

# **High Fidelity Machine Learning-Assisted False Positive Discrimination in Loop-Mediated Isothermal Amplification Using Nanopore-Based Sizing and Counting**

Ming Dong<sup>1</sup>, Aneesh Kshirsagar<sup>1</sup>, Anthony J. Politza<sup>2</sup> and Weihua Guan<sup>1,2\*</sup>

<sup>1</sup> Department of Electrical Engineering, Pennsylvania State University, University Park, Pennsylvania 16802, United States

<sup>2</sup> Department of Biomedical Engineering, Pennsylvania State University, University Park, Pennsylvania 16802, United States

\* Correspondence should be addressed to wzg111@psu.edu

## **Abstract**

Loop-mediated isothermal amplification (LAMP) is a rapid, sensitive, and cost-effective method for developing point-of-care nucleic acid testing due to its isothermal nature. Yet, LAMP can suffer from the issue of false positives, which can compromise the specificity of the results. LAMP false positives typically arise due to contamination, non-specific amplification, and non-specific signal reporting (intercalating dyes, colorimetric, turbidity, etc.). While dye-labeled primers or probes have been introduced for multiplexed detection and enhanced specificity in LAMP assays, they carry the risk of reaction inhibition. This inhibition can result from the labeled primers with fluorophores or quenchers, and probes that do not fully dissociate during reaction. This work demonstrated a nanopore-based system for probe-free LAMP readouts by employing amplicon sizing and counting, analogous to an electronic version of gel electrophoresis. We first developed a model to explore LAMP kinetics, and verified distinct patterns between true and false positives via gel electrophoresis. Subsequently, we implemented nanopore sized counting, and calibrated the event charge deficit (ECD) values and frequencies to ensure a fair analysis of amplicon profiles. This sized counting method, integrated with machine learning, achieved 91.67% accuracy for false positive discrimination, enhancing LAMP's reliability for nucleic acid detection.

## **Keyword**

Loop-mediated isothermal amplification, false positive, nanopore, sizing, counting

Rapid and accurate nucleic acid detection methods are crucial for disease diagnostics and environmental monitoring. Among the existing nucleic acid detection methods, loop-mediated isothermal amplification (LAMP) stands out due to its isothermal assay temperature and high sensitivity. LAMP-based assays have found extensive application in the identification of pathogens like malaria<sup>1</sup> and salmonella,<sup>2</sup> rapid detection of viral RNA such as human immunodeficiency virus (HIV) using reverse transcriptase LAMP (RT-LAMP),<sup>3</sup> detection of genetically modified crops contamination,<sup>4</sup> and forensic science for the precise detection of human DNA.<sup>5</sup> However, false positive results have been frequently documented.<sup>6-8</sup> LAMP assays mainly employ indirect methods for detecting amplification. Those methods include incorporating intercalating dyes for fluorescence or colorimetric determination, generating turbidity<sup>9</sup> through pyrophosphate precipitation, and utilizing quenched Calcein<sup>10</sup> or hydroxy-naphthol blue.<sup>11</sup> However, these methods are not sequence-specific and instead measure total amplification. Consequently, they will also generate signals for non-specific amplification, a common issue with LAMP. Therefore, a significant limitation of these approaches is the inability to distinguish signals between true positives and false positives, leading to inaccurate interpretations.

Several strategies incorporating fluorophore-labeled probes or primers have been suggested to enhance the specificity of LAMP and broaden the way for multiplexed detection. The detection of amplification by the release of quenching (DARQ) is one of the probe-based approaches, which attaches a quencher to the 5' end of a FIP primer to inhibit a fluorescent probe that is then released upon amplification and provides enhanced specificity.<sup>12</sup> One-step strand displacement (OSD) LAMP amplification is also developed by employing a fluorescent probe that binds with loop sequences and displaces a pre-hybridized quencher strand during amplification.<sup>13</sup> The quenching of unincorporated amplification signal reporter (QUASR) is another assay that uses a quencher probe to hybridize with fluorophore-labeled primers that remain unincorporated in amplicons after the reaction cools.<sup>14</sup> While multiplexed detection using these probes labeled with fluorophores is possible, specific issues arise. Techniques like DARQ and OSD relying upon strand displacement of a probe can potentially inhibit the LAMP reaction.<sup>12,15,16</sup> In the QUASR method, choosing the proper probe sequence is essential since the probes are supposed to remain unbound and not disrupt the reaction. It's crucial that the melting temperature of the quenching probe, when hybridized to the labeled primer, is substantially lower than the reaction temperature to avoid inhibiting the reaction.

In this work, we developed a nanopore-based approach for probe-free LAMP readouts to differentiate true positives and false positives. Compared with the fluorophore-labeled probes-based LAMP readout methods that could lead to reaction inhibition and probe-induced non-specific amplification, this approach offers endpoint detection that does not interfere with assay efficiency. We first developed a LAMP kinetics model for the specific amplification pathway and verified distinct patterns between true and false positives via gel electrophoresis. Minor model parameter adjustments led to varied amplicon distributions, underscoring the distinct patterns between true and false positives arising from the variations in pathways and efficiencies. Observing this, we then utilized nanopore-sized counting to acquire the amplicon information. Calibrated ECD values and normalized event frequency with internal calibrators ensured consistent and nanopore size-independent comparisons. Integrating nanopore-sized counting with machine learning yielded a 91.67% accuracy, thereby improving the reliability of LAMP in the detection of nucleic acids.

## Results and Discussion

### Commonly Observed False Positives in LAMP Assays

LAMP assays are emerging as an outstanding method for nucleic acid testing due to the rapid turnaround and simplified isothermal operational conditions. Despite its evident benefits, LAMP assays are challenged by the recurring issue of false positives. Although acknowledged within the scientific realm, these false positives' underlying kinetics in LAMP assays remain insufficiently explored. To assess the incidence of these inaccuracies, we conducted LAMP assays targeting malaria parasites,<sup>17</sup> and expanded them to RT-LAMP assays tailored for detecting SARS-CoV-2<sup>18</sup> and RSV<sup>19</sup> viruses. These pathogens were selected based on their epidemiological significance and the diagnostic outcomes in associated public health frameworks. In those three LAMP and RT-LAMP assays, SYTO-9 was used for fluorescence signals since intercalating dye is one of the most common indirect methods for LAMP readout. The real-time fluorescence curves in **Figure 1** showed that all three LAMP and RT-LAMP assays have false positives that occurred stochastically. The analysis of the cumulative results in **Figure 1d** underscored that each assay exhibited an appreciable false positive rate. Quantitatively, the false positive incidences for malaria, SARS-CoV-2, and RSV assays were 16%, 20%, and 28%, respectively. Additionally, the time-to-threshold ( $T_t$ ) varied randomly between 20 and 60 minutes with a standard deviation of up

to 18 minutes, suggesting the stochastic nature of false positives and complicating their differentiation from true positives. The results highlight that false positives are a consistent problem within the LAMP method rather than isolated instances. Based on the evidence, it suggests that LAMP assays have a notable tendency to produce false positives. This issue could directly impact the clinical outcomes and public health measures. As such, it is imperative to profoundly investigate the underlying mechanics of the LAMP procedure, identify the causes of these inaccuracies, and develop refined techniques to interpret LAMP readout with enhanced accuracy.

