IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 August 2022, accepted 31 August 2022, date of publication 1 September 2022, date of current version 9 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3203846

RESEARCH ARTICLE

A Machine Learning Enabled Near Infrared Tracking Scheme for Localization of Gastrointestinal Smart Capsule

HONGJIE JIANG[1], (Member, IEEE), YI MA[1], JIANCONG YE[1], CHAOFAN LING[1], AND JUNPEI ZHONG[2]

[1]Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou 511442, China
[2]Department of Rehabilitation Sciences, Hong Kong Polytechnic University, Hong Kong, China

Corresponding authors: Hongjie Jiang (jiang1029@scut.edu.cn) and Junpei Zhong (joni.zhong@polyu.edu.hk)

This work was supported by the National Natural Science Foundation of China under Grant 42177440.

ABSTRACT Localizing a smart capsule within the gastrointestinal (GI) tract is essential for high performance, accurate sensing as well as efficacious drug delivery at designated locations. In this work, we describe a data-driven framework that employs a near infrared (NIR) tracking scheme to achieve the localization of smart capsule in GI tract. A prototype of the tracking system consists of a single NIR LED of 940 nm incorporated with an array of readout device that integrated with an array of NIR photodiodes. A data-driven approach was applied to build the non-linear estimation model and estimate the capsule's localization by interpolating the outputs of the photodiodes in response to the movement of the NIR LED. Three different machine learning models: support vector regression (SVR), KNN, adaptive boosting (AdaBoost), and the trilateration positioning approach were trained on calibration data. These models were validated in ex vivo experiments with different thick of porcine tissue (e.g. 20 mm thick) using a NIR LED (three types of intensities) and three photodiodes in different patterns. The capsule localization predicted by these machine learning models showed best estimation results ($R^2 = 96.00\%$, and RMSE = 5.13 mm) when using Adaboost; a second best performance was achieved by SVR with a tradeoff on accuracy and time saving. These results suggest that the proposed machine learning data-driven enabled NIR tracking system can be an effective tool for measuring real-time location of gastrointestinal capsule.

INDEX TERMS Smart capsule, location estimation, near infrared, machine learning.

I. INTRODUCTION

Diagnostic and therapeutic approaches in the human gastrointestinal (GI) tract can benefit from smart capsules with on-board miniaturized electronics that are able to perform various sensing and actuating functionalities with no discomfort or pain to the patients [1], [2], [3]. One of their main applications is non-sedative and non-invasive capsule endoscopy, which couples an imaging sensor with an illumination unit to capture and transmit (wirelessly) images of the GI tract to an off-body or body-worn device. The PillCam™ is one of the best-known endoscopic capsules and is used as the gold standard for collecting images from hard-to-reach areas throughout the GI tract [4]. In addition to image capturing, pharmaceutical companies have also been interested in site-specific drug release within the GI tract for increased therapeutic efficacy in preferential absorption positions. One of such modalities was the InteliSite® capsule supplied by Innovative Devices (Raleigh, NC, USA), which uses an external radiofrequency to activate a bent shape-memory alloy actuator to rotate and align a series of holes in an outer sleeve and drug reservoir, enabling the drug release through the holes (location tracking was accomplished by x-ray fluoroscopy) [5]. Meanwhile, many researchers continue to investigate and improve the fundamental performance and capabilities of smart capsules along various fronts, including more precise control of locomotion and tracking of position/movement, integration of multiple sensors, and

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

incorporation of actuator systems for medication release. For instance, IntelliCap®, produced by Philips, employs the embedded commercial pH and temperature sensors to profile its real-time location in the gut and release therapeutic agents to the targeted position via a programmable stepper motor [6]. Despite these efforts, challenges remain for the development of an affordable, autonomous, and practically invisible (no user intervention) smart capsule for human GI health.

Among the aforementioned problems, one of the main difficulties for their applications in clinics is the lack of continuous or periodical real-time location estimation of such capsules; for gastroenterologists, the GI tract is regarded as a black box in terms of visual localization. Common localization methods include magnetic tracking [7] or imaging techniques using X-ray or gamma scintigraphy [8]. The magnetic method has a short interrogation range and requires complicated instrumentations and accurate setups; while, imaging techniques are not real-time and require the patient to be in a controlled clinic setting and moreover pose health hazards if used frequently. For positioning and triggering drug release, physiological biomarkers, such as pH, can be utilized, as is the case in IntelliCap®, which releases medication at specific sites of GI tract by means of continuously monitoring and interrogating the surrounding pH change with the help of an external recorder. A similar concept was developed by the Ziaie group from Purdue University [9], wherein a 3D printed capsule was incorporated with a pH-sensitive hydrogel for autonomously triggering drug release in the small intestine at a tunable time delay after encountering a low-to-high pH transition. However, the GI tract, specifically the small intestine, has a narrow pH range from slightly acidic to neutral [10] varying from time to time and from one person to another, thus reducing the reliability of pH-driven capsule localization. Other techniques, such as radio frequency (RF), can capture the feedback and concurrently process the unique features of electromagnetic waves (e.g., intensity, angle, RF identification) with a dedicated algorithm to determine the relative locations of capsules. Although they are able to achieve a high accuracy, their availability for *in vivo* tracking of endoscopic capsules suffers several shortcomings, including vulnerability to large attenuation through the human body, susceptibility to surrounding environmental RF interference, and undesired negative effects on biological cells [11], [12], [13]. On the other hand, ultrasound at low frequency of $\sim$2 MHz can achieve an improved soft tissue penetration of $>$10 cm [14]; however, its major application in biomedical areas is still for non-invasive powering or clinically relevant imaging instead of instantaneous localization [15], [16], [17].

Recently, visual-based estimation techniques have gained much scientific interests to tackle capsule tracking issues, as they can translate surrounding sensory features into updates of location with a statistical estimation framework. One such method is visual odometry (VO), which has widely been used in robotics. Its principle is to estimate the most optimal result by fusing relative and absolute sensory information together with a joint statistical distribution. The selection of the distribution should take the consideration of the statistical properties of different kinds of sensors together to get a more accurate result. As such, typical estimation methods, such as Kalman filtering or Particle filter, are used [18]. To accelerate the computational speed and lower the cost of the solution, vector quantization (VQ) neural networks, or VQ with reduced dimension can be applied [19], [20]. In this regard, the estimation of capsule location inside the GI tract should also benefit from these filtering methods, whereas, the efficiency of using these algorithms in imaging processing could be significantly compromised by the limits of transmission rate or powering of capsule.

