Conditional Neural ODE Processes for Individual Disease Progression Forecasting: A Case Study on COVID-19

Ting Dang
td464@cam.ac.uk
University of Cambridge
Cambridge, UK

Jing Han[+]
University of Cambridge
Cambridge, UK

Tong Xia[+]
University of Cambridge
Cambridge, UK

Erika Bondareva[†]
University of Cambridge
Cambridge, UK

Chloë Siegele-Brown[†]
University of Southampton
Southampton, UK

Jagmohan Chauhan[†]
University of Southampton
Southampton, UK

Andreas Grammenos[†]
University of Cambridge
Cambridge, UK

Dimitris Spathis[†]
University of Cambridge
Cambridge, UK

Pietro Cicuta
University of Cambridge
Cambridge, UK

Cecilia Mascolo
University of Cambridge
Cambridge, UK

ABSTRACT

Time series forecasting, as one of the fundamental machine learning areas, has attracted tremendous attentions over recent years. The solutions have evolved from statistical machine learning (ML) methods to deep learning techniques. One emerging sub-field of time series forecasting is *individual disease progression forecasting*, e.g., predicting individuals' disease development over a few days (e.g., deteriorating trends, recovery speed) based on few past observations. Despite the promises in the existing ML techniques, a variety of unique challenges emerge for disease progression forecasting, such as irregularly-sampled time series, data sparsity, and individual heterogeneity in disease progression. To tackle these challenges, we propose novel Conditional Neural Ordinary Differential Equations Processes (CNDPs), and validate it in a COVID-19 disease progression forecasting task using audio data. CNDPs allow for irregularly-sampled time series modelling, enable accurate forecasting with sparse past observations, and achieve individual-level progression forecasting. CNDPs show strong performance with an Unweighted Average Recall (UAR) of 78.1%, outperforming a variety of commonly used Recurrent Neural Networks based models. With the proposed label-enhancing mechanism (i.e., including the initial health status as input) and the customised individual-level loss, CNDPs further boost the performance reaching a UAR of 93.6%. Additional analysis also reveals the model's capability in tracking individual-specific recovery trend, implying the potential usage of the model for remote disease progression monitoring. In general, CNDPs pave new pathways for time series forecasting, and provide considerable advantages for disease progression monitoring.

KDD '23, August 6–10, 2023, Long Beach, CA, USA
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0103-0/23/08.
https://doi.org/10.1145/3580305.3599792

CCS CONCEPTS

• **Computer methodologies** → **Machine learning**; • **Applied computing** → *Computers in other domains*; • **Human centered computing** → Ubiquitous and mobile computing.

KEYWORDS

time series forecasting, disease progression, audio and signal processing, COVID-19, neural ODE

ACM Reference Format:
Ting Dang, Jing Han[+], Tong Xia[+], Erika Bondareva[†], Chloë Siegele-Brown[†], Jagmohan Chauhan[†], Andreas Grammenos[†], Dimitris Spathis[†], Pietro Cicuta, and Cecilia Mascolo. 2023. Conditional Neural ODE Processes for Individual Disease Progression Forecasting: A Case Study on COVID-19. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23), August 6–10, 2023, Long Beach, CA, USA.* ACM, Long Beach, CA, USA, 12 pages. https://doi.org/10.1145/3580305.3599792

1 INTRODUCTION

Time series forecasting is of particular interest in a diverse range of applications, e.g., weather forecasting [34], COVID-19 upcoming number of infection forecasting [32], etc. While traditional methods focused on statistical machine learning techniques such as autoregressive (AR) [2], exponential smoothing [15] and state space and structural models [12, 9, 18], recently, attention has been paid to modern machine learning techniques [25], especially deep learning such as Recurrent Neural Networks (RNNs) and its variants [46, 26, 38].

[+] These authors contributed equally to this work
[†] These authors contributed equally to this work

One emerging sub-field of time series forecasting that attracted tremendous attention is forecasting individuals' disease progression, which can help doctors and physicians make better decisions and reduce the burden on healthcare systems. Especially with the outbreak of COVID-19, developing such a model to forecast pandemic progression has become pivotal. Despite the success of existing methods for time series forecasting, a multitude of unique challenges emerged in disease progression monitoring.

Firstly, it is hard to obtain individuals' health status (i.e., samples or diagnosis results) with time regularity, leading to irregularly-sampled time series which are difficult to model using traditional methods such as AR models or RNNs. Secondly, individuals may also not visit the clinics or provide their data often, resulting in data sparsity. The lack of historical samples for forecasting presents an additional challenge. Thirdly, individuals' disease progression varies greatly, therefore, models should be carefully designed to address this heterogeneity.

To tackle these challenges, we propose novel Conditional Neural Ordinary Differential Equations Processes (CNDPs) for time series forecasting, and validate it in a COVID-19 disease progression forecasting task. Here we focused on audio signals (e.g., cough, speech, breathing) for COVID-19 progression forecasting, due to its numerous advantages in flexible and scalable data collection scheme, as well as extensive evidence of its potential for COVID-19 detection (i.e., distinguishing positive and negative COVID-19 audio samples) [17, 3, 7, 45].

The proposed CNDPs are motivated by a recent approach of Neural Ordinary Differential Equations Processes (NDPs) [27]. NDPs composed of the Ordinary Differential Equation (ODE) are capable of modelling a continuous and dynamic process. Further, ODE can deal with any irregularly-sampled time series, serving as a good fit to model audio time series. However, NDPs are designed for single time series forecasting, i.e., forecasting future $y(t)$ via modelling its past samples. This is less feasible when $y(t)$ is hard or costly to measure (e.g., COVID-19 PCR test result). Therefore, CNDPs first introduce a conditional variable $\mathbf{X}(t)$ which is closely associated with $y(t)$ but easy to measure and quantify. CNDPs model the dynamics of $\mathbf{X}(t)$ and use it implicitly to aid the forecasting of $y(t)$. Furthermore, CNDPs introduce two additional modelling mechanisms to tackle disease progression forecasting specifically. One is a label-enhancing mechanism that includes the initial infection status as an additional input to CNDPs, providing accurate past infection information to aid the forecasting. The other is the individual-level disease progression loss function that accounts for the differences in individuals' heterogeneity. The key contributions of this work are summarised as follows:

- Novel CNDPs are proposed, which enable reliable modelling of irregularly-sampled and sparse time series, for disease progression forecasting.
- A validation of the proposed CNDPs using a COVID-19 sound data set, consisting of 212 participants and 3714 audio samples. It outperforms a variety of RNNs-based models and Transformers and yields the best performance, with a UAR of 93.6%, a sensitivity of 90.6%, and a specificity of 96.7%.
- In-depth analysis further demonstrates that CNDPs are effective in a longer forecasting horizon and the individual-specific recovery rate prediction, suggesting its potential in the remote and long-term monitoring of individuals' different disease progression.
- To the best of the authors' knowledge, this is the first study investigating COVID-19 disease progression forecasting using audio signals.

2 RELATED WORK

2.1 Time Series Forecasting

Autoregressive (AR) models [2] and Recurrent Neural Networks (RNNs) such as Gated Recurrent Units (GRUs) [4] are the two most commonly adopted time series forecasting models. They have been employed for a variety of tasks and showed strong performance, including weather forecasting [33], COVID-19 infection cases forecasting [29], traffic flow forecasting [14, 47], etc. However, AR models are linear models that cannot accommodate complex nonlinear dynamics, such as unknown disease progression processes, and RNNs are discrete models that may not be optimal for continuous time series modelling, and also require massive data to develop the model. Though different strategies for RNNs have been proposed to process irregularly-sampled data, they manually overlay additional mechanisms such as input augmentation and time delay factors instead of explicitly accommodating the irregularity in the time series [43]. Moreover, RNNs are only good at short-term forecasting with reliable predictions at a few time steps ahead but do not perform well in long-range forecasting [47]. Transformers have been one of the recent solutions for time series modelling, but it requires a large amount of training data and computation resources, less applicable for limited health data [48].

Recently, NODEs were proposed in [5, 31], describing a dynamical system by an ordinary differential equation (ODE) with the governing function parameterized by a neural network. It allows for modelling of irregularly-sampled time series, and shows promises in different applications [30, 42, 28, 20, 6, 1]. However, they also require a fair amount of data to develop the model, which may fail in modelling sparse time series commonly existing in clinical data.

A new family of stochastic processes, NDPs, has been proposed in [27] to achieve reliable forecasting with just a few data points. However, it has only been validated on rotating MNIST dataset, and has not been investigated in real-world applications. Both NODEs and NDPs mainly focus on modelling dynamics of a single time series, and further lack the capability to deal with individual heterogeneity. The proposed CNDPs, instead, are able to model the disease progression implicitly via audio sequence and can forecast individuals' disease progression accurately with limited information.

