Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Dual attention-based sequential auto-encoder for Covid-19 outbreak forecasting: A case study in Vietnam

Phu Pham a,1, Witold Pedrycz b,c,2, Bay Vo a,3,*

a Faculty of Information Technology, HUTECH University, Ho Chi Minh City, Viet Nam
b Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2V4, Canada
c Warsaw School of Information Technology, Nowelska 6, Warsaw, Poland

ARTICLE INFO

Keywords:
Deep learning
Auto-encoding
Bi-LSTM
Attention
Covid 19

ABSTRACT

For preventing the outbreaks of Covid-19 infection in different countries, many organizations and governments have extensively studied and applied different kinds of quarantine isolation policies, medical treatments as well as organized massive/fast vaccination strategy for over-18 citizens. There are several valuable lessons have been achieved in different countries this Covid-19 battle. These studies have presented the usefulness of prompt actions in testing, isolating confirmed infectious cases from community as well as social resource planning/optimization through data-driven anticipation. In recent times, many studies have demonstrated the effectiveness of short/long-term forecasting in number of new Covid-19 cases in forms of time-series data. These predictions have directly supported to effectively optimize the available healthcare resources as well as imposing suitable policies for slowing down the Covid-19 spreads, especially in high-populated cities/regions/nations. There are several progresses of deep neural architectures, such as recurrent neural network (RNN) have demonstrated significant improvements in analyzing and learning the time-series datasets for conducting better predictions. However, most of recent RNN-based techniques are considered as unable to handle chaotic/non-smooth sequential datasets. The consecutive disturbances and lagged observations from chaotic time-series dataset like as routine Covid-19 confirmed cases have led to the low performance in temporal feature learning process through recent RNN-based models. To meet this challenge, in this paper, we proposed a novel dual attention-based sequential auto-encoding architecture, called as: DAttAE. Our proposed model supports to effectively learn and predict the new Covid-19 cases in forms of chaotic and non-smooth time series dataset. Specifically, the integration between dual self-attention mechanism in a given Bi-LSTM based auto-encoder in our proposed model supports to directly focus the model on a specific time-range sequence in order to achieve better prediction. We evaluated the performance of our proposed DAttAE model by comparing with multiple traditional and state-of-the-art deep learning-based techniques for time-series prediction task upon different real-world datasets. Experimental outputs demonstrated the effectiveness of our proposed attention-based deep neural approach in comparing with state-of-the-art RNN-based architectures for time series based Covid-19 outbreak prediction task.

1. Introduction

In the end of year 2019, the rapid spread of novel corona-virus pandemic, named as: Covid-19, over the world has causes tremendous pressures on multiple social aspects. This pandemic also leads to major challenges for researchers in various disciplines (Khan et al., 2021; Kucharski et al., 2020; Liu, Magal, Seydi, & Webb, 2020; Wu, Leung, & Leung, 2020). Recently, the appearance of Covid-19’s delta variant (Li, Lou, & Fan, 2021) has dramatically increased the number of infected patients who need intensive medical treatments. These public health associated disasters has led to severe burdensome for the hospitals even with countries that had high-levelled public healthcare systems and infrastructures (Micah et al., 2021). In fact, this variant of Covid-19 is considered as the most contemporary dangerous one. It has flagged out a
new pandemic wave through the world with an unimaginative speed. Many real-world case studies in different countries have showed that the tremendous growth in the healthcare requirements for Covid-19 patients. It is the results of fast person-to-person transmission (Fidan & Yuksel, 2021; Pitchaimani & Devi, 2021) of the Covid-19 delta variant as well as shortage of suitable strategies for preventing the spreads of virus among groups of confirmed Covid-19 cases to other healthy ones. In addition, the shortage of proper data modelling and short/long-term Covid-19 outbreak forecasting solutions also led to challenges for the governments to effectively manage as well as plan for social resource optimization (Nascimento et al., 2021). The accurate pandemic forecasting mechanism also supports for the governments to properly impose suitable policies to simultaneously deal with the expansion of Covid-19 as well as ensure the social/economic stability (Miao, Last, & Litvak, 2022). Due to the severe influences of this pandemic in multiple social aspects, it is considered as necessary for building data analysis systems which support to capture the spreading temporal patterns of this pandemic. These systems can explicitly support for governments to have suitable plans in both economic recovery and daily life renormalization. Therefore, it is needless to say that the precise forecasting in number of new COVID-19 cases is considered as an important problem. The prediction results can be utilized to facilitate the planning of available resources in public healthcare system. Moreover, for the long run it supports to efficiently optimize the social management strategies/policies for Covid-19 spread prevention as well as treatments for infected patients. To deal with the pandemic outbreak prediction problem, many researchers have proposed different statistical/mathematical and machine learning (ML)-based approaches as data-driven predictive solutions (Khan et al., 2021) for fighting against the rapid escalations of new confirmed Covid-19 cases.

1.1. Statistical and machine learning based approach for Covid-19 outbreak forecasting

Since, the medical reports of confirmed infectious cases are collected and stored as time series data form, most of predictive models are designed to learn capture temporal and sequential patterns of Covid-19 outbreaks. Commonly in most of countries, the outbreaks of Covid-19 frequently occurred at different spreading levels which are tightly respected to specific time-dependent periods. Moreover, recent studies also show that the spreading level of this pandemic is also relied on the changes in some natural aspects, like as temperature, seasonal weather, etc. (Chin et al., 2020; Zhou, Gao, Xie, & Xu, 2020) Therefore, the fluctuation patterns of these outbreaks are naturally non-linear and dynamic which are mainly depended on multiple natural/non-natural aspects. Moreover, the traditional linear and statistical prediction approaches are considered as unable to capture non-linear temporal information from the reported daily confirmed Covid-19 cases. This type of dataset might contain high level of chaotic observations, noises and disturbances due to the influences of multiple internal/external aspects. In general, most of classical linear time-series data-driven prediction models highly relied upon the regression paradigm without the capability in considering non-linear data patterns. Therefore, within the Covid-19 outbreak prediction problem, they might totally fail to sufficiently capture the dynamism of Covid-19 infectious transmission through the community at different time-dependent periods. Moreover, from the perspective of real-world applications, the Covid-19 pandemic prediction system should be able to continuously learn from previous historical data in order to obtain the long-range and temporal features from data to achieve better and more accurate predictions. From the past, multiple statistical models, like as: Auto Regressive (AR), Moving Average (MA), Auto Regressive Moving Average (ARIMA), Nonlinear Auto-regression Neural Network (NARNN) etc. have been widely applied to capture the linear patterns of routine number of confirmed infectious case data to conduct forecasting. However, these models are still far from sufficiently preserving the real-time and temporal information from reported Covid-19 routine data.

1.2. Recent achievements & challenges

Recently, there are several studies presented the efficiency of applying ARIMA model in the Covid-19 pandemic outbreak prediction problem (Katris, 2021). In this approach, the reported Covid-19 data is modelled as the time series pattern to achieve useful information for predicting the new Covid-19 infections. In very recent years, the integrated ARIMA-based approach is utilized as the predictive model is widely applied in different time series datasets. The ARIMA and its variants are widely used due to its flexibility in analyzing temporal information from a given time series dataset through using different statistical ordered parameters. Recent works have demonstrated success of ARIMA integrated predictive systems in handling prediction task for multiple types of infectious diseases (Gao et al., 2020; He & Tao, 2018; Liu, Liu, Jiang, & Yang, 2011), including the Covid-19 pandemic (Benvenuto, Giovanetti, Vassallo, Angeletti, & Ciccozzi, 2020; Dehesh, Mardani-Fard, & Dehesh, 2020). However, most of ARIMA-based predictive technique are still unable to preserve temporal information as well as perform the non-linear regression to deal with complex and highly-noise routine reported Covid-19 data. This type of infectious data is majorly influenced by different external aspects. In order to deal with the non-linear and complex sequential data pattern learning, there are several studies (Ceylan, 2021; Wu et al., 2015; Yu et al., 2014) have applied the NARNN-based approach to enable the capability of predictive system in performing non-linear and time-series based data analysis problem. In this approach, the temporal information is efficiently preserved from the input time-dependent observations through the neural network based learning paradigm. In general, the NARNN is considered as a neural network based approach which explicitly supports to obtain the temporal information from time series data through utilizing different sequences of a given dataset and train them with corresponding parameter weights through the back-propagation learning procedure. Recent enhancements (Ceylan, 2021; Hansun et al., 2021) in the integration of NARNN with predictive systems have demonstrated remarkable performances in dealing with short-term Covid-19 outbreak forecasting problem. The successes of NARNN in complex time series based prediction task has clearly indicated that deep neural network is the potential and key factor for achieving better performance in temporal/sequential data representation learning and prediction problem.