On the other hand, data-driven methods have been proven to be an efficient method for estimation problems. Given an appropriate model and sufficient data, the localization problem can be solved in real-time and accurately. For instance, a Multi-Layer Perceptron (MLP) was used to estimate the 3-axis location of a Pillcam™ capsule with respect to anatomic landmarks of an artificial bowel [21]. Nevertheless, neural network methods, such as MLP or deep learning, usually require a large amount of labelled data and dedicated-designed model tailored for applications, which may create other problems for clinical practitioners who are usually lack of technical background.

As a simple and cost-effective method to solve these above-mentioned problems, we propose a near infrared (NIR) estimation system that is able to measure the 3D spatial position of a gastrointestinal capsule in real-time. In recent several decades, NIR optical technologies become a fast and non-destructive analytical technique that offers many advantages for a large variety of medical purposes. For instance, NIR spectroscopy can be applied in the pharmaceutical industry from raw material identification to final product release [22]. The use of NIR in microscopic fluorescence imaging for in vivo molecular imaging has also been investigated because of its capacity to penetrate deep tissue at a wavelength of 650 to 900 nm [23], [24]. The development of NIR sensitive nanoparticles, which combine with NIR light to trigger the photothermal drug release and thereby improve therapeutic effectiveness, also benefits remotely controlled drug delivery systems [25]. Additionally, NIR LEDs (light emitting diodes) can be integrated in an ingestible lab-on-chip device that functions as an optical-mechanical analytical tool [26] for gastrointestinal bleeding detection [27].

In this work, the capsule can equip an array of NIR LEDs to allow NIR light to penetrate human bowel tissue and be detected in real-time by a handheld device incorporated with an n $\times$ m array of high-performance photodiodes. To target a more precise estimation, we implement a data-driven framework to learn the non-linear mapping properties between the LED-to-photodiode distance and the intensity of received NIR light irradiance. The data-driven methods are chosen over the filtering methods because the latter is more suitable for fitting in linear situations. The novelties of our approach are: (1) the use of on-board and commercially available NIR-LED arrays with appropriate

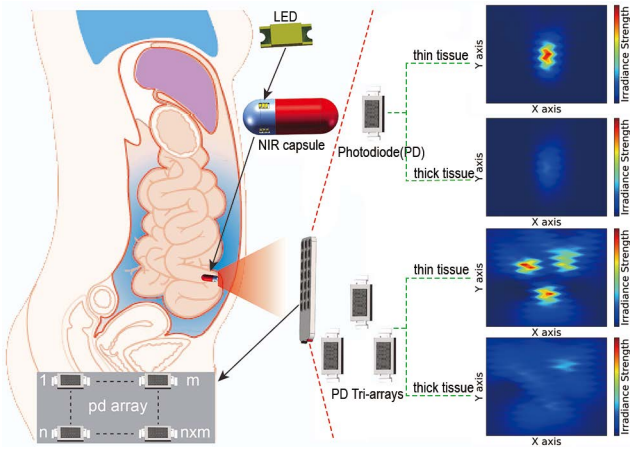

FIGURE 1. A smart capsule emits NIR light source to be detected by the photodiodes array (n × m), which generates various received NIR irradiance gradient at the XY plane for positioning algorithm analysis to enable the real-time location tracking of the capsule in human GI-tract.

TABLE 1. Parameters of light emitter diode.

Parameters	LED-L	LED-M	LED-H
Vishay code	VSMB1940X01	VSMY1940X01	VSMY5940
Peak wavelength	940 nm	940 nm	940 nm
FC-Radiant Intensity	6 mW/sr	10 mW/sr	13 mW/sr
Power dissipation	160 mW	180 mW	190 mW

wavelengths (750 to 1400 nm) that allow deep penetration within soft tissue [28] and are safer to humans as compared to X-ray or gamma ray [24]; and (2) a data-driven framework to estimate the 3D location of the gastrointestinal capsule inside the GI tract which clinical practitioners adapt to different individuals. Fig. 1 shows the conceptual schematic of the proposed capsule tracking method in the GI-tract.

II. EXPERIMENTAL SECTION

A. NIR-LED LOCALIZATION SCHEME

Three NIR LEDs (VSMB194X01, VSMY140X01, VSMY5940, from Vishay) at a wavelength of 940 nm were chosen as the light source and denoted as LED-L (L for low), LED-M (M for median), LED-H (H for high) based on their irradiance strength, respectively, Table 1. The receiving electronics consists of the photodiode (denoted as PD, VBP104S, from Vishay) as the light receiver and a photoelectric circuit configured with a four-channel op amp (operational amplifier, AD8618, from Analog Devices) as the transimpedance amplifier for converting and amplifying the PD's light-dependent current into voltage capable of being finely resolved. Note that the non-inverting input of the op amp receives a modest bias voltage (0.1 V) obtained from the positive supply. This could prevent the output from saturating

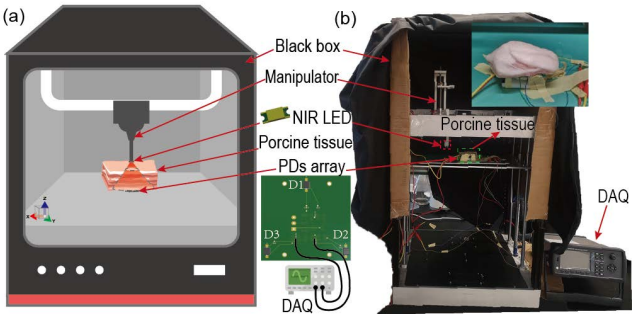

FIGURE 2. (a) Conceptual illustration and (b) photograph of experiment setting to evaluate the tracking performance in various thick porcine tissue (mimicking human soft tissue) by fixing a Tri-PDs array at the bottom and moving LED at the 3D space by a XYZ manipulator.

at the negative supply rail under the absence of input current circumstance.

Fig. 2 illustrates the experiment setup. Assuming the thickness of human abdominal tissue is at a fixed value (e.g., 20 mm) by a small variation [29], given the external readout from each photodiode, the LED position in the XY plane can be estimated. In this work, a group of porcine tissue was used to mimic human abdominal tissue of different thickness. The NIR LED and PD array were placed on the top and bottom of the layered porcine tissue, respectively. The characterization of the PD outputs as a function of the LED positions was performed by scanning the LED over the PD array with a motor-controlled manipulator at a step size of 5 mm over a cubic space of $300 \times 300 \times 300$ mm^3.