2.2 Audio-based COVID-19 Detection

A number of studies have explored audio signals for COVID-19 detection and shown promise. Different sound types have been explored [10, 22, 3], and a variety of deep learning techniques or training strategies have also been analysed, such as ResNet [7], self-supervised learning[45], etc. However, they are designed to detect COVID-19 status given only the current test sample, which ignores the long-term disease progression and is not applicable for forecasting.

A most recent study [8] models the temporal dynamics of past and current audio sequences to aid COVID-19 detection. Although both past and current audio information were explored in this work, the target was still to infer the current COVID-19 infection status, and it is not capable of multi-step forecasting for disease progression ahead of time. Specifically, study [8] predicts a single point estimation at a time, whereas our work deals with trajectory forecasting, which is theoretically a more challenging and clinically important task.

2.3 Disease Progression Modeling

Modelling and predicting the progression of diseases is of great importance for healthcare as it enables early intervention and timely personalised treatment [41]. Disease progression modelling is basically a special case in time series modelling. However, it is more changeling than other problems regarding the health data sparsity and irregularity issues, as discussed in Sec. 1. Some preliminary efforts have been put into modelling chronic disease progressions like Alzheimer's and Parkinson's diseases from clinical images [24, 13, 21]. For example, *Zhou* et al. proposed a multi-task learning-based framework to predict the dementia stage from a sequence of brain image biomarkers. In [16], a recurrent neural network (RNN) based method was deployed on magnetic resonance imaging (MRI) biomarkers to predict dementia degrees at each time point. A recent study [36] investigated NODEs for COVID-19 forecasting, but it focuses on a different applications, e.g., forecasting future disease caseload (infections deaths) at the population level.

Our study is different from those existing works in three aspects, ranging from the application, task definition, and the technical aspects. Firstly, from the application perspective, our work focuses on forecasting individuals' disease progression at the user level, which sets it apart from other studies that primarily examine future disease caseload or address different health issues. Secondly, from the task perspective, our work involves forecasting individual disease progression based on the user's past audio samples, incorporating two series of data. This differs from most forecasting problems that typically involve a single time series. The use of audio as a new data modality in health applications allows for more cost-effective sensing through mobile devices, enabling the monitoring of disease progression in out-of-hospital environments. Thirdly, from a technical perspective, existing research has predominantly explored deterministic models such as conditional latent ODE, whereas our work utilizes a stochastic process with our proposed conditional NDPs. This approach proves advantageous in handling the challenges posed by health data sparsity and irregularity, outperforming commonly used time series models like RNNs.

3 METHODS

3.1 Problem Statement

We aim to study the COVID-19 disease progression forecasting using audio samples as illustrated in Figure 1. Specifically, it aims to forecast the sequence of each individual's COVID-19 status $y_i, i \in [t_{n+1}, t_N]$ simultaneously given only the past audio representations and the corresponding time $\{\mathbf{x}_i, t_i\}, i \in [t_0, t_n]$, referred to as CNDPs. We refer to the past longitudinal audio representations within $[t_0, t_n]$ as context vectors, and the future samples within $[t_{n+1}, t_N]$ as target vectors. We mainly analyse the case where the number of context vectors is limited to validate the model capability in forecasting with extremely limited information. It is worth noting COVID-19 is one of the potential applications, but the proposed CNDPs can generalise to any disease progression forecasting scenario.

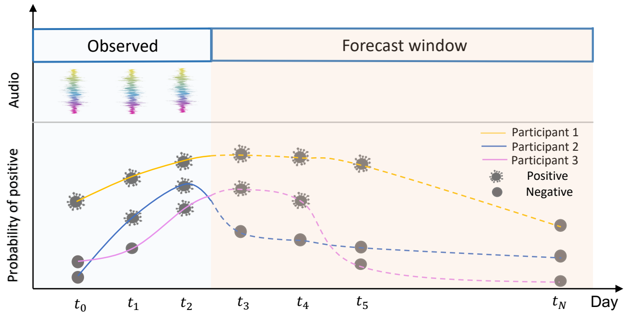


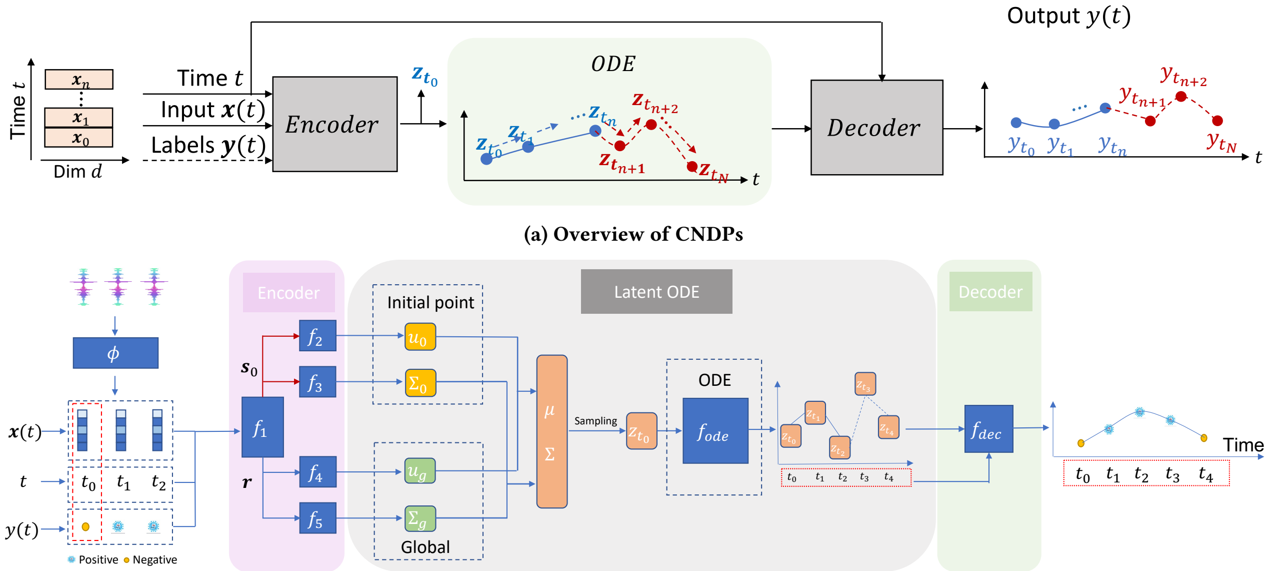
Figure 1: Forecasting model pipeline. The model aims to forecast individuals' COVID-19 progression, e.g., predicting individuals' disease development over a few days (e.g., deteriorating trends, recovery speed) by modelling a few past longitudinal audio samples. A high probability prediction indicates positive, and low probability indicates negative (healthy).

3.2 Proposed CNDPs

3.2.1 Overview. The overview of the proposed method is shown in Figure 2a. It adopts a variational encoder-decoder structure. CNDPs consist of three processing stages, namely, an encoder, a latent ODE and a decoder. The encoder concatenates the past audio representations $\mathbf{x}(t) = [\mathbf{x}_{t_0}, \mathbf{x}_{t_1}, \cdots, \mathbf{x}_{t_n}]$ and the corresponding time $t = [t_0, t_1, \cdots, t_n]$ as input $\{\mathbf{x}_{t_n}, t_n\}$, which adds an additional time dimension to the features and can be denoted as $[[x_{t_0}, t_0], [x_{t_1}, t_1], \cdots, [x_{t_n}, t_n]]$. They are then mapped to a latent representation $\mathbf{z}_{t_0}$, which captures the global dynamics in the past disease progression. The additional proposed label-enhancing mechanism (cf. Section 3.3) further includes the past labels, $y(t) = [y_{t_0}, y_{t_1}, \cdots, y_{t_n}]$, and use $\{\mathbf{x}_{t_n}, t_n, y_{t_n}\}$ as the input. Latent ODE following the encoder produces a trajectory $\mathbf{z}(t) = [\mathbf{z}_{t_0}, \mathbf{z}_{t_1}, \cdots, \mathbf{z}_{t_N}]$ for all desired time steps $[t_0, t_N]$, $t_N \geq t_n$, by solving an ODE initial value problem with $\mathbf{z}_{t_0}$ served as the initial point. A decoder finally maps the trajectory $\mathbf{z}(t)$ to the disease progression $y(t), t \in [t_0, t_N]$.

A detailed model structure is given in Figure 2b. Features $\mathbf{x}(t)$ are first extracted from raw audio waveform by the network ϕ, serving as the input to CDNPs. Each component in CDNPs is discussed in the following sections.

