<q<\frac{1}{p_0}}\prod_{s_i \in S}Pr(c_i \sim Bernoulli(qp_0))\prod_{s_i \notin S}Pr(c_i \sim Bernoulli(p_0))}{\prod_{s_i}Pr(c_i \sim Bernoulli(p_o))}$$

$$= \log\frac{\max_{1<q<\frac{1}{p_0}}\prod_{s_i \in S}Pr(c_i \sim Bernoulli(qp_0))}{\prod_{s_i \in S}Pr(c_i \sim Bernoulli(p_0))}.$$

Now, given the values of (W,P,V) and number of base-level pixels N for subset S, we compute the effective number of violet pixels $C = V*N$ and the effective number of nonwhite pixels $N^{eff} = (1-W)N = (V+P)N$. Substituting the parametric equations for the above formulation we get:

$$F(S) = \max_{1<q<\frac{1}{p_0}} \log\prod_{s_i \in S}\frac{(qp_0)^{C_i}(1-qp_0)^{N_i^{eff}-C_i}}{p_0^{C_i}(1-p_0)^{N_i^{eff}-C_i}}$$

$$= \max_{1<q<\frac{1}{p_0}} C \log q + (N^{eff}-C)\log\frac{1-qp_0}{1-p_0}.$$

We compute the maximum likelihood estimate as $\hat{q} = \dfrac{C}{N^{eff}p_0}$, and hence the q that maximizes the above equation is $max(\hat{q}, 1)$. Hence, our scoring function is reduced as follows:

$$F(S) = C \log\frac{C}{N^{eff}p_0} + (N^{eff}-C)\log\frac{1-\frac{C}{N^{eff}}}{1-p_0} \qquad \text{if } C > N^{eff}p_0$$

$$= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise.}$$

This can be further rewritten as

$$F(S) = N\times\left(V\log\frac{V}{(V+P)p_0} + P\log\frac{P}{(V+P)(1-p_0)}\right) \qquad \text{if } \frac{V}{V+P} > p_0$$

$$= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise.}$$

With this notion of color brought into the scoring function, we can interpret $\dfrac{p_0}{1-p_0}$ as the ratio of the number of violet pixels to pink pixels at the base-level image in a normal (benign) slide.

2.2 | EBB scoring function and LTSS

In the above section, we describe the EBB scoring function for measuring the anomalousness of a subset S of pixels. However, in an image with N base-level pixels, there are 2^N possible subsets to consider. To use a scoring function, including EBB, in our efficient hierarchical framework, it must satisfy the LTSS property,[3] as any function satisfying LTSS can be efficiently and exactly maximized over all subsets of the data. More precisely, for a pair of functions $F(S)$ and $G(s_i)$, which represent the "score" of subset S and the "priority" of data record s_i, respectively, the LTSS

property guarantees that the only subsets with the potential to be optimal are those consisting of the top k highest priority records $\{s_{(1)},...,s_{(k)}\}$, for some k between 1 and N. This property enables us to search only N of the 2^N subsets of records, while still guaranteeing that the highest-scoring subset will be found.

Theorem 1. *The EBB scoring function* $F(S)$ *satisfies the LTSS property with priority function* $G(s_i) = \dfrac{C_i}{N_i^{eff}}$.

Proof We note that Binomial distributions are members of the exponential family of distributions; therefore, we can write the log-likelihood function as follows:

$$\log P(C_i|p_i, N_i^{eff}) = h(C_i) + T(C_i)(\mu_i) - \psi(\theta(\mu_i)),$$

where $\qquad h(C_i) = \begin{pmatrix} N_i^{eff} \\ C_i \end{pmatrix} \qquad , \qquad T(C_i) = C_i, \qquad \theta(\mu_i) = \log\left(\dfrac{p_i}{1-p_i}\right) = \log\dfrac{\mu_i}{N_i^{eff} - \mu_i}, \qquad$ and

$$\psi\{\theta(\mu_i)\} = N_i^{eff}\log\left(\dfrac{1}{1-p_i}\right) = N_i^{eff}\log\dfrac{N_i^{eff}}{N_i^{eff} - \mu_i}.$$

Here, C_i represents the observed number of violet base-level pixels, N_i^{eff} represents the observed number of nonwhite (pink or violet) base-level pixels, p_i represents the assumed probability that a nonwhite base-level pixel is violet, and $\mu_i = N_i^{eff}p_i$ represents the expected number of violet base-level pixels. Under the null hypothesis H_0, we assume $p_i = p_0$ everywhere for some constant p_0, and under the alternative hypothesis $H_1(S)$, we assume $p_i = qp_0$ for pixels in subset S and $p_i = p_0$ for pixels outside subset S, for some constant $q > 1$. Given these assumptions, we can write the generalized log-likelihood ratio statistic for EBB scoring function as follows:

$$F(S) = sup_{q>1}(T(C_i)[\theta(q\mu_i) - \theta(\mu_i)] + \psi\{\theta(\mu_i)\} - \psi\{\theta(q\mu_i)\}).$$

The EBB belongs to the separable exponential family of distributions (as defined in Neill[3]), as we can write

$$\theta(q\mu_i) = \log\dfrac{q\mu_i}{N_i^{eff} - q\mu_i} = \log\dfrac{qp_0 N_i^{eff}}{N_i^{eff} - qp_0 N_i^{eff}} = \log\dfrac{qp_0}{1-qp_0} = z_i\theta_0(q) + v_i,$$

where $z_i = 1$, $v_i = 0$ and $\theta_0(q) = \log\left(\dfrac{qp_0}{1-qp_0}\right)$. Note that $\theta_0(q)$ is only a function of q as p_0 is assumed to be a known constant. Therefore, from Theorem 2 of Neill,[3] the EBB scoring function satisfies the LTSS property with priority function of a given pixel s_i as $G(s_i) = \dfrac{T(C_i)}{\mu_i} = \dfrac{C_i}{p_0 N_i^{eff}} \propto \dfrac{C_i}{N_i^{eff}}$. $\qquad\square$

However, an unconstrained scan on the base-level pixels is not necessarily what we want, both because of the large number of pixels (~10 billion) to be scanned and the fact that such a scan fails to incorporate spatial information from nearby pixels on the slide. Hence, we propose the HLTSS framework to incorporate hierarchical constraints in the search and provide sublinear time algorithms for detection.

