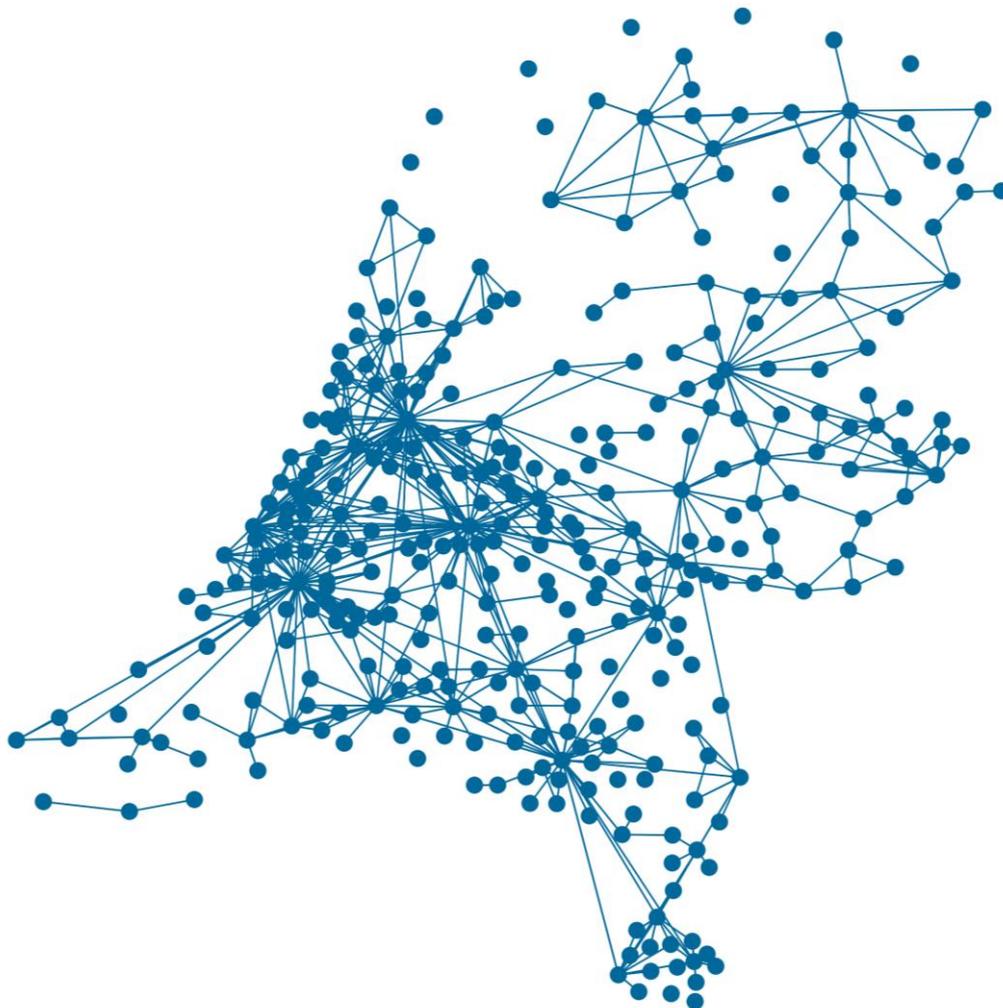


Forecasting of SARS-Cov-2 Infections within Dutch Municipalities using Spatio-Temporal Graph Neural Networks

V. Maxime Croft

Master of Science Thesis



Forecasting of SARS-Cov-2 Infections within Dutch Municipalities using Spatio-Temporal Graph Neural Networks

THESIS

For the degree of Master of Science in Robotics at Delft University of
Technology

V. Maxime Croft
4390024

Thesis committee:	Dr. C. Della Santina	TU Delft, supervisor, chair
	S.C.J.L. van Iersel, MSc	RIVM, supervisor, member
	Prof.dr. R. Babuska	TU Delft, external member

December 31, 2022

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of
Technology

Forecasting of SARS-CoV-2 Infections within Dutch Municipalities using Spatio-Temporal Graph Neural Networks

V. Maxime Croft^{1,2}

Abstract—This paper presents a novel approach to regional forecasting of SARS-Cov-2 infections one week ahead, which involves developing a municipality level COVID-19 dataset of the Netherlands and using a spatio-temporal graph neural network (GNN) to predict the number of infections. The developed model captures the spread of infectious diseases within municipalities over time using Gated Recurrent Units (GRUs) [1] and the spatial interactions between municipalities using GATv2 layers [2]. To the best of our knowledge, this model is the first to incorporate sewage data [3], the stringency index [4], and commuting information [5] into GNN-based infection prediction.

In experiments on the developed real-world dataset, we demonstrate that the model outperforms simple baselines and purely spatial or temporal models for the COVID-19 wild type, alpha, and delta variants. In combination with an average R^2 of 0.795 for forecasting infections and of 0.899 for predicting the associated trend of these variants, we conclude that the model is well suited for predicting the spread of infectious diseases with similar disease dynamics in real world applications. To increase prediction performance and to improve the generalizability of the model for infectious diseases with more complex disease dynamics, we recommend using additional (synthetic) data or expanding the regional forecasting scale in future work.

Index Terms—epidemic prediction, deep learning, spatio-temporal graph neural networks, real world evidence, COVID-19,

regions when there are large differences in infection prevalence [10]. Schools in affected regions could, for instance, encourage parents and children to pay extra attention to disease symptoms and to test preventively. In addition, regional epidemic forecasts can be beneficial for capacity planning within a country. Hospitals could prepare for the potential need to relocate patients to regions with fewer infections.

Extensive research has been devoted to developing a wide range of effective epidemic forecasting models which can be mechanistic [11], [12], based on statistics or machine learning [13], [14], or both [15]–[17]. Mechanistic models are grounded on theoretical principles of disease spread, whereas statistical and machine learning approaches are heavily data-driven [18].

Because of the enormous global impact of COVID-19 and the technologies we have today, this is the first time that epidemic data is available on such a large scale. Therefore, COVID-19 offers more opportunities for epidemic research and data-driven modelling than ever before. As a result of the lessons learned during this period, it is likely that for more and more epidemics and future pandemics, necessary data will be available. Deep learning methods in particular have an outstanding ability to discover complex patterns from large amounts of data [19], [20]. Lastly, they can be easily extended to various temporal and geographical scales, and other diseases, in the presence of collected data. For these reasons, we decided to focus on this machine learning technique.

Current deep learning approaches mainly consider epidemic forecasting as a time-series problem. They usually assume that forecasts for a given location are dependent only on information from that location, without incorporating inter-regional movements and interactions. This is despite the fact that research has indicated that human movement between regions contributes significantly to the transmission and spread of infectious diseases [10]. Because of this, we believe that data on these interregional interactions could be leveraged to increase the prediction performance of (purely temporal) epidemic forecasting models. This warrants a natural graph-based representation of the problem, allowing the application of a subcategory of deep learning called graph neural networks (GNNs).

GNNs are capable of dealing with the irregular nature of graphs and the complex relationships and interdependencies between their objects [20]. The core idea of GNNs is that the representation of each graph’s node is updated based on an

I. INTRODUCTION

Epidemics of infectious diseases are occurring more frequently, and are spreading faster and further than ever before [6]. This has become abundantly clear between late 2019 and early 2020, when COVID-19 quickly progressed from a local outbreak to a global pandemic. The SARS-Cov-2 virus has infected more than 600 million individuals, resulting in over 6 million deaths worldwide [7], and causing significant economic damage due to large-scale quarantining and country-wide lockdowns [8].

Prevention, containment, and mitigation of the spread of infectious diseases is therefore key to humanity [9]. To ensure that policymakers can impose measures and manage the allocation of scarce medical resources to combat the spread of the virus, the ability to accurately forecast epidemics is of paramount importance [8]. On a regional scale, forecasts can be used to justify local measures tailored to particular

¹Department of Cognitive Robotics, Delft University of Technology, Delft, The Netherlands.

²COVID-19 Epidemiology and Surveillance Team, Dutch National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands.

aggregation of messages received from its connected neighbors [21]. Spatio-temporal GNNs can additionally handle data in the temporal dimension. GNNs have successfully been applied to a wide variety of domains and tasks where interaction between different components is important, reaching state-of-the-art performance [20], [22]–[24]. For example, Derron-Pinion et al. [25] developed a GNN that predicts the estimated time of arrival of traffic in Google Maps, and Ying et al. [26] introduced the large-scale GNN recommender system PinSage that is developed and deployed at Pinterest. Due to their proven effectiveness, we would like to examine the use of GNNs for region-based forecasting of epidemic infections.

In collaboration with the Dutch National Institute for Public Health and the Environment (RIVM), we decided that this study focuses on regional SARS-Cov-2 infections in the Netherlands. While earlier work on the regional prediction of disease infections using GNNs exists [8], [15], [17], [21], [27], [28], GNNs for SARS-Cov-2 infection prediction have never been applied in the Netherlands and worldwide not at a small municipality-level scale. We will clarify our contributions with respect to the mentioned related works in detail in Section II. In light of the current research gap, the objective of this research is to develop a model for one week ahead forecasting of reported SARS-Cov-2 infections within municipalities of the Netherlands using GNNs.

Research has shown that for spatio-temporal GNN forecasting, the performance decreases as the prediction horizon increases [17], [21], [27]. Predicting one week ahead gives local governments time to anticipate and take required action, without adding unnecessary uncertainty to the model’s predictions. Since no information on the actual number of infections is available, we assume that a region’s number of officially confirmed reported cases represents this.

This work’s main contributions are:

- Developing a COVID-19 dataset containing municipality level information of the Netherlands, including COVID-19 statistics, demographics, and information on interactions of municipalities.
- Creating a novel spatio-temporal GNN model that is able to capture relationships over time and between municipalities, for predicting the number of SARS-Cov-2 infections per municipality one week ahead using the developed COVID-19 dataset.
- Evaluating the proposed approach on the developed COVID-19 dataset. We observe that on average, the spatio-temporal GNN outperforms the baselines.

The remaining part of this paper is organized as follows: Section II reviews the related work on the use of GNNs for epidemic forecasting. Section III describes the proposed methodology used for forecasting the number of SARS-Cov-2 infections. Section IV explains the performed experiments, and Section V presents the obtained results. These results are further discussed in Section VI. Finally, Section VII concludes this paper.

II. RELATED WORK

Numerous studies on the application of GNNs in the field of epidemiology have been performed. These studies are focused either on forecasting epidemic spreading [8], [15], [17], [21], [27]–[31], extracting the full state of a spreading epidemic [32], reconstructing their evolution [33], [34], generating mobility-control policies [35], or prioritizing vaccine or test receivers [9], [36]. Below, the existing methods on location-based spatio-temporal forecasting of the number of infections are discussed.

A. Spatio-temporal Graph Neural Networks for Epidemic Forecasting

Deng et al. [27] developed a graph neural network called Cola-GNN to predict weekly influenza-like illness cases in the United States and Japan, without the use of an underlying graph. Their framework combines learned temporal feature embeddings with a cross-location attention matrix that captures how locations influence each other.

La Gatta et al. [15] take a different approach. Their proposed hybrid model consists of a GNN framework that estimates the contact rate parameter used to predict infections with the epidemiological SIR and SIRD models. The authors evaluate their approach using COVID-19 data from the regions and provinces of Italy.

Kapoor et al. [8] propose another spatio-temporal GNN method for next-day COVID-19 case prediction. Their model learns from a single spatio-temporal graph, where the spatial edges capture United States county-to-county movement, and a county is connected to a number of past instances of itself with temporal edges. This is the first paper that uses a graph based on data from GPS-enabled mobile devices to model how regions affect each other based on interregional mobility.

Furthermore, Murphy et al. [28] propose a GNN architecture that can learn contagion dynamics on a network from time series data. The approach is demonstrated to be accurate for different contagion dynamics of increasing complexity and can be used to simulate dynamics on arbitrary network structures. The applicability of the approach is demonstrated using real data for predicting infections during the COVID-19 outbreak in Spain.

Most similar to our method are the works of Gao et al. [17] and Panagopoulos et al. [21]. The hybrid spatio-temporal attention network (STAN) of Gao et al. [17] uses real-world COVID-19 data of US counties, and information on demographic similarity and geographical proximity between different forecasting locations as input. Different to our work, the network integrates pandemic transmission dynamics into a deep learning model for enhancing long-term predictions.

