DNA-Based Nanonetwork for Abnormality Detection and Localization in the Human Body

Jorge Torres Gómez, *Senior Member, IEEE,* Bige Deniz Unluturk, *Member, IEEE*, Florian-Lennert Lau, Jennifer Simonjan, *Member, IEEE,* Regine Wendt, Stefan Fischer, and Falko Dressler, *Fellow, IEEE*

***Abstract*—This study presents an innovative deoxyribonucleic acid (DNA)-based nanonetwork designed to detect and localize abnormalities within the human body. The concept for the architecture integrates nanosensors, nanocollectors, and a gateway device, facilitating the detection and communication of disease indicators through molecular and intra-body links. Modeling DNA tiles for signal amplification and fusion rules (`AND`, OR, `MAJORITY`), the system enhances detection accuracy while enabling real-time localization of health anomalies via machine learning models. Extensive simulations demonstrate the efficacy of this approach in the dynamic environment of human vessels, showing promising detection probabilities and minimal false alarms. This research contributes to precision medicine by offering a scalable and efficient method for early disease detection and localization, paving the way for timely interventions and improved healthcare outcomes.**

***Index Terms*—DNA Tiles, Nanosensors, Nano Communication, Human Circulatory System, Precision Medicine, Machine Learning, Markov Model**

I. Introduction

NANOTECHNOLOGIES for healthcare practices may read as a sensational story, but various research effects a closing gap between possibility and reality. Nanomaterials are becoming a technological innovation in the pharmaceutical sector, improving the efficiency of drugs [1] and allowing for manufacturing nanothings for medical diagnosis [2]. Biofunctionalized nanosensors provide alerts on specific biomarkers, enabling the detection of cancer tumors or pathogens like viral or bacterial infections in the early stage [3]. However, isolated nanosensors only have limited sensing and actuation capabilities. Therefore, interconnection internet of bio-nanothings (IoBNT) platforms for nanosensors among themselves and to out-of-body monitors are crucial to boosting the practical benefits of this new technology.

Basically, today, there are two architectures to deploy IoBNT platforms [4], namely (i) reporting abnormal detection directly to an external gateway (sensor-gateway architecture) [5] and (ii) via a fusion node (FC) node (sensor-fusion node-gateway architecture) [6]. Reporting through a FC node, local detection capabilities of biomarkers are studied in [6]–[8] along specific vessel segments. However, little research has been conducted to account for the mobility of nanosensors and FC nodes together within the bloodstream. Reported literature assumes the fixed location of fusion nodes in the human vessels; see [9], [10], or models local vessel segments without evaluating the human circulatory system (HCS) as a whole system; see, for instance, [6], [7]. In a more realistic scenario, the blood flow embeds a time-variable character on nanonetworks' detection and localization capabilities impacting the detection and localization performance.

Including the dynamic of the blood flow in the HCS, this paper aims to evaluate the IoBNT architecture targeting the detection and localization of abnormalities in the human body. Based on a sensor-fusion node-gateway architecture, the nanonetwork comprises nanosensors and nanocollectors (performing as FC nodes) continuously flowing along the human vessels. The role of the nanosensors and nanocollectors is to realize the detection of abnormalities. We also conceive an external gateway attached to the surface of the skin, to

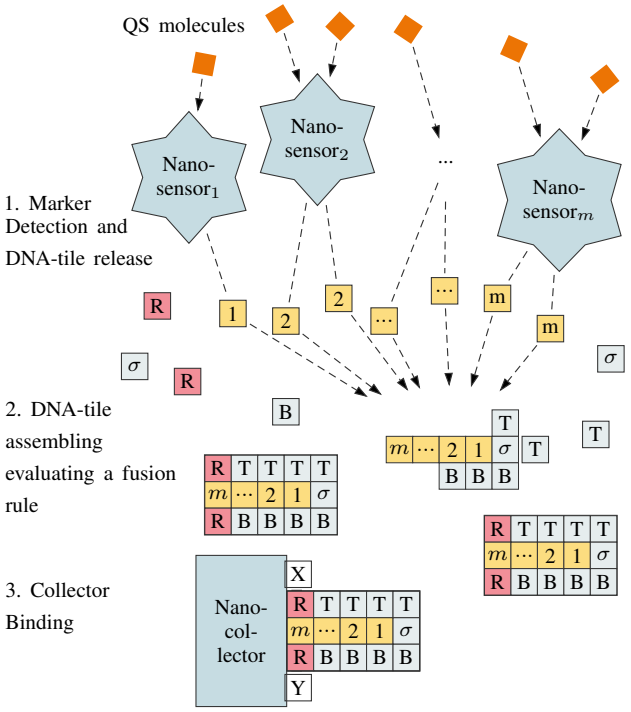

Fig. 1: The process of computing a fusion rule using DNA nanosensor networks using an m-bit `AND` as an example.

Jorge Torres and Falko Dressler are with TU Berlin, Einsteinufer 25, FT 5, 10587 Berlin, Germany, e-mail: {torres-gomez,dressler}@ccs-labs.org

Jennifer Simonjan is with Technology Innovation Institute, Masdar City, Abu Dhabi, UAE, e-mail: Jennifer.Simonjan@tii.ae

Bige Deniz Unluturk is with Michigan State University, 775 Woodlot Dr, East Lansing, MI, USA, e-mail: unluturk@msu.edu

Florian-Lennert Lau, Regine Wendt, and Stefan Fischer are with Universität zu Lübeck, Germany, e-mail: {lau,wendt,fischer}@itm.uni-luebeck.de

Manuscript received March 28, 2024; revised Month Day, Year.

display the detection results and estimate the abnormality's localization; see a representation in [11, Fig. 1].

Within this nanonetwork architecture, we illustrate the detection and localization of quorum sensing (QS) molecules released by harmful bacteria (abnormality) at a given body region. As visualized in Fig. 1, the nanosensors release deoxyribonucleic acid (DNA)-tiles as indicators of their presence upon detection. The released DNA-tiles will also have self-assembly capabilities while flowing in the bloodstream [12], allowing to compute predefined fusion rules *on the fly*, leading to reduced detection errors. Nanocollectors will collect the assembled DNA-tiles to report detection results to the external gateway later.

The sensor-fusion node-gateway architecture will result in a more modular design for deployment than the sensor-gateway architecture, as detection and reporting tasks are split apart. Nanocollectors solely collect and report data to the external gateway using terahertz (THz), optical, or ultrasonic waveforms. Nanosensors solely detect biomarkers and report their detection to the nanocollectors using molecular communication, with message molecules assembled from DNA-tiles [13]. In this way, the network lifetime can also be potentially extended as fewer nanothings perform the more expensive communication link with the gateway. This contrasts the sensor-gateway architecture, where all the nanosensors emit their detection.

Extending the previous work in [11], we research the detection and localization capabilities for the sensor-fusion node-gateway architecture, where nanosensors and fusion nodes (nanocollectors) are passively driven by the blood flow. Our main contributions can be summarized as follows:

- We formulate the detection and localization capabilities of nanosensor-fusion node-gateway architecture with a probabilistic scheme.
- We report a fusion mechanism with DNA-tiles to implement `OR`, `AND` and `MAJORITY` fusion rules.
- We investigate the localization capabilities of abnormalities with machine learning (ML) methods. We comparatively illustrate the performance of low complex unsupervised methods like k-means to predict the localization of abnormalities.

The remainder of this paper is structured as follows: Section II discusses in further detail the previous work on abnormalities detection architectures. Section III introduces the system model and explains all the components of the proposed nanoscale system. Section IV delves into the collaborative detection and reporting to the external gateway. Section V introduces the ML method to localize the reported abnormality. Section VI evaluates our approach using simulations and analytical methods to derive the success probability of the system. Finally, Section VIII summarizes the concluding remarks.

II. Related Work

Future health applications foresee deploying nanosensors inside the human body, enabling earlier and more precise disease detection. Detecting and localizing abnormalities with sensors floating inside the HCS is a highly complex task due to severe resource constraints, high mobility, and limited communication capabilities. Therefore, traditional localization schemes cannot be applied, and new approaches towards in-body sensor localization are under development [4]. To the best of our knowledge, there are very few attempts towards localize and track in-body nanosensors and abnormalities in human vessels. Amidst the reported literature, common factors include the unrealistic assumption concerning the fixed location of fusion nodes in human vessels. Besides, simulation models are limited to focusing on local vessel segments instead of comprising the complete human cardiovascular system. In this Section, we discuss related work on in-body nanosensors, abnormality detection, and localization techniques.

Addressing a local vessel segment, Varshney et al. [6] report detecting abnormalities using nanosensors and fusion centers. Each nanosensor performs abnormality detection with a certain detection and false alarm probability in their work. Nanosensors report their local decisions to a FC node over a diffusion-advection blood flow channel. This study reports using different types of molecules in the presence of inter-symbol interference, multi-source interference, and counting errors. The FC employs the `OR` and `AND` logic-based fusion rules to make the final decision after decoding the local decisions.

The work of Rogers and Koh [7] considers a detection environment in the capillaries, where biomarkers flow from fixed source locations to the nanosensors, distributed as a cone pattern along the flow direction. A FC node is used to merge data from nanosensors detection evaluating performance with the receiver operating characteristics (ROCs) curves. From the various fusion rules, the counting one is preferred because it balances complexity and detection performance.

Mosayebi et al. [8] studied anomaly detection inside blood vessels where nanosensors sense the presence of an anomaly by detecting the biomarkers secreted by cancer cells. The final decision regarding the presence of an anomaly is made at a FC node based on the activation levels of the observed nanosensors. The authors first considered a single cancerous blood vessel from which they deduced the spatial distribution of the biomarkers as a function of time. Then, the study extends to the HCS and analyzes the distribution of biomarkers in the connected blood vessels. The authors derive the optimal decision rule and a simple sum detector for the FC and compare their performances. The study was only able to model a sample network of the HCS containing 16 edges and 5 nodes, and they did not explore how to localize the body region in which the cancerous cells have been detected.

Khaloopour et al. [9] propose a theoretical framework for cooperative abnormality detection and localization via molecular communications. The system consists of mobile sensors in a fluidic medium, which are injected into the medium to search the environment for an abnormality. Some FCs are placed at specific locations in the medium, which absorb all sensors that arrive there, and by observing its state, each FC decides on the abnormality's existence and its location. In a real-world scenario, where sensors will be deployed inside the HCS, the assumption of static FCs with known positions is quite unrealistic and difficult to achieve.

Vasisht et al. [10] introduced a backscatter design, which

is particularly customized to work inside human tissues. It considers the interference generated by the body in the intrabody communication link and develops the localization of devices via time-of-flight signal estimation. The authors implemented and tested their approach in animal tissue with an average localization error of $1.4\,\mathrm{cm}$. This work brings forward the ideas of communication and localization inside the human body. However, it considers the setup of a static implant rather than flowing nanosensors.