2.3 | Hierarchical linear time subset scanning (HLTSS)

The HLTSS methodology is designed to find anomalous patterns in very large scale data, taking advantage of the hierarchy of aggregated information at multiple levels to efficiently and accurately find anomalous subsets. We apply HLTSS to analyze digital pathology images stored at multiple resolutions, finding anomalous groups of pixels that are indicative of cancer. We use the EBB scoring function, which satisfies the LTSS property, to identify relevant clusters of pixels that contain a greater proportion of violet pixels than expected.

We assume that our data is aggregated in a hierarchical manner. In Figure 2, individual pixels are at the bottom level and are very large in number. In a WSI, each pixel at a coarser resolution is an aggregation of pixels at a more granular level. We assume that some small subset of these elements are generated by an anomalous process, and our goal is to find this subset in the large image as quickly and accurately as possible.

Our approach to anomalous pattern detection is through subset scanning. Given a scoring function, eg, a log-likelihood ratio statistic, to quantify the anomalousness of a subset, we want to efficiently find the small subset of most
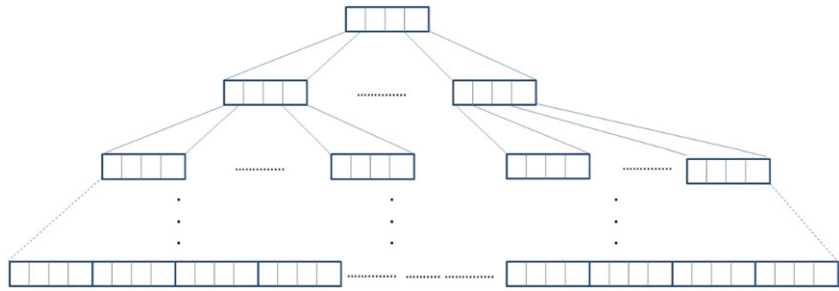
FIGURE 2 A general perspective of large scale data and its hierarchical aggregated information [Colour figure can be viewed at wileyonlinelibrary.com]

anomalous elements. We can employ existing techniques like LTSS[3] or Kulldorff's circular scan.[17] First, we note that these techniques are mostly linear or quadratic in terms of number of elements. For very large scale data, this might have prohibitively high cost both in terms of time and memory. Moreover, without the right constraints in the search procedure, the subsets returned might contain a lot of noise. Hence, we propose the HLTSS framework to incorporate hierarchical constraints in the search and provide sublinear time algorithms for detection.

Given a scoring function that satisfies the LTSS property, we propose the following HLTSS algorithm (Algorithm 1). Note that $F(S)$ is our scoring function, the start level L_s is some coarser level where the search procedure starts, and the end level L_e is the most granular aggregation in our analysis. In the experiments below, $L_s = 8$ (each start-level pixel represents a 256×256 block of base-level pixels) and $L_e = 2$ (each end-level pixel represents a 4×4 block of base-level pixels). Thus, a total of 7 levels were considered, with each parent node representing a 2×2 grid of child nodes at the next level.

Note that we are not individually processing all pixels at level L_e. Due to the hierarchical representation, HLTSS only processes a sublinear number of base-level pixels to detect anomalous regions, thus maintaining computational efficiency with the additional advantage of not returning random noise pixels (like individual violet pixels) that are typically lost in the hierarchical aggregation.

Algorithm 1 Algorithmic framework for Hierarchical Linear Time Subset Scanning

---

 1: **INPUT**: Scoring function $F(S)$, start level L_s, end level L_e, number of subsets returned k
 2: **OUTPUT**: Top k subsets of anomalous elements at level L_e
 3: **for** each pixel p_{start} at level L_s **do**
 4:     $list = \{p_{start}\}$
 5:     **while** any pixels in $list$ are of level $> L_e$ **do**
 6:         p_{list} = highest-priority pixel of level $> L_e$
 7:         Remove p_{list} and insert all children of p_{list} into $list$.
 8:         Find the best subset of the list $S^* = \arg\max_{S \subset list} F(S)$ using LTSS.
 9:         Remove all pixels $p \notin S^*$ from $list$
10:     **end while**
11:     Record $list(p_{start})$
12: **end for**
13: For each pixel p_{start} at level L_s we have $list(p_{start})$, the best subset of finer-grained pixels found by starting the hierarchical search at p_{start}.
14: Regroup the pixels among these various lists based on spatial adjacency. Combine two groups only if the score $F(S)$ of the combined group is greater than either of the individual scores.
15: Output the top k highest scoring regrouped pixels.

In step 9, removing all pixels that are not in the best subset S^* dramatically improves computation time without loss of accuracy. See Section 2.4 for details. Step 14 enforces connectivity of the resulting clusters, merging clusters across the fixed partitions imposed by the hierarchical decomposition of the slide, which helps to correctly identify cancerous regions that lie across these partitions.

2.4 | List pruning

In this section, we provide justification for removing all pixels that are not in the best subset S^* in Step 9 of Algorithm 1. We show that any pixel p that is not in the best subset in a given iteration of the algorithm will not be in the best subset

in any further iterations of the algorithm. Hence, p can be pruned from the list, thereby improving computational efficiency without loss of accuracy.

Theorem 2. *For any pixel* p, *if* $p \notin S_i^*$, *then* $p \notin S_{i'}^*$, *in iteration* i' *for any* $i' > i$, *where* S_i^* *and* $S_{i'}^*$, *are the highest scoring subsets of pixels, as defined by the* EBB *scoring function, in iterations* i *and* i', *respectively.*

Proof. Let us assume that the elements in list at iteration i are $list_i = \{p_1,...,p_N\}$, where the pixels are assumed to be ordered by the priority function G (ie, pixel p_1 has highest priority). Further, let the highest scoring subset in iteration i be $S_i^* = \{p_1, \cdots, p_k\}$ for some k, $1 \le k \le N$. (As shown in Theorem 1, the EBB scoring function satisfies the LTSS property,[3] guaranteeing that the highest scoring subset will be the top-k highest priority pixels). In iteration $i+1$, we expand the highest-priority pixel p_1 and include all of its M children pixels $\{p_{11},...,p_{1M}\}$ in the list. Now, given the new list on iteration $i+1$, $list_{i+1} = \{p_2, \cdots, p_N\} \cup \{p_{11}, \cdots, p_{1M}\}$, it is sufficient to show that $p_q \notin S_{i+1}^*$ for any $q > k$.

