Multi-Modal Diagnosis of Infectious Diseases in the Developing World

Girmaw Abebe Tadesse, Hamza Javed, Nhan Le Nguyen Thanh, Hai Duong Ha Thi, Le Van Tan, Louise Thwaites, David A. Clifton, and Tingting Zhu

***Abstract*—In low and middle income countries, infectious diseases continue to have a significant impact, particularly amongst the poorest in society. Tetanus and hand foot and mouth disease (HFMD) are two such diseases and, in both, death is associated with autonomic nervous system dysfunction (ANSD). Currently, photoplethysmogram or electrocardiogram monitoring is used to detect deterioration in these patients, however expensive clinical monitors are often required. In this study, we employ low-cost and mobile wearable devices to collect patient vital signs unobtrusively; and we develop machine learning algorithms for automatic and rapid triage of patients that provide efficient use of clinical resources. Existing methods are mainly dependent on the prior detection of clinical features with limited exploitation of multi-modal physiological data. Moreover, the latest developments in deep learning (e.g. cross-domain transfer learning) have not been sufficiently applied for infectious disease diagnosis. In this paper, we present a fusion of multi-modal physiological data to predict the severity of ANSD with a hierarchy of resource-aware decision making. First, an on-site triage process is performed using a simple classifier. Second, personalised longitudinal modelling is employed that takes the previous states of the patient into consideration. We have also employed a spectrogram representation of the physiological waveforms to exploit existing networks for cross-domain transfer learning, which avoids the laborious and data intensive process of training a network from scratch. Results show that the proposed framework has promising potential in supporting severity grading of infectious diseases in low-resources settings, such as in the developing world.**

Manuscript received July 18, 2019; revised October 26, 2019; accepted November 19, 2019. Date of publication January 9, 2020; date of current version July 2, 2020. This work was supported in part by the Wellcome Trust under Grants 107367/Z/15/Z and 089276/B/09/7, and in part by the Royal Academy of Engineering (RAEng) Frontiers of Engineering for Development under the Global Challenges Research Fund. The work of T. Zhu was supported by the RAEng Research Fellowship. The work of D. A. Clifton was supported by the RAEng and the Engineering and Physical Sciences Research Council. *(Corresponding author: Girmaw Abebe Tadesse.)*

G. A. Tadesse is with the Department of Engineering Science, University of Oxford, OX1 2JD Oxford, U.K., and also with IBM Research, Africa, 00100, Nairobi, Kenya (e-mail: girmaw.abebe@eng.ox.ac.uk).

H. Javed, D. A. Clifton, and T. Zhu are with the Department of Engineering Science, University of Oxford, OX1 2JD Oxford, U.K. (e-mail: hamza.javed@eng.ox.ac.uk; davidc@robots.ox.ac.uk; tingting.zhu@eng.ox.ac.uk).

N. L. N. Thanh is with Children Hospital No. 1, Ho Chi Minh City 774, Vietnam (e-mail: nhanlnt@oucru.org).

H. D. H. Thi, L. V. Tan, and L. Thwaites are with the Oxford Clinical Research Unit, Ho Chi Minh City 774, Vietnam (e-mail: haduong200385@yahoo.com.vn; tanlv@oucru.org; lthwaites@oucru.org).

Digital Object Identifier 10.1109/JBHI.2019.2959839

***Index Terms*—Multi-modal, infectious diseases, deep learning, fusion, developing world.**

I. Introduction

INFECTIOUS diseases, such as tetanus and hand, foot and mouth disease (HFMD), can still be life-threatening conditions for patients in low and middle income countries [1]. Tetanus often affects the poorest in society in low and middle income countries, and it was estimated to have caused 48-80,000 deaths in 2015 [1]–[4]. Tetanus cases can progress to severe conditions in many cases, and the subsequent hospital treatment is often lengthy (up to six weeks). Comparatively, HFMD is typically a benign self-limited illness in infants and young children. In recent years, HFMD outbreaks that affected millions of children have been reported in the Asia Pacific region [5], [6]. Although most cases are mild, a small number of affected children progress rapidly to severe or fatal manifestations of the disease. Predicting those few children who will progress to severe disease is challenging in HFMD and as a result huge numbers of children are admitted to hospital as a precautionary measure, placing an enormous burden on healthcare systems [5]–[7].

Autonomic nervous system dysfunction (ANSD) is the main cause of death in the aforementioned infectious diseases [2], [6], [7]. ANSD particularly affects the cardiovascular system and can be detected by examining the autonomic control of the heart. Early detection of ANSD is often challenging clinically, yet treatment becomes difficult once the condition is established. Thus, early ANSD detection is an important task that could improve patient outcomes.

Existing approaches to automatically evaluate the severity of ANSD mainly require the detection of each QRS complex in the electrocardiogram (ECG) waveform followed by the extraction of vital signs such as heart rate and RR intervals, i.e. intervals between adjacent QRS complexes [8], [9]. However these features are not generalisable across different modalities, e.g. photoplethysmogram (PPG), and are prone to movement artefacts. Moreover, QRS detection incurs additional computational cost. Previously, we have demonstrated that generic time and frequency features [10] outperformed the traditional heart rate variability (HRV) features [8] across multiple datasets of infectious disease patients. However, the diagnosis performance still has room for improvement, particularly using latest modeling techniques (e.g. deep learning) and transferable features from other domains with minimum computational expenditure.

Convolutional neural networks (CNNs) are shown to achieve accurate malaria diagnosis in the works of [11], [12]. Additionally, the benefits of transfer learning to facilitate feature learning is also demonstrated in [12]. However, similar deep-learning approaches are not common in the diagnosis of tetanus and HFMD patients.

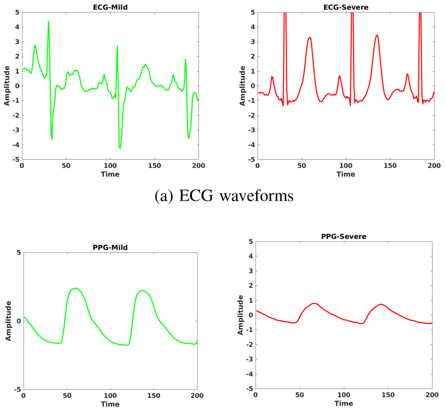
In this paper, we propose a proof-of-principle approach that aims to fuse multi-modal physiological data, collected using low-cost wearable devices, for the diagnosis of infectious disease (i.e. tetanus and HFMD) patients. The diagnosis maps physiological patient data to the severity level of ANSD. Our contributions are as follows.

First, we propose a multi-layer decision making diagnosis step which is decomposed into an on-site triage process that provides rapid diagnosis, followed by a longitudinal model for personalised diagnosis. Second, a multi-modal or -stream framework and different fusion strategies are developed. Third, we validated the proposed approach on multiple infectious disease datasets, i.e. tetanus and HFMD, collected in intensive care units of hospitals in Vietnam. Finally, we applied cross-domain transfer-learning by mapping time-series physiological signals to images using spectrogram representations, thereby enabling the use of existing computer vision networks.

The motivation for this work is the need to fuse multiple modalities of physiological data to obtain multi-stage screening of tetanus and HFMD patients in low-resource settings. To do so, we employ cross-domain transfer learning to spectral representations of time-series signals using existing deep networks designed for natural images. Though there exists a domain gap between spectral and natural images, a spectrogram of a time-series signal gives a visual representation of dynamic information that can be thought of as being composed of low level component features such as edges, lines and general shapes, which are also common low level components in natural images. Therefore despite the final 2D representations being quite different, the earlier layers of computer vision architectures trained on natural images can be thought of as containing capability useful for discriminating spectral images. Furthermore, our approach can be used to exploit other computationally lighter versions of existing networks, which can be employed on ubiquitous devices such as a smartphone. The use of such networks also provides the benefit of dimensionality reduction from images into a vector that best represents the spectral image. Moreover, for multi-stream physiological bio-signals, e.g. 12-lead ECG, the proposed framework could help encode the spatial relationship among multiple leads.

The proposed approach could provide data-driven insights in the diagnosis of such infectious diseases, and hence improve patient care in hospital settings with limited resources. Furthermore, automatic diagnosis of infectious disease may also reduce unnecessary use of antibiotics and therefore limit antimicrobial resistance since patients suspected of having these diseases are often given antibiotics as a precautionary measure.