Enabling the localization and communication with nanosensors through THz backscattering is another approach recently proposed in [14]. The work assumes that each nanosensor can communicate via electromagnetic communication ($0.5 - 1.5\,\mathrm{THz}$) with the anchor nodes outside the body. Additionally, backscattering communication in tissue suffers an extremely high signal power loss which restricts the communication distance to only a few mm.

Lemic et al. [15] have outlined an architecture for in-body THz-operating nanosensors and integrating energy-harvesting schemes. Their goal was to enable localization of the passively flowing and energy-harvesting nanosensors and their two-way communication with the outside world. To do so, the authors propose to utilize location-aware and wake-up radio (WuR)-based wireless nanocommunication paradigms, as well as software-defined metamaterialss (SDMs). As a simulation model, they use a $1\,\mathrm{cm}$ thick slice of the torso area, which is approximated by a circle with the radius of $30\,\mathrm{cm}$.

In line with the published survey by Etemadi et al. [4], present work still miss to include end-to-end realistic system models accounting for the HCS as a whole. In this direction, our previous work [11] researched how to model the traveling paths of nanosensors along the HCS and how to estimate nanosensors' locations based on their traveling time and concentration level. This approach assumes that each nanosensor reports its sensor readings to an external gateway via electromagnetic communication.

However, since nanonetworks for health applications will employ hundreds or even thousands of nanosensors, see [16, Table II], we research an approach to reduce the amount of communication required to make the system operable. We extend our previous system in [11] including FC nodes that collect the nanosensor readings through DNA-based communication. As we employ a small amount of nanocollectors, this architecture dramatically reduces the costly electromagnetic communication reports to the external gateway. Next sections detail the network architecture and formulate performance with the corresponding detection and localization probabilities.

III. Detection and Localization Network Architecture

This section explains the overarching system architecture of the proposed detector and localization network. We describe the architectural components and their communication capabilities as nanothings. This section also includes the flow diagram for detecting and localizing abnormalities and the modeling for the flow of nanosensors and nanocollectors through the HCS.

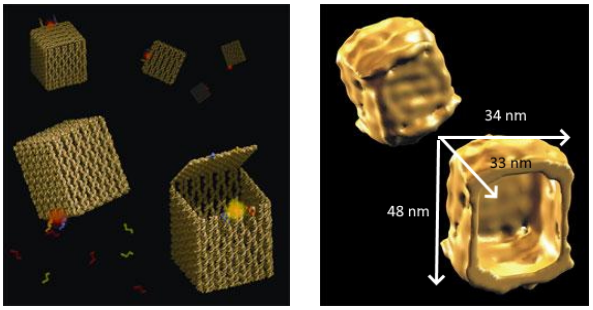

Fig. 2: A controllable DNA box concept (left) and an atomic force image (right). [17]

A. Architecture components

Five components comprise the detection and localization architecture listed as the nanosensors, the molecular communication (MC) channel, nanocollectors, the external gateway, and the intra-body communication link between the nanocollectors and the gateway. Within this architecture, information flows via the MC channel from bacteria to nanosensors using QS molecules and from nanosensors to nanocollectors in the local vessel segment using DNA-tiles.[1] Finally, the information is carried by the nanocollectors through the HCS and finds its destination at the external gateway via the intra-body communication link. The component and their functionalities are listed below:

1) Nanosensors: We assume nanosensors passively flow with the blood in the HCS. Nanosensors are equipped with biosensors to detect QS molecules. Upon detection, nanosensors will release DNA-tiles to the medium, indicating the presence of abnormalities. To illustrate, nanosensors can be conceived as boxes (see Fig. 2) pre-filled with a capacity to store 450 DNA molecules [12]. Those boxes can be opened with specific marks like the QS molecules. Once the QS molecules bind to a lid (represented as red spots in the boxes in Fig. 2), the box opens and release the DNA tiles to the medium; in this way implementing the bio-sensor mechanism.

2) Molecular communication channel between nanosensors and nanocollectors: The nanosensors communicate with the nanocollectors via the MC channel in the blood vessels. On the one hand, the communication can be direct: nanosensors will release DNA-tiles into the blood flow, propagating by convection from nanosensors to the nanocollectors. We assume that the nanocollector detects the presence of DNA-tiles whenever nanosensors and nanocollectors are located in the same vessel segment. This occurs with some probability; its calculation is illustrated later in Section IV.

Besides, the DNA-tiles in the channel will also be able to compute pre-defined fusion rules for more complex detection procedures. The tiles are designed in a way that they can compute the detection rules `AND`, `OR`, and `MAJ` rules, by self-assembling the released DNA-tiles from the various nanosensors. Nanocollectors can then detect the presence of

[1]We remark that the link between nanosensors and nanocollectors occurs in the small-scale distance, whenever nanosensors and nanocollectors are in the same vessel segment. Further details are given in Section IV.

these complex structures just the same way as single DNA-tiles. This process is the one interesting to implement fusion rules, and it is explained in more detail in Section IV-A.

3) Nanocollectors: In this architecture, we conceive nanocollectors performing as FC nodes upon detecting assembled DNA-tiles. The fusion is already computed with the assembled DNA-tiles, and nanocollectors just detect their presence in the blood flow. For instance, a nanocollector detecting assembled DNA-tiles, these implementing the AND rule among N_s nanosensors, will report that N_s nanosensors detected the presence of QS molecules; see further details in Section IV. The nanocollector will also account for the time-instant when this event happens. We assume the nanocollector has an internal clock with $1\,\mathrm{s}$ period, and also report their counting with the detection event. This timestamp will be used later in the gateway as a metric to localize the source of QS molecules in the body (see Section V). The nanocollector will also be able to report the total of neighboring nanocollectors along the vessel segments. We will assume that nanocollectors emit *beacon* messages, with low-power electromagnetic pulses, that reach other nanocollectors in the same vessel segment. Such a beacon signal will allow detection of the presence of neighboring nanosensors later, allowing to report the concentration of those along their path. The concentration is also used as a feature to localize the QS source later (further details in Section V).

4) Gateway: The gateway is basically a computing device that is attached to the body in the skin's surface. Gateways collect all the nanocollectors' report through electro magnetic (EM) [18] or ultrasonic communication links [19], for instance. When the nanocollector is in the vicinity of the gateway, it reports the reading, and upon reception, the gateway performs the ML methods proposed in Section V to localize the abnormality. As the gateway device has significantly higher processing power and memory, all data can be easily collected and processed there.[2]

5) Intra-body communication link between nanocollectors and the gateway: Nanocollectors collect the molecular message and transmit the information to the gateway outside of the body. For this transmission, we consider an ideal link error-free, as we want to isolate the network's fusion and localization capabilities from communication errors.

B. Assumptions for the architecture deployment

In practice, we need to consider electromagnetic communication, optical [20], [21], ultrasound [19], [22], [23] or (THz) communication [14], [24] to enable a communication link between the nanothings and the external gateway. These communication links have been demonstrated as potential technologies to be used for intra-body communication scenarios. However, THz communication suffers from high molecular absorption noise, limiting the communication distance to a few mm. A THz link will impose constraints on the placement of the gateway device, allowing for locations where the intrabody distance between the skin surface and nanocollectors in the human vessels is shorter, such as the wrists. Compared to that, ultrasound and optical communication can cover longer distances with significantly less path loss. It is subject to advances in nanotechnology to develop nanoscale antennas that can communicate in these bands.

Furthermore, we assume the deployed nanothings are biocompatible with the human body. As for the nanosensors, they are implemented with hydrophobic molecules, i.e., long DNA strands self-assemble boxes-like structures through a process called DNA-origami [25]. As for the nanocollectors, we assume no direct contact between radiator elements (antennas, optical interfaces) or ultrasonic generators with the blood. To avoid the potential toxicity to the human body produced by contact of blood with these elements, we assume these radiators and generators surfaces are coated with the biocompatible material polydimethylsiloxane (PDMS).[3]

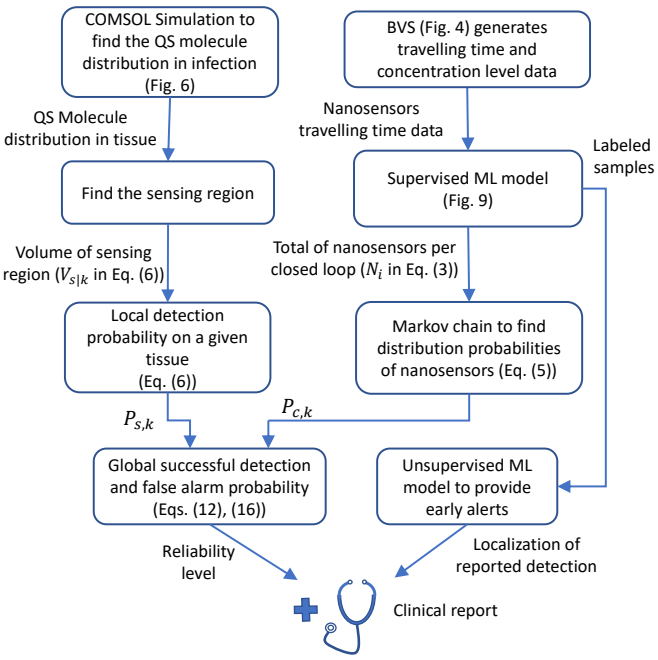

Fig. 3: Flow diagram to detect and localize abnormalities in the HCS.

C. Detection and localization flow diagram

Using the components above, we devise a detection and localization scheme as depicted in the flow diagram in Fig. 3, which follows our previous work in [11]. As the primary outcome of this flow, we aim to provide early alerts and the location of abnormalities, including its reliability with the successful detection probability as a metric.

We implement the flow diagram with two branches to evaluate the detection probability. The left branch computes the local detection probability of an abnormality on a given tissue, assuming that nanosensors and nanocollectors (nanothings) are in the region of interest. We evaluate this probability using

[2]We remark, that the processing in the gateway device can be realized in a remote monitor device instead. That is, when the gateway comprises an interface to the Internet, samples can be transferred and processed at a remote location.

[3]PDMS is a bio-compatible material typically used for implanted antennas; see [26], [27].

