Deep Domain Adaptation Enhances Amplification Curve Analysis for Single-Channel Multiplexing in Real-Time PCR

Ye Mao, *Graduate Student Member, IEEE*, Ke Xu, *Student Member, IEEE*, Luca Miglietta, Louis Kreitmann, Nicolas Moser, *Member, IEEE*, Pantelis Georgiou, *Senior Member, IEEE*, Alison Holmes, and Jesus Rodriguez-Manzano

***Abstract*—Data-driven approaches for molecular diagnostics are emerging as an alternative to perform an accurate and inexpensive multi-pathogen detection. A novel technique called Amplification Curve Analysis (ACA) has been recently developed by coupling machine learning and real-time Polymerase Chain Reaction (qPCR) to enable the simultaneous detection of multiple targets in a single reaction well. However, target classification purely relying on the amplification curve shapes faces several challenges, such as distribution discrepancies between different data sources (i.e., training vs testing). Optimisation of computational models is required to achieve higher performance of ACA classification in multiplex qPCR through the reduction of those discrepancies. Here, we proposed a novel transformer-based conditional domain adversarial network (T-CDAN) to eliminate data distribution differences between the source domain (synthetic DNA data) and the target domain (clinical isolate data). The labelled training data from the source domain and unlabelled testing data from the target domain are fed into the T-CDAN, which learns both domains' information simultaneously. After mapping the inputs into a domain-irrelevant space, T-CDAN removes the feature distribution differences and provides a clearer decision boundary for the classifier, resulting in a more accurate pathogen identification. Evaluation of 198 clinical isolates containing three types of carbapenem-resistant genes (*bla*$_{\mathrm{NDM}}$, *bla*$_{\mathrm{IMP}}$ and *bla*$_{\mathrm{OXA\text{-}48}}$) illustrates a curve-level accuracy of 93.1% and a sample-level accuracy of 97.0% using T-CDAN, showing an accuracy improvement of 20.9% and 4.9% respectively. This research emphasises the importance of deep domain adaptation to enable high-level multiplexing in a single qPCR reaction, providing a solid approach to extend qPCR instruments' capabilities in real-world clinical applications.**

***Index Terms*—Domain adaptation, transformer, ACA, amplificaion curve analysis, PCR, multiplex.**

Manuscript received 12 October 2022; revised 27 February 2023; accepted 10 March 2023. Date of publication 15 March 2023; date of current version 6 June 2023. This work was supported in part by the Imperial COVID-19 Research Fund under Grant WDAI.G28059, in part by the Department of Health and Social Care funded Centre for Antimicrobial Optimisation (CAMO) at Imperial College London. The work of Ke Xu was supported by the Imperial College President's PhD Scholarships 2021. The work of Louis Kreitmann was supported by the Scholarship from bioMerieux, S.A. *(Ye Mao and Ke Xu contributed equally to this work.) (Corresponding author: Jesus Rodriguez-Manzano.)*

This work did not involve human subjects or animals in its research.

Ye Mao and Nicolas Moser are with the Department of Electrical and Electronic Engineering, Faculty of Engineering, Imperial College London, SW7 2AZ London, U.K. (e-mail: ym399@cam.ac.uk; n.moser@imperial.ac.uk).

Ke Xu is with the Department of Infectious Disease, Faculty of Medicine, Imperial College, W12 0NN London, U.K. (e-mail: ke.xu21@imperial.ac.uk).

Luca Miglietta is with the Department of Infectious Disease, Faculty of Medicine, Imperial College, W12 0NN London, U.K., and also with the Department of Electrical and Electronic Engineering, Faculty of Engineering, Imperial College London, SW7 2AZ London, U.K. (e-mail: l.miglietta@imperial.ac.uk).

Louis Kreitmann is with the Department of Infectious Disease, Faculty of Medicine, Imperial College, W12 0NN London, U.K., and also with the Research & Development of bioMérieux, 69280 Auvergne-Rhone-Alpes, France (e-mail: l.kreitmann21@imperial.ac.uk).

Pantelis Georgiou is with the Department of Electrical and Electronic Engineering, Faculty of Engineering, Imperial College London, SW7 2AZ London, U.K., and also with the NIHR Health Protection Research Unit (HPRU), Healthcare Associated Infections and Antimicrobial Resistance, Imperial College, W12 0NN London, U.K. (e-mail: pantelis@imperial.ac.uk).

Alison Holmes and Jesus Rodriguez-Manzano are with the Department of Infectious Disease, Faculty of Medicine, Imperial College, W12 0NN London, U.K., and also with the NIHR Health Protection Research Unit (HPRU), Healthcare Associated Infections and Antimicrobial Resistance, Imperial College, W12 0NN London, U.K. (e-mail: alison.holmes@imperial.ac.uk; j.rodriguez-manzano@imperial.ac.uk).

This article has supplementary downloadable material available at https://doi.org/10.1109/JBHI.2023.3257727, provided by the authors.

Digital Object Identifier 10.1109/JBHI.2023.3257727

I. Introduction

SINCE the invention of Polymerase Chain Reaction (PCR) in 1985, this technology has become the paradigm in the clinical diagnosis of infectious diseases by enabling the rapid and effective detection of DNA and RNA from pathogens [1]. By specifically amplifying the target nucleic acids, PCR is a well-established tool for detecting genetic material even at small concentrations in an extremely sensitive manner [2]. Furthermore, as highlighted by the recent pandemic of COVID-19, real-time quantitative PCR (qPCR) is an essential tool for the surveillance and control of highly infectious diseases at both individual and population levels [3].

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Although widely accepted as a gold-standard diagnostic tool, qPCR is commonly used for single-target amplification reactions (singleplex), which means detecting only one pathogen in a single well each time [4]. With the growing global challenges of public health, detection of multiple pathogens simultaneously has become a demanding issue and raised attention from researchers of several fields. Such identification capabilities can see broad applications in a variety of clinical scenarios, for example, in the screening and diagnosis of influenza-like illnesses caused not by influenza but by other pathogens (e.g., rhinoviruses, coronaviruses, human respiratory syncytial virus, adenoviruses, and human parainfluenza viruses) in patients with comparable symptom, which provides reliable guidance for patient treatments and public-health policy making, in a fast, low-cost, and easy-to-operate manner [5]; or, in the rapid and accurate identification of bacterial infections carrying antimicrobial-resistance (AMR) genes to support better clinical decisions about the use of antimicrobials and improve patient outcomes [6].

One solution to this challenge is the detection of multiple nucleic acid targets simultaneously using PCR or qPCR (multiplexing). Conventional multiplexing approaches require a certain number of reaction wells which increases the cost and the volume of clinical samples required from patients. In the effort of developing cost and time-effective solution for multi-pathogen detection, single-well multiplex PCR conducts several amplification reactions of many targets in a single well setting by combining all the reaction agents [7]. Theoretically, a positive reaction can be observed in the amplification curve of qPCR when any of the targets presents inside the sample.