The false positives in LAMP assay can occur for various reasons. The primary sources of false positives are cross-contamination during sample preparation or amplification, and non-specific amplification induced by LAMP primers. Cross-contamination in amplification assays is often caused by residual DNA from previous LAMP reactions or airborne sources. To minimize this, strict experimental protocols are crucial. Additionally, incorporating dUTP during initial amplification and treating subsequent reactions with uracil DNA glycosylase (UDG) followed by its thermal inactivation can further prevent carry-over contamination.<sup>20-23</sup> In addition to cross-contamination, LAMP primers can bind to non-target sequences, leading to non-specific amplification and false positive results. Primer design and conditions like melting temperature, GC content, secondary structures, and concentration ratios could influence the reaction's specificity. It is essential to optimize the primer design and reaction conditions to minimize non-specific amplification. Primer-formed dimers or polymers could also trigger amplification due to specific properties of *Bst* enzymes, such as the ability for template switching or non-templated synthesis.<sup>24-27</sup> This primer-induced non-specific amplification is also observed in our SARS-CoV-2 RT-LAMP assay, as depicted in **Figure S1**. The non-specific amplification occurred at a delayed time compared with true positives, which means lower amplification efficiency than true positives. Overall, LAMP consistently encounters false positives across various assays, a phenomenon that is not yet fully understood. The results of LAMP assays for significant pathogens, observing false positive rates up to 28% and highly variable time-to-thresholds. The leading causes include cross-contamination and non-specific amplification, necessitating rigorous protocols and designs. These findings stress the need for a deeper understanding of LAMP mechanics and more accurate result interpretation to enhance diagnostic precision.

### **Exploration of LAMP Kinetics: Modeling and Gel Verification of Amplicon Profiles**

The LAMP assay, introduced by Notomi *et al.*, has evolved into an essential method for nucleic acid amplification.<sup>28</sup> This technique relies on a set of four core primers: two inner (Forward Inner Primer (FIP) and Backward Inner Primer (BIP)) and two outer (F3 and B3) that target six unique DNA regions (F1c, F2c, F3c, B1, B2, and B3). Operating at isothermal temperatures between 60-65 °C, the DNA target undergoes spontaneous breathing, creating single-stranded pockets for primer annealing. The process initiates with the FIP primer binding to its complementary F2c sequence, followed by strand displacement and extension. This process ultimately leads to the formation of a tremendous number of cauliflower-like DNA amplicons. An enhancement to the four primers system was proposed by Nagamine *et al.*, introducing a pair of loop primers to accelerate the reaction kinetics and amplify yield.<sup>29</sup> The primers, specifically designed to target the DNA sequence, ensure a well-characterized pathway for the true positive amplification.<sup>29</sup>

**Figure 2a** illustrates a simplified LAMP amplification pathway for true positives. The mother amplicon undergoes periodic self-replication at intervals  $t_R$ , facilitated by the inner and outer primers. The time required for the mother amplicon to complete the extension of one unit spacing ( $L$ , the region between F3 and B3) is proportional to its length and can be expressed as  $(2L_i+1)*t_E$ . In parallel, strand displacement generates a daughter amplicon during the extension of the mother strand, a process timestamped at  $(3L_i+1)*t_E$ . Leveraging these parameters, we have constructed a simplified kinetic model to elucidate amplicon size and population evolution from this distinct amplification pathway in true positives. The details of this model are provided in the supplementary materials. By adjusting the parameters of  $t_R$  and  $L$ , the model's simulation reveals distinct amplicon profiles. In **Figure 2b**, the replication time ( $t_R$ ) is fixed at 2 seconds for simplicity, and the unit spacing ( $L$ ) ranges from 100 bp to 200 bp (A typical length range when designing primers). Since longer spacings require more time to extend, the amplicon counts for each length exhibit an exponential distribution, with even minor changes of 25 bp leading to noticeable differences in amplicon length distribution and population variances exceeding two orders of magnitude. Additionally, **Figure 2c** demonstrates that by keeping  $L$  constant at 100 bp and adjusting  $t_R$  between 2 to 6 seconds, there are evident variations in the relative population of amplicons of different lengths. Specifically, a longer  $t_R$  leads to a slower replication rate of the initial mother amplicon (200 bp), resulting in a reduced population. Those simulation results from the LAMP kinetics model suggest that minor model parameter adjustments led to distinct amplicon

profiles. Given that primer-induced non-specific amplification depends on chance and occurs randomly in actual LAMP assays, the pathways and efficiencies of true and false positives are expected to differ significantly.<sup>24-27</sup> This distinction can lead to pronounced variations in replication time and unit spacing. Consequently, such variability is likely to manifest as notable differences in the distribution of amplicon lengths and significant variances in their populations.

Gel electrophoresis was employed to verify the different profiles between amplicons from true and false positive results, as depicted in **Figure 2d**. Notably, consistent patterns are observed in the true positives, while the false positives display varied band positions and intensities. This divergence in gel patterns suggests that false positives could arise from different amplification routes, yielding amplicons of differing lengths and concentrations compared to true positives. Existing studies suggest that primer-induced non-specific amplifications tend to be random, often resulting in reduced efficiency compared to true positive target-driven amplifications.<sup>8</sup> Given the probabilistic nature of false positives and their variable amplification pathways, our analysis will focus on the true positive amplicon patterns. Any deviations from this pattern can be classified as false positives. For quantitative analysis of amplicon distribution, we employed Image J to measure band intensities from the gel images. As band intensity reflects the overall nucleic acid mass, we derived the relative count of amplicons in each band by dividing the band intensities by the corresponding amplicon lengths. Subsequently, the relative count is normalized by the total counts in each lane to facilitate direct comparison across different gel lanes. Additionally, the amplicon lengths, estimated based on primer positions, may appear to correspond with longer ladder bands in gel images. This is attributed to the reduced mobility caused by the loop structures present in the amplicons.<sup>30</sup> **Figure 2e** summarizes the normalized distribution of the amplicons, revealing distinct patterns between true and false positives. True positives display a consistent exponential distribution pattern, whereas false positives exhibit varying length and population profiles. Despite gel images revealing discernible patterns between true and false positives, their time-consuming nature and limited sensitivity prompt the need for the exploration of alternative techniques.

### **Nanopore Sized Counting with Internal Calibrators for Amplicon Analysis**

The concept of nanopore sized counting relies on the ability of nanopores to detect and analyze individual DNA molecules as they translocate through the pore.<sup>17,18,31</sup> This translocation produces a characteristic electrical signal, ECD, representing the net excluded charges caused by an ionic current blockade event.<sup>32,33</sup> Studies have shown that, despite DNA conformational variations, ECD

values remain relatively stable and exhibit an almost linear relationship with DNA length when measured by the same nanopore device.<sup>34</sup> Therefore, by analyzing these ECD signals, the size distribution of the analyzed molecules can be deduced, effectively functioning as an electronic form of gel electrophoresis and serving as a powerful tool for assessing nucleic acid lengths.<sup>31</sup> Yet, achieving consistent nanopore dimensions across distinct batches presents challenges, as they are susceptible to variations due to ambient temperature, humidity, and characteristics inherent to the pipette puller. Previous studies indicate that this inconsistency in nanopore dimensions can directly influence the ECD values of molecules during translocation.<sup>35</sup>

To investigate the effect of nanopore size on molecular translocation and ECD values, we analyzed a DNA mixture comprising fragments of three lengths. We combined 0.5 nM each of 2 kbp, 10 kbp, and 20 kbp dsDNA, using the 2 kbp and 20 kbp fragments as calibrators to confirm the linear relationship between DNA length and ECD values. The 10 kbp fragments served as a proxy target for LAMP amplicons, which will be replaced by genuine LAMP amplicons when analyzing actual LAMP samples. The mixture was tested using four glass nanopore devices of varied sizes, all produced under identical pipette puller parameters. **Figure 3a** depicts the current traces from four nanopore devices, indicating the varying baseline currents at a consistent voltage bias of 400 mV. The ECD for translocation events was derived and summarized in **Figure 3b**. The histograms (with a bin size of 10 fC) across the four devices show a clear trend of decreasing ECD values with increasing pore size, likely due to the reduced interaction between the DNA fragments and the interior surface of larger pores.<sup>35</sup> Furthermore, the frequency of molecule translocation tends to increase with larger pore sizes. These variations in ECD readings and event frequencies present challenges for directly comparing ECD distributions and molecular populations across samples tested on diverse nanopore devices.

