Knowledge and Information Systems
This is authors' *accepted* version. Publisher's version is available at
http://link.springer.com/article/10.1007/s10115-016-0963-7.

Active Inference for Dynamic Bayesian Networks with an Application to Tissue Engineering

Caner Komurlu · Jinjian Shao · Banu Akar · Elif S. Bayrak · Eric M. Brey · Ali Cinar · Mustafa Bilgic

Received: Jun 04, 2015 / Revised: Apr 20, 2016 / Accepted: May 14, 2016

Abstract In temporal domains, agents need to actively gather information to make more informed decisions about both the present and the future. When such a domain is modeled as a temporal graphical model, what the agent observes can be incorporated into the model by setting the respective random variables as evidence. Motivated by a tissue engineering application where the experimenter needs to decide how early a laboratory experiment can be stopped so that its possible future outcomes can be predicted within an acceptable uncertainty, we first present a dynamic Bayesian network (DBN) model of vascularization in engineered tissues and compare it with both real-world experimental data and agent-based simulations. We then formulate the question of "how early an experiment can be stopped to guarantee an acceptable uncertainty about the final expected outcome" as an *active inference* problem for DBNs and empirically and analytically evaluate several search algorithms that aim to find the ideal time to stop a tissue engineering laboratory experiment.

Keywords Active Inference · Dynamic Bayesian Networks · Tissue Engineering

Caner Komurlu[1]
E-mail: ckomurlu@hawk.iit.edu

Jinjian Shao[1]
E-mail: jshao3@hawk.iit.edu

Banu Akar[2]
E-mail: bakar@hawk.iit.edu

Elif S. Bayrak[3]
E-mail: ebayrak@amgen.com

Eric M. Brey[2]
E-mail: brey@iit.edu

Ali Cinar[3]
E-mail: cinar@iit.edu

Mustafa Bilgic[1]
E-mail: mbilgic@iit.edu

[1] Department of Computer Science
[2] Department of Biomedical Engineering
[3] Department of Chemical and Biological Engineering
Illinois Institute of Technology
Chicago, IL, USA

1 Introduction

People lose tissue due to injuries, treatments, and illnesses. The body can heal itself when the tissue loss is minor. When the tissue loss is severe, however, the body cannot completely heal itself on its own and a potential solution is to replace the lost tissue by artificially engineering it by seeding cells and biomaterials [28]. The formation of tissue needs vascularization to support cells by bringing nutrients and oxygen to them and collecting debris from them, and a scaffold that provides structural and biochemical support to the newly formed tissue. Tissue engineering is an active research area where researchers conduct in vitro (outside of a living organism) and in vivo (within a living organism) experiments to gain an understanding of healthy tissue growth and vascularization.

A crucial component of tissue engineering is proper formation of blood vessels (vascularization) that support life of the cells. In this article, we introduce a dynamic Bayesian network (DBN) model of vascularization in engineered tissues. The DBN enables tissue engineering researchers to perform spatio-temporal reasoning about tissue formation process and allows them to try various experimental settings on computer first, before trying the most promising settings in the lab. Furthermore, DBNs enable researchers to "inject" information into the reasoning process by providing values of some of the random variables as evidence. This is especially helpful when the researchers analyze the lab experiment as it progresses and provide the latest status of the experiment as evidence to the DBN, enabling more accurate predictions for future time steps. The ability to inject information into the system is also helpful for what-if analyses such as "what if I introduce more nutrients into the tissue right now?"

The laboratory experiments require expensive materials and they take time, usually weeks to months. Therefore, tissue engineers would like to i) conduct only the most promising experiments to save cost and time, and ii) stop an experiment early to save time and make predictions as to how it would continue hadn't it been stopped. When an in-vivo laboratory experiment is stopped to collect data, usually a lab animal is sacrificed and the tissue is dissected for analysis; hence a stopped experiment cannot resume where it left off. Typically, many animals are used in an experiment and each animal is sacrificed at a specific time during the course of the experiment to analyze the status of an experiment at that particular time.

Researchers need to determine the ideal time to stop an experiment: stopping too early makes the uncertainty in predictions about the future progress of the tissue unacceptably large whereas stopping the experiment later than necessary costs valuable time and resources. We formulate this question as an *active inference* problem [8] for DBNs, where limited information about the statuses of a subset of the random variables can be gathered at inference time to decrease prediction uncertainty for future time slices. The decision question is then when is the ideal time, t^* to stop an experiment and gather information to guarantee an acceptable uncertainty for a target time slice in the future. The algorithmic question is how to compute this ideal time t^* with minimum computational cost possible. We tackle both questions in this article.

In collaboration with tissue engineering researchers, we developed the first prototype of our DBN model of vascularization in [26]. In this article, we build on this prototype and we significantly expand it. Our *additional* contributions in this article are:

- We present an improved and more realistic DBN model of vascularization.
- We provide real-world experimental results [1] supporting the DBN predictions.
- We compare our DBN model to the agent-based model of Artel et al [2].
- We formulate and present results for active inference for DBNs.
- We perform an analytical evaluation, providing closed-form solutions whenever possible, for the computational cost of several search algorithms applied to active inference.

The main contributions of this paper revolve around the tissue engineering application. However, the active inference formalism we provide for DBNs in this study is general enough that it is potentially applicable to other

real-world temporal domains. We discuss in Section 6 the necessary conditions under which the formulations and the solutions in this article could be utilized for those domains.

The rest of the article is organized as follows: we first provide background information on tissue engineering and active inference in Section 2. Then, we detail our DBN and active inference approach in Section 3, followed by a discussion of our experimental findings in Section 4. Next, we analytically evaluate several search algorithms for active inference in Section 5. Finally, we discuss the limitations of our work and potential applications to other domains in Section 6, related work in Section 7,and then conclude in Section 8.

2 Background and Motivation

In this section, we provide a brief background on the tissue engineering process and active inference.

2.1 Tissue Engineering

Although the human body has a great capability to heal, sometimes the tissue loss is so severe that the body cannot heal completely. For such cases, the missing tissue is replaced with autologous tissue or donor tissue. These approaches are limited, and tissue engineering offers a potential alternative of growing new tissue using cells on biomaterial scaffolds[6, 30, 31]. The process starts with a small number of cells seeded into a biomaterial scaffold, the structure that holds the newly formed tissue together. If stem cells are used, they differentiate into specific tissue cells and the tissue cells multiply to fill the space. At the same time, blood vessels from surrounding tissue grow into the scaffold, carrying oxygen and nutrients to the cells and removing carbon-dioxide and other waste from them. The scaffold that holds the tissue cells degrades over time, allowing the cellular matrix to develop with the blood vessels to generate a fully functional tissue.

In this article, we focus on the vascularization aspect of the tissue engineering process, which is illustrated in Figure 1. When a tissue cell is far from the service range of a blood vessel, it cannot receive the oxygen and nutrients it needs; hence, it enters into a "distressed" phase, called *hypoxia*. Such a cell signals its distress by releasing solubale chemical signals such as vascular endothelial growth factor (VEGF). VEGF diffuses into the region and upon binding the outer layer of a blood vessel, this chemical initiates the process of sprouting a new blood vessel. The tip of the new blood vessel sprout, called the `Tip Cell`, grows stochastically in the direction of the VEGF. `Stalk Cells` proliferate behind the `Tip Cell` forming a new blood vessel as it elongates. When two different blood vessel sprouts meet, they connect (*anastomosis*) and the blood circulation starts through the newly formed blood vessel loop.

The vascularization process is both a temporal and stochastic process. Over time, the tip of the blood vessel elongates stochastically in the direction of the VEGF gradient, searching for the distressed cell. The direction that the blood vessels grow through is stochastic because i) the blood vessels tend to do some exploration in their search for the distressed cell, ii) scaffold structures can sterically hinder movement of the cells up the gradient, and iii) there are other unaccounted and unknown factors besides VEGF that affect vascularization. Because the process is both temporal and stochastic, we model the process using a dynamic Bayesian network, which we explain in detail in Section 3.1.

2.2 Active Inference

As was discussed in the introduction section, tissue engineering researchers conduct costly (time and material) experiments in the lab and because experiments take a long time researchers often need to stop the experiments

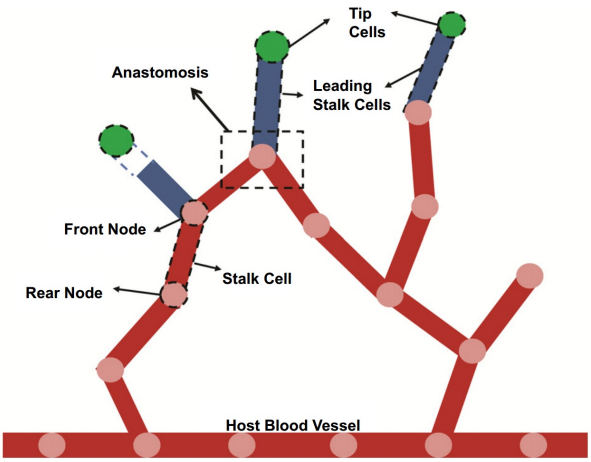


Fig. 1 Illustration of vascularization, including `Tip Cell`, `Stalk Cell`, and anastomosis. Figure courtesy of Mehdizadeh et al [33].

early and make predictions about their future state. Once an experiment is stopped, because the tissue is dissected, the experiment cannot resume and hence researchers need to determine the ideal time to stop an experiment so that reliable predictions about its future states can be made. We formulate this question through active inference for dynamic Bayesian networks.

Active inference [8] deals with the problem of selective gathering of information for some of the variables of a model with the objective to improve prediction for the remaining variables. The main task is to gather more information at inference time to increase the predictive performance of the underlying model. The underlying models tend to be graphical models or relational models where observing the values of a subset of random variables helps with the prediction for the remaining variables. For example, Chen et al [15] manually analyze a few short segments of a video and let the underlying model condition on the observed information to improve prediction on the remaining segments of the video.

In this article, information on the initial settings of an experiment is provided by the tissue engineering researchers; hence, the random variables for the initial time, $t = 0$, are observed. Stopping an experiment at time t and dissecting the tissue is equivalent to observing the values of the random variables at time t. The researchers would like to predict the status of the experiment at a future time T. Hence, the active inference problem here is how early can we stop the experiment (i.e., what is the smallest value of t) so that the uncertainty of our prediction about time T is acceptable, i.e., less than a provided threshold?

Active inference has been discussed in the context of iterative classification algorithm [9], pairwise Markov Random Fields [8], and Hidden Markov Models [27, 14]. Motivated by the real-world application of tissue engineering, we formulate and tackle the active inference problem for dynamic Bayesian networks in this article. We next introduce our DBN model of vascularization and then present the active inference problem more formally.

3 Modeling Approach and Problem Formulation

In this section, we first describe our DBN model for vascularization and then we discuss the active inference formulation for DBNs.