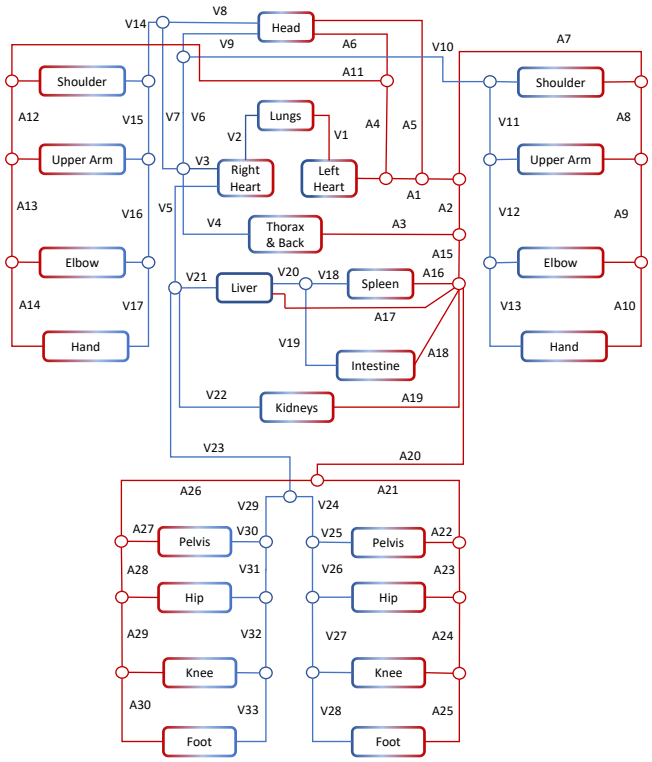

Fig. 4: Human circulatory systems represented in the BVS simulation framework.

COMSOL simulations and formulating analytic expressions for the detection probability. Further details are given in Section IV.

We evaluate the total detection probability when including the probability of nanothings in the region of interest (as evaluated with the right branch). To evaluate the probability of finding nanothings in the region of interest, we model their time-varying location as a Markov model. As explained in Section III-E, the traveling path of nanothings is assumed to follow a markovian process. The Markov transition probabilities are found with an unsupervised ML model using data reported by the BloodVoyagerS (BVS) simulator [28], as explained in Section III-D. We estimate the number of nanothings per tissue with the corresponding stationary vector probability through the Markov formulation. In this way we complete the formulation for the total successful detection probability in Section IV.

The flow diagram in Fig. 3 also includes the localization of the abnormality. Towards its estimation, we will train a second ML-model (supervised) with the labeled samples from the unsupervised ML model, as explained in Section V-C. The samples consist of traveling time and concentration of neighbor nanothings along the traveling path of nanocollectors; the labels will denote the traveling tissue. The trained, supervised method will provide the location of detected abnormalities upon the nanocollector report.

D. Nanosensors and Nanocollectors Mobility Model

We model the mobility of the nanothings in the HCS with the BVS simulator. BVS is a model of the human cardiovascular system that simulates the mobility of nanothings traveling in the vessels [28]. We implemented the mobility of nanosensors and nanocollectors in BVS along 21 different streams in 94 vessels segments; see a representation of vessels in [11, Fig. 4]. Each stream implements a given speed emulating the laminar flow profile in the vessels [29]. BVS implements a constant blood speed along the human vessels, with the parameters listed in Table I. This is a realistic assumption for the capillaries and veins, where the pressure is rather constant, see [30, Fig. 14-2 pp. 172]. However, pressure in the arteries varies with time, influencing the blood speed. This limitation will impact mostly the localization performance, as later discussed in Section VII.

TABLE I: Simulation parameters.

Variable	Description	Value
v_{A}	Arteries' speed	$0.1\,\mathrm{m/s}$
v_{C}	Capillaries' speed	$0.01\,\mathrm{m/s}$
v_{V}	Veins' speed	$0.037\,\mathrm{m/s}$
r_{ns}	Nanosensors' radius	$22\,\mathrm{nm}$
r_{nc}	Nanocollectors' radius	$500\,\mathrm{nm}$

We selected the nanosensor and nanocollector sizes according to the nanometer scale of reported nanotechnologies; see [16]. As for the blood speed blood speed, these are the values coded at the BVS simulator, see https://github.com/RegineWendt/blood-voyager-s/tree/master, which follow the range of reported values in [31, 31, Table I].

Dragged by the blood flow, we implement the traveling nanothings' speed with their size and geometry. According to Stokes' law, the velocity v at which a sphere moves by the action of a force F is given by $v = \frac{F}{f}$, where f is the frictional drag coefficient [29, Eq. (4.2)].[4]

In a given vessel segment, the net force due to a pressure difference at the ends yields $F = \Delta P\, L \Delta h$, where L is the length of the vessel and Δh denotes a thin layer across the vessel where the fluid speed is constant (cf. Fig. 5) [29]. Furthermore, modeling the nanothings as spheres, where $f = 6\pi\eta r$, the resulting speed will be given by

$$v = \frac{\Delta P\, L \Delta h}{6\pi\eta r}, \tag{1}$$

where η is the fluid's dynamic viscosity, and r is the sphere's radius. Using this relation for a given vessel segment, the resulting nanocollectors' speed is related to that of the nanosensors as a factor of

$$\frac{v_{\mathrm{nc}}}{v_{\mathrm{ns}}} = \frac{r_{\mathrm{ns}}}{r_{\mathrm{nc}}}. \tag{2}$$

With the above formulation, BVS is expanded in such a way that any number of nanocollectors and nanosensors can be simulated with their respective speed. The simulation in BVS produces the raw data to later train the ML models (see the right branch in Fig. 3).

E. Nanosensors and Nanocollectors Markov Model

We use the Markov model to compute the stationary distribution of nanothings in the HCS and evaluate the probability $P_{c|k}$ represented in Fig. 3. The varying positions of the nanothings might be modeled as a Markovian process, where the states

[4]The law is valid in incompressible liquids like the blood in the vessels.

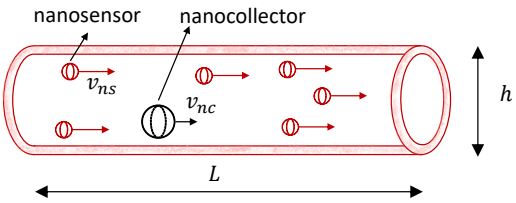

Fig. 5: Representation of traveling nanosensors and nanocollectors inside the blood vessels.

represent the vessel segments as in [32]. The 94 vessel segments represented in BVS are directly mapped into the same number of states; see a representation in [11, Fig. 4]. In this way, the movement of the nanothing can be modeled through random transitions between these states as it moves in the HCS; see further details in our previous work in [33, Fig. 4].

We estimate the transition probabilities with the number of nanothings flowing in the HCS [34, Eq. (1)]. According to their flow in the bifurcations, the transition probability might be evaluated as the corresponding ratio of flows. Having the transition matrix $\Pi = \{p_{i,j}\}$ of the Markov model, we later evaluate the stationary probability vector $\boldsymbol{\nu}$ after solving the equation

$$\boldsymbol{\nu} = \boldsymbol{\nu}\Pi, \tag{3}$$

where the components of ν_k will readily provide the probability to locate a nanothing in the given vessel segment k as

$$P_k = \nu_k. \tag{4}$$

IV. Detection Scheme

The abnormalities in the body are detected in three steps as follows from Fig. 1. In the first step, the nanosensors locally detect the presence of QS molecules (marker detection in Fig. 1) and release DNA-tiles upon detection. Secondly, the released DNA-tiles self-assemble and realize a given fusion rule. In the last step, nanocollectors detect the presence of the assembled DNA-tiles reporting the global detection result to the external gateway. This section details these three steps and provides closed-form expressions for local and global detection probabilities.

We take a probabilistic approach for the detection scheme since when nanosensors are injected into the body, they will follow a random path while travelling, which may or may not pass through the infection region. As a first step, using the BVS simulator and the Markov Chain approach, we calculate the probability of a nanosensor being in the organ of infection representing the right branch of Fig. 3. As a second step, since the BVS simulator operates on the scale of organ and the infection sensing region is much smaller than an organ, we need to calculate the probability that a nanosensor is visiting the sensing region when we determine it is in that organ representing the left branch of Fig. 3. As the last step, we need to calculate the global probability of detection where we aggregate the data from nanosensors at the nanocollectors which then report the aggregated data to the gateway. In this layered architecture, the global probability of successful detection

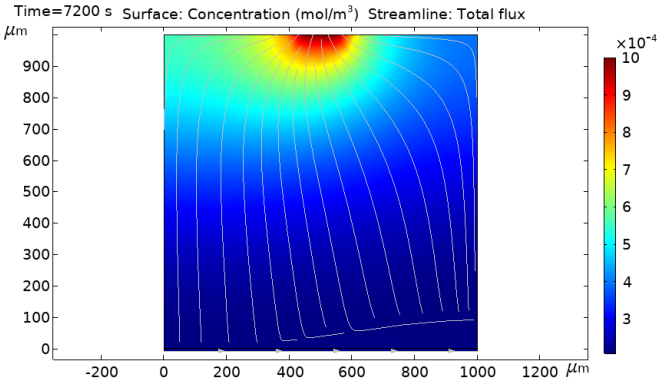

Fig. 6: Concentration of QS molecules in the tissue as provided by COMSOL simulator.

depends on the probability of detection of a single nanosensor and how the data of multiple nanosensors is aggregated, i.e., fusion rules. This layered architecture allows us to utilize less nanocollectors equipped with communication capabilities that can report to the gateway and simpler and more nanosensors patrolling the body and communicating with the nanocollectors only.

Bacteria communicate with each other using QS molecules to establish coordination among them. Since infection in the body is the rapid reproduction of pathogenic bacteria, by eavesdropping to QS it is possible to detect the severity of infection. To carry out this functionality, mobile nanosensors can be equipped with electrochemical sensors measuring QS molecules of *Pseudomonas aeruginosa* pyocyanin [35] and N-Acyl homoserine lactones (AHL) [36] through electrodes.

In our model, adapted from our previous work in [37], the distribution of QS molecules in the tissue follows diffusion in a porous medium where the interstitial space is considered as pores through which QS molecules are diffusing. A representative tissue volume is considered as the simulation domain where the infection is assumed to be on top, and the capillaries are at the bottom of this unit volume.

Using the COMSOL simulations, we determine the region where the concentration of QS molecules is above the threshold that the nanosensors can detect, which we call *sensing region*. Assuming the detection limit of nanosensors to be 1×10^{-5} mol/L [38], we use an ellipse to approximate the sensing region which is elongated in the blood flow direction with axes lengths of 0.1-0.5 cm and 1.75-3.5 cm.

We assume that the nanosensor locally detects the presence of the QS molecules whenever it is located in the sensing region (of volume $V_{s,k}$) and conditioned to be in the given blood vessel segment (of volume V_k), where $V_{s,k} \subset V_k$. Accordingly, the conditional probability can be evaluated as the ratio of these two volume metrics as

$$P_{s|k} = \text{Prob}(\vec{l_n} \in V_{s,k} | l_n \in V_k) = \frac{V_{s,k}}{V_k}, \tag{5}$$

where $\vec{l_n}$ is the (x, y)-location of nanosensors.