Differentiating targets in single-well multiplex reactions is not trivial. To identify the reaction products, several methods have been proposed [8], including melting curve analysis (MCA) [9], multi-channel detection [9], [10], and Final Fluorescent Intensity (FFI) Modulation [11], [12], each of which shows drawbacks for their clinical use. When using intercalating dyes (e.g., SYBR green) which are non-specific to targets [13], melting curves can be generated by conducting a melting step at the end of the PCR reaction. As the chemical composition of each amplification product (amplicon) differs from the others, analysis of melting peaks can be applied to distinguish targets. However, the target identification ability is limited by the temperature resolution of equipment and also subject to dedicated chemical optimisation [9], [14]. MCA is also unavailable for many Point-of-Care (PoC) devices where isothermal chemistries are applied [15], [16]. Besides intercalating dyes, fluorescent probe-based solutions can also be used for multiplexing by detecting each target in a separate fluorescence channel. However, the number of targets is constrained by the number of colour channels in the qPCR equipment (4-6 for commercial machines), and the accuracy also suffers from the unavoidable spectrum leakage among channels [17]. Recently, new methods such as High Definition PCR (HDPCR) tried to differentiate targets based on their FFI by conducting sophisticated optimisations on probe and/or primer concentrations [11], [12]. However, FFI-based modulation can be challenged in sensitivity in case of noisy reactions and become less reliable on real-world clinical samples [18]. All these drawbacks of existing approaches indicate the significance of developing a robust, cheap, and easy-to-use solution for single-well and single-channel multiplex PCR.

In recent years, our group has proposed a number of machine learning-based methods for Amplification Curve Analysis (ACA), which utilise the kinetic information encoded in the amplification curves of PCR reactions to classify different clinical targets [16], [18], [19], [20]. In particular, the first ACA method applied the K-Nearest Neighbour (KNN) algorithm on the entire amplification curves and showed promising results in identifying three AMR genes on synthetic DNA [18]. Coupling KNN-based ACA with the MCA method, a newly developed Amplification and Melting Curve Analysis (AMCA) further extended the single-well single-channel detection capability to nine targets on synthetic DNA [19], and validations on over 250 clinical isolates have highlighted the potential of the AMCA clinical use [20]. However, as we previously mentioned, melting curves are not always available: they can be generated neither when using probe-based chemistries (e.g., TaqMan) nor in many PoC devices. Therefore, we explored the feature extraction of amplification curves to further improve the ACA performance without the necessity of introducing thermodynamic information from melting curves. Specifically, by applying five-parameter sigmoidal fitting to amplification curves, we represented each reaction with several fitting parameters, which were further used as features for a Machine Learning (ML) classifier [21].

The aforementioned previous efforts demonstrated the promising application of the ACA method. However, problems still exist: (1) Unlike some other biomedical signals, the nature of amplification curves is not fully revealed, which brings difficulties to the empirical design of manually extracted features [22]. Therefore, an end-to-end solution with automatic feature extraction capability is required. (2) We noticed a significant difference in the distributions of amplification curve data between synthetic DNA and real clinical samples, that are caused by the inherent complexity of real-world samples, in which not only target pathogen genes but also undesired background genetic materials such as human DNA and other bacteria's genes exist [23]. While synthetic DNA can be easily acquired from manufacturers to generate data for ML model training, annotated clinical samples can be difficult to obtain, especially during the emergence of new pathogens where safety and ethical consent need to be considered. It is ideal that we could train ACA deep classifiers with synthetic DNA data rapidly without the need of having access to clinical samples, which requires additional restrictions such as appropriate safety level facilities, and apply the model directly to unannotated clinical samples in clinical facilities. However, when training an ML model on synthetic DNA data and testing on clinical data, there is a significant performance drop compared to cross-validation results on either synthetic or clinical data [21]. From the machine learning perspective, the discrepancy between two data distributions can be regarded as domain differences, where synthetic DNA and clinical specimens are from source and target domains, respectively. In clinical settings, the true identity (true label) of the target domain

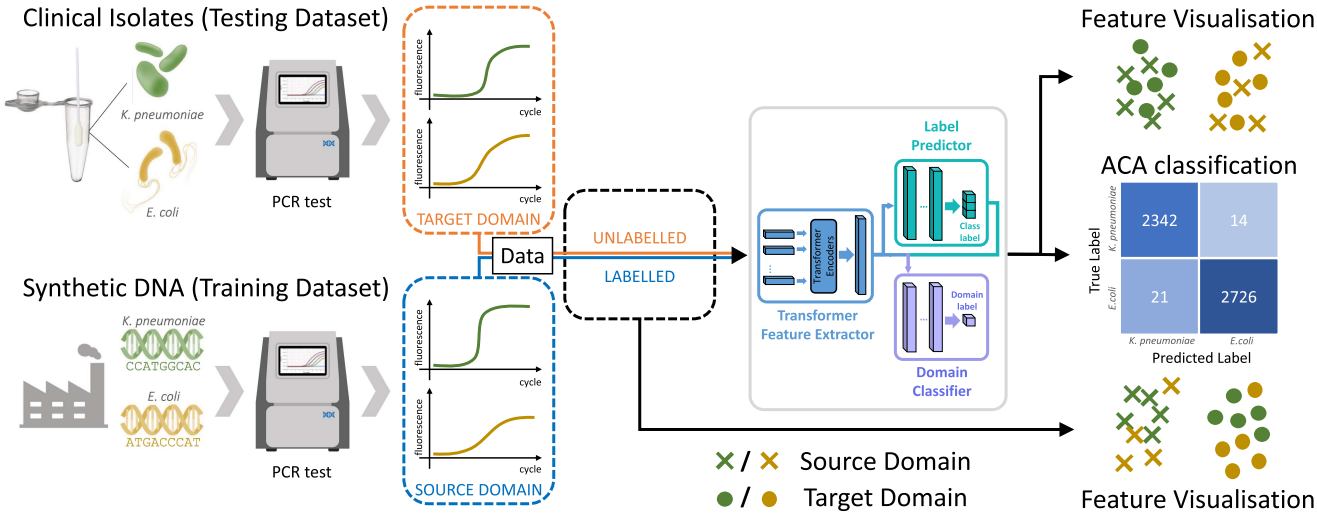


Fig. 1. Concept workflow of the proposed T-CDAN based strategy. To identify different targets within the collected sample, synthetic DNAs of these targets are ordered from the manufacturer and used to generate the training dataset of the deep learning model. The domain shifts between the training (source domain) and testing (target domain) data cause the distant feature distributions in the two domains, and thus result in a dropped performance when applying the source-domain trained model on target-domain data. To eliminate this domain discrepancy, labelled training data and unlabelled testing data are fed into the T-CDAN network, which learns the target and domain information simultaneously. After mapping the inputs into a domain-irrelevant space, T-CDAN can remove the feature distribution differences and provide a clearer decision boundary for the classifier, enabling a better classification result of targets.

is not available, therefore the detection of multiple targets require a robust and unsupervised domain adaptation algorithm.

Over the past few years, deep learning-based approaches have achieved state-of-the-art performance in various classification tasks. One major advantage of utilising deep learning to categorise time series is that it eliminates the need to extract features manually and instead automatically derives high-dimensional feature representations from data while training [24]. Among the existing deep learning structures, transformers have demonstrated excellent modelling abilities for long-range dependencies and interactions in sequential data, making them ideal for time series analysis [25]. Conditional Domain Adversarial Network (CDAN) [26] is an unsupervised domain adaption framework that combines the concepts of conditional Generative Adversarial Network (cGAN) [27] and Domain Adaptation (DA). It has emerged as a powerful tool for reducing marginal distribution discrepancies between domains in the vision field [28]. CDAN outperforms other adversarial learning-based DA algorithms (e.g., Domain-adversarial neural networks or DANN [29] and Adversarial Discriminative Domain Adaptation or ADDA [30]) on the alignment of multimodal distributions, allowing it to work better on multi-classification problems. In this work, our contributions are two-fold: (1) we introduced a state-of-the-art deep transformer-based network which delivers automatic feature extraction through the attention mechanism to classify amplification curves belonging to different pathogens or targets, providing an end-to-end solution for ACA. (2) A novel transformer-based conditional domain adversarial network (T-CDAN) is proposed to eliminate data distribution discrepancies between source domain (synthetic DNA data) and target domain (clinical data).