1.3. Deep learning based approach for Covid-19 outbreak forecasting

Recently, there are notable achievements of deep learning in multiple disciplines of computer science domain, like as natural language process (NLP), computer vision (CV), etc. These deep learning’s progresses have shed lights on reaching higher level of sequential/temporal data representation learning as well as time series based prediction task (Lim & Zohren, 2021). Among advanced deep neural architectures, recurrent neural network (RNN), like as: Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), etc. have played an indispensable role in time series data modelling and representation learning problem. Among RNN-based architecture, GRU and LSTM are considered as the most popular deep neural architectures. These RNN-based architectures have been widely utilized in different types of time series based modelling and fine-tuning processes for effectively handling prediction problem. Specifically, GRU and LSTM can effectively learn and capture the long-range temporal information in forms of generated hidden states through different sequence-connected neural state cells. There are several recent studies (Chimmula & Zhang, 2020; Zerosal, Harrou, Dairi, & Sun, 2020) have shown the outperformances of LSTM-based neural architecture in time series based Covid-19 new case prediction task. The LSTM model enables to capture temporal information representation learning and time-dependent training for data prediction. However, most of recent RNN-based time
series based predictive models still encountered several major limitations. The limitations are mainly related to problems in handling highly noisy/non-smooth sequential data as well as long-range dependency. Since, the reported Covid-19 cases are sometime reported as a cumulative sequence with multiple upward/downward trends which produce a non-smooth sequential data pattern. Thus, it might challenge the capability of LSTM in logically preserving these disturbances to make correct predictions. Moreover, the most of RNN-based architectures like as GRU/LSTM is unable to learn and properly interpret long input sequences due to the vanish/explosion of gradient during the back-propagation learning process. In order words, LSTM-based predictive models tend to likely focus and remember much on it last processed sequences rather than preserve the long-range dependencies of all input sequences. Therefore, it might fail to sufficiently capture the trends of Covid-19 outbreaks within long-range periods.

1.4. Our motivations & contributions

Inspired from recent literature studies and proposed Covid-19 predictive models (Chimmula & Zhang, 2020; Hu, Ge, Li, & Xiongd, 2020; Zeroual et al., 2020), in this paper, we proposed an enhanced sequential auto-encoding architecture with the dual self-attention mechanism, called as DAttAE. The proposed DAttAE efficiently supports to improve the performance of temporal information representation learning process.

1.4.1. The utilization of sequential auto-encoding (AE) mechanism in Covid-19 outbreak prediction

In this study, for achieving better performance in Covid-19 case prediction problem, we utilize a sequential (Bi-LSTM) based auto-encoder to encode and generate rich-semantic temporal representations of historical observations. The learnt rich temporal information from sequences are later used to assist the fine-tuning process of Covid-19 trend forecasting problem. In general, the application of AE enables to learn a latent representation of the data which is split into different neural components (encoding/decoding). It later allows us to efficiently disentangle the latent time-dependent features from the given set of historical observations which are currently evaluated. Specifically our proposed DAttAE model in this paper is designed upon the inheritance of the well-known sequence-to-sequence (seq2seq) architecture (Sutskever, Vinyals, & Le, 2014). To deal with noises/lagged observations from time-series dataset, it is integrated with the attention mechanism (Bahdanau, Cho, & Bengio, 2015; Vaswani, 2017) which is mainly inspired from the NLP domain. In general, the historical confirmed Covid-19 case data in forms of time series are passed through an auto-encoding based architecture. The given sequential AE is composed with two separated Bi-LSTM based encoder and decoder as a sequential self-learning approach (as illustrated in the main block of Fig. 1-A). Within the given proposed AE architecture, first of all, we achieve the concatenated (backward/forward) hidden states of the input sequences in forms of reported number of Covid-19 cases within a specific time-dependent range through the encoding component. Then, the encoder’s outputs are fed into another Bi-LSTM architecture to simultaneously interpret and produce the transformed hidden states of the sequence.
original input sequences. Next, both learnt hidden states of the encoding and decoding components of our given AE are passed through separated attention layers to generate the combined attention weights (as illustrated in the Fig. 1-B). The ultimate purpose of applying the dual self-attention filtering mechanism in the outputs of the Bi-LSTM based AE architecture is to estimate the importance levels of all data entries in each input sequence. It directly supports to enhance the prediction fine-tuning process in the after all. Finally, the concentrated attention weights of both encoder and decoder are fed into a full-connected layer to predict the number of Covid-19 cases.

1.4.2. Our findings and real-world case studies in this paper

In general, the Covid-19 case forecasting task can be formulated as non-smooth/high-noised time-series based data analysis problem in which routine reported Covid-19 cases are abnormally fluctuated over the time. The rapid/abnormal changes in number of infectious cases are come from different reasons, including: impacts of environmental/ geographical aspects, the inconsistency in reported data of different organizations/governments, etc. Thus, recent regression as well as modern RNN-based predictive models might be not powerful enough to effectively analyze and reduce noises/temporal disturbances within the representation learning and prediction processes. In order to deal with this problem, in our approach the neural attention mechanism has been adapted to achieve noise-reduced sequential representation of a given input sequence. By doing this, the importance levels of all routine reported Covid-19 data entries in the given sequence are taken into considered during the embedding process through the Bi-LSTM based auto-encoder. Thus, it supports to deliver better prediction results in the after all. To demonstrate the effectiveness of our findings in this paper, we have tested our proposed method with other baselines in real-world Covid-19 datasets and achieved outperformed results for both Covid-19 andConfirmed Covid-19 datasets which are reported within our proposed model with different deep learning based baselines within another Covid-19 dataset which is reported in different states of United States (US). The experimental outputs demonstrated the outperformance of our proposed DAttAE model in comparing with recent state-of-the-art RNN-based architectures.

1.4.3. Difference between our approach and previous works

For many years, researchers have tried to find different approach to effectively handle as well as preserve temporal dynamic features from the time-series based datasets to produce better predictions. However, most of traditional approaches like as the popular auto-aggression based techniques (ARMA, ARIMA, etc.), Naïve Bayesian probabilistic additive decomposition techniques for time-series, etc. have failed to sufficiently capture the long-range time-dependent features between consecutive observation entries. Moreover, within the short/long-term time series prediction approach through deep learning, multiple complex recurrent neural network based architectures have been utilized to dealing with the long-ranged historical observation representation learning. These RNN-based models have sufficiently supported to capture the rich-semantic time-dependent features within time-series dataset to performance significant improvements. However, these deep learning based predictive model normally performed inaccurate predictions while dealing highly noise/chaotic sequential datasets with unpredicted growing patterns like as the routine reported Covid-19 case dataset. In recent times, there are few works have been concentrated on the utilization of attention mechanism within the sequential auto-encoding paradigm. The utilization of neural attention based mechanism with different full-connected components have effectively assisted to capture the short-term growing patterns of disturbing/chaotic time series datasets. Majored inspired from previous works for the case study of Covid-19 outbreak prediction, we dedicated our efforts on discovering on how the integration between dual self-attention neural architectures can efficient support the sequential AE-based architecture. Our proposed AE with a dual self-attention mechanism can support to efficiently preserve the unsmooth sequences and improve the performance of time series prediction problem, specifically for the routine reported Covid-19 confirmed cases in Vietnam.