Let us assume that there exists some $q > k$ such that $p_q \in S_{i+1}^*$. Then by the LTSS property, the highest scoring subset is of the form $S_{i+1}^* = \{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_q\}$, where the r highest-priority children of pixel p_1 (for some r, $1 \le r \le M$) are included. We show that this leads to a contradiction, as $F(\{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_q\}) < F(\{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_k\})$, and hence the assumed S_{i+1}^* cannot be the highest scoring subset.

To see this, we first note that the EBB scoring function is convex. Thus, it follows that $F(\{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_k\} \cup \{p_{k+1}, \cdots, p_q\})$ + $F(\{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_k\} \cup \{p_{1(r+1)}, \cdots, p_{1M}\})$ $\le$ $F(\{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_k\})$ + $F(\{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_k\} \cup \{p_{k+1}, \cdots, p_q\} \cup \{p_{1(r+1)}, \cdots, p_{1M}\})$, and therefore $F(\{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_q\})$ + $F(\{p_1, \cdots, p_k\})$ $\le$ $F(\{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_k\})$ + $F(\{p_1, \cdots, p_q\})$.

Now, we know that $F(\{p_1,...,p_q\}) < F(\{p_1,...,p_k\})$, since $\{p_1,...,p_k\}$ is the highest scoring subset in iteration i, and thus $F(\{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_q\}) < F(\{p_{11}, \cdots, p_{1r}\} \cup \{p_2, \cdots, p_k\})$. Therefore, for any $p \notin S_i^*$, we know that $p \notin S_{i+1}^*$, and hence by induction $p \notin S_{i'}^*$ for any $i' > i$. □

Thus, we can prune the list by removing any pixels not in the best subset S^*, keeping the list to a manageable size and dramatically improving the efficiency of the algorithm. Figure 3 shows a comparison of run times (in minutes) between the HLTSS algorithm with and without list pruning, keeping the end level $L_e = 2$ fixed and varying the start level L_s. We can see that the difference in computation times with and without pruning increases drastically as the number of levels processed ($L_s - L_e + 1$) increases, while results produced by pruned and unpruned HLTSS were identical.

2.5 | Color mapping

Our generative model assumes that, at the base level, each pixel is either white, pink, or violet, where $White = (R_w, G_w, B_w)$, $Pink = (R_p, G_p, B_p)$, and $Violet = (R_v, G_v, B_v)$ can be represented as 3-dimensional points in RGB space. Given the aggregated color (R, G, B) for subset of pixels S, under the assumption of perfect aggregation in the hierarchical image,
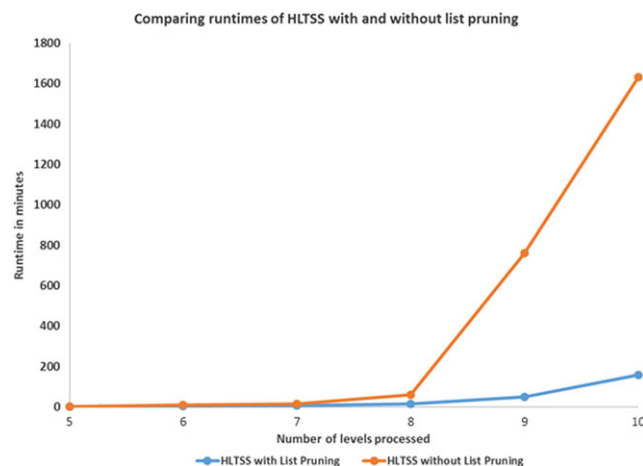

FIGURE 3 Runtime comparison of Hierarchical Linear Time Subset Scanning (HLTSS) algorithm with and without list pruning [Colour figure can be viewed at wileyonlinelibrary.com]

we can compute the proportions of white (W), pink (P), and violet (V), such that $(R,G,B) = (W*White + P*Pink + V*Violet)$, where $W,P,V \geq 0$ and $W + P + V = 1$:

$$
\begin{bmatrix} W \\ P \\ V \end{bmatrix} = \begin{bmatrix} R_w & R_p & R_v \\ G_w & G_p & G_v \\ B_w & B_p & B_v \end{bmatrix}^{-1} * \begin{bmatrix} R \\ G \\ B \end{bmatrix}.
$$

Typically, in the case of digital pathology, the glass slide is scanned at multiple levels of resolution by a high resolution camera with various levels of zooming. Therefore, in reality, not every color (R,G,B) may not be on the plane formed by *White*, *Pink*, and *Violet*. Thus, we first project each point (R,G,B) to a corresponding point (R^*,G^*,B^*) on the plane:

$$
(R^*, G^*, B^*) = (R, G, B) - ((R - R_w, G - G_w, B - B_w) \cdot \overrightarrow{n}) \overrightarrow{n},
$$

where

$$
\overrightarrow{n} = \frac{(Pink - White) \times (Violet - White)}{|(Pink - White) \times (Violet - White)|}.
$$

We use the projected point (R^*,G^*,B^*) in order to compute W, P, and V. This projection guarantees $W + P + V = 1$, but in rare cases (less than 1% of pixels in our data) where the projected point is not a convex combination of *White*, *Pink*, and *Violet*, the constraint $W,P,V \geq 0$ may not be satisfied. In these cases, we use the point (R^a,G^a,B^a), which is closest to (R^*, G^*,B^*) and still satisfies these conditions. We note that both the projection from (R,G,B) to (R^*,G^*,B^*) and the mapping from (R^*,G^*,B^*) to (W,P,V) are linear, and thus we expect the observed (W,P,V) values for a given pixel at coarse resolution to be very close to the mean of the (W,P,V) values of the corresponding base-level pixels. Two potential sources of error are that the (R,G,B) values do not aggregate perfectly because of the physical process of creating an image from a glass slide, as noted above, and the rarely occurring noisy pixels for which the projected point (R^*,G^*, B^*) does not map to nonnegative (W,P,V).

3 | EXPERIMENTAL SIMULATION AND RESULTS

The goals of the experimental simulation are to demonstrate the performance of our method for identifying anomalous regions in large digital pathology images and to compare its performance with competing methods across various experimental settings.