The overall concept workflow of the proposed strategy is depicted in Fig. 1. To identify different targets within the collected samples, synthetic DNAs of these targets are ordered from the manufacturer and used to generate the training data set of the deep learning model. The domain shifts between the training (source domain) and testing (target domain) data cause the distant feature distributions among them, resulting in a dropped performance when applying the source-domain trained model on target-domain data. To eliminate this domain discrepancy, labelled training data and unlabelled testing data are fed into the T-CDAN network, which learns the target and the domain information simultaneously. After mapping the inputs into a domain-irrelevant space, T-CDAN can remove the feature distribution differences and provide a clearer decision boundary for the classifier, resulting in a better target identification.

To verify the effectiveness of our methods, we evaluated T-CDAN on 198 clinical isolates containing three types of carbapenem-resistant genes (bla_{NDM}, bla_{IMP}, $bla_{\mathrm{OXA\text{-}48}}$), achieving the curve-level accuracy of 93.1% and the sample-level accuracy of 97.0%. Compared to previously published methods, T-CDAN shows an accuracy improvement of 20.9% and 4.9% at curve and sample levels, respectively, with clearer target cluster boundaries and fewer inter-domain distribution differences. This is the first work of using a deep feature generator to extract high-level amplification curve features for target identifications in multiplex PCR, and it emphasises the importance of deep domain adaptation in tackling the real-world clinical problem of molecular diagnostics.

This article is organised in the following way: Section II describes the data and methodology in detail; Section III presents the comparison results, the discussion of which is provided in Section IV; Section V concludes the paper.

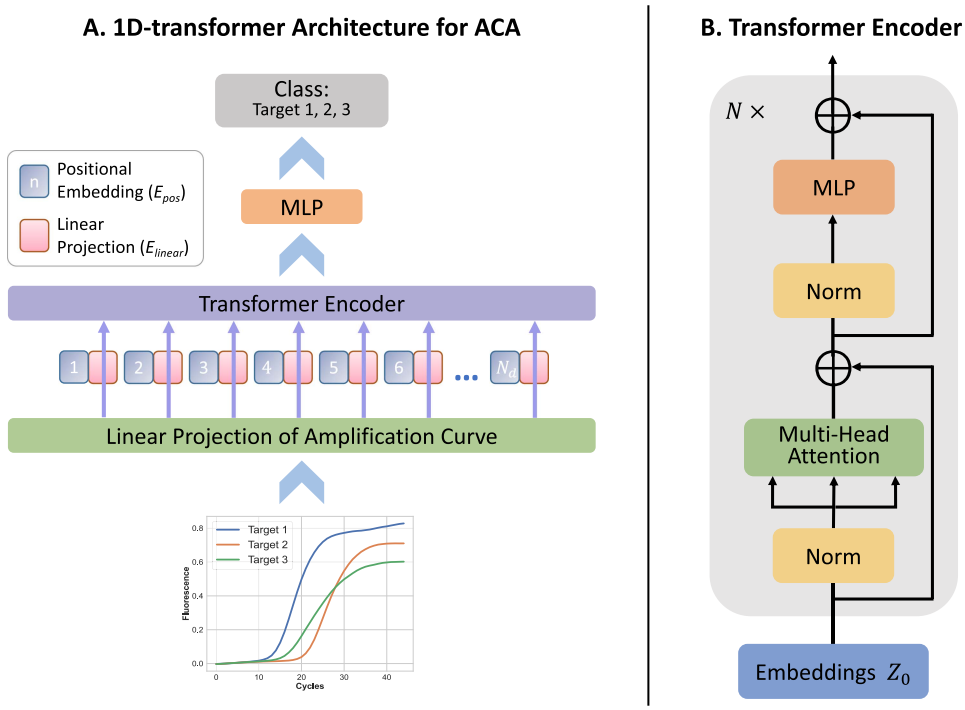


Fig. 2. Part A: Illustration of the proposed 1D-transformer network for amplification curve analysis (ACA). The input amplification curve is embedded using the addition of linear projection and positional embedding. The resulting embedding sequence Z is then fed into N conventional transformer encoders, where its detailed structure is depicted in Part B. An MLP neural network finally utilises the features resulting from the encoders to perform target identification.

II. Methodology

A. Data Information

The data used in this work are originated from Miglietta et al. [21]. The dataset includes amplification events from synthetic DNA (gBlocks™ gene fragments, IDT) of three different carbapenemase genes: bla_{NDM} (N = 18,480), bla_{IMP} (N = 17,710), and $bla_{\text{OXA-48}}$ (N = 17,710). Synthetic target sequences were used as the training/source domain, and 198 clinical isolates containing these three genes were used as the testing/target domain. A total of 152,460 amplification events (each containing 45 data points as the number of PCR cycles performed) were pre-processed accordingly with the pipeline described in Miglietta et al. [21].

B. Transformer-Based Amplification Curve Analysis

Amplification curves are typical 1D time-series data; nevertheless, extracting handcrafted features from these curves is largely underexplored. After preliminary experiments on a number of deep network structures (Table S1), we presented a transformed-based method to process amplification curves in our work due to its strength in processing bio-signal time series and capability in automatic feature extraction.

An overview of the proposed transformer model for classifying amplification curves is shown in Fig. 2. We first applied standardization to set the mean of the data to 0 and the standard deviation to 1. Let $X \in \mathbb{R}^{C\times 1}$ be a standardised amplification curve, where C is the length of the input curve (i.e., number of cycles). X is subsequently projected to a N_d-dimensional latent vector using a trainable linear projection layer. Furthermore, because a transformer encoder lacks recurrence or convolution to leverage sequential information, a typical learnable position embedding E_{pos} were element-wise added to the linear projection results E_{linear} to retain essential sequential information within the curve. The resulting embedding sequence $Z_0 \in \mathbb{R}^{C\times N_d}$ were then fed into N standard transformer encoder layers [31] to extract features (i.e., kinetic information) encoded in the entire amplification curve.

As illustrated in Fig. 2 Part B, the structure of the transformer encoder block is composed of two sub-layers: (1) multi-head self-attention (MSA) and (2) position-wise multilayer perceptron (MLP) sub-layer. In order to speed up the model convergence and avoid the gradient vanishing problem, a residual connection module followed by a Layer Normalization (LN) Module is inserted around each sub-layer. The representation matrix produced by the i_{th} encoder block is of the consistent size as the input embedding matrix Z_0.

Instead of following the original Transformer architecture [31], we replaced the decoder part of the transformer with a feed-forward neural network to allow our model to categorise k clinically relevant targets. Mathematically, the entire workflow of our transformer model can be described as below:

$$Z_0 = E_{\text{linear}} \oplus E_{\text{pos}} \tag{1}$$

$$Z'_i = \text{MSA}\left(\text{LN}\left(Z_{i-1}\right)\right) + Z_{i-1}, \quad \forall i = 1, 2 \ldots N \tag{2}$$

$$Z_i = \text{MLP}\left(\text{LN}\left(Z'_i\right)\right) + Z'_i, \quad \forall i = 1, 2 \ldots N \tag{3}$$

$$y = \text{Softmax}\left(\text{MLP}\left(\text{Flatten}\left(Z_N\right)\right)\right) \tag{4}$$