3.2.2 Encoder. The encoder first transforms the context vectors $\{\mathbf{x}_i, i\}, i \in [t_0, t_n]$ to a latent variable $N(\mu, \Sigma)$ by sub-networks f_1 to f_5 (Figure 2b), and a latent vector $\mathbf{z}_{t_0}$ is randomly sampled from this distribution and used for the following latent ODE. The latent variable $N(\mu, \Sigma)$ consists of two parts, with the first part learning representations from the initial context vector at $t_0 = 0$ denoted as $N(\mathbf{u}_0, \Sigma_0)$ and the second part learning the global



(a) Overview of CNDPs

(b) Detailed model structure

Figure 2: (a) Overview of CNDPs, consisting of an encoder, a latent ODE and a decoder. The encoder maps the input of audio representations $\mathrm{x}(t)$, corresponding time t, and test results $y(t)$ within $[t_0, t_n]$ to an initial point z_{t_0} and evaluate a sequence $\mathrm{z}(t), t \in [t_0, t_N]$ in the latent space using an ODE. The output $y(t), t \in [t_0, t_N]$ is mapped from $\mathrm{z}(t)$ through a decoder. Further, the label-enhancing mechanism takes past label sequence $y(t) = [y_{t_0}, \cdots, y_{t_n}]$ as an additional input (the dashed line), providing accurate past disease progression information to the model. (b) A detailed CNDPs model structure for disease progression forecasting. ϕ represents feature extraction, and f_* represents network structures in CNDPs. The output at each time step is predicted as a Bernoulli distribution, representing the probability of positive test results. The time series within $[t_0, t_5]$ is an example, but CNDPs can process any irregular-sampled time series.

information from the entire context vector sequence denoted as $N(\mathbf{u}_g, \Sigma_g)$. Specifically, the first part $N(\mathbf{u}_0, \Sigma_0)$ is computed as:

$$\begin{aligned} \mathbf{s}_0 &= f_1(\{\mathbf{x}_{t_0}, t_0\}), t_0 = 0 \\ \mathbf{u}_0 &= f_2(\mathbf{s}_0); \Sigma_0 = f_3(\mathbf{s}_0) \end{aligned} \tag{1}$$

where $\mathbf{s}_0$ is the intermedia representation at $t_0 = 0$ learnt from the concatenation of $\{\mathbf{x}_{t_0}, t_0\}$. $N(\mathbf{u}_0, \Sigma_0)$ is particularly learnt to enforce a more accurate representation of $\mathbf{z}_{t_0}$, providing important information for individuals' initial infection status, and highly determining how the disease progression $\mathbf{z}(t)$ is decoded using ODE.

The second part $N(\mathbf{u}_g, \Sigma_g)$ learns the representation of the global information of the context vectors within $[t_0, t_n]$ as:

$$\begin{aligned} \mathbf{r} &= \frac{1}{n+1} \sum_{i=t_0}^{t_n} f_1(\{\mathbf{x}_i, i\}), i \in [t_0, t_n] \\ \mathbf{u}_g &= f_4(\mathbf{r}); \Sigma_g = f_5(\mathbf{r}) \end{aligned} \tag{2}$$

Similarly, $\mathbf{r}$ is the intermedia representation of global information, averaged over all time steps. The final latent variable $N(\mu, \Sigma)$ can be viewed as a concatenation of $N(\mathbf{u}_0, \Sigma_0)$ and $N(\mathbf{u}_g, \Sigma_g)$, and a sample $\mathbf{z}_{t_0}$ can be randomly selected as:

$$\mathbf{z}_{t_0} \sim N(\mu, \Sigma) = N(\begin{bmatrix} \mu_0 \\ \mu_g \end{bmatrix}, \begin{bmatrix} \Sigma_0, 0 \\ 0, \Sigma_g \end{bmatrix}) \tag{3}$$

Different from NDPs, CNDPs learn $N(\mu, \Sigma)$ and obtain $\mathbf{z}_{t_0}$ via modelling conditional variable $\mathbf{X}(t)$, i.e., audio sequence, instead of directly modelling the target variable $y(t)$. Our approach adopts variational inference (VI) to sample $\mathbf{z}_{t_0}$ from the learnt posterior distribution instead of using a deterministic variable, as VI generally yields robustness to overfitting, particularly for small datasets as in our case. Moreover, using such a stochastic process can quickly adapt to new data points [27]. This allows for predicting future values based on a limited number of past samples.

3.2.3 *Latent ODEs.*

A latent ODE model describes a continuous process by an ODE in the latent space as:

$$\frac{d\mathbf{z}(t)}{dt} = f_{ode}(\mathbf{z}(t), t; \theta) \tag{4}$$

where f_{ode} is the neural network used to approximate the governing function in the ODEs, parameterised by the weights θ. Instead of learning $\mathbf{z}(t)$ directly, latent ODE learns the dynamics $\frac{d\mathbf{z}(t)}{dt}$ that are governed by an ODE. If the temporal dynamics of audio and disease progression do not change significantly, this approach explicitly learning the dynamics may be a simpler task, and can potentially show powerful generalisation capabilities.

To obtain the sequence $\mathbf{z}(t)$, it requires solving an ODE initial value problem as:

$$\mathbf{z}_{t_0}, \cdots, \mathbf{z}_{t_N} = ODESolve(f_{ode}, \theta, \mathbf{z}_{t_0}, t_0, \cdots, t_N) \tag{5}$$

A Runge-Kunta method is used to solve the ODE [11]. One of the advantages of ODE manifests in long-range forecasting, while most of the existing time series models show significant deterioration after a short period of forecasting.

3.2.4 Decoder. The decoder transforms the sequence $\mathbf{z}(t)$ to test labels $y(t)$. The input to the decoder is a concatenation of latent vectors and their corresponding time steps $\{\mathbf{z}_i, i\}, i \in [t_0, t_N]$. The decoder maps $\{\mathbf{z}_i, i\}$ to a Bernoulli distribution $p(y_i = 1)$ at each time step i as:

$$p_i = p(y_i = 1) = f_{dec}([\mathbf{z}_i, i]), i \in [t_0, t_N] \tag{6}$$

where $y_i = 1$ represents positive and $y_i = 0$ represents negative. Further, the predicted probability $p(y_i = 1)$ at each time step can be converted to a binary output of positive or negative, with $p_i \geq 0.5$ as positive and $p_i < 0.5$ as negative.

3.3 Label-enhancing Mechanism

Additionally, a label-enhancing mechanism is proposed, by including the past infection status as the input to CNDPs (shown with dash line in Figure 2a), referred to as CNDP$^{\hat{y}}$. This is practical as individuals generally know their past health conditions or infection status. Specifically, the input to the encoder of CNDPs$^{\hat{y}}$ is a concatenation of the past audio representations, the time and, additionally, the past labels $\{\mathbf{x}_i, t_i, y_i\}, i \in [t_0, t_n]$. The label-enhancing CNDPs$^{\hat{y}}$ learn the underlying dynamics of the audio and labels simultaneously in a joint space, with the past infection status serving as an accurate reference to aid the forecasting.

3.4 Individual-level Loss

The model is trained using an amortised variational inference procedure, which jointly optimises the feature extractor ϕ, the encoder, the latent ODE, and the decoder. The typical loss in NDPs consists of Cross Entropy (CE) and KL divergence (shown in Appendix E), which mainly targets point estimation for classification problems and ignores the temporal dynamics. Moreover, it is computed over the entire dataset or batches, which does not consider individual heterogeneity. To guarantee reliable forecasting for each individual, we proposed a customized loss that considers the disease progression over time and also optimized for each individual, defined as:

$$\begin{aligned}\mathcal{L}(\theta,\phi) = \frac{1}{J}\sum_{j}[-\Gamma_{pb}(p^{j}_{\theta}(y_{\mathbb{T}}|\mathbf{z}_{t_0}, t, \theta), \hat{y}_{\mathbb{T}})+ \\ + D_{KL}(q^{j}_{\phi}(\mathbf{z}_{t_0}|\mathbf{x}_{\mathbb{C}}, t_{\mathbb{C}}, y_{\mathbb{C}})||q^{j}_{\phi}(\mathbf{z}_{t_0}|\mathbf{x}_{\mathbb{T}}, t_{\mathbb{T}}, y_{\mathbb{T}}))]\end{aligned} \tag{7}$$

where j represents each individual, and the final loss is an average across all individuals J. It consists of the negative point-biserial correlation $-\Gamma_{pb}(*)$ and the KL divergence D_{KL} between the posterior distributions. Specifically, $\Gamma_{pb}(*)$ computes the correlation between the predicted probability $p^{j}_{\theta}(y_{\mathbb{T}}|\mathbf{z}_{t_0}, t, \theta)$ and the test labels $\hat{y}_{\mathbb{T}}$. The smaller the -$\Gamma_{pb}(*)$, the better the predicted disease progression matching the test labels. KL divergence D_{KL} captures the difference between the posterior distribution $q_{\phi}(*)$ learnt for the context vectors $*_{\mathbb{C}}$ and target vectors $*_{\mathbb{T}}$ (cf. section 3.1), and $q_{\phi}(*)$ is a Gaussian distribution as commonly used in VI. In general, $\Gamma_{pb}(*)$ enhances the learning of disease progression for each individual.