Nonetheless, internal dsDNA calibrators (2 kbp and 20 kbp dsDNA) with predefined lengths and concentrations can provide stable reference standards, enabling the calibration of ECD values and event frequencies. **Figure 3c** compiles comprehensive details of the nanopore sized counting, such as the size of the nanopores (estimated with nanopore conductance<sup>36,37</sup>), the mean ECD values for these calibrators, and the DNA event frequency at the mean ECD. The frequency at the mean ECD, rather than aggregating frequencies within three standard deviations, was chosen to represent the corresponding DNA frequency to simplify the analysis and aims to minimize the noise potentially caused by the overlap of adjacent peaks. With the gathered information, we could then

calibrate the ECD values and normalize the frequencies. For ECD calibration, we applied the mean ECD values of the 2 kbp and 20 kbp fragments to establish a linear relationship ( $ECD = \alpha_i * \text{DNA length}$ ), as shown in **Figure 3d**. Additionally, the event frequencies and their ratio between the 10 kbp target and the 20 kbp calibrator remained consistent across different pore devices, maintaining a steady 1:1 ratio due to their identical molecular concentrations ( $f_{10k} / f_{20k} = C_{10k} / C_{20k}$ ), as illustrated in **Figure 3e**. This consistency enables us to normalize the frequencies in the histograms using the frequency of the 20 kbp calibrator. Normalizing the frequency using the internal calibrator mitigates the effects of nanopore variation and reduces uncertainty in determining target concentration via its event frequency.<sup>38</sup> With those derived correlations, we converted ECD values to DNA length based on  $\alpha$  values and normalized the original frequency with 20 kbp calibrator's frequency. The calibrated DNA distribution (with a bin size of 1 kbp), as shown in **Figure 3f**, exhibits a consistent pattern across various nanopore devices, underscoring the efficacy and reliability of the calibration process. Incorporating these calibrators enables a standardized comparison of the amplicon histograms, rendering the analysis independent of nanopore size. This approach could ensure that any observed differences in amplicon profiles among devices reflect intrinsic variations in the sample molecules rather than inconsistencies in nanopore dimensions.

Before performing sized counting of the Malaria LAMP amplicons, we analyzed a calibrators-only solution (0.5 nM each of 2 kbp and 20 kbp dsDNA) to establish the value of  $\alpha$  and 20 kbp calibrator's frequency for ECD calibration and frequency normalization, as depicted in **Figure 4a**. Subsequently, a mixture of the Malaria LAMP amplicons and calibrators was analyzed with the same nanopore device (**Figure 4b**). The calibrators within the mixture provide a consistent internal reference for sized counting experiments, with a uniform calibrator profile (20 kbp calibrator's ECD and  $f_{20k}$ ) to verify the nanopore's steady functionality did not change over time.<sup>38</sup> Then, subtracting the frequency values of each column in **Figure 4b** from the corresponding values in **Figure 4a** allowed us to establish the amplicon's profile, removing the background signal associated with the calibrators (**Figure 4c**). Notably, the distribution of the LAMP true positive amplicons exhibited a similar exponential pattern observed in the gel image. Employing this calibrated method, we analyzed 24 true positive and 24 false positive samples with multiple nanopore devices, as shown in **Figure 4d&e**. Generally, the true positives exhibited a higher event frequency and a tendency to produce longer-length amplicons. Yet, the differences were subtle, complicating the interpretation of samples by visual inspection alone. Therefore, there's a need for



an improved method to increase the reliability of interpreting LAMP assay results from nanopore sized counting.

### **Machine Learning-Assisted Classification of LAMP Nanopore Readout**

The nanopore sized counting with internal calibrators provides a detailed amplicon distribution, which could potentially help differentiate between true positives and false positives in LAMP assays. We've observed that true and false positives from LAMP assays exhibit distinct efficiencies, resulting in diverse patterns in gel images and nanopore readouts. With this insight, we aim to develop a machine learning model for classifying these nanopore readouts, thus enhancing the reliability of the outcomes. The workflow of the machine learning-assisted classification of LAMP results is shown in **Figure 5a**. The procedure starts with calibrators-assisted nanopore sized counting of samples that exhibit amplified fluorescence signals, from which we obtain the calibrated and normalized histogram of DNA amplicon distribution. Subsequently, the histogram analyzes critical features such as mean, standard deviation (SD), skewness (Skew), kurtosis (Kurt), peak number ( $\#_{\text{Peak}}$ ), and peak relative intensities ( $\text{RI}_{\text{Peak}}$ , ratio of mean to maximum peak intensity). Then, a feature matrix is constructed by aggregating features from the histogram data of both true and false positive samples, after which the model is trained to classify input DNA amplicon histogram profiles.

Our model was trained with several classifiers: decision tree, logistic regression, random forest, naive Bayes, and linear support vector machine (SVM). Further details on the model are available in the methods section and supporting notes. The linear SVM emerged with a high accuracy of 91.67%, as shown in **Figure 5b**. While the naive Bayes classifier achieved the same level of accuracy, it is typically used in text classification and operates under the assumption that all features are independent. This assumption may not be appropriate for our analysis of LAMP amplicon profiles, where features like mean and peak numbers could be interrelated. Therefore, we choose the linear SVM classifier for the LAMP readout for the following reasons. First, the SVM is recognized for its effectiveness with small sample sizes, rendering it well-suited for LAMP classification without requiring numerous samples.<sup>39</sup> On the other hand, the linear SVM model is selected based on the expectation that features would display linear separability, reflecting the differing amplicon profiles stemming from the varied amplification efficiencies of true and false positives. This variation leads us to expect more abundant and lengthier amplicons in true positive samples compared to false positives in the nanopore counting results. We constructed a parallel

coordinate plot to demonstrate this linear separability of features (**Figure 5c**). In this plot, features of mean, standard deviation, peak number, and peak relative intensities exhibit relatively distinct two populations between true positives and false positives, reinforcing the choice of the linear SVM model. Besides, the model's learning curve (**Figure 5d**) indicates improved accuracy with increasing sample size and superior performance when utilizing six features. Preliminary analysis using 48 samples yielded a promising 91.67% accuracy. The model's performance was further validated using a confusion matrix, as depicted in **Figure 5e**. At a probability threshold of 50%, the model could achieve a 91.67% sensitivity and 91.67% specificity. Taking into account that the intercalating dye-based readout method has a false positive rate of 16% for the Malaria LAMP assay, the machine learning-assisted nanopore readout method could achieve a substantially lower false positive rate of 1.33% (calculated as 16% multiplied by the complement of 91.67% specificity), effectively reducing the false positive rate by a factor of 10.

This approach's systematic methodology offers a reliable way to classify true and false positives based on the described features. It is worth noting that this machine learning-assisted nanopore sized counting method can also be utilized as a direct endpoint test for assays, eliminating the need for intercalating dyes or fluorophore-labeled probes. Our previous study showed that positive and negative results could be effectively distinguished based on the event frequency, given that negligible amplicons are produced in negative cases.<sup>17,18</sup> Yet, the presented nanopore sized counting method cannot differentiate true negatives from false negatives, as neither condition generates a significant quantity of amplicons detectable by the nanopore. Despite this limitation, our advancements in distinguishing between true and false positives advance the potential for more accurate, rapid, and probe-free interpretations of LAMP outcomes, reducing the dependency on human discernment. This method has the potential for integration into clinical diagnostics. Initially, raw samples such as saliva and plasma are collected from patients to extract pathogen nucleic acids. These samples are then subjected to the LAMP reaction. Following this, the probe-free nanopore sizing and counting method is employed to analyze the calibrated amplicon profiles. Utilizing a pre-built database, an accurate readout of the samples could be achieved through machine learning-assisted classification.