1.4.4. Managerial implications of our studies in this paper

In recent years, the COVID-19 pandemic as well as it associated aspects have been considered as a hot/top priority research topic for the researchers worldwide. To prevent the rapid spreads of this pandemic, early infectious case forecasting is a crucial task which can facilitate multiple management activities in different social sectors, especially public administration and healthcare system. The accurate Covid-19 case forecasts can directly support to optimize multi-sector social resources for preventing the spreads of this pandemic as well as planning for quarantine isolation strategy. Moreover, due to serve impacts of Covid-19 on human health, the effective public healthcare resource anticipation and distribution are extremely important to provide fast and proper treatments for patients with serve COVID-19 symptoms. Therefore, within the computer science domain, there are various researches have been conducted in the literature on different mathematical models/deep learning based paradigms to predict the spread of the Covid-19. Normally, routine Covid-19 infectious cases are reported in forms of complex and highly-fluctuated time-series datasets. Therefore, to effectively analysis and learn the dynamic feature representations of these datasets, different RNN-based models have been applied in this research direction, however, they still be considered as insufficient to cope with abnormal/lagged observations to convey better prediction results. In our works, we proposed a novel integrated Bi-LSTM based
auto-encoding architecture with dual attention mechanism to effectively deal with complex and non-smooth/abnormal fluctuations of routine reported Covid-19 infectious cases, thus deliver more accurate predictions. The accuracy forecasts in number of new Covid-19 cases and 21 thousand of deaths have been reported up to this time. In fact, the spreads of this Covid-19 pandemic in different countries are varied and it seems unpredictable. This pandemic is influenced by multiple aspects related to both natural (seasonal weather, regional temperature, etc.) as well as unnatural (population, immigration, etc.) aspects. Due to the severe effects of this pandemic in multiple social and economic aspects, it has been considered as a multidisciplinary issue which requires the involvements of many organization and governments. It requires higher efforts for the governments in imposing new social management strategies, pharmaceutical/epidemiological organizations in medical treatment proposal as well as data modeling/analysis solutions for early forecasting and planning.

In Vietnam, our government and VMH have proposed different effective management strategies which enable to flexibly impose different quarantine policies for different regions. There are different policies are imposed depending on the levels of Covid-19 infectious risk at different Vietnam’s regions and cities. Recent new proposed management strategies support to jointly optimize social resources for preventing Covid-19 outbreaks in high-risked regions as well as ensure the social and economic stability in low-risked/safe regions. Following the VMH, the level of risk in each region is identified upon two criteria, which are: the daily/weekly number of confirmed Covid-19 cases and vaccination rate (%) in that region. Table 2 shows the detailed assessment criteria which are provide by VMH \[^5\] for the Covid-19 risk zone classification in Vietnam. Our works in this paper mainly focused on the solution of Covid-19 data as time series modelling and prediction, specifically in learn the evolution pattern of confirmed infected Covid-19 cases. Then, we utilize the predicted infectious case results to conduct Covid-19 risk zone classification following the listed assessment criteria in Table 2. The Fig. 2 illustrates the Covid-19 risk zone prediction and classification problem within 62 districts of Vietnam.

2.2. Recent achievements in Covid-19 forecasting task

2.2.1. The traditional statistical and machine learning based approach

Recently, there are several works focused on applying statistical/mathematical and traditional predictive approach, like as ARIMA (Benvenuto et al., 2020; Dehesh et al., 2020). These traditional time series based techniques support to efficiently model and predict the transmission trends of Covid-19. These works have demonstrated the success in modelling and correctly predicting the linear growth of new Covid-19 infectious cases. However, real-world reported Covid-19 is much complex and dynamic in different nations in which the growth...
trends in number of Covid-19 reported cases are observed as non-linear growing patterns. Therefore, traditional statistical methods might be unable to sufficiently fit non-linear patterns of the given dataset which leads to the downgrade in the accuracy for short-term prediction of Covid-19 transmission growth. Same on the statistical approach, recently Zhang et al. (Zhang et al., 2020) proposed a the R0-based statistical technique for forecasting the Covid-19 infection rate of the outbreak which was occurred in the Diamond Princess cruise ship\[7\]. Similar to that, Varotsos, C. A. et al. proposed a novel probabilistic model, called as: Covid-19 Decision-Making System (CDMS) to measure the spread of Covid-19 disease for prompt actions and further planning activities (Varotsos & Krapivin, 2020). Also considered as an probabilistic approach, in recent studies (Amaral, Casaca, Oishi, & Cuminato, 2021), Amaral, F. et al. proposed a Covid-19 spreading predictive model which is relied on the Spread of Disease (SIR) model (Cooper, Mondal, & Antonopoulos, 2020) to evaluate and forecast the Covid-19 outbreaks in São Paulo, Brazil and Brazil. However, it was still unable to accurately predicting complex non-linear escalations of Covid-19 cases in different time steps. Moreover, traditional mathematical and statistical techniques are also limited in retaining the temporal information from the routine reported infectious cases in forms of time series data. To effectively deal with the non-linear patterns of daily reported infectious cases, there are recent attempts of Ceylan (Ceylan, 2021) and Hansun et al. (Hansun et al., 2021) have demonstrated the effectiveness of integrating NARNN architecture. These NARRN-based models (Ceylan, 2021; Hansun et al., 2021) support to model the flexible time-dependent growth patterns of Covid-19 diseases in different real-world datasets. In general, the NARNN approach is basically considered as an initial artificial neural network based technique. It mainly relies on neuron-varied learning paradigm to deal with the time series based modelling and prediction problem. On the same approach of applying multi-layered neural architecture for handling Covid-19 spreading prediction problem, recent works of Wieczorek, M. et al. proposed an application of stacked multi-layered full-connected neural architecture.

\[7\] Covid-19 outbreak in Diamond Princess cruise ship (BBC news): https://www.bbc.com/news/world-asia-51555420.
to preserve the temporal information from the routine reported Covid-19 datasets (Wieczorek, Silka, Woźniak, 2020). However, the simplicity of NARNN architecture is considered as unable to preserve the long-range temporal information of complex time series datasets to effectively perform the short-terms prediction.

2.2.2. The deep learning based approach

In recent years, the tremendous raises of deep learning in multiple disciplines have shed some lights for enhancing the performance of Covid-19 outbreak forecasting problem (Shorten, Khoshgoftaar, & Furht, 2021). Within the problem of time series data modelling and representation learning tasks, there are popular RNN-based architectures (e.g., GRU, LSTM, Bi-LSTM, etc.) are widely applied to achieve significant performance in multiple data analysis tasks, including the time-series prediction. Recent attempt of Chimmula et al. (Chimmula & Zhang, 2020) and Hu et al. (Hu et al., 2020) in applying advanced deep neural architectures in handling complex non-linear and temporal information capturing from time series data. In more specifics, in this work (Chimmula & Zhang, 2020), Chimmula et al. have presented the advantages of LSTM-based neural architectures in modelling reported Covid-19 time series data for predicting short-term number of infection cases in Canada, US and Italy. For the general literature reviews of potential applications of different deep neural architectures in sequential data representation learning and Covid-19 outbreak forecasting problem, Zeroual et al. (Zeroual et al., 2020) have conducted extensive comparative studies between different RNN-based and auto-encoding (AE) based architectures to present the potentiality of these advanced neural architectures in handling real-world time series data based Covid-19 datasets. Similar to that, recently Chatterjee, A. et al. (Chatterjee, Gerdes, & Martinez, 2020) studied on using multiple LSTM-based architectures to efficiently preserve the dynamic temporal information from reported Covid-19 spreading data to conduct accurate predictions. Also related to Covid-19 time series based data evaluation and learning, Nascimento et al. (Nascimento et al., 2021) recently propped a novel dynamic graph-based analysis technique with multi-regression dynamic model (MDM) approach. It supports to find relationships between time series routine Covid-19 reported data and financial market trends. These recent well-known studies have placed strong backgrounds for building efficient Covid-19 prediction systems.

However, as aforementioned issues related to the limitations of RNN-based architectures in remembering the long-range temporal information from input sequences as well as problem associated with the highly chaotic noisy/non-smooth time series datasets, there are existing challenges which have still required extra researching efforts. Mainly motivated by recent studies in future Covid-19 case forecasting, our works in this paper concentrate on integrating the Bi-LSTM based auto-encoding architecture with dual self-attention mechanism to improve the performance of Covid-19 case prediction task.