3.5 Learning and Inference

During the training phase, the model takes the entire audio sequence of each individual. Using the entire sequence guarantees the model to observe the complete disease progression process. Instead of training the model to forecast future values given limited past samples, the model is trained on interpolation tasks, where a subset of the audio samples in the sequence are randomly selected as context vectors and the entire sequence is used as the target vectors, following conventional choices in [3] and [4]. This improves the model's understanding of overall dynamics. We only use a small number of context vectors, as it can force the model to forecast accurately using limited past information. During the test phase, only the past samples are provided for forecasting purposes, whereas only the initial audio sample and initial label (in the label-enhancing setting) are provided and used to forecast future disease progression.

4 EXPERIMENTAL SETTINGS

4.1 Data

The dataset was collected via our mobile app, released on multiple platforms including Android, iOS and a webpage. The project and data collection were approved by the Ethics Board of the Department of Computer Science and Technology at the University of Cambridge. Each participant was encouraged to record three different sound types via smartphone built-in microphones, including breathing, coughing, and speech, where each participant was asked to read a short phrase displayed on the screen. The COVID-19 test results were self-reported, chosen from a positive test result, and a negative test result, as well as a separate option for users who had not been tested. 212 participants were selected with balanced positive and negative participants. Each participant reports 5 to 385 days of samples, covering 3714 days in total.

Missing data occurs when users record their audio data irregularly. Specifically, the app collects users' daily audio recordings and their test results. However, users may not always remember to record their audio every day, or users may occasionally record useless data which can introduce conflicts in the reported test results when compared to data from other days. To ensure data integrity, such recordings are manually removed from our dataset. These scenarios result in audio sequences with varying day intervals, giving rise to irregular time series within our data.

Age and gender are relatively balanced between positive and negative groups, with 110 female participants (55 positive and 55 negative), 90 male participants (49 positive and 41 negative), and 12 unknown. There are 142 participants aged between 30-59 (75 positive and 67 negative). The median reported duration of the 212 participants is of 35 days, and the median number of samples is 9. This time duration is able to cover the period of disease progression, and aligns with the reported disease progression duration [40, 44]. More details of data, preprocessing and feature extraction can be found in Appendix A.

4.2 Data Processing

Audio recordings were first resampled to 16kHz and converted to mono channel. Silence periods were removed at the beginning and the end of the recording. Normalisation was performed for the data to have a maximum amplitude of 1. Data augmentation was used to increase the number of samples, where Gaussian noise was added to the audio recording. The data was split into training, validation, and test partitions with 70%, 10% and 20% balanced positive and negative participants respectively, as well as the relatively balanced gender and age.

4.2.1 Data augmentation. As negative participants generally contribute more samples than positive participants, and positive participants report both positive and negative samples, this leads to the number of negative samples being significantly larger than that of positive samples. In order to relatively balance the number of positive and negative samples, noise augmentation is carried out three times for positive participants, but only one time for negative participants, resulting in 6862 samples in total. This can potentially balance the positive and negative classes as well as increase the data size for model developments.

4.2.2 Feature extraction. Transfer learning is applied for feature extraction, where a pre-trained network of VGGish [19] is used, as it is trained for audio event detection and can learn good acoustic feature representations. This can potentially help capture better audio representation via transfer learning. Three modalities including breathing, cough, and speech recordings were adopted. For each modality, spectrogram for each audio recording was first computed and passed to the VGGish to learn a 128-dimension feature embedding. The embeddings converted by VGGish from the three modalities were then concatenated to form a multi-modal input vector, leading to a final 384-dimensional feature embedding $\mathbf{x}(t)$ at each time step.

4.3 Baselines

We compared our proposed model to state-of-the-art systems for time series forecasting. To understand the impact of each component of the CNDPs, we compared a range of different models. First, without the label-enhancing mechanism, six systems were analysed:

- **RNN** Δt_1: a classic RNN based autoregressive model for one-step-ahead forecasting, i.e., forecasting $y_{t_{n+1}}$ given input as $[\mathbf{x}_{t_0}, \mathbf{x}_{t_1}, \cdots, \mathbf{x}_{t_n}]$. The time difference between consecutive days of audio recordings Δt is used as an additional input to model irregular-sampled time series.
- **RNN** Δt_{all}: a similar structure as RNN Δt_1, but for multi-step-ahead forecasting, i.e., forecasting $y(t) = [y_{t_{n+1}}, \cdots, y_{t_N}]$ given $[\mathbf{x}_{t_0}, \mathbf{x}_{t_1}, \cdots, \mathbf{x}_{t_n}]$.
- **RNN-VAE**: an RNN based encoder-decoder structure [31] for multi-step-ahead forecasting. The past longitudinal audio sample $[\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$ is used as the context vector to forecast the future COVID-19 test labels $y(t) = [y_{n+1}, \cdots, y_N]$.
- **Transformer**: a Transformer structure [39] with variations in decoder, where 2 Transformer encoder layers are used instead of the default 6, as our dataset size is relatively small. Due to the limited data size, dropout ratio was set to 0.9.
- **CNDPs (ours)**: proposed model with typical CE loss.
- **CDNPs$_l$(ours):** proposed model with proposed individual-level loss L.

Table 1: Optimized model hyperparameters

Systems	Latent	Lr	Loss weight	Decay
RNN Δt_1	50	1e-3	2.5	0.95
RNN Δt_{all}	50	1e-3	2	
RNN $\Delta t_{all}^{\hat{y}}$			1	
RNN-VAE	50/100	1e-3	1	
RNN-VAE$^{\hat{y}}$			1	
Transformer	2048	1e-5	2	0.98
Transformer$^{\hat{y}}$			2.4	
CNDPs$_l$	25	1e-4	1.5	0.95
CNDPs$_l^{\hat{y}}$				

Label-enhancing models were further studied by including additional input of past longitudinal labels $y(t) = [y_0, \cdots, y_{t_n}]$ for all multi-step-ahead forecasting systems, represented with $*^{\hat{y}}$ (e.g., CDNPs$_l{}^{\hat{y}}$). More details can be found in Appendix B.

4.4 Model Training

For the encoder in CNDPs (cf. Section 3.2), sub-networks f_1 to f_5 are all fully-connected (FC) layers. The function f_{ode} in the latent ODE is approximated using 3 FC layers with Tanh activation. Similar to f_1, the decoder also employs 3 FC layers with ReLU activation, followed by the output layer. A Gaussian prior is used for the variational inference, with a mean 0 and a standard deviation 0.01. Details of the optimized hyperparameters are shown in Table 1. More information can be found in Appendix C.

The maximum number of context vectors during the training phase is set to 5, which is randomly determined at each batch within the range of [1,5]. Only the initial audio sample is used as the context vector during the test phase. The initial label is also used in the label-enhancing setting. This is consistent across all the systems. All models were implemented by Pytorch and trained using one GPU with 64G memory. The code can be found in Github repository[1].

4.5 Evaluation Metrics

We evaluate two aspects of the systems for COVID-19 disease progression forecasting. The first one validates the forecasting performance in distinguishing positive and negative samples, as in conventional classification approaches. Unweighted Average Recall (UAR), sensitivity (also named true positive rate or recall), and specificity (also referred to as the true negative rate) are used. Sensitivity and specificity are metrics that evaluate the performance of a binary classifier for one group (i.e., positive or negative group) at a time, and there is a trade-off between the two metrics.

The second evaluates the system performance in tracking and forecasting each individual's disease progression over time. Therefore, point-biserial correlation coefficient γ_{pb} [37] between forecasted trajectory and test labels were computed. It is a special case of the Pearson correlation coefficient, and commonly used to measure

[1] Code: https://github.com/TingDang90/CNDP

Table 2: Comparison of CNDPs and state-of-the-art systems in terms of Unweighted Average Recall (UAR), Sensitivity, and Specificity in percentage (%). 95% confidence interval is estimated using Bootstrap and reported in parenthesis.