3.1 DBN Modeling of Vascularization in Tissue Growth

In a typical tissue engineering laboratory experiment, there are many factors considered: the shape and porosity of the scaffold, seed cells that are in various cycles of their lives, and the distribution of the VEGF that is released

by distressed cells, etc. In this article, we focus on the following question: given a group of distressed cells that signal their distress through VEGF, which we simply refer to as the *VEGF source*s (VS), and the closest host blood vessel, how likely are the VSs to be reached by new blood vessels that sprout from the host blood vessel?

In this article, we zoom in the region that is not vascularized yet. Hence, given one or more VSs and a host blood vessel, we model the vascularization using random variables that represent each location in a tissue grid as `Empty` (`E`), the tip of a new blood vessel `Tip Cell` (`TC`), or the body of the blood vessel `Stalk Cell` (`SC`). We manually construct the DBN structure basing it on our experimental and simulation expertise and evidence. We experiment with numerous parameter settings allowing us to try various real and hypothetical scenarios. We next describe the structure and the parameters for the DBN.

3.1.1 The DBN Structure

In vascularization, at each step in time, the tip of a blood vessel elongates stochastically in the direction of the gradient of the VEGF, forming the body of the blood vessel. This corresponds to the following transition in our DBN: a location that is `TC` at time t becomes `SC` at time $t+1$ forming the body of the blood vessel, and the location in the direction the blood vessel elongation becomes the new `TC`. Given a 2D grid of the space, following the assumptions made in [33], we assume that the VSs are located at north whereas the host blood vessel is located in south, and hence the blood vessel elongates towards north-east (NE), north (N), or north-west (NW). Figure 2 illustrates the tissue grid, the dependencies between the random variables, and the corresponding 2-slice DBN.

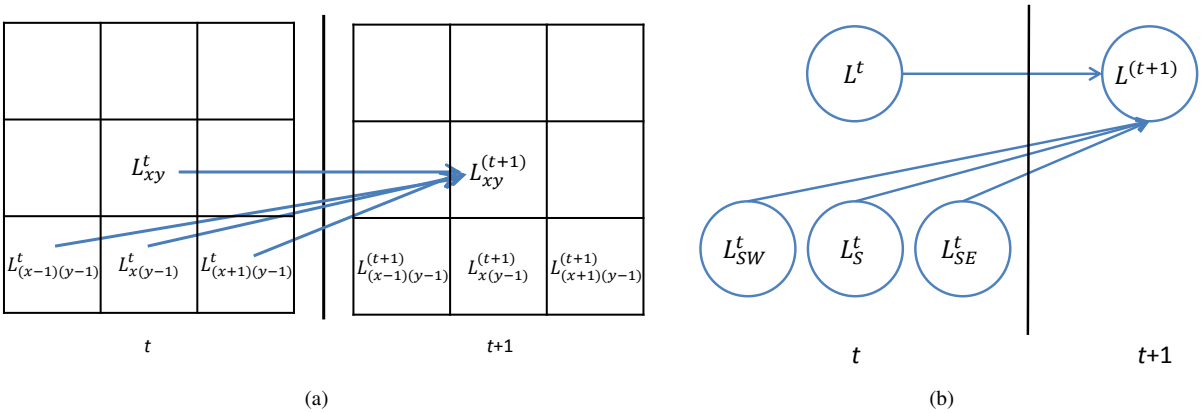


Fig. 2 a) The tissue grid. The status of a location at time $t+1$ depends on the previous time t statuses of itself and its neighbors at its south-west, south, and south-east. b) The corresponding 2-slice DBN.

In this DBN, the status of a location at time $t+1$, L^{t+1}_{xy}, depends on the statuses of itself, L^t_{xy}, and the neighbors at its south-west (SW), $L^t_{(x-1)(y-1)}$, south (S), $L^t_{x(y-1)}$, and south-east (SE), $L^t_{(x+1)(y-1)}$, at previous time t. We simply refer to these neighbors generically as L^t_{SW}, L^t_S and L^t_{SE}. We next describe the parameter settings, i.e., the conditional probability densities (CPDs), for this DBN.

3.1.2 The DBN Parameters

The typical vascularization process is as follows: a `Tip Cell` elongates stochastically in the direction of the VEGF source, VS, occupying an `Empty` location and forming the body of the blood vessel (`Stalk Cell`)

during the process. Upon touched by VEGF, the body of the blood vessel might sprout a new blood vessel, i.e., a `Stalk Cell` might turn into a `Tip Cell`. Hence, if a location is

- a `Tip Cell` at time t, it elongates at time $t+1$ to an `Empty` location at one of its NW, N, or NE location, and leaves a `Stalk Cell` behind.
- a `Stalk Cell` at time t, it either remains a `Stalk Cell` or turns to a new `Tip Cell` to sprout a new blood vessel at time $t+1$.
- `Empty` at time t, whether it remains `Empty` or gets occupied by `Tip Cell` at time $t+1$ depends on whether there was a `Tip Cell` that can elongate to this location from at least one of its SW, S, and SE locations at time t.

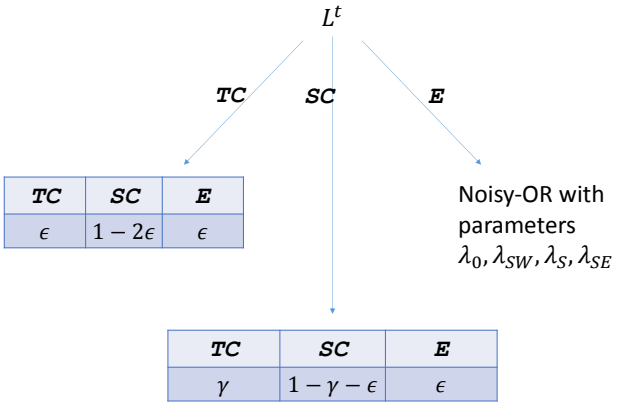


Fig. 3 The tree CPD representation for $P(L^{(t+1)}|L^t, L^t_{SW}, L^t_S, L^t_{SE})$.

These transitions are best represented through a tree-CPD, which is shown in Figure 3. In this representation, we present the parameters in the order of $\langle \texttt{TC}, \texttt{SC}, \texttt{E} \rangle$. More formally, the probability of $L^{(t+1)}$ being `TC`, `SC` or `E`, given $L^{(t)} = \texttt{TC}$ is:

- $P(L^{(t+1)} = \texttt{TC}|L^t = \texttt{TC}, L^t_{SW}, L^t_S, L^t_{SE}) = P(L^{(t+1)} = \texttt{TC}|L^t = \texttt{TC}) = \epsilon$
- $P(L^{(t+1)} = \texttt{SC}|L^t = \texttt{TC}, L^t_{SW}, L^t_S, L^t_{SE}) = P(L^{(t+1)} = \texttt{SC}|L^t = \texttt{TC}) = 1 - 2\epsilon$
- $P(L^{(t+1)} = \texttt{E}|L^t = \texttt{TC}, L^t_{SW}, L^t_S, L^t_{SE}) = P(L^{(t+1)} = \texttt{E}|L^t = \texttt{TC}) = \epsilon$

That is, a `TC` at time t is most likely turn into a `SC` at time $t+1$. Likewise, the probability of $L^{(t+1)}$ being `TC`, `SC` or `E`, given $L^{(t)} = \texttt{SC}$ is:

- $P(L^{(t+1)} = \texttt{TC}|L^t = \texttt{SC}, L^t_{SW}, L^t_S, L^t_{SE}) = P(L^{(t+1)} = \texttt{TC}|L^t = \texttt{SC}) = \gamma$
- $P(L^{(t+1)} = \texttt{SC}|L^t = \texttt{SC}, L^t_{SW}, L^t_S, L^t_{SE}) = P(L^{(t+1)} = \texttt{SC}|L^t = \texttt{SC}) = 1 - \gamma - \epsilon$
- $P(L^{(t+1)} = \texttt{E}|L^t = \texttt{SC}, L^t_{SW}, L^t_S, L^t_{SE}) = P(L^{(t+1)} = \texttt{E}|L^t = \texttt{SC}) = \epsilon$

That is, a `SC` at time t can become active with probability γ at time $t+1$, can stay as `SC` with probability $1-\gamma-\epsilon$, or can become `E` with a noise parameter of ϵ.

Finally, probabilities of $L^{(t+1)}$ being `TC`, `SC` or `E`, given $L^{(t)} = \texttt{SC}$ are computed with a noisy-OR function, and values are given in the compact form $\langle \texttt{TC}, \texttt{SC}, \texttt{E} \rangle$ Probabilities of $L^{(t+1)}$ given $L^{(t)} = \texttt{E}$:

- $P(L^{(t+1)}|L^t = \texttt{E}, L^t_{SW}, L^t_S, L^t_{SE}) = \text{noisy-OR}(L^t_{SW}, L^t_S, L^t_{SE})$

Whether an E location at time t gets occupied with the tip of a blood vessel (TC) at time $t+1$ is modeled as a noisy-OR of its parents located at its SW, S, and SE neighbors: $P(L^{(t+1)} = \texttt{TC} | L^t = \texttt{E}, L^t_{SW}, L^t_S, L^t_{SE})$ is a noisy-OR of L^t_{SW}, L^t_S, L^t_{SE}, with parameters λ_0, λ_{SW}, λ_S and λ_{SE}. λ_0 is a leak parameter, and λ_{SW}, λ_S and λ_{SE} represent the possibility that a TC in the SW, S, or SE will elongate to this location.

Noisy-OR is probabilistic generalization of logical OR, where an effect can be caused by the presence of one or several of its causes [24]. For example, fever can be caused by several diseases, such as cold, flu, infection, etc. Moreover, the effect can occur without any explanation as well, which is called the leak. Typically, the probability of false is computed, where none of the causes are present and probability of true is simply calculated by subtracting it from 1. That is, $P(Effect = false) = (1-\lambda_0)\prod(1-\lambda_i)^{C_i}$, where C_i are boolean variables representing whether cause i is present and λ_i is the probability that cause i can cause the effect independently, and λ_0 is the probability that the effect can happen without any accounted factors, i.e., the leak parameter. This Noisy-OR approach fits our case because the probability that an empty location at time t is occupied with a TC is a noisy-OR of any one of its S, SE, and SW TC neighbors elongating to this location.

In our model, the magnitude of λ_{SW}, λ_S and λ_{SE} are determined by i) how far the VEGF can travel before it completely dissipates and ii) the magnitude of the VEGF gradient. If the VEGF cannot reach a TC, then TC follows a path uniformly at random, exploring its surroundings. If TC is reached by VEGF, then the closer the TC to the source of VEGF , VS, the higher the VEGF gradient, and hence the λ values become more skewed towards the VS. The further away from the VS, the more uniform the λ values get.

More formally, let d_{xy} be the distance of location L_{xy} to VS and let d_V represent the maximum distance that the VEGF can travel before it dissipates (or becomes negligible). If $d_{xy} > d_V$, then because VEGF cannot reach L_{xy}, the tip cell at L_{xy} has no clue as to whether it should grow towards NW, N, or NE, and hence it has equal probability in either direction: $\lambda_{(x-1)(y+1)} = \lambda_{(x)(y+1)} = \lambda_{(x+1)(y+1)} = \frac{1}{3}$. If $d_{xy} \leq d_V$, then the λ values become skewed towards VS. Without loss of generality, assume that the VS is located at NE with respect to xy. Then $\lambda_{(x-1)(y+1)} = \lambda_{(x)(y+1)} = \frac{d_{xy}}{3d_V}, \lambda_{(x+1)(y+1)} = \frac{d_{xy}}{3d_V} + (1 - \frac{d_{xy}}{d_V})$, and hence the TC has a higher probability moving NE.