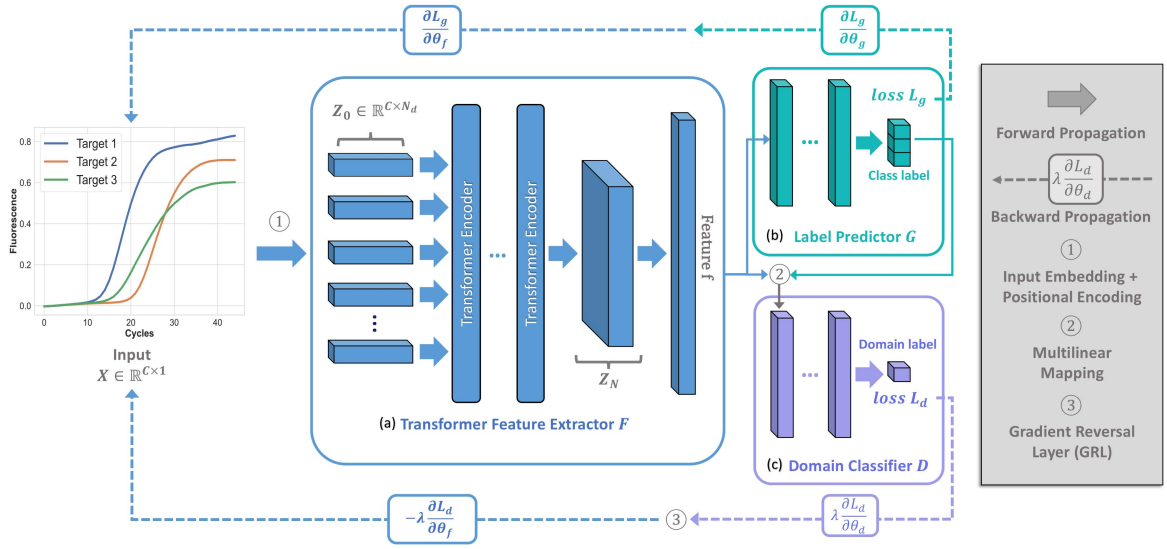


Fig. 3. The proposed T-CDAN framework is composed of three main components: **(a)** deep feature generator F, **(b)** label predictor G, and **(c)** domain classifier D. A sequence of transformer encoder blocks compose F to generate a feature representation f from a standardized amplification curve X. Label predictor is a standard feed-forward neural network to classify source domain data. D is a 2-layer perceptron to determine whether X is from the source or target domain. Multilinear mapping is applied to condition D on label predictions obtained from G, capturing multi-modal information behind the feature distribution. During the training procedure, the label prediction loss L_g and domain classifier loss L_d obtained from G and D are iteratively minimized, enabling F to learn domain-invariant features finally. The Gradient reversal layer inserted between F and D is utilized to discover shared representations between source and target domains, neatly combining domain-specific and domain-invariant shared features [33].

where $\oplus$ denotes element-wise addition and $E_{linear}, E_{pos} \in \mathbb{R}^{C\times N_d}$, Z_i' is the output of i_{th} MSA block and Z_i represents the i_{th} point-wise MLP output. Z_N stands for the output of the last transformer encoder layer and $y \in \mathbb{R}^k$ indicates categorical distribution over class labels.

The presented transformer model in this article is composed of $N = 4$ transformer encoder blocks, where each of them contains a 4-head self-attention module and a position-wise network with 128 hidden neurons. The expected number of embedding vectors generated using linear projection and a position encoding layer is $N_d = 16$. The presented MLP network consists of three hidden layers, where a batch normalisation and a dropout layer are added in between. To address the class imbalance of datasets, we employed focal loss [32] as our loss function.

C. Conditional Adversarial Domain Adaptation for Amplification Curve Analysis

Let synthetic DNA data be a source domain $\mathcal{S} = \{(x_1^s, y_1^s), \ldots, (x_{n_s}^s, y_{n_s}^s)\}$ of n_s labelled amplification curves, where (x_i^s, y_i^s) denotes the i_{th} source domain example, and clinical isolates dataset be a target domain $\mathcal{T} = \{x_1^t, \ldots, x_{n_t}^t\}$ of n_t unlabelled curves, where x_i^t is the i_{th} target domain example. The problem of bridging the domain shift between synthetic DNA and clinical isolates datasets can be intuitively modelled as an unsupervised domain adaptation problem. In our work, we presented a transformer-based conditional domain adversarial network (T-CDAN), to incorporate the aforementioned transformer-based network with Conditional Domain Adversarial Networks (CDAN) [26] strategy, formally mitigating the discrepancy in data distributions across domains.

An overview of the T-CDAN is presented in Fig. 3, including a transformer-based feature extractor F with parameters θ_f, a label predictor G with parameters θ_g, and a conditional domain classifier D with parameters θ_d. In the forward pass, F extracts domain-specific feature representation denoted as $\mathbf{f} = F(\mathbf{x})$ from input curve $\mathbf{x}$. G takes captured features $\mathbf{f}$ as input to output label predictions $\mathbf{g} = G(\mathbf{f})$. D is a binary domain classifier conditioned by the cross-covariance of $\mathbf{f}$ and $\mathbf{g}$, resulting in a domain label which indicates whether $\mathbf{f}$ is from the source domain or target domain. Throughout the training procedure, F grows to learn domain-invariant features that can successfully confuse the domain classifier. D, on the other hand, seeks a rule to distinguish whether the features extracted by F are from the source or target domains.

1) Feature Extractor and Label Predictor: The purpose of the feature extractor F is to learn high-level feature representations from amplification curves, which are implemented in a manner that is consistent with the transformer model described in Section II-B without the last MLP network. The output of the last transformer encoder blocks will be reshaped into a vector and utilised directly as the generated feature representation $\mathbf{f}$. The MLP network attached to the transformer encoder block (see Fig. 2) serves as the label predictor G in the proposed T-CDAN framework. As illustrated in Fig. 3, F and G comprise a conventional feed-forward architecture. Providing the source and target domain data $\mathcal{S}$ and $\mathcal{T}$, the loss function of label predictor G is a negative log-likelihood, expressed as:

$$L_g(\theta_f, \theta_g) = -\mathbb{E}_{(x_i^s, y_i^s)\sim\mathcal{S}} \sum_{c=1}^{n_c} \tau(y_i^s, c) \log G(c; \mathbf{f}_i^s) \quad (5)$$

where $\tau(y_i^s, c) = 1$ if $y_i^s = c$ and $\tau(y_i^s, c) = 0$ otherwise. $\mathbf{f}_i^s$ represents the i_{th} source domain features, $G(c; \mathbf{f}_i^s)$ denotes is the probability of predicting $\mathbf{f}_i^s$ to the c_{th} class, and n_c is the number of class.

2) Conditional Domain Classifier: The aim of the conditional domain classifier is to distinguish the high-level features in amplification curves across the source and target domains. When we deal with multi-classification problems, the feature distribution is multimodal. In this case, the failure of the classical domain classifier in identifying domain types does not necessarily demonstrate that features which belong to the same class but in different domains have identical distributions, resulting in non-ideal classification performance. Therefore, an ideal domain classifier can only be deceived when it differentiates completely transferable and discriminative features. Inspired by conditional GANs [27], the domain classifier D presented in this paper can capture multimodal structures behind the feature distribution, by conditioning D on the class information included in the label prediction $\mathbf{g}$ when adapting the feature representation $\mathbf{f}$. In particular, the conditioning strategy adopted in D is multilinear mapping, where we compute the outer product of $\mathbf{f}$ and $\mathbf{g}$, as defined in (6), to capture multiplicative interactions between the feature representation and the label prediction.

$$M(\mathbf{h}) = M((\mathbf{f}, \mathbf{g})) = \mathbf{f} \otimes \mathbf{g} \tag{6}$$

where $\mathbf{h}$ represents the joint variable of $\mathbf{f}$ and $\mathbf{g}$, and $M(\cdot)$ denotes the multilinear mapping function. Based on binary cross-entropy, the loss function of D is modified and formulated as follows:

$$\begin{aligned} L_d(\theta_f, \theta_d) = & -\mathbb{E}_{x_i^s \sim \mathcal{S}} \log\left[D\left(M(\mathbf{h}_i^s)\right)\right] \\ & -\mathbb{E}_{x_j^t \sim \mathcal{T}} \log\left[1 - D\left(M(\mathbf{h}_i^t)\right)\right] \end{aligned} \tag{7}$$

where $\mathbf{h}_i^s$ and $\mathbf{h}_i^t$ denote the joint variable of $\mathbf{f}_i^s$ and $\mathbf{g}_i^s$, $\mathbf{f}_i^t$ and $\mathbf{g}_i^t$ at i_{th} source or target domain label prediction, respectively. The proposed domain classifier D, like the label predictor, is a feed-forward neural network with two fully connected layers, each of which comprises 256 neurons. Sigmoid activation is adopted in the output layer, since identifying domain type is a binary classification problem.