The remainder of this paper is organised as follows. Section II reviews related works in the diagnosis of infectious diseases in the developing world. Section III provides the problem formulation followed by a step-by-step analysis of the proposed approach. Section IV describes the tetanus and HFMD datasets used for validation. Section V details the design parameters in the proposed approach, baseline methods selected for comparison, and the metrics employed to evaluate the classification performance. We then present and discuss the results in Section VI. Finally, concluding remarks are provided in Section VII.



(a) ECG waveforms

(b) PPG waveforms

Fig. 1. Examples of, (a) ECG and, (b) PPG waveforms extracted from tetanus patients with Mild (green colour) and Severe (red colour) diagnosis of autonomous nervous system dysfunction.

II. Related Work

The early identification of severe infections is a task that has begun receiving considerable interest from researchers, due to the obvious practical benefits early detection could have for patient outcomes. In the work of [8], the relationship between a patient's health status and irregularities exhibited in their ECG readings was highlighted. This was through markers like HRV, which can be derived and computed from the aforementioned signal waveforms. A comparative analysis was carried out in [9], in which the use of PPG readings to derive HRV was investigated, due to the fact that the PPG is easier to measure than ECG. From the results presented, it is clear that estimating such markers from either waveforms remains challenging in practice, due to the difficulty of robustly extracting RR intervals. However, there is clear utility in developing methods that can determine the health status of a patient from signals that are easily acquired from inexpensive sensors.

More recently, significant research has been conducted in the use of ML for identifying severe infection, a prominent example being the detection of sepsis, which is one of the leading causes of mortality for hospital inpatients. Despite this, it remains extremely challenging to detect using developed approaches [13], [14]. Yet the ability of machine learning (ML) methods to outperform heuristic and rule-based methods currently in use for detecting the disease, illustrates the promise of data driven modelling approaches at discovering complex patterns and insights from medical data.

By comparison, research into the use of ML for predicting risk of developing severe cases of infections such as tetanus and HFMD, which primarily affect low and middle income countries, has to date been limited. In the work of Zhang and

Liu [15], [16], random forest and gradient boosting tree models are developed to classify severe cases of HFMD from mild ones, for a study dataset of 530 paediatric patients, taken from Guangdong hospital in China. Impressive predictive performance was achieved in both models. Additionally, a feature importance analysis was also conducted to identify covariates that had the most influence on whether a severe case of HFMD went onto manifest itself.

However, it is worth noting that many of the features used, particularly those ranked highest in terms of importance, were those obtained using advanced hospital facilities and resources. For example, MRI scan results and laboratory blood tests. In a resource-constrained setting, these features are not available or affordable. By contrast, in this work we propose the use of features collected from inexpensive wearable sensors, i.e. PPG and ECG signals, which can be acquired at an earlier stage of a patient care pathway through a triage process, potentially removing the need for hospitalisation altogether. The practical benefits of such an approach would be most apparent for citizens in lower and middle income countries, where healthcare systems are not as well resourced.

Another area of related work includes the early and rapid diagnosis of malaria, a parasitic infection that is estimated to be responsible for 400,000 deaths a year, primarily in developing countries [19], [20]. Notable studies include the use of Naive Bayes and SVM for multi-class classification of different stages of malaria infection, on a dataset of 230 samples obtained from a hospital in India [17]. Using handcrafted features, both ML methods were able to obtain reasonable predictive accuracy. By contrast, in the works of [11], [12], deep learning methods in the form of CNNs are successfully employed to make diagnoses with higher accuracies. The latter work also explores the use of transfer learning, which would be appropriate when dataset sizes are small.

In many of the aforementioned papers, particularly those concerning healthcare tasks, classical ML methods are frequently considered and employed [15]–[17], [21], [22]. Although reasonable to very good performance can be achieved using classical techniques (e.g. SVMs), deep learning is recognised as the current state of the art within the ML field. This is due to the impressive performances deep neural network architectures have obtained across multiple application domains. In addition to the performance achievable using deep learning, one other important advantage they offer is the ability to avoid laborious and time consuming feature engineering, rather 'raw' data can instead usually be fed into the network directly from which features are learnt.

However, due to the many parameters that need to be learned by deep architectures, the ability to train robust models that generalise well, large amounts of data are required. For many healthcare problems, the amount of data available can be seriously limited, as is the case with many of the studies reviewed in this section. In such contexts, deep neural network architectures can still be exploited through the use of transfer learning. That is through employing a model trained on a similar but considerably larger dataset (such as for general image recognition), which is then fine tuned on the smaller dataset for the task in question. Transfer learning has been successfully employed for a range of medical tasks, particularly medical imaging diagnoses, through using popular computer vision architectures like Inception and ResNet [12], [23].

In previous work, we demonstrated the utility of the PPG and ECG physiological signals at predicting ANSD severity levels arising from tetanus and HFMD infections, using hand-crafted features [10], [18]. In this paper we leverage multi-modal ML and transfer learning to produce deep learning architectures to better predict infection severity using signals that could be obtained from inexpensive wearable sensors. Multi-modal ML research to date has primarily focused on applications such as audio-visual speech recognition, gesture identification, video captioning and affect analysis [24].

By comparison, limited research has been conducted into multi-modal/-stream ML for healthcare. The techniques developed for audio-visual fusion however, were successfully shown to be capable of identifying emotional and mental health well being [25]. A different application task was considered in the work of [26], where video, accelerometer and GPS data streams were successfully fused for activity and fall detection for elderly patients in home settings. Taking into account the improvements that can be achieved by considering multiple data modes, in this work we propose fusing different data modalities and streams using Fourier analysis to create 2D representations of the data. This enables the use of transfer learning through existing architectures like Inception [27], thereby providing the possibility of achieving superior diagnosis ability. The mapping of different modes to image representation was shown to be effective for the fusion of accelerometer data for activity detection in the work of [28].

A summary of the key relevant work in the detection of infectious diseases, in the context of the developing world, is provided in Table II. The table highlights the different approaches used for identifying infection severity, and how the proposed approach in this paper distinguishes itself from these works.

III. Proposed Method

Let $\mathcal{C} = \{c_l\}_{l=1}^{L}$ be a set of L ANSD severity levels of infectious disease patients. Let $\mathcal{P}_n$ be the nth sample window that contains a set of multi-modal (-stream) physiological data, i.e. $\mathcal{P}_n = \{\mathbf{p}_n^m\}$, where $m \in \{1, 2, \ldots, M\}$ and M is the total number of modalities or streams. We aim to provide resource-aware triage process by predicting the ANSD severity level in $\mathcal{P}_n$, i.e. $\mathbf{s}_n \in \mathcal{C}$, using a fusion of the multi-modal information, transfer learning from existing vision-based deep convolutional networks, and recurrent neural network for longitudinal modelling (see Fig. 2). The proposed approach consists of *multi-modal physiological data acquisition*, *cross-domain transfer learning*, *on-site triage process* and *personalised longitudinal modelling*. This section describes the details of each block in the proposed framework.