The localization experiment is proceeded in two phases. Phase I employs a pair of LED-L and PD as transmitting and sensing unit via a variety of porcine tissue at a step of approximate 5 mm from 7 to 23 mm thick to obtain the tendency curve between three-dimensional distance and receiving irradiance. Phase II utilizes a NIR LED and an equilateral triangle PDs array (Tri-PDs array, side length 50 mm) across a fixed porcine tissue of 10 mm or 20 mm thick to investigate the NIR tracking characterization by only varying the transverse directions. These experiment settings mimic most practical environments, where the capsule locomotion is driven by natural peristalsis and the tracking system (i.e., photodiode array with integrated electronics, shows in the center of Fig. 2) is fixed at some place upon the abdomen skin or mounted on an elastic belt around the abdomen. In phase II, to compensate the declined accuracy induced by light attenuation over porcine tissue, not only LED-L but also LED-M and LED-H were applied. Note that an air gap of 1 mm is added upon the porcine tissue for both phases for moving LED in the controlled path undisturbed by avoiding physical contact with the tissue.

All the experiments used a constant power supply to fix the emitting irradiances of NIR and a data acquisition system (DAQ 970A, Keysight) to continuously measure and record the output from the photoelectric conversion circuit in terms of voltage signal for data analysis. All the tests were

conducted in a black box to eliminate the interference of ambient light; the background noise, i.e., dark current, was analyzed using the same setting but with the LED off. All the experiments were performed in triplicate.

B. DESCRIPTION OF THE DATA

In the experiment of Phase I, the NIR LED-L was carefully controlled to move upon different thick porcine tissue at a rate of 1000 mm/min in x and 800 mm/min in y direction respectively, covering a scanning area of 250 mm × 96 mm. Whereas, as to the experiment of Phase II, the NIR LED moves upon the fixed porcine tissue (thickness = 10 or 20 mm) at a rate of 1000 mm/min in both x and y direction, covering a scanning area of 150 mm × 150 mm. At the same time, the sampling frequency was set to be 10/3 Hz, thus allowing a spatial sampling size of 5 mm in each horizontal scan. In phase I, such manner resulted in 51 data points per horizontal scan and a total of 1275 points over the entire sampling area; while, in phase II, it generated 31 data points per horizontal scan and a total of 961 points per PD (thus a total of 961 × 3 = 2883 points for the Tri-PDs array) over the entire sampling area. As all the output irradiance strength data was measured in triplicate, a low-pass filter was applied to reduce random electrical noise [30] resulting in the average of three samples at each x and y position. Note that the LED moving rate is set to approximately 1.6 cm/s to meet the practical peristalsis rate, ranging from 0.5 cm/s to 2 cm/s [31].

The photodiode reads the NIR irradiance strength in units of amperes. The sensor outputs were converted to irradiance strength (denoted as symbol I) as follows:

$$I_{(x,y)} = \Delta A_{(x,y)}/40,$$

where irradiance strength has units of mW/cm^2, ΔA is the difference between the measured reverse light current and the dark current (i.e., ΔA=A_final–A_(dark current)), 40 is the constant conversion factor given by the photodiode manufacturer for a large range of received light current of 0.4 μA to 80 μA, and (x,y) is the coordinate pair of the recorded data point in the x and y direction at each step along 5 mm.

III. DATA DRIVEN FRAMEWORK

This data-driven framework aims to estimate the location of the digestible capsule in the human GI tract by tracing the spatial irradiance strength change of the NIR LEDs with an array of photodiodes held outside the human body. In practice, the optical signal path between the LEDs and photodiodes is affected by variations of the three-dimensional distance between LEDs and photodiodes due to the capsule moving towards or away from the photodetectors or due to the capsule rotating/vibrating, or by changes in abdominal tissue thickness along the GI tract. As the first step in demonstrating this concept, however, we demonstrate the feasibility of using an on-board NIR light source to localize a gastrointestinal capsule in a semi-three dimensional space (like various XY coordinates but fixed Z position) before considering a more complicated, three-dimensional one. For this purpose, an estimation based on data driven method is proposed in this work. We will firstly formulate it as a static estimation problem in a noisy scenario by a statistical model. In order to solve this problem, a regression analysis will be used. Practically, to solve it as a non-liner regression problem, three learning-based regression methods will be adopted in this framework.

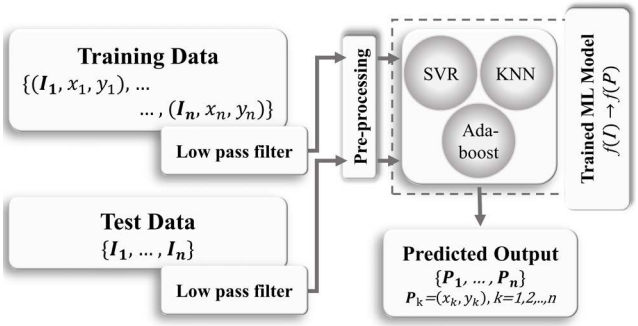

FIGURE 3. Schematic diagram of data-driven algorithm to estimate the capsule position at the XY plane through the received irradiance strength of NIR-LEDs.

Fig. 3 shows a schematic diagram of the data-driven framework, all data acquired from Phase II. Calibration data consisting of the received irradiance strength was measured via scanning the XY space of the porcine tissue (10 mm or 20 mm thick) with the Tri-PDs array underneath by using three types of NIR LED. Next, the ML models were trained using position and irradiance strength data. After training, the ML-enabled models were able to predict the three-dimensional distance between LED and PDs array from irradiance strength data. Three different machine learning algorithms, including support vector regression (SVR), Adaptive Boosting (AdaBoost), and k-nearest neighbors (KNN), were arranged to evaluate the tracking performance in Phase II experiment. A comparison of the prediction scores is then conducted to assess the performance of using one configuration over the rest. Besides, a traditional trilateral algorithm was conducted in the same procedure and used to compare its results to those from ML.

A. DATA COLLECTION

Consider a sample of measurements of irradiance strength $\boldsymbol{I} = \{\boldsymbol{I_k}\}_{k=1}^{n} = \{i_{1k}, i_{2k}, i_{3k}\}_{k=1}^{n}$ representing a physical process of each photodiode receiving NIR light signal via porcine tissue within the sampling area of tracking experiment. Assuming the NIR light direction is always pointing vertically downward in the experiment, the received irradiance $\boldsymbol{I_k}$ can be described as the function of the three-dimensional distance between the NIR LED and the photodiode array, i.e., $\boldsymbol{I_k} = f(x_k, y_k, z_k)_{k=1}^{n}$. For a uniform tissue thickness, e.g., z = 20 mm, the formula can be further simplified to $\boldsymbol{I_k} = f(x_k, y_k)_{k=1}^{n}$. This process will generate a subset of data, each with $n = 961$ tuples, in the format $D =$

$\{(\boldsymbol{I_1}, x_1, y_1), \ldots, (\boldsymbol{I_n}, x_n, y_n)\}$ for each photodiode of the Tri-PDs array configuration, resulting in 3 subsets in total.