## Conclusions

In this study, we presented a nanopore-based approach for probe-free LAMP readouts,

leveraging amplicon sizing and counting, which serves a role similar to an electronic adaptation of gel electrophoresis. Given the observed false positive rates of 16% to 28% in our LAMP and RT-LAMP assays, we explored the LAMP reaction kinetics by developing a model for the specific amplification pathway of true positives. Minor model parameter adjustments led to varied amplicon lengths and population variances exceeding two orders of magnitude between true and false positives. These distinct patterns highlight the potential difference in amplicon profiles between false and true positives, attributed to variations in replication time and unit spacing, a finding confirmed by the gel electrophoresis and nanopore sized counting. Although nanopore size variability during fabrication could alter ECD values and event frequencies, implementing internal dsDNA calibrators enables consistent calibration, rendering the measurements independent of nanopore size. The DNA amplicon profiles could be acquired with calibrator-assisted nanopore sized counting, and then features for machine learning classification could be extracted. Our approach achieved a 91.67% accuracy in identifying true versus false positives, substantially refining the accuracy of the LAMP assay readout for more reliable disease diagnostics. As an endpoint detection readout, the probe-free nanopore sizing and counting method eliminates the risk of LAMP reaction inhibition or non-specific amplification associated with fluorophore-labeled probes. Given its adaptability and demonstrated precision, this method holds potential promise for broad application across various LAMP assays.

## **Methods and Experimental Section**

### **Materials and Chemicals**

Quartz capillaries (QF100-50-7.5) with inner and outer diameters of 0.5 and 1 mm were used in our experiment (Sutter Instrument). The pipette holder (QSW-T10N) was purchased from Warner Instruments. Ag/AgCl electrodes were homemade with 0.2 mm Ag wires (Warner Instruments). The micro-injector (MF34G-5) with 34 gauge was purchased from World Precision Instruments. 2 kbp (SM1701), 10 kbp (SM1751), 20 kbp (SM1541) DNA fragments, and SYTO-9 green fluorescent nucleic acid stain (S34854) were purchased from Thermo Fisher Scientific. LiCl (L9650) and Tris-EDTA buffer solution (pH 8.0, 93283) were purchased from Sigma Aldrich. All solutions were filtered with a 0.2  $\mu$ m syringe filter (WHA67802502, Whatman). The malaria *Plasmodium vivax* (Pv.) genomic DNAs (5 ng/ $\mu$ l) were gifts from Dr. Cui's lab at Penn State, extracted by the phenol-chloroform-based procedure. *Bst* 2.0 DNA polymerase (M0537),

WarmStart RTx reverse transcriptase (M0380), nuclease-free water (B1500), isothermal amplification buffer (B0537), deoxynucleotide solution mix (N0447), magnesium sulfate solution (B1003) were purchased from New England Biolabs. Heat-inactivated SARS-CoV-2 (VR-1986HK) RNA and quantitative genomic RNA from human respiratory syncytial virus strain A2 (VR-1540DQ) were purchased from ATCC.

### **Glass Nanopore Fabrication and Measurement**

The quartz capillaries were initially cleaned with piranha solution ( $\text{H}_2\text{SO}_4$ :  $\text{H}_2\text{O}_2$  at a 3:1 ratio) at 95 °C for 30 minutes to remove organic contaminants. They were then thoroughly rinsed with deionized water and dried in an oven at 100 °C for 15 minutes. A laser pipette puller (P-2000, Sutter Instruments) was used to fabricate the nanopore using a two-line program: (1) Heat 750, Filament 5, Velocity 50, Delay 140, and Pull 50; (2) Heat 715, Filament 4, Velocity 30, Delay 145, and Pull 225. This recipe typically produces nanopore size around 10 nm. After nanopore fabrication, it is loaded with 2 M LiCl using a micro-injector. The calibrators-only solution comprises 0.5 nM of 2 kbp and 0.5 nM of 20 kbp dsDNA fragments in 2M LiCl. For the mixed sample, the same concentrations of calibrators are combined with 500× diluted LAMP amplicons in 2M LiCl. This dilution aims to minimize nanopore clogging, with specifics illustrated in **Figure S2** (The samples tested in this instance are Malaria LAMP true positives, which were analyzed without the addition of 2 kbp or 20 kbp dsDNA calibrators). If clogging was encountered, five cycles of IV sweeps ranging from -700 to 700 mV were conducted to clear and restore the nanopore. A consistent 400 mV voltage was applied to the nanopore using a 6363 DAQ card (National Instruments) during sized counting experiments. Ionic current recordings were captured by an Axopatch 200B amplifier (Molecular Device). These recordings were digitized using the same DAQ card, processed through a custom LabVIEW program at a sampling frequency of 100 kHz, and subjected to a 10 kHz low-pass filter. The measurement setup was housed within a custom-made Faraday cage to minimize environmental electrical noise. A MATLAB program was developed to reconstruct data and perform single-molecule event analysis, which encompasses calculating event rates, measuring current blockage amplitudes, determining event durations, and acquiring ECD values. Each sample underwent a 10-minute nanopore measurement during which more than 1000 events were captured for the subsequent ECD analysis. For the machine learning classification training of LAMP samples, calibrated amplicon profiles were obtained using a total

of 14 different nanopore devices.

## **Machine Learning Model Development**

Based on calibrated DNA amplicon histogram data, this study uses a classification model using linear support vector machines (SVM) to differentiate between true positive and false positive cases. The data was sourced from multiple Excel files within two distinct directories representing the two classes. Essential statistical features were extracted from the frequency column of the data, including mean, standard deviation, skewness, kurtosis, peak number, and peak relative intensities. The model was subjected to the stringent 'Leave-One-Out' cross-validation. The probability threshold was set at the conventional mark of 50% for assessing accuracy and analyzing the confusion matrix. The complete code for the classification model is available in the supplementary materials.

## **Author Information**

### **Corresponding Author**

Weihua Guan

\* Email: wzg111@psu.edu

### **ORCID**

Weihua Guan: 0000-0002-8435-9672

### **Author Contributions**

W.G. conceived the concept and supervised the study. D.M. performed nanopore and LAMP experiments. A. K. and A. P. helped develop the LAMP and RT-LAMP assays. W.G. and D.M. co-wrote the manuscript, with discussion from all authors.

### **Notes**

The authors declare no competing financial interest.

## **Acknowledgments**

This work was partially supported by the National Institutes of Health (R33AI147419) and the National Science Foundation (1902503, 1912410, 2045169, 2319913). Any opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the National Science Foundation and National Institutes of Health.

## **Associated Content**

The Supporting Information is available: The analysis of primer-induced false positives in LAMP assay, the real-time fluorescence curves for Malaria *Pv*. LAMP assay, examines the effect of amplicon dilution on nanopore's time-to-clog amplicon dilution, assays and primer sets information, details of the kinetic model depicting LAMP's true positive pathway, and details for machine learning-assisted classification.