3. Methodology

In this section, we formally present the methodology and detailed implementations of our proposed DAttAE model in this paper. In general, the proposed DAttAE model is designed as a Bi-LSTM based auto-encoder which support to softly capture the temporal information from the input time series sequences as a self-supervision representation learning approach. Then, the concatenated hidden states which are produced by the encoder and decoder parts are fed into a dual self-attention mechanism to produce the combined weighted attention vectors which support to estimate the importance of all data entries in the given input sequences to assist for the data temporal forecasting process. To do this, the concatenated output attention weights of both encoder and decoder are passed through a linear full-connected layer to conduct prediction.

3.1. Problem formulation & background concepts

3.1.1. Short-term time series prediction task

In general, large number of real-world datasets in multiple disciplines can be considered as time series data for in which data entries are collected at specific time intervals with (N) data entries, denoted as: \( X = \{x_1, x_2, \ldots, x_N\} \), with \( x_i \) being a single data entry at a specific \( i^{th} \) time. Due to the popularity of time series based datasets, there are numerous studies have presented notable achievements as well as existing challenges for the time series prediction problem. The time series prediction model is formally designed to forecast/predict the upcoming sequential trends/patterns of a given time series dataset through analyze the latent temporal features of the historical data entries. For the short-term time series prediction problem, the given time series dataset \( X \) is normally split into (n) smaller observation sequences, denoted as: \( x_t \), or: \( X_t = \{x_{t-1}, x_{t-2}, \ldots, x_0\} \), each observation sequence, \( x_t = \{x_{t-1}, x_{t-2}, \ldots, x_0\} \) which are used to learn and predict the upcoming trend/pattern of the consecutive data entry, as: \( x_{t+1} \), with (L) is the pre-defined observation length or look-back parameter. In general, given time series predictive method, denoted as: \( f_{RNN}(\cdot) \), is designed to optimize the following learning objective: \( L_{RNN}(X) \approx X_{t+1} \).

3.1.2. Recurrent neural network (RNN)

From the past, most of stacked full-connected linear neural network architectures are widely applied to model and learn context latent features from different types of datasets. However, these stacked multi-layered neural architectures are unable to model and retain the consecutive information from sequential/time-series datasets. In order to deal with the continuous and temporal information present in different sequential data forms, the novel deep neural architecture is required. RNN is considered as the most popular neural architecture which is designed to deal with the sequence/time-dependent latent feature representation learning problem. In general, the RNN is developed upon the principle of considering the effects of consecutive data entries’ information to generate the corresponding outputs. To do this, a neural state cells are organized as the sequence-ordered structure. Each neural cell in a given RNN-based architecture contains multiple logic gates which are used to control the influences of historical observations of previous data to the current input entry, as: \( x \) and generate the corresponding output in form of hidden state, denoted as \( h \). For a specific time-step \( (t^n) \), the generated hidden state, denoted as \( h \), for a specific input \( x_t \) is generally obtained as follows: \( h_t = \begin{cases} 0, & \text{if } t = 0 \\ \left( W_{RNN}^h h_{t-1}, \sigma(.) \right), & \text{if } t \geq 1 \end{cases} \), standing for the set of trainable parameters (weights, bias) and the corresponding activation functions of different gates in a given RNN architecture. Among RNN-based architectures, GRU and LSTM are the most commonly and widely applied to model and analyze different types of sequence/time series based datasets in multiple disciplines like as NLP explain abbreviation, short-term recommendation, multi-media information retrieval, etc.

3.2. DAttAE model

3.2.1. Bi-LSTM based auto-encoding architecture

To effectively model the input sequences coming from a non-smooth/high noisy time series dataset like as routine reported Covid-19 infectious cases, we majorly inspired from the well-known seq2seq-based approach (Sutskever et al., 2014) in the NLP domain in which the proposed DAttAE model is designed as a sequential auto-encoding architecture with two components encoder and decoder. The encoding part is composed as a Bi-LSTM architecture which support to effectively encode all data entries in each input sequence in both forward and backward directions. In more details for a specific data entry \( x_t \) in a given observation sequence \( X \), each LSTM architecture at each direction
supports to produce the corresponding hidden state $h_i$ as the following (as shown in equation (1)). In this equation, the $U$, $W$ and $b$ present for the weighting and bias parameter matrices of different gates of the given LSTM architecture in each direction which is identified by the operator: $[+,-]$ ( $+$ , $-$ used for denoting the forward and backward directions, respectively).

$$\begin{align*}
\hat{h}_i^{(+)} &= \sigma(U_x + W_i h_i^{(-)} + b) \\
\hat{h}_i^{(-)} &= \sigma(U_x + W_i h_i^{(+)} + b) \\
\hat{h}_i &= \tanh(U_x + W_i h_i^{(+)} + b) \\
h_i &= \hat{h}_i^{(+)} \odot \tanh(\hat{h}_i^{(-)})
\end{align*}$$

(1)

Then, the last ($k^{th}$) generated hidden states of the given Bi-LSTM architecture are then combined to form the final sequential embedding of the given input sequence ($X$). These learnt temporal representations are formulated as the output hidden state embedding vectors of Bi-LSTM based last layers, denoted as $e_X = h_k \in \mathbb{R}^{N \times 2k}$, with $|X|$ is the length of input sequence ($X$).

### 3.2.2. The utilization of LSTM/Bi-LSTM and sequential auto-encoding mechanism

In general, the utilization of a LSTM/Bi-LSTM based sequential embedding mechanism in our work is to sufficiently capture the range-varied time-dependent features from a given time-series dataset. The equation (1) and (2) generally illustrated a basic LSTM-based neural cell which is implemented in our proposed sequential encoding mechanism. In our encoding and decoding parts, a dual hierarchical LSTM based architecture (Smagulova & James, 2019) is applied to generate the bi-directional temporal embeddings of each input entry in forms of last output hidden states of each LSTM architecture. These rich-structural temporal representations are combined to produce the unified representations of consecutive observation entries.

The overall process of hidden state combination process is illustrated as shown in the equation (2), with $[\ldots]$ presenting for the vector concatenation operation. Then, the generated latent embedding vectors for each input sequence in the encoding part will be passed to the decoding part. The decoding component is also designed as another Bi-LSTM based architecture to transform them into another sequential representation form, denoted as ($e_X^{\text{det}}$). The ultimate purpose of using different Bi-LSTM based embedding mechanism to learn the sequential representation of input sequence ($X$). It supports to create a longer consecutive neural learning architecture. By using the given Bi-LSTM based auto-encoding mechanism, it assists to prevent the problems related to gradient explosion as well as efficient trainable weighting parameter calculation. Moreover, the Bi-LSTM architecture also provides better temporal information preserving for chaotic time-series data in which generated latent features in each neural cell can be shared and remembered for sets of long/non-smooth input sequences.

#### 3.2.3. Dual attention mechanism and time series prediction

Next, in order to enhance the capability of the given Bi-LSTM based auto-encoder to achieve better sequential representation of a given input sequence in which the importance levels of all data entries in the given sequence are taken into considered during the embedding process, we implement a custom hierarchical dual self-attention mechanism in our proposed DAttAE model. There are two self-supervised attention layers are placed at the end of encoder and decoder components which take the output embedding vectors of each component and help to overall model to selectively concentrate on different parts of the given sequence for better prediction-driven fine-tuning process at the end. Specifically, the self-supervised attention layer which is located at the output of encoder and decoder components is designed as follows.

$$\begin{align*}
score(a, e_X) &= \text{ReLU}(W^{\text{lin}} X_{\text{det}}) \\
\alpha_X &= \frac{\exp(score(a, e_X))}{\sum_{i=1}^{N} \text{score}(a, e_X)} \\
e_X &= \sum_{i=1}^{N} \alpha_X e_X
\end{align*}$$

(3)

(4)

In the equation (3), the attention score of each data entry at a specific $i^{th}$ index, denoted as: $score(a, e_X)$, is computed as the linear transformation of the attention alignment paradigm which is majorly inspired in previous works (Bahdanau et al., 2015; Vaswani, 2017). Being different from previous proposed attention mechanisms which are mainly applied for machine translation task, in our developed dual self-supervised attention mechanism, we used the ReLU activation function to perform the linear transformation of the given output sequential embedding $e_X$ with trainable weighting parameter matrix, as: $W^{\text{lin}}$.