A. Multi-Modal Physiological Data Acquisition

Different physiological data streams (Fig. 1) are acquired from patients of infectious diseases using low-cost wearable

TABLE I
SUMMARY OF KEY EXISTING WORKS THAT EMPLOY DATA-DRIVEN APPROACHES TO SUPPORT DECISION MAKING IN THE DIAGNOSIS OF INFECTIOUS DISEASE PATIENTS IN THE DEVELOPING WORLD

				Features Extraction			Fusion			
Existing Work [Ref.]	Diseases	Modalities	Mobile sensors	Manual	Deep learning	Transfer learning	Feature fusion	Decision fusion	Multi-layer modelling	Multiple datasets
Zhang et al. [15]	HFMD	EHR	x	✓	x	x	x	x	x	x
Liu et. al [16]	HFMD	EHR	x	✓	x	x	x	x	x	x
Das et al. [17]	Malaria	Image	✓	✓	x	x	x	x	x	x
Delahunt et. al [11]	Malaria	Image	✓	x	✓	x	x	x	x	✓
Liang et al. [12]	Malaria	Image	✓	x	✓	✓	x	x	x	x
Duong et al. [18]	Tetanus	ECG	✓	✓	x	x	✓	x	x	x
Abebe et al. [10]	Tetanus, HFMD	ECG, PPG	✓	✓	x	x	✓	x	x	✓
Proposed	Tetanus, HFMD	ECG, PPG	✓	x	✓	✓	✓	✓	✓	✓

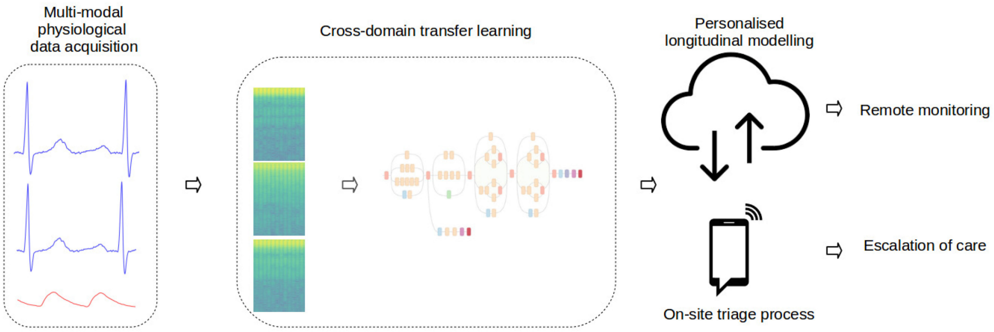
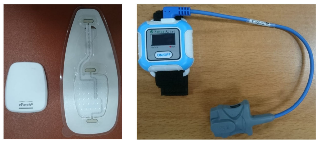
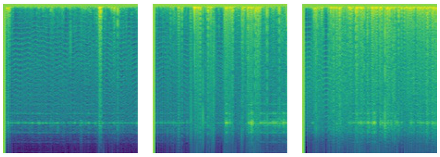

Multi-modal physiological data acquisition
Cross-domain transfer learning
Personalised longitudinal modelling
Remote monitoring
Escalation of care
On-site triage process

Fig. 2. Block diagram of the proposed approach. Multi-modal/-stream physiological data are collected and pre-processed. A cross-domain transfer learning is applied using frequency-time representation and existing networks such as Inception. On-site triage process is then employed to determine the escalation of care for a patient. Personalised longitudinal modelling is applied in a cloud that can help to remotely monitor the patient.



(a) E-patch (b) Pulse oximeter

Fig. 3. Wearable devices which could be used for ECG and PPG data collection, respectively, (a) E-Patch with mounting electrolyte and (b) Pulse oximeter with adjustable wrist-strap.



(a) Mild (b) Intermediate (c) Severe

Fig. 4. Spectrogram examples extracted from ECG waveforms of HFMD patients with Mild, Intermediate and Severe ANSD cases.

devices such as a pulse oximeter (see Fig. 3). These streams could be from different modalities such as ECG and PPG. In addition, a single modality may contain multiple channels, e.g. the conventional 12-lead ECG. ECG signals are generated by electrical activity of the sinoatrial node, which controls the expansion and contraction of the heart (see Fig. 1(a)). PPG signals represent the changes in light absorption of the skin, measured using an infrared sensor emitting light on the skin, when the heart pumps blood into peripheral vessels (see Fig. 1(b)).

B. Cross-Domain Transfer Learning

Physiological signals collected using wearable sensors are often susceptible to noise and movement artefacts. Hence, a high pass filter followed by a Gaussian filter is applied to mitigate these issues. Following the noise filtering, we encode signal variation (dynamics) using a fast Fourier transform (FFT), $\mathcal{F}(\cdot)$. The output of the FFT is then rearranged to obtain a frequency-time representation, i.e. spectrogram, which contains the frequency response magnitude at different frequency bins (see Fig. 4). Let $\mathbf{p}_n^m$ be the mth modality patient data of $\mathcal{P}_n$, the spectrogram can be presented as $S_n^m = \mathcal{F}(\mathbf{p}_n^m)$. We normalise S_n^m by its maximum value and bound its values between [0, 255] similar to the intensity values of natural images as follows:

$$\bar{S}_n^m = \log\left(\frac{S_n^m}{max(S_n^m)} * 255\right). \quad (1)$$

Logarithmic-scale is applied to smooth the spectrogram values since much of the energy lies in the lower frequency bins.

Raw time-series data could also be fed directly to a Conv1D-based deep network. Compared to the proposed spectrogram-based Conv2D approach, Conv1D applied directly on raw time series data, however, has the following limitations. 1) Heart rate variability is the main marker to diagnose ANSD in clinical practice. However, feeding the raw data to a network makes temporal encoding difficult compared to using spectrograms that already contain pre-processed temporal information. 2) It requires more training data to effectively extract the variability from raw time series by training a network from scratch. 3) Compared to raw data, spectrogram representation provides robustness against variations in device specifications and mounting positions of wearable sensors [26, 27]. 4) In addition to its effective temporal encoding using a short-time Fourier transform, a spectrogram representation enables transfer learning by using existing computer vision networks, which in turn enable multi-modal learning; 5) Finally, spectrograms of multiple channels, e.g. incase of 12 leads ECG, help to exploit their spatial relationships via stacking.

Examples of spectrograms obtained from the ECG waveforms of Mild, Intermediate and Severe ANSD cases of HFMD patients are shown in Fig. 4. We utilised existing vision models (e.g. GoogLeNet [27]) that are pre-trained on large image datasets (e.g. ImageNet [29]) to achieve cross-domain transfer learning between time-series physiological data and natural images. Thus, a pre-trained deep network could be used to extract hidden-layer features, $\mathbf{d}_n^m \in \mathbb{R}^D$ (where D is the feature dimension), from each normalised spectrogram of a modality, $\bar{S}_n^m$. The features are then fed into on-site triage process presented below.

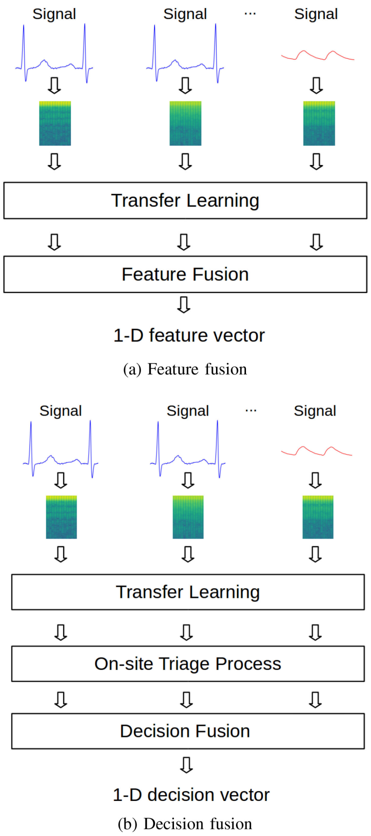
C. On-Site Triage Process

On-site triage process refers to the modelling and decision making steps that are done locally, i.e. the triage process pre-screens patients on-site, e.g. on a mobile device. As a result, we propose to employ simple classifier, such as logistic regression, due to constrained on-site resources and the high dimension of CNN features. This approach is also feasible in wearable system settings where an initial decision is required to be made on-site (on the wearable device) with limited computational resources.

Information fusion is another issue associated with using multi-modal/stream physiological data for decision making. Fusion could be applied at different stages and hence can be categorised as either feature or decision fusion (See Fig. 5). In the proposed framework, feature fusion refers to the concatenation of the CNN features, and it results in single dimensional feature vector as input for the on-site triage process. Decision fusion, on the other hand, involves the concatenation / accumulation of the on-site triage outputs of each modality (see Section III-D). Feature fusion is more plausible for on-site triage process, which allows features from different modalities to be combined before severity modelling. Thus, feature fusion helps to avoid separate modelling for each modality, and it is beneficial when the modalities are highly correlated, e.g. when considering multiple ECG leads. In the proposed approach, features extracted from the hidden layers of existing CNNs could be concatenated for simple classifier-based triage process. The drawback of feature-level fusion is evident when the feature dimension of each modality is already high as in the case of CNN features. Thus, the concatenation of M CNN feature vectors (each D-dimensional) results in an even higher feature dimension, i.e. $[\mathbf{d}_n^m]_{m=1}^M \in \mathbb{R}^{M*D}$. The on-site triage process outputs a decision vector $\mathbf{l}_n \in \mathbb{R}^L$, with $L << D << M * D$. The decision vector, $\mathbf{l}_n$, consists of the score for each class, and the class with the highest score becomes the predicted label. The longitudinal model, on the hand, is relatively intensive computationally and hence proposed to be done remotely.