## Reference

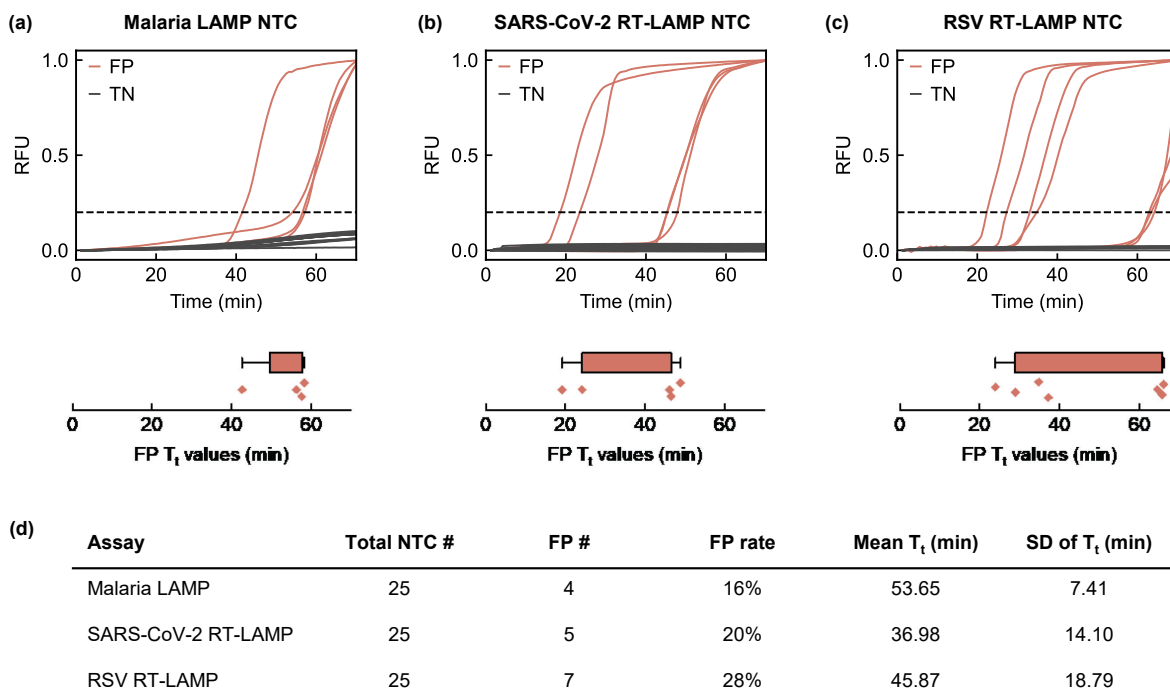
- (1) Poon, L. L.; Wong, B. W.; Ma, E. H.; Chan, K. H.; Chow, L. M.; Abeyewickreme, W.; Tangpukdee, N.; Yuen, K. Y.; Guan, Y.; Looareesuwan, S. Sensitive and Inexpensive Molecular Test for Falciparum Malaria: Detecting Plasmodium Falciparum DNA Directly from Heat-Treated Blood by Loop-Mediated Isothermal Amplification. *Clin. Chem.* **2006**, *52*, 303–306.
- (2) Ohtsuka, K.; Yanagawa, K.; Takatori, K.; Hara-Kudo, Y. Detection of Salmonella Enterica in Naturally Contaminated Liquid Eggs by Loop-Mediated Isothermal Amplification, and Characterization of Salmonella Isolates. *Appl. Environ. Microbiol.* **2005**, *71*, 6730–6735.
- (3) Curtis, K. A.; Rudolph, D. L.; Owen, S. M. Rapid Detection of HIV-1 by Reverse-Transcription, Loop-Mediated Isothermal Amplification (RT-LAMP). *J. Virol. Methods* **2008**, *151*, 264–270.
- (4) Kiddle, G.; Hardinge, P.; Buttigieg, N.; Gandelman, O.; Pereira, C.; McElgunn, C. J.; Rizzoli, M.; Jackson, R.; Appleton, N.; Moore, C. GMO Detection Using a Bioluminescent Real Time Reporter (BART) of Loop Mediated Isothermal Amplification (LAMP) Suitable for Field Use. *BMC Biotechnol.* **2012**, *12*, 1–13.
- (5) Watthanapanpituck, K.; Kiatpathomchai, W.; Chu, E.; Panvisavas, N. Identification of Human DNA in Forensic Evidence by Loop-Mediated Isothermal Amplification Combined with a Colorimetric Gold Nanoparticle Hybridization Probe. *Int. J. Legal Med.* **2014**, *128*, 923–931.
- (6) Moehling, T. J.; Choi, G.; Dugan, L. C.; Salit, M.; Meagher, R. J. LAMP Diagnostics at the Point-of-Care: Emerging Trends and Perspectives for the Developer Community. *Expert Rev. Mol. Diagn.* **2021**, *21*, 43–61.
- (7) Hardinge, P.; Murray, J. A. H. Reduced False Positives and Improved Reporting of Loop-Mediated Isothermal Amplification Using Quenched Fluorescent Primers. *Sci. Rep.* **2019**, *9*, 7400.
- (8) Meagher, R. J.; Priye, A.; Light, Y. K.; Huang, C.; Wang, E. Impact of Primer Dimers and Self-Amplifying Hairpins on Reverse Transcription Loop-Mediated Isothermal Amplification Detection of Viral RNA. *Analyst* **2018**, *143*, 1924–1933.
- (9) Mori, Y.; Nagamine, K.; Tomita, N.; Notomi, T. Detection of Loop-Mediated Isothermal Amplification Reaction by Turbidity Derived from Magnesium Pyrophosphate Formation. *Biochem. Biophys. Res. Commun.* **2001**, *289*, 150–154.
- (10) Tomita, N.; Mori, Y.; Kanda, H.; Notomi, T. Loop-Mediated Isothermal Amplification (LAMP) of Gene Sequences and Simple Visual Detection of Products. *Nat. Protoc.* **2008**, *3*, 877–882.
- (11) Goto, M.; Honda, E.; Ogura, A.; Nomoto, A.; Hanaki, K.-I. Colorimetric Detection of Loop-Mediated Isothermal Amplification Reaction by Using Hydroxy Naphthol Blue. *Biotechniques* **2009**, *46*, 167–172.
- (12) Tanner, N. A.; Zhang, Y.; Evans Jr, T. C. Simultaneous Multiple Target Detection in Real-Time Loop-Mediated Isothermal Amplification. *Biotechniques* **2012**, *53*, 81–89.
- (13) Jiang, Y. S.; Bhadra, S.; Li, B.; Wu, Y. R.; Milligan, J. N.; Ellington, A. D. Robust Strand Exchange Reactions for the Sequence-Specific, Real-Time Detection of Nucleic Acid Amplicons. *Anal. Chem.* **2015**, *87*, 3314–3320.
- (14) Ball, C. S.; Light, Y. K.; Koh, C.-Y.; Wheeler, S. S.; Coffey, L. L.; Meagher, R. J. Quenching of Unincorporated Amplification Signal Reporters in Reverse-Transcription Loop-Mediated Isothermal Amplification Enabling Bright, Single-Step, Closed-Tube, and Multiplexed

