A Machine Learning Approach for Abnormality Detection in Blood Vessels via Mobile Nanosensors

Jorge Torres Gómez
torres-gomez@ccs-labs.org
TU Berlin
Berlin, Germany

Anke Kuestner
anke.kuestner@ccs-labs.org
TU Berlin
Berlin, Germany

Ketki Pitke
pitke@ccs-labs.org
TU Berlin
Berlin, Germany

Jennifer Simonjan
Jennifer.Simonjan@ieee.org
Lakeside Labs
Klagenfurt, Austria

Bige Deniz Unluturk
unluturk@msu.edu
Michigan State University
East Lansing, MI, USA

Falko Dressler
dressler@ccs-labs.org
TU Berlin
Berlin, Germany

ABSTRACT
Early detection of diseases in the human body is of utmost importance for the diagnosis and medical treatment of patients. Supported by recent advancements in nanotechnology, diseases may be detected by patrolling nanosensors, even before symptoms appear. This paper explores the detection capabilities of nanosensors flowing through the human circulatory system (HCS). We model the HCS through a Markov chain and propose the use of machine learning (ML) methods to learn the corresponding transition probabilities. Doing so, we propose a methodology to develop an early detection mechanism of quorum sensing (QS) molecules released by bacteria. Simulation results indicate the suitability of our machine learning approach as a basis for in-body precision medicine.

CCS CONCEPTS
• Computing methodologies → Classification and regression trees; • Mathematics of computing → Markov networks; • Applied computing → Health care information systems.

KEYWORDS
Biosensors, Machine Learning, Human Circulatory System, Markov Model, Precision Medicine

ACM Reference Format:
https://doi.org/10.1145/3485730.3494037

1 INTRODUCTION
Medical applications are evolving through unprecedented development in the field of nanotechnology. Nanoscale sensing devices, capable of continuously monitoring health parameters [4, 7], are foreseen to travel through the human circulatory system (HCS), enabling next generation precision medicine solutions to actuate in hard-to-reach areas due to their tiny dimensions. Equipped with synthetic units made of biological components [27], nanodevices are expected to carry out embedded computing capabilities to sense and actuate on targets. Furthermore, such nanosensors can be equipped with networking capabilities in order to enable inter-device communication [5, 12]. Exploiting these sensing and communication capabilities, novel applications like early disease detection and smart targeted drug delivery are in today’s research focus [11, 26].

In the case of infectious diseases, patrolling nanosensors help avoiding detection delays introduced by conventional methods, which typically take 48–72 h [3]. Nanonetworks, leveraging the advances in nanosensors and molecular communication (MC), are envisioned to detect anomalous concentrations of biomarkers in vivo. Once detected, observed data will be sent to external wearable hubs for further analysis and diagnosis by healthcare professionals, allowing proactive treatment procedures. This idea has been studied in the context of detecting cancer cells [21], tumors [32], and other abnormalities [18], even before symptoms appear.
Following this direction, we explore the possibility of early infection detection by mobile nanosensors. There exist already first studies and concepts on potential shapes and sizes of such nanoscale sensing devices [25]. These nanosensors, flowing through the vessels of the HCS, will inherently make use of the in body networking capabilities [24]. As depicted in Fig. 1, nanosensors with sensing capabilities will detect molecules emitted during quorum sensing (QS) communications among bacteria, therefore preventing further infection in the given body region (cf. Fig. 1c). Upon successful detection, the nanosensors will report their sensing data to an external gateway when passing, e.g., the right heart chamber (cf. Fig. 1b). Although we assume an ideal connection link between nanosensors and the gateway, in practice, communication through human tissue can be achieved by exploiting sub-terahertz [22, 26] or ultrasonic frequencies [8, 23].