	Forecasting	Systems	UAR	Sensitivity	Specificity
Audio only	One-step-ahead	RNN Δt_1	75.5(71.0-79.5)	73.4(65.6-80.5)	77.6(73.9-81.2)
	Multi-step-ahead	RNN Δt_{all}	74.7(70.3-78.8)	73.4(65.5-81.0)	75.9(72.2-79.8)
		RNN-VAE	74.8(70.5-78.9)	75.0(67.0-82.2)	74.7(70.8-78.3)
		Transformer	75.3(70.3-78.9)	72.7(64.8-80.0)	76.8(73.0-80.7)
		CNDPs	**77.1(72.6-80.9)**	**76.6(68.6-83.6)**	77.6(73.8-81.3)
		CNDPs$_l$	**78.1(74.0-81.8)**	**78.9(71.2-85.7)**	77.2(73.1-80.9)
Audio + Labels	Multi-step-ahead	RNN $\Delta t_{all}^{\hat{y}}$	82.5(78.8-85.9)	84.4(78.0-90.4)	80.5(76.9-83.9)
		RNN-VAE$^{\hat{y}}$	75.1(70.5-79.2)	73.4(65.4-80.6)	76.8(73.0-80.7)
		Transformer$^{\hat{y}}$	77.9(73.6-81.5)	79.7(72.5-86.2)	75.9(72.1-79.9)
		CNDPs$^{\hat{y}}$	**88.3(84.8-91.5)**	**84.4(78.0-90.3)**	**92.3(89.7-94.5)**
		CNDPs$_l{}^{\hat{y}}$	**93.6 (90.8-96.1)**	**90.6(85.2-95.2)**	**96.7(94.9-98.2)**

the relationship between a continuous variable (e.g., the predicted probability of positive) and a binary variable (such as COVID-19 status in terms of positive or negative). It is computed as:

$$\gamma_{pb} = \frac{\mu_1 - \mu_0}{s_n}\sqrt{\frac{n_1 n_0}{n^2}}. \tag{8}$$

Here μ_1 and μ_0 are the mean values of the predicted probabilities p_i for the positive and negative samples of the participant, s_n is the standard deviation of the predicted probabilities for all the samples of the participant. n_1 and n_0 are the numbers of samples in the positive and negative classes of each participant, while n is the total number $n = n_1 + n_2$. A higher γ_{pb} indicates a stronger correlation, thus a better-forecasted disease progression trajectory.

For the participants who consistently report positive or negative test results, γ_{pb} is not applicable. Therefore, the ratio of correctly predicted samples γ is computed. Details of evaluation metrics can be found in Appendix D.

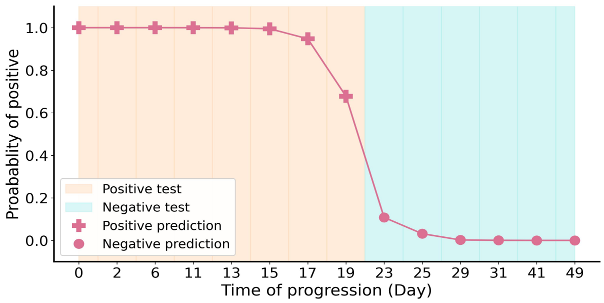


Figure 3: An example of the forecasted disease progression in terms of probability of positive over time. Only the sample at day 0 is given and the disease progression from day 2 is predicted. The forecasted decreasing trend matches the true disease development from day 0 to day 49, with a high correlation of 0.98.

5 RESULTS AND DISCUSSION

5.1 Comparison with Baselines

5.1.1 Classification. The results are shown in Table 2. In terms of the systems without label-enhancing mechanism (i.e., audio only), four baselines show similar performance. The proposed CNDPs outperform four baselines in terms of UAR and sensitivity, and show comparable or better performance for specificity. Surprisingly, CNDPs outperform RNN Δt_1, where the latter as a one-step-ahead forecasting problem is expected to be a simpler task. This is possible as CNDPs capture the continuous disease progression using ODE, while RNN Δt_1 models it discretely. The inferior performance of the Transformer is possibly due to the large number of model parameters which are hard to be optimized with our limited data. We have observed overfitting on the training set even with a high dropout ratio of 0.9. This further shows the advantages of our proposed model when limited data is available, which is common in real healthcare scenarios. Further, with only the initial point as observed, the multiple attention heads in Transformers yield no impact. CNDPs$_l$ with individual-level loss perform best, with relative improvements of 3.4%, 4.6%, 4.4% and 3.7% over four baselines in terms of UAR, and 7.5%, 7.5%, 5.5% and 8.5% in terms of sensitivity. This suggests that incorporating the individual-level temporal dynamics in disease progression aids the positive and negative classification.

Regarding the systems with label-enhancing mechanism (i.e., audio + labels), they generally display better performance compared to non-label-enhancing systems (i.e., audio only), suggesting the benefits of including past health status in the proposed model for forecasting. The label-enhancing CNDPs$^{\hat{y}}$ and CNDPs$_l^{\hat{y}}$ significantly outperform three baselines RNN Δt_{all}^{y}, RNN-VAE$^{\hat{y}}$ and Transformer$^{\hat{y}}$, with CNDP$_l^{\hat{y}}$ yielding the best UAR of 93.6%, sensitivity of 90.6% and specificity of 96.7%. Further, it can be seen that RNN-VAE$^{\hat{y}}$ with the most similar structure to CNDPs$_l^{\hat{y}}$ and the advanced Transformer$^{\hat{y}}$ do not show significant improvements over their non-label-enhancing versions. Although the inclusion of the initial health status in RNN-VAE$^{\hat{y}}$ can lead to a better $\mathbf{z}_{t_0}$ in the latent space (cf. Section 3.2), the decoder in RNN-VAE$^{\hat{y}}$ is still discrete and the lack of continuity may not generate better $\mathbf{z}(t)$. On the other hand, $\mathbf{z}_{t_0}$ in the CNDPs based models (cf. Figure 2a)

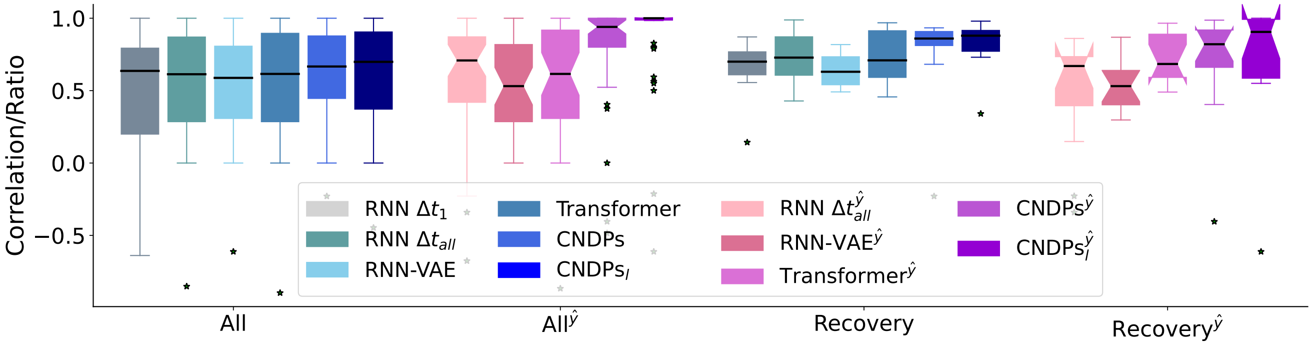
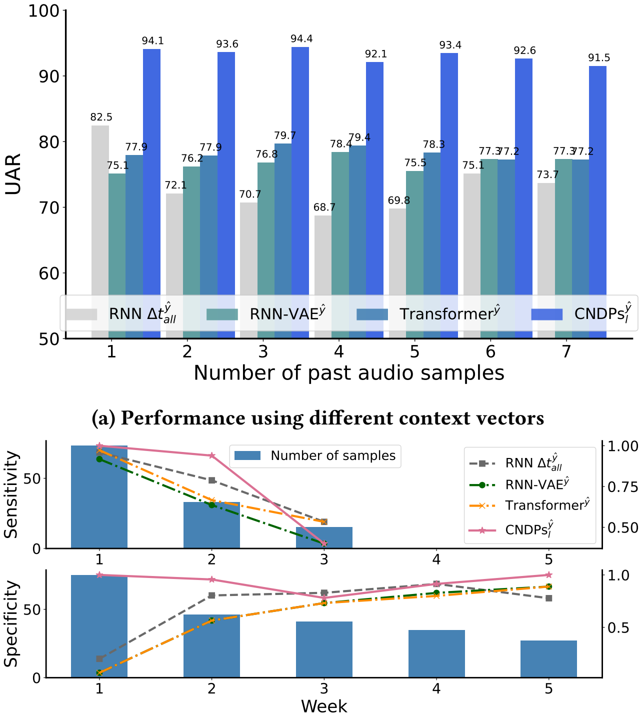


Figure 4: Comparison in terms of the correlation/ratio between the forecasted progression with test results for the entire test cohort and the recovery subgroup. CNDPs$^{\hat{y}}$ and CNDPs$_l^{\hat{y}}$ significantly outperform baselines in tracking disease progression.

highly determines $\mathbf{z}(t)$ governed by the continuous ODE, thus a better $\mathbf{z}_{t_0}$ is obtained, leading to a better $\mathbf{z}(t)$ to aid forecasting.