3.1 | Simulation setup

To replicate realistic conditions, we used the parameters of the real data (Section 4) to generate the simulated data. More specifically, we created images that at the most granular level have $2^{16} \times 2^{16}$ pixels, with each pixel being either pink or violet color. The probability of a pixel being violet (p_0), and the RGB combinations of a typical pink and violet color were observed from the actual digital pathology images. Note that for the purpose of this simulation, we have not generated any white pixels at the most granular level. After generating images at the most granular level, we built the hierarchical aggregations, where the RGB combination of each pixel at a given level is an aggregation of the corresponding RGB values of 4×4 pixels at the lower level. This process was repeated to generate 8 levels of hierarchically aggregated data.

Next, we injected simulated anomalous regions into these hierarchically aggregated images. These anomalous regions are created at the most granular level such that the probability of violet pixels is increased by a multiplicative factor inside the injected box. There are 2 parameters for injected regions: the size of the injected box and the effect size (q). Increasing the box size and/or the effect size makes the detection problem easier. The box sizes are $2^b \times 2^b$ pixels, where b is varied between 8 and 10 with increments of 0.25. Similarly, we vary the effect size q from 1.25 to 3 with increments of 0.25, that is, for a given value of q, the probability of the violet pixels inside the box is increased to $q*p_0$. For a given box size and effect size, we create multiple boxes (100 in our experimental setting) randomly injected at the most granular level. Note that the offsets of each box's x and y coordinates are chosen uniformly at random, and thus the injected box is very unlikely to align perfectly with the previously created image hierarchy.

3.2 | Methods

We are not aware of any previous methods that have been applied to large, raw digital pathology data for finding regions of interest. However, we compare the performance of our method with the recently proposed deep learning-based cancer classification algorithm described by Janowczyk and Madabhushi.[14] We will refer to this approach as the "Janowczyk and Madabhushi Deep Learning" (JMDL) algorithm throughout the discussion below, and note that additional details of the JMDL network architecture and input parameters can be found in reference.[14] The JMDL method uses a multilayer Convolutional Neural Network architecture which minimizes a softmax loss function. The standard "AlexNet" network architecture used by JMDL is described in detail in Krizhevsky et al[18] and summarized in Tables 2 and 3 of Janowczyk and Madabhushi.[14] We used the code provided by Janowczyk and Madabhushi,[14] with the default parameters, implemented on an open source framework (Caffe). Further, the JMDL algorithm expects the input image to be much smaller in scale (32×32 pixels). Hence, we partition the image at end level L_e into multiple images. Furthermore, we compare the performance of HLTSS with 2 variants of the basic LTSS algorithm: LTSS-granular, an unconstrained subset scan on the fine-grained pixels at end level L_e, and LTSS-coarse, an unconstrained subset scan on the coarse-grained pixels at start level L_s. In each case, after obtaining the highest-scoring subsets, we combine groups of pixels that are connected and then return the top k connected subsets. These methods are compared in order to understand whether and how the use of hierarchical structure benefits the performance of HLTSS.

3.3 | Evaluation metrics

We apply our HLTSS algorithm and the competing methods to the simulated data and evaluate the performance on 2 metrics: detection power and spatial overlap.

For a given combination of simulation parameters (q and box size), detection power measures a method's ability to detect the presence of an anomalous region in a given image. We generate scores for each competing method under the null hypothesis (H_0), where there are no anomalous regions injected in the data. More specifically, we generate 1000 random images under H_0 and compute the score for each method to get the distribution of null scores. Then, for each method and each combination of simulation parameters, we produce a score F_{alt} and identify the proportion of null scores that are lower than F_{alt}. Essentially, we are finding the probability that a method correctly rejects the null hypothesis given that H_1 is true, at a threshold corresponding to a 5% false positive rate (ie, the 95th percentile of null scores). Therefore, detection power is computed by calculating the proportion of the 100 datasets with injected regions that score above the 95th percentile of the distribution of null scores. For HLTSS, LTSS-coarse, and LTSS-granular, the score is the maximum (over subsets S) of the EBB scoring function for discoloration, $\max_S F(S)$, as defined in Section 2.1. For JMDL, the score is the average estimated probability of being anomalous across all small-scale images.

Spatial overlap is a measure of detection accuracy. While detection power measures how well each method identifies whether or not an anomalous region is present in an image, spatial overlap measures how well a method can identify the exact anomalous region within the image. For a given combination of simulation parameters, each method identifies the anomalous region S_{detected}, and from the simulation experiment, we know the true affected region S_{affected}. Therefore, the spatial overlap is defined as

$$\text{spatial overlap} = \frac{|S_{\text{detected}} \cap S_{\text{affected}}|}{|S_{\text{detected}} \cup S_{\text{affected}}|},$$

where $|S|$ is defined as the number of pixels in the region S. The above measure of spatial overlap lies between [0,1] and is a combination of precision and recall. The higher the spatial overlap, the higher the accuracy in exactly identifying the true affected region.

3.4 | Simulation results

We compare the performance of various methods by varying both the effect size (q) and the box size of the injected anomalous region. Figure 4 shows the detection power of each method for various effect sizes and box sizes. We observe that, for a given box size, detection power typically increases with the effect size. Similarly, the ability to detect anomalous images increases with increasing box size. Figure 4 demonstrates that the HLTSS algorithm has significantly
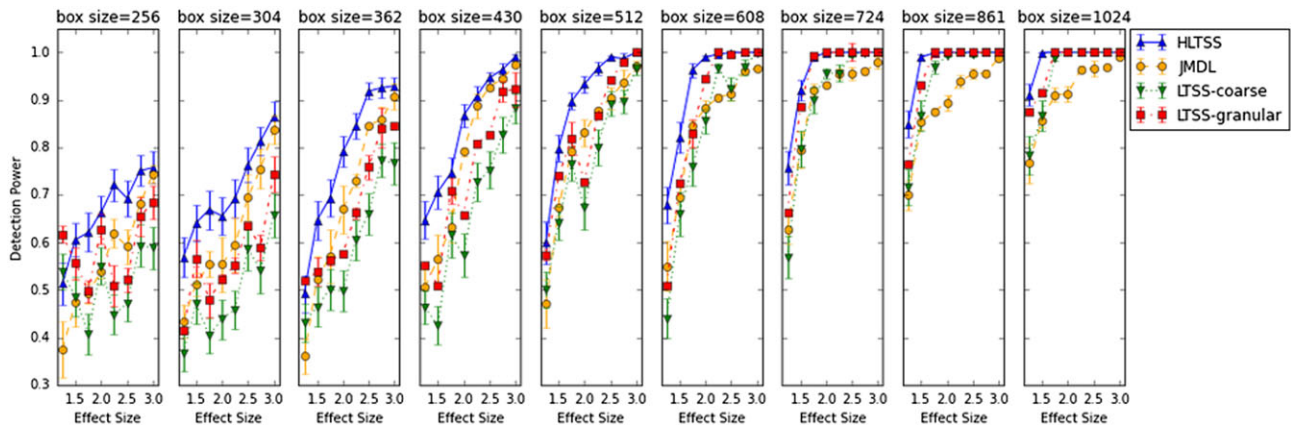
FIGURE 4 Detection power, for differentiating between anomalous images and benign images of Hierarchical Linear Time Subset Scanning (HLTSS), Janowczyk and Madabhushi Deep Learning (JMDL),[14] Linear Time Subset Scanning (LTSS)-granular, and LTSS-coarse [Colour figure can be viewed at wileyonlinelibrary.com]