Once the capsule is swallowed, the only trackable signal, NIR in this context, will have its irradiance strength varying as a function of the spatial distance between the NIR LED and the photodiodes. In general, this type of problem can be classified into one of tracking estimation using remote sensing methods (e.g., NIR, magnetic field, radio frequency, etc.), whose input is the received wireless signal and whose output is the corresponding estimated position of the target being tracked. In our case, the input is the received NIR irradiance strength, denoted as $\boldsymbol{I_k}$, and the output is the 2D-coordinates position between the moving NIR LED and the fixed photodiodes, denoted as $\boldsymbol{P_k} = (x_k, y_k)$ for $k = 1$ to n. Therefore, the trained estimation model can be expressed as $\boldsymbol{P_k} = f(i_k), k = 1 \; to \; n$.

B. PRE-PROCESSING METHOD

Since the data points collected usually include noisy data which may result from the reflection of infrared light or other reasons, a pre-processing method following the mean filtering (in II.B) should be used to filter out the noises. As the NIR light crosses through the porcine tissue, the receiving irradiance typically shows the Beer–Lambert profile, in which a majority of light strength will attenuate exponentially to be within a small range once over a certain distance from the LED to the PD (more details shown in Fig. 4 and section IV). Therefore, a thresholding filter could be applied to eliminate the data if it is smaller than the threshold; or, the data will be maintained if it is larger than the threshold. To this end, three thresholds were set to a value of 0%, 50%, or 75%, referring to the percentage of the received maximum irradiance. It is worth to notice that as the threshold increases, the detection range decreases while the detection accuracy increases, which, later, is investigated in section IV.

C. REGRESSION ALGORITHMS

The main component of the prediction framework is the regression models, where three machine learning (ML) algorithms: SVR, AdaBoost, and KNN were implemented as standard baseline approaches. For this purpose, the Python programming language and the scikit-learn ML library [32] were employed to model the data using three ML approaches. GridSearchCV in the library was used to search for and tweak the hyper-parameters of the best cross validation score in all ML algorithms by exhaustive searching over specified parameter values over the estimator.

Support Vector Machine (SVM) is a supervised learning algorithm, based on statistical learning theory [33] and focusing on the data points (i.e., support vector) that are closest in the feature space to the optimal boundary between classes. The aim in SVM is to find the optimal boundary which maximizes the separation, or margin, between the support vectors. It is one of the most commonly used machine learning algorithms suitable for classification or regression analysis of low-throughput, high-dimensional datasets such as Support Vector Classification (SVC) and Support Vector Regression (SVR). Since our output data illustrated a certain level of linear regression characteristics, we employed SVR as our first ML algorithm. In this regard, we set the default hyperparameters of kernel trick as Gaussian Function, c = 1 and gamma = auto.

KNN is one of the simplest ML algorithms based on Supervised Learning technique. It can store all the available data and predict estimations by the 'feature similarity' measurement, which refers to the nearest neighbors of data labels [34]. Unlike the KNN classification, the regression based on KNN can be employed in the context of the data labels are continuous rather than discrete variables. Here, the KNN hyperparameters k (number of neighbors) = 5 and weights (distance from these neighbors) were set to uniform as default.

AdaBoost is one type of learning method that seeks to boost the accuracy of a given learning algorithm by converting weak learners to strong learners [35]. In AdaBoost, the prediction model is trained to improve the weak learners by assigning higher weights to incorrectly classified observations and concurrently force the weak learner to focus on the hard samples in the training set. This process is performed in an iteration procedure to grow weaker learners sequentially until weak learners are converted into strong ones., Therefore, we conducted AdaBoost algorithm with a regression model as another effective method to improve the accuracy of the prediction. In this study, the initial hyperparameters were configured to n_estimator = 50 and Loss Function as linear.

D. THE TRAINING PROCESS

At first, the data were normalized and randomly split into training set and testing set with ratio of 8:2. Next, the parameters of the ML algorithms were optimized using a k-fold cross-validation method, allowing for randomly sorting and splitting the training data into k subsets of mutually exclusive datasets (k = 5). Then, k - 1 datasets were randomly selected for training the ML algorithm, and one remaining dataset (thus a 20% split) was used for performing the validation of the training model. This optimization process was iterated k times by changing the held-out dataset. At last, the testing dataset was employed for generating the prediction value.

In this work, the performance of the trained model for the specific training parameters were evaluated by coefficient of determination (R^2), mean squared error (MSE) and root mean squared error (RMSE). The R^2 value expresses the ratio of the variance explained by the ML model and the total variance of the data, which indicates how close the data samples are to the fitted regression line. The range of R^2 value is always between 0 and 1, where a value of approaching to 1 indicates a good amount of agreement between most of the predicted values and the reference values. The MSE measures the average squared difference (defined as ''errors) between the estimated values and the actual value, which helps to indicate how close a set of data points are to the regression line by the predicted model. It also gives more