- Detection of RNA Viruses. *Anal. Chem.* **2016**, *88*, 3562–3568.
- (15) Tanner, N. A.; Evans Jr., T. C. Loop-Mediated Isothermal Amplification for Detection of Nucleic Acids. *Curr. Protoc. Mol. Biol.* **2014**, *105*, 15.14.1–15.14.14.
  - (16) Fan, Q.; Xie, Z.; Wei, Y.; Zhang, Y.; Xie, Z.; Xie, L.; Huang, J.; Zeng, T.; Wang, S.; Luo, S.; Li, M. Development of a Visual Multiplex Fluorescent LAMP Assay for the Detection of Foot-and-Mouth Disease, Vesicular Stomatitis and Bluetongue Viruses. *PLOS ONE* **2022**, *17*, e0278451.
  - (17) Tang, Z.; Choi, G.; Nouri, R.; Guan, W. Loop-Mediated Isothermal Amplification-Coupled Glass Nanopore Counting Toward Sensitive and Specific Nucleic Acid Testing. *Nano Lett.* **2019**, *19*, 7927–7934.
  - (18) Tang, Z.; Nouri, R.; Dong, M.; Yang, J.; Greene, W.; Zhu, Y.; Yon, M.; Nair, M. S.; Kuchipudi, S. V.; Guan, W. Rapid Detection of Novel Coronavirus SARS-CoV-2 by RT-LAMP Coupled Solid-State Nanopores. *Biosens. Bioelectron.* **2022**, *197*, 113759.
  - (19) Takayama, I.; Nakauchi, M.; Takahashi, H.; Oba, K.; Semba, S.; Kaida, A.; Kubo, H.; Saito, S.; Nagata, S.; Odagiri, T.; Kageyama, T. Development of Real-Time Fluorescent Reverse Transcription Loop-Mediated Isothermal Amplification Assay with Quenching Primer for Influenza Virus and Respiratory Syncytial Virus. *J. Virol. Methods* **2019**, *267*, 53–58.
  - (20) Tang, Y.; Chen, H.; Diao, Y. Advanced Uracil DNA Glycosylase-Supplemented Real-Time Reverse Transcription Loop-Mediated Isothermal Amplification (UDG-rRT-LAMP) Method for Universal and Specific Detection of Tembusu Virus. *Sci. Rep.* **2016**, *6*, 27605.
  - (21) Manajit, O.; Longyant, S.; Sithigorngul, P.; Chaivisuthangkura, P. Development of Uracil-DNA-Glycosylase-Supplemented Loop-Mediated Isothermal Amplification Coupled with Nanogold Probe (UDG-LAMP-AuNP) for Specific Detection of *Pseudomonas Aeruginosa*. *Mol. Med. Rep.* **2018**, *17*, 5734–5743.
  - (22) Longo, M. C.; Berninger, M. S.; Hartley, J. L. Use of Uracil DNA Glycosylase to Control Carry-over Contamination in Polymerase Chain Reactions. *Gene* **1990**, *93*, 125–128.
  - (23) Zen, L. P. Y.; Lai, M. Y.; Lau, Y. L. Elimination of Contamination in Loop-Mediated Isothermal Amplification Assay for Detection of Human Malaria. *Trop. Biomed.* **2020**, *37*, 1124–1128.
  - (24) Zyrina, N. V.; Antipova, V. N.; Zheleznyaya, L. A. Ab Initio Synthesis by DNA Polymerases. *FEMS Microbiol. Lett.* **2014**, *351*, 1–6.
  - (25) García, P. B.; Robledo, N. L.; Islas, Á. L. Analysis of Non-Template-Directed Nucleotide Addition and Template Switching by DNA Polymerase. *Biochemistry* **2004**, *43*, 16515–16524.
  - (26) Ramadan, K.; Shevelev, I. V.; Maga, G.; Hübscher, U. De Novo DNA Synthesis by Human DNA Polymerase  $\lambda$ , DNA Polymerase  $\mu$  and Terminal Deoxyribonucleotidyl Transferase. *J. Mol. Biol.* **2004**, *339*, 395–404.
  - (27) Zyrina, N. V.; Zheleznyaya, L. A.; Dvoretzky, E. V.; Vasiliev, V. D.; Chernov, A.; Matvienko, N. I. N.BspD6I DNA Nickase Strongly Stimulates Template-Independent Synthesis of Non-Palindromic Repetitive DNA by Bst DNA Polymerase. *Biol. Chem.* **2007**, *388*, 367–372.
  - (28) Notomi, T.; Okayama, H.; Masubuchi, H.; Yonekawa, T.; Watanabe, K.; Amino, N.; Hase, T. Loop-Mediated Isothermal Amplification of DNA. *Nucleic Acids Res.* **2000**, *28*, E63.
  - (29) Nagamine, K.; Hase, T.; Notomi, T. Accelerated Reaction by Loop-Mediated Isothermal Amplification Using Loop Primers. *Mol. Cell. Probes* **2002**, *16*, 223–229.
  - (30) Biegeleisen, K. Methods for Non-Destructively Separating or Reannealing the Strands of Circular Duplex DNA Chromosomes. *Open Access Libr. J.* **2017**, *4*, 1–31.

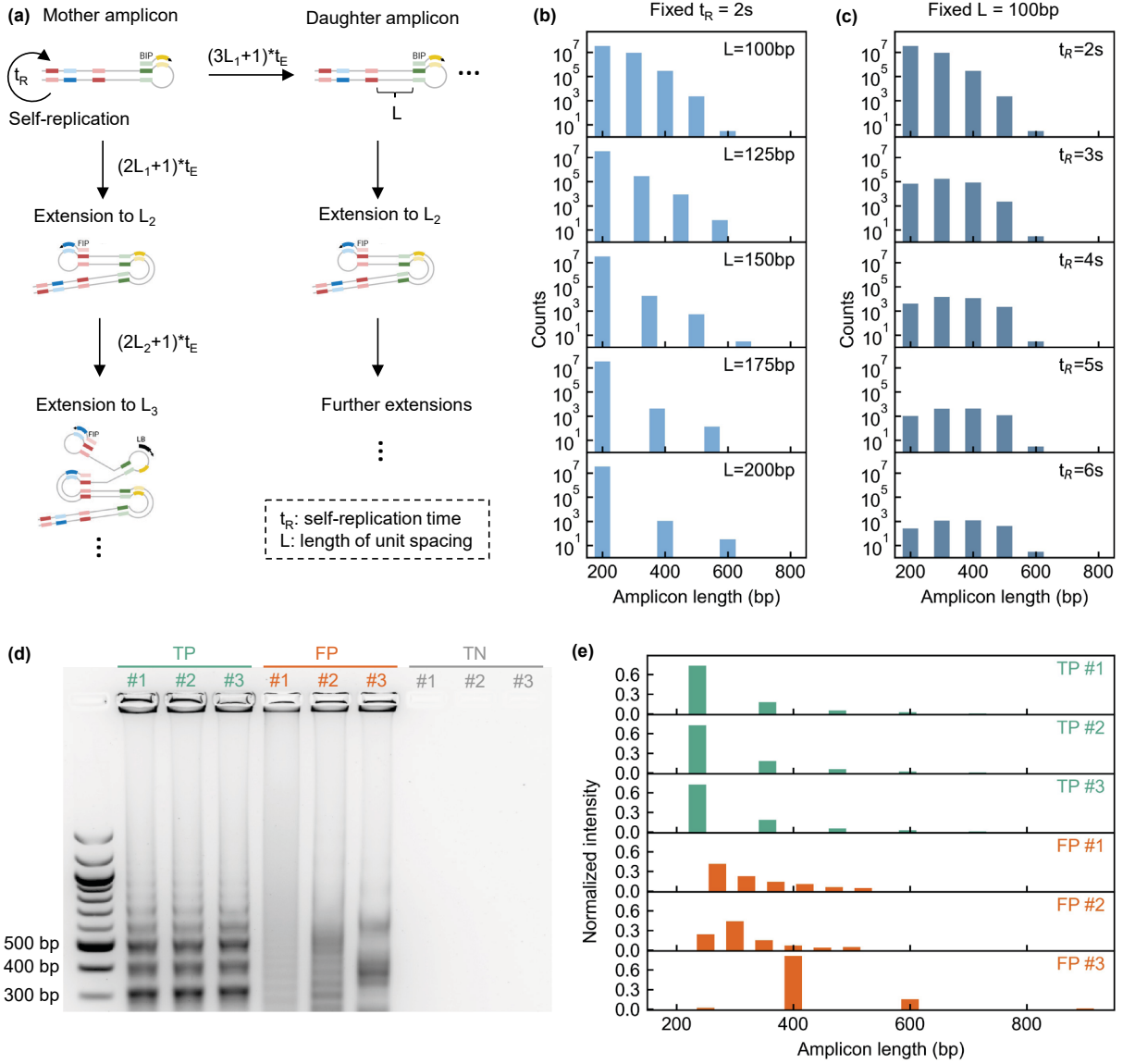


- (31) Bell, N. A. W.; Muthukumar, M.; Keyser, U. F. Translocation Frequency of Double-Stranded DNA through a Solid-State Nanopore. *Phys. Rev. E* **2016**, *93*, 022401.
- (32) Fologea, D.; Gershow, M.; Ledden, B.; McNabb, D. S.; Golovchenko, J. A.; Li, J. Detecting Single Stranded DNA with a Solid State Nanopore. *Nano Lett.* **2005**, *5*, 1905–1909.
- (33) Storm, A. J.; Chen, J. H.; Zandbergen, H. W.; Dekker, C. Translocation of Double-Strand DNA through a Silicon Oxide Nanopore. *Phys. Rev. E* **2005**, *71*, 051903.
- (34) Fologea, D.; Brandin, E.; Uplinger, J.; Branton, D.; Li, J. DNA Conformation and Base Number Simultaneously Determined in a Nanopore. *Electrophoresis* **2007**, *28*, 3186–3192.
- (35) Rivas, F.; DeAngelis, P. L.; Rahbar, E.; Hall, A. R. Optimizing the Sensitivity and Resolution of Hyaluronan Analysis with Solid-State Nanopores. *Sci. Rep.* **2022**, *12*, 4469.
- (36) Nouri, R.; Tang, Z.; Guan, W. Quantitative Analysis of Factors Affecting the Event Rate in Glass Nanopore Sensors. *ACS Sens.* **2019**, *4*, 3007–3013.
- (37) Steinbock, L. J.; Bulushev, R. D.; Krishnan, S.; Raillon, C.; Radenovic, A. DNA Translocation through Low-Noise Glass Nanopores. *ACS Nano* **2013**, *7*, 11255–11262.
- (38) Charron, M.; Briggs, K.; King, S.; Waugh, M.; Tabard-Cossa, V. Precise DNA Concentration Measurements with Nanopores by Controlled Counting. *Anal. Chem.* **2019**, *91*, 12228–12237.
- (39) Meng, M.; Zhao, C. Application of Support Vector Machines to a Small-Sample Prediction. *Adv. Pet. Explor. Dev.* **2015**, *10*, 72–75.