Then, the computed attention score for a specific $i^{th}$ data entry in the sequence is utilized to calculate the corresponding importance score, denoted as: $\alpha_X$. Finally, we compute the context attention vector for the given input sequence $X$, denoted as: $e_X$ as the weight sum of all attention scores for all data entries in $X$ which is illustrated in the equation (4). In general, the application of this self-supervised attention mechanism in our DAttAE model directly supports to the soft alignment between the generated sequential embedding vectors in both encoding $e_X^{\text{enc}}$ and decoding $e_X^{\text{det}}$ parts which are simultaneously learned and controlled by the corresponding context vectors: $e_X^{\text{enc}}$ and $e_X^{\text{det}}$, respectively. Then, we concatenate the calculated context vectors of both encoding and decoding part to produce the final dual attention-based weighting embedding vector of input sequence $X$, denoted as: $e_X^{\text{att}} = [e_X^{\text{enc}}, e_X^{\text{det}}]$. Finally, to let the given architecture predict the value of consecutive data entry of the given sequence, as: $x_{(i+1)}$, we feed the achieved attention-based embedding vector $e_X^{\text{att}}$ to a linear full-connected layer. The general calculation of this process is formulated as shown in form.

$$\begin{align*}
x_{(i+1)} = \text{Linear}(e_X) = W^C e_X^{\text{att}} + b^C
\end{align*}$$

(5)

$$\begin{align*}
J_{\text{DAttAE}} &= \sum_{i=1}^{N} \left\| x_{(i+1)} - x_{(i)} \right\|^2
\end{align*}$$

(6)

As shown in (5), a linear full-connected layer supports to linearly transform the previous obtained attention-based embedding vector of input sequence $X$ into the distributions of all possible entry values in a given time series dataset $X$. Then, we optimize DAttAE model’s parameters through the mean square error (MSE) loss strategy with the defined learning objective, as: $J_{\text{DAttAE}}$ as shown in the equation (6).

### 4. Experiments & discussions

In this section, we present extensive experiments to evaluate the performance of the proposed DAttAE model in new infection case prediction task within two real-world Vietnam reported Covid-19 datasets. Moreover, we also provide thorough comparative studies of the performances of our model with other state-of-the-art RNN-based baselines which are assessed under standard evaluation metrics used for the time series forecasting problem.
4.1. Experimental datasets & setups

4.1.1. Dataset descriptions & usage

In order to ensure the adaptation of the proposed model in real-world application, we mainly utilize realistic daily reports of confirmed Covid-19 cases in forms time series data resources within Vietnam. The experimental datasets in this paper are collected from official source of VMH. There are two main types of datasets which are used for all experiments in this paper, which are:

- **VN-62P**: this dataset contains the daily reported Covid-19 cases of 62 provinces in Vietnam. For each province, we collected the number of new daily Covid-19 infection cases which have been confirmed and reported by VMH. The quantity of new daily cases in each province has been collected within the duration of April 27, 2021 - October 21, 2021. We used this dataset to test the capability of deep learning baselines in predicting new infectious Covid-19 cases at 62 provinces. To train deep learning predictive baselines, we split the set of data entries of each province into two parts: training set (70 %) and test set (30 %). The training set is used to train and validate the given predictive model. Then, the trained predictive model is evaluated again with the test set under different evaluation metrics.

- **HCMC-12D**: similar to the VN-62P dataset, we also collected a set of Covid-19 reports in Ho Chi Minh city (HCMC) provide by VMH which include the daily reported number of infection cases in 12 large and high-populated districts of HCMC. This time series dataset is collected within the time range of [June 22, 2021 - October 19, 2021]. Similar to experiments with VN-62P dataset, deep learning based predictive techniques will be trained to predict the number of new Covid-19 cases at specific time steps. For the experiments in time series predictive model learning and evaluation, the set of data entries in each district is divided into training and test sets.

4.1.2. Dataset usage for experiments

For evaluating the performance of long-range Covid-19 case prediction task within two datasets, we applied the time-dependent splitting strategy in which the training set are taken from the period of April 27, 2021 to the end of (35th) week (September 5, 2021) for the VN-62P dataset and from June 22, 2021 to September 5, 2021 for the HCMC-12D dataset. For experiments with different traditional and deep learning based predictive techniques which are applied for Covid-19 prediction task, we all applied the same test/split ratio in which the training set is utilized to learn and extract temporal features from historical observation which are later applied to predict the future number of Covid-19 cases against the test set. The detailed information and statistics about these two datasets as well as usage which are used for all experiments in our paper can be found in Table 3.

4.1.3. Configurations & evaluation methods

**Experimental environment & configurations.** To implement the DAttAE model, we mainly used the Python programming language under the PyTorch\(^8\) machine learning framework. We set up our DAttAE model and other traditional deep learning based comparative baselines for all experiments on a same computer with Intel Xeon CPU E5-2620 v4 2.10 GHz (8 cores – 16 threads) CPU and 64 Gb memory.

**Setups of our proposed DAttAE model for data training and evaluation.** For the Bi-LSTM based auto-encoding mechanism (as described in sub-section 3.2.1), we set the number of used LSTM-based cells (k\(^{\text{LSTM}}\)) for each Bi-LSTM architecture as 32. For the experiments with LSTM/Bi-LSTM related techniques which later are described in sub-section 4.1.3, we also used the same configured number of hidden states for all datasets. For all predictive techniques which are related to the RNN-based approach for dealing with the long-range Covid-19 prediction task, the default observation/look-back length (L) are set as 5 for all datasets. For the setup of dual self-supervised attention mechanism (as described in the sub-section 3.2.2), the default weighting parameter values of all attention layers is initialized by using Xavier initialization. The hidden size layer of these attention-based architectures is set as the same with the number of configured (k\(^{\text{LSTM}}\)) parameters for the encoder and decoder components. In more specifics, Table 4 details configured parameters which are set up for our DAttAE model in all experiments.

**Experimental evaluation criteria.** Similar to recent works (Chimmula & Zhang, 2020; Hu et al., 2020; Zeroual et al., 2020 ), to evaluate the performance of different deep learning based time series predictive model, we three main standard evaluation metrics which are: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). These evaluation methods support to assess how much a given predictive model make wrong predictions upon a context of time series dataset. These evaluation methods are calculated upon a given test set, denoted as: \( \mathcal{X}_{\text{test}} \) as the following (as shown in the equations (7) and (8)). For each time series predictive model in each dataset, we run experiments 10 times and reported the average output of each model as the final result.

\[
\begin{align*}
\text{MAE} & = \frac{1}{n} \sum_{t=1}^{\mathcal{X}_{\text{test}}} |x_t - \hat{x}_t| \quad \text{(7)}
\end{align*}
\]

\[
\begin{align*}
\text{RMSE} & = \sqrt{\frac{1}{n} \sum_{t=1}^{\mathcal{X}_{\text{test}}} (x_t - \hat{x}_t)^2} \quad \text{(8)}
\end{align*}
\]

\[
\begin{align*}
P & = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN} \\
F-1 & = 2 \cdot \frac{P \cdot R}{P + R} \quad \text{(9)} \]
\]

Table 3

| Dataset | Number of evaluated region | Data reported time range | No. data entries for each region | Training size | Testing size |
|---------|---------------------------|-------------------------|----------------------------------|--------------|-------------|
| VN-62P  | 62                        | April 27, 2021 - October 21, 2021 | 177                             | 129          | 48          |
| HCMC-12D | 12                        | June 22, 2021 - October 19, 2021 | 110                             | 64           | 46          |

| Parameter | Value |
|-----------|-------|
| Number of LSTM-based cells which are used in both encoder and decoder components (k\(^{\text{LSTM}}\)) | 32 |
| Default observation/look-back length (L) for each input sequence. | 5 |
| Training/validation(%) split ratio. | 90/10 |
| Default number of training epochs. | 500 |
| Default general dropout rate for all neural architectures in our model. | 0.5 |
| Default learning rate (\( \eta \)) for all datasets. | \( 1 \times 10^{-2} \) |
| Default model’s optimizer. | Adam (weighting decay: \( 5 \times 10^{-5} \)) |

---

\( ^8 \) PyTorch: https://pytorch.org/.
In our experiments, we also conducted the extended Covid-19 zone categorization problem in each assessed geographical region (provinces in VN-64D dataset and districts in HCMC-12D dataset) which is considered as the classification problem. To evaluate the accuracy performance of this Covid-19 risk zone categorization task, we mainly used two metrics which are the Accuracy and F-1 measure (as shown in equation (9) and (10)). In this equation, the values of precision (P) and recall (R) are calculated by identifying the TP (true positive), FP (false positive) and FN (false negative). Specifically, the TP (true positive) presents the number of regions which are correctly classified to their ground-truth Covid-19 risk zone labels. The FP (false positive) and FN (false negative) indicate the number of expected regions which are categorized to specific Covid-19 risk zone based classes but not correct and the number of regions which are not classified by their actually Covid-19 risk zone ground-truth labels, respectively. To identify the Covid-19 risk zone-based labels for each region (district/province) in each dataset, we mainly relied on the standard assessment criteria of VMH which are mentioned in Table 2. Due to the lack of information in vaccination rate in each region, thus we only considered the daily/weekly number of new reported Covid-19 infection cases to identify the ground-truth/predictive labels for each region.