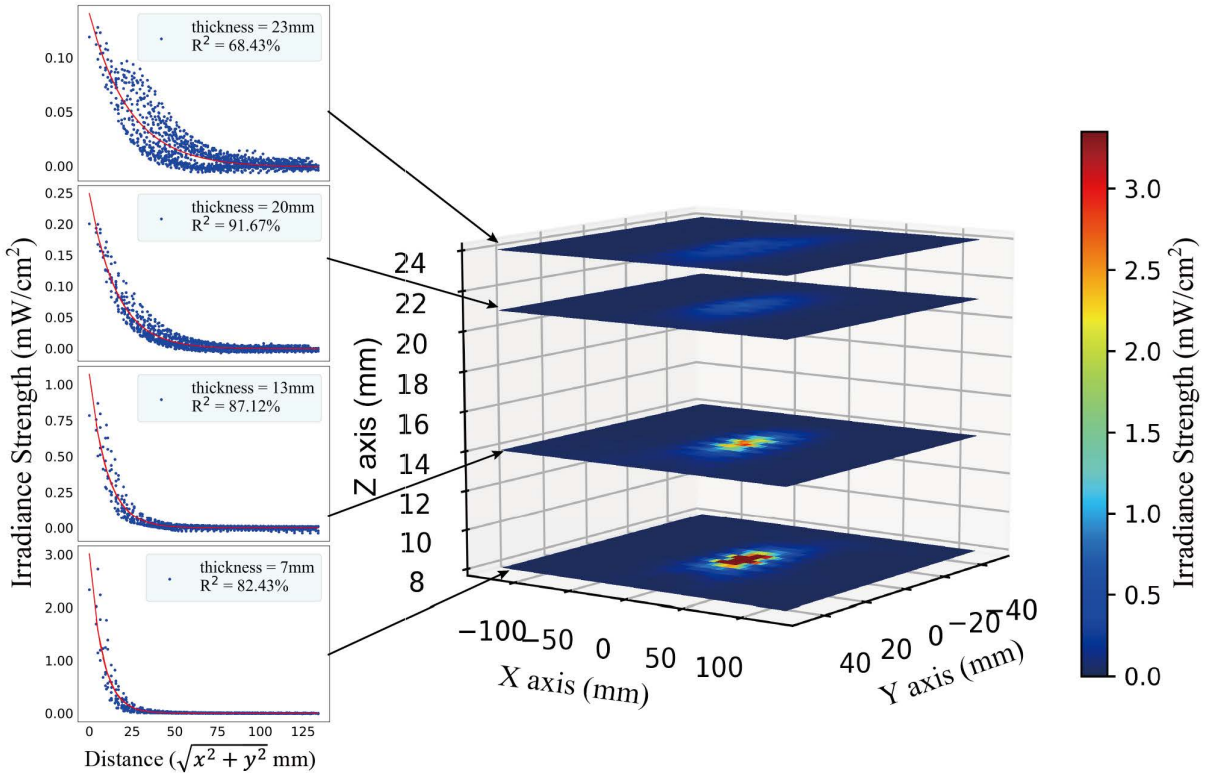

FIGURE 4. Top-view of three-dimensional heat-maps resulted from Phase I, i.e., one LED-L moving on the top of porcine tissue and one PD fixed at the center of bottom over a XY plane of 250 mm × 96 mm with various tissue thickness. Insets (from bottom to top) indicating receiving irradiance strength versus XY distance between LED and PD by increasing tissue thickness from 7 mm to 23 mm at a step of approximate 5 mm.

weight to larger differences. The RMSE was calculated as the root mean square of the absolute errors of the distance between the predicted data and corresponding reference data in three-dimensional space.

E. TRILATERAL POSITIONING ALGORITHMS

Trilateral positioning algorithm is a commonly used localization method in wireless sensor networks applications [36]. Based on the attenuation of wireless signal strength with propagation distance, the distance can be deduced according to the attenuation model formula, so as to carry out target positioning [37]. In this regard, the distance between light source and receiving site can be calculated by a mathematical model of the received NIR irradiance profile in a three-dimensional space (inset of Fig. 4) as follows:

$$d_{ik} = -a \log((i_{lk} - b)/c), \begin{cases} l = 1, 2, 3 \\ k = 1, 2, \ldots, n \end{cases}, \quad (1)$$

where d_{ik} is the distance between the LED and the three PDs, i is the receiving NIR irradiance by each PD, l is the PD number, k is the sampling number, and a and b is the constant from the irradiance mathematical modelling. Afterward, the 3D coordinate of the light source can be obtained by Eq. (2), where (x_1, y_1), (x_2, y_2), and (x_3, y_3) are the predetermined coordinates of each PD in the Tri-PDs array.

$$\begin{cases} d_1^2 = (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 \\ d_2^2 = (x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 \\ d_3^2 = (x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2 \end{cases} \quad (2)$$

IV. RESULTS AND DISCUSSION

A. NIR TISSUE PENETRATION CHARACTERIZATION

Fig. 4 presents a 3D heat-map of the received irradiance on XY plane of 250 mm x 96 mm with layers of porcine tissue by using a pair of LED-L (at the top) and PD (at the bottom). The result shows 4 layers of rectangular heat-map that each layer represents the irradiance distribution at an escalated tissue thickness of 7 to 23 mm at an approximate step of 5 mm. We first checked the relationship between the received irradiance and the thickness of porcine tissue. When porcine tissue thickness increased from 7 to 13, 20, and 23 mm, the irradiance strengths at XY coordinates of (0, 0) of each layer varied at a range of 2.76 to 0.87, 0.2, and 0.13 mW/cm^2, in a descending order. The observation of received irradiance inversely-proportional to porcine tissue thickness was also found in our previous work [38] that the received irradiance demonstrates a similar exponential pattern in response to the penetration depth, in which a maximum penetration can approach 62 mm porcine tissue at a received

power of 0.3 μW/cm^2. Moreover, the received irradiance over the XY plane also demonstrates a one-directional exponential decline for the increase of NIR travelling path, Fig. 4, inset. For example, upon 7 mm thick porcine tissue, the received irradiance reduced sharply from 2.76 mW/cm^2 at the center to 0.012 mW/cm^2, which was the average value at the distance of $\sqrt{x^2 + y^2} = 28$ *mm*. The rapid decrease, nevertheless, slows down as the porcine tissue raises to 23 mm thick, at which the received irradiance reduced from 0.13 to 0.039 mW/cm^2 by moving LED-L 28 mm away from the center. The shrinking of received irradiance range induced by the increase of porcine tissue thickness indicates that compared to the line-of-sight distance between LED and PD, porcine tissue plays a more important role in the tissue-attenuation of NIR LED. As a result, at porcine tissue of 23 mm thick, the received irradiance was narrowed down to a small range of 0.06 to 0.012 mW/cm^2 at a ring region from inner radius of 65mm to outer radius of 125 mm. Therefore, to increase the detection sensitivity (i.e., the slope of irradiance strength over distance, Fig. 4, inset), a filter on a threshold irradiance is applied at a cost of the decrease of detection range. Overall, the results quantitatively identified the amplitude of NIR irradiance as an exponential function of the three-dimensional spatial distance between LED and PD, whose reciprocal, thus, could be used for the monitoring of real-time position of a smart capsule.

B. NIR LOCALIZATION CHARACTERIZATION

As previously discussed, LED-L illustrates a compromised exponential irradiance profile that as the porcine tissue increases over 20 mm, the peak received power at (0, 0) coordinate decreases to 0.2 mW/cm^2 with only a small different to that of value of at 5 mm away from the center. Such small difference strongly degenerates the system sensitivity and concurrently confines the detection range in a small region of few millimeters radius to the center. Thus, to compensate the thick tissue dominated attenuation when using LED-L, two more LEDs with larger initial transmission power, i.e., LED-M and LED-H, were employed in NIR localization test. Furthermore, at the receiving site, three PDs were configured into an equilateral triangle with side length of 50 mm to achieve the trilateral positioning.