higher detection power as compared with the 3 competing methods. Most importantly, HLTSS is able to perform significantly better on the most challenging injects where both effect size and box size are small.

Figure 5 shows the variation in each method's detection accuracy, as measured by spatial overlap, as a function of the simulation parameters. Again, we observe that the spatial overlap of all methods typically increases with increasing effect size and box size. The spatial overlap of the JMDL algorithm[14] is similar to our HLTSS algorithm, though HLTSS performs slightly better across all simulation parameter combinations. One reason for this could be because of the small scale images (32×32 pixels) that are given as inputs to JMDL. These small scale images could be noisy and have a higher chance of being incorrectly identified as anomalous by the method, whereas, due to the nature of top down search in our HLTSS algorithm, it is more robust to the noise in the data. Finally, the LTSS-coarse and LTSS-granular methods have very low accuracy results. LTSS-coarse performs poorly in terms of spatial accuracy because it is limited to detecting coarse-grained regions, which are rough approximations of the true injected region, while LTSS-granular incorrectly identifies many fine-grained regions that appear anomalous because of noise, leading to low precision.

4 | EMPIRICAL EVALUATION AND RESULTS

We now demonstrate the performance of our method for identifying regions of interest in digital pathology images, identifying the most anomalous groups of pixels in images of prostate biopsy samples of patients suspected to have prostate cancer. This method of identifying cancerous regions can be used to first differentiate the cancerous images from
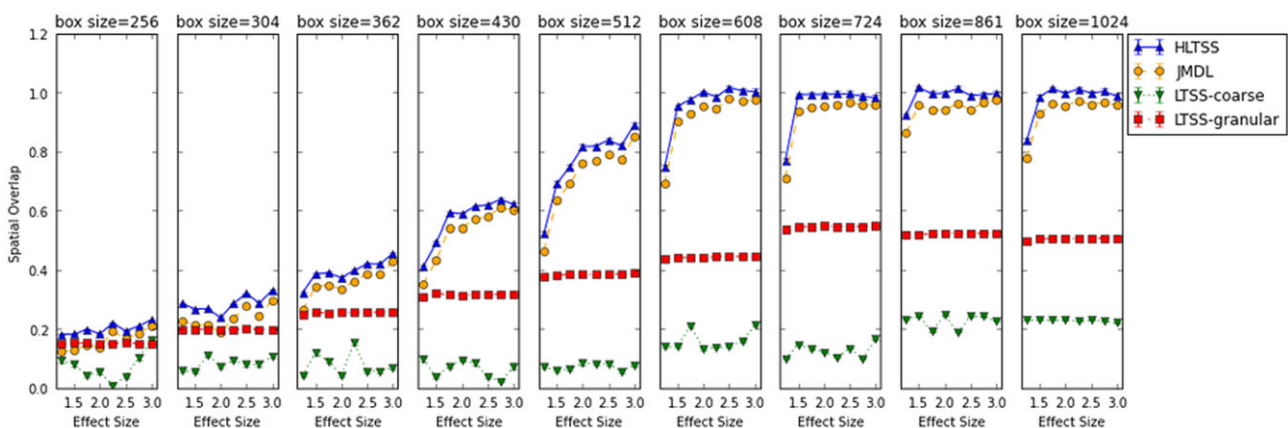

FIGURE 5 Spatial overlap, for correctly identifying the anomalous region within an image of Hierarchical Linear Time Subset Scanning (HLTSS), Janowczyk and Madabhushi Deep Learning (JMDL),[14] Linear Time Subset Scanning (LTSS)-granular, and LTSS-coarse [Colour figure can be viewed at wileyonlinelibrary.com]

benign images. Once we have identified the cancerous images, we can find regions of interest within each large-scale image, which aids a pathologist to make further decisions. As in the experimental simulations described above, we compare the performance of our HLTSS algorithm with the JMDL method described in Janowczyk and Madabhushi,[14] LTSS-coarse and LTSS-granular methods.

4.1 | Data

We have digital pathology images of prostate biopsy samples from the Department of Pathology, UPMC, including 35 images that contained cancerous locations and 55 images from benign prostate biopsy samples. The 35 cancerous images were also annotated indicating the locations of cancer. These annotations were used for evaluating the performance of our method in detecting regions of interest within a digital image. Each image contains about 5 billion pixels at the most granular level and takes up approximately 10 GB of space in uncompressed form.

4.2 | Methods and metrics

We apply our HLTSS algorithm and the competing methods to analyze the images obtained from UPMC. We evaluate the ability of each method to distinguish between benign and cancerous slides by showing the Receiver Operator Characteristic (ROC) curve. As the JMDL method requires labeled training examples, we partition the dataset into 2 halves (using the first half for training and second half for testing, and vice versa) and report the results as an average of these 2 folds. Further, we evaluate the methods' ability to pick out the malignant areas on a slide, which contains cancer by showing Precision-Recall curves. Precision provides the percentage of detected pixels that were actually cancerous, while Recall gives the percentage of cancerous pixels that were successfully detected.

We note that the ground truth regions (known to contain cancerous cells) were provided by the pathologist at a very coarse level, as shown by the yellow rectangles in Figure 6. As our HLTSS algorithm finds the affected pixels at a more granular level than the ground truth data, we draw a bounding box around each subset detected by HLTSS and consider the bounding box as our detected cluster for comparison to the ground truth. We use the same procedure for the LTSS-coarse, LTSS-granular, and JMDL algorithms.