It is important to note that the λ formalism discussed above is only an approximation to the reality. The VEGF gradient is one of many factors that affect the direction of the blood vessel growth. Hence, the probabilistic framework allows researchers to "sweep the unaccounted and uncontrolled factors under the probability rug." Additionally, how far the VEGF travels depends on a number of factors including the scaffold porosity. Therefore, we do not set d_V parameter to a fixed value; rather, we experiment with various hypothetical values to see how it effects vascularization.

3.2 Active Inference for DBNs

3.2.1 Motivation and Objective Function

In this section, we formalize the question of "*given an uncertainty threshold of σ, how early can a tissue engineering experiment be stopped so that the prediction uncertainty over the tissue grid for a target time slice T is below σ*?" More formally, we need to find t^* such that:

$$t^* = \underset{t<T}{\operatorname{argmin}} \, UNC\left(P\left(\mathcal{L}^T | l^0, l^t\right)\right) < \sigma$$

where $UNC\left(P\left(\mathcal{L}^T | l^0, l^t\right)\right)$ is the prediction uncertainty over the target time slice T, σ is the maximum uncertainty acceptable to the researcher, l^t is the observation (i.e., the values of all the random variables L^t_{xy} at time t) we would get once we stop the lab experiment and dissect the tissue, and $\mathcal{L}^T$ is the set of all locations at time

T. Of course, we do not know what the status of the experiment would be at time t unless we actually stop the experiment. Therefore, a standard technique is to take an expectation over all possible outcomes l^t:

$$t^* = \underset{t<T}{\operatorname{argmin}} \sum_{l^t} P\left(\mathcal{L}^t = l^t | l^0\right) UNC\left(P\left(\mathcal{L}^T | l^0, l^t\right)\right) < \sigma$$

A challenge here is, however, that if the number of all locations at time t is m, then l^t ranges over 3^m possible assignments (each location can be TC, SC, or E), which makes taking the expectation clearly intractable. There are two viable approximations that are common in the literature. One is to sample potential assignment values using the probability distribution $P\left(\mathcal{L}^t | l^0\right)$ and take the mean of $UNC\left(P\left(\mathcal{L}^T | l^0, l^t\right)\right)$. A second approach is to use l^t where $P\left(\mathcal{L}^t | l^0\right)$ is maximum (i.e., the MAP assignment). Note that finding l^t where $P\left(\mathcal{L}^t | l^0\right)$ is maximum does not necessarily require us to enumerate all possible l^t values, when i) each location at time t is independent of other locations at the same time t given their parents, and ii) there are no future, $t' > t$, observations. Both assumptions hold for our DBN structure.

3.2.2 Uncertainty Definition

There are a number of possible approaches for defining uncertainty over the target time slice, $UNC\left(P\left(\mathcal{L}^T | l^0, l^t\right)\right)$. One typical approach is to use entropy. However, entropy cannot capture correctly the cases that are misclassified with high certainty. For example, a model that predicts all locations as empty with high uncertainty would have low entropy but high error. In our domain, the researchers are interested in predicting the path that a new sprout from the host blood vessel follows and hence we formulate uncertainty as the conditional error of the most probable blood vessel path that originates from the host blood vessel.

More formally, let $p = \langle p_1, p_2, ..., p_{|path|} \rangle$ be a path of length $|path|$ that represents a connected sequence of locations representing a blood vessel, whose tip $p_{|path|}$ is a TC and whose body $\langle p_1, p_2, ..., p_{|path-1|} \rangle$ are SCs. Then, the conditional error over this path p being a blood vessel is:

$$UNC(p | l^t, l^0) = \left(\sum_{i=1}^{|path|-1} \left(1 - P(p_i = \texttt{SC} | l^t, l^0)\right) \right) + \left(1 - P(p_{|path|} = \texttt{TC} | l^t, l^0)\right)$$

The uncertainty over the target time locations $\mathcal{L}^T$ is then defined as the conditional error of the most probable blood vessel path at time T:

$$UNC\left(P\left(\mathcal{L}^T | l^0, l^t\right)\right) = \min_{p \in \mathcal{L}^T} UNC(p | l^t, l^0) \tag{1}$$

3.2.3 Search

To find the optimal t^* to stop a laboratory experiment so that uncertainty for target time T is below threshold σ, we need to search for the $0 < t < T$ for which $UNC\left(P\left(\mathcal{L}^T | l^0, l^t\right)\right) < \sigma < UNC\left(P\left(\mathcal{L}^T | l^0, l^{t-1}\right)\right)$. However, such a search can be computationally expensive. For each t value we try, we need to:

- *Step 1:* Find the most-likely assignment l^t for $\mathcal{L}^t$, i.e., find $\underset{l^t}{\operatorname{argmax}} P(\mathcal{L}^t = l^t | l^0)$. Given the Bayesian network structure, this requires us to run inference from only from $t' = 0$ up to $t' = t$; any variables for $t' > t$ are irrelevant. Moreover, given two times $t_1 < t_2$, the computations done for $\underset{l^{t_1}}{\operatorname{argmax}} P(\mathcal{L}^{t_1} = l^{t_1} | l^0)$ can be reused for computing $\underset{l^{t_2}}{\operatorname{argmax}} P(\mathcal{L}^{t_2} = l^{t_2} | l^0)$. For example, if an approximate inference technique

such as sampling is used for inference, the counts can be stored and reused for all $0 \leq t \leq T$. If variable elimination is used, respecting the temporal order of the variables, the computations for eliminating variables need not to be repeated. Therefore, this step can be done once for the whole search process.

– *Step 2:* Compute $UNC\left(P\left(\mathcal{L}^T|l^0,l^t\right)\right)$, which requires us to run inference with l^t as observed, find the most-probable blood vessel path of length T at time T, and compute conditional error for this blood vessel path. Unfortunately, for this phase, computations needed for $P\left(\mathcal{L}^T|l^0,l^t\right)$ cannot be shared between different t values, and hence, this step needs to be repeated for every t value the search algorithm tries.

In this active inference setting, the observations are always for past time slices; i.e., we need to compute $P\left(\mathcal{L}^t|l^0\right)$ where $t > 0$ and $P\left(\mathcal{L}^T|l^0,l^t\right)$ where $t < T$. We assume that once the laboratory experiment is stopped at time t and the tissue is dissected, all locations at time t can be observed by the experimenter. Hence, forward sampling can be conveniently utilized for both probability computations. Then, the inference cost for computing $P(\mathcal{L}^{t'}|l^t)$ is

$$IC(t'|t) = m \times (t' - t) \times s$$

where m is the number of locations on each time slice and s is the number of samples per location. The inference cost for *Step 1* is independent of t and it is $IC(T|0)$ because sampling is done once for all time slices and the computations are shared between different time slices. The inference cost for *Step 2* depends on t and it is $IC(T|t)$. Hence, a search algorithm that tries multiple time steps will incur the first cost $IC(T|0)$ once and it will incur the second cost $IC(T|t)$ for each t tried. More formally, the aggregated inference cost for a search algorithm to find the optimal t^* is:

$$ASC(t^*|search) = IC(T|t=0) + \sum_{t \in search} IC(T|t) \tag{2}$$

Assuming the uncertainty over the target time slice decreases as the evidence is gathered later in time (i.e., the later the experiment is stopped, the more reliable the predictions about the target time slice get), viable search algorithms are:

– Forward search (`ForS`), starting with $t = 1$ and ending when the uncertainty goes below σ,
– Backward search (`BackS`), starting with $t = T - 1$ and ending when the uncertainty goes above σ,
– Binary search (`BinS`), starting with the end points, and halving the space at each iteration, and
– Line search (`LineS`), iteratively fitting a line to the uncertainties of the latest known (computed) time points and trying to pinpoint the optimal time t^* accordingly.

We compare these search algorithms in terms of their aggregated inference costs (Equation 2) both empirically and analytically in the rest of the article.

4 Empirical Evaluation

In this section, we present inference results of our DBN model on various real and hypothetical settings that are obtained by varying the location of the VEGF source, VS, the number and locations of initial sprouts as `Tip Cells` (`TCs`), and the maximum distance the VEGF can travel, d_V.

4.1 Experimental Setup

Given an initial setup specified by the experimenter, which is specified as the values of all the random variables at time $t = 0$, i.e., the locations that are `TC`, `SC`, and `E` at $t = 0$, given the initial location of the distressed cell VS, and the maximum distance that the VEGF can travel, d_V, we compute the probability distribution of

$\langle \texttt{TC}, \texttt{SC}, \texttt{E} \rangle$ for each location for each time slice $0 \leq t \leq T$. We present detailed results for 5×5 and 9×9 grids for illustrative purposes first. Then we present active inference results using a much larger 51×51 grid, which is more realistic. Note that these grids are relatively small compared to the full tissue space, because each grid does not represent the full tissue space but rather represents a zoomed region of a single VS, because the task is to figure out the likelihood of this single VS being reached by a blood vessel.

The grid is drawn in such a way that the VS is at the top/north of the grid ($y = Y - 1$) and the blood vessel sprout(s) ($\texttt{TC}$) are located at the bottom/south of the grid ($y = 0$). We experimented with various conditions where the VS is located in the middle versus in the corners, the $\texttt{TC}$ is in the middle or the corners, and there is only one or more $\texttt{TC}$s in the grid. We also varied how far the VEGF can travel (d_V). We chose three representative cases where the VEGF can travel short, medium, and long (in comparison to the grid height) distances $d_V = Y/4$, $d_V = Y$, and $d_V = 4Y$. In all of our experiments, the likelihood of a $\texttt{SC}$ becoming $\texttt{TC}$ again (γ) and the noise parameter (ϵ) are both set to a very small number, 10^{-5}.

4.2 Prediction Results

Figure 4 shows the probability of each location being occupied by a blood vessel cell (i.e., $P(L^t = \texttt{TC}|l^0) + P(L^t = \texttt{SC}|l^0)$) for $t = 0, 1, 2, 3, 4$ for a 5×5 grid. The case where VEGF can travel only a short distance, $d_V = Y/4$, is shown at the top and the case where VEGF can travel a long distance, $d_V = 4Y$, is shown at the bottom. For both cases one VS is located at the middle north location and one $\texttt{TC}$ is located in the middle south.