(a) Performance using different context vectors

(b) Performance of different forecasting horizon

Figure 5: Performance of RNN $\Delta t_{all}^{\hat{y}}$, RNN-VAE$^{\hat{y}}$ and CNDPs$_l^{\hat{y}}$ (a) with different number of context vectors, ranging from 1 to 7, and (b) for week 1 to 5 forecasting. CNDPs$_l^{\hat{y}}$ achieve accurate forecasting with limited past information, and can also reliably forecast for a 2-3 week horizon.

5.1.2 Progression prediction. We first presented a case study of one participant, showing the forecasted probability over 49 days using CNDPs$_l^{\hat{y}}$ in Figure 3. A decrease in the predicted probabilities of positive (pink curve) can be clearly observed, aligning with the reported recovery trend (shaded background). The correlation γ_{pb} between the predicted curve and test results is 0.98. Further, we can classify the probability at each time step as positive or negative with a threshold of 0.5, which also matches the test results.

The correlation γ_{pb}/γ for the entire test cohort and the recovery subgroup are displayed in Figure 4. The recovery subgroup includes participants who reported infection first and recovered after a few days. For the entire test cohort (All and All$^{\hat{y}}$ in Figure 4), CNDPs and CNDPs$_l$ show better performance over four baselines, and the smaller interquartile range of CNDPs indicates a smaller variability across different participants. Moreover, CNDPs$^{\hat{y}}$ and CNDPs$_l^{\hat{y}}$ with label-enhancing mechanism display significant improvements over their non-label-enhancing versions compared to that of other baselines. For the recovery subgroup, a similar trend is also observed. The superior performance using CNDPs (Recovery in Figure 4 without label-enhancing mechanism) further suggests the great benefits of using audio biomarkers only to capture disease progression and, therefore, the promises in recovery monitoring.

5.2 Time Dependencies and Forecasting Horizons

5.2.1 How many days are required for reliable forecasting? We analysed whether using more past longitudinal audio samples can improve the forecasting performance for RNN $\Delta t_{all}^{\hat{y}}$, RNN-VAE$^{\hat{y}}$, Transformer$^{\hat{y}}$ and CNDPs$_l^{\hat{y}}$. If the model is capable of capturing the disease progression using limited past audio samples, the system performance may not be improved significantly when more past audio samples are used for forecasting. Figure 5a shows the effect of varied days on the performance in terms of UAR, since it captures the performance in both positive and negative classes. RNN $\Delta t_{all}^{\hat{y}}$ shows decreased performance, possibly due to the mismatch between training and test, where the first sample is used to forecast the future values during the training and the first 2 to 7 samples are used during the test. Performance of RNN-VAE$^{\hat{y}}$ increases from 75.1 to 78.4 when the number of context vectors increases from 1 to 4, showing that increasing past information is beneficial for forecasting using RNN-VAE$^{\hat{y}}$. Oppositely, Transformer$^{\hat{y}}$ and CNDPs$_l^{\hat{y}}$

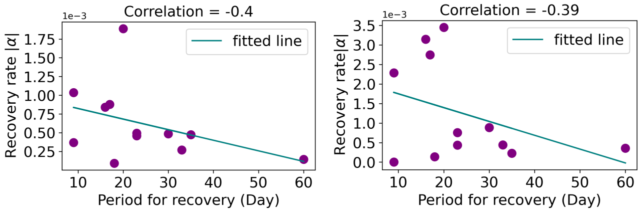
show relatively stable performance, with a slight increase from 77.9 to 79.4 for Transformer$^{\hat{y}}$ and from 94.1 to 94.4 for CNDPs$_l^{\hat{y}}$ respectively, when context vectors increase from 1 to 3. The proposed CNDPs$_l^{\hat{y}}$ consistently outperform Transformer$^{\hat{y}}$. This suggests that the proposed model can achieve reliable COVID-19 forecasting with limited past information, e.g., even with only one initial audio sample.

5.2.2 How long can it effectively forecast? We further investigated the forecasting horizon of the model, i.e., how long the model can reliably forecast the disease progression. The forecasting performance for different time periods (i.e., weeks) is reported in Figure 5b, ranging from the 1st week to the 3rd/5th week in terms of sensitivity and specificity respectively. Sensitivity after 3rd week was not analysed due to the extremely scarce number of samples collected afterwards. The blue bars represent the number of samples for each week, and the coloured lines represent the sensitivity and specificity of different systems. In terms of sensitivity, it can be seen a decreasing trend for all four systems. This is reasonable as forecasting progression in the near future is an easier task compared to that in the far future. The performance drop of CNDPs$_l^{\hat{y}}$ from week 1 to week 2 is smaller compared to other systems, which still yields 93.9% for week 2, suggesting its reliability in forecasting the progression in the next 2 weeks.

In terms of specificity, it is surprising that three baselines show an increasing trend while CNDPs$_l^{\hat{y}}$ displays a relatively stable performance from week 1 to week 5. The low specificity at the first week for three baselines suggests that the models tend to classify more false positives in the first week, indicating a bias towards positive classes. This bias could be due to the imbalanced reporting period for positive and negative users. Specifically, positive users tend to report for a shorter period (i.e., 25 days on average) than negative users (i.e., 40 days on average). The baseline models may have learned this information unintentionally and used it for forecasting. Consequently, for the first week, the model takes a short duration of the sequence and is more likely to predict positive classes, yielding a higher sensitivity and a lower specificity. The increasing specificity from weeks 1 to 5 could be due to capturing the skewed distribution in the data instead of capturing the underlying disease progression. On the other hand, the CNDPs$_l^{\hat{y}}$ remained relatively stable from week 1 to 5, validating its reliability in capturing the disease progression instead of the skewed data distributions. CNDPs$_l^{\hat{y}}$, in general, displays better performance compared to other baselines and also remains stable from weeks 1 to 5. Combining both sensitivity and specificity, these results suggest the potential of the proposed systems for forecasting in a 2 to 3 weeks horizon.

5.3 Individual-specific Recovery Rate

The recovery rate (returning to normal from COVID-19 infection) varies among individuals, due to a variety of factors such as comorbidities and age [35, 40]. We aim to investigate the capability of our models in capturing individual-specific recovery rates. As shown in Figure 3, we can estimate the recovery rate by estimating the sharpness of the decrease in the predicted trajectory (pink line). This is achieved by fitting a second order polynomial function to the forecasted trajectory, and using the absolute value of the coefficient of the square term $|\alpha|$ in the polynomial function to reflect the recovery rate. A larger $|\alpha|$ indicates a sharper curve thus a fast recovery, and vice versa.



(a) CNDPs$_l$

(b) CNDPs$_l^{\hat{y}}$

Figure 6: Scatter plot of the predicted recovery rate from (a) CNDPs$_l$ and (b) CNDPs$_l^{\hat{y}}$ and reported period of recovery days. The longer the recovery days, the smaller the recovery rate $|\alpha|$. Negative correlations of -0.4 and -0.39 are observed for CNDPs$_l$ and CNDPs$_l^{\hat{y}}$ respectively, suggesting the capability of the models in forecasting the individual-specific recovery rate.

A scatter plot between the predicted recovery rates $|\alpha|$ and the true recovery days for the recovery subgroup is shown in Figure 6a and 6b for CNDPs$_l$ and CNDPs$_l^{\hat{y}}$ respectively. It is observed that i) the predicted recovery rates vary among individuals, suggesting the potential in predicting individual-specific recovery rate; ii) a smaller $|\alpha|$ (i.e., slow recovery) generally corresponds to a longer period of recovery days, matching the expectations. The Pearson's correlation coefficients between $|\alpha|$ and the recovery days yield -0.40 and -0.39 for CNDP$_l$ and CNDP$_l^{\hat{y}}$, respectively, further validating the effectiveness of proposed models in forecasting the individual-specific recovery.

6 CONCLUSION

Novel CNDPs have been proposed for time series forecasting, and shown great promise in irregular-sampled time series modelling and achieved accurate forecasting with limited past information. The model validated on a crowdsourced audio dataset for COVID-19 disease progression outperforms state-of-the-art time series modelling approaches, and in-depth analysis further reveals its effectiveness in relatively long-range forecasting and individual-specific recovery trend tracking.

In general, the proposed CNDPs can be potentially employed for any type of time series, and they particularly benefit chronic disease progression forecasting in a remote monitoring context, due to its mechanism in individual-level tracking. Future work includes personalised model development that incorporates personal information (e.g. medical history, smoking habits, etc.) or adapts the universal model to each individual, to better account for the differences in the individuals' disease progression (e.g., genetic variations).

ACKNOWLEDGEMENT

This work was supported by ERC Project 833296 (EAR). We thank everyone who volunteered their data.

REFERENCES

[1] Willem Bonnaffé, Ben C Sheldon, and Tim Coulson. 2021. Neural ordinary differential equations for ecological and evolutionary time-series analysis. *Methods in Ecology and Evolution*, 12, 7, 1301–1315.

[2] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time series analysis: forecasting and control.* John Wiley & Sons.