4.3 | Experimental results

Figure 6 shows an example of our detected regions as compared with the ground truth results. The green regions are the top 10 highest-scoring regions identified using HLTSS and yellow are the coarse-level cancerous locations marked by a pathologist. We can see that HLTSS identified most of the regions with high accuracy.
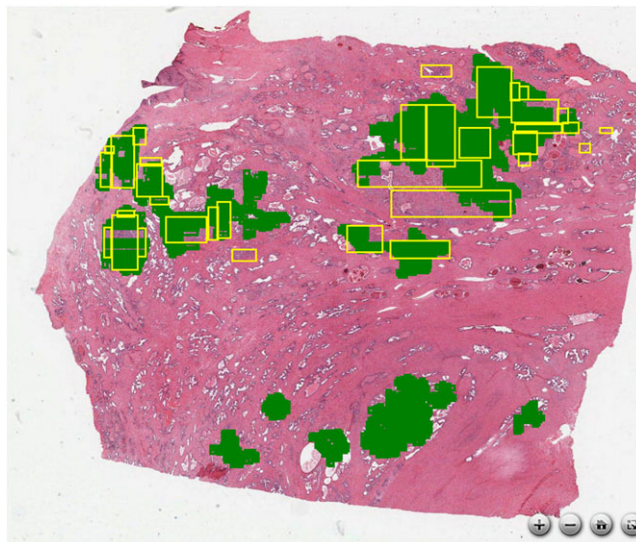

FIGURE 6 Example of regions detected by Hierarchical Linear Time Subset Scanning (green) as compared with ground truth boxes drawn by a pathologist (yellow) [Colour figure can be viewed at wileyonlinelibrary.com]

4.3.1 | Accuracy in identifying benign slides from cancerous slides

Figure 7 provides the Receiver Operating Characteristic curve for differentiating cancerous from benign images. This is very important for a pathologist as they have a huge number of slides to analyze per day and effectively identifying potentially cancerous slides helps them in prioritizing their work. Similar to the evaluation above using simulated data, for HLTSS, LTSS-coarse, and LTSS-granular, each image is scored based on the highest region score found in that image. For the JMDL method, the score is the average estimated probability of being cancerous across all small-scale images. For each method, we sort its scores in decreasing order and step through the sorted list to determine the false positive rate (fraction of noncancerous images labeled as cancerous) and true positive rate (fraction of cancerous images labeled as cancerous) corresponding to each score threshold.

From Figure 7, we can see that HLTSS has higher true positive rates for most false positive rates in comparison with the competing methods. Hierarchical Linear Time Subset Scanning has over 75% true positive rate for a 20% false positive rate, while the nonhierarchical LTSS methods perform no better than chance, suggesting that they are picking up scattered violet pixels rather than potentially relevant clusters.

4.3.2 | Accuracy in identifying cancerous regions

We use the spatial precision-recall curve to evaluate and compare the accuracy of detected regions within a cancerous image. To calculate this metric, we record the list of scores returned for each pixel in an image. For HLTSS, the score of each pixel is the score of the highest scoring detected group of which it is a member. Any pixels that are not detected in any group are scored individually using the EBB score function and ranked below all detected pixels. For JMDL, the score of each pixel is the estimated probability that its corresponding 32×32 image is cancerous. We sort the list of scores and use each unique score value as a threshold for classifying locations and calculate the precision (ratio of correctly identified cancerous pixels to total number of pixels in the detected region) and the recall (ratio of correctly identified cancerous pixels to total number of cancerous pixels) at each threshold. These precision and recall values are plotted on the curve for each image and are averaged over all slides to obtain a single curve for each method with standard errors shown in Figure 8.

From Figure 8, we see that HLTSS maintains high precision while capturing most of the cancerous locations in an image. The precision of JMDL degrades substantially for high recall values, while nonhierarchical LTSS methods incorrectly flag many small violet areas as cancerous, resulting in low precision. Hence, our results show that HLTSS can not only differentiate cancerous from benign images, but also accurately identifies regions of interest within a cancerous image.
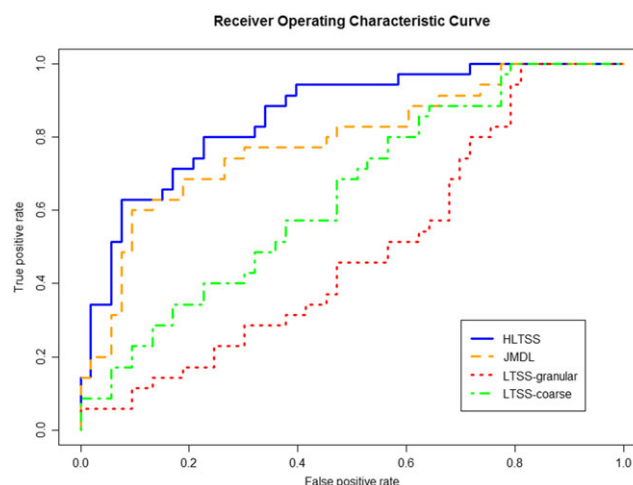

FIGURE 7 ROC curves, for differentiating cancerous images from benign images of Hierarchical Linear Time Subset Scanning (HLTSS), Janowczyk and Madabhushi Deep Learning (JMDL),[14] Linear Time Subset Scanning (LTSS)-granular, and LTSS-coarse [Colour figure can be viewed at wileyonlinelibrary.com]
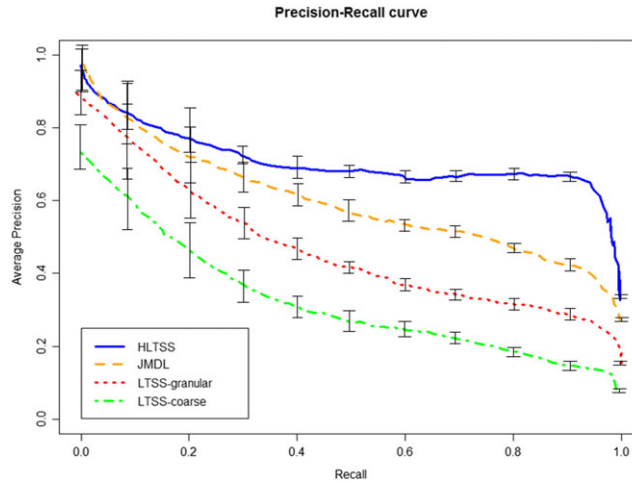
FIGURE 8 Precision and Recall curves, for finding cancerous regions within a cancerous image of Hierarchical Linear Time Subset Scanning (HLTSS), Janowczyk and Madabhushi Deep Learning (JMDL),[14] Linear Time Subset Scanning (LTSS)-granular, and LTSS-coarse [Colour figure can be viewed at wileyonlinelibrary.com]