A. DNA Fusion Rule Mechanism

Once nanosensors detect the presence of QS molecules, they release DNA-tiles as an indicator. These tiles have been designed in a way that they can compute fusion rules like OR, AND, and MAJORITY. These rules are implemented by resorting to the inherent binding properties of DNA. The molecule consists of four amino acids: adenine, thymine, cytosine, and guanine. Interestingly, adenine and thymine as well as cytosine and guanine automatically form a stable binding in a process called self-assembly. By artificially manufacturing DNA strands, the self-assembly process of DNA can be controlled to implement these rules. The temperature stability of the binding follows the Wallace rule where each A or T contributes $2\,^\circ\mathrm{C}$ of temperature stability and every C or G adds $4\,^\circ\mathrm{C}$. As such, suitable tiles may be designed for the most realistic temperature ranges.

This technology is even capable of computing mathematical functions and entire nanonetworks may be created from DNA tiles that have already been created and used in wet-lab experiments [12]. For instance, as depicted in Fig. 1, the released DNA strands by the nanosensors (blocks 1 to m) are manufactured in a way that the sequence of strands 1, 2 to m can only bind with the corresponding neighbor molecules T, σ, B, and R and with the specific order illustrated in the figure. This block, will be the one detected by the nanocollector, as the sequence of molecules R, m, and R will be the ones to bind to the nanocollector receptor.

We now briefly introduce the DNA-based sensor systems we propose and explain in depth how they can also compute fusion rules.[5] A fusion rule is a function that is directly executed on the measured data of a sensor network. It is akin to a preprocessing step to separate valid measurements from false alarms.

Since the nanosensors are envisioned to be simple devices to detect the presence of infection rather than an its level, they operate in a binary fashion, i.e., they either generate an alert for detecting infection or not rather than reporting a continuous measurement. To aggregate their binary data at the nanocollector, we use several different kinds of fusion rules that are also demonstrated to be possible using DNA tiles as a communication mechanism between the nanosensors and the nanocollector.

We consider the:

1) OR-rule,
2) AND-rule,
3) MAJORITY-rule.

using DNA-based self-assembly mechanisms.

The OR-rule, AND-rule, and MAJORITY-rules describe how many different types of nanosensors have to agree upon the detection of a disease before the entire system is activated. For the OR-rule, a single nanosensor would be sufficient but false positive detections might frequently occur. In this case, each nanosensor emits a single unique tile that may be collected. The AND-rule requires n nanosensors to agree upon a detection. Given a suitable n, false positives become arbitrarily unlikely. The physical implementation is illustrated in Fig. 1. The MAJORITY-rule is a compromise between OR and AND and requires a subset of nanosensors to agree.

As the physical implementation is rather complex, we only offer an intuition; see further details in [39]. For each possible way to match a threshold, we construct a unique message molecule that behaves akin to an AND. As a single way to exceed a threshold is sufficient, the concurrent presence of possibly different message molecules works akin to a logical OR. Deciding on a specific fusion rule and respective input sizes allows us to fine-tune the specificity of the system.

An example of the entire process from detection to fusion can be seen in Fig. 1, where the AND-rule is computed upon detecting certain markers based on the approach in [12]. The fusion rule's computing process is separated into four phases: marker detection, tile release, fusion, and collector binding.

Upon detection of the QS molecules (Phase 1 in Fig. 1), each nanosensor releases DNA-tiles out of m-distinct types of *tiles* (Phase 2). The released tiles will then compute a preprogrammed fusion rule function by assembling it into a message molecule, in this case, an m-bit-AND fusion rule. Only when all m specific tile types are present can the *message molecule* in the middle fully form, including a receptor (R block in Fig. 1). The very special property of this process is that the computation occurs inside a message molecule and not inside a nanocollector. Notice that the message molecule can only be detected by a nanocollector if the fusion has been fully computed/assembled without errors (Phase 3 in Fig. 1).

With this binding mechanism, the other two rules, OR and MAJORITY, can be implemented as well; see [39]. The OR-rule is the simplest type of fusion rule. Just a single positive measurement of any of the nanosensors detecting the QS molecules is sufficient to signal the detection of a marker or disease. In this case, no real computation is necessary and the single tile just acts as a regular molecule for communication. As a result, there are no additional sources for errors in the further propagation of a detection.

The MAJORITY rule can be implemented with a threshold mechanism, as shown in [39]. If a pre-defined threshold is exceeded, all involved devices can decide on a single course of action. As an example, a threshold of three out of four possible tiles is used to compute the MAJORITY-rule. A distinct message molecule for all relevant combinations of tiles that reach the defined threshold is created. Once a single message molecule is finished, the threshold has been met and the nanocollector can collect the DNA-tiles.

The fifth combination that includes all four tiles is unnecessary as a smaller message molecule that includes just three would be collected before. Simulations conducted in [39] have shown that a 3.5 % error is to be expected while the assembly time is estimated to be nine minutes. That said, the process can be much faster/slower depending on the ambient temperature and the concentration of tiles.

B. Global probability of successful detection

To compute the global detection probability, we will keep the total number of injected nanothings (nanocollectors and

[5]An extensive definition framework is beyond the scope of this paper and we refer the interested reader to [12].

nanosensors) a constant, here denoted as $N_t = N_s + N_{nc}$, where the ratio of nanocollectors to nanosensors is evaluated as $\eta = N_{nc}/N_t$. We assume that all the nanosensors in a given vessel segment can connect to all the nanocollectors in the same vessel segment, where the number of nanosensors in the vessel segment can be found as

$$N_s^k = (1-\eta) N_t P_k. \tag{6}$$

where P_k is the probability of finding a nanonothing in the given vessel segment k, and evaluated with their stationary distribution as indicated in Eq. (4). Similarly, the number of nanocollectors in the vessel segment k can be found as

$$N_{nc}^k = \eta N_t P_k. \tag{7}$$

In both equations, we find the number of nanosensors or nanocollectors multiplying the total amount with the probability of them being in that vessel segment k.

For simplicity, we also assume zero errors in the assembling process of DNA-tiles. As for the OR-rule

$$P_{nc}^{\text{OR}} = 1 - \prod_{i=1}^{N_c^k} (1 - P_{s|k}), \tag{8}$$

for the AND-rule:

$$P_{nc}^{\text{AND}} = \prod_{i=1}^{N_c^k} P_{s|k}, \tag{9}$$

and as for the MAJORITY-rule:

$$P_{nc}^{\text{MAJ}} = \sum_{i=0}^{N_c^k - \lceil \frac{N_c^k}{2} \rceil} \binom{N_c^k}{\lceil \frac{N_c^k}{2} \rceil} (1 - P_{s|k})^{N_c^k - \lceil \frac{N_c^k}{2} \rceil - i} P_{s|k}^{\lceil \frac{N_c^k}{2} \rceil + i}. \tag{10}$$

where $P_{s|k}$ denotes the local detection probability, as given in Eq. (5).

Using these expressions, we calculate the probability of successful detection if at least one of the nanocollectors reports a detection as

$$P_{d,k}^{\text{X}} = 1 - (1 - P_{nc}^{\text{X}})^{N_{nc}^k}. \tag{11}$$

where X refers to OR, AND, or MAJ to represent the mode of fusion can be found as.

Although very unlikely due to the specific binding of QS molecules to the sensor, there is still a probability that the nanosensors will falsely detect due to the randomness of the reading process. To incorporate this effect, we will assume the false alarm constant of value 10^{-4} [37]. The false alarm probability at a single nanocollector, $P_{nc,f}^{\text{X}}$, can be found by replacing $P_{s|k}$ with P_f in the above expressions.

As for the OR-rule

$$P_{nc,f}^{\text{OR}} = 1 - \prod_{i=1}^{N_c^k} (1 - P_f), \tag{12}$$

for the AND-rule

$$P_{nc,f}^{\text{AND}} = \prod_{i=1}^{N_c^k} P_f, \tag{13}$$

and as for the MAJORITY-rule

$$P_{nc,f}^{\text{MAJ}} = \sum_{i=0}^{N_c^k - \lceil \frac{N_c^k}{2} \rceil} \binom{N_c^k}{\lceil \frac{N_c^k}{2} \rceil} (1 - P_f)^{N_c^k - \lceil \frac{M}{2} \rceil - i} P_f^{\lceil \frac{N_c^k}{2} \rceil + i}. \tag{14}$$

Similar to the probability of detection, for the false alarm calculation, we assume an alarm is produced whenever a single nanocollector reports an anomaly, yielding

$$P_{f,k}^{\text{X}} = 1 - \left(1 - P_{nc,f}^{\text{X}}\right)^{N_{nc}^k}. \tag{15}$$

V. Localization Scheme using Machine Learning

We use ML models to localize the source of abnormality upon nanocolectors report to the gateway. The ML model is trained with the traveling time and concentration level as delivered by nanocollectors. Similarly to our previous work in [11], the reported data is used to distinguish the nanothings' traveling circuit, the Markov model's transition probabilities, and also to provide the location for detected abnormalities (see the right branch in Fig. 3).

Two ML models are used as unsupervised and supervised methods. As represented in Fig. 7, the raw data is generated with BVS in Step 1. In Steps 2 and 3, the unsupervised model evaluates the transition probabilities for the Markov model on the one hand. On the other hand, it is used to train the supervised model and provide early alerts with the detection and localization of the abnormality, as represented in Step 4.

The unsupervised model distinguishes the different data sets (clusters) according to each tissue. Each cluster will correspond to the human body's tissues, e.g., spleen, liver, and intestine, as given by BVS. The supervised model will be trained to promptly distinguish the tissue once an abnormality is reported using the labeled samples from the unsupervised method, as represented by Step 4 in Fig. 7. That is, whenever a nanocollector reports a detection, the supervised method will provide the location of the source, e.g., on the liver, hands, feet, etc.

We comparatively illustrate results between low and high complex clustering ML methods. Specifically, we develop in this section the unsupervised methods k-means (low complex) and the deep neural network (NN) self-organizing feature map (SOFM) (high complex) method for comparison purposes. As illustrated in the section below, both methods perform similarly regarding the positive predicted and false negative metrics, allowing us to recommend deploying the k-means method. As the supervised method, we implement the Decision Tree algorithm, which results in a high-accuracy and low-complexity classifier.

A. Training dataset

The ML model is trained with two datasets: the traveling time of nanocollectors and their concentration level along the vessel segments. As for the concentration level, we refer to the total of neighbor nanosensors around each nanocollector per vessel segment. In this regard, we assume that nanocollectors might also detect the presence of neighbor nanosensors while traveling (see Section III-A3).