4.1.4. Comparative baselines

To demonstrate the effectiveness of our proposed DAttAE model in comparing with recent state-of-the-art deep learning based predictive model for time series dataset, we implement several techniques to conduct Covid-19 infection cases in two VN-62P and HCMC-12D datasets, which are:

- **Naïve Bayes (NB):** for further evaluation on how traditional probabilistic model can facilitate on the problem of time-series based Covid-19 prediction problem, we implemented a simple NB-based regression model which supports to product the predictions through the additive decomposition approach.

- **ARIMA:** is considered as the most well-known and traditional approach for time series data forecasting problem. In this study, we implement the ARIMA model to predict the number of Covid-19 cases within two VN-62P and HCMC-12D datasets. In our experiments, we used the ARIMA model to conduct the long-range time-dependent prediction task.

- **GRU (Zeroual et al., 2020):** is considered as the classical and earliest RNN-based neural architecture in which neural state cells are organized as the sequence-ordered structure to enable the capability in modelling and representation learning for sequence/time series-based data. Each GRU’s neural cell contains two types of logic gates: reset and update which support to control the preserved sequential information of each data entry in a sequence when it is passed through it. For the implementation of GRU based predictive model in our experiments, the number of used neural state cells (kGRU) is set as the same our DAttAE model.

- **LSTM (Zeroual et al., 2020):** is the recent most advanced sophisticated sequence-ordered neural network architecture which is designed to prevent problems related to the vanishing/exploration of gradient while handling long-dependent sequence/time series-based datasets. Different from the GRU, LSTM contains three types of logic gates: input, forget and output. The advanced designs of LSTM enable it to extend the capability in preserving the long-term dependencies between input sequences and thus perform great ability to dealing with sequence/time-series based data analysis and representation learning problems.

- **Bi-LSTM (Zeroual et al., 2020):** is the enhanced version of the LSTM in which two different LSTM-based architectures are vertically aligning to learn the sequential representation of all data entries in each input sequence at both direction (forward/backward). Then, the last concatenated hidden state vector of each sequence is used to predict the next data entry value. For all experiments, the Bi-LSTM architecture is setup as the same configurations with the encoder part of our proposed DAttAE model.

- **COVID-ANN (Wieczorek, Silka, & Woźniak, 2020):** on the same view of applying deep learning based approach for the Covid-19 forecasting problem, Wieczorek, M. et al. proposed a novel stacked full-connected neural architecture which is similar to the traditional multi-layered perceptron (MLP) architecture to capture temporal information from time-series based Covid-19 dataset. In this work, the authors utilized a 7-layered stacked linear layers with the tanh(·) activation function between two layers. Then, the hidden states of last layer are fed into a ReLU activation to conduct prediction. For comparative studies, we have constructed the same neural architecture in this work to conduct the number of Covid-19 case prediction.

- **COVID-RNN (Chatterjee et al., 2020):** is considered as a recent deep learning approach for Covid-19 prediction problem within the recurrent neural network based temporal learning paradigm. In this work, Chatterjee, A. et al. proposed a novel Covid-19 predictive model which is composed as a multiple LSTM-based architecture. The dynamic temporal representations of historical observation entries which are aggregated through hidden states of different LSTM based neural networks are then fed into task-driven layer to conduct prediction. Chatterjee, A. et al. have utilized different-layered LSTM architecture to conduct Covid-19 prediction problem within different Asian countries like as India, Iran, Singapore and Turkey. For experiments in this paper, we implemented a 3-layered LSTM architecture to deal with the routine Covid-19 case forecasting task within our datasets.

For other similar configurations of these comparative techniques, we set them as the same DAttAE model’s configurations which are shown in Table 4.

4.2. Experimental results & discussions

4.2.1. Daily number of Covid-19 infection case forecasting task

In this experimental section, we present extensive comparative studies between different deep learning based predictive models for handling the daily number of Covid-19 infection case forecasting problem within the VN-64D and HCMC-12D datasets.

The routine Covid-19 infectious case forecasting is considered as a challenging task within non-smooth time-series databases like as VN-64D and HCMC-12D. The Fig. 3 and Fig. 4 show the average experimental outputs for daily Covid-19 case prediction problem through different traditional and deep learning based techniques, within the VN-62P and HCMC-12D datasets, respectively. The experimental outputs as shown in these tables are evaluated under the MAE and RMSE standard metrics. In general, as briefly take a look at output experimental charts our proposed DAttAE model outperforms most of recent baselines for the Covid-19 case forecasting problem.

In general, as shown from the experimental outputs (Fig. 3 and Fig. 4), our proposed DAttAE model significantly achieved better performances than traditional time-series and probabilistic/regression models, like as: NB and ARIMA. As shown from the experimental outputs, it achieved better predictive MAE/RMSE-based performances than the NB and ARIMA, about 116.21 %/104.51 % and 135.64 %/120.77 % within the VN-64P dataset (as shown in Fig. 3) which is similar in the HCMC-12D dataset (as shown in Fig. 4) for 79.33 %/66.52 % and 53.19 %/50.7 %, respectively. Generally, these experimental outputs have demonstrated the outperformances of deep learning based approaches in comparing with classical probabilistic and regression methods in which better long-range time dependent features are captured with deep neural architectures. In comparing with other common sequential deep neural baselines, including: GRU, LSTM and Bi-LSTM in terms of MAE and RMSE evaluation metrics for this task, our proposed DAttAE model also demonstrated significant improvements. Specifically, for the Covid-19
case prediction task in the VN-62P dataset, our proposed DAttAE model remarkably outperforms previous RNN-based architectures (GRU, LSTM and Bi-LSTM) about: 136.51 %, 116.67 % and 17.97 % in terms of MAE and 117.76 %, 99.62 % and 14.26 % in terms of RMSE, respectively. Similar to that within the HCMC-12D dataset, our approach also notably improves the accuracy performances in terms of MAE and RMSE evaluations about: 19.72 %/20.95 %, 80.21 %/67.84 % and 64.23 %/59.16 % in comparing with Bi-LSTM, LSTM and GRU, respectively. In addition, for comparisons with recent deep learning based methods for Covid-19 case prediction problem like as the COVID-ANN and COVID-RNN, our proposed DAttAE model also slightly improved the accuracy performances in terms of MAE/RMSE metrics approximately 86.85 %/75.23 % and 4.67 %/5.98 % within the VN-62P dataset and 24.7 %/29.53 % and 10.21 %/12.85 % within the HCMC-12D dataset, respectively.

Moreover, as shown from some forecasting outputs (Fig. 5 and Fig. 6) in the VN-62P and HCMC-12D datasets, our proposed DAttAE model perform better predictions than the traditional Bi-LSTM architecture. In general, the ground-truth Covid-19 daily reported cases are extremely undulating, thus presenting how challenging of the routine infectious case prediction problem. The experimental outputs present that the daily predictive values are close to the actual values than the Bi-LSTM based predictive ones. By utilizing the custom dual attention mechanism within the Covid-19 data representation learning and prediction tasks, our DAttAE model have significantly achieved better sequential representations of Covid-19 case observation data entries in both VN-62P and HCMC-12D and deliver nearly-fit predictions. Through experiments in Covid-19 data in Vietnam, the experimental outputs have demonstrated the effectiveness of our proposed DAttAE model in handling the Covid-19 outbreak forecasting problem in form of time-series based prediction approach.