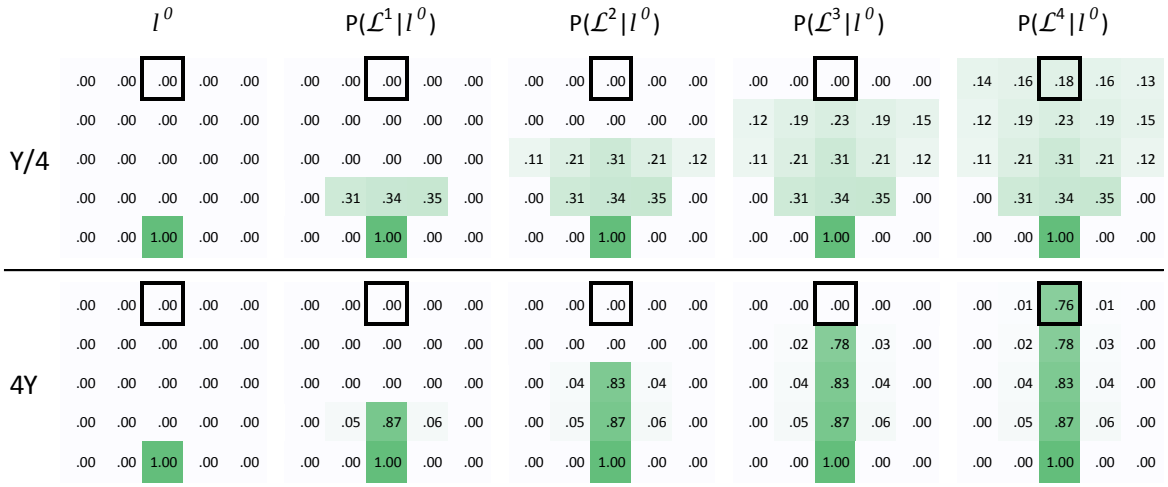


Fig. 4 Cell probabilities (the sum of $\texttt{TC}$ and $\texttt{SC}$) for the 5×5 grid at all time slices for $d_V = Y/4$ and $d_V = 4Y$.

Note that λ parameters denote the likelihood of $\texttt{TC}$ elongating in one of NW, N, or NE directions and the values of these parameters depend whether the $\texttt{TC}$ is reached by VEGF and if so, how close the $\texttt{TC}$ is to the VS, the source of the VEGF. For $d_V = Y/4$, any location that is further away from the VS as much as $Y/4$ or more has uniform λ values and hence the blood vessel has equal likelihood of growing in all three directions, resulting in uniform probabilities (the top of Figure 4). For $d_V = 4Y$, all locations are in the VEGF diffusion range, causing the λ values to be more skewed in the gradient of the VEGF. In this case, we see that the $\texttt{TC}$ simply follows the gradient to reach the distressed cell. For $d_V = Y/4$, the probability that the $\texttt{TC}$ reaches the distressed cell is only 0.18 whereas for $d_V = 4Y$, the probability is much higher: 0.76.

We next present results of different placings of VS and $\texttt{TC}$ on a 9×9 grid (we present only the $d_V = 4Y$ case as the $d_V = Y/4$ case simply results in uniform probabilities). The top of Figure 5 shows a setting where

the VS is on the north-west corner and a single TC is on the south-east corner. The bottom of the same figure shows the results for the case of one VS in the middle north, one TC in the south-east corner and another TC in the south-west corner.

Comparing to the 5×5, the 9×9 grid is larger in both dimensions and furthermore, on top the top figure, the VS and TC are placed on the opposite ends of a diagonal. This setting doubles the distance between VS and TC comparing with the previous case and hence the likelihood of VS being reached by a blood vessel is much smaller, 0.37 versus 0.76. At the bottom, however, the chance of VS being reached by a blood vessel is much higher than the case on top since there are two initial TCs and VS is not as further away. These results illustrate the obvious: that as the VEGF travels longer distances, the chance of a blood vessel finding the VS increases. When multiple initial sprouts are involved, the chance rises even more. What is perhaps more important, however, is that the DBN model associates a probability with these scenarios that can be acted upon by the tissue engineering researchers.

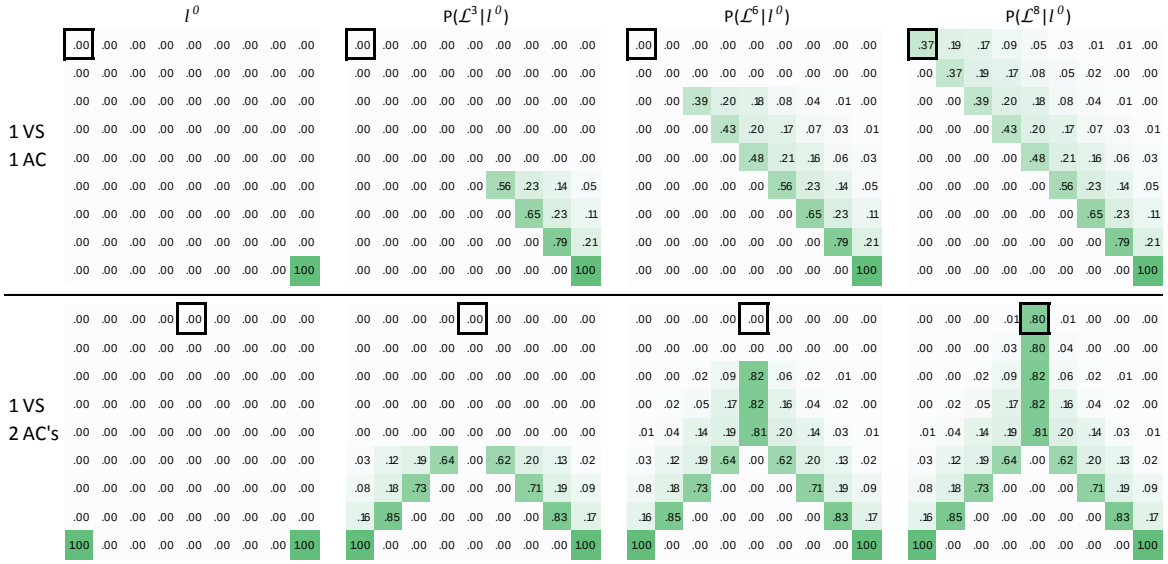


Fig. 5 Cell probabilities for the 9×9 grid at 4 selected time slices for maximum travel distance = $4Y$.

4.3 Discussion

Evaluating mathematical models of tissue engineering is fairly challenging for several reasons: there are many uncontrolled and unaccounted factors that mathematical models simply do not take into account. Therefore, the process looks just too stochastic from the modeling perspective, and comparison by a set of controlled experiments is not straightforward. Therefore, rather than evaluating precision of locations of the blood vessels, general characteristics are evaluated, such as how much branching occurs, the mean and the maximum blood vessel invasion depth, blood vessel length density, etc.

4.3.1 Laboratory Experiments

We next present laboratory experimental results where we focus on the invasion depth, i.e., the depth of vascularized tissue growth into the environment. An experimental study was previously performed using porous polymer scaffolds containing gradients of growth factors *in vivo* [1]. More detail on the experimental conditions can be

found in [1]. Briefly, gradient scaffolds with varying growth factor doses (0, 2, 20 and 200 ng) of PDGF-BB (a growth factor, of which VEGF is a subfamily) were investigated on a rodent model and vascularization was evaluated at 1, 3, and 6 weeks post implantation. Harvested samples were stained, i.e., the tissue was colored with dyes (in this case with hematoxylin and eosin (H&E)) to enhance visibility and contrast in the microscopic imaging. H&E stain cell nuclei and the tissue structure purple and pink, respectively. The dataset corresponding to this experiment will soon be available at `http://share.iit.edu/`.

Stained samples were imaged to quantify tissue invasion depth using an Axiovert 200 inverted microscope (1.10 μm/pixel). The depth of tissue invasion was measured as the straight-line distance from the underlying host tissue to the deepest location where tissue could be seen within the scaffolds. For these experiments, five animals were sacrificed at each time point (1 week, 3 weeks, and 6 weeks). Each animal received four implants each one corresponding to a different growth factor concentration (0, 2, 20, and 200ng). This procedure resulted in five samples per growth factor concentration per time point. We took measurements from three different locations (right, middle, and left) within each sample. Therefore, we have $5 \times 3 = 15$ measurements for each condition (dose & time). In each measurement, we calculated tissue invasion depth statistics for various threshold ranges (Table 1). Statistics were calculated by dividing the number of times tissue invasion reaches an identified threshold to the total number of measurements of the condition.

Table 1 Statistics of tissue invasion depth.

0 ng	**<1000 μm**	**1000 μm**	**1250 μm**	**1500 μm**	**>1750 μm**
Week 1	1	0.2	0.13	0.07	0
Week 3	1	0.53	0.13	0	0
Week 6	1	0.73	0	0	0
2 ng	**<1000 μm**	**1000 μm**	**1250 μm**	**1500 μm**	**>1750 μm**
Week 1	1	0.53	0.33	0.2	0.07
Week 3	1	0.8	0.73	0.53	0.27
Week 6	1	0.67	0.27	0.07	0
20 ng	**<1000 μm**	**1000 μm**	**1250 μm**	**1500 μm**	**>1750 μm**
Week 1	1	0.73	0.4	0	0
Week 3	1	0.93	0.67	0.4	0.2
Week 6	1	0.87	0.73	0.27	0.2
200 ng	**<1000 μm**	**1000 μm**	**1250 μm**	**1500 μm**	**>1750 μm**
Week 1	1	0.93	0.8	0.47	0.2
Week 3	1	1	0.87	0.67	0.6
Week 6	1	1	1	0.4	0.33

Blood vessels within the scaffold were labeled using a fluorescent dye (Alexa Fluor 647-isolectin) and imaged using confocal microscopy. Representative images of blood vessels within gradient scaffolds (20 ng PDGF-BB) for weeks 1, 3 and 6 are shown in Figure 6.

As these results in Table 1 illustrate, the higher the growth factor dose (the lower rows), the higher the likelihood of deeper invasion of the tissue (the columns on the right). For example, at week 6, the ratios of samples where an invasion depth of 1250 μm was reached are 0%, 27%, 73%, and 100% corresponding to 0 ng, 2 ng, 20 ng, and 200 ng doses of growth factor. Note that in some cases, the invasion depth decreases between 3 and 6 weeks due to regression and pruning, i.e., cell death.

We cannot directly compare these numbers (Table 1) with the numbers in Figure 4, because our model focuses on specific regions whereas these experimental results show aggregate information. However, these data support the claim that our DBN model captures the phenomenon that the higher the growth factor concentration is (Figure 4 bottom figure), the higher the likelihood of reaching distressed cells.

