



Automated optimization of double heater convective polymerase chain reaction devices based on CFD simulation database and artificial neural network model

Seong Hyeon Hong¹ · Jung-Il Shu¹ · Yi Wang¹ · Oktay Baysal²

Accepted: 10 February 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

This paper presents a framework for automated optimization of double-heater convective PCR (DH-cPCR) devices by developing a computational fluid dynamics (CFD) simulation database and artificial neural network (ANN) model. The optimization parameter space that includes the capillary tube geometries and the heater sizes of DH-cPCR is established, and a database consisting of nearly 10,000 CFD simulations is constructed. The database is then used to train a two-stage ANN models that select practically relevant data for modeling and predict PCR device performance. The trained ANN model is then combined with the gradient-based and the heuristics optimization approaches to search for optimal device configuration that possesses the shortest DNA doubling time. The entire design process including model meshing and configuration, parallel CFD computation, database organization, and ANN training and utilization is fully automated. Case studies confirm that the proposed framework can successfully find the optimal device configuration with an error of less than 0.3 s, and hence, representing a cost-effective and rapid solution of DH-cPCR device design.

Keywords Polymerase chain reaction (PCR) · Point of care (POC) · Neural network · Device optimization · Computational fluid dynamics (CFD) · Machine learning

1 Introduction

Polymerase chain reaction (PCR) is an effective technique to amplify a few copies of deoxyribonucleic acids (DNA) to a detectable level (Li et al. 2016; Shu et al. 2019a), and has found widespread applications in biomedical research, medicine, criminal forensics, molecular archaeology, and others that require genomic information (Shu et al. 2019a). Especially during the recent COVID-19 outbreak, PCR has been used extensively for testing viral infections throughout the world, substantiating its significance in biomedicine (Powledge (2004)). Despite popularity of bulky and high-throughput PCR devices in centralized laboratories, there is a strong demand for integrated and miniaturized PCR technologies that can be utilized economically with a minimal requirement of operator experience and deployed

easily outside the lab environment near the patients (Petralia and Conoci 2017). Such devices are normally referred to as “Point-of-Care” (POC) systems, which have already been established for testing of diabetes, pregnancy, cardiac disease, HIV, and others (Luppa et al. 2011). Similarly, the use of PCR devices for POC testing would be advantageous as it enables rapid scale-up of medical tests in community and immediate responses to emergent situations (Sia and Kricka 2008; Niemz et al. 2011).

However, design and development of the POC PCR device is challenging due to its stringent requirements on reliability, cost, ease-of-use, portability, and speed of DNA amplification (Weidemaier et al. 2015). The existing process for POC PCR design suffers from several limitations, such as high cost, long development cycles, and intensive labors, since most of them solely rely on experiments and trial-and-error. Moreover, it is also formidable to maintain the same level of PCR performance in the field due to severe environmental uncertainties and limited functionalities and resources. Hence, exploiting the computational methods to enhance the POC PCR performance and design process efficiency is highly desirable (Shu et al. 2019a). One of

✉ Yi Wang
yiwang@cec.sc.edu

¹ University of South Carolina, Columbia, SC 29208, USA

² Old Dominion University, Norfolk, VA 23529, USA

the methods that recently attracts significant research attention is to use the computational fluid dynamics (CFD) to accelerate the design process of PCR devices. Krishnan et al. and Chen et al. utilized CFD to provide an insight into the buoyancy-driven flow induced in a PCR reactor, including velocity and temperature fields (Krishnan et al. 2002; Chen et al. 2004). Li et al. studied the flow conditions of several geometries of PCR capillary reactor design using CFD (Li et al. 2016), and Qiu et al. numerically analyzed the flow changes inside the PCR reactor in the vertical and horizontal positions (Qiu et al. 2019). Yariv et al. developed a mathematical model for DNA amplification, applied it to a very simple nondimensional geometry, and showed its potential to be used for estimating PCR performance (Yariv et al. 2005). Allen et al., Muddu et al., and Shu et al. adopted Yariv's mathematical model and applied it to more practical problems for DNA amplification (Allen et al. 2009; Muddu et al. 2011; Shu et al. 2019b). These studies have made exceptional contributions to various aspects of numerical analysis of PCR devices, and demonstrated the promise of partially replacing experimentation with CFD to improve the cost- and time-effectiveness of POC PCR development.

Besides the reduced cost, labor, and time, when combined with optimization, numerical simulations can find not only a feasible, but an optimal (or at least a pseudo-optimal) design with dramatically improved performance. However, in many cases the CFD-based design optimization is not preferred because each CFD simulation may easily take tens of minutes to several hours depending on the system complexity, which renders the unaffordable iterative optimization process that needs a large number of function calls to CFD. To tackle this issue, fast-running surrogate models have been developed to replace the demanding CFD simulations, and thus, the optimization can be performed in a manageable manner on the resource-limited computing platform (Carrillo et al. 2018; Ou et al. 2019; Wang et al. 2020; Herten et al. 2017). Among various surrogate modeling techniques, the machine learning approach based on the artificial neural network (ANN) has won the spotlight for their salient applicability and accuracy in high-dimensional parameter space. In addition, with the advent of deep learning, it is possible for the ANN to accurately model a highly complex system by utilizing deep architectures, e.g., multiple hidden layers (Stoecklein et al. 2017; Malkiel et al. 2018; Hong et al. 2020; Kasim et al. 2001). ANN-based optimization has proved its vast values in diverse areas, including hydrogen purification, design of a solar power plant, model predictive control, and prediction of smart grid energy consumption (Ye et al. 2019; Boukelia et al. 2016; Hong et al. 2019; Muralitharan et al. 2018).

In this paper, we propose a framework to construct a CFD simulation database for the double-heater convective PCR (DH-cPCR) device, to train ANN models using

the data from the database, and to perform ANN-based optimization to search for the optimal device configuration. Key optimization parameters, including the diameter, the aspect ratio, and the heights of the top and bottom heating surfaces that govern the thermofluidic behavior within the DH-cPCR and its DNA amplification performance (quantified by DNA doubling time), are considered in order to examine the feasibility and efficiency of ANN-based optimization. Contributions of the present study can be summarized in three aspects. First, a database of CFD simulation is generated using the in-house developed codes, which have already been verified by the experiments (Shu et al. 2019a, b; Qiu et al. 2019; Shu 2019) in our prior studies. The parameter space has been selected carefully in the practically relevant ranges (Li et al. 2016; Krishnan et al. 2002; Shu et al. 2019b; Qiu et al. 2017; Miao et al. 2020), and partitioned by the Latin hypercube sampling using nearly 10,000 sample points, where CFD simulations are conducted. A procedure and numerical programs have been developed to automate and parallelize the large number of CFD simulations and post-processing of the simulation results. Second, the fast-running ANN model is constructed based on the CFD database, which captures the varying landscape of the underlying CFD data and allows to explore the entire design domain continuously and economically, leading to an optimal solution. More importantly, a two-stage ANN modeling approach that includes a classifier and a regressor model is developed to analyze the data in series and address the issue associated with extreme data ranges of CFD database. Specifically, the classifier filters out infeasible configurations, and the regressor predicts the PCR performance (doubling time) of feasible configurations only. The curtailed data range ensures excellent predictive accuracy of the ANN model. Third, two device design case studies that combine the trained two-stage ANN models with different optimization methods are presented. In one of them, the genetic algorithm (GA), a global optimization technique is used to search for the optimal DH-cPCR configuration within the entire parameter space. GA is one of the evolutionary algorithms and finds the optimum by performing selection, crossover, and mutation on a set of population for numerous generations. The optimum and the data landscape around its vicinity are portrayed and confirmed with additional CFD simulations to ensure the robustness and the accuracy of the solution.

It should be noted that generating training CFD database and performing ANN-based optimization is more computationally efficient than the traditional, brute-force optimization method that solely relies on CFD simulation. This is because when the cost and constraint functions of the optimization are changed, the generated CFD database can be re-processed to create new surrogate models accordingly without the need for additional simulation, which however is

not feasible by the traditional method. CFD simulation when combined with the gradient-based optimization is either not able to fully exploit the high-performance computing facility or computationally prohibitive on the resource-restricted platform. In contrast, the process of constructing the CFD database for ANN training is highly parallelizable, resulting in effective resource utilization. Last, the compact size and ultrafast speed of the ANN model renders it well-suited for the global optimization method, in which lots of evaluations need to be conducted in parallel and in serial, nevertheless direct CFD simulation is almost impossible.

The remainder of this paper is organized as follows. Section 2 describes in detail the DH-cPCR device and its CFD model and numerical simulation. The process to construct a database through automated, parallelized CFD simulation and analysis is presented in Sect. 3. The two-stage ANN modeling and training process is introduced in Sect. 4. In Sect. 5, two device design case studies are presented to validate the proposed ANN-based optimization framework. The paper is concluded in Sect. 6.