3) Optimization of T-CDAN: During the training stage, the optimal parameters of the feature extractor ($\hat{\theta}_f$) and the label predictor ($\hat{\theta}_g$) are achieved by maximizing $L_d(\theta_f, \theta_d)$ (7), while minimizing $L_g(\theta_f, \theta_g)$ (6) leads to the optimal parameters of the domain classifier $\hat{\theta}_d$. It yields adversarial learning between the feature extractor/label predictor and the classifier. Based on the above analysis, the optimization of T-CDAN can be formulated as a minimax problem of minimizing the below overall loss function $L(\theta_f, \theta_g, \theta_d)$,

$$L(\theta_f, \theta_g, \theta_d) = L_g(\theta_f, \theta_g) - \lambda L_d(\theta_f, \theta_d) \tag{8}$$

where $0 \leq \lambda \leq 1$ is a hyperparameter for balancing the two objectives. Motivated by previous work [26], [29], a gradient reversal layer (GRL) is added before the conditional domain classifier, as shown in Fig. 3. GRL takes no actions in the forward pass, whereas it reverses the gradient when performing a backpropagation operation, which explains why (8) uses a minus sign to combine two losses into a single overall loss function. The optimal parameters $(\hat{\theta}_f, \hat{\theta}_g, \hat{\theta}_d)$ can be finally learned by iteratively minimising $L(\theta_f, \theta_g, \theta_d)$ in (8). More specifically, $\hat{\theta}_f$ and $\hat{\theta}_g$ can be obtained by addressing the following problem:

$$(\hat{\theta}_f, \hat{\theta}_g) = \arg\min_{\theta_f, \theta_g} L(\theta_f, \theta_g, \hat{\theta}_d) \tag{9}$$

Once the optimal parameters $\hat{\theta}_f$ and $\hat{\theta}_g$ are obtained, solving the below problem to search the optimal parameters $\hat{\theta}_d$ for domain classifier.

$$\hat{\theta}_d = \arg\max_{\theta_d} L(\hat{\theta}_f, \hat{\theta}_g, \theta_d) \tag{10}$$

Adaptive Moment Estimation (Adam) optimizer [34] with the $L2$ regularization of 0.0001 was used to train T-CDAN. The batch size of each domain was 128. As for the learning rate μ, an annealing strategy according to [29] was adopted, defining as $\eta = 0.001(1 + 10p)^{-0.75}$, where $0 \leq p \leq 1$ denotes the training progress. Note that the learning rate of 15μ was applied to train the domain classifier D, since D has a more lightweight design than feature extractor G, making it harder to discriminative features compared to G. The hyperparameter λ for balancing the two objectives gradually increases from 0 to 1 using the following formula:

$$\lambda = \frac{1 - \exp(-10p)}{1 + \exp(-10p)} \tag{11}$$

This method was trained using 10^4 iterations for at least 10 times to reduce the impact of the randomness, and its complete training procedure of T-CDAN for amplification curve analysis is summarized in Algorithm 1. Note that the feature distributions over the source and target domains will be unidentifiable for the domain classifier at the end of the training phase, resulting in domain-invariant feature representation.

III. Results

A. Experimental Setup

In this work, we focus on evaluating the performance of the proposed transformer-based Conditional Domain Adversarial Network (T-CDAN) on distinguishing three carbapenem-resistant genes: bla_{NDM}, bla_{IMP}, and $bla_{\text{OXA-48}}$. All codes were implemented in PyTorch.

We compared the proposed one-dimensional (1D) transformer and T-CDAN with state-of-the-art machine learning methods for amplification curve analysis: K-Nearest Neighbour (KNN) [18] and Random Forest (RF) [21]. It is worth noting that KNN used curve-cycle dimensional data as input, whereas RF used handcrafted features (i.e., sigmoid curve-fitting parameters) as input to the model. 1D-transformer and T-CDAN were trained using the standardised whole curve as inputs.

We used over 14,000 double-stranded synthetic DNA reactions as the source domain curves and over 116,000 clinical isolate reactions as the target domain curves. T-CDAN followed the standard protocol for unsupervised domain adaptation [29] by using labelled source domain data and unlabelled target domain data for training. All other comparison algorithms solely exploited the source domain data during the training procedure.

Algorithm 1: CDAN Model for Amplification Curve Analysis - Training Procedure.

Input :
- Labelled amplification curves from synthetic DNA (source domain) $\mathcal{S} = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^{n_s}$,
- Unlabelled amplification curves from clinical isolates (target domain) $\mathcal{T} = \{\mathbf{x}_j^t\}_{j=1}^{n_t}$,
- Source domain label set $\mathcal{D}_s = \{0\}$ and target domain label set $\mathcal{D}_t = \{1\}$,
- Parameter λ for weighting loss and learning rate μ;

Output: Trained parameters $\hat{\theta}_f, \hat{\theta}_g, \hat{\theta}_d$;

Initialize model parameters and learning rate μ.
while *pre-defined maximum iterations is not met* **do**
  Compute label prediction loss $L_g(\theta_f, \theta_g)$
  Compute domain classifier loss $L_d(\theta_f, \theta_d)$
  `// Update feature generator parameters`
  $\theta_f \leftarrow \theta_f - \mu(\frac{\partial L_g}{\partial \theta_f} - \frac{\partial L_d}{\partial \theta_f})$
  `// Update label predictor parameters`
  $\theta_g \leftarrow \theta_g - \mu\frac{\partial L_g}{\partial \theta_g}$
  `// Update conditional domain classifier parameters`
  $\theta_d \leftarrow \theta_d - \mu\lambda\frac{\partial L_d}{\partial \theta_d}$

Feature extractor and label predictor in T-CDAN, together, compose the architecture of 1D-transformer to fairly evaluate the effectiveness of the domain adaptation.

The two-dimensional t-distributed Stochastic Neighbourhood Embedding (t-SNE) [35] plot was utilised in this research to visualise the difference between feature representations. Besides, the Proxy $\mathcal{A}$-distance [36] is used as a metric to assess the discrepancy between the representations of source and target features. It is computed as $\text{dist}_{\mathcal{A}} = 2(1 - 2\epsilon)$, where ϵ is the test error of a KNN classifier when distinguishing the source domain from the target domain. Furthermore, because each sample in a panel is divided into 770 microwells and utilised to generate 770 curves in our digital PCR chip, we assessed the models' classification performance at both the curve and sample levels (panel-level). The latter is computed by applying hard votes on predicted positive curve results inside the panel. The model's sample-level classification performance is better than its curve-level performance in most cases because it ensembles hundreds of independent predictions. This ensemble strategy ensures that the sample will be distinguished correctly as long as the majority of the curves in a panel are correctly classified.