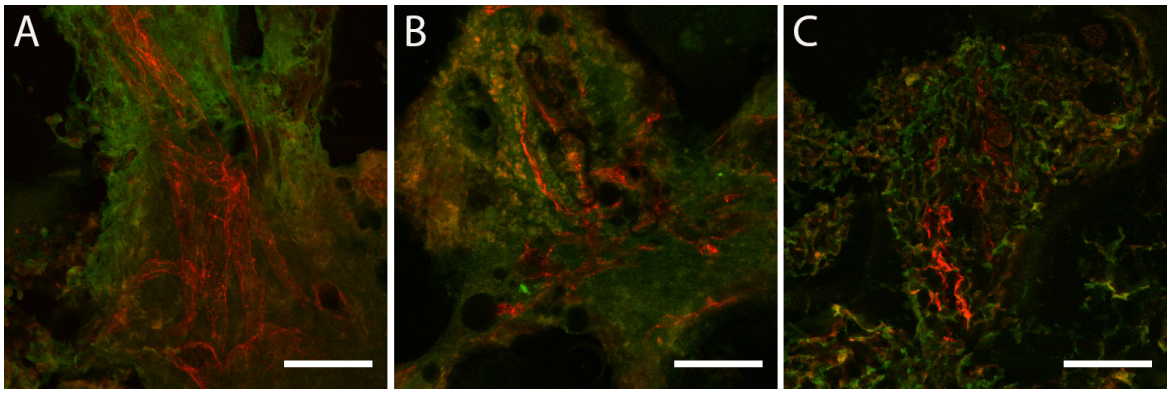


Fig. 6 Blood vessel invasion within gradient scaffolds at (A) week 1, (B) week 3, and (C) week 6. All images are for the 20 ng PDGF-BB case. Red shows isolectin labeled blood vessels and green is autofluorescence of the tissue. The large black areas are the scaffold structure. Scale bars are 100 μm.

Note that the end result of the DBN model of vascularization is not to predict the positive correlation between the amount of growth factor concentration and the likelihood of reaching distressed cells. The similarity of the results obtained in the laboratory (Table 1) and the results obtained through the DBN model (Figure 4) simply validates that the DBN model is indeed able to represent the actual phenomenon observed in the lab and hence the DBN model can be used for various probabilistic queries, such as i) given an initial condition (various placements of distressed cells, host blood vessel, VEGF concentration, etc.), what is the probability that the distressed cells will be reached? ii) given an initial setting, what will be the uncertainty about the condition of the tissue at a target time T?, and iii) given an uncertainty threshold for the target time T, how early can we stop the wet lab experiments?, and so on. We next compare the DBN results with the results obtained through a validated agent-based model of vascularization.

4.3.2 Agent-Based Models

A direct comparison of our DBN model with the laboratory experiments is challenging, as was explained above. However, there are other mathematical models of tissue engineering applications where one-to-one evaluation is more straightforward. One such mathematical modeling approach is agent-based modeling. For example, Artel et al [2] developed an agent-based model to investigate blood vessel formation in biomaterial scaffolds, and the model was improved progressively through constant revision and incorporating new experimental findings [2, 32, 33, 49].

The agent-based models have many advantages over dynamic Bayesian networks. Different cell types such as stem cells, tissue cells, and the cells forming blood vessels can be represented through different agents with specific rules governing their behavior. This allows the modeler to impart domain knowledge easily into the system. For example, a cell going through its life cycle can exist in different states and perform actions such as migration, proliferation, or differentiation. The cell state such as being hypoxic or not can be determined by the presence (or lack thereof) of nearby blood vessels that are also represented by agents that have their own rule set.

A big challenge in using agent-based models is the need for many simulations to observe the average behavior of the system. In order to observe the average behavior, the researcher repeats the same simulation multiple times, starting with the same initial conditions but varying the random seed. The number of trials is often decided on an ad-hoc manner. Graphical models have the obvious advantage over such simulations in observing average behavior because that is what the probabilistic graphical models are obviously designed for. Further, and more

importantly, graphical models allow for "what if" analysis more effectively by enabling one to provide hypothetical evidence values for future time slices and then one can reason both backward and forward in time. Starting an agent-based model in future and running it backwards, however, is almost impossible, unless one designs a complete set of new rules that takes agents back in time.

We next provide a comparison of DBN inference results to average behavior of the agent based model of Artel et al [2]. The average behavior of the agent-based model is obtained by running multiple simulations (200 in this case) and averaging the observed results. In Figure 7, on the left, we present cell probabilities at 5^{th} time tick at each location after 200 runs of the agent-based model simulation and in the middle we present cell probabilities at 5^{th} time slice of the DBN model. These results show that both models result in the same average behavior. We computed Jensen-Shannon divergence between the results of these two models and present it as a table on the right of Figure 7. As these results show, the differences between the two models are very small, and the differences stem from the approximation errors of the average behavior from multiple simulations.

Cell probabilities based on simulation

.14	.20	.21	.14	.16
.15	.18	.22	.27	.11
.14	.19	.34	.24	.10
.00	.34	.31	.36	.00
.00	.00	1.00	.00	.00

Cell probabilities based on DBN

.14	.16	.18	.16	.13
.12	.19	.23	.19	.15
.11	.21	.31	.21	.12
.00	.31	.34	.35	.00
.00	.00	1.00	.00	.00

Jensen-Shannon Divergence

3.80E-05	1.52E-03	1.03E-03	5.66E-04	9.23E-04
1.39E-03	1.20E-04	2.35E-04	5.79E-03	3.30E-03
1.49E-03	7.12E-04	5.16E-04	9.31E-04	7.38E-04
0	5.16E-04	7.40E-04	1.98E-05	0
0	0	0	0	0

Fig. 7 The comparison of cell probabilities obtained from 200 runs of simulation with our DBN model.

In addition to matching the results of the agent-based model, the DBN allows for more complicated queries that the agent-based models cannot readily handle. For example, we can observe the behavior of the underlying vascularization laboratory experiment at time t and enter the observations as evidence into the DBN model. Further, this ability to enter the state of the experiment as evidence into DBN allows us to perform active inference, which we discuss next in the following section.

4.4 Active Inference Results

In this section, we provide empirical results for investigating the question of "*given an acceptable uncertainty threshold of σ, how early can a tissue engineering experiment be stopped so that the prediction uncertainty for a target time T is below σ?*" We first present results exploring how uncertainty, as defined in Equation 1, is affected by d_V and how it varies by t. Then, we compare the computation cost of various search algorithms that search for the optimal time to stop an experiment.

4.4.1 The Uncertainty Distribution

In this experiment, we computed uncertainty as defined in Equation 1 on a 51×51 grid. We placed one VS in the middle north and one `TC` in the middle south. We experimented with $d_V = Y/4$, $d_V = Y$, and $d_V = 4Y$ to

observe the relation between d_V and uncertainty. Figure 8 shows the uncertainty in the y axis, assuming the lab experiment is stopped at time t (the x axis).

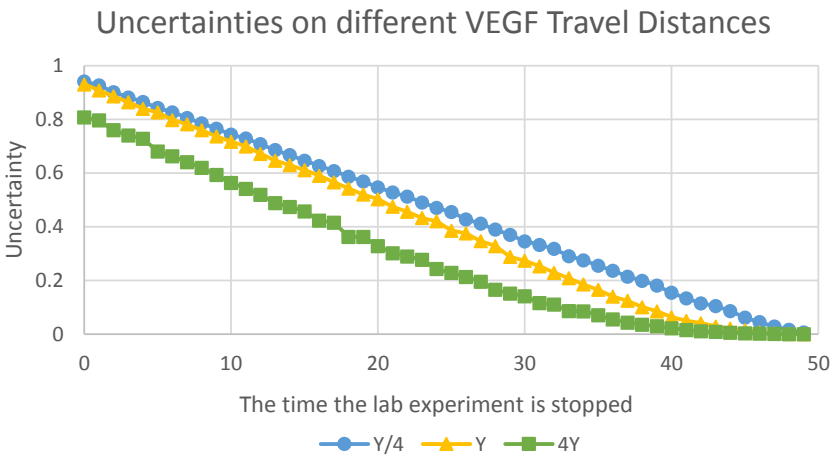


Fig. 8 Uncertainty at time slice T for each time slice candidate of observation, using various d_V values.

As expected, as we stop the tissue engineering experiment at later time slices, the uncertainty for the target time goes down. Additionally, the longer the VEGF can travel, the more skewed the λ values become and hence the blood vessel path becomes more predictable, i.e., less uncertain.

Next, we tackle the question of finding the optimal t^*, the earliest time t that guarantees a prediction uncertainty below the threshold of σ. This question can be easily answered if we compute the expected uncertainty for each candidate time t, like we did in Figure 8. However, computing uncertainty for a candidate time t is computationally expensive, as was discussed and formulated in the approach section. Next, we compare various search algorithms on how they fare in minimizing this computational cost.

4.4.2 Uncertainty Search

We compared the search methods we described in the approach section: forward search (`ForS`), backward search (`BackS`), binary search (`BinS`), and line search (`LineS`) in terms of the amount of computation (Equation 2) they would require to find t^*. We tested how the search algorithms would compare if the ideal time t^* was $t = 1$, $t = 2$, ..., $t = T - 1$. We experimented with both $d_V = Y/4$ and $d_V = 4Y$ cases, but we show only the $d_V = Y/4$ case, as the other case is similar.

Figure 9 shows the aggregated computation cost of each search method. The x axis represents the hypothetical ideal time to stop an experiment and the y axis shows the aggregated search cost a search algorithm would incur to find that ideal time. As expected, `ForS` incurs more cost and `BackS` incurs less cost as we move t^* to later time slices, and `BinS` outperforms both `ForS` and `BackS`. `LineS` performs the best because it fits a function to the so-far-computed uncertainty values and tries to pinpoint t^* using this function. We also see that `LineS` outperforms `BinS` for each t^*.

5 Analytical Evaluation

In this section, we provide closed-form solutions to the aggregated search costs (Equation 2) for the search algorithms `ForS`, `BackS`, `BinS`, and `LineS`. We make the reasonable assumption that uncertainty decreases

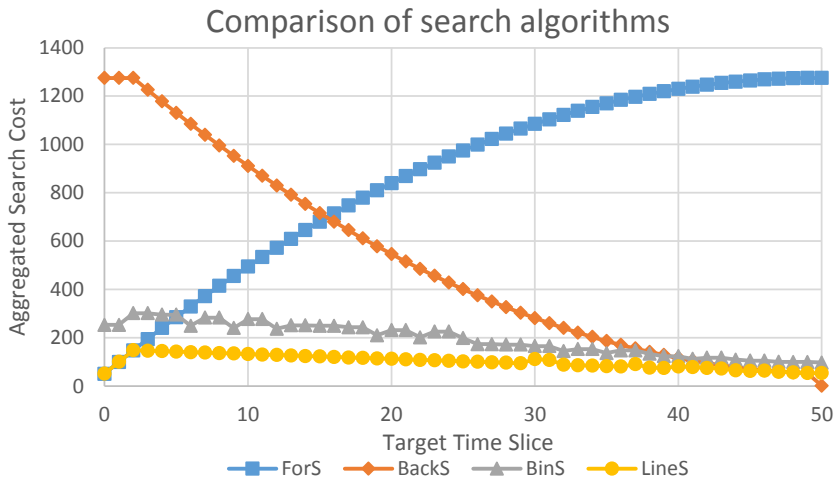


Fig. 9 Aggregated computation cost for each search method when $d_V = Y/4$.

as we provide evidence for later time slices. That is, $UNC\left(P\left(\mathcal{L}^T|l^0, l^{t+1}\right)\right) < UNC\left(P\left(\mathcal{L}^T|l^0, l^t\right)\right)$ for $0 < t < T$. This assumption is verified to hold for our domain as was shown in Figure 9.