2 Numerical simulation

The DH-cPCR device under consideration is composed of a capillary tube and a thermal/control module, as depicted in Fig. 1. The thermal/control unit has the top and bottom heating modules, and the thermal bridges connect the two

heating modules and transfer heat from the bottom to the top. The device is designed to maintain approximately 368.15 K near the bottom and 328.15 K near the top. Thermostats are utilized to read local temperatures and transmit the signal to the thermal controller. Since only one resistive heater is used as the heating source, the fan can be turned on to cool down the top heating module as needed and to maintain a local temperature of 328.15 K. The exposed surface of the capillary tube is enclosed by an insulator to minimize heat loss.

DNA amplification using PCR is based on the thermal cycling processes of three distinct temperatures: (1) denaturation, (2) annealing, and (3) extension. As demonstrated in Fig. 1, the denaturation process occurs near the bottom of the capillary tube, where the fluid temperature is maintained at approximately 368.15 K. During this process, a double-stranded DNA is separated into two single-stranded DNAs. Next an annealing process takes place near the top of the capillary tube, where the temperature is cooled down to approximately 328.15 K. At this stage, primers bind to the ends of the two single-stranded DNAs. Lastly, an extension process occurs in the middle area of the tube with the temperature varying between 365.15 K and 350.15 K, where enzymes actively support DNA synthesis, transforming single-stranded DNAs into double-stranded DNAs.

Two CFD models are then developed to, respectively, capture two important phenomena during the PCR amplification, (1) momentum and conjugate heat transfer (MCHT), and (2) DNA species transport and reaction kinetics. A 3-dimensional (3-D)

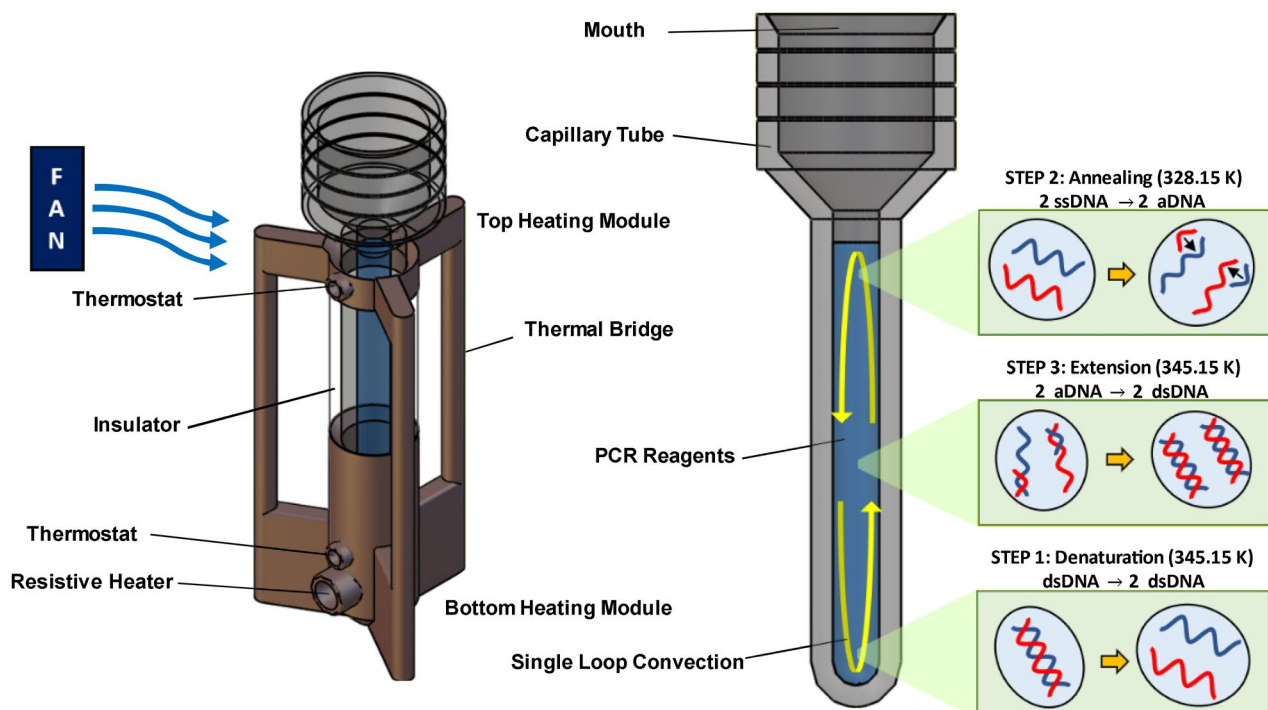


Fig. 1 The double heater convective polymerase chain reaction (DH-cPCR) device and DNA amplification process

computational domain consisting of structured meshes for fluid and solid regions is constructed as shown in Fig. 2. The MCHT model is computed in a coupled manner in both the solid and fluid regions to obtain the steady-state, spatial distribution of flow quantities and temperatures. Specifically, the natural convection-induced flow and both convective and conductive heat transfer is simulated in the fluid domain. In the solid domain, only heat conduction is considered. The temperature at the fluid–solid interface is determined by the balanced heat flux through the interface. In the second CFD model, the species transport and the DNA amplification kinetics are solved to determine the spatiotemporal fields of species concentrations in the fluid domain.

In the first step of CFD simulation, 3D heat transfer equations are used in both the fluid and solid domains, while the natural convection flow is only solved in the fluid domain. The PCR reagents are pure water basis used for DNA amplification (Li et al. 2016); thereby, in the numerical simulation, fluid motions are assumed to be Newtonian, steady-state, incompressible, and laminar. The fluid is governed by continuity, momentum, and energy equations as follows (Çengel et al. 2015):

$$\frac{\partial(\rho u_j)}{\partial x_j} = 0 \tag{1}$$

$$\frac{\partial(\rho u_i u_j)}{\partial x_j} = \rho g_i - \frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \left\{ \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \right\} \tag{2}$$

$$\rho C_p u_j \frac{\partial T_f}{\partial x_j} = \frac{\partial}{\partial x_j} \left(\kappa_f \frac{\partial T_f}{\partial x_j} \right) \tag{3}$$

where ρ denotes the fluid density dependent on the temperature; u is the flow velocity; x is the spatial coordinate; p is the pressure; g is the gravitational acceleration; μ stands for the fluid dynamic viscosity; C_p is the isobaric specific heat, which is temperature dependent; T_f is the fluid temperature; κ_f represents the fluid thermal conductivity, which is a function of the temperature; the subscripts i and j are based on the Einstein notation. All thermal properties (i.e., ρ , μ , C_p , and κ_f) used for the fluid domain are treated as the polynomial functions of the temperature, as shown in Table 1 (Shu et al. 2019a, b).

The heat transfer in the solid domains is solved by the thermal conduction equation (Holman 2002)

$$\frac{\partial}{\partial x_i} \left(\kappa_s \frac{\partial T_s}{\partial x_i} \right) = 0 \tag{4}$$

where κ_s and T_s denote the constant thermal conductivity and temperature applied for the solid domain, respectively. Note that both κ_s and T_s are a scalar quantity, and the subscript s does not imply Einstein summation. The capillary tube used in this study is made of the polymethyl methacrylate (PMMA), and therefore, a constant thermal conductivity of 0.22 W/(m K) is applied (Shu et al. 2019a, b).

The following boundary conditions are employed to solve Eqs. (1) to (4)

1. No-slip flow velocity condition at wall, $u_{wall} = 0$;

Fig. 2 Computational domains made up of structured meshes in both the solid and fluid regions