More recently, Panagopoulos et al. [21] employed a GNN to predict the number of future COVID-19 cases in the regions of four European countries. To account for the low quantity of available training data, their method utilizes transfer learning to shift disease-spreading models from countries where the epidemic has been stabilized to other countries where the virus is in its early stages.

TABLE I: Features in Related Works

Paper	Inf	Hos	Dec	Sew	Rul	Dem	Dis	Com	GPS
[27]	☒	☐	☐	☐	☐	☒	☒	☐	☐
[15]	☒	☒	☒	☐	☒	☒	☐	☐	☒
[8]	☒	☐	☒	☐	☐	☐	☐	☐	☒
[28]	☒	☐	☐	☐	☐	☒	☐	☐	☒
[17]	☒	☒	☒	☐	☐	☒	☒	☐	☐
[21]	☒	☐	☐	☐	☐	☐	☐	☐	☒
Our work	☒	☒	☐	☒	☒	☒	☐	☒	☐

TABLE II: Models in Related Works

Paper	Temporal	Spatial
Deng et al. [27]	RNN & CNN	MPNN [37]
Kapoor et al. [8]	MLP	GCN [38]
Murphy et al. [28]	RNN	GCN [38]
La Gatta et al. [15]	LSTM	GCN [38]
Gao et al. [17]	GRU	GAT [39]
Panagopoulos et al. [21]	LSTM	MPNN [37]
Our work	GRU	GATv2 [2]

In addition to being the first epidemic prediction GNN applied to the Netherlands at the fine-grained municipality scale, we show in Table I that we also introduce new dataset features to improve predictions. In the Table, the infections (**Inf**) category includes various measures of infection status, such as the number of reported infections, active infections, incidence rate or recovered infections. Other categories include hospitalizations (**Hos**), deceased patients (**Dec**), virus loads in sewage water (**Sew**), and features related to the strictness of the COVID-19 rules (**Rul**). Demographic (**Dem**) features such as the population size and density are also included. Finally, either the distance (**Dis**) between municipalities, commuting information (**Com**), or GPS-based mobility data (**GPS**) are used to connect regions. In this work, we are the first to incorporate virus loads in sewage water and commuter information into our analysis. Additionally, our approach to incorporating the strictness of COVID-19 rules more accurately reflects real-world conditions. Further details about feature selection can be found in Section III-A. Furthermore, our study distinguishes itself by the fact that we have collected data over a period of almost two years, allowing us to compare the results of different COVID-19 variants.

In Table II, we compare the temporal and spatial layers of the models used in previous works with our own approach. All the spatio-temporal GNN methods discussed consist of temporal layers that model the evolution of the epidemic over time and spatial layers that capture the interactions between different locations. In addition to these core components, these models may also incorporate additional deep learning components such as fully connected layers, dropout layers, and skip connections to improve performance. Our approach is the first to use GATv2 [2] as a spatial component, and we propose a unique and optimized combination of deep learning components, which will be further described in Section III-B.

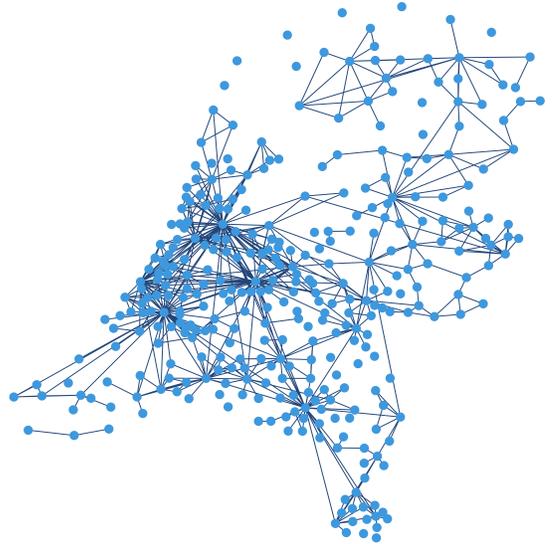


Fig. 1: Input graph G consisting of a set of nodes \mathcal{V} corresponding to the municipalities of the Netherlands and a set of edges \mathcal{E} representing the connections between them.

III. METHODOLOGY

In this section, we present the methods used for developing the COVID-19 dataset and predicting the course of the disease. More specifically, Section III-A presents the constructed graph, Section III-B reviews the proposed model architecture, Section III-C clarifies how we optimize this model, and Section III-D describes how the optimized model is used for making predictions.

A. Graph Construction

In order to make municipality-level predictions for the Netherlands using GNNs, we developed a graph-structured dataset containing information on COVID-19 statistics, demographics, and information on interregional interactions. The resulting undirected graph $G(\mathcal{V}, \mathcal{E})$ represents the model's input data and consists of a set of nodes \mathcal{V} with node features, and a set of edges \mathcal{E} with associated weights. In addition, we use the graph's node labels as a ground truth for supervised learning.

Graph nodes: In the graph G , all 344 Dutch municipalities are modeled as an individual node. Here, $n = |\mathcal{V}|$ denotes the number of nodes. The graph is visualized in Figure 1.

Graph edges and weights: We construct the edges of the graph based on the assumption that mobility rates between pairs of municipalities influence each other's infection rates [10]. We use a proxy to quantify mobility, as travel data of people between municipalities is not available for privacy reasons. To determine the most effective proxy for our application, we constructed and evaluated various graphs with different edge configurations and weights. Besides a randomly chosen graph configuration, the edge locations and weights we tested are based on distance between municipalities, distance

in combination with population sizes [40], a fitted gravity model [41], and commuter information [5]. We adopted the last graph configuration for the remainder of this research, since this was found to be the most effective. The data about the place of residence and work of employees is collected by Statistics Netherlands (CBS). In our graph, edges connect municipality pairs with at least 1000 daily commuters. Please note that, limited by the available data, we assume that this measure of mobility remains constant over time. Furthermore, we presume that travelling individuals visit only one region and return to their home region directly afterward.

As a result, we obtain the adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. This matrix indicates which municipalities are connected and how strong these connections are, and is defined as

$$\mathbf{A}_{ij} = \begin{cases} w_{ij} & \text{if } e_{ij} \in \mathcal{E} \text{ and } i \neq j \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where w_{ij} is the normalized weight for each two regions i and j with an existent edge e_{ij} between them.

Node features: Each node or municipality of the graph has a set of associated features \mathcal{F} to provide additional information useful for making more accurate predictions. The node features we include in our developed dataset are:

- *Incidence:* the daily number of reported COVID-19 cases per unit of population [3], [42].
- *Hospitalizations:* the daily number of COVID-19 hospital admissions [3].
- *Virus load in sewage water:* the average concentration of SARS-CoV-2 RNA, converted to the daily amount of sewage water (flow rate) and displayed per 100,000 inhabitants [3].
- *Stringency Index:* a country-wide composite measure indicating the strictness of the applicable COVID-19 control measures, based on nine response indicators including school closures, workplace closures, and travel bans [4], [43].
- *Population density:* the number of inhabitants per square kilometer of land area [44].

The feature selection process started with an initial set of ten features, which was based on data and expertise available within the RIVM. Five features were not included in the dataset due to their relatively lower correlation with the ground truth or their negative impact on the model’s performance. These features include the population size, daily number of reported infections and deaths, vaccination rate, and day of the week.

Each feature is normalized to ensure that all features are on the same scale, ranging from 0 to 1. Except for the static population density, we assume that the above-mentioned node attributes are dynamic and change on a daily basis. Subsequently, it is important to note that at each time step t (in days) the features are represented as vectors that stretch back d days, including day t . This means that the values in this feature vector correspond to days $t+1-d$ to t . Together, these feature vectors are presented in a time dependent feature matrix

$\mathbf{X}^t \in \mathbb{R}^{n \times d \times f}$, where n denotes the number of municipalities and $f = |\mathcal{F}|$ the number of features.

Node labels: The graph has a time-varying associated node label per municipality, which is equal to the ground truth. While the goal of this paper is to forecast the number of reported SARS-Cov-2 infections seven days ahead, we train the model on the normalized incidence of this day instead. The reason for training on incidence is the disproportionate distribution of municipalities’ population sizes, and hence also of their infection numbers. By optimizing on incidence instead of infections, we compensate for the population size and prevent the model from focusing too much on the forecasts for the largest municipalities of the Netherlands. Training on incidence means that our node label vector $\mathbf{y}^{t+7} \in \mathbb{R}^n$ on prediction day $t+7$ is calculated as

$$\mathbf{y}^{t+7} = \text{normalize} \left(\frac{\mathbf{I}^{t+7}}{\mathbf{p}} \right), \quad (2)$$

where $\mathbf{I}^{t+7} \in \mathbb{R}^n$ represents the daily number of reported infections per municipality on day $t+7$ and $\mathbf{p} \in \mathbb{R}^n$ denotes the region’s population sizes [3], [42].

When evaluating the performance of our trained model, we use the data on the observed daily number of reported infections per municipality \mathbf{I}^{t+7} directly. For real-world application, it is particularly important that epidemic forecasting models can properly predict the trend of infection rates within municipalities. Initiating local measures or allocating resources will in practice not occur daily, which allows policymakers to act on a prediction of the infection trend. Therefore, we also evaluate the model’s performance relative to the trend of the observed infection numbers $\mathbf{I}_{\text{trend}}^{t+7} \in \mathbb{R}^n$ instead of these numbers themselves. As a trend, we define the moving average of the observed infection numbers.

B. The Model

Since we would like to capture COVID-19 spread both over time and between municipalities, we propose a spatio-temporal GNN model. Because recurrent neural networks provide a way to extend deep learning to sequential data, we first apply deep learning layers of this type to find temporal patterns in the changing node features of municipalities. Gated recurrent units (GRUs) and long short-term memories (LSTMs) are the most effective sequence models used in practical applications, due to their ability to effectively and accurately retain long-term dependencies in sequential data [45]–[47]. Based on the observation that the GRU outperforms the LSTM for our application, we decided to include this technique in our model.

After applying the GRU and obtaining the resulting node embedding $\mathbf{h}_{\text{GRU}}^t$, we employ a GNN to model the spatial disease transmission patterns caused by the relationship between municipalities. Essentially, the GNN updates each municipality’s node embedding by aggregating the information municipalities receive from their adjacent neighbors. Nowadays, graph attention networks (GATs) as proposed by Veličković et al. [39] are one of the most popular and widespread GNN architectures [48]. Additionally, Brody et

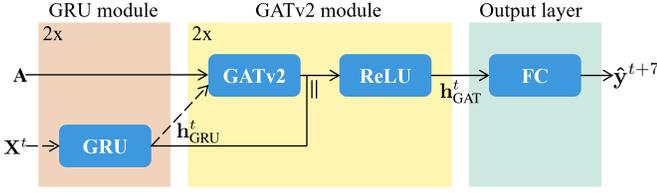


Fig. 2: The proposed architecture is made up of three stages, being the GRU module in orange, the GATv2 module in yellow, and the output layer in green. Both the GRU module and the GATv2 module are repeated once. Therefore, the dashed input arrows X^t and h^t_{GRU} imply that for the second layer of their corresponding module, they are replaced by the output embeddings of the module’s first layer.

al. [2] have proposed the closely related GATv2 recently, which outperforms GAT on all their benchmarks. Based on these results and the performance evaluation on our developed dataset, we decided to use GATv2 to capture the spatial trends in disease transmission. We modify the GATv2 model with skip-connections between layers to avoid diluting the self-node embeddings. Specifically, we concatenate the output of each GATv2 layer with the learned temporal embedding, before we apply the ReLU activation function.

Finally, we need to map the output of the final GATv2 layer h^t_{GAT} to the desired output dimension, for which we use a fully connected (FC) linear output layer.

We visualize our proposed model in Figure 2, and describe each of its elements in more detail below. Please note that we optimized the number of layers of the architecture based on performance, resulting in a model consisting of two GRU modules, two GATv2 modules, and one fully connected output layer.

1) *Gated Recurrent Unit (GRU) module*: The two layer GRU module uses the node feature matrix $X^t \in \mathbb{R}^{n \times d \times f}$ as an input. The goal is to retrieve an output embedding $h^t_{GRU} \in \mathbb{R}^{n \times \text{GRU-dim}}$, which captures the temporal patterns of the dynamic node features over a window of the past d days for each node. This new representation for the municipalities corresponds to the hidden state of the last time step t of the second GRU layer, and is calculated as follows:

$$h^t_{GRU} = \text{GRU}(\text{GRU}(X^t, \mathbf{h}_0), \mathbf{h}_0), \quad (3)$$

where $\mathbf{h}_0 \in \mathbb{R}^{n \times \text{GRU-dim}}$ is a matrix filled with zeros, corresponding to the initial hidden state before the first day of the time window.