Doing so, we provide a methodology to evaluate the detection performance, which will ultimately depend on the distribution of nanosensors in the human blood vessels. For this purpose, we simulate the motion of nanosensors in the HCS using the BloodVoyagerS simulation framework [15] and compute their distribution through a Markov model. Furthermore, we incorporate a machine learning (ML)-based approach to evaluate the transition probabilities of the Markov model. This approach provides a low-complex and widely applicable solution avoiding the specifics of physiological parameters (e.g., vessel lengths, blood viscosity, blood pressure [16]) and the use of complex blood flow mechanisms in the vessels (e.g., turbulence [29]). Unsupervised and supervised methods are combined to predict the distribution of nanosensors in the HCS based on their traveling time in the vessels. The use of ML models has already been reported to support the detection of diseases from many different types of data such as genomics [9]. In this work, we report its use to infer the location of nanosensors in the HCS for the first time. Our simulation results clearly show the feasibility of our approach, paving the way for future precision medicine solutions.

2 DETECTION OF INFECTIOUS DISEASES

To detect infectious diseases, we consider that mobile nanosensors are capable of sensing QS molecules emitted by bacteria when establishing communication among them. Even though the human body contains a large number of harmless bacteria, some pathogenic bacteria growing out of control can cause serious health conditions. Early detection of infections and timely administration of antibiotics can reduce symptoms and improve the cure rate. QS molecules are unique to each species of bacteria and play a role in the onset of infection. Since they diffuse into the bloodstream from the infection site, it is possible to detect them in bodily fluids such as blood, saliva, and sputum in correlation with the infection status [6, 19].

According to COMSOL simulations, based on our previous work [3], the distribution of QS molecules in the vicinity of an infection is shown in Fig. 2. The figure depicts a tissue section with an infection site on top and the capillaries here located at the bottom of this figure. Based on this result, we can define the region where the QS molecule concentration is above the detection threshold of the nanosensors as the sensing region. Considering the detection limit of nanosensors to be $1 \times 10^{-5}$ million [20], the sensing region can be approximated to be an ellipse elongated towards the direction of blood flow with axes lengths of 0.1–0.5 cm and 1.75–3.5 cm depending on the blood velocity in the capillaries in different regions of the body (0.01–1 cm/s).

Given the region of interest $V_k$ (e.g., capillaries in the arms, center body, or legs), we can express the conditional probability for a nanosensor to visit the sensing region $V_{k,s} \subset V_k$, given that the nanosensor is in the region $V_k$, as

$$P_{s|k} = \text{Prob}(I_n \in V_{k,s} | I_n \in V_k) = \frac{V_{k,s}}{V_k},$$

where $I_n$ is the location of nanosensors, $V_{k,s}$ is the volume of the sensing region, defined via the COMSOL simulations for the organ $k$, and $V_{T,k}$ is the total volume for the given organ.

We set our observation time to 20 min, which is the reproduction time of many bacteria. During this observation window, QS molecule concentration will not go through a dramatic change and nanosensors will go through several loops in the cardiovascular system, thereby reporting multiple samples.

We measure the probability of successful detection, $P_{d,k}$, as the probability of at least $N_{th}$ nanosensors reporting sensed QS molecules as

$$P_{d,k} = \sum_{i=N_{th}}^{N} B(i, P_{s|k} P_{c,k}),$$

where $B(\cdot)$ denotes the binomial distribution, $N_{th}$ is the arbitrary threshold for the number of samples needed to decide towards detection of infection, $P_{s|k}$ is given in Eq. (1), and $P_{c,k}$ is the probability of a nanosensor being located in the capillaries of organ $k$.

The binomial distribution, we report in Eq. (2) accounts for the probability of $i$ nanosensors visiting the sensing region out of a total of $N$. Summing from $N_{th}$ to $N$, will consider that at least $N_{th}$ sensors are visiting the given sensing region. The two main variables to evaluate this expression are given by $P_{s|k}$ and $P_{c,k}$, with the latter still to be determined. In order to estimate $P_{c,k}$, the probability of a nanosensor visiting the capillaries of organ $k$, we exploit Markov chains and ML models (cf. Sections 3.2 and 4).

\(^{1}\)On average it takes 1 min for the blood to circulate through the body [16].
Following this procedure, we conceive the methodology illustrated in Fig. 3 to provide alerts with a given reliability level for the clinical reports. The successful detection probability (reliability level) comes from the COMSOL simulation and Markov chain results, whereby the latter is derived from evaluating the corresponding transition probabilities with the supervised ML model. The alert comes after combining the supervised and unsupervised ML models. The next sections address the specific details concerning the right branch in Fig. 3.