D. Personalised Longitudinal Modeling

The on-site triage process provides the instantaneous ANSD severity level prediction for short-duration physiological data from an infectious disease patient. Longitudinal modelling helps encode the temporal dependency among subsequent samples of the patient and predict the inference level while considering the previous states of the same patient, i.e. provide a personalised prediction. This can naturally be extended to predict the severity of a patient in the future.



(a) Feature fusion

(b) Decision fusion

Fig. 5. Fusion types employed in the proposed framework: (a) feature fusion in the on-site triage process and (b) decision fusion in the longitudinal modelling.

To this end, we employ a recurrent neural network (RNN) to model the longitudinal dependency among subsequent samples drawn from a single patient. Vanilla RNN is often characterised by its inability to easily capture or encode long-term temporal patterns, due to *vanishing* and *exploding* gradient problems [30]. Gated networks, such as long short-term memory (LSTM) networks, have been introduced to control the flow of information from the input to the output in recurrent networks and thus avoid vanishing and exploding gradients.

The input feature vector to the RNN-based longitudinal modelling, $\mathbf{q}_n$, consists of the output of the on-site triage process, i.e. $\mathbf{q}_n = \mathbf{l}_n \in \mathbb{R}^L$. Note that if feature fusion was not applied during the triage process, longitudinal modelling would be applied on the decision fusion of triage outputs from different modalities/streams.

Decision fusion addresses the limitations of feature-level fusion when the feature dimension of each modality is high. It also exploits each modality separately during the triage step, which is beneficial particularly when the modalities are less correlated and each provides different discriminative characteristics, e.g. ECG and PPG (See Fig. 5(b)). The on-site triage process provides a low-dimension feature vector per modality, $\mathbf{l}_n \in \mathbb{R}^L$, which is a vector of class probabilities. In decision fusion, triage outputs from multiple modalities could be fused either through decision concatenations (DC) or decision accumulation (DA). DC concatenates the class probability vectors of different modalities that results in $\mathbf{q}_n = [\mathbf{l}_n^m]_{m=1}^M \in \mathbb{R}^{M*L}$, whereas DA sums up the prediction vector per class across the modalities and resulting in $\mathbf{q}_n \in \mathbb{R}^L = \sum_{m=1}^M \mathbf{l}_n^m$.

Finally, we compute ANSD severity level, $\mathbf{s}_n$, using softmax normalisation on the current hidden state of the longitudinal model, i.e. $\mathbf{h}_n$, as follows:

$$\mathbf{s}_n = \frac{e^{W_{hs}\mathbf{h}_n}}{\sum_{l=1}^L e^{W_{hs}\mathbf{h}_n}}, \quad (2)$$

where $W_{hs} \in \mathbb{R}^{L\times\nu}$ is the wrapping matrix.

IV. Datasets

We used two infectious disease datasets for the validation of the proposed approach. The datasets are of HFMD and tetanus patients admitted in hospitals in Vietnam. The data collection was approved by the relevant Ethical Committees and carried out in line with the declaration of Helsinki.

A. HFMD Dataset

We collected HFMD dataset from 74 patients, the majority of whom were children less than three years old, at Children Hospital No. 1, Ho Chi Minh City, Vietnam. We used unobtrusive commercial devices, i.e. E-patch,[1] in order to collect ECG waveforms from HFMD patients, with a sampling rate of 256 Hz. The collection is designed to acquire patient data at least twice during hospital stay. First, a 24-hour ECG recording is performed when a patient is admitted to the infectious disease department. Then, another round of recording is done on the penultimate day of hospitalisation. The clinical diagnosis of the patients is used as a ground truth for the proposed approach. The clinical score contains five labels: $2a$, $2b_1$, $2b_2$, 3 and 4 in the order of increasing severity. However, there are obvious class imbalances in the HFMD dataset as the number of patients (per class) is as follows: $2a(33)$, $2b_1(9)$, $2b_2(11)$, $3(20)$ and $4(1)$. As a result, we decided to merge adjacent classes, i.e. level-$2b_1$ and level-$2b_2$ as Intermediate and level-3 and level-4 as Severe.

B. Tetanus Dataset

Ten tetanus patients, all adults, were recruited for a proof-of-principle study at the Hospital for Tropical Diseases, Ho Chi Minh City. Four of the patients were diagnosed with Mild ANSD and the remaining six as Severe cases. Temporally synchronised ECG (300 Hz) and PPG (100 Hz) waveforms, each approximately lasting up to 24 hours, were collected from a patient. A Datex Ohmeda monitor and a pulse oximeter were employed for data acquisition. In order to download the waveforms from the monitor, we employed VS Capture software.

V. Experiment Setup

In this section, we describe the setups of parameters in each step of the pipeline, the baseline methods used for comparison and the performance metrics used to evaluate the proposed approach.

A. Parameter Setup

A time-series physiological data stream is decomposed into a sequence of non-overlapped windows (samples) upon which training and testing is performed. We set the duration of window length to be five minutes, similar to the duration in the clinical baseline method [8]. The total number of samples extracted from each modality of the tetanus dataset is 3,141, which consists of 1,117 Mild and 2,024 Severe samples. From HFMD dataset, a total of 60,373 samples are extracted from each ECG lead, which consists of 20,151 Mild, 19,594 Intermediate and 20,628 Severe samples. For the spectrogram generation, we applied a short-time Fourier transform on each chunk of five seconds in a window. Overlapping percentage of 95% is applied on subsequent chunks to obtain smooth frequency-time (spectrogram) representation. The ECG waveforms in both HFMD and tetanus datasets are characterized by higher sampling rates, i.e. 256 Hz and 300 Hz respectively, compared to 100 Hz PPG waveforms. Thus, a decimation (with a factor of 2) is performed during spectrogram computation on the ECG waveforms. After normalization and logarithmic scaling of the magnitude of the frequency response values (see Eq. 1), the spectrogram is stored as an image in JPG format using the default 'viridis' color map. We set the parameters (e.g. window length and percentage of overlapping for the FFT calculation) to obtain a square-like spectrogram of 146×161 pixels. Thus, no significant change on the input spectrogram occurs due to the resizing method employed by an existing network, e.g. Inception.

For the cross-domain transfer learning, we experimented with Inception-v3 [27], MobileNet [31] and MnasNet [32] trained on ImageNet [29] to extract the CNN features on the spectrogram images. We extracted the Inception features from the next-to-last layer of Inception-v3, i.e. '$pool\_3:0$,' which provides a $D = 2,048$ dimensional feature vector for each spectrogram. Similarly, we extracted MobileNet and MnasNet features with dimensions $D = 1,280$ and $D = 1,056$, respectively. We applied fixed batch normalisation after features are extracted from the hidden layer of an existing network (e.g. Inception and MobileNet) prior to on-site triage process and/or longitudinal modelling. In our approach, we did not opt to fine-tune any hyper-parameters of the existing architectures. Different modalities are expected to contribute an equivalent number of samples from each patient. Thus, replication of samples is applied in order to achieve data balance among modalities when lower samples are available from a specific modality.

Logistic regression and SVM are experimented with for the on-site triage process. They take the CNN features as inputs and each outputs a score for each class. For the longitudinal modelling, we employ an LSTM recurrent network that contains three additional gates (input, output and forget) in order to control flow of information and hence avoid vanishing gradients. We used a single layer LSTM (without stacking) to

[1] [Online]. Available: epatch.madebydelta.com

keep the framework simple, factoring in the limited size of the HFMD and tetanus datasets in addition to the high dimensional CNN feature input. The number of neurons in each gate of the LSTM is set to $\nu = 64$ neurons trained with a batch size of 12 and with 100 epochs. We set the recursive duration to contain $T = 6$, i.e. 30 minutes for a 5-min window duration, which means the LSTM can utilise the information from previous 5 samples when it makes decision. Adam-optimizer is used for training with an initial learning rate of 0.01. Before training, we split the data from each patient sequentially to train and test sets with a ratio of 80% and 20%, respectively. Each window (sample) in the train and test sets is classified independently during on-site triage process, e.g. using SVM, and later on the temporal relationships among subsequent samples is exploited using the longitudinal model, e.g. LSTM. Though, the level of severity hardly changes in the validation datasets, this approach would help to predict the deterioration of a patient in advance. Both train and test sets are normalized to force the scale of each feature element to unit variance.