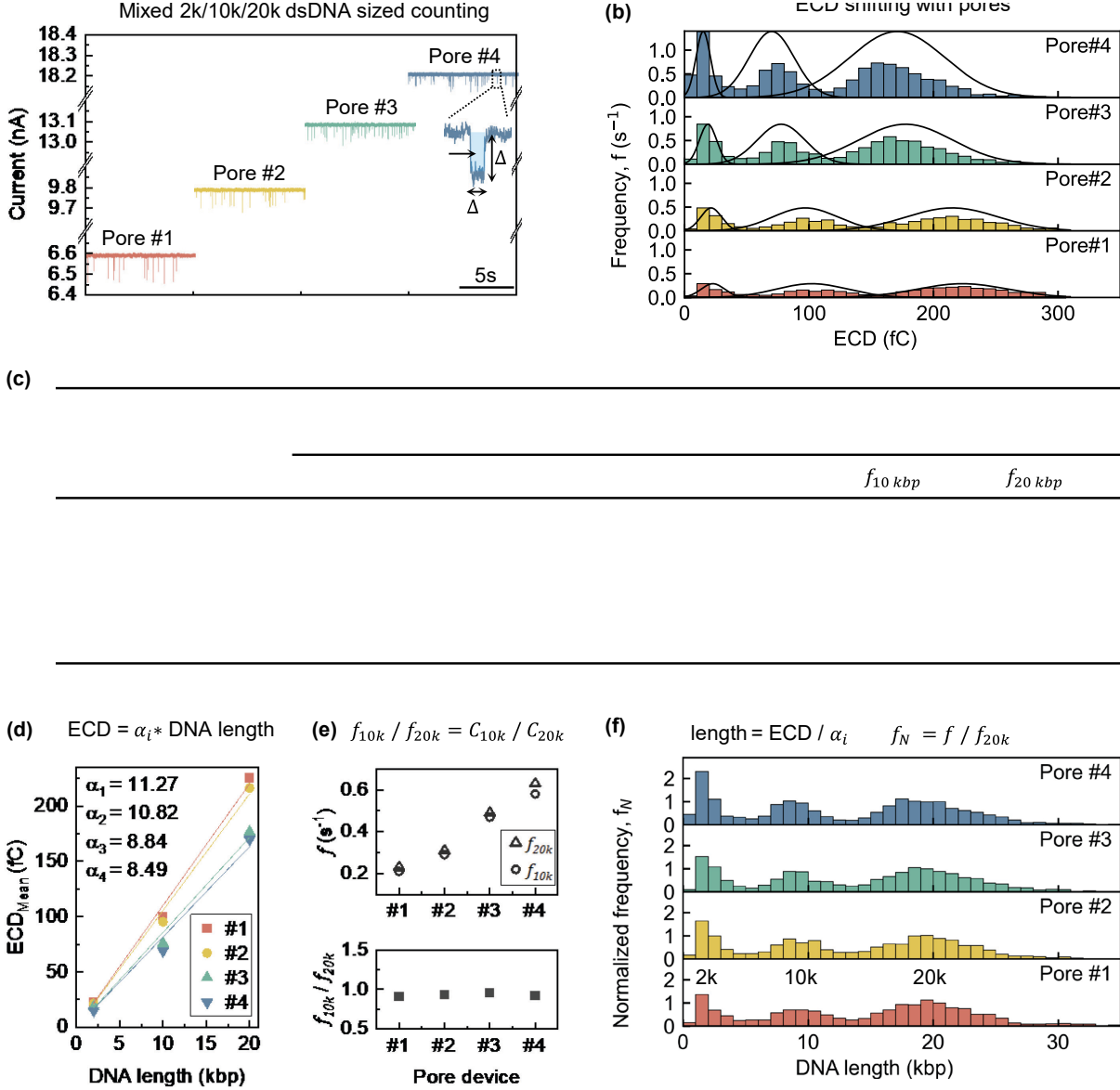
## Figures and Captions



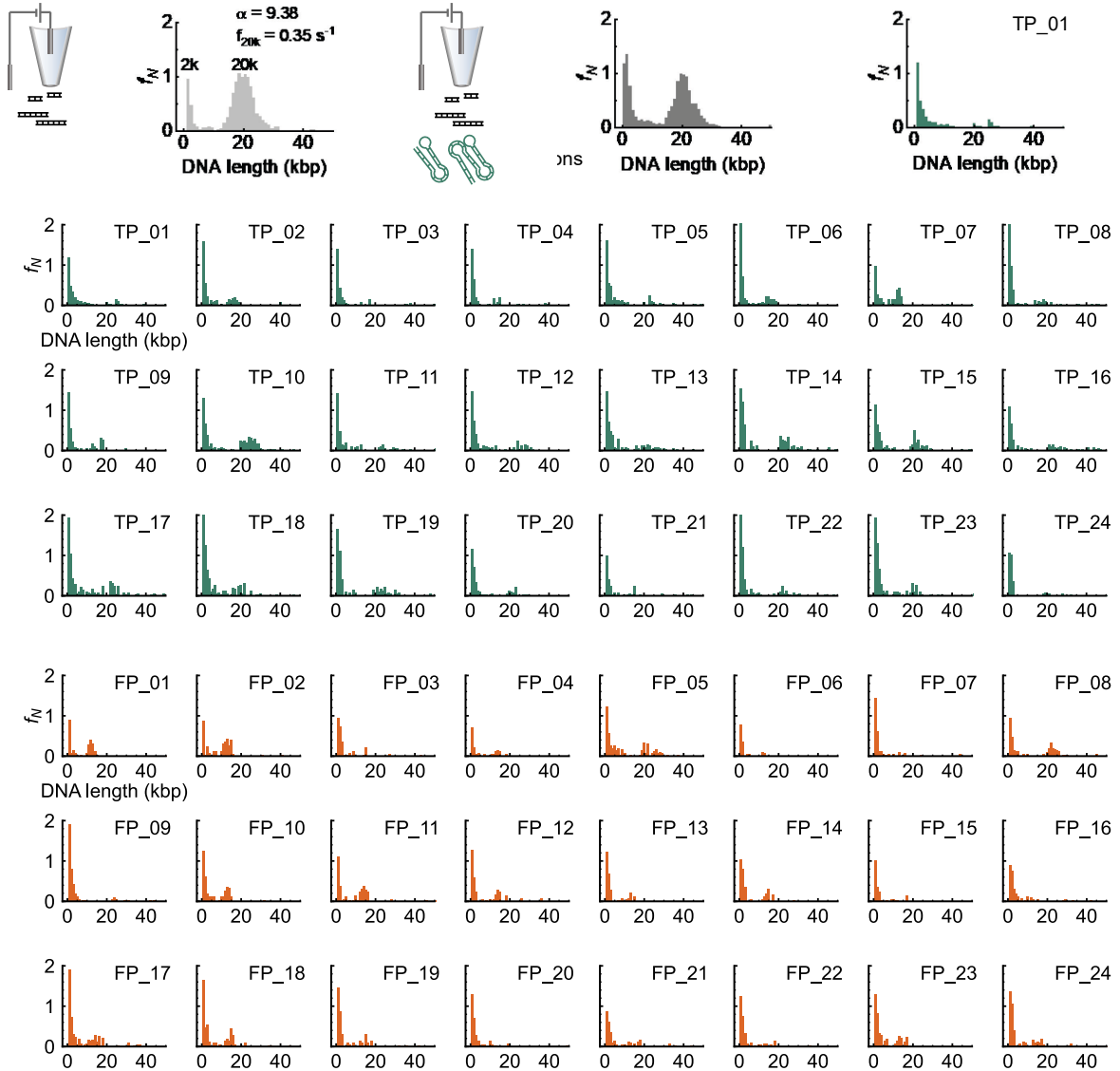
**Figure 1. Results of testing the no template controls (NTC) for three different LAMP assays.** (a) The real-time fluorescence curve of 25 no-template controls (NTC) in the Malaria LAMP assay. 4 controls displayed false positive (FP) results, with the threshold time ( $T_t$ , defined as the duration to reach a normalized RFU threshold value of 0.2) ranging from 42 to 58 minutes, as highlighted in the accompanying box plot. (b) The real-time fluorescence curve of 25 NTC in the SARS-CoV-2 RT-LAMP assay. 5 controls displayed false positive results, with the  $T_t$  ranging from 20 to 49 minutes. (c) The real-time fluorescence curve of 25 NTC in the RSV RT-LAMP assay. 7 controls displayed false positive results, with the  $T_t$  ranging from 24 to 66 minutes. (d) Statistical overview of LAMP assay outcomes from all NTC experiments.