In the active inference problem, we are searching for t^* where $t^* = \text{argmin}_{t<T}\, UNC\left(P\left(\mathcal{L}^T|l^0, l^t\right)\right) < \sigma$. `ForS` tries $t = 1, 2, \cdots, T-1$ in increasing order till it finds the first t that satisfies the uncertainty requirement. Therefore, `ForS` will try $t = 1, 2, \cdots, t^*$. `BackS` tries $t = T-1, T-2, \cdots, 1$ in decreasing order till it finds the first t that violates the uncertainty requirement. That is, `BackS` will try $t = T-1, T-2, \cdots, t^*, t^*-1$. `BinS` will perform search by halving the search space at each iteration. That is, it will first try $t = T/2$ and depending on the uncertainty at this time, it will try either $t = T/4$ or $t = 3T/4$, and so on, till it tries t^* and either one of t^*-1 or t^*+1. `LineS` tries $t = T-1$ first and then fits a line to the uncertainty at time $t = 0$, which is given, and tries to pinpoint the exact uncertainty threshold. If the pinpoint uncertainty is larger than the threshold, it fits a new line between the pinpoint and $t = T-1$, otherwise between the pinpoint and $t = 0$. It iterates narrowing the search interval down until σ is clamped between two latest pinpoints.

5.1 Forward Search

The aggregated search cost for `ForS`, $ASC(t^*|\texttt{ForS})$ is:

$$\begin{aligned} ASC(t^*|\texttt{ForS}) &= IC(T|t=0) + IC(T|t=1) + ... + IC(T|t=t^*) \\ &= m \times s \times [(T-0) + (T-1) + ... + (T-t^*)] \\ &= m \times s \times [(t^*+1) \times T - \frac{t^*(t^*+1)}{2}] \\ &= m \times s \times (t^*+1)(T - \frac{t^*}{2}) \end{aligned}$$

Note that $0 < t^* < T$ and hence each of the terms $(T-0), (T-1), \ldots, (T-t^*)$ in the second line of this equation are positive and the number of terms grows in t^*. Therefore, the ASC for `ForS` grows as the ideal time to stop the experiment is closer to the target time T as expected and as has also been empirically validated and shown in Figure 9.

5.2 Backward Search

Similarly, the aggregated search cost of `BackS` is:

$$\begin{aligned}
ASC(t^*|\texttt{BackS}) &= IC(T|t=0) + IC(T|T-1) + IC(T|T-2) + \cdots + \\
&\quad IC(T|t^*) + IC(T|t^*-1) \\
&= m \times s \times [(T-0) + (T-T+1) + (T-T+2) + \cdots + \\
&\quad (T-t^*) + (T-(t^*-1))] \\
&= m \times s \times \big[T + \big(T-(T-1)\big) + \big(T-(T-2)\big) + \cdots + \\
&\quad \big(T-(t^*-1)\big)\big] \\
&= m \times s \times \big[T + T \times \big((T-1) - (t^*-1) + 1\big) - [(T-1) + \\
&\quad (T-2) + \cdots + (t^*-1)]\,\big] \\
&= m \times s \times \left[T + T[T-t^*+1] - \frac{(T-t^*+1)(T+t^*-2)}{2}\right] \\
&= m \times s \times \left[T + \frac{(T-t^*+1)(T-t^*+2)}{2}\right]
\end{aligned}$$

Following similar reasoning, each of terms $(T-0), (T-T+1), (T-T+2), \ldots, (T-(t^*-1))$ in the second line are positive and the number of terms decreases in t^* resulting lower cost as the ideal time to stop the experiment gets closer to the target time T.

The ASC for `ForS` is monotonously increasing in t^* whereas it is monotonously decreasing for `BackS`. Next, we determine t^* for which `ForS` and `BackS` have equal cost.

$$\begin{aligned}
(t^*+1)\left(T - \frac{t^*}{2}\right) &= T + \frac{(T-t^*+1) \times (T-t^*+2)}{2} \\
-\frac{1}{2}t^{*2} + (T-\frac{1}{2})t^* + T &= T + \frac{(T-t^*+1) \times (T-t^*+2)}{2} \\
-\frac{1}{2}t^{*2} + (T-\frac{1}{2})t^* + T &= T + \frac{t^{*2} - (2T-3)t^* + T^2 + 5T + 2}{2} \\
2t^{*2} + (-4T-2)t^* + T^2 - 3T + 2 &= 0
\end{aligned}$$

The solution of this quadratic equation is:

$$\begin{aligned}
t^* &= \frac{4T + 2 \pm \sqrt{(-4T-2)^2 - 4 \times 2 \times (T^2 - 3T + 2)}}{4} \\
&= T + \frac{1}{2} \pm \sqrt{\frac{T^2}{2} - \frac{T}{2} + \frac{5}{4}}
\end{aligned}$$

The feasible solution for which $t^* < T$ is:

$$t^* = T + \frac{1}{2} - \sqrt{\frac{T^2}{2} - \frac{T}{2} + \frac{5}{4}}$$

For large T values, the t^* for which `ForS` and `BackS` have equal aggregated inference cost is equal to:

$$\begin{aligned}
t^* &\approx T - \frac{T}{\sqrt{2}} \\
&= T(1 - 1/\sqrt{2}) \\
&\approx 0.3T
\end{aligned} \tag{3}$$

That is, for $t^* < 0.3T$, `ForS` is more efficient than `BackS`, and for $t^* > 0.3T$, it is better to use `BackS`. For $T = 50$, `ForS` and `BackS` are equal on $t^* = 15$ which is roughly the value we see in Figure 9.

5.3 Binary Search

For binary search, because the actual steps depend on where t^* lies, we provide closed-form solutions for i) the worst-case, ii) the best-case, and iii) average-case scenarios.

5.3.1 Worst-Case Scenario

Because the inference cost of earlier steps is larger than the cost of the later time steps, in the worst case, the binary search keeps trying earlier time steps $t = \frac{T}{2}, \frac{T}{4}, \cdots, \frac{T}{2^{\lceil \log_2 T \rceil}}$. Let $n = \lceil \log_2 T \rceil$. The inference cost is then:

$$\begin{aligned} ASC(t^* = 0|\texttt{BinS}) &= m \times s \times \left[T + \left(T - \frac{T}{2} \right) + \left(T - \frac{T}{4} \right) + \cdots + \left(T - \frac{T}{2^n} \right) \right] \\ &= m \times s \times \left[T + nT - T\left(\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} \right) \right] \\ &= m \times s \times \left[T + nT - T\left(\frac{1}{2} + \frac{1}{4} + ... + \frac{1}{2^n} \right) \right] \\ &= m \times s \times \left[T + nT - T\left(\frac{2^n - 1}{2^n} \right) \right] \end{aligned}$$

$\frac{2^n - 1}{2^n} \approx 1$ for large n. So the equation becomes:

$$\begin{aligned} ASC(t^* = 0|\texttt{BinS}) &\approx m \times s \times [T + nT - T] \\ &= m \times s \times nT \\ &= m \times s \times \lceil \log_2 T \rceil T \end{aligned}$$

5.4 Best Case

Because the inference cost of later time steps is smaller than the cost of the earlier time steps, in the best case, the binary search keeps trying later time steps $t = \frac{T}{2}, \frac{3T}{4}, \cdots, \frac{(2^{\lceil \log_2 T \rceil} - 1)T}{2^{\lceil \log_2 T \rceil}}$. Let $n = \lceil \log_2 T \rceil$. The inference cost is then:

$$\begin{aligned} ASC(t^* = T - 1|\texttt{BinS}) = m \times s \times T + \Bigg[&\left(T - \frac{T}{2} \right) + \left(T - \frac{3T}{4} \right) + \cdots + \\ &\left(T - \frac{(2^n - 1)T}{2^n} \right) \Bigg] \\ = m \times s \times &\left[T + T\left(\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} \right) \right] \\ = m \times s \times &\left[T + T\left(\frac{2^n - 1}{2^n} \right) \right] \\ \approx m \times s \times &\, 2T \end{aligned}$$

5.4.1 Average Case

We analyze the average case in Figure 10. Binary search makes a decision on two options at each step. Without prior knowledge, two options are equally likely to be selected. Since each option in the previous step is equally likely, all possible decisions are also equally likely in following steps. For example in Figure 10, in the third step, four time slices can be dissection point and each of them can be reached with equal probability. Therefore, knowing inference cost for each time slice, we can compute expected inference cost at each step. Computing and summing expected inference cost at each search step gives the expected inference cost of binary search as a function of T , which is equal to $m \times s \times \lceil log_2 T \rceil T/2$.

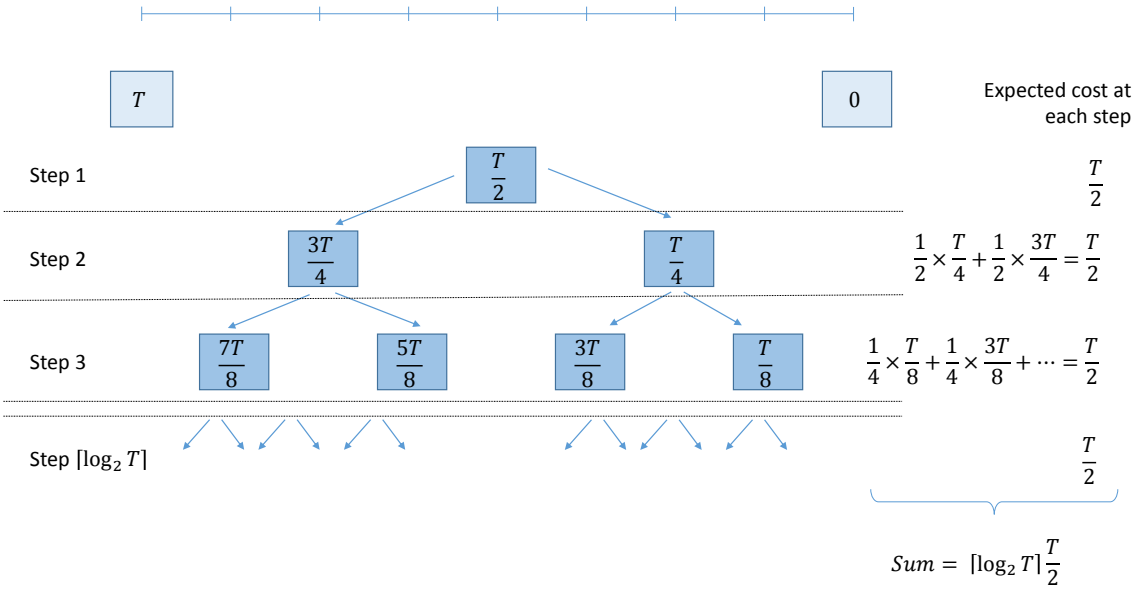


Fig. 10 Indices of potentially explored time slices by binary search at each step of the search phase. Each of these time slices may or may not be explored. Hence, probability of a time slice for being explored is equal to other time slices of the same row.