4.3.3 | Comparison of execution times

We also compared the execution times of the various methods. We executed each method on a quad-core Intel(R) Xeon(R) CPU 5160 @ 3.00 GHz machine. To be fair in comparing the execution times, we did not use any Graphics Processing Unit (GPU) components for the execution of the JMDL algorithm.[14] Table 1 provides the comparison, including

TABLE 1 Comparison of average execution times (in minutes) of various methods

Method	Pre/Post Processing Time	Algorithm Time	Total Time
HLTSS	10.23	34.49	44.72
JMDL[14]	13.57	21.34	34.91
LTSS-granular	17.12	9.43	26.55
LTSS-coarse	2.45	3.38	5.83

Abbreviations: HLTSS, Hierarchical Linear Time Subset Scanning; JMDL, Janowczyk and Madabhushi Deep Learning; LTSS, Linear Time Subset Scanning
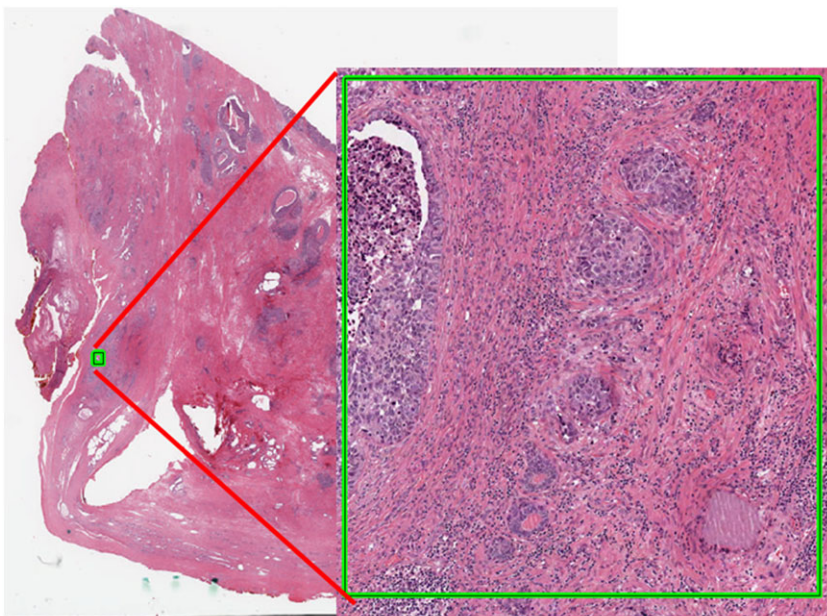

FIGURE 9 Case study: Hierarchical Linear Time Subset Scanning successfully identified prostate cancer locations on this digital pathology image [Colour figure can be viewed at wileyonlinelibrary.com]

both the pre/post processing times and the algorithm execution times. There is a substantial preprocessing time for JMDL in partitioning the image into smaller scale 32×32 pixel images and storing these images in the database, required by the algorithm. Further, our HLTSS algorithm requires postprocessing to identify all the connected regions detected by the algorithm. We can see that the execution times are very similar: HLTSS is slightly slower than JMDL, while LTSS-coarse and LTSS-granular are faster but have substantially lower accuracy than HLTSS.

4.4 | Case studies

Given the digital images of prostate biopsy samples from UPMC, we show several examples of regions of interest that were detected automatically by HLTSS and then further investigated by a pathologist.

In our first case study, 2 of the top-scoring regions detected by HLTSS are shown in Figures 9 and 10, respectively. Follow-up analysis (by Dr. Anil Parwani) of the region of interest in Figure 9 confirmed that HLTSS was successful in detecting prostate cancer. The region of interest in Figure 10 contained not only prostate cancer but also bladder cancer,
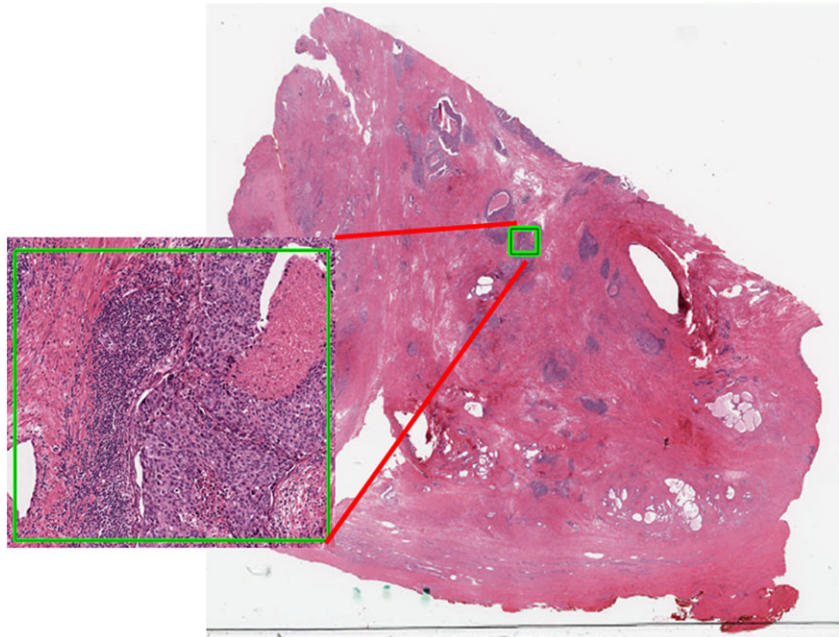