2) *Graph Attention Network v2 (GATv2) module*: In the GATv2 module, the temporal node embedding h^t_{GRU} is updated per municipality by the GATv2 mechanism [2] using the update rule

$$h'_i = \sum_{j \in \mathcal{N}(i)} \alpha_{ij} \mathbf{W} h_j, \quad (4)$$

where h'_i is the updated embedding of node i , and h_j is the node representation of node j of the previous layer, initially

being $h^t_{GRU,j}$. We inject the graph structure as represented by adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ into the mechanism by exclusively summing over the nodes j that are part of the set of neighbors of node i , $\mathcal{N}(i)$. We consider node i itself to be part of this neighborhood as well. Furthermore, \mathbf{W} is a learnable weight matrix associated with each node’s linear transformation and α_{ij} is the normalized attention coefficient. This value quantifies the connective strength between the node i and its neighbor j using an attention mechanism

$$\alpha_{ij} = \frac{\exp(\mathbf{a}^\top \text{LeakyReLU}(\mathbf{W} [h_i || h_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\mathbf{a}^\top \text{LeakyReLU}(\mathbf{W} [h_i || h_k]))}, \quad (5)$$

where \mathbf{a}^T is a transposed vector of learnable parameters, and $||$ is the concatenation operation.

To stabilize the self-attention learning process and to improve its expressive power, the GATv2 model uses multi-head attention [2], [39]. This implies that K independent attention mechanisms perform the transformation described above, and subsequently concatenate their outputs to create the final node representation.

The two layered GATv2 module and its resulting spatio-temporal embedding $h^t_{GAT} \in \mathbb{R}^{n \times \text{GAT-dim}}$ can thus be represented as

$$\begin{aligned} h^t &= \text{ReLU}(\text{GATv2}(h^t_{GRU}, \mathbf{A}) || h^t_{GRU}) \\ h^t_{GAT} &= \text{ReLU}(\text{GATv2}(h^t, \mathbf{A}) || h^t_{GRU}). \end{aligned} \quad (6)$$

3) *Output layer*: Finally, we need to map the output of the final hidden layer to the desired output dimension. Therefore, we feed the hidden state obtained by the GATv2 module h^t_{GAT} to a fully connected (FC) linear output layer

$$\hat{y}^{t+7} = \text{FC}(h^t_{GAT}). \quad (7)$$

As a result, we predict the normalized incidence vector $\hat{y}^{t+7} \in \mathbb{R}^n$ containing predictions for each municipality seven days ahead.

C. Optimization

We propose to optimize the model for seven day ahead prediction directly [27], following the pseudocode described in Algorithm 1. We use the mean squared error (MSE) as our loss function

$$\mathcal{L} = \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (y_i^{t+7} - \hat{y}_i^{t+7})^2, \quad (8)$$

since it is commonly used in regression tasks and encourages the model to minimize large errors [49]. Here, n denotes the number of municipalities and T the number of days on which we train the model. Furthermore, y_i^{t+7} refers to the reported infection incidence for municipality i at day $t+7$ and \hat{y}_i^{t+7} to the associated predicted infection incidence as obtained by the model.

Algorithm 1: Training the model

Input: Time series node feature and node label data $\{\mathbf{X}, \mathbf{y}\}$, adjacency matrix \mathbf{A} , learning rate η
Output: Model parameters Θ
Initialize Θ randomly
for each epoch do
 for each timestep t do
 $\mathbf{h}_{\text{GRU}}^t = \text{GRU module}(\mathbf{X}^t)$
 $\mathbf{h}_{\text{GAT}}^t = \text{GATv2 module}(\mathbf{h}_{\text{GRU}}^t, \mathbf{A})$
 $\hat{\mathbf{y}}^{t+7} = \text{Output layer}(\mathbf{h}_{\text{GAT}}^t)$
 error $+$ = $\text{MSE}(\mathbf{y}^{t+7}, \hat{\mathbf{y}}^{t+7})$
 end for
 $\mathcal{L}(\Theta) = \frac{1}{T} \text{error}$
 $\Delta \mathcal{L}(\Theta) = \text{BackProp}(\mathcal{L}(\Theta), \Theta)$
 $\Theta = \Theta - \eta \Delta \mathcal{L}(\Theta)$
end for
return Θ

D. Prediction

We apply the trained model to predict the number of reported COVID-19 cases $\hat{\mathbf{I}}^{t+7} \in \mathbb{R}^n$ in all municipalities on day $t+7$ using the equation

$$\hat{\mathbf{I}}^{t+7} = \mathbf{p} * \text{denormalize}(\hat{\mathbf{y}}^{t+7}). \quad (9)$$

Here, $\mathbf{p} \in \mathbb{R}^n$ is the population size per municipality and $\hat{\mathbf{y}}^{t+7} \in \mathbb{R}^n$ the predicted incidence, which is converted back to its original scale.

IV. EXPERIMENTS

This section introduces the datasets we used in Section IV-A, refers to our comparison methods in Section IV-B, describes the evaluation metrics used in Section IV-C, and explains the hyperparameters and implementation details in Section IV-D.

A. The Dataset

We use open data from three sources: the COVID-19 dataset of the National Institute for Public Health and the Environment (RIVM) [3], data from Statistics Netherlands (CBS) [5], [42], [44], and the Oxford COVID-19 Government Response Tracker [4], [43]. We fuse these datasets to obtain the necessary graph edge, node feature, and node label information.

While the first COVID-19 case in the Netherlands was discovered on February 27, 2020, our study period is from June 1, 2020, until April 10, 2022. In the Netherlands, testing capacity was limited before June 1, 2022, so the number of positive tests did not reflect the actual number of SARS-Cov-2 infections. The same applies to all days after April 10, 2022, because from then on, the advice to confirm a positive self-test at the Dutch Municipal Health Service was canceled, except for specific target groups such as healthcare workers [50], [51].

To ensure real-world application of our method, we chose to divide the study period into shorter time intervals of 61 days. By training on the first 53 and validating on the subsequent 7

TABLE III: Time period characteristics

	Variant	Prediction period	Avg new cases total (train/val/test)
W1	Wild type	11/06/2020 - 10/08/2020	0.5 (0.4/1.4/2.3)
W2		15/07/2020 - 13/09/2020	1.5 (1.2/3.0/3.2)
W3		18/08/2020 - 17/10/2020	7.4 (5.4/20.7/23.5)
W4		20/09/2020 - 19/11/2020	17.8 (18.1/15.2/16.5)
W5		23/10/2020 - 22/12/2020	21.3 (19.6/32.7/28.5)
W6		25/11/2020 - 24/01/2021	21.7 (22.6/15.4/14.2)
A1	Alpha	04/03/2021 - 03/05/2021	19.7 (19.6/19.2/26.6)
A2		14/04/2021 - 13/06/2021	14.7 (16.3/4.3/3.0)
D1	Delta	15/07/2021 - 13/09/2021	10.3 (10.8/6.9/5.0)
D2		17/08/2021 - 16/10/2021	6.6 (6.2/8.8/10.7)
D3		19/09/2021 - 18/11/2021	18.7 (13.8/48.7/68.0)
D4		21/10/2021 - 20/12/2021	44.1 (44.6/41.1/35.0)
O1	Omicron	13/01/2022 - 14/03/2022	178.5 (177.7/190.9/133.7)
O2		09/02/2022 - 10/04/2022	124.6 (138.5/34.4/20.6)

days of each interval, we guarantee that our model can be used for prediction on test day 61. Because there have been only four widespread COVID-19 variants (i.e. the wild type, alpha, delta, omicron) [52], it is difficult for the model to capture the differences between them. Therefore, we make sure that in each time period at least 75% of all test samples of SARS-Cov-2 infected persons that are sequenced by the pathogen surveillance consist of the same variant [53].

Although users would in real life retrain the model every day to make a prediction based on the most recent data, we chose to predict for 14 days in this paper. This allows us to balance the need for computational efficiency with the ability to examine differences in performance over time and between variants. We present the associated 14 time periods and their characteristics in Table III.

In Figure 3, we visualize the split in training, validation, and test data by means of a plot of the aggregated number of infections in the Netherlands over period **D3**, corresponding to a time when the delta variant was dominant, and the infection numbers were increasing. Once the model is trained, day t is the day up until which data is used to make $t+7$ predictions for the test day on November 18, 2021. In Appendix A, these plots are presented for the other periods as well.

B. Comparison Methods

Our approach involves applying the proposed method on the newly developed COVID-19 dataset, which is why no baselines yet exist to compare our model to. Therefore, we implemented various simple and commonly used baselines [8], [21]. According to these baselines, the prediction for day $t+7$ is equal to:

- 1) *PD*: the number of reported cases on day t .
- 2) *HA*: the historical average of the number of reported cases up to and including day t .
- 3) *HA_{window}*: the historical average of the number of reported cases in observation window d (day $t+1-d$ to day t).

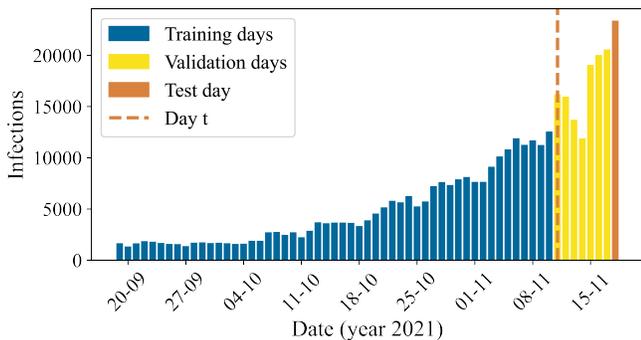


Fig. 3: The aggregated number of reported COVID-19 cases in the Netherlands over the prediction days of time period **D3**, with the days included in the training set in blue, the validation set in yellow, and the test set in orange. Once the model is trained, we use data up to and including day t (dashed orange line) to predict infections on the test day at $t+7$.

- 4) *EB*: the number of reported cases on day t multiplied with λ ($\hat{y}^{t+7} = \lambda y^t$), where λ is determined based on the infection rates of observation window d .

In order to gain a better understanding of the proposed spatio-temporal GNN’s behavior, we perform an ablation study. We investigate the influence of the GRU and GATv2 modules on the performance of the developed model by removing them one at a time, resulting in the following architectures:

- 1) *GRU*: A temporal model consisting of the GRU module and the output layer.
- 2) *GATv2*: A spatial model consisting of the GATv2 module and the output layer.

C. Evaluation Metrics

To quantify the difficulty of making predictions for the developed dataset, we propose the fluctuation size per municipality as a metric that measures the fluctuations in the data. The fluctuation size Fl_i of municipality i is calculated as follows:

$$Fl_i = \frac{1}{T-1} \sum_{t=1}^{T-1} |\tilde{I}_i^{t+1} - \tilde{I}_i^t|, \quad (10)$$

where T is equal to the number of time steps in the considered time period, and \tilde{I}_i^{t+1} and \tilde{I}_i^t are the normalized observed number of reported infections in the municipality i at times $t+1$ and t .

The objective of this study is to predict the number of SARS-Cov-2 infections as accurately as possible, which is defined by the error between the observed infections $\mathbf{I}^{t+7} \in \mathbb{R}^n$ and predicted infections $\hat{\mathbf{I}}^{t+7} \in \mathbb{R}^n$. Lower errors indicate higher prediction accuracy. We especially aim to minimize relatively large errors, as these can result in inaccurate control measures and resource allocation. Therefore, we adopt the root

TABLE IV: Hyperparameters

Parameter	Value(s) / choice of algorithm
epochs	early stopping: minimum 100 and maximum 500 epochs 50 epochs of patience
optimizer	Adam
learning rate	5e-3
batch size	344
hidden dimensions	[32, 32, 64, 64]
# of GATv2 heads	[2, 2]

mean squared error (RMSE) [54] over all municipalities on each test day as an evaluation metric.

Besides the RMSE, we also introduce a scale independent metric, the coefficient of determination (R^2) [54]. This metric assesses the correlation between observed and predicted number of infections. A R^2 close to one suggests a perfectly accurate prediction, while a R^2 close to or below zero indicates that the model fails to make accurate predictions. Please note that we also use the RMSE and R^2 metrics when evaluating the performance of the model relative to the trend of the observed infection numbers $\mathbf{I}_{\text{trend}}^{t+7} \in \mathbb{R}^n$.