5.5 Line Search

If $UNC(P(\mathcal{L}^T|l^0, l^t))$ is assumed to follow a perfect line as a function of t, then line search can try $t = T - 1$ (the cheapest case) and because $t = 0$ case is already given, can fit a line and pinpoint t^* without any further search. In that case, the total cost of `LineS` is:

$$\begin{aligned} ASC(t^*|\texttt{LineS}, perfect\ line) &= IC(T|t = 0) + IC(T|t = T - 1) \\ &= m \times s \times [(T - 0) + (T - (T - 1))] \\ &= m \times s \times (T + 1) \end{aligned}$$

As Figure 8 shows, the uncertainty is not a perfect line but close to a line. In that case, line search needs to verify that its pinpoint is indeed correct. That is, line search tries $t = T - 1$, $t = t^*$, and $t = t^* - 1$ to verify $UNC(\mathcal{L}^T|l^0, l^{t^*}) < \sigma < UNC(\mathcal{L}^T|l^0, l^{t^*-1})$. Then, for each possible t^*, line search incurs the cost of $IC(T|t = 0), IC(T|t = T - 1), IC(T|t = t^*), IC(T|t = t^* - 1)$.

t^*	explored slices	aggregated search cost
1	$0, T-1, 1$	$(T-0)+(T-(T-1))+(T-1)$
2	$0, T-1, 1, 2$	$(T-0)+(T-(T-1))+(T-1)+(T-2)$
3	$0, T-1, 2, 3$	$(T-0)+(T-(T-1))+(T-2)+(T-3)$
...	...	...
t	$0, T-1, t-1, t$	$(T-0)+(T-(T-1))+(T-(t-1))+(T-t)$
...	...	...
$T-2$	$0, T-1, T-3, T-2$	$(T-0)+(T-(T-1))+(T-(T-3))+(T-(T-2))$
$T-1$	$0, T-1, T-2$	$(T-0)+(T-(T-1))+(T-(T-2))$

Fig. 11 Aggregated search cost for each t^* value in the case that uncertainty curve is close to a line.

In order to compute the average case, we need to calculate the summation of aggregated search cost for each of $t^* = 1, 2, \cdots T-1$, as illustrated in Figure 11:

$$\begin{aligned}
\sum_{t^*=0}^{T-1} ASC(t^*|\texttt{LineS}) &= m \times s \times \big[(T-1)\times(T+1) + [(T-1)+(T-2)+...+2] \\
&\quad + [(T-1)+(T-2)+(T-3)+...+2]\big] \\
&= m \times s \times \left[(T-1)(T+1) + 2 \times \left[\frac{(T-1)T}{2} - 1\right]\right] \\
&= m \times s \times \Big[(T^2-1) + (T-1)T - 2\Big] \\
&= m \times s \times \big(T^2 - 1 + T^2 - T - 2\big) \\
&= m \times s \times \big(2T^2 - T - 3\big)
\end{aligned}$$

The average case aggregated search cost is then:

$$\begin{aligned}
E[ASC(t^*|\texttt{LineS})] &= \frac{1}{T} \times \sum_{t^*=0}^{T-1} ASC(t^*|\texttt{LineS}) \\
&= \frac{1}{T} \times m \times s \times (2T^2 - T - 3) \\
&= m \times s \times \left(2T - 1 - \frac{3}{T}\right)
\end{aligned}$$

6 Current Limitations and Generalizations to Other Domains

The DBN model we presented in this article serves as a proof of concept to show the feasibility of using DBNs for vascularization, as has been validated through the similarities of the results with those of agent-based models of vascularization [2], and on an aggregate level through real-world experimental data [1]. The DBN model can be enriched further to reflect reality, such as a 3D instead of a 2D model. Much of these enrichments are engineering problems, where the parents of a node come from a 3D space instead of a 2D space.

The active inference formulation we discussed in this article focused on tissue engineering. Some of the ideas and approaches are, however, general enough to be applicable to several other domains. There are many practical scenarios in which intelligent agents face the question of how much to wait gathering information versus when to act based on what is so far known. This problem is in fact an active inference problem in a temporal domain such as the one we described in this article and hence cost formulations and search results we discussed largely carry over to these problems.

Though our active inference formulation is generic, our inference cost calculations assumed that when an observation is made at time slice t, all variables at that time slice are observed. Hence, expected uncertainty calculations required forward sampling starting from time t to target time T. This meant searching for later time steps incurred less cost (note that `BackS` had less cost than `ForS` for $t^* > 0.3T$ as shown in Equation 3). In practice, however, there might be hidden variables or variables that are too costly to acquire so that not all variables can be observed at time t. Then, a simple forward sampling from t to T would not suffice; instead, other approximations algorithms, e.g., likelihood weighting and Gibbs sampling, including all non-observed variables from time 0 to time T need to be performed and hence search for earlier and later steps would incur the same inference cost. In that case, `ForS` and `BackS` would have equal cost when $t^* = T/2$. Still, `BinS` is expected to outperform `ForS` and `BackS`, and `LineS` is expected to outperform `BinS`, as expected.

7 Related work

Tissue engineering is an active research field where a number of papers tackle different aspects of the problem using in-vivo and in-vitro laboratory experiments (e.g., [1, 16, 17, 23, 22, 51]). There have been also computational models of tissue development, such as multi-agent systems (e.g., [2, 4, 5, 12, 33]). Artel et al [2] propose a multi-agent model to simulate vascularization in polymer scaffolds. They compare effects of various growth factor concentrations and scaffold porosities on vascularization. Mehdizadeh et al [33] observe the contribution of porosity and interconnectivity of scaffold to vascularization. Bentley et al [5] also propose an agent-based model for vascularization in which they observe the effect of different growth factor conditions on blood vessel growth. Bailey et al [4] use an agent-based model to represent an existing blood vessel network and they simulate white blood vessel trafficking. Finally, Byrne et al [12], model tissue differentiation using agent-based models under different settings of scaffold porosity and degradation rate.

In this article, we propose a dynamic Bayesian network modeling to vascularization in engineered tissues. We compared our approach to the agent based model of Artel et al [2] and we found similar results. Though agent-based models provide a more natural fit for modeling various cells and biomaterials as agents, DBNs allow a natural formalism for modeling the spatiotemporal uncertainty present in vascularization. Further, DBNs enable researchers to ask what if questions about past and future through setting a subset of random variables as evidence variables.

DBNs have also been successfully applied to several other real-world applications. A few examples include managing water sources [11], gene regulatory networks [21, 35, 37], figure tracking [36], ranking [13], speech recognition [53], modeling environmental problems [50], and driverless cars [20].

Active inference was previously formulated and utilized for iterative classification of nodes of a network [8, 9], video analysis using hidden Markov models [14], battery optimization in wireless sensor networks [25], and more generally as value of information in graphical models [10, 27]. In this paper, we formulate active inference problem specifically for dynamic Bayesian networks and utilize it for computing the ideal time to stop a laboratory experiment to guarantee an acceptable prediction uncertainty about the final outcome of the experiment.

A closely related area to active inference is active learning [42]. In this field, similar to active inference, the objective is to gather information to help the underlying predictive model, but unlike active inference, active learning collects information during training of the predictive model, not during inference. Common approaches include query-by-committee [18, 44], expected error reduction [41], and uncertainty sampling [29]. Uncertainty sampling is arguably the most frequently utilized approach and hence there has been several approaches that augment uncertainty sampling. For example, in density-weighted uncertainty sampling, Settles and Craven [43] weighed uncertainty of instances by their similarity to other instances so as to avoid picking outliers. Sharma

and Bilgic [45, 46] formulated two types of uncertainties: conflicting-evidence uncertainty where the underlying model has conflicting and strong information about which class the instance belongs to and insufficient-evidence uncertainty where the underlying model has little information about which class the instance belongs to. Recently, several researchers looked at incorporating domain knowledge and different kinds of queries into the learning. Two prominent examples are feature annotation [3, 19, 34, 38, 48] and learning with rationales[47, 52]. Ramirez-Loaiza et al [39, 40] focused on improving annotation effiency by showing a carefully chosen snippet of an instance rather than the full instance.Bilgic [7] performed dynamic dimensionality reduction to improve learning efficiency of the underlying model when trained on scarce data.

8 Conclusion

We presented a dynamic Bayesian network model of vascularization in engineered tissues. The DBN model allows tissue researchers to perform spatial and temporal reasoning for the tissue development process. Additionally, we formulated and evaluated active inference for DBNs in the context of tissue engineering, aimed at answering the question of determining the ideal time to stop a laboratory experiment to guarantee an acceptable uncertainty for the prediction of the future progress of the tissue. We compared several search algorithms and analyzed their inference time complexity, providing closed-form solutions whenever possible. In this article, we focused on the tissue engineering application. However, the active inference formulation for DBNs and the complexity analysis for the search algorithms are general and can potentially be applied to other spatio-temporal domains under some natural assumptions.

Acknowledgements This material is based upon work supported by the National Science Foundation under grant no. IIS-1125412.

References

1. Akar B, Jiang B, Somo SI, Appel AA, Larson JC, Tichauer KM, Brey EM (2015) Biomaterials with persistent growth factor gradients in vivo accelerate vascularized tissue formation. Biomaterials 72:61–73
2. Artel A, Mehdizadeh H, Chiu YC, Brey EM, Cinar A (2011) An agent-based model for the investigation of neovascularization within porous scaffolds. Tissue Engineering Part A 17(17-18):2133–2141
3. Attenberg J, Melville P, Provost F (2010) A unified approach to active dual supervision for labeling features and examples. In: Proceedings of the European Conference on Machine Learning (ECML), pp 40–55
4. Bailey AM, Thorne BC, Peirce SM (2007) Multi-cell agent-based simulation of the microvasculature to study the dynamics of circulating inflammatory cell trafficking. The Journal of the Biomedical Engineering Society 35(6):916 – 936
5. Bentley K, Gerhardt H, Bates PA (2008) Agent-based simulation of notch-mediated tip cell selection in angiogenic sprout initialisation. Journal of Theoretical Biology 250(1):25 – 36
6. Bianco P, Robey PG (2001) Stem cells in tissue engineering. Nature 414(6859):118–121
7. Bilgic M (2012) Combining active learning and dynamic dimensionality reduction. In: Proceedings of the SIAM International Conference on Data Mining (SDM)
8. Bilgic M, Getoor L (2009) Reflect and correct: A misclassification prediction approach to active inference. ACM Transactions on Knowledge Discovery from Data 3(4):1–32
9. Bilgic M, Getoor L (2010) Active inference for collective classification. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI NECTAR Track), pp 1652–1655
10. Bilgic M, Getoor L (2011) Value of information lattice: Exploiting probabilistic independence for effective feature subset acquisition. Journal of Artificial Intelligence Research (JAIR) 41:69–95