B. Feature Visualisation

One practical way to compare classification algorithms is to visualise the feature clusters of different targets. The first two columns of Fig. 4 show the t-SNE plots colour-labelled by domains and targets, respectively. When we look at the overall clustering of plots, only the features generated from T-CDAN can form three clearly recognisable clusters with no apparent domain shift.

The domain-labelled plots in Fig. 4(a)–(b) (column 1) illustrate the shift between source and target feature representations. The less overlapping area of the two domains' feature distributions shown in the plot, the more significant the domain gap is. Circled areas in Fig. 4(a)–(c) clearly show that, without the domain adaptation method, there is at least one visible domain discrepancy between the source and target domains. However, as compared to the RF and KNN techniques, the 1D-transformer can yield more domain overlapping features. This shows that the features captured from the transformer attention mechanism are more domain-invariant than the handcrafted features. Fig. 4(d) demonstrates that both source and target domain features generated from T-CDAN have no apparent domain shift and can be projected at relatively consistent locations in 2D t-SNE space. This can be explained by the fact that non-domain adaptation-based classifiers learnt from the source domain data will be biased by the source distribution and adjusted only according to these data. The T-CDAN, on the other hand, focuses on the feature extractor to consider the (unlabelled) distribution of both source and target domain, resulting in a more generalised mapping from raw curves to the feature space, with a unified distribution of features and no significant density differences, as shown in the first column of Fig. 4(d).

The target-labelled plots in Fig. 4(a)–(d) (column 2) illustrate the distances of target clusters in the feature space, where more separate clusters with clearer boundaries and larger distances usually indicate easier classification. By using handcraft sigmoidal fitting parameters as features, the RF method outperforms the original KNN method with more separate target clusters, but there is still an isolated cluster which mix all targets together. The deep features extracted automatically by the transformer further reduce the overlapping area by enlarging the inter-target distances. Even though, cluster boundaries are still unclear between bla_{NDM} and bla_{IMP} features for the 1D-transformer, with a joint circled on the plot. By finding a common space for both domains, as shown in the second column of Fig. 4(d), T-CDAN further benefits the target clustering and enables much clearer boundaries, with no overlap and only a tiny number of data points being clustered to the wrong group.

In addition, we observed that T-CDAN features achieves the smallest $\mathcal{A}$-distance among all models' features, implying that T-CDAN is the most effective method to reduce the domain gap. The 1D-transformer, with the second-smallest $\mathcal{A}$-distance, outperforms the RF and KNN methods in terms of reducing domain shifts, thanks to its ability to extract features from higher-dimensional space compared to machine learning-based methods. This outcome supports the inference we made based on the t-SNE feature visualisations.

C. Quantitative Evaluations

In this subsection, we further analyse the four methods' quantitative results of target identifications. Confusion matrices at the curve and sample level for the four compared algorithms are illustrated in the last two columns of Fig. 4. Although

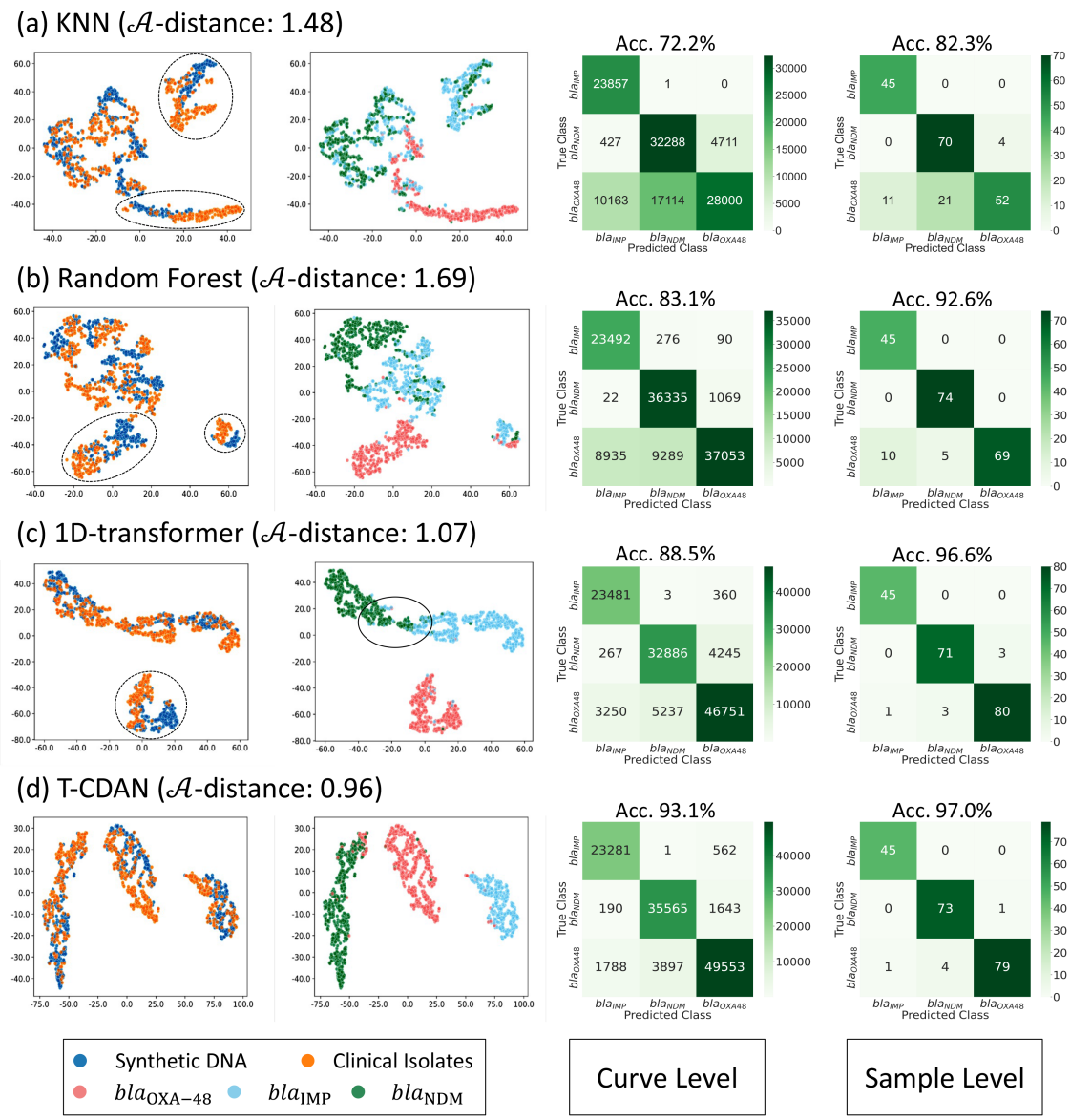


Fig. 4. Visualization and results of considered methods. Each subtitle depicts the name of the evaluated methods and their related $\mathcal{A}$-distance. Column 1&2: t-SNE feature visualisations, where features from different domains are coloured differently in Column 1 plots, and features from distinct AMR targets are coloured differently in Column 2 plots. Column 3: curve-level classification confusion matrix; Column 4: sample-level classification confusion matrix.

all the algorithms, to various extent, struggle in distinguishing $bla_{\text{OXA-48}}$, transformer-based approaches defeat the rest of two classical methods, with much fewer misclassifications shown for this target. The 1D-transformer presents slightly worse performance for bla_{NDM} than the RF method, but these drawbacks are partially eliminated by integrating the transformer into the CDAN framework. When considering sample-level results, the advantages of the newly proposed algorithms in clinical use are emphasised, with only 7 and 6 of 198 samples misclassified for 1D-transformer and T-CDAN, respectively, compared to 38 (KNN) and 15 (RF) for the two previous methods.