[3] Chloë Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, and Cecilia Mascolo. 2020. Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3474–3484.

[4] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. 2018. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8, 1, 1–12.

[5] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. 2018. Neural ordinary differential equations. *arXiv preprint arXiv:1806.07366*.

[6] Jeongwhan Choi, Hwangyong Choi, Jeehyun Hwang, and Noseong Park. 2022. Graph neural controlled differential equations for traffic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* number 6. Vol. 36, 6367–6374.

[7] Harry Coppock, Alex Gaskell, Panagiotis Tzirakis, Alice Baird, Lyn Jones, and Björn Schuller. 2021. End-to-end convolutional neural network enables covid-19 detection from breath and cough audio: a pilot study. *BMJ innovations*, 7, 2.

[8] Ting Dang et al. 2022. Exploring longitudinal cough, breath, and voice data for covid-19 progression prediction via sequential deep learning: model development and validation. *Journal of medical Internet research*, 24, 6, e37004.

[9] Jan G De Gooijer and Rob J Hyndman. 2006. 25 years of time series forecasting. *International journal of forecasting*, 22, 3, 443–473.

[10] Soham Deshmukh, Mahmoud Al Ismail, and Rita Singh. 2021. Interpreting glottal flow dynamics for detecting covid-19 from voice. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1055–1059.

[11] John R Dormand and Peter J Prince. 1980. A family of embedded runge-kutta formulae. *Journal of computational and applied mathematics*, 6, 1, 19–26.

[12] James Durbin and Siem Jan Koopman. 2012. *Time series analysis by state space methods.* Oxford university press.

[13] Saba Emrani, Anya McGuirk, and Wei Xiao. 2017. Prognosis and diagnosis of parkinson's disease using multi-task learning. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 1457–1466.

[14] Jie Feng, Ziqian Lin, Tong Xia, Funing Sun, Diansheng Guo, and Yong Li. 2021. A sequential convolution network for population flow prediction with explicitly correlation modelling. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 1331–1337.

[15] Everette S Gardner Jr. 1985. Exponential smoothing: the state of the art. *Journal of forecasting*, 4, 1, 1–28.

[16] Mostafa Mehdipour Ghazi, Mads Nielsen, Akshay Pai, M Jorge Cardoso, Marc Modat, Sébastien Ourselin, Lauge Sørensen, Alzheimer's Disease Neuroimaging Initiative, et al. 2019. Training recurrent neural networks robust to incomplete data: application to alzheimer's disease progression modeling. *Medical image analysis*, 53, 39–46.

[17] Jing Han et al. 2022. Sounds of covid-19: exploring realistic performance of audio-based digital testing. *NPJ digital medicine*, 5, 1, 1–9.

[18] Andrew C Harvey. 1990. Forecasting, structural time series models and the kalman filter.

[19] Shawn Hershey et al. 2017. Cnn architectures for large-scale audio classification. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)*. IEEE, 131–135.

[20] Jiahao Ji, Jingyuan Wang, Zhe Jiang, Jiawei Jiang, and Hu Zhang. 2022. Stden: towards physics-guided neural networks for traffic flow prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[21] Pengbo Jiang, Xuetong Wang, Qiongling Li, Leiming Jin, and Shuyu Li. 2018. Correlation-aware sparse and low-rank constrained multi-task learning for longitudinal analysis of alzheimer's disease. *IEEE journal of biomedical and health informatics*, 23, 4, 1450–1456.

[22] Jordi Laguarta, Ferran Hueto, and Brian Subirana. 2020. Covid-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, 1, 275–281.

[23] Abdul Ghaaliq Lalkhen and Anthony McCluskey. 2008. Clinical tests: sensitivity and specificity. *Continuing education in anaesthesia critical care & pain*, 8, 6, 221–223.

[24] Baiying Lei et al. 2020. Deep and joint learning of longitudinal data for alzheimer's disease prediction. *Pattern Recognition*, 102, 107247.

[25] Bryan Lim and Stefan Zohren. 2021. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379, 2194, 20200209.

[26] Bryan Lim, Stefan Zohren, and Stephen Roberts. 2020. Recurrent neural filters: learning independent bayesian filtering steps for time series prediction. In *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[27] Alexander Norcliffe, Cristian Bodnar, Ben Day, Jacob Moss, and Pietro Liò. 2020. Neural ode processes. In *International Conference on Learning Representations*.

[28] Sunghyun Park, Kangyeol Kim, Junsoo Lee, Jaegul Choo, Joonseok Lee, Sookyung Kim, and Edward Choi. 2020. Vid-ode: continuous-time video generation with neural ordinary differential equation. *arXiv preprint arXiv:2010.08188*.

[29] Matjaž Perc, Nina Gorišek Miksić, Mitja Slavinec, and Andraž Stožer. 2020. Forecasting covid-19. *Frontiers in Physics*, 8, 127.

[30] Gavin D Portwood et al. 2019. Turbulence forecasting via neural ode. *arXiv preprint arXiv:1911.05180*.

[31] Yulia Rubanova, Ricky TQ Chen, and David K Duvenaud. 2019. Latent ordinary differential equations for irregularly-sampled time series. *Advances in neural information processing systems*, 32.

[32] Furqan Rustam, Aijaz Ahmad Reshi, Arif Mehmood, Saleem Ullah, Byung-Won On, Waqar Aslam, and Gyu Sang Choi. 2020. Covid-19 future forecasting using supervised machine learning models. *IEEE access*, 8, 101489–101499.

[33] Afan Galih Salman, Yaya Heryadi, Edi Abdurahman, and Wayan Suparta. 2018. Single layer & multi-layer long short-term memory (lstm) model with intermediate variables for weather forecasting. *Procedia Computer Science*, 135, 89–98.

[34] Afan Galih Salman, Bayu Kanigoro, and Yaya Heryadi. 2015. Weather forecasting using deep learning techniques. In *2015 international conference on advanced computer science and information systems (ICACSIS)*. Ieee, 281–285.

[35] Adekunle Sanyaolu et al. 2020. Comorbidity and its impact on patients with covid-19. *SN comprehensive clinical medicine*, 2, 8, 1069–1076.

[36] Ruian Shi, Haoran Zhang, and Quaid Morris. 2022. Pan-code: covid-19 forecasting using conditional latent odes. *Journal of the American Medical Informatics Association*, 29, 12, 2089–2095.

[37] Robert F Tate. 1954. Correlation between a discrete and a continuous variable. point-biserial correlation. *The Annals of mathematical statistics*, 25, 3, 603–607.

[38] George Trigeorgis, Fabien Ringeval, Raymond Brueckner, Erik Marchi, Mihalis A Nicolaou, Björn Schuller, and Stefanos Zafeiriou. 2016. Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 5200–5204.

[39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

[40] Irena Voinsky, Gabriele Baristaite, and David Gurwitz. 2020. Effects of age and sex on recovery from covid-19: analysis of 5769 israeli patients. *Journal of Infection*, 81, 2, e102–e103.

[41] Xulong Wang, Jun Qi, Yun Yang, and Po Yang. 2019. A survey of disease progression modeling techniques for alzheimer's diseases. In *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*. Vol. 1. IEEE, 1237–1242.

[42] Yuchen Wang, Matthieu Chan Chee, Ziyad Edher, Minh Duc Hoang, Shion Fujimori, Sornnujah Kathirgamanathan, and Jesse Bettencourt. 2020. Forecasting black sigatoka infection risks with latent neural odes. *arXiv preprint arXiv:2012.00752*.

[43] Philip B Weerakody, Kok Wai Wong, Guanjin Wang, and Wendell Ela. 2021. A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing*, 441, 161–178.

[44] Jian Wu et al. 2020. Early antiviral treatment contributes to alleviate the severity and improve the prognosis of patients with novel coronavirus disease (covid-19). *Journal of internal medicine*, 288, 1, 128–138.

[45] Hao Xue and Flora D Salim. 2021. Exploring self-supervised representation ensembles for covid-19 cough classification. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 1944–1952.

[46] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. 2018. Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine*, 13, 3, 55–75.

[47] Qu Zhaowei, Li Haitao, Li Zhihui, and Zhong Tao. 2020. Short-term traffic flow forecasting method with mb-lstm hybrid network. *IEEE Transactions on Intelligent Transportation Systems.*

[48] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* number 12. Vol. 35, 11106–11115.

Appendix

A DATA COLLECTION AND PARTITION

A mobile app was designed and released in 2020 to collect crowdsourced respiratory sounds. Each participant was encouraged to record three different sound types via smartphone built-in microphones. This includes the cough repeated three times which is forced, deep breath repeated three to five times, and speech by reading a sentence displayed on screen for three times. The corresponding test results are also required, chosen from a list of 'positive', 'negative' and 'not tested'. Additionally, participants' symptoms, medical history, demographics, smoking status and hospitalisation are also collected. The app was released in multiple platforms including Android, iOS and a webpage. Therefore, the data contains different formats of audio files (i.e., .wav, .m4a, etc.) and different sampling rates (i.e., 48kHz, 44.1kHz, 8kHz, 16kHz, etc.). The data will be made publicly available for academic research upon publication.