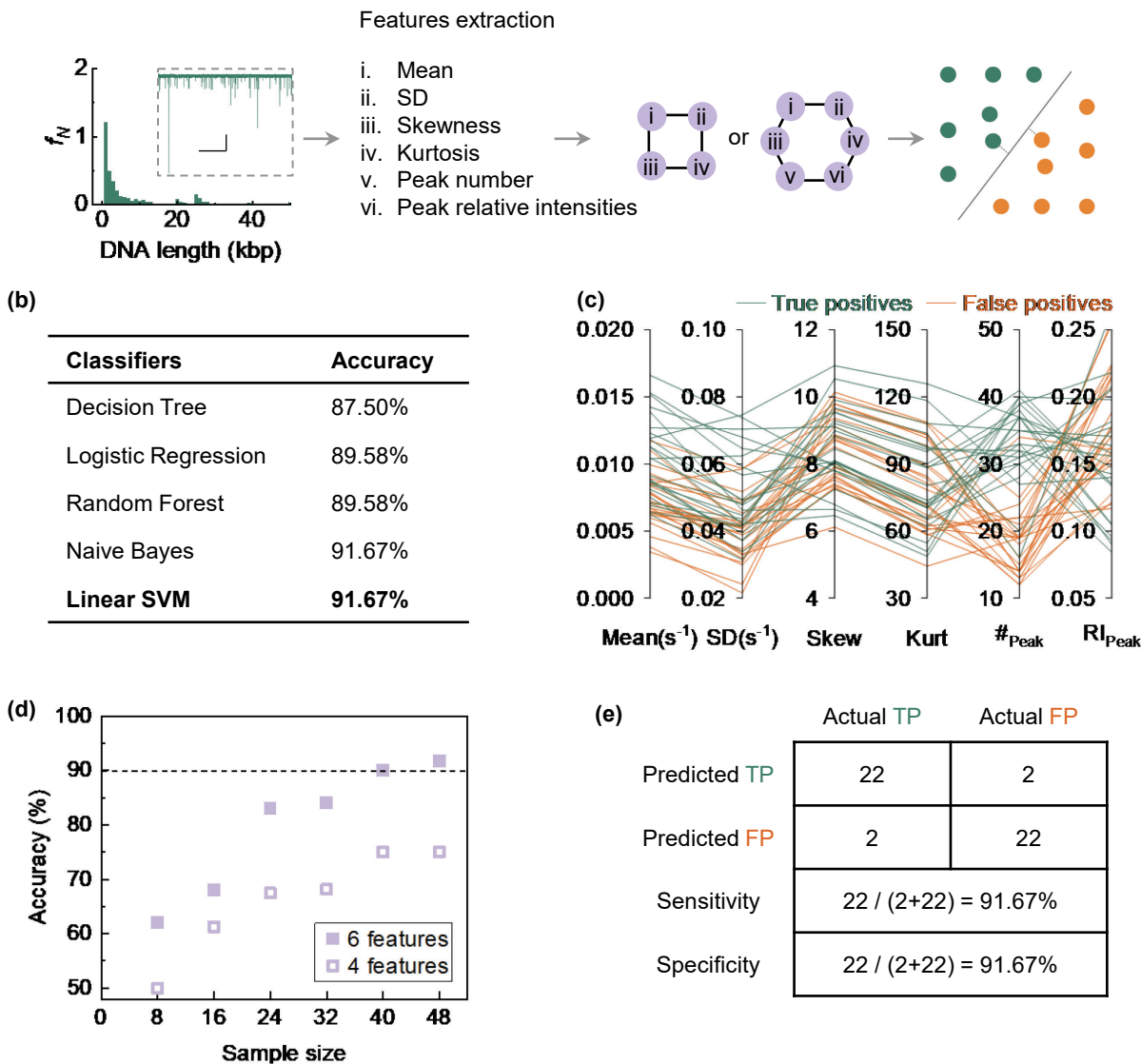
**Figure 2. The LAMP kinetic model demonstrates the factors influencing amplicon profiles, with gel validation revealing distinct patterns between true positives and false positives.** (a) Simplified illustrative diagram of the target DNA-induced specific amplification pathway.  $t_E$  represents the time required for polymerase to extend one unit spacing, established at 1 second per 50 bp extension.  $L_n$  denotes an amplicon containing  $n$  unit spacings. (b) Simulation results from the LAMP kinetic model with a constant  $t_R$  of 2 s and varying  $L$ , demonstrating notable variations in amplicon profiles. (c) Simulation results with a constant  $L$  of 100 bp and varying  $t_R$ . (d) Gel electrophoresis image of true positive (0.05 pg gDNA per reaction), false positive, and true negative samples from Malaria LAMP assay. (2% agarose gel run at 100 V for 80 mins) (e) Comparative analysis of amplicon profiles between true positive and false positive samples.



**Figure 3. Investigating the shifting of ECD values across nanopore devices and developing calibrators-assisted sized counting.** (a) Current traces for sized counting mixture of 2kbp, 10kbp, and 20kbp dsDNA fragments using nanopores of varying sizes.  $ECD = \int \Delta I * \Delta t$ . (b) ECD distribution of the events measured by different nanopores. (c) The table summarizes pore size information, mean ECD for 2kbp, 20kbp dsDNA calibrators, 10kbp dsDNA targets, and event frequencies. (d) Mean ECD correlation with DNA length. Solid lines are linear fittings for four pores, each with an  $R^2$  value of 0.99. (e) The frequency of 10kbp targets and 20kbp calibrators maintained a consistent 1:1 ratio across various pores, attributable to their equal concentrations ( $C_{10k} = C_{20k}$ ). (f) The distribution of DNA events after calibration. DNA lengths are converted based on  $\alpha$  values, and frequency is normalized by  $f_{20k}$ .



**Figure 4. Sized counting of the Malaria LAMP amplicons with internal calibrators.** (a) Nanopore sized counting of calibrators with DNA histogram calibration using  $\alpha$  and  $f_{20k}$ . (b) Nanopore sized counting of a mixture containing calibrators and amplicons, along with the calibrated DNA histogram. (c) The amplicon histogram with the background from calibrators was subtracted. (d) Representative histograms from 24 true positive samples. (e) Representative histograms from 24 false positive samples.



**Figure 5. Advancing LAMP assay result interpretation with machine learning.** (a) Workflow illustrating the integration of machine learning in interpreting LAMP assay results. (b) The comparative accuracy of the model using various classifiers. (c) The parallel coordinate plots derived from six features. (d) The learning curve depicts the improvements in model accuracy with increasing sample size and number of features. (e) The confusion matrix evaluates the model's performance.

## For Table of Contents Only