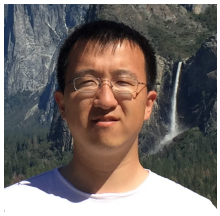
11. Bromley J, Jackson NA, Clymer OJ, Giacomello AM, Jensen FV (2005) The use of Hugin to develop Bayesian networks as an aid to integrated water resource planning. Environmental Modelling & Software 20(2):231–242
12. Byrne DP, Lacroix D, Planell JA, Kelly DJ, Prendergast PJ (2007) Simulation of tissue differentiation in a scaffold as a function of porosity, young's modulus and dissolution rate: Application of mechanobiological models in tissue engineering. Biomaterials 28(36):5544 – 5554
13. Chapelle O, Zhang Y (2009) A dynamic Bayesian network click model for web search ranking. In: Proceedings of the International Conference on World Wide Web (WWW), pp 1–10
14. Chen D, Bilgic M, Getoor L, Jacobs D (2011) Dynamic processing allocation in video. IEEE Transactions on Pattern Analysis and Machine Intelligence 33:2174–2187
15. Chen D, Bilgic M, Getoor L, Jacobs D, Mihalkova L, Yeh T (2011) Active inference for retrieval in camera networks. In: Proceedings of IEEE Workshop on Person Oriented Vision
16. Chiu YC, Cheng MH, Engel H, Kao SW, Larson JC, Gupta S, Brey EM (2011) The role of pore size on vascularization and tissue remodeling in PEG hydrogels. Biomaterials 32(26):6045 – 6051
17. Chiu YC, Kocagoz S, Larson JC, Brey EM (2013) Evaluation of physical and mechanical properties of porous poly (ethylene glycol)-co-(l-lactic acid) hydrogels during degradation. PLoS One 8:4
18. Dagan I, Engelson SP (1995) Committee-based sampling for training probabilistic classifiers. In: Proceedings of the International Conference on Machine Learning (ICML), pp 150–157
19. Druck G, Settles B, McCallum A (2009) Active learning by labeling features. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 81–90
20. Forbes J, Huang T, Kanazawa K, Russell S (1995) The BATmobile: Towards a Bayesian automated taxi. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), pp 1878–1885
21. Husmeier D (2003) Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. Bioinformatics 19(17):2271–2282
22. Jiang B, Waller TM, Larson JC, Appel AA, Brey EM (2013) Fibrin-loaded porous poly(ethylene glycol) hydrogels as scaffold materials for vascularized tissue formation. Tissue Engineering, Part A 19:224–234
23. Jiang B, Akar B, Waller T, Larson J, Appel A, Brey E (2014) Design of a composite biomaterial system for tissue engineering applications. Acta Biomaterialia 10(3):1177 – 1186
24. Koller D, Friedman N (2009) Probabilistic Graphical Models: Principles and Techniques. The MIT Press, Cambridge, Massachusetts
25. Komurlu C, Bilgic M (2016) Active inference and dynamic gaussian Bayesian networks for battery optimization in wireless sensor networks. In: Proceedings of AAAI Workshop on Artificial Intelligence for Smart Grids and Smart Buildings
26. Komurlu C, Shao J, Bilgic M (2014) Dynamic Bayesian network modeling of vascularization in engineered tissues. In: Proceedings of the Eleventh UAI Bayesian Modeling Applications Workshop
27. Krause A, Guestrin C (2009) Optimal value of information in graphical models. Journal of Artificial Intelligence Research 35:557–591
28. Langer R, Vacanti J (1993) Tissue engineering. Science 260(5110):920–926
29. Lewis DD, Gale WA (1994) A sequential algorithm for training text classifiers. In: Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval, pp 3–12
30. Li WJ, Tuli R, Okafor C, Derfoul A, Danielson KG, Hall DJ, Tuan RS (2005) A three-dimensional nanofibrous scaffold for cartilage tissue engineering using human mesenchymal stem cells. Biomaterials 26(6):599–609
31. Mao J, Giannobile W, Helms J, Hollister S, Krebsbach P, Longaker M, Shi S (2006) Craniofacial tissue engineering by stem cells. Journal of dental research 85(11):966–979

32. Mehdizadeh H, Artel A, Brey EM, Cinar A (2011) Multi-agent systems for biomedical simulation: modeling vascularization of porous scaffolds. In: Proceedings of the International Conference on Agents in Principle, Agents in Practice, pp 113–128
33. Mehdizadeh H, Sumo S, Bayrak ES, Brey EM, Cinar A (2013) Three-dimensional modeling of angiogenesis in porous biomaterial scaffolds. Biomaterials 34(12):2875 – 2887
34. Melville P, Sindhwani V (2009) Active dual supervision: Reducing the cost of annotating examples and features. In: Proceedings of the NAACL HLT Workshop on Active Learning for Natural Language Processing, pp 49–57
35. Min Z, Conzen SD (2005) A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics 21(1):71–79
36. Pavlović V, Rehg JM, Cham TJ, Murphy KP (1999) A dynamic Bayesian network approach to figure tracking using learned dynamic models. In: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol 1, pp 94–101
37. Perrin BE, Ralaivola L, Mazurie A, Bottani S, Mallet J, d'Alche Buc F (2003) Gene networks inference using dynamic Bayesian networks. Bioinformatics 19(suppl 2):ii138–ii148
38. Raghavan H, Allan J (2007) An interactive algorithm for asking and incorporating feature feedback into support vector machines. In: Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval, pp 79–86
39. Ramirez-Loaiza ME, Culotta A, Bilgic M (2013) Towards anytime active learning: Interrupting experts to reduce annotation costs. In: Proceedings of the IDEA Workshop at ACM SIGKDD Conference on Knowledge Discovery and Data Mining
40. Ramirez-Loaiza ME, Culotta A, Bilgic M (2014) Anytime Active Learning. In: Proceedings of the AAAI Conference on Artificial Intelligence
41. Roy N, McCallum A (2001) Toward optimal active learning through sampling estimation of error reduction. In: Proceedings of the International Conference on Machine Learning (ICML), pp 441–448
42. Settles B (2012) Active Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool
43. Settles B, Craven M (2008) An analysis of active learning strategies for sequence labeling tasks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp 1070–1079
44. Seung HS, Opper M, Sompolinsky H (1992) Query by committee. In: Proceedings of the Conference on Learning Theory (COLT), pp 287–294
45. Sharma M, Bilgic M (2013) Most-surely vs. least-surely uncertain. In: Proceedings of the IEEE International Conference on Data Mining (ICDM)
46. Sharma M, Bilgic M (2016) Evidence-based uncertainty sampling for active learning. Data Mining and Knowledge Discovery pp 1–39
47. Sharma M, Zhuang D, Bilgic M (2015) Active learning with rationales for text classification. In: Proceedings of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies
48. Small K, Wallace BC, Brodley CE, Trikalinos TA (2011) The constrained weight space svm: learning with ranked features. In: Proceedings of the International Conference on Machine Learning (ICML), pp 865–872
49. Somo SI, Akar B, Bayrak ES, Larson JC, Appel AA, Mehdizadeh H, Cinar A, Brey EM (2015) Pore interconnectivity influences growth factor-mediated vascularization in sphere-templated hydrogels. Tissue Engineering Part C: Methods 21(8):773–785
50. Uusitalo L (2007) Advantages and challenges of Bayesian networks in environmental modelling. Ecological modelling 3(203):312–318

51. Yang S, Leong KF, Du Z, Chua CK (2004) The design of scaffolds for use in tissue engineering. part I. traditional factors. Tissue Engineering 7:679–689
52. Zaidan OF, Eisner J, Piatko CD (2007) Using annotator rationales to improve machine learning for text categorization. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies, pp 260–267
53. Zweig G, Russell S (1998) Speech recognition with dynamic Bayesian networks. In: Proceedings of the national conference on Artificial intelligence/Innovative applications of artificial intelligence, pp 173–180

Author Biographies

Caner Komurlu received his B.S. degrees in computer engineering and applied mathematics and his M.S. degree in computer science from Istanbul Technical University. He is now a PhD candidate in computer science at Illinois Institute of Technology and a member of the Machine Learning Laboratory. His research interests include machine learning, statistics, and probabilistic graphical models. His work is focused on dynamic Bayesian network modeling of real world problems and active inference/intelligent information gathering for dynamic Bayesian networks. Before joining Illinois Tech, he worked in the industry as software engineer for 3 years. (`http:mypages.iit.edu/~ckomurlu/`)



Jinjian Shao is a Masters student at the Computer Science Department at Illinois Institute of Technology. He received his Bachelors of Software engineering degree from Tongji University, Shanghai, China. His research interests are machine learning and probabilistic graphical models.

Banu Akar received her B.S. degree in Chemical Engineering from Hacettepe University in 2011, and is currently pursuing a Ph.D. degree in Biomedical Engineering at the Illinois Institute of Technology. Her research focuses on biomaterial design, modeling, and tissue engineering in the areas of vascularization and bone regeneration.

Elif S. Bayrak, Ph.D., is a Senior Engineer at Amgen in Digital Integration and Predictive Technologies group under Process Development organization. She received her Ph.D. degree from Illinois Institute of Technology in Chemical Engineering in Chicago. She focused on developing high performance agent-based models for bioengineering applications. She obtained her bachelor's degree from Hacettepe University in Ankara, Turkey. Her email address is ebayrak@amgen.com.

Eric M. Brey, Ph.D., is Duchossois Leadership Professor of Biomedical Engineering at the Illinois Institute of Technology and Research Health Scientist at Hines Veterans Administration Hospital. He received B.S. and MEng degrees in Chemical Engineering from the University of Louisville and a Ph.D. in Chemical Engineering from Rice University. He served as a National Institutes of Health post-doctoral Fellow in the Departments of Surgery and Cell Biology at Loyola University Medical Center. Professor Brey has contributed to new methods for engineering vascularized tissues and imaging methods for analysis of engineered tissues. Professor Brey's research has resulted in over 100 peer-reviewed publications, 9 book chapters and one book. He is a fellow of the American Institute for Medical and Biological Engineers.

Ali Cinar, Ph.D., received the Ph.D. degree in chemical engineering from Texas A&M University, College Station, TX, USA. He is a Professor of Chemical Engineering and Biomedical Engineering at Illinois Institute of Technology and since 2004 he has been the director of the Engineering Center for Diabetes Research and Education. His current research includes agent-based techniques for modeling, supervision, and control of complex and distributed systems, modeling of diabetes, angiogenesis and tissue formation, and adaptive control techniques for artificial pancreas systems for patients with diabetes. He has published two books, more than 200 technical papers in refereed journals and conference proceedings. A full list of publications, detailed description of research interests, presentations, and software is available at `http://engineering.iit.edu/faculty/ali-cinar`. Dr. Cinar is a senior member of the IEEE and a fellow of the AIChE and a member of American Diabetes Association and Biomedical Engineering Society.

Mustafa Bilgic, Ph.D., is an Associate Professor at the Computer Science Department and the Director of the Machine Learning Laboratory at Illinois Institute of Technology in Chicago. He received his B.S. degree in Computer Science from the University of Texas at Austin and his M.S. and Ph.D. degrees in Computer Science from the University of Maryland at College Park. Dr. Bilgic's research interests include interactive machine learning, transparent machine learning, active learning, active inference, recommender systems, probabilistic graphical models, and spatio-temporal reasoning. He received several awards including ACM SIGKDD Best Student Paper Award, National Science Foundation CAREER Award, and Illinois Institute of Technology College of Science Junior Researcher Excellence Award.