B. Baseline Methods

We compared the proposed approach with two existing works: our previous work [10] and a baseline method [8], which employed handcrafted features for the classification of severity levels. Abebe *et al.* [10] applied simple time- and frequency-domain features whereas Malik *et al.* [8] strictly required the detection of each QRS complex in ECG waveforms followed by the extraction of vital signs such as heart rate and RR intervals. We employed an SVM with a Gaussian kernel to validate the baseline features.

In the proposed approach different fusion strategies are investigated. These include FC-LSTM: feature-level concatenation of CNN features from different modalities/streams followed by LSTM; FC-LR-LSTM: feature-level concatenation of the feature groups followed by logistic regression and LSTM; LR-DC-LSTM: decision-level concatenation of LR outputs of the feature groups prior to the LSTM; LR-DA-LSTM: decision-level accumulation of LR outputs of the feature groups prior to the LSTM. We have also experimented with SVM for the on-site triage process and compared it with LR.

C. Performance Metrics

We employ the following performance metrics for our evaluation: accuracy (A), precision (P), sensitivity or recall (R), specificity (S) and F-score (F_1).

For multi-class classification as in the HFMD dataset, one-vs-all (OVA) strategy is applied to compute the performance metrics per severity level. Each experiment is repeated 10 times (number of iterations), and each iteration is performed with a new initialisation set of the network parameters. The average performance is computed, first, across the severity levels, followed by another averaging across the iterations. The standard deviation of performance metrics across the iterations is also reported.

VI. Results and Discussion

This section presents the ANSD level classification performance achieved in both tetanus and HFMD patients. Compared to the baseline methods, we discuss the results achieved via transfer learning, feature and decision fusions, and longitudinal modelling. Moreover, the misclassification among severity levels will be discussed. Then follow the experiments that investigate the effect of window duration in the spectrogram generation and hidden layer size in the LSTM-based temporal modelling. Finally, the proposed framework is also validated on mobile architectures, such as MobileNet [31] and MnasNet [32].

TABLE II
On-Site Triage Performance on Tetanus Patients
FC: Feature-Level Concatenation; LR: Logistic Regression;
SVM: Support Vector Machine

			Baseline - SVM performance (%)		
Method	A	P	R	S	F_1
PPG- [10]	70.2 ± 1.0	70.4 ± 0.8	92.6 ± 0.3	29.5 ± 2.9	80.0 ± 0.5
ECG- [10]	80.2 ± 0.7	78.4 ± 0.9	95.3 ± 0.5	53.4 ± 2.5	86.0 ± 0.4
FC- [10]	78.2 ± 1.0	75.3 ± 1.0	98.1 ± 0.3	43.1 ± 3.3	85.2 ± 0.6
			Transfer learning (%)		
	A	P	R	S	F_1
PPG-LR	90.2 ± 0.3	94.1 ± 0.4	90.5 ± 0.2	89.8 ± 0.0	92.3 ± 0.1
ECG-LR	91.5 ± 0.1	93.2 ± 0.4	93.6 ± 0.2	87.6 ± 0.1	93.4 ± 0.3
PPG-SVM	89.4 ± 0.2	87.7 ± 0.4	97.3 ± 0.6	75.2 ± 0.1	92.2 ± 0.2
ECG-SVM	92.4 ± 0.1	95.5 ± 0.5	92.7 ± 0.1	92.0 ± 0.0	94.0 ± 0.6
			Transfer learning + Feature fusion (%)		
	A	P	R	S	F_1
FC-LR	94.5 ± 0.3	97.0 ± 0.8	94.4 ± 0.5	94.7 ± 0.5	95.7 ± 0.6
FC-SVM	92.3 ± 0.2	90.9 ± 0.0	97.8 ± 0.7	82.3 ± 0.5	94.2 ± 0.3

A. On-Site Triage

1) Tetanus: Table II shows the performance of the proposed approach for the on-site triage process compared with the baseline methods. First, we present the performance of hand-crafted time and frequency domain features in the baseline methods using SVM. Second, we present the impact of transfer learning by validating the CNN features extracted from existing networks using LR and SVM for triage process. Third, we assess the performance improvement via feature fusion. We primarily reference the F-score results in our discussion, as it is the harmonic mean of the precision and recall values.

a) Transfer learning: Results in Table II show that the SVM-based validation of hand-crafted features achieves the lowest performance compared to the use of CNN features extracted via cross-domain transfer learning in the proposed framework. This demonstrates the limited generalisability of manually designed features. We experimented with both LR and SVM to validate the CNN features obtained from PPG and ECG waveforms using transfer learning, which results in at least 12% and 7% performance improvements over PPG and ECG waveforms, respectively.

b) Feature fusion for triage: The fusion of CNN features from multiple-modalities has been shown to slightly improve the on-site triage process as FC-LR (95.7%) and FC-SVM (94.2%) achieved the highest F-score values among LR-based and SVM-based methods respectively, as detailed in Table II.

c) LR vs. SVM: LR performs competitively with SVM, and even exploits the feature fusion better than the SVM, i.e. 95.7% vs. 94.2%, respectively. This can likely be attributed, at least in part, to the high dimension of the CNN features ($D = 2,048$).

d) Modalities: Due to its relatively stable acquisition process, ECG waveforms proved to be more discriminant than

TABLE III
ON-SITE TRIAGE PERFORMANCE ON HFMD PATIENTS FC: FEATURE-LEVEL CONCATENATION; LR: LOGISTIC REGRESSION; SVM: SUPPORT VECTOR MACHINE

	Baseline - SVM performance (%)				
Method	A	P	R	S	F_1
ECG- [8]	57.1 ± 0.2	35.0 ± 0.2	35.2 ± 0.2	67.6 ± 0.1	34.6 ± 0.2
ECG-2- [10]	70.9 ± 0.1	60.6 ± 0.1	55.9 ± 0.2	78.0 ± 0.1	55.7 ± 0.2
FC- [10]	70.2 ± 0.1	60.0 ± 0.1	54.5 ± 0.1	77.3 ± 0.1	53.9 ± 0.2
	Transfer learning (%)				
	A	P	R	S	F_1
ECG-1-LR	68.5 ± 0.2	52.8 ± 0.1	52.8 ± 0.0	76.4 ± 0.3	52.8 ± 0.2
ECG-2-LR	69.8 ± 0.0	54.9 ± 0.2	54.8 ± 0.1	77.4 ± 0.2	54.8 ± 0.1
ECG-1-SVM	72.6 ± 0.3	60.1 ± 0.2	58.9 ± 0.0	79.4 ± 0.3	59.1 ± 0.0
ECG-2-SVM	72.8 ± 0.1	60.9 ± 0.0	59.2 ± 0.0	79.5 ± 0.4	59.5 ± 0.1
	Transfer learning + Feature fusion (%)				
	A	P	R	S	F_1
FC-LR	70.6 ± 0.4	56.0 ± 0.2	56.0 ± 0.2	78.0 ± 0.3	56.0 ± 0.2
FC-SVM	76.3 ± 0.2	65.8 ± 0.1	64.3 ± 0.1	82.1 ± 0.4	64.6 ± 0.0

PPG when validated with both hand-crafted features (86.0% vs. 80.0%) and CNN features (93.4% vs. 92.3% using LR and 94.0% vs. 92.2% using SVM).