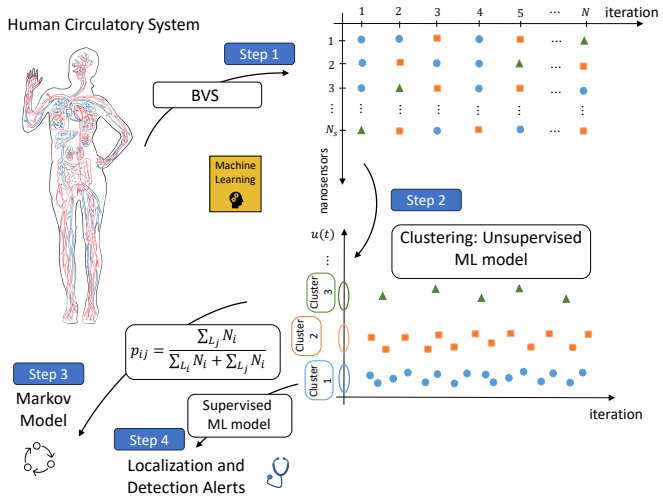

Fig. 7: Interactions between the different components in the system model with the ML models.

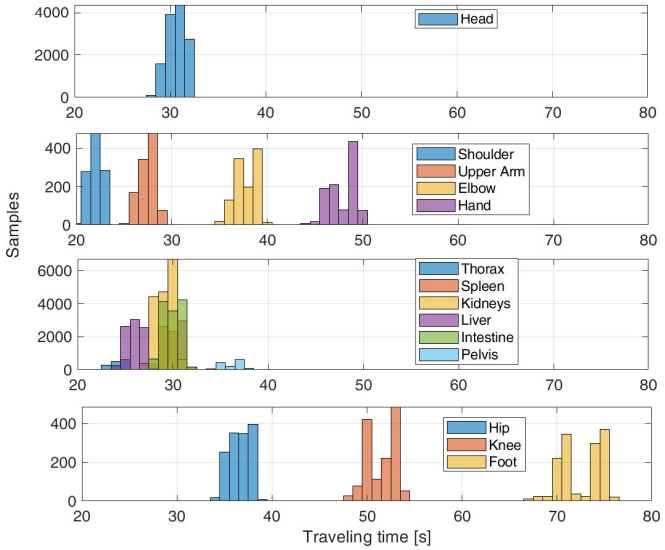

Fig. 8: Representation of the collected data for the traveling time along vessels, as provided by nanocollectors to the gateway device.

We process the output data from the BVS simulator to compile the dataset. BVS provides a file with the coordinates from the traveling nanocollectors and nanosensors with time, which we processed to compute the traveling time and nanosensors concentration level per nanocollector.[6]

Figures 8 and 9 illustrate the histogram plot of the traveling time and the concentration level of nanocollectors along the different tissues. As these plots demonstrate, these two metrics can distinguish the traveling tissue reported by the nanocollector. For instance, the traveling time is larger in the low body than in the center body regions (cf. Fig. 8). Similarly, the average concentration level is lower in the shoulders than in the hands

[6]The Matlab code to process the BVS file is accessible in this link https://github.com/jorge-torresgomez/BVS_data, and it also includes a companion datasheet for the dataset describing the composition of data. The datasheet's content follows the recommendations from Gebru et al. [40].

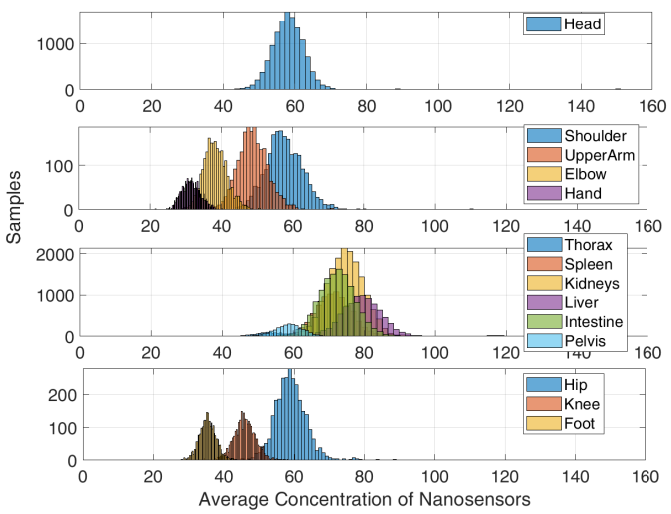

Fig. 9: Representation of the collected data for the average total of nanosensors, as provided by nanocollectors to the gateway device.

(cf. Fig. 9). Although not illustrated here, we also remark on the use of average concentration, as it provides more discernible sets compared to other metrics like the cumulative concentration level.[7]

To reduce errors, we also assume that body regions are distinguishable, i.e., head, center body, upper body, and lower body. Consequently, clustering operations only focus on the corresponding tissues per body region. For instance, in the upper body, the problem will be to cluster samples from the shoulders, upper arms, elbows, and hands. Otherwise, the Elbow will not be distinguishable from the hips when mixing all the body regions into one set only (cf. Figures 8 and 9), for instance. To that end, we assume that anchor nodes, fixed to suitable body locations (e.g., shoulders, hips, and the head), provide the location to the nanocollectors when traveling nearby. Following this assumption, we focus on the center body, where the data is mixed in with the time and concentration dimensions.

Using these raw data for the traveling time and the concentration of nanosensors, three alternative training sets might be used for the ML models. The ML can be trained with 1D data in the time or concentration domains (see Figures 8 and 9). Alternatively, the combined domains as 2D data might be used as a training set; see Fig. 10 for the histogram plot in the center body tissues. The following Sections detail the ML training procedure and their comparative performance in distinguishing the data sets.

B. Unsupervised ML Methods: k-means vs. self-organizing feature map

This section explores the performance of low and high complex ML methods using the dataset, as described in the Section above. We implement in Matlab code the k-means

[7]Further plots for comparison are accessible in https://github.com/jorge-torresgomez/BVS_data.

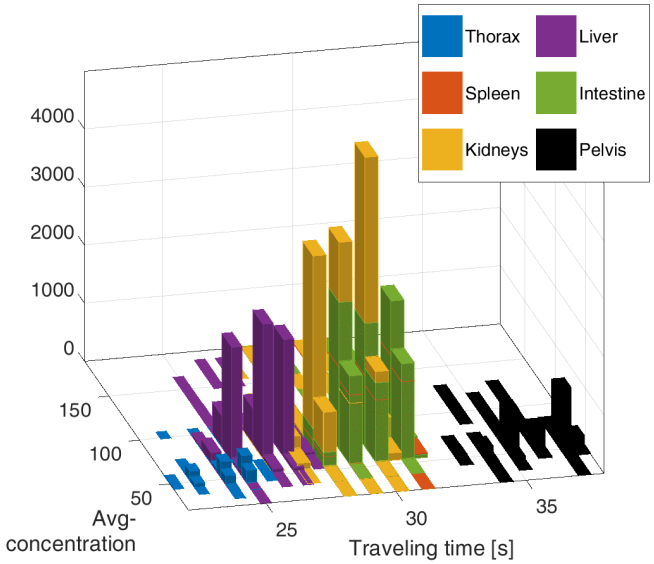

Fig. 10: Representation of the collected data, concerning traveling time and average concentration level, provided by nanocollectors to the gateway device.

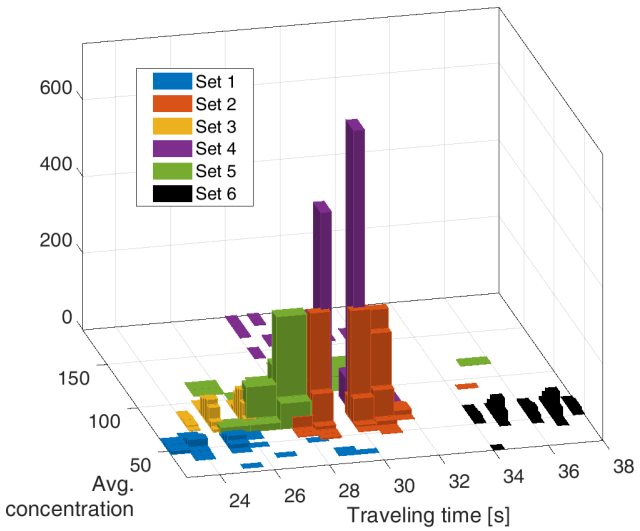

Fig. 11: Unsupervised method after merging clusters with the SOFM SOFM method.

method as the low-complex [41] one, and the deep learning SOFM method as the complex one [41], [42].[8]

Both methods aim to find the partition of samples into clusters, looking for resembling samples in the same cluster that are also dissimilar to the ones in different clusters. The k-means method implements this partition by clustering the data around the centroids and updating the clusters' centroids iteratively. Data is partitioned around the centroids by minimizing a proximity metric as the distance between samples and centroids. Starting with a random distribution of centroids, as proved to converge faster, the centroids are iteratively updated for each cluster; see [41, Sec. II C.].

The training for the k-means method is implemented with 4000 iterations. We set the number of clusters to $k = 6$ as we know in advance the number of capillaries in the center body region, see Fig. 10. The six centroids are initially placed randomly on each iteration, and we run the iterations until the six centroids are located in the time interval corresponding to each circuit. For instance, we expect one of the centroids located in the range $30 - 40\,\mathrm{s}$, in correspondance with the values in Fig. 8. Similarly, we define ranges for the other five centroids.

In a different approach, the SOFM method uses neurons as placed in a grid to cluster samples [42]. Inspired by the cortex of mammals, the neuron's positions are updated according to heuristic rules, i.e., without explicitly minimizing error functions [43]. The resulting location of neurons is then used to partition the data into clusters.

We train the SOFM model using $90\,\%$ of the amount of data and the remaining $10\,\%$ for testing. Six neurons are initially layered as a hexagonal topology during the training phase. Using this minimum amount of neurons, equal amount of clusters are achieved as represented in the 2D histogram plot in Fig. 11. The training phase runs till we observe the centroids are located within the same time interval as for the k-means method, as explained above.

When comparing to the ground truth in Fig. 10, we can directly identify the thorax (Set 1), the pelvis (Set 6) and part of the kidneys (Set 4), while the spleen, intestine and liver are distributed among the Sets 2, 3, and 5, respectively. To account for six clusters (according to the six tissues in the center body), we arbitrarily merged 2 and 8 as the kidneys, 3 and 9 as the intestine, and 5 and 7 as the spleen, resulting in the sets depicted in Fig. 11 for the testing data.

After implementing these methods, we evaluate their clustering performance with the positive predicted and false negative metrics. To illustrate, Fig. 12 depicts the prediction performance when using the k-means method, in average the positive predicted samples are $85.96\,\%$, and the false negative is $14.04\,\%$ of the total. Most false negative bars stem in the center body region, which is the hardest region to distinguish among the different clusters. As Fig. 8 depicts, the overlapping sets in the center body (thorax, spleen, etc) impact the clustering performance.