When evaluating the forecast of the model for an individual municipality on a specified day, we would like to use an evaluation metric that is easy to interpret. Therefore, we use the absolute error between the municipality’s observed number of infections and the predicted number of infections.

D. Hyperparameter Setting and Implementation Details

In Table IV, we present the hyperparameters that have been used for the final model. The mentioned hidden dimensions correspond to the two GRU layers and the two GATv2 modules, and the number of independent attention heads K to the two GATv2 layers. With early stopping, we store the parameters of the model that achieved the highest validation accuracy, and then retrieve it to make predictions about the test samples. The model is implemented with Pytorch [55] and Pytorch Geometric [56]. For all our experiments, we use a look-back window d of 7 days, and the aforementioned prediction horizon h of 7 days.

V. RESULTS

In this section, we present the obtained results, where Section V-A focuses on the dataset and Section V-B on the model.

A. The dataset

To evaluate the proposed approach on the developed COVID-19 dataset, we first determine the difficulty of making predictions for the developed dataset by calculating the fluctuation sizes Fl for our time periods. Taking time period **D3** as an example, we see that during this period of time the fluctuation size for the Netherlands as a whole is equal to 0.031, while the average over the individual municipalities is equal to 0.098, meaning that the fluctuation of individual municipalities is on

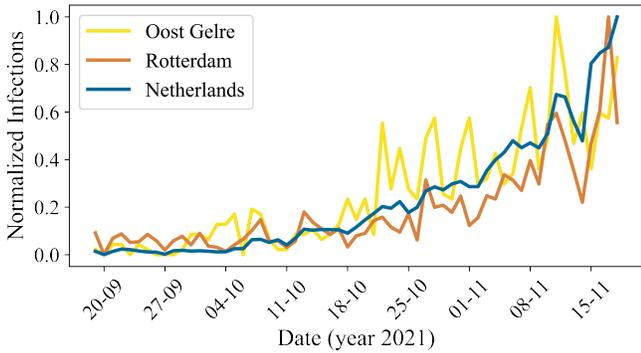


Fig. 4: The normalized actual number of reported COVID-19 cases over time in the Netherlands for period **D3**. The blue line corresponds to the total number of infections in the Netherlands, the yellow line to the municipality *Oost Gelre*, and the orange line to the municipality *Rotterdam*.

average 3.2 times as large as of the Netherlands. Averaged over all time periods, the mean fluctuation size of individual municipalities is 2.3 times as large as of the Netherlands as a whole.

We present a visualization of the normalized actual number of reported infections and its fluctuations during time period **D3** for the whole of the Netherlands and for two example municipalities, *Oost Gelre* and *Rotterdam*, in Figure 4. Here, the fluctuations in *Oost Gelre* are 3.8 times as large, and for *Rotterdam* 2.5 times as large as in the Netherlands.

B. The Model

In Table V and Table VI, we report the RMSE and the R^2 performance metrics for the prediction of our spatio-temporal GNN model, its GRU and GATv2 components, and the baselines for the different time periods of our dataset. Here, the numbers in bold indicate which of these concepts perform best on the test day of each time period, and underlined numbers present the corresponding best baseline. It is notable that both the performance of our proposed model and the baselines vary widely across the different time periods. The proposed model’s minimum coefficient of determination is 0.004 for the first omicron period **O1**, which indicates that the model fails to make accurate predictions during this time period. Meanwhile, the R^2 of 0.934 for the last delta time period **D4** suggests that the model explains as much as 93.4% of the variance of our forecasts.

In Table VII and Table VIII, we report the average performances over all 14 time periods (Avg), and over the four COVID-19 variants being the wild type (WT), alpha (α), delta (δ), and omicron (\omicron). It is not fair to compare the RMSE over the time periods, because the interpretation of a given RMSE is highly dependent on the predicted number of infections. For example, a RMSE of 10 infections while there are 200 infections predicted is not important for policymaking and resource relocation, as opposed to a RMSE of 10 infections with a prediction of 20 infections. Therefore, we only report

the RMSE relative to the baselines and its parts in Table VII. Here, we notice that averaged over all time periods, the RMSE of our model is -9.8% relative to the best performing baseline *PD*, meaning that on average our proposed model achieves a 9.8% lower RMSE than this baseline. Looking at Table VIII, we find that the average coefficient of determination of the proposed model 0.696, which is 3.2% higher than of the best performing baseline *PD*. These results show that on average, our model can conduct more accurate predictions than the baselines, and that the relationship between the actual values and our model’s predictions accounts for 69.6% of the variation in the predictions.

By looking at the performance of the model per COVID-19 variant in Table VII and Table VIII, we notice that our proposed model achieves a lower RMSE and a higher R^2 than all the baselines for the wild type, alpha, and delta variant. Looking at the coefficient of determination, we see that relative to the *PD* baseline, this value increases with 8.7% for the wild type, 8.2% for the alpha variant, and 6.1% for the delta variant. For these variants, we note that the average coefficient of determination is 0.795. Meanwhile, we notice that the performance of the proposed model is worse than all baselines for the omicron variant.

1) *The GRU and GATv2 modules*: Table V and Table VI also contain the RMSE and R^2 evaluation metrics for the exclusively temporal GRU model and the solely spatial GATv2 model. As with our proposed model and the baselines, there is a large variation between the time periods. While the R^2 of the *GRU* and the *GATv2* are both below zero for the second omicron period **O2**, the *GRU* achieves a R^2 of 0.917 for period **A1** and the *GATv2* a R^2 of 0.910 for period **D4**.

Table VII and Table VIII show that on average the RMSE of our spatio-temporal model is 12.9% lower, and the coefficient of determination 8.6% higher than of the temporal GRU model. Similarly, the model’s RMSE is 15.4% lower than the *GATv2* model, and the corresponding R^2 score is 8.2% higher. For our application, the performance of the proposed spatio-temporal GNN is thus higher than of the temporal *GRU* model or the spatial *GATv2*.

2) *Individual municipalities*: To further evaluate the performance of our model on the developed COVID-19 dataset, we also look at our forecasts for individual municipalities. Hence, we visualize the correlation between the observed and the predicted number of reported SARS-Cov-2 infections of period **D3** in Figure 5, which we present for the other time periods in Appendix B. The blue dots indicate the November 18, 2021, forecasts for all 344 municipalities.

It is apparent from the plot that there is a high number of municipalities with low infection rates. Accordingly, there are more small municipalities than large ones in the Netherlands. The dots of most of the municipalities are quite close to the 45 degree yellow line through the origin, which corresponds to the high coefficient of determination of 0.893 that we presented for period **D3** in Table VI. In general, the model thus appears to be both accurate and precise. However, for the municipalities with less than 400 infections, the mean of

TABLE V: The root mean squared errors (RMSE) of the developed model, the *GRU* and *GATv2* components, and the baselines for all 14 time periods. Numbers in bold indicate the lowest RMSE per time period, underlined numbers present the corresponding best performing baseline.

RMSE (\downarrow)	W1	W2	W3	W4	W5	W6	A1	A2	D1	D2	D3	D4	O1	O2
<i>PD</i>	<u>8.04</u>	4.97	<u>13.59</u>	15.57	21.52	10.02	22.39	4.49	4.34	<u>8.16</u>	<u>39.08</u>	17.88	139.81	28.4
<i>HA</i>	10.34	8.08	35.95	18.31	18.41	13.97	20.06	23.29	18.84	8.36	79.28	<u>15.16</u>	93.44	203.69
<i>HA_{window}</i>	8.36	5.98	18.11	<u>12.39</u>	13.08	<u>7.10</u>	<u>13.75</u>	7.82	6.53	8.34	52.20	26.23	96.70	71.49
<i>EB</i>	8.32	5.27	14.44	15.97	23.16	9.74	23.75	<u>4.07</u>	<u>4.31</u>	9.12	39.73	17.47	159.33	<u>24.51</u>
<i>GRU</i>	10.55	8.25	18.79	19.54	17.94	6.93	12.64	3.10	4.87	8.40	32.11	20.65	175.90	55.25
<i>GATv2</i>	9.68	8.20	35.23	10.19	15.23	9.88	20.44	5.28	4.07	7.84	36.02	15.5	133.2	59.97
<i>Model</i>	7.03	5.76	13.02	11.19	16.12	7.92	17.10	4.29	4.60	6.91	26.56	13.22	177.10	30.31

TABLE VI: The coefficients of determination of the developed model, the *GRU* and *GATv2* components, and the baselines for all 14 time periods. Numbers in bold indicate the highest R^2 per time period, underlined numbers present the corresponding best performing baseline.

R^2 (\uparrow)	W1	W2	W3	W4	W5	W6	A1	A2	D1	D2	D3	D4	O1	O2
<i>PD</i>	<u>0.560</u>	0.860	<u>0.926</u>	0.566	0.664	0.723	0.739	0.386	0.849	<u>0.702</u>	<u>0.769</u>	0.880	0.379	0.23
<i>HA</i>	0.274	0.630	0.482	0.400	0.754	0.461	0.790	<0	<0	0.688	0.048	<u>0.914</u>	0.723	<0
<i>HA_{window}</i>	0.525	0.797	0.869	<u>0.725</u>	0.876	<u>0.861</u>	<u>0.901</u>	<0	0.659	0.690	0.587	0.741	0.703	<0
<i>EB</i>	0.529	0.842	0.916	0.544	0.611	0.738	0.706	<u>0.497</u>	<u>0.852</u>	0.628	0.761	0.885	0.194	0.427
<i>GRU</i>	0.243	0.614	0.859	0.317	0.766	0.867	0.917	0.708	0.810	0.685	0.844	0.840	0.018	<0
<i>GATv2</i>	0.362	0.618	0.503	0.814	0.832	0.730	0.782	0.151	0.867	0.725	0.803	0.910	0.437	<0
<i>Model</i>	0.664	0.812	0.932	0.776	0.811	0.827	0.848	0.441	0.831	0.787	0.893	0.934	0.004	0.123

TABLE VII: The average relative RMSE prediction performance of our model with respect to its parts and the baselines for each COVID-19 variant and in total.

RMSE (\downarrow)	WT	α	δ	o	Avg
<i>PD</i>	-12.5%	-14.0%	-16.9%	16.7%	-9.8%
<i>HA</i>	-36.5%	-48.2%	-43.1%	2.2%	-34.5%
<i>HA_{window}</i>	-3.8%	-10.4%	-36.4%	12.8%	-11.7%
<i>EB</i>	-15.8%	-11.3%	-18.7%	17.4%	-11.3%
<i>GRU</i>	-22.1%	36.8%	-19.1%	-22.2%	-12.9%
<i>GATv2</i>	-20.7%	-17.5%	-10.0%	-8.3%	-15.4%

TABLE VIII: The average coefficient of determination of our model, its parts, and the baselines for each COVID-19 variant and in total. Numbers in bold indicate the highest R^2 per variant, underlined numbers present the corresponding best performing baseline.

R^2 (\uparrow)	WT	α	δ	o	Avg
<i>PD</i>	0.717	0.563	<u>0.800</u>	0.305	<u>0.660</u>
<i>HA</i>	0.500	0.395	0.413	0.362	0.440
<i>HA_{window}</i>	<u>0.776</u>	0.451	0.669	0.352	0.638
<i>EB</i>	<u>0.697</u>	<u>0.602</u>	0.782	0.311	0.606
<i>GRU</i>	0.611	0.813	0.795	0.009	0.606
<i>GATv2</i>	0.643	0.467	0.826	0.219	0.610
<i>Model</i>	0.804	0.645	0.861	0.064	0.692

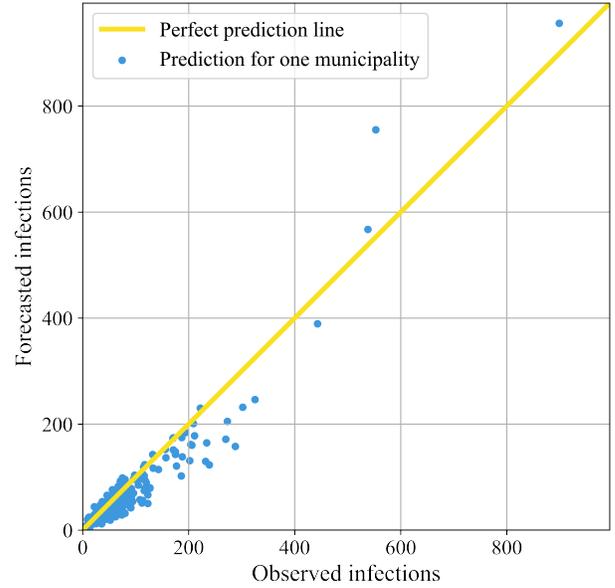


Fig. 5: The correlation between the forecasted and the observed number of reported COVID-19 cases in period *D3*. Here, each blue dot corresponds to the forecast for one municipality November 18, 2021. When a dot lies on the yellow diagonal line, it means that the prediction is perfect.