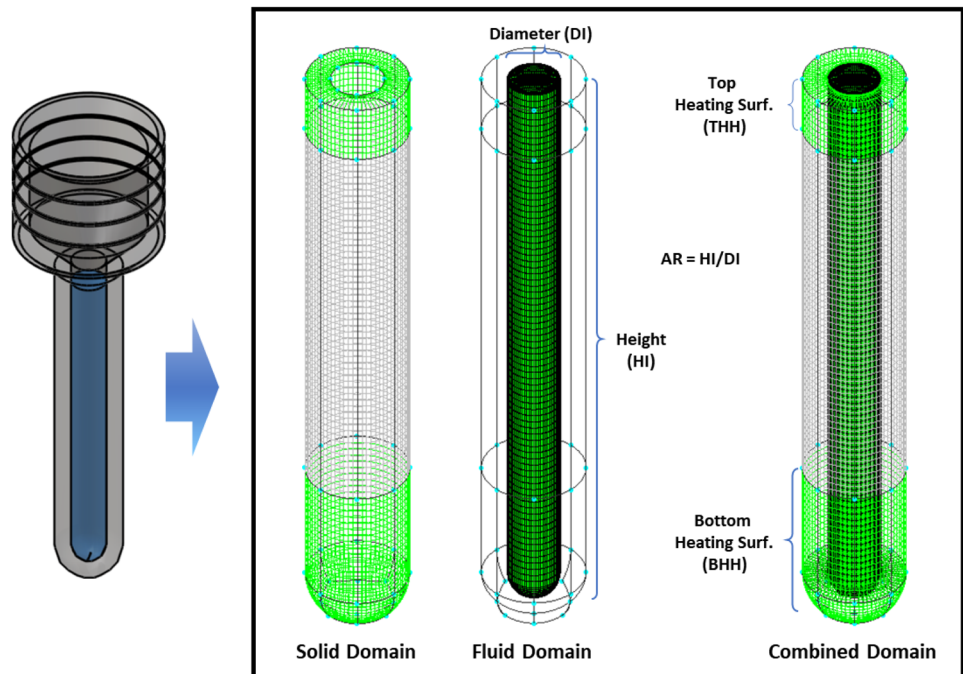


Table 1 Temperature-dependent fluid thermal properties in polynomials

| Thermal Properties | Unit | Polynomial Coefficients | | | | |
|----------------------------------|-------------------|-------------------------|-----------------|-----------------|-----------------|-----------------|
| | | 0 th | 1 st | 2 nd | 3 rd | 4 th |
| Density, ρ | kg/m ³ | 784.5 | 1.673 | -3.234e-03 | - | - |
| Specific Heat, C_p | J/(kg K) | 2.665e+04 | -2.567e+02 | -1.099 | -2.097e-03 | 1.507e-06 |
| Dynamic Viscosity, μ | Pa s | 8.764e-02 | -7.216e-04 | -1.993e-06 | -1.840e-09 | - |
| Thermal Conductivity, κ_f | W/(m K) | 4.581 | -5.784e-02 | 2.895e-04 | -6.071e-07 | 4.603e-10 |

2. Isothermal and isoflux conditions at the interface between the fluid and solid domains, i.e., $T_f = T_s$ and $\dot{q}_f = \dot{q}_s$ at the interface, where \dot{q} is the heat flux and the subscript f and s denotes the fluid and the solid, respectively;
3. Constant temperature $T_s = 368.15K$ and $T_s = 328.15K$ at the bottom and top boundaries of the solid domain, respectively;
4. Zero temperature gradient along the outer surface of the tube due to insulation, $\nabla T_s = 0$.

In the second step of CFD simulation, the unsteady convection–diffusion–reaction equations are solved to produce the spatiotemporal field of species concentrations and evaluate the performance of a convective PCR reactor. The equations are given by (Shu et al. 2019a; Yariv et al. 2005; Allen et al. 2009)

$$\frac{\partial c_{ss}}{\partial t} + u_j \frac{\partial c_{ss}}{\partial x_j} = D \frac{\partial^2 c_{ss}}{\partial x_j^2} + 2k_{df}(\vec{x})c_{ds} - k_{af}(\vec{x})c_{ss} \tag{5}$$

$$\frac{\partial c_a}{\partial t} + u_j \frac{\partial c_a}{\partial x_j} = D \frac{\partial^2 c_a}{\partial x_j^2} + k_{af}(\vec{x})c_{ss} - k_{ae}(\vec{x})c_a \tag{6}$$

$$\frac{\partial c_{ds}}{\partial t} + u_j \frac{\partial c_{ds}}{\partial x_j} = D \frac{\partial^2 c_{ds}}{\partial x_j^2} + k_{de}(\vec{x})c_a - k_{df}(\vec{x})c_{ds} \tag{7}$$

where c is the DNA species concentration; D denotes the diffusive coefficient; k is the constant rate of reaction kinetics, $f(\vec{x})$ is the reaction intensity at a local position of $\vec{x} = (x_1, x_2, x_3)$; the subscripts ss , a , and ds denote the single-stranded, annealed, and double-stranded DNAs, respectively; the subscripts d , a , and e represent the denaturation, annealing, and extension processes, respectively.

The diffusivity of typical DNA templates (100 base-pairs) is approximately $10^{-7} \text{ cm}^2/\text{s}$ (Yariv et al. 2005). The constant rates of the denaturation, annealing, and extension processes are 0.1 s^{-1} , 0.1 s^{-1} , and 0.05 s^{-1} , respectively for a conventional PCR cycle (Yariv et al. 2005). A Gaussian function is employed to map the reaction intensities onto the temperature field in the fluid domain (Allen et al. 2009). The details can be found in the authors’ previous work (Shu et al. 2019a, b).

The initial and boundary conditions applied to Eqs. (5) through (7) are as follows:

1. The initial concentrations of the double-stranded, single-stranded, and annealed DNAs are 100, 0, and 0 copies, respectively.
2. The species concentration do not penetrate the wall: $\hat{n} \cdot \nabla c_i = 0$, where the subscript i denotes the individual DNA species.

Once Eqs. (5)–(7) are solved to produce the species concentrations, the doubling time can be computed as follows (Allen et al. 2009; Shu et al. 2019b):

$$t_d = \frac{\ln(2)\Delta t}{\ln\left(\frac{c_{ds,t=t_f}}{c_{ds,t=0}}\right)} \tag{8}$$

where Δt is the duration, and $c_{ds,t=t_f}$ and $c_{ds,t=0}$ denote the concentrations for the double-stranded DNA at the final and initial time, respectively. The DNA doubling time is utilized to evaluate the performance of convective PCR reactors. The physical time simulated by the CFD model is set to 30 min (Shu et al. 2019a, b).

3 CFD database construction and analysis

As stated in the previous section, there are four optimization parameters of the double-heater cPCR (DH-cPCR) considered in this research: (1) diameter (DI) and (2) aspect ratio (AR) of the capillary tube, (3) top heater height (THH), and (4) bottom heater height (BHH). AR is defined as the ratio of the tube height to the diameter, and both heater heights are measured in percentage with respect to the height. For example, if THH and BHH are 10% and 5%, the height of THH and BHH is 10% and 5% of the tube height, respectively. The parameter space of the DH-cPCR is summarized in Table 2, and selected in such a manner that the doubling time of the majority of the configurations falls into the feasible category. Within this parameter space, Latin hypercube sampling technique was implemented to select 9,555 samples, which were processed by CFD simulations to compute the doubling time and evaluate the performance.

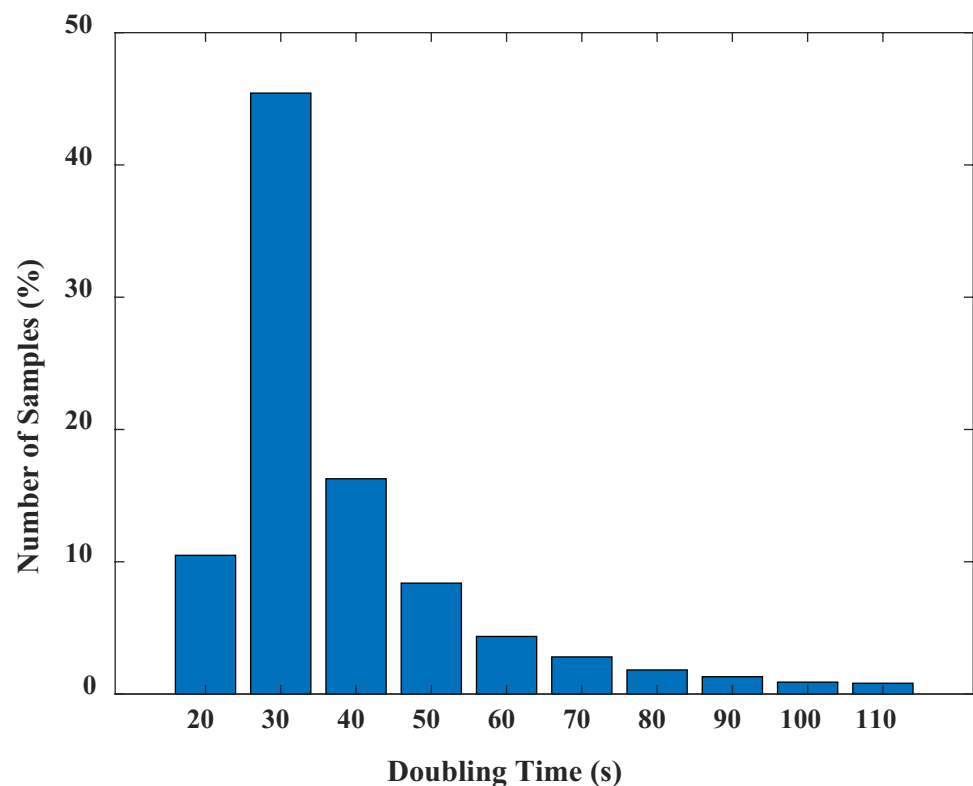
Table 2 Optimization parameter space of the DH-cPCR

| | Diameter (mm) | Aspect Ratio | Top Heater Height (%) | Bottom Heater Height (%) |
|---------|---------------|--------------|-----------------------|--------------------------|
| Minimum | 1.5 | 8 | 5 | 15 |
| Maximum | 2.0 | 12 | 35 | 35 |