2) HFMD: Table III shows the on-site triage process performance of the baseline features, CNN features and feature-level fusion for HFMD patients. Compared to classifying tetanus severity (Table II), the overall performance of classifying the severity levels of HFMD patients is significantly lower, as detailed in Table III. This can be attributed to the noise introduced through the wearable sensors, due to the motion artefacts as the HFMD patients are children, who often make spurious movements during data collection. Moreover, compared to the binary (Mild vs. Severe) ANSD prediction for tetanus patients, the task by definition is more challenging for HFMD patients as it involves three class prediction, i.e. Mild, Intermediate and Severe.

a) Transfer learning: Similarly to tetanus results (see Table II), the baseline features are inferior to the CNN features for ANSD prediction on HFMD patients, particularly when SVM is employed for both feature types. Baseline features from ECG-2 resulted in 55.7% whereas the CNN features from ECG-2 achieved 59.5%.

b) Feature fusion: FC-LR (56.0%) and FC-SVM (64.6%) achieved the highest LR- and SVM-based F-scores, respectively, for the on-site triage process of HFMD patients. This reflects the advantage of feature fusion in the process. However, the improvement is not as large as it is for tetanus severity prediction, due to the difficulty of the prediction task in the HFMD dataset. This is also partly due to the limitation of feature-level fusion when the individual modalities or streams are highly correlated, as in the case of the ECG-1 and ECG-2 streams of the HFMD dataset.

c) LR vs. SVM: LR is shown to be inferior to SVM during the individual validation of the ECG streams as F-scores of ECG-1 and ECG-2 increased, respectively, from 52.8% and 54.8% (using LR) to 59.1% and 59.5% (using SVM). Moreover, the SVM exploited the feature-level fusion of the CNN features better than the LR: 64.6% vs. 56.0%.

d) Modalities: The two streams of ECG waveforms have achieved competitive performance across different methods and classifiers. This is expected as the two channels of the ECG data are highly correlated to each other.

TABLE IV
MILD AND SEVERE LEVEL CLASSIFICATION USING LONGITUDINAL MODELLING FOR TETANUS PATIENTS. FC: FEATURE-LEVEL CONCATENATION; DC: DECISION-LEVEL CONCATENATION, DA: DECISION-LEVEL ACCUMULATION; LR: LOGISTIC REGRESSION; SVM: SUPPORT VECTOR MACHINE; LSTM: LONG SHORT-TERM MEMORY NETWORK

	Feature fusion + Longitudinal model (%)				
	A	P	R	S	F_1
FC-LR-LSTM	97.3 ± 0.3	98.0 ± 0.3	97.8 ± 0.5	96.5 ± 0.6	97.9 ± 0.2
FC-SVM-LSTM	95.6 ± 0.5	94.7 ± 0.5	98.7 ± 0.4	90.1 ± 1.0	96.7 ± 0.4
	Decision fusion + Longitudinal model (%)				
	A	P	R	S	F_1
LR-DC-LSTM	97.2 ± 0.3	97.0 ± 0.3	98.8 ± 0.3	94.4 ± 0.6	97.9 ± 0.2
LR-DA-LSTM	97.1 ± 0.1	96.7 ± 0.0	98.9 ± 0.2	93.8 ± 0.0	97.8 ± 0.1
SVM-DC-LSTM	97.0 ± 0.6	95.9 ± 0.9	99.5 ± 0.1	92.3 ± 1.8	97.7 ± 0.4
SVM-DA-LSTM	96.9 ± 0.4	95.7 ± 0.6	99.6 ± 0.1	92.0 ± 1.1	97.6 ± 0.3

TABLE V
ANSD LEVEL CLASSIFICATION USING LONGITUDINAL MODELING FOR HFMD PATIENTS. FC: FEATURE-LEVEL CONCATENATION; DC: DECISION-LEVEL CONCATENATION; DA: DECISION-LEVEL ACCUMULATION; LR: LOGISTIC REGRESSION; SVM: SUPPORT VECTOR MACHINE; LSTM: LONG SHORT-TERM MEMORY NETWORK

	Feature fusion + Longitudinal model (%)				
	A	P	R	S	F_1
FC-LR-LSTM	74.8 ± 0.5	62.2 ± 0.7	62.4 ± 0.7	81.1 ± 0.4	61.0 ± 0.9
FC-SVM-LSTM	77.7 ± 0.2	68.3 ± 0.4	66.3 ± 0.4	83.2 ± 0.2	66.6 ± 0.4
	Decision fusion + Longitudinal model (%)				
	A	P	R	S	F_1
LR-DC-LSTM	72.7 ± 0.8	61.5 ± 0.6	59.3 ± 1.1	79.5 ± 0.6	55.3 ± 1.9
LR-DA-LSTM	71.6 ± 0.9	61.6 ± 0.6	57.7 ± 1.3	78.7 ± 0.7	52.8 ± 2.3
SVM-DC-LSTM	75.7 ± 0.2	65.8 ± 0.4	63.3 ± 0.4	81.7 ± 0.2	63.6 ± 0.3
SVM-DA-LSTM	75.4 ± 0.6	65.7 ± 0.8	62.9 ± 0.9	81.5 ± 0.5	63.2 ± 0.9

B. Longitudinal Modeling

This subsection describes the results achieved using the personalised longitudinal model and discusses the effects of feature-fusion and decision-fusion on the performance improvement.

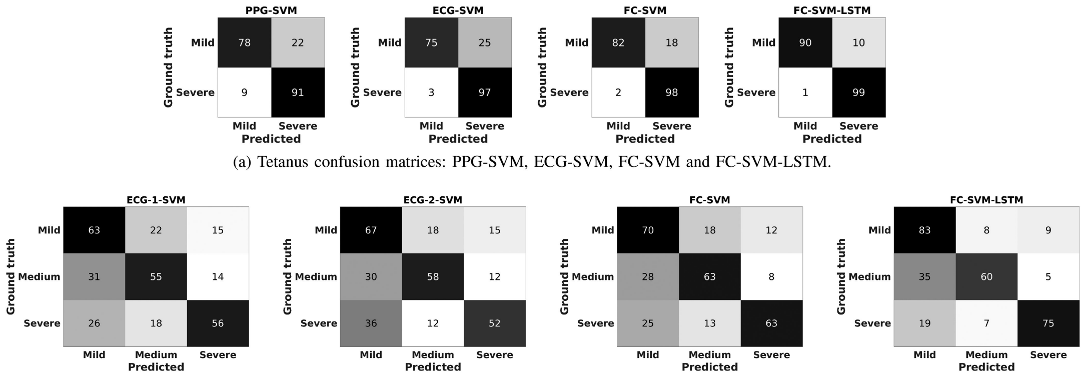
1) Tetanus:

a) Improved performance with longitudinal models: The LSTM-based longitudinal model improved the performance for severity level classification for the tetanus patients as shown in Table IV. The highest LR-based and SVM-based F-scores in Table II, i.e. 95.7% (FC-LR) and 94.2% (FC-SVM), have been improved to 97.9% (FC-LR-LSTM) and 96.7% (FC-SVM-LSTM), respectively. Particularly noteworthy is the fact that the tetanus dataset contains longer duration physiological waveforms. Thus, the temporal model was likely able to exploit this information and hence improve performance.

b) Feature fusion vs. decision fusion: Table IV also shows that the type of fusion is less significant when the longitudinal modelling is applied, particularly for LR-based approaches as FC-LR-LSTM (feature fusion) and LR-DC-LSTM and LR-DA-LSTM (decision-fusions) achieved similar F-scores. SVM, on the other hand, has shown a slight improvement with decision fusions. The two decision fusions, decision concatenation and decision accumulation, also achieved similar performance.

2) HFMD:

a) Improved performance with longitudinal models: Table V shows performance improvement due to the longitudinal model for HFMD patients (as Table IV did for tetanus patients). The LSTM-based temporal model improved the LR-based performance from 56% (FC-LR) to 61.0% (FC-LR-LSTM).



(a) Tetanus confusion matrices: PPG-SVM, ECG-SVM, FC-SVM and FC-SVM-LSTM.

(b) HFMD confusion matrices: ECG-1-SVM, ECG-2-SVM, FC-SVM and FC-SVM-LSTM.

Fig. 6. Confusion matrices of severity-level detection in the (a) HFMD and (b) tetanus datasets that show a step-by-step performance improvement. First, individual performance of CNN features from each modality using transfer learning, followed by feature fusion and then LSTM-based longitudinal modelling (1-min temporal windows is employed).