3 MODELING SCHEMES

The detection capabilities of the flowing nanosensors are modeled by two primary schemes, addressing both, their traveling behavior and their stationary location in the HCS. The traveling behavior is modeled using the simulation framework BloodVoyagerS [15]. The stationary distribution of nanosensors is derived after modeling the process as a Markov chain.

3.1 Modeling the flow of nanosensors in the human circulatory system

Modeling all characteristics of the HCS is non-trivial as it comprises approx. 4900 cm$^3$ of blood volume and 120 000 km of blood vessels [31]. To achieve a realistic model of the traveling behavior of nanosensors, we use BloodVoyagerS [15], which is a nanonetwork simulation module, capable of simulating the blood flow of all major vessels in the HCS. BloodVoyagerS is modeled upon a simplified HCS to realize the movement of nanosensors within the human bloodstream. Fig. 4 depicts all uniquely numbered vessels and organs considered by the simulator. It comprises a model including 94 vessels and their respective blood flow rates (20 cm/s in the aorta, 10 cm/s in the arteries, 2–4 cm/s in the veins), adding up to a total simulated vessel length of 12717 m (vessel length measurements are based on a person with 1.72 m height and 69 kg weight). The blood flow in a healthy circulatory system is laminar, consisting of ordered layers that slide past each other. BloodVoyagerS incorporates this effect by implementing separate streams in each vessel. These streams have different sizes and an additional speed factor that adapts the basic speed in a vessel to the respective stream. In streams on the outer part of a vessel the speed is lower than in the center streams. The simulator allows to inject an arbitrary number of nanosensors into any vessel as starting point, from where they then flow along the blood with the respective speed.

3.2 Modeling the distribution of nanosensors as a Markov model

The varying location of the flowing nanosensors through the HCS can be modeled as a discrete chain of transitions between vessel segments. For instance, a traveling path can be modeled as the transition $A1 \rightarrow A2 \rightarrow A3 \rightarrow V4$ through the Arcus aorta, Thoratica, and Intercostales aortas to the thorax in Fig. 4. The transition on the bifurcations ($A1 \rightarrow A2$) can be also modeled as a random variable to specify the next visiting stage. That is, a given nanosensor flowing through the Arcus aorta will randomly jump to the Thoratica or to the Subclavia sinistra aorta (cf. $A1$ to $A2$ or $A7$ in Fig. 4). Therefore, when assuming that random transitions in the bifurcations are independent of the previous visited stage, the transitions between vessel segments can be modeled by a Markov chain [14].

Fig. 5 depicts the equivalent Markov chain model according to the vessel segments and transitions from Fig. 4. A total of 30 nodes are defined for the arteries ($A1$ to $A30$), 25 nodes for the...
capillaries (C1 to C25), and 33 nodes for the veins (V1 to V33). Given the direction of the blood flow, the transition probabilities for the nodes are mostly given as 1 except for the transitions between segment vessels on the arteries. This is the case of \( P_{A1,A2} \), used to describe the random jump from the Arcus to the Thoratica aorta. Furthermore, relying on the circuit representation of the Markov chain \[17\], such transition probabilities at the arteries’ bifurcations can be computed based on the flow rate of each vessel segment. For instance, the probability to jump for the Arcus to the Thoratica aorta (A1 \(\rightarrow\) A2) yields [30]

\[
P_{A1,A2} = \frac{I_{A2}}{I_{A1} + I_{A2}},
\]

where \( I_{A1} \) and \( I_{A2} \) denote the flow through the corresponding vessel segments Arcus and Thoratica aortas, respectively. Similar to this relation, the expressions for the remaining transition probabilities can be derived as well [30].

The relation in Eq. (3) can also be estimated by computing the flow of nanosensors when considering that they become part of the blood flow. This can be achieved by identifying the total of nanosensors traveling per closed-loop, as introduced in the current-mesh method in [30]. To illustrate, the total of nanosensors flowing through the vessel segment A2 is given by the number of nanosensors through the variety of loops starting and ending at C1 and passing through A2. This is the case of loops L1 and L2 represented in Fig. 5, as well as the others defined on the lower body region.