Fig. 5 shows the NIR localization within a XYZ space of 150 mm $\times$ 150 mm $\times$ 20 mm. As moving a LED over the XY plane emitting NIR light downwards to the Tri-PDs array cross layers of porcine tissue, each PD of the array receives an attenuated NIR irradiance, thus creating a three-PDs overlaid heat-map at each layer of porcine tissue. As expected, the NIR tracking system exhibits better penetration characterization when increasing the initial emission power of NIR LED. For instance, upon the layer of 20 mm thick porcine tissue, the Tri-PDs array received the peak irradiances of 0.306 or 0.317 mW/cm^2 emitted from LED-M or LED-H, which are 1.53 or 1.585 times of that of 0.2 mW/cm^2 from LED-L. On the other hand, each XY plane reveals three regions surrounding the three PDs, respectively, each with the received irradiances in an exponential decay pattern. These overlaid irradiance distributions indicate that the 50 mm side length of the Tri-PDs array provides sufficient spatial space for each PD to maintain its own irradiance profile by attenuating the optical interferences from the other two PDs. Thus, such arrangement is of great benefit for guaranteeing the system at a high SNR (signal to noise) for applying the trilateral positioning. Note that the received irradiance pattern of each PD appears slightly different, which is caused by the non-uniformity of porcine tissue.

C. MACHINE LEARNING ANALYSIS

As mentioned previously, a threshold filtering was applied to the sampled dataset of 961 points, which generated the reduced dataset of 481 or 241 points by using 50% or 75% threshold, respectively. To this end, a random selection of 20% of filtered dataset was utilized for the evaluation of ML analysis in terms of the regression indexes, e.g., RMSE and R^2 scores. In this regard, the high R^2 value indicates a strong correlation between the predicted data and test data, and the RMSE value represents the Euclidean distance between the predicted data and test data. Note that only the experiment with the Tri-PDs covered with 10 or 20 mm thick porcine tissue were tested as 0 mm thick porcine tissue means no cover.

Three machine learning (ML) were assessed in this work. For SVR model, the results were shown in Table S1. For the scenario of LED-L positioning via 10 mm thick porcine tissue, the prediction scores of (R^2, RMSE) were (74.3%, 22.49 mm), (83.02%, 14.33 mm), and (85.08%, 10.98 mm) with respect to the threshold of 0%, 50%, and 75%, respectively. The scores were improved (i.e., increase of R^2 but decrease of RMSE) by using LED-M or LED-H. In response to the threshold of 0%, 50%, or 75%, the former presents a set of prediction scores (R^2, RMSE) of (84.27%, 17.62 mm), (89.13%, 10.35 mm), or (89.76 %, 7.92 mm); whereas, the latter shows the scores of (76.33%, 21.60 mm), (88.79%, 11.50 mm), or (88.98 %, 9.28 mm). Similarly, the improvement of prediction scores by using high power LED was also observed at 20 mm porcine tissue. For example, LED-L shows the prediction RMSE of (19.3 mm, 10.82 mm, 9.03 mm), which are (130%, 107%, 109%) of that of (14.83, 10.07, 8.26) by using LED-M. On the other hand, it is interesting to note that compared to LED-H, LED-M illustrates a better statistical performance, e.g., the highest prediction score of (R^2 = 91.04%, RMSE = 8.26 mm) at 20 mm porcine tissue was obtained by using LED-M but not LED-H. This result might be understood by the increase of LED power causing the raise of interference level from one LED to the other two. In brief, the comparison reveals that (1) the threshold filtering is an effective data-processing method to eliminate the outliers and concurrently improve the accuracy (particularly in reducing the RMSE value) of SVR analysis; (2) increasing the NIR intensity can compensate the tissue penetration induced decrease of tracking performance, but

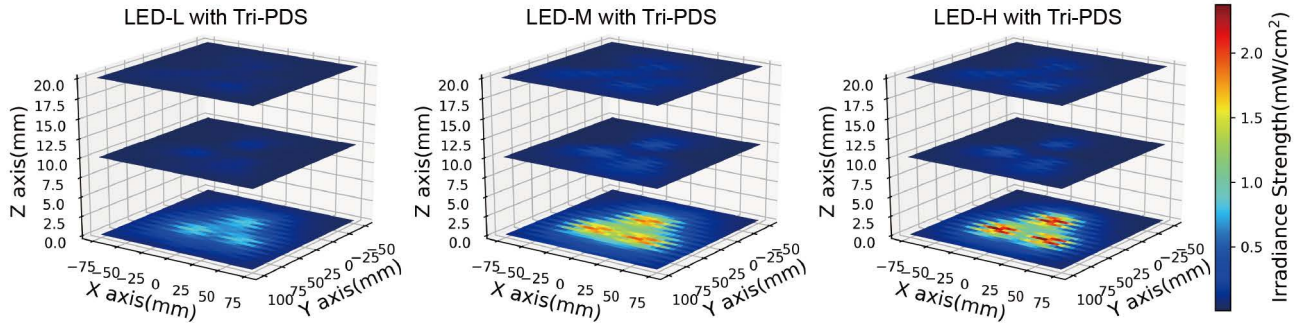

FIGURE 5. 3D heat-map generated from the Tri-PDs array receiving NIR irradiance of a LED moving on XYZ plane of 150 mm × 150 mm × 20 mm (Z referring to tissue thickness), showing an enhanced NIR tissue penetration profile by increasing NIR emission power (from left to right).

requires a careful design of spatial distance among the triangle patterned PDs.

The statistical analysis of KNN model were presented in Table S2. Unlike SVR, KNN shows the R^2 scores for all the scenarios as a one-direction function of irradiance threshold but in descending order. One reason for this might be that because KNN model is proportional to the data density, as the prediction dataset is generated from local interpolation of the test data relevant to the training dataset's nearest neighbors. When reducing the sample size via the threshold filtering, it's, nevertheless, inevitable to cause the raise of statistical errors, e.g., the decrease of R^2 value in this context. However, the RMSE score shows a different pattern. For instance, the RMSE scores of using LED-H at 0% threshold at 20 mm porcine tissue are 9.55 mm, which increase to 12.32 mm at 50% threshold and then decrease to 9.29 mm at 75% threshold. Other configurations, such as LED-M at 20 mm porcine tissue and LED-H as well as LED-L at 10 mm porcine tissue, exhibit the same pattern—RMS score rising initially but falling later as threshold rises. Hence, it is reasonable to draw the conclusion that R^2 is better suited for KNN evaluation when compared to RMSE.

The third ML analysis is Adaboost model, who demonstrates a better performance than both SVR and KNN. As shown in Table S3, Adaboost had all the R^2 score at approximately 95% and almost the RMSE below 8 mm. While compared to Adaboost, either SVR or KNN had their prediction scores deteriorated, like almost all R^2 scores less than 90% or RMSE large than 9 mm. The overall improvement with Adaboost model could be attributed to its nature as the Adaboost regressor is able to autonomously adjust the weights of fitting data until to obtain the best optimized value of the weights. However, Adaboost requires more computation cost than the other two.