TABLE IX: The absolute error for the proposed model and its baselines for the best performing municipalities for prediction on 18/11/2021 (period **D3**).

	Oost Gelre	Meerijstad	Opsterland
# infections	39	96	42
<i>PD</i>	8.0	27.0	14.0
<i>HA</i>	28.2	69.6	30.7
<i>HA</i> _{window}	14.0	29.6	13.0
<i>EB</i>	15.8	28.3	26.6
<i>Our model</i>	0.2	0.2	0.2

TABLE X: The absolute error for the proposed model and its baselines for the worst performing municipalities for prediction on 18/11/2021 (period **D3**)

	Rotterdam	Maastricht	Heerlen
# infections	553	288	239
<i>PD</i>	36.0	198.0	144.0
<i>HA</i>	351.1	245.6	195.0
<i>HA</i> _{window}	148.6	184.4	141.7
<i>EB</i>	84.4	200.0	144.1
<i>Our model</i>	202.2	130.0	115.9

the dots is in an increasing manner slightly below the perfect prediction line, indicating that the accuracy is not perfect and that the model increasingly underestimates the number of infections. There are only four municipalities with more than 400 observed infections, being the four biggest municipalities of the Netherlands; *Amsterdam*, *Rotterdam*, *'s-Gravenhage*, and *Utrecht*. In contrast to the smaller municipalities, their predictions are on average above the actual infections.

As an illustration of the variation between municipalities, Table IX and Table X present the absolute difference between the predicted and the actual number of infections for individual municipalities on November 18, 2021. Here, Table IX shows the three municipalities with the smallest error, being *Oost Gelre*, *Meerijstad*, and *Opsterland*, and Table X the three worst performing municipalities, being *Rotterdam*, *Maastricht*, and *Heerlen*. As we have seen before with the variation between the time periods, it appears here that the variation on the same day between municipalities is also large, with a minimum error of approximately zero infections in *Oost Gelre*, *Meerijstad*, and *Opsterland*, and a maximum error of 202 infections in Rotterdam. Relatively speaking, the differences are also large. Indeed, the error relative to the number of infections is below 1% for Oost Gelre, Meerijstad and Opsterland, while it is 36.6% for Rotterdam, 45.2% for Maastricht and 48.5% for Heerlen.

We present plots corresponding to the predictions of these municipalities in Figure 6. Please note that, next to our proposed model and the actual values, we chose to visualize only the best performing baseline in each prediction for clarity. It is noticeable here that the data contains a lot of noise due to the rapid fluctuations. Therefore, it seems that while our model often captures the right trend, it is unable to interpret

TABLE XI: The performance of the proposed model for trend predictions compared to actual value predictions.

	WT	α	δ	\omicron	Avg
RMSE (\downarrow)	-31.4%	-40.8%	-26.3%	-20.7%	-29.8%
R^2 (\uparrow)	9.4%	18.6%	7.3%	49.8%	15.9%

TABLE XII: The relative prediction performance for $T+7$ with respect to $T+1$

	WT	α	δ	\omicron	Avg
RMSE (\downarrow)	27.0%	14.9%	9.3%	125.7%	34.3%
R^2 (\uparrow)	-5.9%	-10.5%	-2.1%	-78.1%	-15.3%

these rapid changes.

3) *The trend*: As explained in before, we would also like to analyze the performance of our proposed model for infection trend prediction. We present an example of the actual infection curve and the trend for Rotterdam on November 18, 2021, in Figure 7.

If we evaluate our model against the trend rather than actual infection values, the performance improves significantly. Where the average coefficient of determination for actual infection prediction is 0.692, it increases to 0.851 for prediction of the trend, indicating a 15.9% gain. For each variant specifically, the coefficients of determination for trend prediction are 0.898 for the wild type, 0.831 for alpha, 0.934 for delta, and 0.562 for omicron. We give an overview of the improvements of the trend predictions compared to the actual value predictions of our model in Table XI. It is evident that, across all variants and on average, our model exhibits superior performance in predicting the trend of infection rates compared to the highly variable actual rates.

4) *The forecasting horizon*: We would like to understand if our varying performance is caused by the prediction horizon h of 7 days, or whether the cause is in our dataset or model itself. Therefore, we also made $t+1$ forecasts for the exact same days as with the $t+7$ forecasts. For these $t+1$ forecasts, the coefficients of determination are 0.845 on average, 0.863 for the wild type, 0.750 for alpha, 0.882 for delta, and 0.845 for omicron. We present the relative performances of our 7-day multistep prediction with respect to the performance of the model when predicting only one day ahead in Table XII. Comprehensive results of the one-day ahead forecast can be found in Appendix C.

As presented earlier in literature [17], [21], [27], we share the observation that for both the individual COVID-19 variants and the average, the performance of the model decreases as the prediction horizon increases. On average, the RMSE increases with 34.3%, and the coefficient of determination decreases with 15.3%. By looking at the individual variants, we particularly notice that the results of our model for the omicron variant deteriorate relatively much, with a RMSE increase of 125.7% and a coefficient of determination decrease of 78.1%.

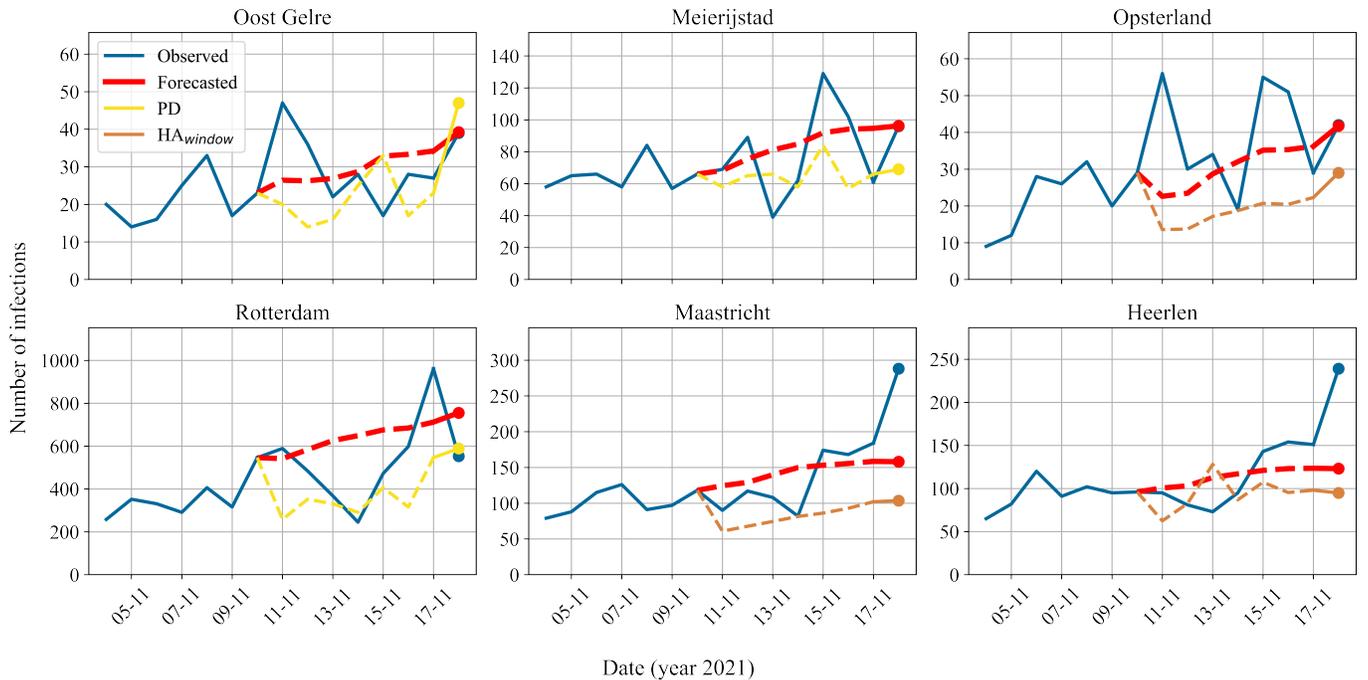


Fig. 6: The actual and predicted reported number of infections for individual municipalities on 11/18/2021 (period **D3**). Based on the absolute difference between the prediction and the actual value, the upper row corresponds to the best performing municipalities (*Oost Gelre*, *Meerijstad*, *Opsterland*) and the bottom row to the worst performing municipalities (*Rotterdam*, *Maastricht*, and *Heerlen*). The blue line indicates the actual number of infections reported, the red line the model’s predictions, and either the yellow or the orange line the predictions of the baselines. Note that for clarity of illustration, the plots only visualize the best performing baseline in each prediction. In the plots, the dashed lines represent predictions in the validation set, and the dots represent the prediction on test day 18/11/2021.

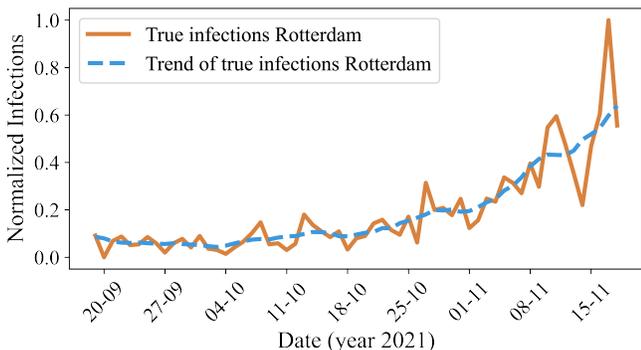


Fig. 7: The normalized actual number of reported COVID-19 cases (orange) and its trend (blue) over time in *Rotterdam* for period **D3**.

VI. DISCUSSION

The variation in results as demonstrated in Section V-B, can be explained by the observation that the daily numbers of reported SARS-Cov-2 infections per municipality have a high fluctuation size. We also note that there are significant differences in the data between the chosen time periods and

between municipalities, such as the order of magnitude of the infection rates and the dynamic node features (e.g. the virus load in sewage water), and the shape of the infection curve. Between time periods, the variants circulating vary as well. Additionally, there are variations between municipalities in population size, population density, and the number of neighbors. These aspects of the dataset make it challenging to capture the resulting complex relationships, and to obtain accurate predictions for all time period and municipality combinations.

Despite these challenges, the average performance of our model in predicting the spread of SARS-CoV-2 at the municipal level in the Netherlands is strong. We observed that the performance of the proposed spatio-temporal GNN outperforms its temporal and spatial components, proving the effectiveness of modelling the spreading of epidemic diseases as a spatio-temporal problem. Nevertheless, we hypothesize that if related data would be available, the performance of the spatial GATv2 module could be enhanced through the use of a more representative strategy for selecting edge locations and weights.

As discussed in Section V-B, both our model and the baselines struggle to make accurate predictions for the omicron variant. This suggests that omicron is a particularly chal-

lenging variant to predict. This may be due to the fact that omicron is more contagious and produces milder symptoms compared to earlier variants, leading to more rapid changes in the number of cases [57]. Another potential explanation could be the occurrence of natural immunity in individuals due to the high number of infections during the omicron period [3], leading to a decrease in transmission during this period as a result of the increased protection.

In order to improve the overall prediction performance (including that of the omicron variant), a model capable of capturing more complex relationships would be required. To obtain such models, we either increased the number of GRU layers, the number of GATv2 layers, or the hidden dimension sizes. Evaluation of these models demonstrated that utilizing these adapted models with the limited available data a decrease in performance. To effectively utilize a more complex model, an increased amount of data, such as synthetic data or data from other countries, would be necessary.

In future work, it may also be worth considering a larger regional scale, such as the 25 security regions of the Netherlands instead of the 344 municipalities. This approach could potentially decrease the fluctuation size and complexity of relationships, making it easier for the model to capture the relationship of infection numbers over time and between regions, while still allowing for the implementation of regional measures or resource distribution.

VII. CONCLUSION

In this paper, we introduced a novel model architecture for forecasting reported SARS-Cov-2 infections one week ahead using a spatio-temporal GNN. We applied our model to a newly introduced COVID-19 dataset containing municipality-level information for the Netherlands, including COVID-19 statistics, demographic features, and municipality interaction data.