We compare the performance of the ML models (k-means and SOFM) using the 2D-data and 1D data, as depicted in Fig. 12. In addition, we also illustrate the Gaussian method to cluster data in the concentration domain due to its similarity to a Gaussian distribution, see Fig. 9.[9] As illustrated in Fig. 12, the low complex k-means method scores with the highest performance when using the 2D data. Remarkably, clustering in the time domain only with the k-means method does not

[8]The k-means and SOFM methods are implemented according to the description in https://www.mathworks.com/help/stats/kmeans.html and https://www.mathworks.com/help/deeplearning/gs/cluster-data-with-a-self-organizing-map.html, respectively.

[9]The Gaussian model is implemented according to the description in this link https://www.mathworks.com/help/stats/clustering-using-gaussian-mixture-models.html.

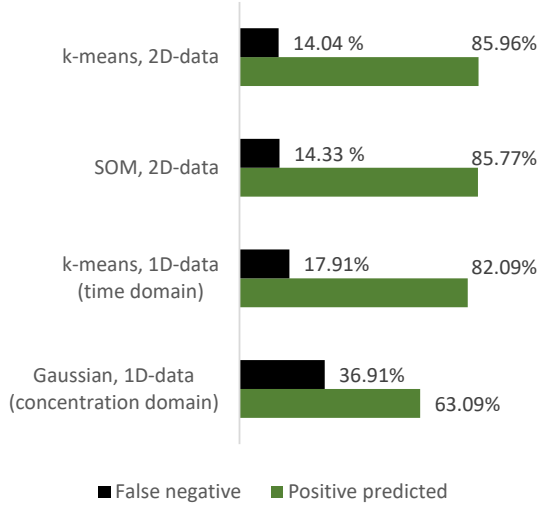

Fig. 12: Comparative prediction performance of the various ML models using the 2D and 1D data.

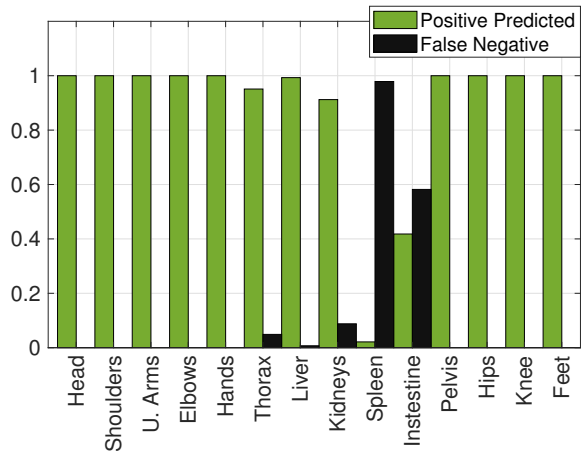

Fig. 13: Supervised method: prediction performance of the supervised method to provide early alerts upon detection. Positive predicted samples are $87.83\,\%$ while false negative are $12.17\,\%$ of the total.

result in that amount different from the 2D data.

C. Supervised ML method: Decision Tree

We implement a supervised ML method to promptly alert the location of detected abnormalities, as follows from the right branch in Fig. 3. As for the supervised model, we apply the Decision Tree method with the 2D labeled data provided by the unsupervised k-means method, as it resulted in the best performance; see Fig. 13. Using Gini's diversity index as the splitting rule, this method minimizes the errors using the resubstitution estimate [44], resulting in a high accuracy and low complexity classifier. [10]

The Decision Tree is implemented with 100 splits using Gini's diversity index as the split criterion. The training is performed with the $90\,\%$ of the BVS data set, and the testing is made with the remainder $10\,\%$. The resulting prediction performance is illustrated in Fig. 13 and assumes the labeled data from the unsupervised method as ground truth. Similarly to the unsupervised methods, there are non-negligible errors in detecting the source of the abnormality in the center body, specifically in the Spleen and the Intestine. This is due to the similarity of reported traveling time and concentration level in the raw data; see Figures 8 and 9.

VI. Results

This section illustrates the localization and detection capabilities in two subsections with the proposed methodology in Fig. 3. We evaluate the localization capability with the right branch of this methodology. In this respect, we illustrate the ML capabilities to localize the location of reported abnormalities. Besides, we also present the results for the detection performance with the left branch of the methodology. We illustrate the improvement in detection due to the fusion capabilities of DNA-tiles.

A. Distribution of nanothings in the human vessels

Using the Markov model formulation in Eq. (4), we determine the stationary distribution of nanothings along the vessels, according to the right branch in Fig. 3. The stationary vector ν is numerically computed as the solution to Eq. (3) in Matlab code.[11] The transition matrix Π is determined by evaluating the ratio of nanosensors at bifurcations ([34, Eq. (1)]), where the total of nanosensors per circuit N_i is estimated with the k-means method using the 2D data, see the step 3 in Fig. 7. The k-means method, trained with the 2D data, is selected due to its better prediction performance (cf. Fig. 13).

Fig. 14 illustrates the stationary distribution of nanosensors along the various vessel segments in the arteries, capillaries, and veins. As expected, the highest probability is found in the Arcus Aorta (A1), in the heart, and in the inferior and superior vena Cava, as these are segments where most circuits pass by. A low probability is obtained in the capillaries of the legs, for instance, as the blood concentration is expected to be lower.

To compare results, we also depict the solution to the stationary vector but compute it with the ground truth labeled data from BVS. We evaluate the transition probabilities of the Markov model without errors in the prediction performance using an ideal estimator. Most of the probabilities differ in the half for the vessel segments, which is a consequence of the errors introduced by the predicted sample; see the prediction performance in Fig. 13.

B. Detection Probabilities

To account for the detection capabilities of our system, comprised of nanosensors and nanocollectors, we evaluate the global detection probability with Eq. (11) and the global false alarm probability with Eq. (15). These probabilities are derived by merging three components in simulation: 1) the local detection probability of a nanosensor using COMSOL

[10]The Decision Tree method is implemented in Matlab code according to the description in https://www.mathworks.com/help/stats/decision-trees.html.

[11]To solve for the stationary distribution of the Markov chain, we follow the description in https://www.mathworks.com/help/econ/dtmc.asymptotics.html.

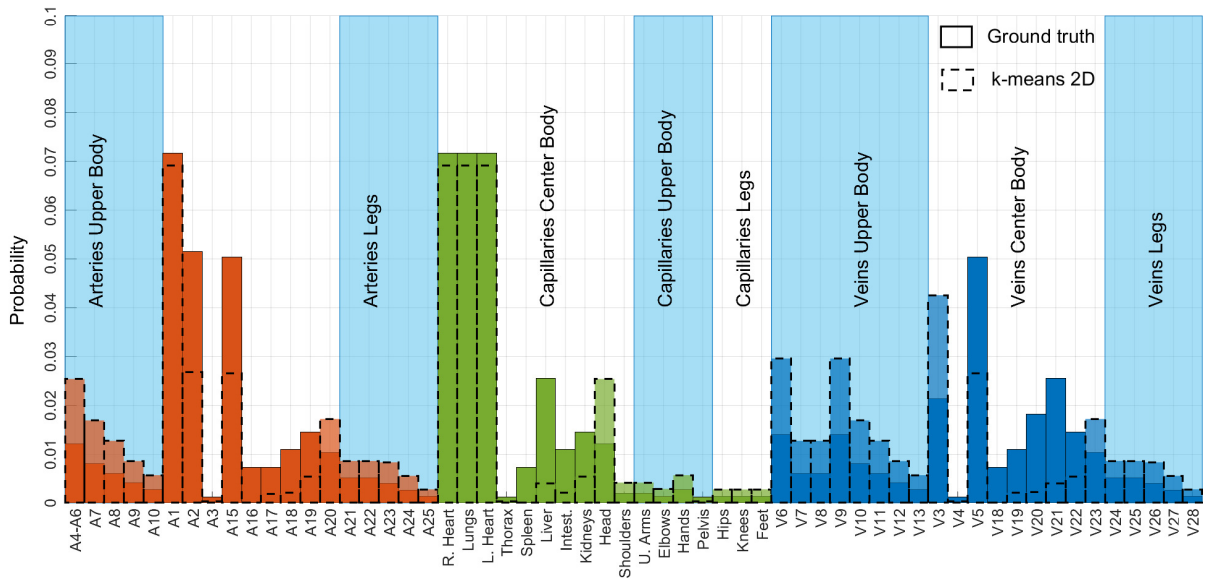

Fig. 14: Distribution of nanosensors in the HCS as derived by the k-means method and the ground truth as using the labeled data from BVS. As for the labels in the arteries and veins; see [11, Fig. 4].

simulations; 2) the probability of a nanosensor visiting a given sensing region as evaluated by the ML algorithms; and 3) the sensor fusion rules, namely, OR, AND, and MAJ rules to determine the detection probability at the nanocollector. We choose to evaluate these probabilities for the upper shoulders region, but they can be easily extended to any other part in the body.

Fig. 15 depicts the global probability of detection and probability of false alarm versus the total number of sensors for the AND rule.[12] We plot the results for varying nanocollector-to-nanosensor ratios, denoted as η, when an AND rule is applied. We evaluate these probabilities with Equations (9) and (11) as for the detection probability and as for the false alarm probability with Equations (13) and (15).

We observe that although increasing the number of nanothings helps with the detection probability up to a certain point ($N_t \approx 10^4$), it does not yield better detection probabilities after that point. This is because it gets more difficult to have a correct reading out of all the nanosensors to decide on detection in the AND rule; even a single nanosensor missing detection blocks a correct decision. Also, even at its peak, the AND rule gives a probability of detection of 0.073, which is very low for our application scenario.

Fig. 16 depicts the global probability of detection and false alarm with the MAJ rule, evaluating the Equations (10) and (11) and Equations (14) and (15), respectively. Similarly to the AND rule, we observe peaks with the detection probability; see for instance the peak at $N_t = 13 \times 10^3$ when $\eta = 0.5$. The observed peak is since, with the MAJ rule, more than half of the nanosensors in that organ should detect the anomaly and report it to the nanocollector. We also observe that the highest probability of detection as 0.14 is better than the AND

[12]We remark that P_d also refers to the sensitivity, and P_{fa} to the specificity as a common nomenclature in the medical field to evaluate infection detection mechanisms.