FIGURE 10 Case study: Hierarchical Linear Time Subset Scanning identified bladder cancer which has spread to the prostate gland. This bladder cancer has color characteristics similar to prostate cancer [Colour figure can be viewed at wileyonlinelibrary.com]
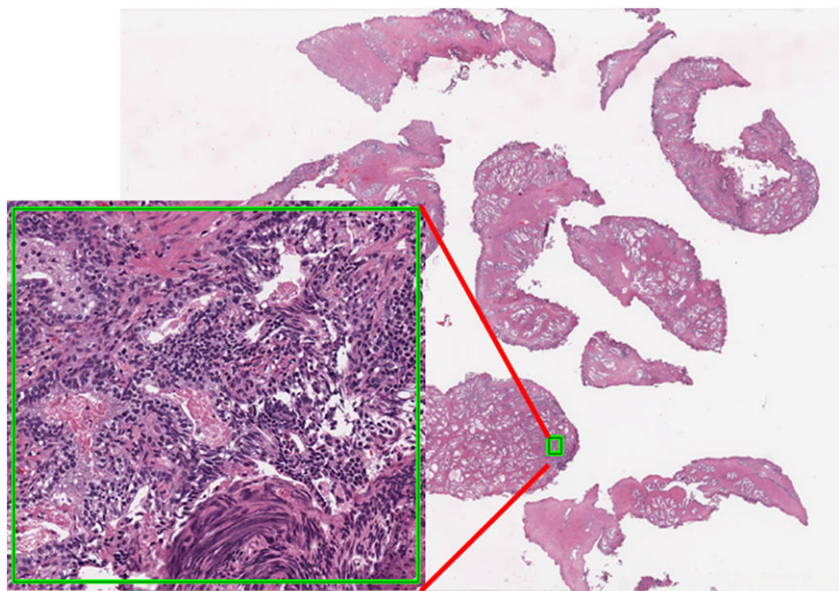

FIGURE 11 Case study: Hierarchical Linear Time Subset Scanning identified a region of interest for pathologists. In this case, the location was a benign mimic of prostate cancer and has color characteristics similar to prostate cancer [Colour figure can be viewed at wileyonlinelibrary.com]

which had spread to the prostate gland. The bladder cancer cells in the prostate gland exhibit color characteristics similar to prostate cancer, and hence our method detected this location as a region of high interest.

Our second case study is also of a prostate biopsy sample, and one of the highest-scoring detected regions of interest is shown in Figure 11. This region is interesting for a pathologist, as it is a benign mimic of cancer. Though it does not contain cancerous cells, it has all the color characteristics of prostate cancer, and it is difficult for a human pathologist to differentiate it from cancer using the shape characteristics. We believe that providing such regions of interest is important, allowing a well-trained pathologist to make final decisions on each of these regions and to report all relevant findings (both benign and malignant). Also, these regions could be used as examples for educational purposes in training pathology residents to make subtle distinctions between cancerous regions and benign mimics.

5 | CONCLUSIONS

In this work, we proposed a novel HLTSS framework for detecting regions of interest in massive, multiscale digital pathology slides. Such images typically consist of 10 billion pixels or more at the finest resolution but are stored as multiple layers, each representing a hierarchical aggregation of data from the previous layer. Hierarchical Linear Time Subset Scanning exploits this hierarchical structure inherent in data produced through virtual microscopy in order to accurately and quickly identify regions of interest for pathologists to review. We evaluated the performance of HLTSS for identifying cancerous locations on digital slides of prostate biopsy samples and demonstrated that our methods detect regions of cancer in minutes with high accuracy, as measured by the ROC curve (measuring ability to distinguish between benign and cancerous slides) and the spatial precision-recall curve (measuring ability to pick out the malignant areas on a slide that contains cancer).

Identifying regions of interest in pathology slides based on color characteristics may be useful in many other applications, such as identifying regions of inflammation in the gastrointestinal tracts for Crohn's disease; finding gastritis in the stomach, which may be indicative of colonization by *Helicobacter pylori*; and diagnosing prostatic intraepithelial neoplasia, which may lead to prostate cancer. Other cases where the shape characteristics of cells are more important than their color characteristics would not be appropriate for this method and might be better addressed using existing image processing methods. Furthermore, we would like to apply our fast detection techniques to other domains, which rely on massive image data, such as analyzing multiresolution satellite imagery to automatically detect forest fires.

ORCID

Sriram Somanchi http://orcid.org/0000-0002-3153-1248

REFERENCES

1. Mettlin C, Murphy GP, Lee F, et al. Characteristics of prostate cancers detected in a multimodality early detection program. The investigators of the american cancer society-national prostate cancer detection project. *Cancer*. 2013;72(1):1701-1708.

2. Romero Lauro G, Cable W, Lesniak A, et al. Digital pathology consultations-a new era in digital imaging, challenges and practical applications. *J Digital Imaging*. 2013;26(4):668-677.

3. Neill DB. Fast subset scan for spatial pattern detection. *J R Stat Soc (Ser B: Stat Method)*. 2012;74(2):337-360.

4. Neill DB, McFowland E, Zheng H. Fast subset scan for multivariate event detection. *Stat Med*. 2013;32(13):2185-2208.

5. Neill DB. Expectation-based scan statistics for monitoring spatial time series data. *Int J Forecast*. 2009;25:498-517.

6. Parwani AV, Park S, Pantanowitz L. Digital imaging in pathology. *Clin Lab Med*. 2012;32(4):557-584.

7. Amer. Cancer Soc. Cancer Facts and Figures; 2013.

8. Parwani A. http://pathologyconference.upmc.edu/2013/0418b2013/0418b2013.html; 2013.

9. Tahir MA, Bouridane A. Novel round-robin tabu search algorithm for prostate cancer classification and diagnosis using multispectral imagery. *IEEE Trans Inf Technol Biomed*. 2006;10(4):782-793.

10. Bouatmane S, Roula MA, Bouridane A, Al-Maadeed S. Round-robin sequential forward selection algorithm for prostate cancer classification and diagnosis using multispectral imagery. *Mach Vision Appl*. 2011;22(5):865-878.

11. Peng Y, Jiang Y, Chuang S-T, Yang XJ. Computer-aided detection of prostate cancer on tissue sections. *Appl Immunohistochem Mol Morphol*. 2009;17(2):442-450.

12. Tabesh A, Teverovskiy M, Pang HY, et al. Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Trans Med Imaging*. 2007;26(10):1366-1378.

13. Eberlin LS, Tibshirani RJ, Zhang J, et al. Molecular assessment of surgical-resection margins of gastric cancer by mass-spectrometric imaging. *Proc Nat Acad Sci*. 2014;111(7):2436-2441.

14. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *J Pathol Inf*. 2016;7(1):29.

15. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol*. 2001;23(4):291-299.

16. ImageJ. Imagej software. http://rsb.info.nih.gov/ij/; 2004.

17. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Stat Med*. 1995;14:799-810.

18. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. Lake Tahoe, Nevada; 2012:1097-1105.