Our model demonstrates strong performance in predicting the spread of SARS-Cov-2 at the Dutch municipal level for the wild type, alpha, and delta variants. The model outperforms the baselines and the exclusively spatial and temporal models. Additionally, with an average coefficient of determination of 0.795 and the ability to capture the trend of the disease with a R^2 score of 0.899, we conclude that the model is well-suited for predicting the spread of these variants in real-world applications, allowing policymakers to anticipate and take necessary action in response to the pandemic. However, for more complex disease dynamics, such as those exhibited by the omicron variant, we recommend using more (synthetic) data or expanding the regional forecasting scale. In real-world applications, we recommend using our proposed model in combination with a mechanistic model in an ensemble approach to enhance long-term predictions and to improve prediction accuracy and robustness.

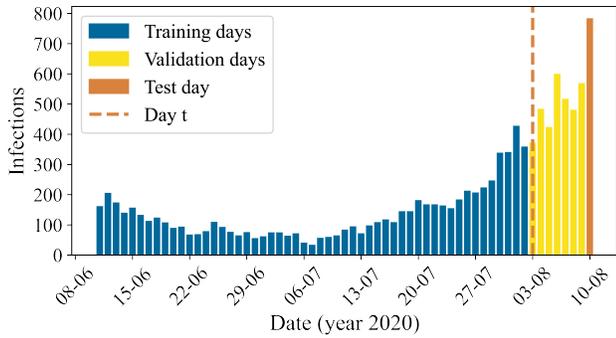
REFERENCES

- [1] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," 6 2014. [Online]. Available: <http://arxiv.org/abs/1406.1078>
- [2] S. Brody, U. Alon, and E. Yahav, "How Attentive are Graph Attention Networks?" 5 2021. [Online]. Available: <http://arxiv.org/abs/2105.14491>
- [3] Rijksinstituut voor Volksgezondheid en Milieu, "COVID-19 dataset," 2022. [Online]. Available: <https://data.rivm.nl/covid-19/>
- [4] T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, and H. Tatlow, "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)," *Nature Human Behaviour*, vol. 5, no. 4, pp. 529–538, 4 2021.
- [5] Centraal Bureau voor de statistiek, "Banen van werknemers naar woon- en werkregio." 2020. [Online]. Available: <https://opendata.cbs.nl//CBS/nl/dataset/83628NED/table?dl=489D>
- [6] World Health Organization, *Managing epidemics: key facts about major deadly diseases*. Geneva: World Health Organization, 2018.
- [7] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," pp. 533–534, 5 2020.
- [8] A. Kapoor, X. Ben, L. Liu, B. Perozzi, M. Barnes, M. Blais, and S. O'Banion, "Examining COVID-19 forecasting using spatio-temporal graph neural networks." *arXiv preprint arXiv:2007.03113*, 2020.
- [9] B. Jhun, "Effective vaccination strategy using graph neural network ansatz," *arXiv preprint arXiv:2111.00920*, 2021. [Online]. Available: <http://arxiv.org/abs/2111.00920>
- [10] M. Schoot Uiterkamp, M. Gösgens, H. Heesterbeek, R. Van Der Hofstad, and N. Litvak, "The role of inter-regional mobility in forecasting SARS-CoV-2 transmission," *Journal of the Royal Society Interface*, vol. 19, no. 193, 8 2022.
- [11] W. Kermack and A. McKendrick, "A Contribution to the Mathematical Theory of Epidemics," *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, vol. 115, no. 772, pp. 700–791, 1927.
- [12] M. J. Keeling and K. T. Eames, "Networks and epidemic models," *Journal of the Royal Society Interface*, vol. 2, no. 4, pp. 295–307, 2005.
- [13] S. Kandula and J. Shaman, "Near-term forecasts of influenza-like illness. An evaluation of autoregressive time series approaches.," *Epidemics*, vol. 27, pp. 41–51, 2019.
- [14] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos, Solitons and Fractals*, vol. 135, 6 2020.
- [15] V. La Gatta, V. Moscato, M. Postiglione, and G. Sperli, "An epidemiological neural network exploiting dynamic graph structured data applied to the COVID-19 outbreak," *IEEE Transactions on Big Data*, vol. 7, no. 1, pp. 45–55, 2021.
- [16] R. Hinch, W. J. Probert, A. Nurtay, M. Kendall, C. Wymant, M. Hall, K. Lythgoe, A. Bulas Cruz, L. Zhao, A. Stewart, L. Ferretti, D. Montero, J. Warren, N. Mather, M. Abueg, N. Wu, O. Legat, K. Bentley, T. Mead, K. Van-Vuuren, D. Feldner-Busztin, T. Ristori, A. Finkelstein, D. G. Bonsall, L. Abeler-Dörner, and C. Fraser, "OpenABM-Covid19 - An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing," *PLoS Computational Biology*, vol. 17, no. 7, 2021.
- [17] J. Gao, R. Sharma, C. Qian, L. M. Glass, J. Spaeder, J. Romberg, J. Sun, and C. Xiao, "STAN: spatio-temporal attention network for pandemic prediction using real-world evidence," *Journal of the American Medical Informatics Association*, vol. 28, no. 4, pp. 733–743, 4 2021.
- [18] A. Adiga, D. Dubhashi, B. Lewis, M. Marathe, S. Venkatramanan, and A. Vullikanti, "Mathematical Models for COVID-19 Pandemic: A Comparative Analysis," *Journal of the Indian Institute of Science*, vol. 100, no. 4, pp. 793–807, 2020.
- [19] Z. Zhang, P. Cui, and W. Zhu, "Deep Learning on Graphs: A Survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 249–270, 1 2022.
- [20] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 1 2021.
- [21] G. Panagopoulos, G. Nikolentzos, and M. Vazirgiannis, "Transfer Graph Neural Networks for Pandemic Forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 6, pp. 4838–4845, 2021.
- [22] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.

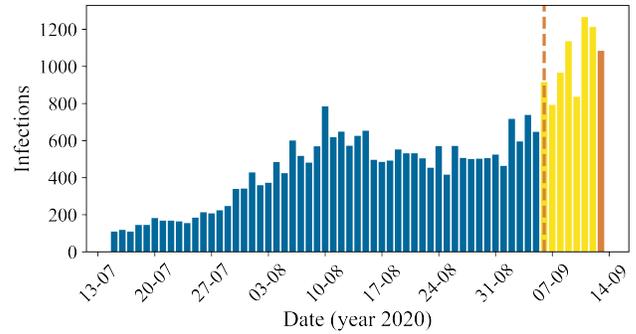
- [23] Z. Liu and J. Zhou, *Introduction to Graph Neural Networks*, R. Brachman, F. Rossi, and P. Stone, Eds. Morgan & Claypool, 2020, vol. 14, no. 2.
- [24] J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackerman, V. M. Tran, A. Chiappino-Pepe, A. H. Badran, I. W. Andrews, E. J. Chory, G. M. Church, E. D. Brown, T. S. Jaakkola, R. Barzilay, and J. J. Collins, “A Deep Learning Approach to Antibiotic Discovery,” *Cell*, vol. 180, pp. 688–702, 2 2020.
- [25] A. Derrow-Pinion, J. She, D. Wong, O. Lange, T. Hester, L. Perez, M. Nunkesser, S. Lee, X. Guo, B. Wiltshire, P. W. Battaglia, V. Gupta, A. Li, Z. Xu, A. Sanchez-Gonzalez, Y. Li, and P. Veličković, “ETA Prediction with Graph Neural Networks in Google Maps,” in *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, 2021, pp. 3767–3776.
- [26] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, “Graph convolutional neural networks for web-scale recommender systems,” *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 974–983, 2018.
- [27] S. Deng, S. Wang, H. Rangwala, L. Wang, and Y. Ning, “Graph Message Passing with Cross-location Attentions for Long-term ILI Prediction,” *arXiv preprint arXiv:1912.10202*, 2019. [Online]. Available: <http://arxiv.org/abs/1912.10202>
- [28] C. Murphy, E. Laurence, and A. Allard, “Deep learning of contagion dynamics on complex networks,” *Nature Communications*, vol. 12, no. 1, pp. 1–11, 2021.
- [29] Z. Li, X. Luo, B. Wang, A. L. Bertozzi, and J. Xin, “A Study on Graph-Structured Recurrent Neural Networks and Sparsification with Application to Epidemic Forecasting,” in *world congress on global optimization*. Springer, 2019, pp. 730–739. [Online]. Available: <http://arxiv.org/abs/1902.05113>
- [30] S. Mežnar, N. Lavrač, and B. Škrj, “Prediction of the Effects of Epidemic Spreading with Graph Neural Networks,” in *International Conference on Complex Networks and Their Applications*. Springer, 12 2020, pp. 420–431.
- [31] Y. Wang, D. D. Zeng, Q. Zhang, P. Zhao, X. Wang, Q. Wang, Y. Luo, and Z. Cao, “Adaptively temporal graph convolution model for epidemic prediction of multiple age groups,” *Fundamental Research*, 2021.
- [32] A. Tomy, M. Razzanelli, F. Di Lauro, D. Rus, and C. Della Santina, “Estimating the State of Epidemics Spreading with Graph Neural Networks,” 5 2021. [Online]. Available: <http://arxiv.org/abs/2105.05060>
- [33] G. Cutura, B. Li, A. Swami, and S. Segarra, “Deep demixing: Reconstructing the evolution of epidemics using graph neural networks,” *arXiv preprint arXiv:2011.09583*, 11 2020. [Online]. Available: <http://arxiv.org/abs/2011.09583>
- [34] C. Shah, N. Dehmamy, N. Perra, M. Chinazzi, A.-L. Barabási, A. Vespignani, and R. Yu, “Finding patient zero: Learning contagion source with graph neural networks,” *arXiv preprint arXiv:2006.11913*, 2020. [Online]. Available: <http://arxiv.org/abs/2006.11913>
- [35] S. Song, Z. Zong, Y. Li, X. Liu, and Y. Yu, “Reinforced Epidemic Control: Saving Both Lives and Economy,” *arXiv preprint arXiv:2008.01257*, 8 2020. [Online]. Available: <http://arxiv.org/abs/2008.01257>
- [36] E. A. Meiron, H. Maron, S. Mannor, and G. Chechik, “How to Stop Epidemics: Controlling Graph Dynamics with Reinforcement Learning and Graph Neural Networks,” 10 2020. [Online]. Available: <http://arxiv.org/abs/2010.05313>
- [37] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, “Neural Message Passing for Quantum Chemistry,” in *Proceedings of the 34th International Conference on Machine Learning*, Sydney, 2017.
- [38] T. N. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” 9 2016. [Online]. Available: <http://arxiv.org/abs/1609.02907>
- [39] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph Attention Networks,” 10 2017. [Online]. Available: <http://arxiv.org/abs/1710.10903>
- [40] J. M. Barrios, W. W. Verstraeten, P. Maes, J. M. Aerts, J. Farifteh, and P. Coppin, “Using the gravity model to estimate the spatial spread of vector-borne diseases,” *International Journal of Environmental Research and Public Health*, vol. 9, no. 12, pp. 4346–4364, 2012.
- [41] G. Lagerweij, “Spatial / prediction model,” Tech. Rep., 2021.
- [42] Centraal Bureau voor de Statistiek, “Bevolking op 1 januari en gemiddeld; geslacht, leeftijd en regio,” 2022. [Online]. Available: <https://www.cbs.nl/nl-nl/cijfers/detail/03759ned?dl=39E0B>
- [43] E. Mathieu, H. Ritchie, L. Rodés-Guirao, C. Appel, C. Giattino, J. Hasell, B. Macdonald, S. Dattani, D. Beltekian, E. Ortiz-Ospina, and M. Roser, “Coronavirus Pandemic (COVID-19),” *Our World in Data*, 2020. [Online]. Available: <https://ourworldindata.org/coronavirus>
- [44] Centraal Bureau voor de Statistiek, “Inwoners per gemeente,” 2022. [Online]. Available: <https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/regionaal/inwoners>
- [45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>
- [46] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” *arXiv preprint arXiv:1412.3555*, 2014. [Online]. Available: <http://arxiv.org/abs/1412.3555>
- [47] R. Chandra, S. Goyal, and R. Gupta, “Evaluation of Deep Learning Models for Multi-Step Ahead Time Series Prediction,” *IEEE Access*, pp. 83 105–83 123, 2021.
- [48] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges*, 5 2021. [Online]. Available: <http://arxiv.org/abs/2104.13478>
- [49] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into Deep Learning*, 2020. [Online]. Available: <https://d2l.ai>
- [50] Rijksoverheid, “Februari 2020: Eerste coronabesmetting in Nederland.” [Online]. Available: <https://www.rijksoverheid.nl/onderwerpen/coronavirus-tijdlijn/februari-2020-eerste-coronabesmetting-in-nederland>
- [51] —, “Ontwikkeling van het virus - Positieve testen.” [Online]. Available: <https://coronadashboard.rijksoverheid.nl/landelijk/positief-getestemensen>
- [52] Rijksinstituut voor Volksgezondheid en Milieu, “Varianten van het coronavirus SARS-Cov-2,” 2022. [Online]. Available: <https://www.rivm.nl/coronavirus-covid-19/virus/varianten>
- [53] —, “Covid-19 rapportage van SARS-CoV-2 varianten in Nederland via de aselece steekproef van RT-PCR positieve monsters in de nationale kiemsurveillance.” 2022. [Online]. Available: <https://data.rivm.nl/meta/srv/dut/catalog.search/metadata/4678ae0b-2580-4cdb-a50b-d229575269ae>
- [54] M. Z. Naser and A. H. Alavi, “Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences,” *Archit. Struct. Constr.*, 2021.
- [55] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, 2019. [Online]. Available: <http://arxiv.org/abs/1912.01703>
- [56] M. Fey and J. E. Lenssen, “Fast Graph Representation Learning with (PyTorch Geometric),” in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- [57] A. D. Iuliano, J. M. Brunkard, T. K. Boehmer, E. Peterson, . S. Adjei, A. M. Binder, S. Cobb, P. Graff, P. Hidalgo, M. J. Panaggio, J. J. Rainey, P. Rao, K. Soetebier, S. Wacaster, C. Ai, V. Gupta, N.-A. M. Molinari, and M. D. Ritchey, “Trends in Disease Severity and Health Care Utilization During the Early Omicron Variant Period Compared with Previous SARS-CoV-2 High Transmission Periods - United States, December 2020-January 2022,” US Department of Health and Human Services/Centers for Disease Control and Prevention, Tech. Rep., 1 2022.