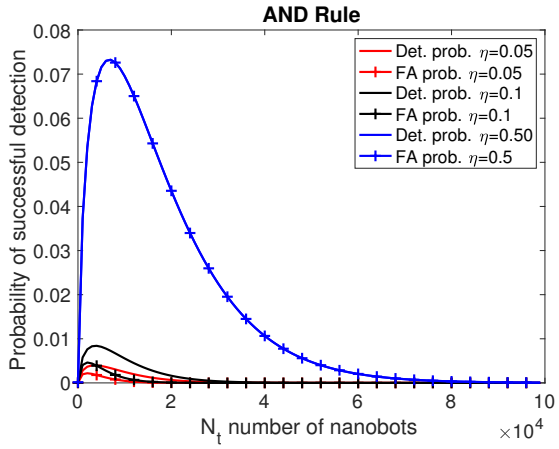

Fig. 15: Resulting detection and false alarm probabilities for the AND rule for the varying ratio of nanocolectors and nanosensors η.

rule with similar probabilities of false alarm. However, the detection probability is still far from the unit (best performance) for reasonable numbers of nanothings.

Finally, Fig. 17 depicts the global probability of detection and probability of false alarm with the OR rule; see Equations (8) and (11) as for the detection probability and Equations (12) and (15) as for the false alarm. In contrast to the AND and MAJ rules, we observe this rule gives a probability of detection of 0.99 with 6×10^3 nanothings. This result is very promising for the application that we consider. We also observe that within the region of the number of nanosensors where the detection probability approaches the maximum performance (1), with N_t in the range 6×10^3-10^4, while the false alarm probability is still below 0.1.

Another major conclusion is that we observe the maximum detection performance when the amount of nanosensors and

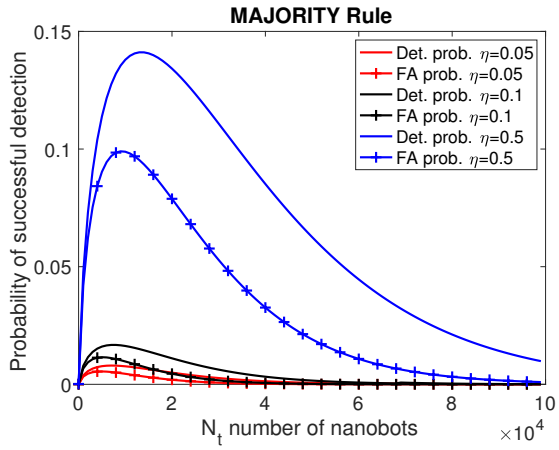

Fig. 16: Resulting detection and false alarm probabilities for the MAJ rule for the varying ratio of nanocolectors and nanosensors η.

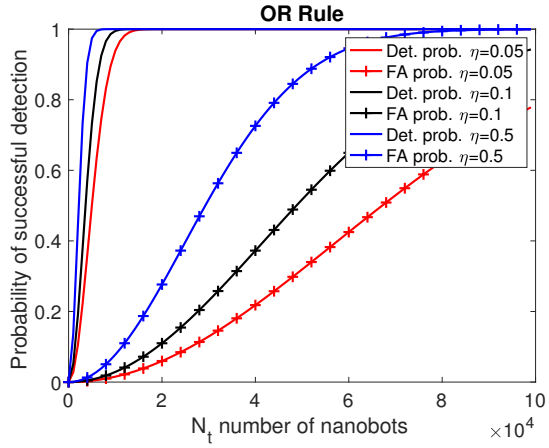

Fig. 17: Resulting detection and false alarm probabilities for the OR rule for the varying ratio of nanocolectors and nanosensors η.

nanocollectors is balanced ($\eta = 0.5$) in the above three Figures 15 to 17. For the fusion rules, it is important to locally detect and have a sufficient amount of FC nodes. These plots show that both nanothings should be in the same amount in the vessel segment where the infection happens.

In summary, we observe a promising application when merging the localization prediction of the ML methods with the fusion mechanism to detect abnormalities in the human body. The localization with the k-means method positively predicts the location of the $86\,\%$ of the samples. Besides, it estimates the location of the abnormalities and provides the spatial probabilities needed to evaluate the fusion rules. Furthermore, the OR rule implements the highest detection probability as 0.99. Although at the expense of the highest false alarm probability compared to the other two fusion rules, it yields an acceptable performance.

VII. Future Work & Open Questions

Within the IoBNT framework, there are more challenges to address in future work. Despite of the suporting testbed technologies (see [45], [46]), the framework requires more development as a concept. Research directions related to more realistic assumptions for the system model, the communication links, and the development of goal-oriented communication strategies open new development pathways. This section remarks on these topics, elaborating on possible research directions.

BVS implements a constant blood speed with time along the different vessel segments. Although this is a realistic assumption in the capillaries and the veins, see [30, Fig. 14-2 pp. 172], will not be the case for the arteries where the pressure variability with time is significant. This time variability also yields a varying traveling time for the nanosensors along the HCS. In that regard, we expect a broader distribution for the traveling time than the case depicted in Fig. 8; thereby, a reduced performance for the localization mechanism introduced in Section V. We let for future work further developments in this direction.

A second direction for future work is related to the impact of the communication mechanism in the interfaces with external devices. This work assumes an ideal connection between the nanocollectors and the gateway. Still, in a real-case scenario, there will be some errors due to noise in the communication link and due to synchronization performance. See, for instance, the impact of noise in the bit error rate (BER) metric for intrabody links in [47, Figures 4 and 5], and the impact of synchronization in [18, Figures 12 and 13].

The operation of the IoBNT network is also prone to failures that require further investigation. The nanothings' limited battery capacities or the chemical reaction with blood components can prevent the system from operating. Although there are preliminary studies on the topic, which includes the impact of a limited lifetime in the network, see [48], further research is needed in this topic to account for realistic parameters as the battery capacity or the resistance of the nanodevice to chemical interaction with blood components.

The impact of noise in the MC channel between the nanosensors and the nanocollectors is also relevant as well. The process in the MC channels follows a diffusion-advection mechanism, where the diffusion of molecules also introduces a noisy component; see the diffusion noise description in [49, Eq. (87)]. Current literature evaluates the reliability of deployed ML modules for prediction detection, see [50], however, the impact of noise, as reflected within the input data is still a topic for further development.

Further, it is yet unclear how the DNA-computation of the fusion rule fares under more realistic circumstances. While DNA quickly and reliably assembles under laboratory conditions, the complex environments in the human body might interfere in undesired ways. As an example, the human immune system is sensitive to DNA outside of the proper places and might attack artificial DNA structures. Further, certain tumors might emit enzymes like DNase that destroy double-stranded DNAs. Future wet lab experiments are necessary to validate the functionality of such systems under gradually more realistic circumstances.

Further, DNA-based computations are stochastic systems where potentially millions of tiles self-assemble in parallel,

which is supportive for scalability. However, there is need of more research when it comes to complex or large computations as the assembly process tends to slow down significantly for some computations. Yet, as the nanoscale is extremely resource-constrained anyway, we do not think that those limits will be approached any time soon.

Finally, we will consider goal-oriented communication strategies within this framework to achieve early alerts and prompt actuation constraints on potential diseases. In this regard, the update's timeliness at the gateway device should allow for following the evolution of potential diseases; see some initial work in [48], [51], [52]. Aiming to maximize information freshness, metrics related to information age can provide a way to evaluate it with a balance between nanosensors and nanocollectors.

VIII. Conclusion

This paper outlines a detection and localization framework of abnormalities in the human body following the sensor-fusion node-gateway architecture. The architecture integrates nanosensors and fusion nodes implemented with nanocollectors in the human vessels scenario, where both components are mobile. Furthermore, leveraging on the DNA-tiles computational capabilities, we realize the standard `OR`, `AND`, and `MAJ` fusion rules. Such a conception supersedes computation tasks in nanothings with natural processes in the biological domain. Besides adding flexibility to the architecture, we integrate ML modules to evaluate physiologic-based parameters. The ML fulfills two purposes: it computes transition probabilities in the vessels' bifurcations and jointly localizes the abnormalities' source. The ML module accounts for a flexible design as it can be trained for the specificity of individual subjects.

Our findings illustrate a valuable balance between detection capabilities and false alarm performance. Implementing the fusion `OR` rule achieves a high detection ratio with a low false alarm probability. Although remarkably, this is obtained with an equal balance of fusion nodes and nanosensors. Reducing the number of nanocollectors has little impact on the detection ratio but has a larger impact on the false alarm probability. These results illustrate the usefulness of including mobile fusion nodes within the architecture.

Acknowledgment

The reported research was supported by the project NaBo-Com/BodyTalks, funded by the German Research Foundation (DFG) under grants DR 639/21-2 and FI 605/21-2 as well as by the project IoBNT, funded by the Federal Ministry of Education and Research (BMBF, Germany) under grants 16KIS1986K and 16KIS1991.

References

[1] A. Czapka, C. Grune, P. Schädel, et al., "Drug delivery of 6-bromoindirubin-3'-glycerol-oxime ether employing poly(D,L-lactide-co-glycolide)-based nanoencapsulation techniques with sustainable solvents," *Journal of Nanobiotechnology*, vol. 20, no. 1, Jan. 2022.

[2] M. Kuscu and B. D. Unluturk, "Internet of Bio-Nano Things: A review of applications, enabling technologies and key challenges," *ITU Journal on Future and Evolving Technologies*, vol. 2, no. 3, pp. 1–24, Dec. 2021.

[3] R. G. Kerry, K. E. Ukhurebor, S. Kumari, et al., "A comprehensive review on the applications of nano-biosensor-based approaches for non-communicable and communicable disease detection," *Biomaterials Science*, vol. 9, no. 10, pp. 3576–3602, 2021.

[4] N. Etemadi, M. Farahnak-Ghazani, H. Arjmandi, M. Mirmohseni, and M. Nasiri-Kenari, "Abnormality Detection and Localization Schemes Using Molecular Communication Systems: A Survey," *IEEE Access*, vol. 11, pp. 1761–1792, 2023.

[5] M. Stelzner, F.-L. A. Lau, K. Freundt, et al., "Precise Detection and Treatment of Human Diseases Based on Nano Networking," in *11th International Conference on Body Area Networks (BODYNETS 2016)*, Turin, Italy: EAI, Dec. 2016.

[6] N. Varshney, A. Patel, Y. Deng, W. Haselmayr, P. Varshney, and A. Nallanathan, "Abnormality Detection Inside Blood Vessels With Mobile Nanomachines," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 4, no. 3, pp. 189–194, Sep. 2018.

[7] U. Rogers and M.-S. Koh, "Parallel Molecular Distributed Detection With Brownian Motion," *IEEE Transactions on NanoBioscience*, vol. 15, no. 8, pp. 871–880, Dec. 2016.

[8] R. Mosayebi, A. Ahmadzadeh, W. Wicke, V. Jamali, R. Schober, and M. Nasiri-Kenari, "Early Cancer Detection in Blood Vessels Using Mobile Nanosensors," *IEEE Transactions on NanoBioscience*, vol. 18, no. 4, pp. 103–116, Oct. 2019.