Detailed numerical results are shown in Tables I and II. For either curve or sample level, an increasing trend in both F1-score and Accuracy from KNN to T-CDAN methods for all targets can be observed (except for the F1-score between RF and 1D-transformer for bla_{NDM}, where the results of the new method slightly dropped), indicating a better overall performance when utilising the new strategies. Compared to KNN and RF, the transformer increases the mean accuracy by 16.3% and 5.4% at the curve level, and by 14.3% and 4% at the sample level. After applying the domain adaptation, T-CDAN further widens the performance gap by 20.9% and 10% at the curve level, 14.7% and 4.4% at the sample level. Providing the 97% of sample-level accuracy of T-CDAN, it is very promising for this algorithm to be introduced to real-world clinical diagnosis after certain optimisations [37]. Although KNN and RF-based methods show slightly higher sensitivity and specificity for some targets, their performance is severely biased and unbalanced among targets.

In Fig. 5, where the micro-averaged Receiver Operating Characteristics (ROC) of all the algorithms are shown, it shows that

TABLE I
CURVE-LEVEL PERFORMANCE OF FOUR METHODS IN THREE AMR TARGET CLASSIFICATIONS

		KNN	Random Forest	1D-transformer	**T-CDAN**
bla_{IMP}	Specificity (%)	69.3	72.4	87.0	**92.2**
	Sensitivity (%)	**99.9**	98.5	98.5	97.6
	F1-score	0.818	0.834	0.924	**0.948**
bla_{NDM}	Specificity (%)	65.4	79.2	86.3	**90.1**
	Sensitivity (%)	86.3	**97.1**	87.9	95.1
	F1-score	0.744	0.872	0.870	**0.925**
$bla_{\text{OXA-48}}$	Specificity (%)	85.6	**97.0**	91.0	95.7
	Sensitivity (%)	50.7	67.0	84.6	**89.7**
	F1-score	0.636	0.793	0.878	**0.926**
	Mean Accuracy (%)	72.2	83.1	88.5	**93.1**

TABLE II
SAMPLE-LEVEL PERFORMANCE OF FOUR METHODS IN THREE AMR TARGET CLASSIFICATIONS

		KNN	Random Forest	1D-transformer	**T-CDAN**
bla_{IMP}	Specificity (%)	80.4	81.8	**97.8**	**97.8**
	Sensitivity (%)	**100.0**	**100.0**	**100.0**	**100.0**
	F1-score	0.891	0.900	**0.989**	**0.989**
bla_{NDM}	Specificity (%)	76.9	93.7	**95.9**	94.8
	Sensitivity (%)	94.6	**100.0**	95.9	98.7
	F1-score	0.848	**0.967**	0.959	**0.967**
$bla_{\text{OXA-48}}$	Specificity (%)	92.9	**100.0**	96.4	98.8
	Sensitivity (%)	61.9	82.1	95.2	**94.0**
	F1-score	0.743	0.902	0.958	**0.963**
	Mean Accuracy (%)	82.3	92.6	96.6	**97.0**

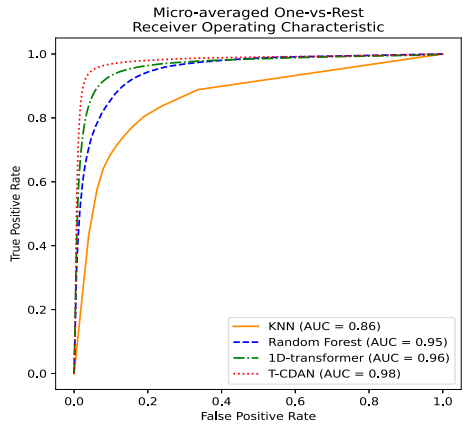


Fig. 5. Micro-averaged Receiver Operating Characteristics of all the compared algorithms.

the proposed T-CDAN provides the highest Area under Curve (AUC) among all the compared methods.

IV. DISCUSSION

We can plainly see from the findings that transformer models extract more discriminative features that are more robust to domain shift than machine learning-based methods. The primary reason is that the positional encoding module enables the transformer to extract additional positional information from the curve. Furthermore, unlike other types of time series, qPCR amplification data from different targets represent a complex classification problem because of difficulties in manipulating the shape of this biological signal. As a result, the local features retrieved from the original dimensional space may be insufficiently discriminative. The encoding block in the transformer model, on the other hand, can extract deeper features by mapping curves into a higher dimensional space. In addition, providing that the amplification curve can be further divided into three different phases: (1) Initiation phase, (2) Exponential phase and (3) Plateau phase, the kinetic information from different phases can be effectively captured and aggregated by the multi-head attention mechanism of the encoding block.

T-CDAN outperforms all other methods in the AMR target classification because non-domain-adaptation-based methods only aim to perform well on the source domain data, whereas T-CDAN also ensures that the feature distributions across the two domains are as similar as possible. Besides, T-CDAN exploits multilinear mapping to control the uncertainty of label predictions, which guarantees the features' transferability and discriminability simultaneously. In addition, T-CDAN is also robust in the convergence, which is illustrated in Fig. S1.

Unlike conventional machine learning methods, the proposed T-CDAN provides balanced performance for all the target categories, which will benefit the real-world clinical application where observations of different pathogens follow various distributions [38]. In this research, we demonstrated the proposed algorithm on digital PCR data. An apparent advantage of using dPCR for learning-based multiplexing is that numerous amplification reactions that happened simultaneously generate millions of amplification curves in a short period of time, providing the huge training dataset required by deep neural networks. Leveraging dPCR, we can further extend the training size and introduce more variability easily in our future work.

As all other deep learning models, T-CDAN can be deployed on standard PC or workstations after the training progress is

finished, and the inference time for classifying new samples can be negligible, since the network size of the transformer backbone is not large. These facts make the proposed algorithm accessible to every PCR lab with basic computing resources.

We are also aware of the demand for implementing the proposed deep ACA algorithm in conventional qPCR devices, which will widen the application scenarios and increase the clinical value of the methodology by integrating it into the current diagnosis flow without any hardware modification. Understandably, data from dPCR and qPCR will show heterogeneous structures with inconsistent distributions. The potential performance gap when transferring a dPCR trained network to a qPCR application can also be filled by the proposed deep domain adaptation approach, inside which the qPCR data are considered coming from the target domain. Future work will be focused on validating the T-CDAN on qPCR while gaining benefits from the large-scale dPCR training data.

Under certain circumstances, the data distributions shift among different experiments due to the variance of chemical agents, operation procedures, and manufacturer bias, which can become a significant factor causing the inter-experiment discrepancy in performance. If data from each experiment are considered belonging to a separate domain, this issue can be described as a multi-domain adaption problem. Our future work will further extend T-CDAN into the multi-source one-target (multiple training experiments vs one testing) and one-source multi-target (one training experiment vs multiple testing) manners, by integrating with other techniques such as knowledge distillation to learn domain-invariant features on multiple domains.

We demonstrated in this article the precise classification of three AMR genes. Our next step is to enlarge the target number in the multiplex assay while maintaining a high accuracy level. Furthermore, research will be carried on for co-amplification situations, for example, when double infection occurs and more than one target show up in a single sample. Our preliminary results suggest the possibility of identifying co-infections by including them as additional categories and applying ACA accordingly [18], [20]. Efforts will also be made regarding the optimisation of chemical agents, such as modifying primer and probe concentrations [12], to generate amplification curves with more significant shape discrepancy when co-amplifications happen.