Scripts and application programming interfaces (API) were developed to automate the entire process of mesh generation, CFD simulation, and database construction to reduce manual efforts and time. Specifically, a script developed using Pointwise Glyph 2 generates a total of 9,555 computational meshes at a speed of approximately 10 s/mesh, one for each configuration. Then, a bash script creates a number of file structures corresponding to each configuration, including the mesh and model configuration files for the first step of the CFD simulation, where the MCHT equations are solved. The script automatically runs `chtMultiRegionSimpleFoam`, one of the OpenFOAM CFD solvers to obtain temperature and velocity fields inside capillary tubes. Similarly, another group of file structures are automatically created using the script to prepare for the second step of CFD simulations. The temperature and velocity fields obtained from the first step are transferred to the directory of the second CFD simulation and used as the background flow and thermal field for the PCR amplification reaction. The bash script automatically runs `convDiffFoam`, OpenFOAM based in-house

CFD code, and the DNA doubling time is computed based on the DNA concentration as discussed above. High performance computing (HPC) was utilized to perform such a large number of computations. The hardware is made up of Intel® Xeon® CPU @ 2.1 to 2.5 GHz of 16 to 32 processors per node with 128 GB of RAM. In this study, a single CFD simulation employed 16 processors for parallel computing.

Then the library of DH-cPCR CFD simulation was constructed, which consists of various device configurations and their corresponding doubling time, t_d . As discussed previously, configurations with shorter t_d is preferred since the purpose of a PCR is to duplicate DNAs rapidly, and therefore, the speed is the primary measure of its performance. It is also essential to examine if the selected parameter space is targeted towards the smaller t_d since it is not worthwhile to have a database full of infeasible device designs that cannot duplicate DNAs within a reasonable amount of time. Figure 3 presents the histogram of the sampled CFD simulations, i.e., the number of samples in different groups of the doubling time stored in the database. Considering the fact that the shortest t_d is approximately 20 s, device designs with the doubling time less than 60 s will be of our interest. From Fig. 3, we can observe that approximately 85% of the total samples fall into this region. This ensures that the selected parameter space was able to generate the database of sampled CFD simulations that are mostly useful. It is important to note that there are samples with t_d larger than 110 s, which are not shown in the figure. They are omitted since they only

Fig. 3 Histogram of the number of samples in different categories of doubling time

make up of approximately 7.4% of the entire data, the highest of which reaches up to 200,000 s and loses the practical relevance.

In order to have more insight into the database, the relationship between the optimization parameters and the doubling time is analyzed in four sections/sub-ranges as shown in Fig. 4. The four sub-ranges are: (1) $t_d \leq 30$ s; (2) $30 < t_d \leq 40$ s; (3) $40 < t_d \leq 60$ s; and (4) $60 < t_d$, which are colored in red, green, blue, and cyan, respectively. Figure 4a shows the scatter plots of the diameter and the aspect ratio of the sampled designs in the four sub-ranges, and Fig. 4b is for the top and bottom heater heights. It indicates that t_d less than 60 s can be achieved with any size of the capillary tube explored in this study. However, it is evident from Fig. 4 that to achieve t_d less than 30 s, a smaller diameter is preferred. Moreover, fewer cases with large t_d (greater than 60 s) is observed near the diameter of 1.6 mm. This is important since it not only provides key insight into the design factors, leading to a shorter t_d , but also identifies the regions within the parameter space where the designs are more robust, and large t_d is less likely to occur. Similar trends can be observed with heater heights. Although it is possible to achieve feasible t_d for various heater heights, for the desirable t_d (less than 30 s), the top heater height must be larger than approximately 15%. Also for small values of the top heater height, it is likely that the design will yield t_d larger than 60 s.

4 Artificial neural network modeling

As described in Sect. 3, analysis of the scatter plots showing the relationship between the sampled optimization parameters and the doubling time using the CFD simulation data, implies the qualitative rules how the DH-cPCR needs to be

designed. Nevertheless, building machine learning models using the library of CFD data allows accurate and quantitative representation of the relationship, which is fast-to-evaluate and can be integrated with optimization engines to iteratively search for optimal configuration. Among various machine learning techniques, artificial neural network (ANN) has emerged as one of the most popular and reliable approaches due to its ability to identify complex behaviors for a variety of engineering applications. Consequently, CFD database above is employed to construct the ANN model that can predict DH-cPCR performance at different optimization parameters. The methods and procedures for ANN model construction is detailed below.

4.1 Data preprocessing

Upon ANN training, it is necessary to examine and pre-process the data to improve the model accuracy. Figure 5 shows t_d of all the sampled CFD simulations in an ascending order. The largest t_d of the samples has the value close to 260,000 s, which is too large relative to the feasible design objective. In addition, beyond the data index of 9,000, t_d begins to rise substantially, even when the figure is plotted in a log scale. This small portion of unrealistic data with excessively large t_d has no utility to practically relevant design and could even jeopardize the model accuracy if included. Therefore, it is beneficial to build a classifier and filter out samples and corresponding data that are distant from the regime of feasible device configuration before quantitative modeling. Therefore, the first step is to identify and filter out the distant samples. Here, a K-means clustering algorithm is applied, which clusters the data into groups in an unsupervised fashion based on the sample distances from the cluster centers. Moreover, it is able to identify the outliers.

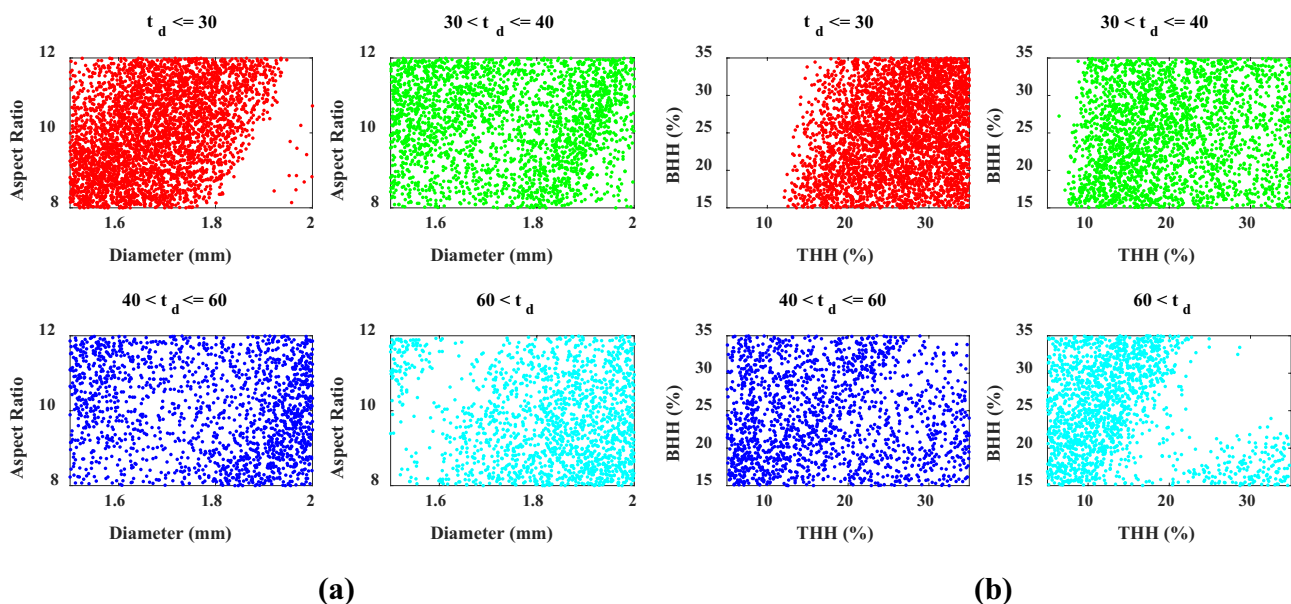


Fig. 4 Scatter plots of the (a) diameter and aspect ratio, and (b) top and bottom heater heights in the four subranges of the doubling time