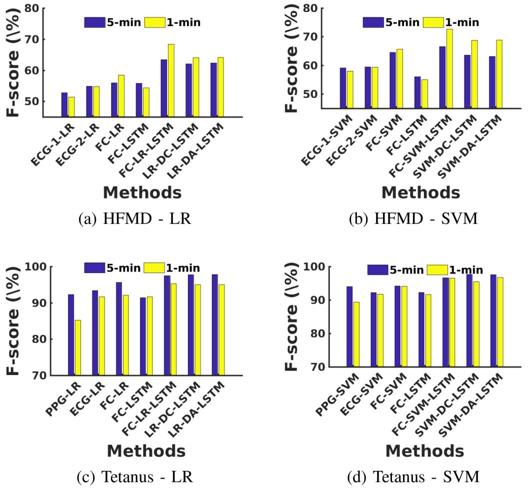
Similarly, the SVM-based performance increased from 64% (FC-SVM) to 66.6% (FC-SVM-LSTM).

b) Feature fusion vs. decision fusion: The impact of high correlation among individual streams is demonstrated in Table V, where feature fusion achieved significantly higher performance compared to decision fusion approaches when the longitudinal model is employed. For example, 61% by the LR-based feature fusion (FC-LR-LSTM) is higher than the decision fusions: LR-DC-LSTM (55.3%) and LR-DA-LSTM (52.8%). Similarly, 66.6% by the SVM-based feature fusion (FC-SVM-LSTM) is superior to the decision fusions, SVM-DC-LSTM (63.6%) and SVM-DA-LSTM (63.2%). It is evident that the two decision fusion schemes performed competitively with each other. Decision concatenation helped to expand the feature space fed to the longitudinal model when compared to decision accumulation, and hence resulted in higher performance, especially when LR is employed.

C. Misclassification

The misclassification among ANSD severity levels is shown in Fig. 6(a) and (b) for tetanus and HFMD datasets, respectively. High sensitivity is achieved on Mild vs. Severe classification of tetanus patients. ECG and PPG waveforms are shown to perform competitively, and their concatenation (FC-SVM) somewhat reduced the misclassification of Mild samples to Severe, and FC-SVM-LSTM further reduces the misclassification from 19% to 11%.

The confusion matrices in Fig. 6(b) reveal that, on the HFMD dataset, the individual ECG leads (ECG-1 and ECG-2) are shown to perform equivalently due to the high correlation between the streams, i.e. they are not different modalities as ECG and PPG but multiple streams of a single modality. Similarly to the tetanus dataset, the feature fusion, i.e. concatenation, of the CNN features validated with SVM (FC-SVM) increases the recall performance of each severity level by an average of 6.3%. As expected, the full pipeline of the proposed approach, i.e. combination of SVM-based on-site triage followed by an LSTM-based longitudinal model, has significantly increased the recall values of Mild and Severe levels to 83% and 75%, respectively.



(a) HFMD - LR

(b) HFMD - SVM

(c) Tetanus - LR

(d) Tetanus - SVM

Fig. 7. Comparison of two different window durations, i.e. 1-min vs. 5-min, on the F-score (%) of severity-level classification in (a) HFMD and (b) tetanus datasets.

D. Window Duration

So far, we have employed a 5-minutes window duration as recommended in a clinical baseline method [8]. To assess the robustness of our method to window size, we also experimented with much shorter window lengths, i.e. 1-minute. This would help to provide much more frequent inference of ANSD severity levels for both tetanus and HFMD patients. Fig. 7 shows results of both LR and SVM for the on-site triage process on HFMD and tetanus patients. On the HFMD dataset, having a shorter window duration helps to increase the number of training samples and hence improves performance for the majority of the methods. This has been replicated with both LR- and SVM-based classifiers (see Fig. 7(a) and (b)). However, similar findings are not observed on the tetanus dataset (see Fig. 7(c) and (d)), which suggests the difficulty associated with encoding discriminative variability in short duration signals.

TABLE VI
MILD AND SEVERE LEVEL CLASSIFICATION OF TETANUS PATIENTS USING MOBILENET ARCHITECTURE

	Transfer learning (%)				
	A	P	R	S	F_1
PPG-SVM	82.0	87.9	83.6	79.2	85.7
ECG-SVM	89.0	86.3	98.5	71.7	92.0
	Transfer learning + Feature fusion (%)				
	A	P	R	S	F_1
FC-SVM	88.8	88.6	94.9	77.9	91.6
	Feature fusion + Longitudinal model (%)				
	A	P	R	S	F_1
FC-SVM-LSTM	86.0	93.3	83.4	90.3	88.0
	Decision fusion + Longitudinal model (%)				
	A	P	R	S	F_1
SVM-DC-LSTM	90.5	95.0	89.6	92.1	92.0
SVM-DA-LSTM	91.2	94.5	91.1	91.4	92.7

TABLE VII
MILD AND SEVERE LEVEL CLASSIFICATION OF TETANUS PATIENTS USING MNASNET ARCHITECTURE

	Transfer learning (%)				
	A	P	R	S	F_1
PPG-SVM	84.4	91.0	84.1	85.0	87.4
ECG-SVM	89.1	88.0	96.3	76.1	92.0
	Transfer learning + Feature fusion (%)				
	A	P	R	S	F_1
FC-SVM	89.9	91.6	92.9	84.5	92.2
	Feature fusion + Longitudinal model (%)				
	A	P	R	S	F_1
FC-SVM-LSTM	89.1	95.3	86.6	93.1	90.7
	Decision fusion + Longitudinal model (%)				
	A	P	R	S	F_1
SVM-DC-LSTM	93.7	95.6	94.3	92.8	94.9
SVM-DA-LSTM	94.2	95.7	95.0	92.9	95.3

TABLE VIII
ANSD LEVEL CLASSIFICATION OF HFMD PATIENTS USING MOBILENET ARCHITECTURE

	Transfer learning (%)				
	A	P	R	S	F_1
ECG-1-SVM	71.3	57.8	57.0	78.5	57.2
ECG-2-SVM	71.6	58.0	57.3	78.7	57.6
	Transfer learning + Feature fusion (%)				
	A	P	R	S	F_1
FC-SVM	71.3	57.8	57.0	78.5	57.2
	Feature fusion + Longitudinal model (%)				
	A	P	R	S	F_1
FC-SVM-LSTM	75.1	62.6	62.7	81.3	62.6
	Decision fusion + Longitudinal model (%)				
	A	P	R	S	F_1
SVM-DC-LSTM	74.7	61.9	62.0	81.0	61.9
SVM-DA-LSTM	75.7	63.4	63.5	81.7	63.3

TABLE IX
ANSD LEVEL CLASSIFICATION OF HFMD PATIENTS USING MNASNET ARCHITECTURE

	Transfer learning (%)				
	A	P	R	S	F_1
ECG-1-SVM	71.3	58.0	56.9	78.4	57.1
ECG-2-SVM	71.7	59.4	57.4	78.7	57.8
	Transfer learning + Feature fusion (%)				
	A	P	R	S	F_1
FC-SVM	74.5	63.4	61.7	80.8	62.1
	Feature fusion + Longitudinal model (%)				
	A	P	R	S	F_1
FC-SVM-LSTM	80.7	71.8	70.9	85.4	71.1
	Decision fusion + Longitudinal model (%)				
	A	P	R	S	F_1
SVM-DC-LSTM	79.7	70.6	69.5	84.7	69.7
SVM-DA-LSTM	78.7	68.8	68.1	84.0	68.3

E. Hidden Layer Size

Finally, we also experimented with different sizes of the single hidden layer in the LSTM network. We validated the choice of 64 and 128 neurons for both tetanus and HFMD patients and we found out these two layer sizes do not cause significant performance changes across both datasets.

F. Mobile Architectures

We presented the benefit of cross-domain transfer learning for pre-screening of tetanus and HFMD patients using Inception architecture (as a proof-of-concept) on time-series physiological waveforms. Though Inception is known to be robust, it could be computationally more intensive compared to other light architectures, such as MobileNet [31]. To this end, we selected two other networks, MobileNet [31] and MnasNet [32], to demonstrate that the proposed approach generalises across other deep learning architectures. Thus, feature extraction from the hidden layers of MobileNet and MnasNet resulted in feature vectors of 1,280 and 1,056 units long, respectively, both of which are significantly shorter than that of Inception's 2,048 units long feature vector. The experiments are conducted using SVM (with Gaussian kernel) for on-site triage process with the LSTM hidden layer size of 64 neurons with 5-minutes window duration. The results of MobileNet and MnasNet on tetanus dataset are shown in Tables VI and VII, whereas the results on HFMD dataset are shown in Tables VIII and IX, respectively.