4.2.2. Daily/weekly Covid-19 risk zone classification task

For the Covid-19 risk zone identification problem (as mentioned in sub-section 2.1), in this experimental section we demonstrated the extensive comparative studies in using deep neural time series predictive techniques to conduct the Covid-19 risk zone classification. For the dataset VN-62P is the zone classification for 62 provinces of Vietnam. Similar to that, for HCMC-12D dataset is the zone classification for 12 high-populated/large districts in HCMC, Vietnam. In order to conduct Covid-19 zone identification depending on criteria which have been described in previous section, similar to the daily Covid-19 case prediction problem, we used the trained deep learning-based models, including our proposed DAttAE model, to predict the daily number of infection cases in each region (provinces for the VN-62P dataset and districts for the HCMC-12D dataset). Then, we relied on the assessment criteria for Covid-19 risk zone categorization in Table 2 to identify the corresponding classes (red, orange, yellow and green zones) for all regions at specific time-steps. Finally, the Covid-19 risk zone categorization results are evaluated under the Accuracy and F-1 evaluation metrics for the comparison purposes between different baselines. We evaluated the performance of each model in Covid-19 risk zone classification task with two types of time-interval, which are: daily and weekly.
The Fig. 7 and Fig. 8 show the experimental outputs for both daily and weekly Covid-19 zone risk classification task through different techniques, within the VN-62P and HCMC-12D datasets, respectively. Specifically, as shown from the experimental outputs, our proposed DAttAE model also slightly leverage the accuracy performance of Covid-19 risk-zone classification task in both VN-62P and HCMC-12D datasets. In more specific, our proposed DAttAE model explicitly achieves better accuracy performances in terms of F-1 based accuracy than the GRU, LSTM and Bi-LSTM approximately 3.92 %, 2.79 % and 1.57 %, within the VN-62P dataset and 12.33 %, 14.15 % and 1.53 % within the HCMC-12D dataset, respectively. In comparing with recent deep learning based studies, such as: COVID-ANN and COVID-RNN, about proposed DAttAE model also achieve better performances for this task about 5.41 % and 1.79 % in the VN-62P, 5.06 % and 8.01 % in the HCMC-12D dataset.

Similar to previous empirical studies in the Covid-19 case forecasting problem (sub-section 4.2.1), experimental results in this section have proved the effectiveness of integrating with attention mechanism for long-varied sequential data representation learning task in forms of daily/weekly Covid-19 risk-zone classification problem.

4.3. Model evaluation on other time-series Covid-19 dataset

For further performance evaluation of our proposed DAttAE model within different times-series based Covid-19 dataset, we conducted extensive experiments on a popular United States (US) reported Covid-19 confirmed cases within different states (named as US-Covid19). This dataset contains different types of information related to the reported death and infectious cases in forms of time-series and routinely
updated by collecting data from the well-known Johns Hopkins University Center for Systems Science and Engineering (CSSE)\(^9\). For experiments in this dataset, we extracted the number of confirmed Covid-19 infectious cases within the top-5 highest populated states in the US, including: California, Texas, Florida, New York and Pennsylvania. The extracted data related to the confirmed cases in these states are within the period of April 16, 2021 to January 15, 2022. For dealing with the infectious case prediction problem within this dataset, we utilized different deep learning based methods including: LSTM, Bi-LSTM, COVID-ANN, COVID-RNN and our proposed DAttAE model. We applied the same train/test data splitting strategy with previous datasets as well as general configurations (as described in Table 4) for all deep learning techniques which are studies in the experiments with the US-Covid19 dataset. The average prediction performances in terms of MAE and RMSE evaluation metrics of all states are reported as shown in Fig. 9.

As shown from the final experimental results in Fig. 9, our proposed DAttAE model has demonstrated effectiveness as well as outperforms in comparing with other deep learning based time-series predictive models. In more specific, our proposed DAttAE model explicitly achieved better performance in terms of MAE/RMSE for the Covid-19 confirmed case prediction about 34.6 %/36.3 %, 20.84 %/18.61 %, 29.01 %/27.63 % and 14.5 %/14.89 % in comparing with LSTM, Bi-LSTM, COVID-ANN and COVID-RNN, respectively. This extensive comparative result has proved the efficiency of utilizing the dual attention mechanism within RNN-based architecture for dealing with non-smooth and chaotic time-series based dataset, like as the routine reported Covid-19 infectious cases.

4.4. Parameter sensitivity analysis

To furtherly study on the influences of different model’s parameters upon the accuracy performance of our proposed DAttAE model, in this section we presented extensive ablation studies on important fine-tuning parameters of our model, like as the number of training epochs, the used LSTM-based cells for the Bi-LSTM based auto-encoding architecture (\(k_{\text{LSTM}}\)) and the length of observation sequence (\(l\)) for the Covid-19 case forecasting problem. As shown from the experimental outputs in Fig. 10, our proposed DAttAE model are quite insensitive with these parameters. Specifically, for the experiment with number of training epochs, we trained our model with different number of epochs within range of [10, 500] and reported the changes in accuracy performances for the Covid-19 case prediction task in terms of RMSLE assessment metric. The experimental outputs in Fig. 10-A presented that the DAttAE model achieved the stability in accuracy performance within 400–500 training epochs for both VN-62P and HCMC-12D datasets. For the \(k_{\text{LSTM}}\) parameter, the experiments in Fig. 10-B showed the ideal number of used LSTM’s cells for our model is about \(\geq 32\). Similar to that with the default length of observation sequence (\(l\)) parameter, our proposed DAttAE model reached the highest accuracy performance with the value of \(l\) parameter is in range (Cao et al., 2020; Chin et al., 2020); in both VN-62P and HCMC-12D datasets.

5. Conclusion & future works

In this paper, we study the problems of routine Covid-19 infection case forecasting and risk zone classification in forms of deep learning based analysis and representation learning for time series dataset. The accurate daily Covid-19 infection case forecasting and risk zone categorization provide crucial information to governments for effectively planning social resources and imposing suitable policies to prevent the Covid-19 pandemic escalation in different regions. To effectively model and retain the temporal and growing patterns of time series based reported Covid-19 confirmed infection cases, we proposed a novel Bi-LSTM based auto-encoding mechanism with the custom dual self-supervised attention mechanism, called as DAttAE. The proposed DAttAE model not only enables to sufficiently preserve them temporal
information from complex/non-smooth time series dataset but also ensures the readiness of learnt sequential embedding vector for dealing with the short-term prediction problem with chaotic noise/disturbance through the application of self-supervised attention mechanism. The utilization of attention mechanism for integrating with the archived sequential embedding vectors which are produced by encoder and decoder supports to estimate the important level of all data entries in each input sequence. Then, the calculated attention weights of input sequences are later used to explicitly facilitate the forecasting-driven fine-tuning process in the after all. The application of attention mechanism with RNN-based architecture such as Bi-LSTM enables to better preserve longer chaotic time series in which temporal latent features are sufficiently capture within deeper neural network architectures. Extensive experiments in real-world reported Covid-19 datasets in Vietnam demonstrated the effectiveness of the proposed ideas.

5.1. Achievements/advantages of our works & remained shortages.

Our studies in this paper mainly focused on the application of deep learning approach for dealing with Covid-19 pandemic outbreak prediction problem. The proposed model in this paper might be useful for the governments to effectively forecast as well as optimize multi-sector social resources for preventing the Covid-19 pandemic spreading. To effectively deal with complex and non-smooth fluctuations of routine reported infectious cases and deliver more accurate prediction results, we applied the custom dual attention mechanism within our Bi-LSTM based sequential auto-encoding model to efficiently reduce noises and lagged observations from the Covid-19 reported data. The effectiveness of our proposed ideas in this paper have been demonstrated through extensive experiments in real-world Covid-19 datasets. However, our proposed DAttAE still be unable to incorporate the routine reported Covid-19 data with other associated information resources, such as external environmental/geographical aspects in order to achieve better prediction results. The environmental/geographical aspects like temperature, contamination, geographical locations, etc. (Bashir et al., 2020; Rasheed, Rizwan, Javed, Sharif, & Zaidi, 2021) are considered as important direct/indirect factors which might lead to the raise in number of Covid-19 infectious cases. Therefore, the capability of integrating with exogenous/external information resources of Covid-19 predictive models could be also considered as a potential improvement direction for further studies in this domain.