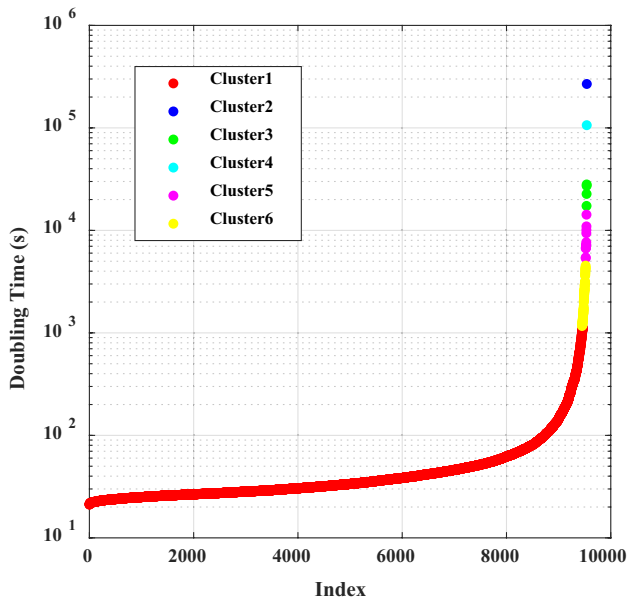


Fig. 5 K-means clustering result of the doubling time data, expressed in a log scale

Figure 5 also portrays six different groups of t_d classified by the K-means clustering algorithm. The total number of clusters was chosen simply by repeating the K-means clustering while increasing the number of clusters, until the range of feasible configuration is reduced by at least two orders of magnitude and the number of infeasible samples (outliers) remains less than 1% of the entire dataset. Samples in the 1st cluster that make up 99% of the dataset and accommodate the feasible configuration regime were labeled positive, and used to build a quantitative regression model. Samples from other 5 clusters were labeled negative, filtered out, and excluded from the regression model training. Furthermore, data was also normalized since all optimization parameters have distinctly different ranges. Min–max scaling was performed on all the parameters to put them in the range of 0 and 1. However, for the doubling time, which is the output of the regression model, log scaling was performed before min–max scaling to more evenly distribute the data in the normalized range.

4.2 ANN model training

As described in Sect. 4.1, two different ANN models; classifier and regressor, were trained sequentially in two stages. Again the former is used to filter out unrealistic device configuration candidates; and only the performance (doubling time) of the realistic configuration is predicted by the regressor. The structures of ANN models trained are illustrated in Fig. 6. Since the classifier is comparably easier to model, it is assigned with only one hidden layer, whereas three hidden layers are used for the regressor because the quantitative relationship between the inputs and the output is more complicated. In fact, a simple iterative analysis (not shown) was performed and it was found that three hidden layers produced better predictions compared to one or two hidden layers. For both ANNs, inputs are the four optimization parameters. The output of the classifier represents a probability of being in Cluster 1, denoted as P_{C1} . If P_{C1} is greater than or equal to a threshold value, the configuration belongs to Cluster 1 and vice versa. On the other hand, the output of the regressor is t_d , which is the performance metric of the DH-cPCR. Activation functions of all hidden layers are hyperbolic tangent functions. A sigmoid activation function is implemented for the output layer of the classifier to restrict the range to be between 0 and 1. No activation function is applied to the output layer of the regressor since the output value of the doubling time must be continuous.

For both training processes, the Levenberg–Marquardt optimization algorithm was used with a maximum of 500 training epochs. In addition, the training process was designed to stop whenever the validation performance stops improving for 6 consecutive iterations to prevent from overfitting. From the entire data set, 8,122 samples (equivalent to 85%) were used for training, and 1,433 samples (remaining 15%) were used for testing. Within the training set, 80% and 20% were used for training and validation, respectively. Due to the randomness of ANN training associated with network weight initialization, the training process was repeated 100 times (for both classifier

Fig. 6 Structures of ANN models

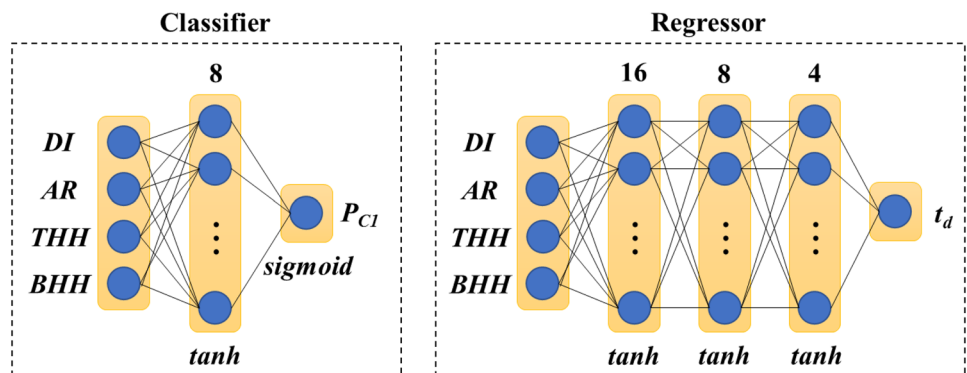


Table 3 Confusion matrix for feasibility classification; positive if feasible and negative if infeasible

| | Actual Positive | Actual Negative |
|--------------------|-----------------|-----------------|
| Predicted Positive | 1417 | 1 |
| Predicted Negative | 1 | 14 |

and regressor) and the networks with the best accuracy were selected in order to mitigate the issue of the local minima.

The test accuracy of the ANN classification model is shown by a confusion matrix in Table 3. If the configuration belongs to Cluster 1, then it is classified as positive. Otherwise, it is classified as negative. As seen from the table, only two configurations are misclassified, resulting in an accuracy of greater than 99%. The two misclassified configurations possess t_d of 634 and 1,176 s, which are far away from the doubling time of interest. The result of t_d prediction by the ANN regression model is depicted in Fig. 7, where the actual t_d is graphed against the predicted value. In other words, close alignment of the scatter points along $y = x$ line (solid black curve) indicates salient accuracy of prediction. Figure 7 shows the result over the entire range on the left, and on the right, it illustrates the solution within the range of interest, which is less than 60 s. Within the range up to approximately 1,000 s, the predictions are accurate with a mean absolute error (MAE) of less than 1 s. Within the range of interest (less than 60 s), some outliers of prediction are observed. However, the majority of the predictions remain within a small distance from the black curve and the MAE of these points is evaluated to be less than 0.5 s.