The results showed that MobileNet and MnasNet achieved encouraging classification performance on both tetanus and HFMD datasets. Compared to the SVM-DA-LSTM's F_1 score of 97.6% using Inception architecture on tetanus dataset, MobileNet and MnasNet achieved 92.7% and 95.3%, respectively. Similarly, on the HFMD dataset, MobileNet and MnasNet achieved F_1 scores of 63.3% and 68.3%, respectively, compared to the Inception-based SVM-DA-LSTM's F_1 score of 63.2%, which means MobileNet performed competitively with Inception architecture whilst MnasNet even outperformed it. The results also showed that MnasNet outperformed MobileNet consistently across both tetanus and HFMD datasets expectedly. Generally, we demonstrated that the proposed framework can also be applied across a variety of existing computer vision networks beyond Inception. Encouraging results are achieved using MobileNet and MnasNet with smaller feature dimensions, showing a promising potential to implement the proposed framework in low-cost mobile devices.

VII. CONCLUSION

We presented our proof-of-principle study that applies multi-modal and multi-layer decision making to triage patients with infectious diseases (tetanus and HFMD) using low-cost and unobtrusive wearable sensors that collect artefact-prone physiological data. ANSD is the main cause of death in both tetanus and HFMD, and it is not often apparent until late stage manifestations. Hence, early and automatic diagnosis of its severity level could be used for timely intervention.

We employed spectrogram representations of ECG and PPG waveforms that enable us to exploit existing pre-trained networks, i.e. cross-domain transfer learning. This allows us to avoid training from scratch and hence reduces the requirement for large training data and high computational resource. Later, feature fusion is applied to improve the performance of the on-site triage process, followed by personalised longitudinal modelling that infers the ANSD severity level taking into consideration previous patient states. Thus, the proposed approach would provide efficient hospital resource utilisation in low-resource clinical-settings of the developing world, which could in turn help improve overall patient care. Our approach was validated with three existing networks (Inception, MobileNet and MnasNet) on two independent datasets (tetanus and HFMD) collected from patients in Southern Vietnam, and we achieved significant performance improvement over existing methods. Future avenues of research would concern investigating the impact of incorporating an attention model into the LSTM, as the inclusion of such a model has been shown to greatly improve performances of the LSTM in other application domains.

Acknowledgment

We would like to thank Dr Nguyen Van Vinh Chau, Director of the Hospital for Tropical Diseases, staff in ICU at the Hospital for Tropical Diseases in addition to Drs Nguyen Thanh Hung, Troung Huu Khanh and Pham Van Quang and staff at Children's Hospital Number 1, Ho Chi Minh City.

References

[1] D. B. Thuy *et al.*, "Tetanus in Southern Vietnam: Current situation," *Amer. J. Tropical Med. Hygiene*, vol. 96, no. 1, pp. 93–96, 2017.

[2] V. L. Feigin *et al.*, "Global, regional, and national burden of neurological disorders during 1990–2015: A systematic analysis for the global burden of disease study 2015," *Lancet Neurol.*, vol. 16, no. 11, pp. 877–897, 2017.

[3] P. K. Lam *et al.*, "Prognosis of neonatal tetanus in the modern management era: An observational study in 107 Vietnamese infants," *Int. J. Infectious Diseases*, vol. 33, pp. 7–11, 2015.

[4] H. T. Trieu *et al.*, "Neonatal tetanus in Vietnam: Comprehensive intensive care support improves mortality," *J. Pediatric Infectious Diseases Soc.*, vol. 5, no. 2, pp. 227–230, 2015.

[5] N. T. Le Nguyen Thanh *et al.*, "Severe enterovirus a71 associated hand, foot and mouth disease, Vietnam, 2018: preliminary report of an impending outbreak," *Eurosurveillance*, vol. 23, no. 46, 2018.

[6] T. Y. Lin, S. J. Twu, M. S. Ho, L. Y. Chang, and C. Y. Lee, "Enterovirus 71 outbreaks, Taiwan: occurrence and recognition," *Emerg. Infectious Diseases*, vol. 9, no. 3, pp. 291–3, 2003.

[7] N. J. Schmidt, E. H. Lennette, and H. H. Ho, "An apparently new enterovirus isolated from patients with disease of the central nervous system," *J. Infectious Diseases*, vol. 129, no. 3, pp. 304–309, 1974.

[8] M. Malik *et al.*, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, 1996.

[9] V. Jeyhani, S. Mahdiani, M. Peltokangas, and A. Vehkaoja, "Comparison of hrv parameters derived from photoplethysmography and electrocardiography signals," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 5952–5955.

[10] G. A. Tadesse *et al.*, "Severity detection tool for patients with infectious disease," *IET Healthcare Technol. Letters*, 2020.

[11] C. B. Delahunt *et al.*, "Automated microscopy and machine learning for expert-level malaria field diagnosis," in *Proc. IEEE Global Humanitarian Technol. Conf.*, 2015, pp. 393–399.

[12] Z. Liang *et al.*, "CNN-based image analysis for malaria diagnosis," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2016, pp. 493–496.

[13] T. Desautels *et al.*, "Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach," *JMIR Med. Inform.*, vol. 4, no. 3, 2016, Art. no. e28.

[14] R. A. Taylor *et al.*, "Prediction of in-hospital mortality in emergency department patients with sepsis: A local big data–driven, machine learning approach," *Academic Emergency Med.*, vol. 23, no. 3, pp. 269–278, 2016.

[15] B. Zhang *et al.*, "Machine learning algorithms for risk prediction of severe hand-foot-mouth disease in children," *Nature Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 5368.

[16] G. Liu *et al.*, "Developing a machine learning system for identification of severe hand, foot, and mouth disease from electronic medical record data," *Nature Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 16341.

[17] D. K. Das, M. Ghosh, M. Pal, A. K. Maiti, and C. Chakraborty, "Machine learning approach for automated screening of malaria parasite using light microscopic images," *Micron*, vol. 45, pp. 97–106, 2013.

[18] H. T. H. Duong *et al.*, "Heart rate variability as an indicator of autonomic nervous system disturbance in tetanus," *Am. J. Tropical Medicine Hygiene*, 2019.

[19] H. Albert *et al.*, "Performance of three led-based fluorescence microscopy systems for detection of tuberculosis in uganda," *PLoS One*, vol. 5, no. 12, 2010, Art. no. e15206.

[20] M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, and G. Thoma, "Image analysis and machine learning for detecting malaria," *Transl. Res.*, vol. 194, pp. 36–55, 2018.

[21] H. Yin and N. K. Jha, "A health decision support system for disease diagnosis based on wearable medical sensors and machine learning ensembles," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 3, no. 4, pp. 228–241, Oct.–Dec. 2017.

[22] G. A. Tadesse *et al.*, "Cardiovascular disease diagnosis using cross-domain transfer learning," in *Proc. 41st Annu. Int. Conf IEEE Eng. Med. Biol. Soc.*, Jul. 2019, pp. 4262–4265.

[23] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.

[24] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019.

[25] M. Valstar *et al.*, "AVEC 2013: The continuous audio/visual emotion and depression recognition challenge," in *Proc. 3rd ACM Int. Workshop Audio/Visual Emotion Challenge*, 2013, pp. 3–10.

[26] J. C. Castillo, D. Carneiro, J. Serrano-Cuerda, P. Novais, A. Fernández-Caballero, and J. Neves, "A multi-modal approach for activity classification and fall detection," *Int. J. Syst. Sci.*, vol. 45, no. 4, pp. 810–824, 2014.

[27] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, 2015, pp. 1–9.

[28] G. Abebe and A. Cavallaro, "Inertial-vision: Cross-domain knowledge transfer for wearable sensors," in *Proc. IEEE Int. Conf. Comput. Vision Workshops*, Oct. 2017, pp. 1392–1400.

[29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2009, pp. 248–255.

[30] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *NIPS Workshop Deep Learn.*, Dec. 2014.

[31] M. Sandler *et al.*, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018.

[32] M. Tan *et al.*, "Mnasnet: Platform-aware neural architecture search for mobile," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 2820–2828.