At last, the traditional trilateral positioning was evaluated. The results show all negative R^2 scores and high RMSE values (>5 times larger than those of three ML models), Table S4. The trilateration algorithm is based on the receiving signal irradiance attenuation model in an undisturbed light-of-sight environment. However, the practical scenario is a complicated one mainly due to the scattering of NIR light propagating in the porcine tissue. Thus, it is reasonable to say that without the assistance of ML analysis, the positioning prediction of the system by using simple trilateral algorithm is inaccurate.

In summary, this work successfully demonstrates the capability of localizing a moving NIR-LED array over porcine tissue by using a single light source and Tri-PDs array incorporated with suitable ML algorithms. All the ML models present a significant improvement against traditional trilateral algorithm in NIR positioning. Adaboost presents the best performance, like the largest R^2 score of 96.49% and the lowest RMSE of 5.13 mm, but it costs more computation time, approximately one order of that of the other two methods. Furthermore, the thresholding filter proves to be an efficient procedure to improve the system performances as 75% threshold in received irradiance shows a better prediction score than 0% threshold in almost all the experiments. Besides, enhancement can also be done by increasing the LED power. It is worth noticing that the system works better generally at 20 mm porcine tissue than it does at 10 mm porcine tissue, which indicates the potential of the system available for practical applications as the average abdominal thickness of normal person is slightly higher than 20 mm.

V. CONCLUSION

We have demonstrated a machine learning-enabled NIR tracking scheme to predict the spatial position of a smart capsule performing gastrointestinal diagnostics in GI-tract. A prototype system was developed by coupling a 3D-printed capsule equipped with an array of NIR LEDs to a readout circuit integrated with an array of NIR photodiodes. A data-driven framework was established with three machine learning algorithms. The tracking system is capable of measuring capsule location in real-time over a 2D sensing plane of 150 mm × 150 mm. The configuration of a single NIR LED and a photodiode presents the penetration tendency as decays exponentially with distance through the porcine tissue. When switching to NIR tracking configurations, the ML models (e.g., Adaboost) generate three-dimensional LED localization with accuracy as high as 96% R^2 and 5.13 mm

RMSE. Future work involves validation of functionality of the NIR tracking system in large animal models.

ACKNOWLEDGMENT

The authors would like to thank the help of Dr. Babak Ziaie and Dr. Rahim Rahimi from Purdue University, Dr. Albert Kim from Temple University, and Dr. Manuel Ochoa as an Independent Scholar for their technical support.

REFERENCES

[1] C. McCaffrey, O. Chevalerias, C. O'Mathuna, and K. Twomey, "Swallowable-capsule technology," *IEEE Pervasive Comput.*, vol. 7, no. 1, pp. 23–29, Jan./Mar. 2008.

[2] I. R. Wilding, "Site-specific drug delivery in the gastrointestinal tract," *Crit. Rev. Therapeutic Drug Carrier Syst.*, vol. 17, no. 6, p. 76, 2000.

[3] G. Ciuti, A. Menciassi, and P. Dario, "Capsule endoscopy: From current achievements to open challenges," *IEEE Rev. Biomed. Eng.*, vol. 4, pp. 59–72, 2011.

[4] D. R. Cave, "Technology insight: Current status of video capsule endoscopy," *Nature Clin. Pract. Gastroenterol. Hepatol.*, vol. 3, no. 3, pp. 158–164, Mar. 2006.

[5] M. E. A. McGirr, S. M. McAllister, E. E. Peters, A. W. Vickers, A. F. Parr, and A. W. Basit, "The use of the InteliSite companion device to deliver mucoadhesive polymers to the dog colon," *Eur. J. Pharmaceutical Sci.*, vol. 36, nos. 4–5, pp. 386–391, Mar. 2009.

[6] P. J. van der Schaar, J. F. Dijksman, H. Broekhuizen-de Gast, J. Shimizu, N. van Lelyveld, H. Zou, V. Iordanov, C. Wanke, and P. D. Siersema, "A novel ingestible electronic drug delivery and monitoring device," *Gastrointestinal Endoscopy*, vol. 78, no. 3, pp. 520–528, Sep. 2013.

[7] W. Yu, R. Rahimi, M. Ochoa, R. Pinal, and B. Ziaie, "A smart capsule with GI-tract-location-specific payload release," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 9, pp. 2289–2295, Sep. 2015.

[8] I. R. Wilding and D. V. Prior, "Remote controlled capsules in human drug absorption (HDA) studies," *Crit. Rev. Ther. Drug Carrier Syst.*, vol. 20, no. 6, pp. 405–431, 2003.

[9] H. Jiang, W. Yu, M. Oscai, and B. Ziaie, "A smart capsule with a hydrogel-based pH-triggered release switch for GI-tract site-specific drug delivery," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2808–2813, Dec. 2018.

[10] D. F. Evans, G. Pye, R. Bramley, A. G. Clark, T. J. Dyson, and J. D. Hardcastle, "Measurement of gastrointestinal pH profiles in normal ambulant human subjects," *Gut*, vol. 29, no. 8, pp. 1035–1041, Aug. 1988.

[11] P. Valdastri, A. Menciassi, and P. Dario, "Transmission power requirements for novel ZigBee implants in the gastrointestinal tract," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 6, pp. 1705–1710, Jun. 2008.

[12] Y. E. Mohammed and A. G. Saber, "Estimation of E-field inside muscle tissue at MICS and ISM frequencies using analytic and numerical methods," *J. Biomed. Eng. Technol.*, vol. 2, no. 3, pp. 29–33, 2014.

[13] H. Aubert, "RFID technology for human implant devices," *Comp. Rendus Physique*, vol. 12, no. 7, pp. 675–683, 2011.

[14] H. Basaeri, D. B. Christensen, and S. Roundy, "A review of acoustic power transfer for bio-medical implants," *Smart Mater. Struct.*, vol. 25, no. 12, pp. 1–23, 2016.

[15] H. Jiang, N. M. Carter, A. Zareei, S. Nejati, J. F. Waimin, S. Chittiboyina, E. E. Niedert, T. Soleimani, S. A. Lelièvre, C. J. Goergen, and R. Rahimi, "A wireless implantable strain sensing scheme using ultrasound imaging of highly stretchable zinc oxide/poly dimethylacrylamide nanocomposite hydrogel," *ACS Appl. Bio Mater.*, vol. 3, no. 7, pp. 4012–4024, Jul. 2020.