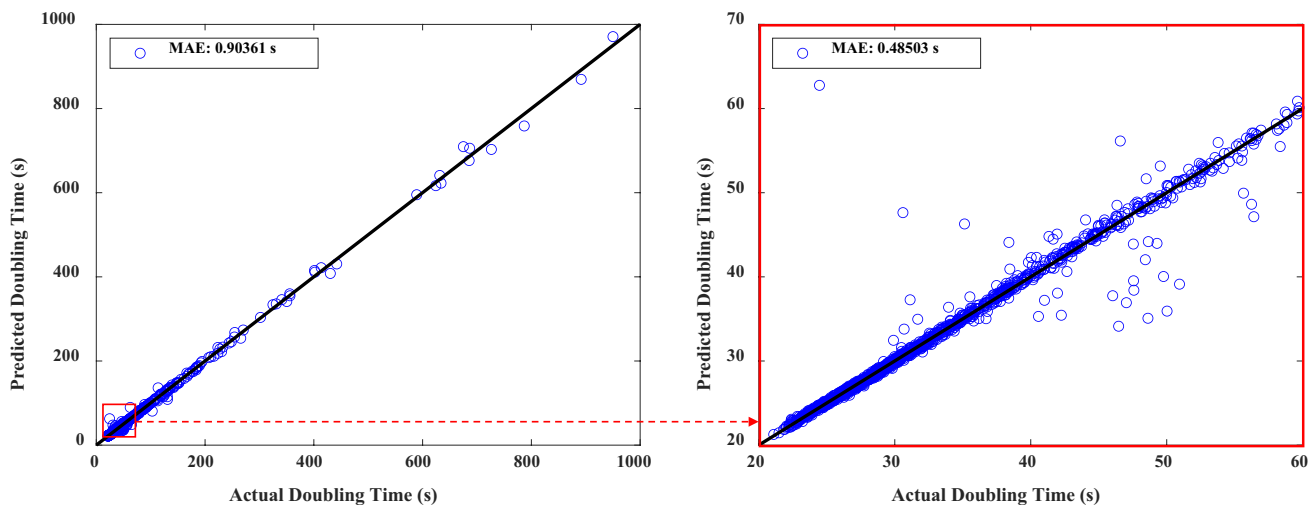


Fig. 7 Doubling time predictions of the ANN regression model

5 Results and discussion

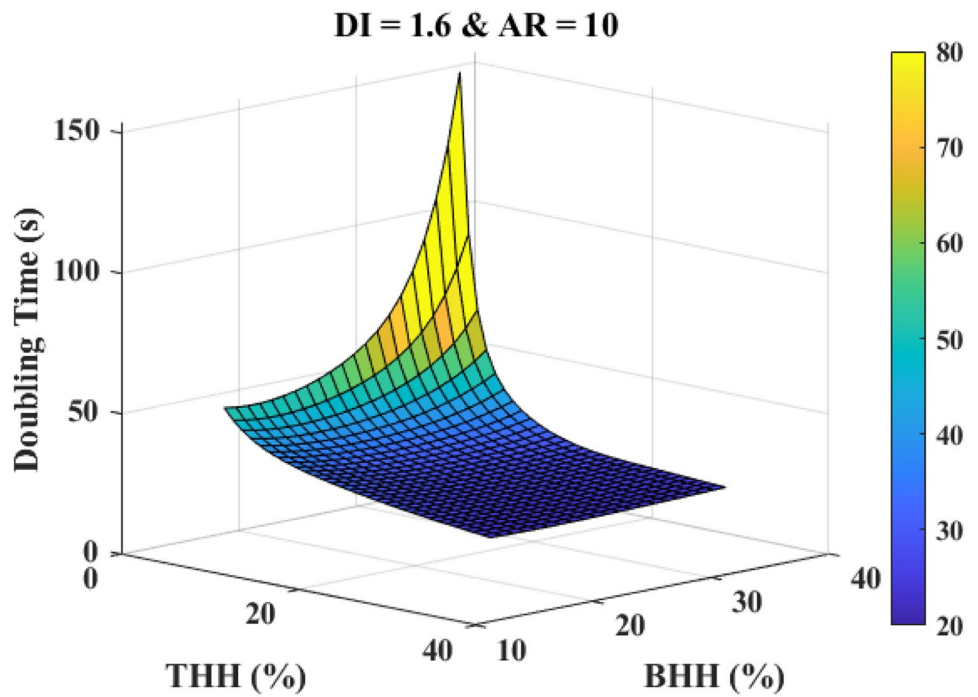
In this section, two different case studies are presented that combine the two-stage ANN models trained in Sect. 4 with different optimization techniques to obtain optimal configuration of DH-cPCR. For the first case study, the task is to find the optimal THH and BHH based on an existing capillary tube. In other words, variables DI and AR are fixed to be 1.6 mm and 10, respectively. The second case study aims to find the optimal configuration that treats all four parameters as device design variables. Moreover, a volume constraint is applied during the optimization to place it within the practical bounds of the capillary tube configurations.

5.1 Case study 1: optimal configuration of heater heights

First, the response surface of the trained ANN model is depicted over the entire parameter range of THH and BHH. Since DI and AR are fixed, the dimension of the optimization parameter space is only 2, and the response can be visualized easily through the surface plot as presented in Fig. 8. From the figure, it can be observed that the optimal configuration, i.e., the smallest doubling time is favored by comparably large THH and small BHH. Furthermore, the surface seems smooth enough to implement gradient based optimization to locate the minimum, which eliminates the need for global optimization techniques that rely on randomization to avoid local optima. Consequently, a constrained gradient-based optimization algorithm—a sequential quadratic programming (SQP) is implemented directly on the ANN model to find the optimum. The optimization problem is formulated as follows:

$$\min_x f(c, x) \text{ s.t. } x_{\min} \leq x \leq x_{\max} \tag{9}$$

Fig. 8 The surface plot of doubling time vs. THH and BHH for fixed DI and AR



where f is the trained ANN model, c is the constant value of DI and AR, and x is the vector of optimization parameters: THH and BHH, and their bounds for optimization is listed in Table 2.

It took 14 iterations to reach the minimum with 46 function evaluations in total. The minimum doubling time is found to be 23.50 s and the corresponding optimal heights are 35.00% and 19.98% for top and bottom heaters, respectively. In order to confirm the optimal solution found by the ANN- and

gradient-based optimization, the optimum and 20 points in its vicinity were sampled randomly and the CFD simulations were performed on these samples. It is essential to note that the neighboring points near the solution must be confirmed by CFD simulation as well to ensure the robustness of the configuration and evaluate model uncertainties. In Fig. 9, we can visualize the surface plot predicted by the ANN model, in which the results of CFD validation are also plotted in circles.

Fig. 9 CFD validation of the ANN-based SQP optimization result. The surface plot represents the ANN-predicted doubling time and the circle markers are the results of CFD validation points. The red marker indicates the optimal configuration and black markers are randomly sampled points in the vicinity of the optimum

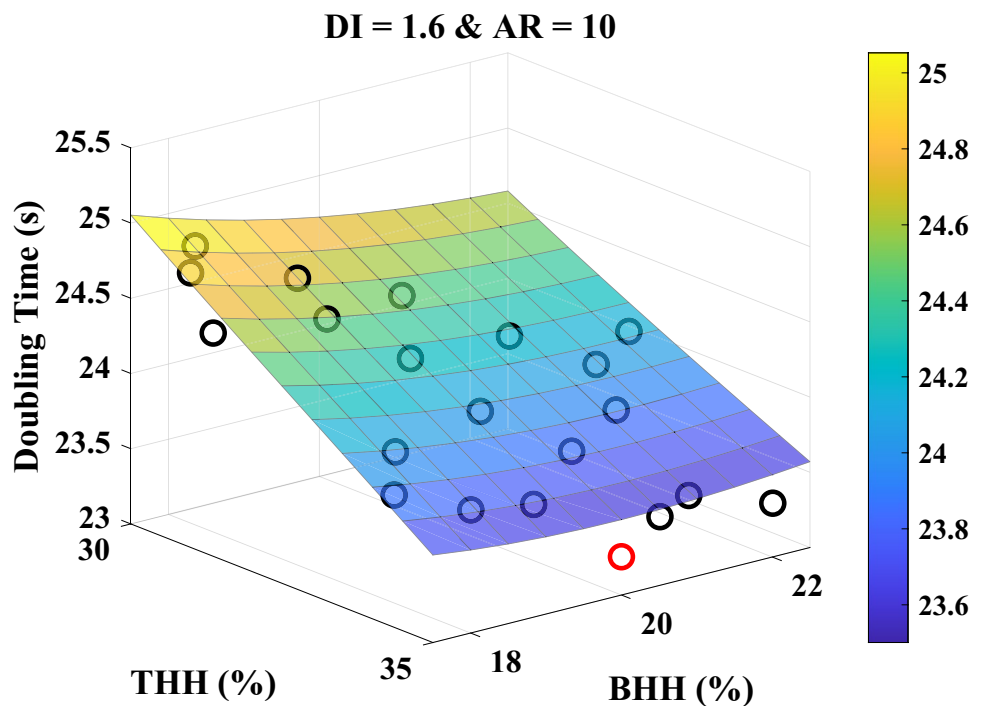
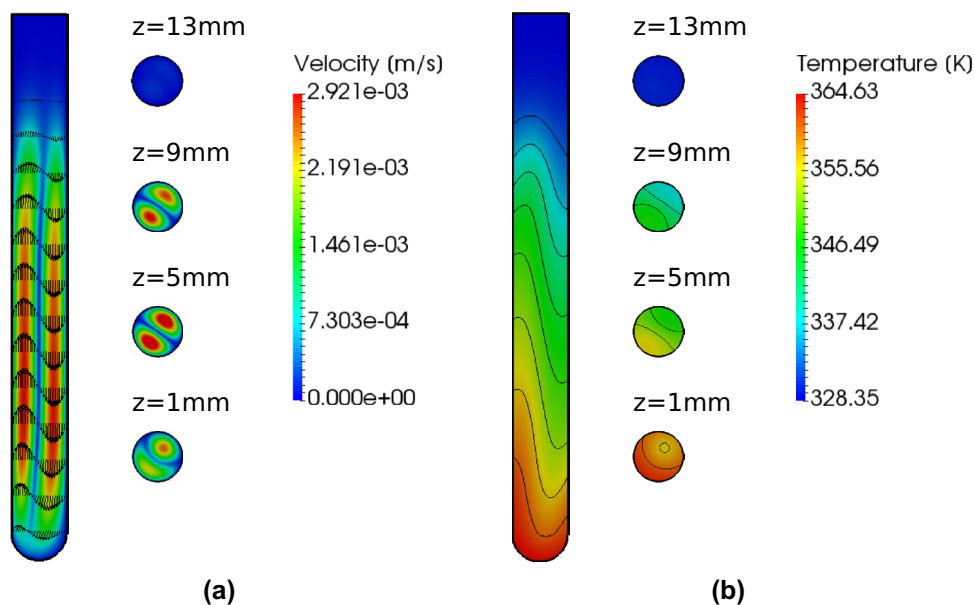


Fig. 10 Results of CFD simulation at the optimal configuration point: (a) velocity contour, and (b) temperature contour



The red circle represents the CFD simulation of the optimal configuration and the black circles are for sample points in the vicinity of the optimum. Compared to the response surface of the ANN model, the CFD validations tend to provide slightly smaller doubling time, as all the circles are below the surface. Despite the minor error, the ANN response surface follows the same trend of the CFD simulations. The average MAE of the validation points is found to be 0.21 s, which is almost negligible relative to the true value, 23.25 s of the doubling time of the optimal configuration. In addition, compared to the optimal configuration, all its neighboring points exhibit larger values of the doubling time, which confirms that the optimal configuration yields the best performance.