APPENDIX
A DATASETS

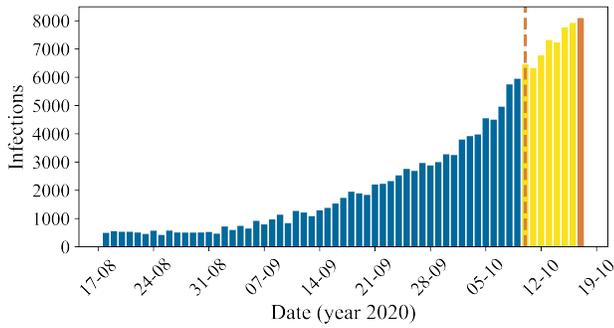
In Figure 8 and Figure 9, we visualize the aggregated number of infections in the Netherlands over time for the 14 time periods we evaluated. As explained in Section VI, we note that there are significant differences between the chosen time periods in the order of magnitude of the infection rates and in the shape of the infection curve.



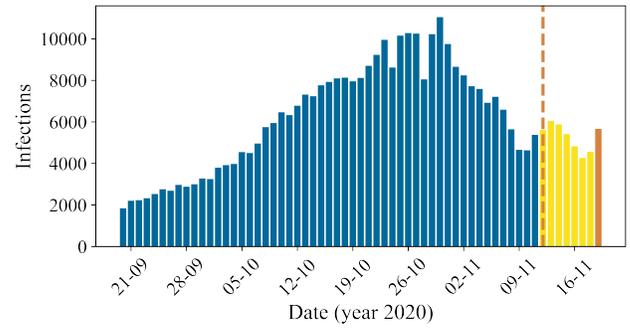
(a) Time period **W1**



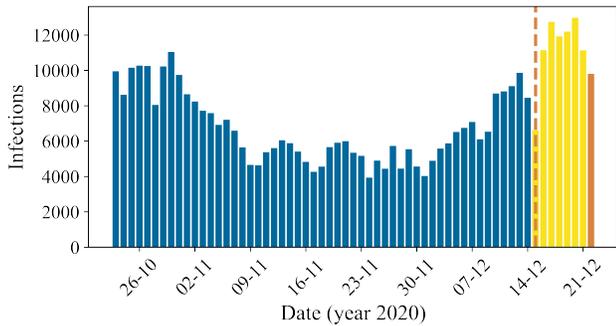
(b) Time period **W2**



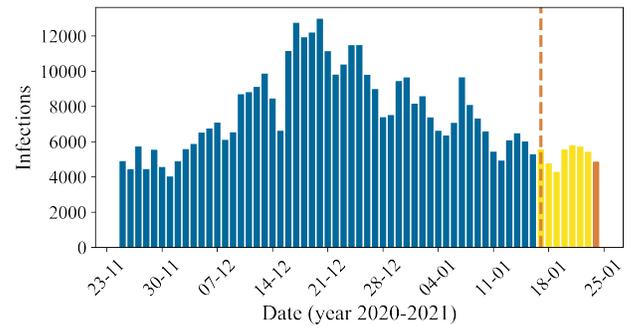
(c) Time period **W3**



(d) Time period **W4**

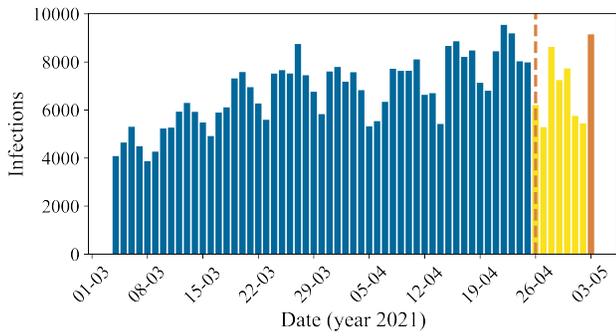


(e) Time period **W5**

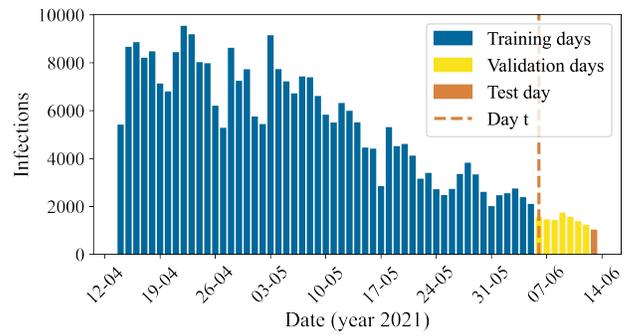


(f) Time period **W6**

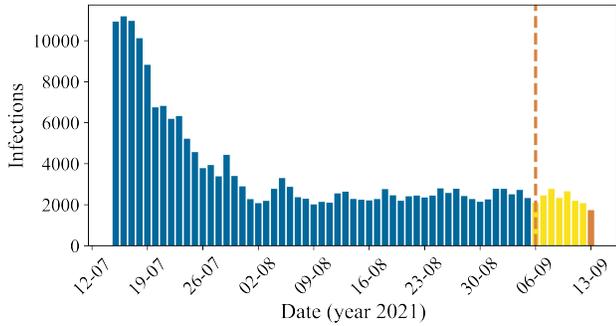
Fig. 8: The aggregated number of reported COVID-19 cases in the Netherlands over the prediction days of the wild type time periods **W1**, **W2**, **W3**, **W4**, **W5**, and **W6**. We present the training set in blue, the validation set in yellow, and the test set in orange. Once the model is trained, we use data up to and including day t (dashed orange line) to predict one week ahead $t+7$ for the test day.



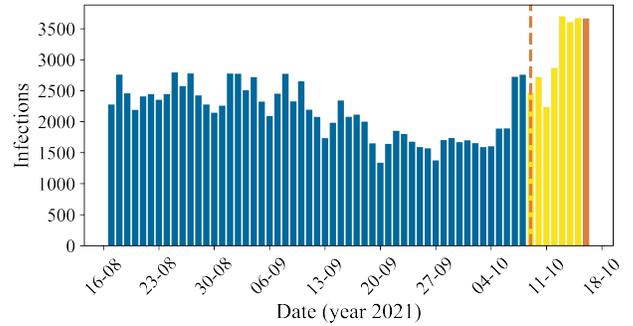
(a) Time period **A1**



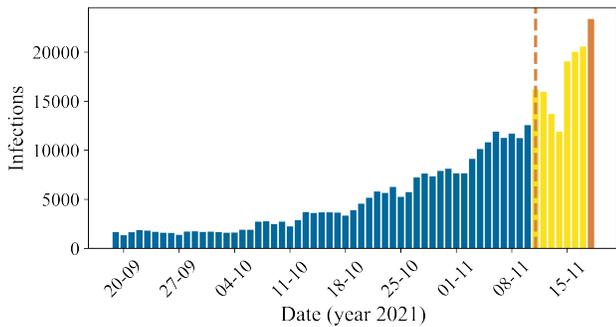
(b) Time period **A2**



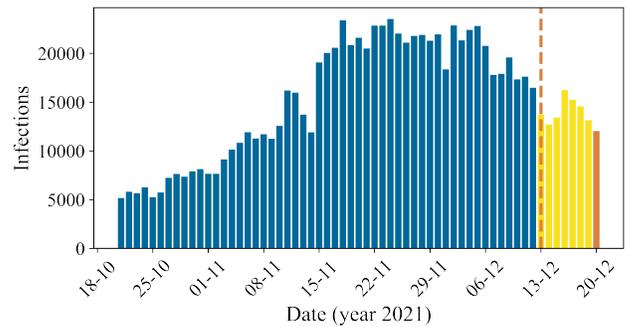
(c) Time period **D1**



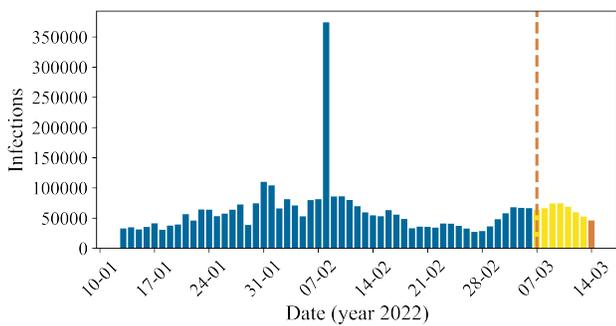
(d) Time period **D2**



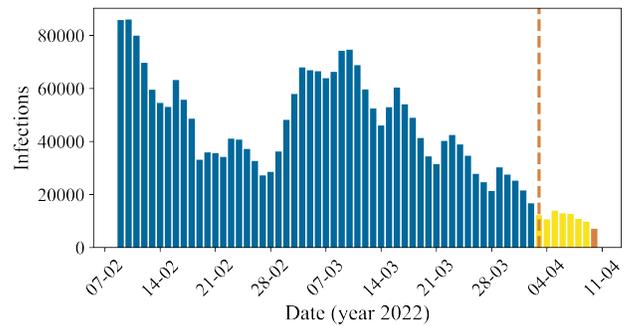
(e) Time period **D3**



(f) Time period **D4**



(g) Time period **O1**



(h) Time period **O2**

Fig. 9: The aggregated number of reported COVID-19 cases in the Netherlands over the prediction days of the alpha, delta, and omicron variant time periods **A1**, **A2**, **D1**, **D2**, **D3**, **D4**, **O1**, and **O2**. We present the training set in blue, the validation set in yellow, and the test set in orange. Once the model is trained, we use data up to and including day t (dashed orange line) to predict one week ahead $t+7$ for the test day.

B CORRELATION PLOTS

In Figure 10 and Figure 11, we visualize the correlation between the observed and the predicted number of reported SARS-Cov-2 infections for the 14 time periods we evaluated. The blue dots indicate the test day forecasts for all 344 municipalities, and the yellow line indicates that the prediction is perfect.