5.2. Our future works

For future works, we intend to expand the studies in GIS-based spatial clustering problem within the context of time series for identifying the spread pattern of Covid-19 hotspots within specific geographical regions. These expansions require extra research effort on integrating geographical spatial clustering techniques with temporal information representation learning to detail with the clustering problem in the context of temporal dynamism. In addition, in order to improve the performance of our proposed DAttAE model in handling chaotic time series prediction task more accurately, we intend to extend the current model to involve a fuzzy-neural inference mechanism (Han, Zhong, Qiu, & Han, 2018; Soto, Castillo, Melin, & Pedrycz, 2019) within the RNN-based architecture. The utilization of fuzzy-neural inference might enable to leverage the performance of time series prediction problem in such chaotic time series-based datasets like as routine Covid-

![Fig. 9. Comparative studies between different deep learning based techniques within the US-Covid19 dataset.](image)

![Fig. 10. Extensive experiments on the model’s parameter sensitivity for Covid-19 case forecasting problem within VN-62P and HCMC-12D datasets.](image)
19 reported cases.

CRediT authorship contribution statement
Phu Pedrycz: Writing – review & editing, Validation. Bay Vo: Writing – review & editing, Validation, Conceptualization.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References
Aamaral, F., Casaca, W., Oishi, C. M., & Cumminato, J. A. (2021). Towards providing effective data-driven responses to predict the COVID-19 in São Paulo and Brazil. Sensors, 21(2), 540.
Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. 39th International Conference on Learning Representations.
Bashir, M. F., Jiang, B., Komal, B., Bashir, M. A., Farooq, T. H., Iqbal, N., et al. (2020). Forecasting disease spreading and death. Sensors, 20(11), 3271.
Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Artificial Intelligence Forecasting of Covid-19 pandemic: The case study of lockdown in New York State. Expert Systems with Applications, 187, Article 115797.
Bashir, M. F., Jiang, B., Komal, B., Bashir, M. A., Farooq, T. H., Iqbal, N., et al. (2020). Tracking social media during the COVID-19 pandemic: A brief study in Californian context. Chaos, Solitons & Fractals, 135, Article 109664.
Cao, L. T., Liu, H. H., Li, J., Yin, X. D., Duan, Y., & Wang, J. (2020). Relationship of relationship of meteorological factors and human brucellosis in Hebei province, China. Science of the Total Environment, 703, Article 135491.
Choi, K. W., Lee, Y. H., Jeong, J., & Jeong, Y. (2020). A new model for the spread of COVID-19 and the improvement of safety. Safety science, 132, Article 104962.
Ceylan, Z. (2021). Short-term prediction of COVID-19 spread using grey rolling model optimized by particle swarm optimization. Applied Soft Computing, 107592.
Chatterjee, A., Gerdes, M. W., & Martínez, S. G. (2020). Statistical explorations and univariate timeseries analysis on COVID-19 datasets to understand the trend of disease spreading and death. Sensors, 20(11), 3089.
Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons & Fractals, 135, Article 109664.
Chin, A. W., Chu, J. T., Peerea, M. R., Hui, K. P., Yen, H. L., Chan, M. C., et al. (2020). Stability of SARS-CoV-2 in different environmental conditions. The Lancet Microbe, 1(1), e10.
Cooper, L., Mondal, A., & Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. Chaos, Solitons & Fractals, 139, Article 110057.
Dehesh, T., Mardani-Fard, H. A., & Dehesh, P. (2020). Forecasting of covid-19 confirmed cases in different countries with arma models. MedRxiv.
Fidan, H., & Yuskel, M. E. (2021). A comparative study for determining COVID-19 risk levels by unsupervised machine learning methods. Expert Systems with Applications, 126243.
Hu, M., Zhong, K., Qiu, T., & Han, B. (2018). Interval type-2 fuzzy neural networks for chaotic time series prediction: A concise overview. IEEE transactions on cybernetics, 49 (7), 2720–2731.
Khan, M., Mehran, M. T., Haq, Z. U., Ullah, Z., Naqvi, S. R., Ihsan, M., et al. (2021). Applications of artificial intelligence in COVID-19 pandemic: A comprehensive review. Expert Systems with Applications, 185, Article 115665.
Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., et al. (2020). Early dynamics of transmission and control of COVID-19: A mathematical modelling study. The lancet infectious diseases, 20(5), 553–558.
Li, M., Lou, F., & Fan, H. (2021). SARS-CoV-2 Variants of Concern Delta: A great challenge to prevention and control of COVID-19. Signal Transduction and Targeted Therapy, 6(1), 1–3.
Lim, B., & Zhouen, S. (2021). Time-series forecasting with deep learning: A survey. Philosophical Transactions of the Royal Society A, 379(2194), 20200209.
Liu, Q., Liu, X., Jiang, B., & Wang, Y. (2011). Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. BMC infectious diseases, 11(1), 1–7.
Liu, Z., Magal, P., Seydi, O., & Webb, G. (2020). Understanding unreported cases in the COVID-19 epidemic outbreak in Wuhan, China, and the importance of major public health interventions. Biology, 9(3), 50.
Miao, L., Mant, M., & Lirvok, M. (2022). Tracking social media during the COVID-19 pandemic: An integrated analysis. The Lancet infectious diseases, 20(10), 1296–1299.
Nascimento, D. C., Pimentel, B. A., Souza, R. M., Costa, L., Gonçalves, S., & Louzada, F. (2021). Dynamic graph in a symbolic data framework: An account of the causal relation using COVID-19 reports and some reflections on the financial world. Chaos, Solitons & Fractals, 153, Article 114440.
Pitchain, M., & Devi, M. B. (2021). Stochastic probical strategies in a delay virus infection model to combat COVID-19. Chaos, Solitons & Fractals, 152, Article 112325.
PloS One, 9(7), e0135492.
Qin, Y., Li, H., & Zhou, D. (2020). Dynamic graph in a symbolic data framework: A concise overview. IEEE transactions on cybernetics, 49 (7), 2720–2731.
Qin, Y., Li, H., & Zhou, D. (2020). Dynamic graph in a symbolic data framework: A concise overview. IEEE transactions on cybernetics, 49 (7), 2720–2731.
Rasheed, R., Rizwan, A., Javed, H., Sharif, F., & Zaidi, A. (2021). Socio-economic and environmental impacts of COVID-19 pandemic in Pakistan—an integrated analysis. Environmental Science and Pollution Research, 28(16), 19926–19945.
Smagulova, K., & James, A. P. (2019). A survey on LSTM memristive neural network architectures and applications. The European Physical Journal. Special Topics, 228(10), 2313–2326.
Soto, J., Castillo, O., Melin, P., & Pedrycz, W. (2019). A new approach to multiple time series prediction using MIMO fuzzy aggregation models with modular neural networks. International Journal of Fuzzy Systems, 21(3), 1629–1648.
Su, J., Ying, S., & Zhang, L. (2021). Sequence to Sequence Learning with Neural Networks. Advances in Neural Information Processing Systems.
Vaswani, A., et al. (2017). Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems, (pp. 6000-6010).
Wang, C. M., Vlassi, J., & Zeng, X. (2020). Neural network powered COVID-19 epidemic forecasting model. Chaos, Solitons & Fractals, 135, Article 109664.
Wu, W., Guo, J., An, S., Guan, P., Ren, Y., Xia, L., et al. (2015). Comparison of two hybrid models for forecasting the incidence of hemorrhagic fever with renal syndrome in Jiangsu Province, China. PLoS One, 10(8), e0135492.
Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. The Lancet, 395(10225), 689–697.
Yu, L., Zhou, L., Tan, L., Jiang, H., Wang, Y., Wei, S., et al. (2014). Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in Shenzhen, China. PloS One, 9(6), e98241.
Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-series data: A Comparative study. Chaos, Solitons & Fractals, 140, Article 110121.
Zhang, S., Xiao, M., Wu, L., Zou, F., Du, Z., & Sun, Z. (2020). Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. International Journal of Infectious Diseases, 93, 201–204.
Zhao, C., Gao, C., Xie, Y., & Xu, M. (2020). COVID-19 with spontaneous pneumomediastinum. The Lancet Infectious Diseases, 20(4), 510.