[16] J. H. Park, A. Kim, H. Jiang, S. H. Song, J. Zhou, and B. Ziaie, "A wireless chemical sensing scheme using ultrasonic imaging of silica-particle-embedded hydrogels (silicagel)," *Sens. Actuators B, Chem.*, vol. 259, pp. 552–559, Apr. 2018.

[17] A. Kim, J. Zhou, S. Samaddar, S. H. Song, B. D. Elzey, D. H. Thompson, and B. Ziaie, "An implantable ultrasonically-powered micro-light-source (μ light) for photodynamic therapy," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.

[18] K. Yousif, A. Bab-Hadiashar, and R. Hoseinnezhad, "An overview to visual odometry and visual SLAM: Applications to mobile robotics," *Intell. Ind. Syst.*, vol. 1, no. 4, pp. 289–311, Nov. 2015.

[19] K. Duda, T. Zielinski, R. Fraczek, J. Bulat, and M. Duplaga, "Localization of endoscopic capsule in the GI tract based on MPEG-7 visual descriptors," in *Proc. IEEE Int. Workshop Imag. Syst. Techn.*, May 2007, pp. 1–4.

[20] J. Bulat, K. Duda, M. Duplaga, R. Fraczek, A. Skalski, M. Socha, P. Turcza, and T. P. Zielinski, "Data processing tasks in wireless GI endoscopy: Image-based capsule localization & navigation and video compression," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, pp. 2815–2818.

[21] G. Dimas, E. Spyrou, D. K. Iakovidis, and A. Koulaouzidis, "Intelligent visual localization of wireless capsule endoscopes enhanced by color information," *Comput. Biol. Med.*, vol. 89, pp. 429–440, Oct. 2017.

[22] J. Luypaert, D. L. Massart, and Y. Vander Heyden, "Near-infrared spectroscopy applications in pharmaceutical analysis," *Talanta*, vol. 72, no. 3, pp. 865–883, May 2007.

[23] V. Ntziachristos, C. Bremer, and R. Weissleder, "Fluorescence imaging with near-infrared light: New technological advances that enable *in vivo* molecular imaging," *Eur. Radiol.*, vol. 13, no. 1, pp. 195–208, Jan. 2003.

[24] M. Gurfinkel, S. Ke, X. Wen, C. Li, and E. M. Sevick-Muraca, "Near-infrared fluorescence optical imaging and tomography," *Dis. Markers*, vol. 19, nos. 2–3, pp. 107–121, 2004.

[25] B. P. Timko, T. Dvir, and D. S. Kohane, "Remotely triggerable drug delivery systems," *Adv. Mater.*, vol. 22, no. 44, pp. 4925–4943, Nov. 2010.

[26] F. Cairone, S. Gagliano, and M. Bucolo, "Experimental study on the slug flow in a serpentine microchannel," *Experim. Thermal Fluid Sci.*, vol. 76, pp. 34–44, Sep. 2016.

[27] M. R. Mohebbian, M. H. A. Sohag, S. S. Vedaei, and K. A. Wahid, "Automated detection of bleeding in capsule endoscopy using on-chip multispectral imaging sensors," *IEEE Sensors J.*, vol. 21, no. 13, pp. 14121–14130, Jul. 2021.

[28] V. Ntziachristos, J. Ripoll, and R. Weissleder, "Would near-infrared fluorescence signals propagate through large human organs for clinical studies?" *Opt. Lett.*, vol. 27, no. 5, pp. 333–335, 2002.

[29] H. Kanehisa, M. Miyatani, K. Azuma, S. Kuno, and T. Fukunaga, "Influences of age and sex on abdominal muscle and subcutaneous fat thickness," *Eur. J. Appl. Physiol.*, vol. 91, nos. 5–6, pp. 534–537, May 2004.

[30] R. A. Yotter and D. M. Wilson, "A review of photodetectors for sensing light-emitting reporters in biological systems," *IEEE Sensors J.*, vol. 3, no. 3, pp. 288–303, Jun. 2003.

[31] A. Tharakan, I. T. Norton, P. J. Fryer, and S. Bakalis, "Mass transfer and nutrient absorption in a simulated model of small intestine," *J. Food Sci.*, vol. 75, no. 6, pp. E339–E346, Aug. 2010.

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2011.

[33] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, May 2018.

[34] S. Sun and R. Huang, "An adaptive k-nearest neighbor algorithm," in *Proc. 7th Int. Conf. Fuzzy Syst. Knowl. Discovery*, Aug. 2010, pp. 91–94.

[35] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[36] G. Mao and B. Fidan, *Localization Algorithms and Strategies for Wireless Sensor Networks*, 1st ed. New York, NY, USA: Information Science Reference, 2009.

[37] W. Xu, J. Wang, H. Shen, H. Zhang, and X. You, "Indoor positioning for multiphotodiode device using visible-light communications," *IEEE Photon. J.*, vol. 8, no. 1, pp. 1–11, Feb. 2016.

[38] H. Jiang, A. Kim, J. Zhou, R. Rahimi, and B. Ziaie, "Real-time tracking of a 3D-printed smart capsule using on-board near-infrared led array," in *Proc. 20th Int. Conf. Solid-State Sensors, Actuat. Microsystems Eurosensors (TRANSDUCERS EUROSENSORS)*, Jun. 2019, pp. 2201–2204.

HONGJIE JIANG (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Purdue University, in 2018. He is currently an Assistant Professor with the Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou, China. His research interests include the development of low-cost, scalable, and flexible sensors for healthcare and environmental applications.

YI MA received the B.E. degree from the East China University of Science and Technology, Shanghai, China. He is currently pursuing the M.E. degree with the Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou, China. His current research interest includes wireless sensors for healthcare based machine learning applications.

CHAOFAN LING received the B.E. degree from the South China University of Technology, Guangzhou, China, in 2020, where he is currently pursuing the M.E. degree with the Shien-Ming Wu School of Intelligent Engineering. His current research interests include computer vision and artificial intelligence.

JIANCONG YE received the B.E. degree from the South China University of Technology, Guangzhou, China, in 2020, where he is currently pursuing the M.E. degree with the Shien-Ming Wu School of Intelligent Engineering. His current research interests include machine learning and artificial intelligence.

JUNPEI ZHONG received the Ph.D. degree from the Department of Computer Science, University of Hamburg, Germany. He currently is a Research Assistant Professor with the Hong Kong Polytechnic University. His research is committed to understand the working mechanisms of humans using neuroscience and psychological measurements, to explain them with mathematical models, and to use them for artificial systems. In applications, these systems can be applied for assistive robotics, interactive robots or other rehabilitation devices.