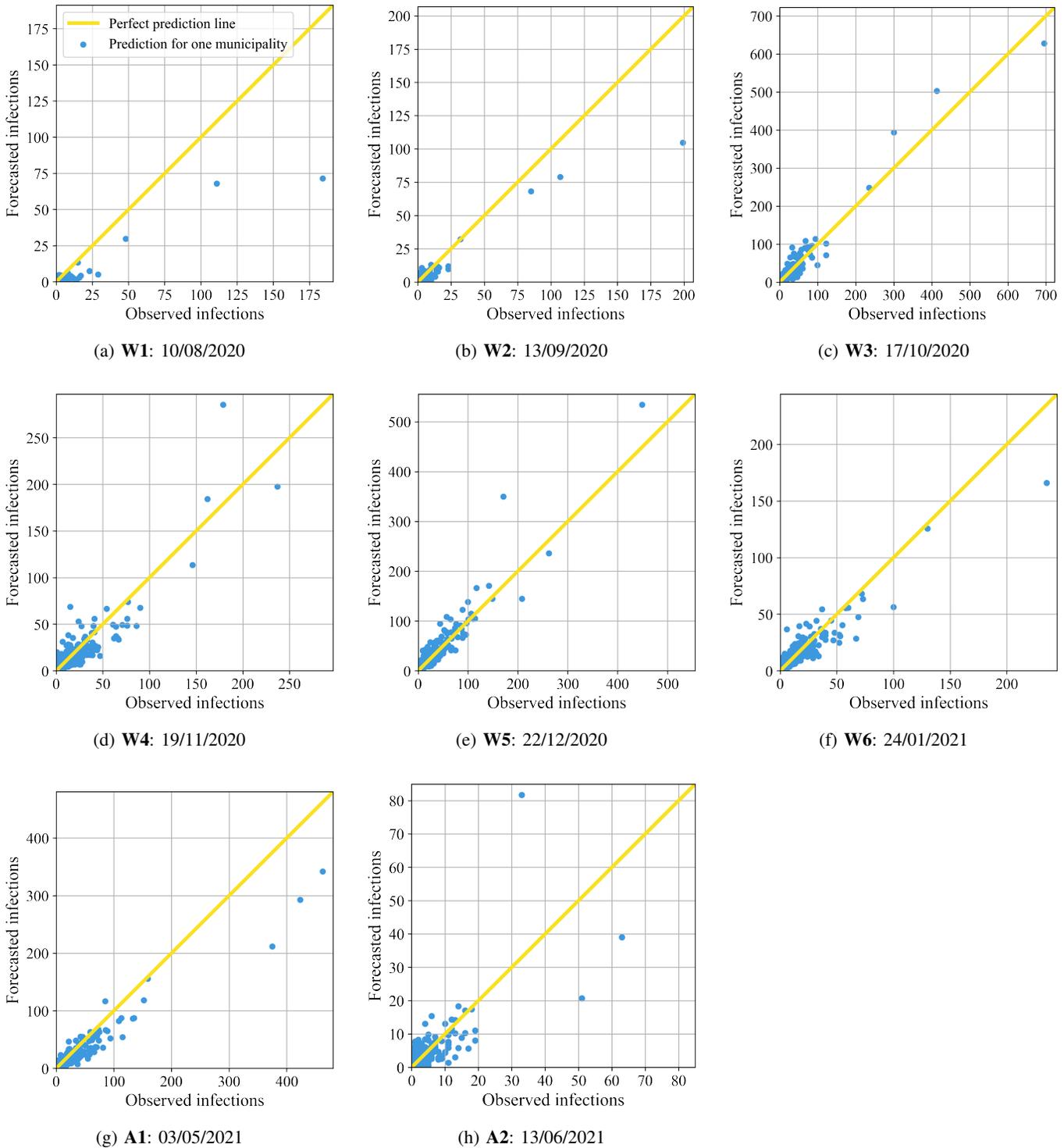
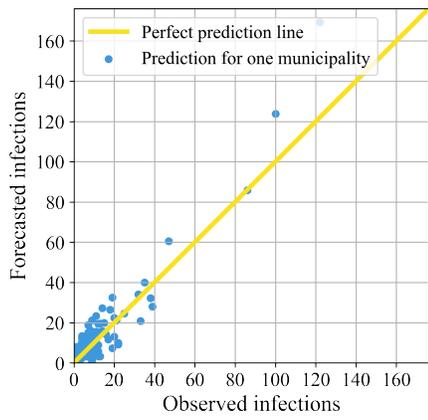
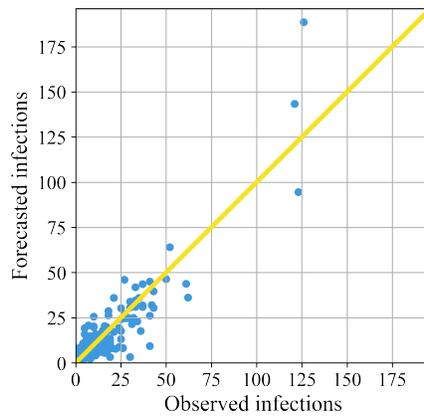


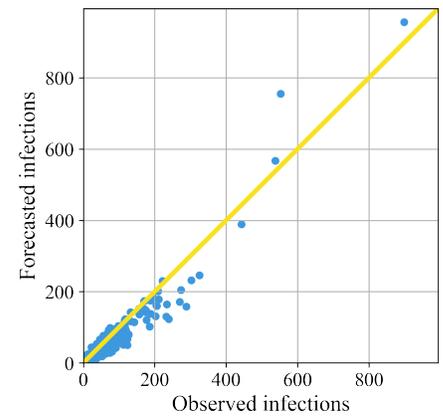
Fig. 10: The correlation between the observed and forecasted number of reported SARS-Cov-2 infections during the wild type and alpha periods **W1**, **W2**, **W3**, **W4**, **W5**, **W6**, **A1**, and **A2**. The blue dots correspond to the forecasts for each municipality on the test day of the related time period. When a dot lies on the yellow diagonal line, it means that the prediction is perfect.



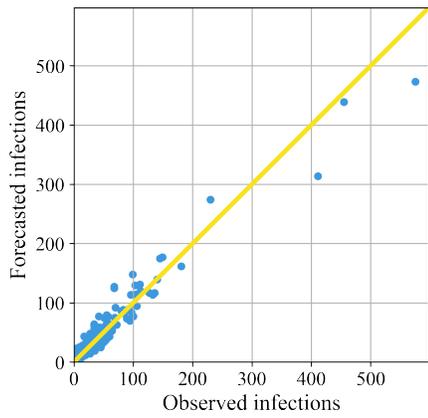
(a) **D1**: 13/09/2021



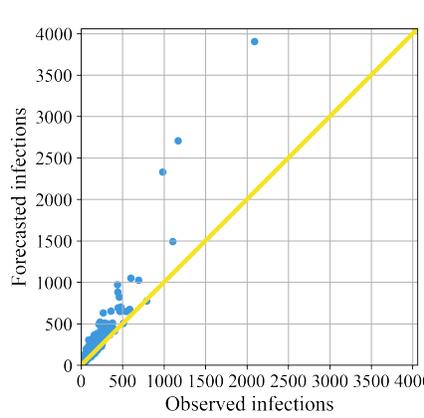
(b) **D2**: 16/10/2021



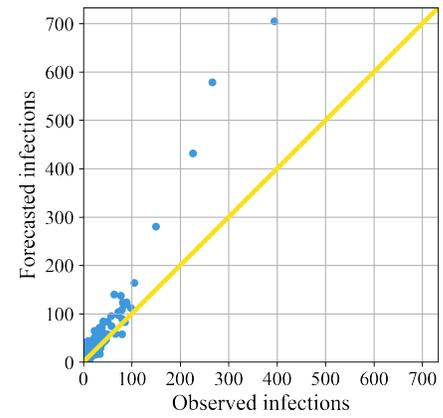
(c) **D3**: 18/11/2021



(d) **D4**: 20/12/2021



(e) **O1**: 14/03/2022



(f) **O2**: 10/04/2022

Fig. 11: The correlation between the observed and forecasted number of reported SARS-Cov-2 infections during the delta and omicron variant periods **D1**, **D2**, **D3**, **D4**, **O1**, and **O2**. The blue dots correspond to the forecasts for each municipality on the test day of the related time period. When a dot lies on the yellow diagonal line, it means that the prediction is perfect.

C ONE DAY AHEAD PREDICTION

In Table XIII to Table XV, we summarize the performance of our model, its components, and the baselines for one day ahead prediction.

TABLE XIII: The root mean squared errors (RMSE) for $t+1$ prediction of our model, its parts, and the baselines for all 14 time periods.

RMSE (\downarrow)	W1	W2	W3	W4	W5	W6	A1	A2	D1	D2	D3	D4	O1	O2
<i>PD</i>	6.23	3.95	11.60	11.46	15.35	8.75	23.40	3.63	<u>4.90</u>	6.44	36.07	15.00	55.58	16.99
<i>HA</i>	10.29	7.61	34.76	19.69	18.17	14.45	19.04	22.31	13.92	8.26	77.86	17.43	101.18	169.68
<i>HA_{window}</i>	<u>5.70</u>	<u>3.17</u>	<u>11.03</u>	<u>10.67</u>	13.02	6.54	<u>18.90</u>	3.89	5.73	<u>5.39</u>	<u>32.89</u>	<u>13.35</u>	102.09	24.54
<i>EB</i>	7.24	3.37	12.97	12.17	16.28	9.29	25.24	<u>3.57</u>	5.05	7.70	44.60	15.69	57.92	<u>14.81</u>
<i>GRU</i>	5.65	2.55	10.58	11.15	16.29	6.70	12.06	3.46	4.81	5.37	31.61	11.09	157.10	40.83
<i>GATv2</i>	8.18	6.57	13.84	13.33	19.82	10.14	15.38	8.08	5.48	7.17	38.87	24.87	126.55	59.97
<i>Model</i>	5.65	2.92	11.41	9.70	17.06	6.81	16.22	3.45	4.87	5.43	23.39	12.95	79.90	13.19

TABLE XIV: The coefficients of determination (R^2) for $t+1$ prediction of our model, its parts, and the baselines for all 14 time periods.

R^2 (\uparrow)	W1	W2	W3	W4	W5	W6	A1	A2	D1	D2	D3	D4	O1	O2
<i>PD</i>	0.736	0.911	0.946	0.765	0.829	0.789	0.714	0.599	<u>0.808</u>	0.815	0.803	0.915	0.902	0.725
<i>HA</i>	0.281	0.671	0.516	0.307	0.761	0.423	0.811	<0	<0	0.695	0.081	0.886	0.675	<0
<i>HA_{window}</i>	<u>0.779</u>	<u>0.943</u>	<u>0.951</u>	<u>0.796</u>	0.877	0.882	<u>0.814</u>	0.541	0.737	<u>0.870</u>	<u>0.836</u>	<u>0.933</u>	0.669	0.425
<i>EB</i>	0.644	0.935	0.933	0.735	0.808	0.762	0.668	<u>0.613</u>	0.796	0.777	0.699	0.907	0.893	<u>0.790</u>
<i>GRU</i>	0.783	0.963	0.955	0.778	0.807	0.876	0.924	0.636	0.815	0.871	0.849	0.954	0.217	<0
<i>GATv2</i>	0.545	0.755	0.923	0.682	0.715	0.716	0.877	<0	0.760	0.771	0.771	0.768	0.492	<0
<i>Model</i>	0.783	0.952	0.948	0.832	0.789	0.872	0.863	0.637	0.810	0.864	0.917	0.937	0.797	0.834

TABLE XV: Average relative RMSE for $t+1$ prediction performance of our model with respect to its parts and the baselines for each COVID-19 variant and in total.

RMSE (\downarrow)	WT	α	δ	σ	Avg
<i>PD</i>	-10.6%	-17.8%	-16.3%	10.7%	-10.2%
<i>HA</i>	-47.3%	-49.7%	-48.7%	-56.6%	-49.4%
<i>HA_{window}</i>	3.5%	-12.8%	-11.5%	-34.0%	-8.5%
<i>EB</i>	-14.9%	-19.6%	-24.5%	13.5%	-14.9%
<i>GRU</i>	2.6%	17.1%	-1.7%	-58.4%	-5.3%
<i>GATv2</i>	-29.7%	-25.9%	-30.8%	-57.4%	-33.4%

TABLE XVI: Average coefficient of determination (R^2) for $t+1$ prediction of our model, its parts, and the baselines for each COVID-19 variant and in total.

R^2 (\uparrow)	WT	α	δ	σ	Avg
<i>PD</i>	0.829	0.657	0.835	0.814	<u>0.804</u>
<i>HA</i>	0.493	0.406	0.416	0.338	0.436
<i>HA_{window}</i>	0.871	<u>0.678</u>	<u>0.844</u>	0.547	0.790
<i>EB</i>	0.803	0.641	0.795	0.842	0.783
<i>GRU</i>	0.860	0.780	0.872	0.109	0.745
<i>GATv2</i>	0.723	0.439	0.768	0.246	0.627
<i>Model</i>	0.863	0.750	0.882	0.816	0.845