Following this reasoning, an alternative way to estimate the corresponding transition probability in Eq. (3) yields

\[
P_{A1,A2} = \frac{\sum_{i \in L_{A2}} N_i}{\sum_{k \in L_{A1}} N_i + \sum_{i \in L_{A2}} N_i},
\]

where \( N_i \) is the total of nanosensors flowing on a given loop \( L_p \), and \( L_{A1} \) and \( L_{A2} \) represent the set of loops which pass through the vessel segments A1 and A2, respectively.

Although the relation in Eq. (4) seems complex to assess at a first glance, its corresponding term \( M_{ij} \) can be readily computed at the gateway. The total of nanosensors per loop can be identified based on their traveling time. Those with the longest traveling times will correspond to the capillaries on the feet (largest path loop), while those with shorter traveling times will correspond to the head (shortest path loop). Based on the collected timestamps, provided by the nanosensors to the gateway, the total of nanosensors per circuit can be identified, and thus their flow per vessel segment.

Considering this methodology, circuits on the right and left sides, accounting for the arms and the legs, will not be resolvable. Due to the symmetry of the human body, the traveling time by the left shoulder will be similar to the one through the right shoulder, for instance. Consequently, the total of nanosensors will be counted twice. Just a total of 14 circuits will be distinguishable according to the upper body (head, upper arms, shoulders, elbows, hands), center body (thorax, spleen, intestine, liver, pelvis), and from the legs (hips, knee, feet), (cf. Fig. 5). To identify the total of nanosensors on both sides, we assume that the distribution of nanosensors will be the same on both (due to the symmetry of the human body). Thus, we just divide by two the total of identified nanosensors for those loops concerning the arms and the legs.

This way, based on these transition probabilities \( p_{i,j} \), the location of a given nanosensor can be derived from the transition matrix \( \Pi = \{p_{i,j}\} \), after solving for the stationary probability vector \( \nu \) in \( \nu = \nu \Pi \). The vector \( \nu \) will have a total of elements equal to the total of stages in the Markov model as 91, where the probability to locate a nanosensor on a given vessel segment \( k \) will be directly given by the vector component as

\[
P_{\nu,k} = \nu_k.
\]

## 4 LEARNING METHODOLOGY

To determine the total number of traveling nanosensors per closed-loop in the HCS, we propose the use of machine learning (ML) methods. Due to their self-learning capacity, ML models can be used to compute the total of flowing nanosensors through the wide range of physiological parameters of the human body (e.g., blood viscosity, vessel lengths, pressure levels) and for a variety of activities (e.g., walking, running, sleeping).
To apply machine learning, we follow the three-steps methodology depicted in Fig. 6. In the first step, we collect the timestamps provided by the nanosensors at the gateway. These timestamps indicate the traveling time on a given loop for one round (cf. Fig. 5). Besides, we assume that nanosensors reset their internal counter after delivering this information to the gateway. In the second step, we apply ML models to self-identify the total of loops and samples per loop. In the last step, we compute the transition probabilities according to Eq. (4), based on the outputs provided by the ML model.

In the first step, samples are collected at the gateway in four different groups (a total of 5882 samples). Through the use of anchor nodes located on the hips, the shoulders, and the head, nanosensors traveling nearby are updated with the corresponding coordinates. As shown in Fig. 7, samples coming from the head, arms, center body, and legs can be distinguished, otherwise overlapped.

In the second step, we combine two ML models from those available in MatLab®. The unsupervised method k-means is used to identify the total of circuits from the collected timestamps at the gateway. Then, the transition probabilities $p_{i,j}$ for the Markov model are computed as shown in Eq. (4). Furthermore, the resulting labeled sequence is used to train a second ML model using Decision Trees. This second model is used to assist an early detection mechanism. Whenever a sample is received, the Decision Tree model will predict the corresponding source location in the human body.