Figure 10 shows the result of the CFD simulation at the optimal configuration point in this case study. As seen in the velocity contour plot, a single convection loop is developed in the clockwise direction. The strongest local velocity is created in the middle and lower regions, which is approximately 2.92 mm/s. Stationary flow, known as the dead zone causing degraded performance, is observed near the top due to the negligible local temperature gradient in order to maintain the isothermal condition. The maximum and minimum temperatures are found to be 364.63 K and 328.35 K in the vicinity of the bottom and the top, respectively.

5.2 Case study 2: optimal configuration of the whole DH-cPCR with volume constraint

In the second case study, all four device configurations are optimized to find the DH-cPCR that results in the shortest doubling time. At the same time, a nonlinear constraint of the reactor volume is imposed. The new optimization problem is formulated as follows:

$$\min_x f(x) \text{ s.t. } x_{\min} \leq x \leq x_{\max} \quad 25 \mu\text{l} \leq V(x) \leq 40 \mu\text{l} \quad (10)$$

where f is the ANN model; x is the vector of all four optimization parameters, DI, AR, THH, and BHH; and V is the function of the capillary tube volume. Again, the bounds of all variables are listed in Table 2. For this particular problem, a genetic algorithm (GA) was selected for use rather than the gradient-based optimization algorithms to ensure the global optimal solution can be found. This is because a higher dimensional space is prone to multiple local optima, and the gradient-based approaches may not reach the global optimum. GA was performed for 50 generations with a population size of 50. In addition, scattered crossover, Gaussian mutation and stochastic uniform selection approaches were implemented for the GA process. The solution, however, converged after the 3rd generation with a total number of 7718 function evaluations. The optimal configuration found is listed in Table 4 with the predicted doubling time.

In order to validate the solution, optimal configuration and the points in its vicinity were selected and confirmed by the CFD simulation, which are shown in Fig. 11. Two surface plots are graphed with two pairs of optimization parameters while keeping the other two constant to facilitate visualization of the responses. For each plot, 20 randomly selected neighboring points for validation are presented. Similar to the previous case study, the ANN predictions tend to overestimate the doubling time slightly. Despite a small bias error, the ANN model

Table 4 Optimal configuration found by GA using the ANN model

| DI (mm) | AR | THH (%) | BHH (%) | Volume (μl) | Estimated t_d (s) |
|---------|------|---------|---------|-------------|---------------------|
| 1.56 | 8.01 | 35.00 | 25.81 | 25.02 | 21.31 |

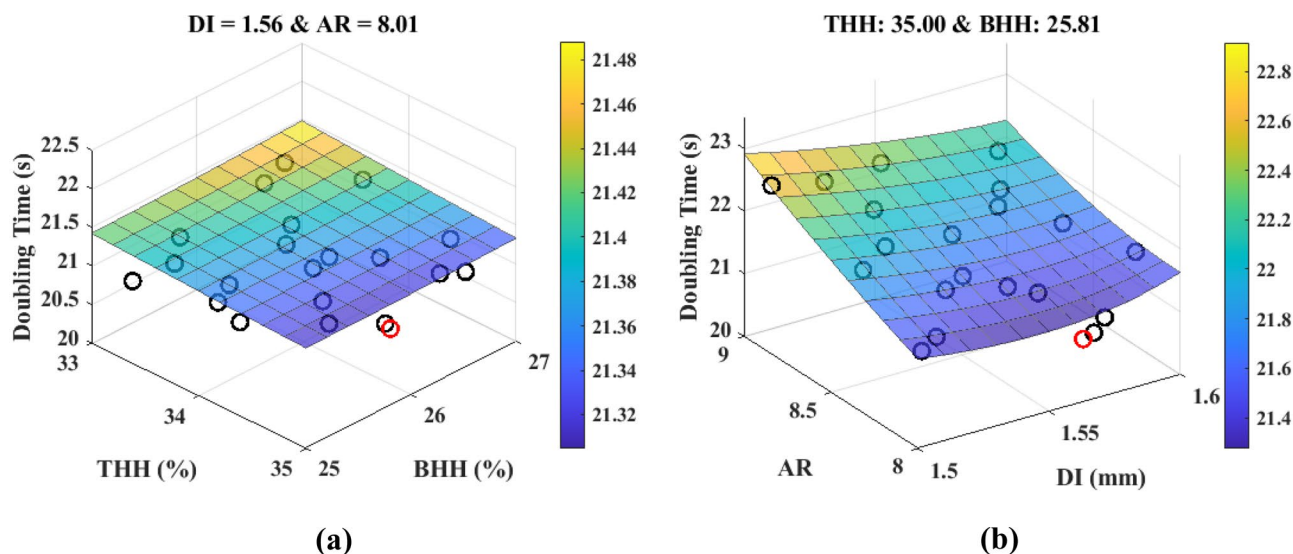


Fig. 11 CFD validation of the GA optimization result: (a) t_d against THH and BHH, and (b) t_d against DI and AR. The surface plot represents the ANN predicted doubling time and the circle markers are the

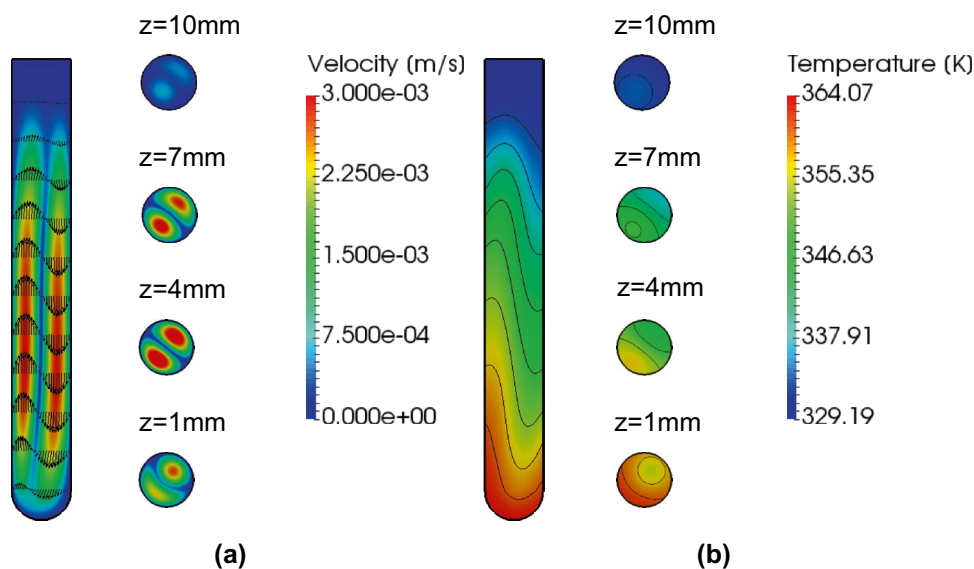
CFD validation points. The red marker indicates the optimal configuration and black markers are its vicinity points

captures the trend of both response surfaces. Moreover, the doubling time of the optimal configuration is found to be smaller than any neighboring points that satisfies the volume constraint. The average MAE computed for all validation points is 0.26 s and the true doubling time of the optimum is 20.99 s. It should be noted that the doubling time in this case study is shorter than that in the previous study because two additional parameters, i.e., DI and AR are included in the optimization, allowing broader search in higher dimensional parameter space for better device design. The results confirm that the ANN

model in combination with GA is able to find the optimal solution.

Figure 12 illustrates the CFD results of the optimal configuration specified in this case study. Again a single convection loop is clearly observed in the fluid domain. The maximum flow velocity is detected in the middle of the capillary tube with a value of 3.0 mm/s. It is found that the area of the stationary flow is relatively smaller than that in the first case study, which again indicates that this configuration outperforms the one in the first case study where only two parameters are included in

Fig. 12 Results of CFD simulation at the optimal configuration point: (a) velocity contour, and (b) temperature contour



optimization. In addition, the temperature is distributed in a desired manner over the entire domain, leading to the shorter DNA doubling time. In this case study, the maximum and minimum temperatures are found to be 364.07 K and 329.19 K, respectively.