[9] L. Khaloopour, M. Mirmohseni, and M. Nasiri-Kenari, "Theoretical Concept Study of Cooperative Abnormality Detection and Localization in Fluidic-Medium Molecular Communication," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17 118–17 130, Aug. 2021.

[10] D. Vasisht, G. Zhang, O. Abari, H.-M. Lu, J. Flanz, and D. Katabi, "In-body backscatter communication and localization," in *ACM SIGCOMM 2018*, Budapest, Hungary: ACM, Aug. 2018, pp. 132–146.

[11] J. Torres Gómez, A. Kuestner, J. Simonjan, B. D. Unluturk, and F. Dressler, "Nanosensor Location Estimation in the Human Circulatory System using Machine Learning," *IEEE Transactions on Nanotechnology*, vol. 21, pp. 663–673, Oct. 2022.

[12] F.-L. A. Lau, F. Büther, R. Geyer, and S. Fischer, "Computation of decision problems within messages in DNA-tile-based molecular nanonetworks," *Elsevier Nano Communication Networks*, vol. 21, p. 100 245, Sep. 2019.

[13] C. Lin, Y. Liu, S. Rinker, and H. Yan, "DNA Tile Based Self-Assembly: Building Complex Nanoarchitectures," *ChemPhysChem*, vol. 7, no. 8, pp. 1641–1647, Aug. 2006.

[14] J. Simonjan, B. D. Unluturk, and I. F. Akyildiz, "In-body Bionanosensor Localization for Anomaly Detection via Inertial Positioning and THz Backscattering Communication," *IEEE Transactions on NanoBioscience*, pp. 1–1, 2021.

[15] F. Lemic, S. Abadal, E. Alarcón, and J. Famaey, "Toward Location-aware In-body Terahertz Nanonetworks with Energy Harvesting," arXiv, cs.ET 2101.01952, Jan. 2021.

[16] F. Lemic, S. Abadal, W. Tavernier, et al., "Survey on Terahertz Nanocommunication and Networking: A Top-Down Perspective," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1506–1543, Jun. 2021.

[17] E. S. Andersen, M. Dong, M. M. Nielsen, et al., "Self-assembly of a nanoscale DNA box with a controllable lid," *Nature*, vol. 459, no. 7243, pp. 73–76, May 2009.

[18] J. Torres Gómez, J. Simonjan, and F. Dressler, "Low-Complex Synchronization Method for Intra-Body Links in the Terahertz Band," *IEEE Journal on Selected Areas in Communications*, May 2024, to appear.

[19] J. Torres Gómez, A. Kuestner, L. Stratmann, and F. Dressler, "Modeling Ultrasonic Channels with Mobility for Gateway to In-Body Nanocommunication," in *IEEE Global Communications Conference (GLOBECOM 2022)*, Rio de Janeiro, Brazil: IEEE, Dec. 2022, pp. 4535–4540.

[20] A.-A. A. Boulogeorgos, S. E. Trevlakis, and N. D. Chatzidiamantis, "Optical Wireless Communications for In-Body and Transdermal Biomedical Applications," *IEEE Communications Magazine*, vol. 59, no. 1, pp. 119–125, Jan. 2021.

[21] I. Ahmed, S. Halder, A. Bykov, A. Popov, I. V. Meglinski, and M. Katz, "In-Body Communications Exploiting Light: A Proof-of-Concept Study Using Ex Vivo Tissue Samples," *IEEE Access*, vol. 8, pp. 190 378–190 389, Jan. 2020.

[22] Z. Kou, R. J. Miller, A. C. Singer, and M. L. Oelze, "High Data Rate Communications In Vivo Using Ultrasound," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 11, pp. 3308–3316, Nov. 2021.

[23] T. Bos, W. Jiang, J. D'hooge, M. Verhelst, and W. Dehaene, "Enabling Ultrasound In-Body Communication: FIR Channel Models and QAM Experiments," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 1, pp. 135–144, Feb. 2019.

[24] H. Elayan, R. M. Shubair, J. M. Jornet, and P. Johari, "Terahertz Channel Model and Link Budget Analysis for Intrabody Nanoscale Communication," *IEEE Transactions on NanoBioscience*, vol. 16, no. 4, pp. 491–503, Jun. 2017.

[25] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns," *Nature*, vol. 440, no. 7082, pp. 297–302, Mar. 2006.

[26] M. L. Scarpello, D. Kurup, H. Rogier, et al., "Design of an Implantable Slot Dipole Conformal Flexible Antenna for Biomedical Applications," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 10, pp. 3556–3564, Oct. 2011.

[27] J. Torres Gómez, J. Engstrand, S. Abadal, R. Augustine, T. Voigt, and F. Dressler, "Implications of Nanodevice Mobility on Terahertz Communication Links in the Human Vessels," in *11th ACM International Conference on Nanoscale Computing and Communication (NANOCOM 2024)*, to appear, Milan, Italy: ACM, Oct. 2024.

[28] R. Geyer, M. Stelzner, F. Büther, and S. Ebers, "BloodVoyagerS: Simulation of the Work Environment of Medical Nanobots," in *5th ACM International Conference on Nanoscale Computing and Communication (NANOCOM 2018)*, Reykjavík, Iceland: ACM, Sep. 2018, 5:1–5:6.

[29] H. C. Berg, *Random Walks in Biology*. Princeton University Press, 1993.

[30] A. C. Guyton and M. E. Hall, *Guyton and Hall Textbook of Medical Physiology*, 14th ed. Elsevier, 2015.

[31] M. Fruchard, L. Arcese, and E. Courtial, "Estimation of the Blood Velocity for Nanorobotics," *IEEE Transactions on Robotics and Automation*, vol. 30, no. 1, pp. 93–102, Feb. 2014.

[32] C. Gatsonis, J. S. Hodges, R. E. Kaas, and N. D. Singpurwalla, *Case Studies in Bayesian Statistics*. Springer, 2012, vol. II.

[33] J. Torres Gómez, A. Kuestner, K. Pitke, J. Simonjan, B. D. Unluturk, and F. Dressler, "A Machine Learning Approach for Abnormality Detection in Blood Vessels via Mobile Nanosensors," in *19th ACM Conference on Embedded Networked Sensor Systems (SenSys 2021), 2nd ACM International Workshop on Nanoscale Computing, Communication, and Applications (NanoCoCoA 2021)*, Coimbra, Portugal: ACM, Nov. 2021, pp. 596–602.

[34] J. Torres Gómez, R. Wendt, A. Kuestner, K. Pitke, L. Stratmann, and F. Dressler, "Markov Model for the Flow of Nanobots in the Human Circulatory System," in *8th ACM International Conference on Nanoscale Computing and Communication (NANOCOM 2021)*, Virtual Conference: ACM, Sep. 2021, 5:1–5:7.

[35] T. A. Webster and E. D. Goluch, "Electrochemical detection of pyocyanin in nanochannels with integrated palladium hydride reference electrodes," *Lab on a Chip*, vol. 12, no. 24, pp. 5195–5201, 2012.

[36] E. Baldrich, F. X. Muñoz, and C. García-Aljaro, "Electrochemical Detection of Quorum Sensing Signaling Molecules by Dual Signal Confirmation at Microelectrode Arrays," *Analytical Chemistry*, vol. 83, no. 6, pp. 2097–2103, Feb. 2011.

[37] I. F. Akyildiz, M. Ghovanloo, U. Guler, T. Ozkaya-Ahmadov, A. F. Sarioglu, and B. D. Unluturk, "PANACEA: An Internet of Bio-NanoThings Application for Early Detection and Mitigation of Infectious Diseases," *IEEE Access*, vol. 8, pp. 140 512–140 523, Jan. 2020.

[38] A. Kumari, P. Pasini, S. K. Deo, D. Flomenhoft, H. Shashidhar, and S. Daunert, "Biosensing Systems for the Detection of Bacterial Quorum Signaling Molecules," *Analytical Chemistry*, vol. 78, no. 22, pp. 7603–7609, Nov. 2006.

[39] F.-L. A. Lau, R. Wendt, and S. Fischer, "Efficient in-message computation of prevalent mathematical operations in DNA-based nanonetworks," *Elsevier Nano Communication Networks*, vol. 28, p. 100 348, Jun. 2021.

[40] K. Gebru, J. Morgenstern, B. Vecchione, et al., "Datasheets for datasets," *Communications of the ACM*, vol. 64, no. 12, pp. 86–92, Dec. 2021.

[41] R. Xu and D. WunschII, "Survey of Clustering Algorithms," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 16, no. 3, pp. 645–678, May 2005.

[42] T. Kohonen, "The self-organizing map," *Neurocomputing*, vol. 21, no. 1-3, pp. 1–6, Nov. 1998.

[43] K.-L. Du, "Clustering: A neural network approach," *Neural Networks*, vol. 23, no. 1, pp. 89–107, Jan. 2010.

[44] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification And Regression Trees*, 1st ed. Albany, NY: Routledge, 1984, p. 368.

[45] S. Lotter, L. Brand, V. Jamali, et al., "Experimental Research in Synthetic Molecular Communications – Part I," *IEEE Nanotechnology Magazine*, vol. 17, no. 3, pp. 42–53, Jun. 2023.

[46] S. Lotter, L. Brand, V. Jamali, et al., "Experimental Research in Synthetic Molecular Communications – Part II," *IEEE Nanotechnology Magazine*, vol. 17, no. 3, pp. 54–65, Jun. 2023.

[47] J. Torres Gómez, J. Simonjan, J. M. Jornet, and F. Dressler, "Optimizing Terahertz Communication Between Nanosensors in the Human Cardiovascular System and External Gateways," *IEEE Communications Letters*, vol. 27, no. 9, pp. 2318–2322, Sep. 2023.

[48] S. Pal, J. Torres Gómez, R. Wendt, S. Fischer, and F. Dressler, "Age of Information-based Abnormality Detection with Decay in the Human Circulatory System," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 2024, to appear.

[49] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel Modeling for Diffusive Molecular Communication - A Tutorial Review," *Proceedings of the IEEE*, vol. 107, no. 7, pp. 1256–1301, Jul. 2019.

[50] G. Angelopoulos and S. Bates, "Conformal Prediction: A Gentle Introduction," *Foundations and Trends® in Machine Learning*, vol. 16, no. 4, pp. 494–591, 2023.

[51] J. Torres Gómez, J. Angjo, and F. Dressler, "Age of Information-based Performance of Ultrasonic Communication Channels for Nanosensor-to-Gateway Communication," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 9, no. 2, pp. 112–123, Jun. 2023.

[52] J. Torres Gómez, K. Pitke, L. Stratmann, and F. Dressler, "Age of Information in Molecular Communication Channels," *Elsevier Digital Signal Processing, Special Issue on Molecular Communication*, vol. 124, p. 103 108, May 2022.