The k-means method has been selected as it is a low-complex solution providing the total number of clusters in advance with a good performance. This corresponds to a total of 13 loops excluding the head (cf. Section 3.2). Its performance is verified by the silhouette diagram depicted in Fig. 8, where the total of clustered samples are close to one on the different body regions, except for the Kidneys and the Spleen. Specifically, in the Center Body the performance is less due to the superposition of received samples. As Fig. 7 c) exhibits, samples coming from the Liver, Kidneys, Intestine, and Spleen are hard to distinguish since their traveling time is pretty similar. However, as presented in the next section, the derived distribution of nanosensors is in accordance with results obtained following different methods, as illustrated in [30, Fig. 9].

The labeled samples are then used to train the Decision Trees model, implemented with 21 nodes [1]. This method is mainly selected due to its low-complexity. The prediction performance is illustrated in Fig. 9, depicting the positive and false negative results per capillaries on the tissues. Results perform well on the upper body and the legs, however, in the center body the false negative is higher than the positive predicted samples for the Liver, the Kidneys, and the Intestine. Samples coming from these body regions are not well distinguishable due to its superposition, as exhibits the histogram plot in Fig. 7 c). Although the method fails to predict the location on these capillaries, still we can distinguish alerts coming from the center body, which in turn target the possible regions to diagnose. As future work, we will analyze further localization methods like fingerprinting [13] or hop counting [10, 28] to improve the localization performance in this specific region.

5 RESULTS

The detection capabilities of the nanosensors, when sensing the released QS molecules by bacteria, will be dependent on the total...

$^{2}$Using the Classification Learner App from MatLab [2].
of nanosensors observing a particular organ will be ultimately defined by their flow in the corresponding vessel segments, and can be described by the transition probabilities given in Eq. (4). We now illustrate the derived detection capabilities of the flowing nanosensors by evaluating Eq. (2). To do so, we first illustrate the stationary probability of nanosensors per capillaries following Eq. (5).

The resulting stationary probability of nanosensors per vessel segment is depicted in Fig. 10. This distribution is derived after computing the transition probabilities using Eq. (4) with the total of nanosensors per circuit computed by the k-means method. As expected, the larger probability to find the nanosensors will be given in those segments where all the loops pass by, i.e., Arcus aorta (A1), heart (C1, C3), lungs (C2), and vena cava (V3). The least probability will be found on the arteries, capillaries, and veins located in the legs. The distribution derived exhibits a similar pattern as shown in [30, Fig. 9]. In both cases, the larger probability is achieved in the arteries and veins directly connected to the heart. Besides, a similar distribution is also obtained for the capillaries located in the center body region.

Fig. 11 depicts the detection probability versus the total of flowing nanosensors in the system. To illustrate, this is obtained when evaluating Eq. (2) in the shoulders and upper arms, although a similar analysis can be conducted in other capillaries as well. As expected, this is a monotonic growing function of the total of nanosensors. Besides, with increasing threshold value, as defined by $N_{th}$ in Eq. (2), the total number of nanosensors required to reach the same probability of detection also increases. Although the velocity of blood flow in capillaries is similar in the shoulders and upper arms, it requires a larger number of nanosensors to detect an infection in the upper arms, due to large volume, and the lower probability of a nanosensor being in upper arms, as depicted in Fig. 10.

6 CONCLUSIONS

In this work, we presented a methodology to evaluate a detection mechanism for infections in the human body. We introduced the use of ML models to predict the location of nanosensors traveling through the blood vessels. Through these mechanisms, we provide a new approach to use the inherent body-networking capabilities to early detect infections. Our approach will have wide applicability for in-body precision medicine solutions due to its adaptability to varying physiological parameters. As future work, we have plan to study the performance of the proposed methodology when considering different activities, e.g., walking, running, sleeping. We will also conceive more accurate probabilistic-based ML methods to distinguish the flowing nanosensors and particular localization mechanisms to improve performance.

ACKNOWLEDGMENTS

Reported research was supported in part by the project MAMOKO funded by the German Federal Ministry of Education and Research (BMBF) under grant number 16KIS0917 as well as the project NaBoCom funded by the German Research Foundation (DFG) under grant number DR 639/21-1.
REFERENCES