6 Conclusion

This paper presents a framework for automated and efficient optimization of the POC DH-cPCR device using CFD simulation database and ANN model to reduce the development cost and efforts. Within DH-cPCR the temperature variation that induces stable single convection loop for DNA amplification is controlled by two heaters located at the top and the bottom of the capillary tube. Thus, four optimization parameters including the diameter (DI) and aspect ratio (AR) of the capillary tube, and the top heater height (THH) and bottom heater height (BHH) are considered. The parameter space is then partitioned by the Latin hypercube sampling, yielding a total of 9,555 DH-cPCR reactor configurations that are evaluated by CFD simulation. A fully automated process is established to streamline multi-blocked mesh generation, model configuration, parallel CFD computation, and post-processing to fully utilize the computing resources and enhance device efficiency. Each configuration is assessed by a two-step CFD simulation, first of which obtains the flow profiles induced by the temperature distributions within the reactor; and the second quantifies the DNA concentrations and doubling time of the individual device configuration. To make best use of the CFD database, a machine learning model, specifically, an artificial neural network (ANN) is developed using the CFD data. The ANN model builds the mapping relationship between the device configuration and the doubling time, which can be evaluated at extremely fast speed (a fraction of a second) to replace CFD simulations during the device optimization. The ANN consists of two sub-models, a classifier and a regressor, which are trained and used in two stages. The classifier first identifies infeasible configurations that manifest inordinately large doubling time and excludes them from the second regressor stage. Then the regressor model evaluates the doubling time of the configurations that are classified as feasible in the first stage. The classification accuracy is >99%, and the mean absolute error (MAE) of the doubling time prediction is less than 1 s.

Two case studies are carried out to verify the automatically constructed CFD database and ANN-based optimization. The first aims to find the optimal heater heights given a fixed configuration of the capillary tube. A sequential quadratic programming is implemented to search for the configuration with the minimum doubling time. The optimal THH and BHH are found to be 35.00% and 19.98%, respectively. Additional CFD simulations are also conducted at the vicinity of the optimal configuration to verify robustness and accuracy of the

proposed approach, and it is found that the average MAE of predicted doubling time is 0.21 s. The second case study is more challenging and seeks the optimal configuration within the four parameter space along with a practical volume constraint of the capillary tube. Because of the ultrafast speed of the ANN model, a GA is adopted for this study. It identifies an optimal solution of $DI = 1.56$ mm, $AR = 8.01$, $THH = 35.00\%$, and $BHH = 25.81\%$. Likewise, the optimal solution and the response surface around its vicinity are verified by the CFD simulations and the MAE of doubling time is 0.26 s. The results of both case studies verify the accuracy of the CFD database and the ANN models, and demonstrate that the optimal device configuration of the DH-cPCR can be successfully found by the ANN-based optimization. The proposed method and framework represents an accurate and efficient solution to accelerate DH-cPCR device design, and can potentially, be extended to different types of POC PCR devices.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- A. Niemi, T.M. Ferguson, D.S. Boyle, Point-of-care nucleic acid testing for infectious diseases. *Trends Biotechnol.* **29**(5), 240–250 (2011)
- B. Wang, B. Xie, J. Xuan, K. Jiao, AI-Based optimization of pem fuel cell catalyst layers for maximum power density via data-driven surrogate modeling. *Energy Convers Manage* **205**, 112460 (2020)
- D. Stoecklein, K.G. Lore, M. Davies, S. Sarkar, B. Ganapathysubramanian, Deep learning for flow sculpting: insights into efficient learning using scientific simulation data. *Scientific Reports* **7**(1), (2017)
- E. Yariv, G. Ben-Dov, K.D. Dorfman, Polymerase chain reaction in natural convection systems: a convection-diffusion-reaction model. *Europhys Lett* **71**(6), 1008–1014 (2005)
- F. Ye, S. Ma, L. Tong, J. Xiao, P. Benard, R. Chahine, Artificial neural network based optimization for hydrogen purification performance of pressure swing adsorption. *Int J Hydrogen Energy* **44**(11), 5334–5344 (2019)
- G. Miao, L. Zhang, J. Zhang, S. Ge, N. Xia, S. Qian, D. Yu, X. Qiu, Free convective PCR: from principle study to commercial applications - a critical review. *Anal Chim Acta* **1108**(29), 177–197 (2020)
- I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, H. Suchowski, Plasmonic nanostructure design and characterization via deep learning. *Light Sci Appl* **7**(1), 1–8 (2018)
- J.A.E. Carrillo, F.J.S. de La Flor, J.M.S. Lissén, Single-phase ejector geometry optimisation by means of a multi-objective evolutionary algorithm and a surrogate CFD model. *Energy* **164**, 46–64 (2018)
- J.I. Shu, O. Baysal, S. Qian, X. Qiu, Computational design of a single heater convective polymerase chain reaction for point-of-care. *J Med Devices* **13**(4), 041007 (2019)
- J.P. Holman, *Heat Transfer*, 9th edn. (McGraw-Hill, New York, 2002).

- J.I. Shu, Computational analysis and design optimization of convective PCR devices. Old Dominion University, (2019)
- J.I. Shu, O. Baysal, S. Qian, X. Qiu, F. Wang, Performance of convective polymerase chain reaction by doubling time. *Int J Heat Mass Transf* **133**, 1230–1239 (2019)
- J.V.D. Herten, T.V. Steenkiste, I. Couckuyt, T. Dhaene, *Surrogate Modelling with Sequential Design for Expensive Simulation Applications* (IntechOpen, In Computer Simulation, 2017).
- J.W. Allen, M. Kenward, K.D. Dorfman, Coupled flow and reaction during natural convection PCR. *Microfluid Nanofluid* **6**(1), 121–130 (2009)
- K. Muralitharan, R. Sakthivel, R. Vishnuvarthan, Neural network based optimization approach for energy demand prediction in smart grid. *Neurocomputing* **273**, 199–208 (2018)
- K. Weidemaier, J. Carrino, A. Curry, J.H. Connor, A. Liebmann-Vinson, Advancing rapid point-of-care viral diagnostics to a clinical setting. *Future Virol* **10**(3), 313–328 (2015)
- M. F. Kasim, D. Watson-Parris, L. Deaconu, S. Oliver, P. Hatfield, D. H. Froula, and J. Topp-Muggleston, "Up to two billion times acceleration of scientific simulations with deep neural architecture search." (2020)
- M. Krishnan, V.M. Ugaz, M.A. Burns, PCR in a rayleigh-benard convection cell. *Science* **298**(5594), 793–893 (2002)
- M. Ou, L. Yan, W. Huang, T.T. Zhang, Design exploration of combinational spike and opposing jet concept in hypersonic flows based on CFD calculation and surrogate model. *Acta Astronaut* **155**, 287–301 (2019)
- P.B. Lippa, C. Müller, A. Schlichtiger, Point-of-care testing (Poct): current techniques and future perspectives. *Trends Anal Chem* **30**(6), 887–898 (2011)
- R. Muddu, Y.A. Hassan, V.M. Ugaz, Chaotically accelerated polymerase chain reaction by microscale rayleigh-benard convection. *Angew Chem Int Ed* **50**(13), 3048–3052 (2011)
- S.K. Sia, L.J. Kricka, Microfluidics and point-of-care testing. *Lab Chip* **8**(12), 1982–1983 (2008)
- S.H. Hong, J. Cornelius, Y. Wang, K. Pant, Fault compensation by online updating of genetic algorithm-selected neural network model for model predictive control. *SN: Applied Sciences* **1**(11), 1488 (2019)
- S.H. Hong, H. Yang, Y. Wang, Inverse design of microfluidic concentration gradient generator using deep learning and physics-based component model. *Microfluid Nanofluid*. **24**(6), (2020)
- S. Petralia, S. Conoci, PCR Technologies for point of care testing: progress and perspectives. *ACS Sensors* **2**(7), 876–891 (2017)
- T.E. Boukelia, O. Arslan, M.S. Mecibah, ANN-based optimization of a parabolic through solar thermal power plant. *Appl Therm Eng* **107**, 1210–1218 (2016)
- T.M. Powledge, The polymerase chain reaction. *Adv Physiol Educ* **28**(2), 44–50 (2004)
- X. Qiu, J.I. Shu, O. Baysal, J. Wu, S. Qian, S. Ge, K. Li, X. Ye, N. Xia, D. Yu, Real-time capillary convective PCR based on horizontal thermal convection. *Microfluid Nanofluid* **23**(3), 39 (2019)
- X. Qiu, S. Ge, P. Gao, K. Li, S. Yang, S. Zhang, X. Ye, N. Xia, S. Qian, A smartphone-based point-of-care diagnosis of H1n1 with microfluidic Convection PCR. *Microsystem technologies-micro- and nanosystems-information storage and processing systems* **23**(7), 2951–2956 (2017)
- Y.A. Çengel, A.J. Ghajar, M. Kanoglu, *Heat and Mass Transfer: Fundamentals and Applications*, 5th edn. (McGraw Hill Higher Education, New York, 2015).
- Z.Q. Li, Y. Zhao, D.W. Zhang, S.L. Zhuang, Y. Yamaguchi, The development of a portable buoyancy-driven PCR system and its evaluation by capillary electrophoresis. *Sens Actuators B Chem* **230**, 779–784 (2016)
- Z.Y. Chen, S.Z. Qian, W.R. Abrams, D. Malamud, H.H. Bau, Thermosiphon-based PCR reactor: experiment and modeling. *Anal Chem* **76**(13), 3707–3715 (2004)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.