Regarding the economic benefit of single-well multiplexing, it can be assumed that by applying multiplex and multi-target detection of N targets in a single well, the total cost of screening these N targets would be reduced to $1/N$ compared to the traditional singleplex setting which requires emphN times the reaction chambers. In addition, preparation time can be reduced, resulting in a simplified and lower workload for the entire testing.

V. Conclusion

This article proposed a novel framework, referred to as a Transformer-based Conditional Domain Adversarial Network (T-CDAN), to address the problem of domain discrepancy in amplification curve analysis. Currently, no published study has applied deep learning techniques to analyse PCR amplification curves. Also, our work is the first to incorporate the idea of a one-dimensional Transformer and CDAN to alleviate domain shifts in time series. Extensive experiments validated the effectiveness of the T-CDAN for the target identification of three AMR genes: bla_{NDM}, bla_{IMP}, and $bla_{\text{OXA-48}}$. A comparison of experimental results with the state-of-the-art ACA methods illustrates that T-CDAN achieves the most promising classification performance, by learning domain-invariant and discriminative feature representations from labelled synthetic DNA data and unlabelled clinical isolates data. T-CDAN provides the lowest $\mathcal{A}$-distance value of all approaches and exhibits the clearest decision boundary between different targets in feature visualisations, showing that it can significantly reduce the discrepancy between feature distributions from both domains while ensuring feature discriminability. We believe that T-CDAN's impressive performance in bridging the domain gap between multiple amplification curve datasets demonstrates its potential to address various types of domain shifts that may occur during PCR multiplexing.

References

[1] S. Yang and R. E. Rothman, "PCR-based diagnostics for infectious diseases: Uses, limitations, and future applications in acute-care settings," *Lancet Infect. Dis.*, vol. 4, no. 6, pp. 337–348, 2004.

[2] M. A. Valasek and J. J. Repa, "The power of real-time PCR," *Adv. Physiol. Educ.*, vol. 29, no. 3, pp. 151–159, 2005.

[3] Y. Wang, H. Kang, X. Liu, and Z. Tong, "Combination of RT-qPCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak," *J. Med. virol.*, vol. 92, no. 6, pp. 538–539, 2020.

[4] J. Parker et al., "Analytical sensitivity comparison between singleplex real-time PCR and a multiplex PCR platform for detecting respiratory viruses," *PLoS One*, vol. 10, no. 11, 2015, Art. no. e0143164.

[5] A. M. Caliendo, "Multiplex PCR and emerging technologies for the detection of respiratory pathogens," *Clin. Infect. Dis.*, vol. 52, no. suppl_4, pp. S326–S330, 2011.

[6] E. Tacconelli et al., "Surveillance for control of antimicrobial resistance," *Lancet Infect. Dis.*, vol. 18, no. 3, pp. e99–e106, 2018.

[7] H. Zhang, M. Ganova, Z. Yan, H. Chang, and P. Neuzil, "PCR multiplexing based on a single fluorescent channel using dynamic melting curve analysis," *ACS Omega*, vol. 5, no. 46, pp. 30267–30273, 2020.

[8] L. Kreitmann et al., "Next-generation molecular diagnostics: Leveraging digital technologies to enhance multiplexing in real-time PCR," *TrAC Trends Anal. Chem.*, vol. 160, 2023, Art. no. 116963.

[9] C. T. Wittwer, "High-resolution dna melting analysis: Advancements and limitations," *Hum. Mutat.*, vol. 30, no. 6, pp. 857–859, 2009.

[10] A. S. Whale, J. F. Huggett, and S. Tzonev, "Fundamentals of multiplexing with digital PCR," *Biomol. Detection Quantification*, vol. 10, pp. 15–23, 2016.

[11] A. Rajagopal et al., "Significant expansion of real-time PCR multiplexing with traditional chemistries using amplitude modulation," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, 2019.

[12] L. Jacky et al., "Robust multichannel encoding for highly multiplexed quantitative PCR," *Anal. Chem.*, vol. 93, no. 9, pp. 4208–4216, 2021.

[13] H. Ihmels and D. Otto, "Intercalation of organic dye molecules into double-stranded dna–general principles and recent developments," in *Supermolecular Dye Chemistry*. Berlin, Germany: Springer, 2005, pp. 161–204.

[14] I. Simko, "High-resolution dna melting analysis in plant research," *Trends Plant Sci.*, vol. 21, no. 6, pp. 528–537, 2016.

[15] J. Rodriguez-Manzano et al., "Handheld point-of-care system for rapid detection of SARS-CoV-2 extracted RNA in under 20 min," *ACS Central Sci.*, vol. 7, no. 2, pp. 307–317, 2021.

[16] K. Malpartida-Cardenas et al., "Single-channel digital lamp multiplexing using amplification curve analysis," *Sensors Diagn.*, vol. 1, no. 3, pp. 465–468, 2022.

[17] S. Tyagi, S. A. Marras, and F. R. Kramer, "Wavelength-shifting molecular beacons," *Nature Biotechnol.*, vol. 18, no. 11, pp. 1191–1196, 2000.

[18] A. Moniri et al., "Amplification curve analysis: Data-driven multiplexing using real-time digital PCR," *Anal. Chem.*, vol. 92, no. 19, pp. 13134–13143, 2020.
[19] A. Moniri, L. Miglietta, A. Holmes, P. Georgiou, and J. Rodriguez-Manzano, "High-level multiplexing in digital PCR with intercalating dyes by coupling real-time kinetics and melting curve analysis," *Anal. Chem.*, vol. 92, no. 20, pp. 14181–14188, 2020.
[20] L. Miglietta et al., "Coupling machine learning and high throughput multiplex digital PCR enables accurate detection of carbapenem-resistant genes in clinical isolates," *Front. Mol. Biosci.*, vol. 8, 2021, Art. no. 775299..
[21] L. Miglietta et al., "Adaptive filtering framework to remove nonspecific and low-efficiency reactions in multiplex digital PCR based on sigmoidal trends," *Anal. Chem.*, vol. 94, no. 41, pp. 14159–14168, 2022.
[22] J. SantaLucia, "Physical principles and visual-OMP software for optimal PCR design," *PCR Primer Des.*, vol. 402, pp. 3–33, 2007.
[23] F. Bollmann, I. Casper, J. Henke, and A. Pautz, "qRT-PCR: A method and its difficulties," *Naunyn-Schmiedeberg's Arch. Pharmacol.*, vol. 349, pp. 949–951, 2012.
[24] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," *Data Mining Knowl. Discov.*, vol. 33, no. 4, pp. 917–963, 2019.
[25] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, 2022.
[26] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 1647–1657.
[27] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
[28] X. Yu et al., "Conditional adversarial domain adaptation with discrimination embedding for locomotive fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2020.
[29] Y. Ganin et al., "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.
[30] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7167–7176.
[31] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.
[32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
[33] A. Kamath, S. Gupta, and V. Carvalho, "Reversing gradients in adversarial domain adaptation for question deduplication and textual entailment tasks," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 5545–5550.
[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[35] G. E. Hinton and S. Roweis, "Stochastic neighbor embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 857–864.
[36] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 513–520.
[37] Z. S. Ulhaq and G. V. Soraya, "The diagnostic accuracy of seven commercial molecular in vitro SARS-CoV-2 detection tests: A rapid meta-analysis," *Expert Rev. Mol. Diagnostics*, vol. 21, no. 7, pp. 733–740, 2021.
[38] K. L. Hopkins et al., "Diversity of carbapenemase-producing enterobacterales in England as revealed by whole-genome sequencing of isolates referred to a national reference laboratory over a 30-month period," *J. Med. Microbiol.*, vol. 71, no. 5, 2022, Art. no. 001518.