We divided the users into training (70%), validation (10%), and testing (20%) sets, with an equal number of positive and negative users in each set. We also aimed to balance the gender, age, and language distribution across these sets. The details of demographic data for each set are shown in Figure 1.

Regarding data size, we implemented data augmentation, leading to a threefold increase in the data size compared to the original size. This approach was expected to reduce overfitting. We appreciate the reviewer's suggestion and will consider implementing cross-validation in future work.

B BASELINES

The state-of-the-arts systems for time series forecasting as comparisons to our proposed model are detailed below:

- RNN Δt_1: It is a classic RNN based autoregressive model. The time difference between consecutive days of audio recordings is used as additional input to model irregular-sampled time series. This baseline uses the past longitudinal audio samples $[\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$ to predict label at next time step $\mathbf{y}_{n+1}$. It is a one-step-ahead forecasting framework.
- RNN Δt_{all}: This employs a similar structure as RNN Δt_1, but used for multi-step-ahead forecasting. The past longitudinal audio sample $[\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$ is used as the context vector to forecast the future COVID-19 test labels $y(t) = [y_{n+1}, \cdots, y_N]$.
- RNN-VAE: This is an RNN based encoder-decoder structure [31]. The past audio samples from t_0 to t_n are first transformed to a latent distribution, and a random sample $\mathbf{z}_0$ is obtained and used as the initial hidden state for the RNN based decoder. The recurrent processing in the decoder achieves the disease progression forecasting within $[y_{t_{n+1}}, y_{t_N}]$.

C MODEL TRAINING

For the encoder in the proposed CNDPs, f_1 to f_5 adopts fully-connected (FC) layers. 3 FC layers are used for f_1, and 2 FC layers are employed for f_2, f_3, f_4 and f_5. ReLU activation is used for input and hidden layers. Each hidden layer consists of 100 neurons. The latent distribution $\mathbf{s}_0$ and $\mathbf{r}$ is empirically set to 25 dimension, leading to $\mathbf{z}_{t_0}$ of dimension 50. For the typical loss, weighted cross entropy (CE) with Sigmoid activation is used. The weights are optimized within [1, 5]. To enable a fair comparison, the network structure in RNN Δt_1, RNN Δt_{all} and RNN-VAE is similar as in CNDPs. For RNN Δt_1 and RNN Δt_{all}, one GRU layer is used with latent dimension setting to 25, as it shows better performance than 50 (used in CNDPs). In terms of the RNN-VAE, the encoder and decoder adopt a similar network structure, which employs one GRU layer with latent dimension setting to 50 and 100, respectively. One FC layer is used as the output layer. The scaling factor for the KL divergence in the loss function is set by 1 initially and decays with a ratio of 0.01 at each epoch. Similarly, weighted CE loss is used.

The parameters of the model have been fine-tuned over a range of values. The Adam optimizer was adopted for all the systems. The initial learning rate (lr) is optimized within [1e-2, 1e-3, 1e-4, 1e-5]. The decay factor is tuned within the range of [0.98, 0.95, 0.9]. The optimal latent dimension for RNN-based models is chosen from [25, 50, 100], while for CNDP-based models it is selected from [25, 50]. Additionally, the weight for the cross-entropy loss is fine-tuned within the range of [1, 4]. 60 epochs were used, and the best model with the highest sum/product of sensitivity and specificity in the validation set is saved and used for the test.

D EVALUATION METRICS

In terms of the classification accuracy in Table 2, UAR represents the average recall for each class. Sensitivity and specificity demonstrate the model's capability in identifying correctly positive and negative samples, respectively [23].

The Point-Biserial Correlation Coefficient γ_{pb} is used to evaluate the forecasting performance in tracking individuals' disease progression and calculated between the forecasted trajectory and test results for each participant. For the participants who continuously reported positive or negative test results, we adopted the accuracy γ computed as the ratio of the correctly predicted samples N_i over the total number of samples for each participant. as:

$$\gamma = \frac{N_i}{N} \tag{1}$$

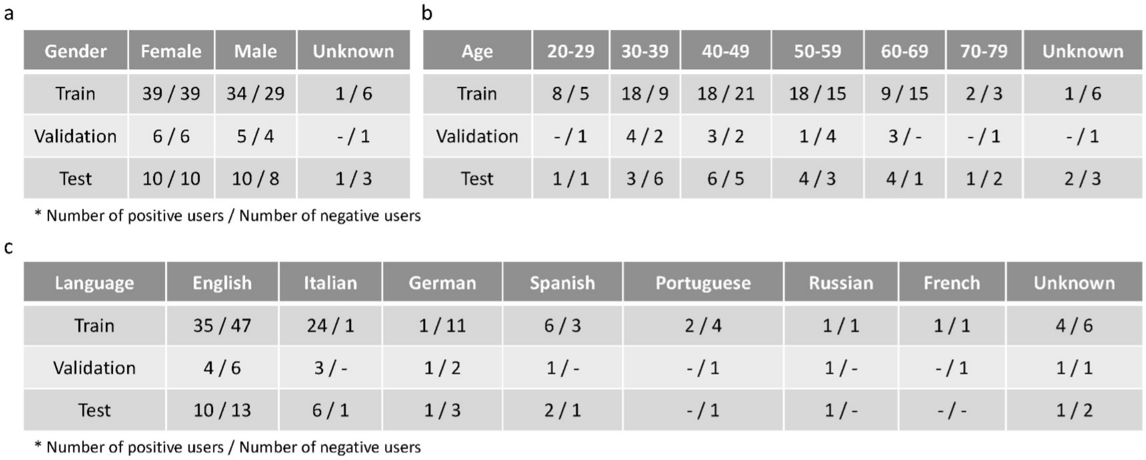
where N_i and N are the correctly predicted samples and the total number of samples of each individual. γ_{pb} ranges within [-1,1], and γ is in the range of [0,1]. A higher value of γ_{pb} or γ indicates a better forecasted trajectory. Therefore, we pool γ_{pb} and γ together for all the participants in the test cohort and reported the performance.

E NDP LOSS

NDPs are trained using an amortised variational inference procedure, which jointly optimised the encoder, the latent ODE and the decoder by maximising the ELBO, equal to minimising the loss function $L(\theta, \phi)$ as:

$$\begin{aligned} L(\theta,\phi) = \mathbb{E}_{\mathbf{z}_{t_0} \sim q_\phi(\mathbf{z}_{t_0}|\mathbf{x}_\mathbb{C}, t_\mathbb{C}, y_\mathbb{C})}[-log p_\theta(y_\mathbb{T}|\mathbf{z}_{t_0}, t, \theta)] \\ + D_{KL}(q_\phi(\mathbf{z}_{t_0}|\mathbf{x}_\mathbb{C}, t_\mathbb{C}, y_\mathbb{C})||q_\phi(\mathbf{z}_{t_0}|\mathbf{x}_\mathbb{T}, t_\mathbb{T}, y_\mathbb{T})) \end{aligned} \tag{2}$$

It consists of the negative log-likelihood and KL divergence, with D_{KL} representing the KL divergence, and q_ϕ representing the variational posteriors learnt for the hidden state $\mathbf{z}_{t_0}$. The subscripts $\mathbb{C}$ and $\mathbb{T}$ represent the context vectors and target vectors, where



a

Gender	Female	Male	Unknown
Train	39 / 39	34 / 29	1 / 6
Validation	6 / 6	5 / 4	- / 1
Test	10 / 10	10 / 8	1 / 3

b

Age	20-29	30-39	40-49	50-59	60-69	70-79	Unknown
Train	8 / 5	18 / 9	18 / 21	18 / 15	9 / 15	2 / 3	1 / 6
Validation	- / 1	4 / 2	3 / 2	1 / 4	3 / -	- / 1	- / 1
Test	1 / 1	3 / 6	6 / 5	4 / 3	4 / 1	1 / 2	2 / 3

* Number of positive users / Number of negative users

c

Language	English	Italian	German	Spanish	Portuguese	Russian	French	Unknown
Train	35 / 47	24 / 1	1 / 11	6 / 3	2 / 4	1 / 1	1 / 1	4 / 6
Validation	4 / 6	3 / -	1 / 2	1 / -	- / 1	1 / -	- / 1	1 / 1
Test	10 / 13	6 / 1	1 / 3	2 / 1	- / 1	1 / -	- / -	1 / 2

* Number of positive users / Number of negative users

Figure 1: Data statistics in the training, validation, and test partitions in terms of gender, age, and language. a: gender, b: age, c: language.

context vectors are past longitudinal samples, and target vectors are the future samples (cf. section 3.1).