IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 August 2023, accepted 6 September 2023, date of publication 25 September 2023, date of current version 12 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3318931

RESEARCH ARTICLE

DSA-BEATS: Dual Self-Attention N-BEATS Model for Forecasting COVID-19 Hospitalization

AMIRHOSSEIN MOTAVALI[1], KIN-CHOONG YOW[1], (Senior Member, IEEE), NICOLE HANSMEIER[2], AND TZU-CHIAO CHAO[3,4]

[1]Faculty of Engineering and Applied Sciences, University of Regina, Regina, SK S4S 0A2, Canada
[2]Department of Biology, Luther College, University of Regina, Regina, SK S4S 0A2, Canada
[3]Institute of Environmental Change and Society, University of Regina, Regina, SK S4S 0A2, Canada
[4]Department of Biology, University of Regina, Regina, SK S4S 0A2, Canada

Corresponding author: Kin-Choong Yow (kin-choong.yow@uregina.ca)

This work was supported by the Health Canada's COVID-19 Safe Restart Agreement Contribution Program.

ABSTRACT The high number of hospitalization cases of COVID-19 made public health providers overloaded. Forecasting the number of hospitalized patients related to COVID-19 can help public health providers make informed decisions for controlling the spread. In this study, we present the Dual Self-Attention NBEATS (DSA-BEATS) model, a novel approach that effectively combines the self-attention mechanism of transformers with the proficiency of the N-BEATS model in dealing with multivariate forecasting problems. We expanded the dataset to a multivariate one by including data from Canadian transportation hub cities and SARS-CoV-2 RNA load in wastewater, which allowed for a more comprehensive modeling of the complex relationships impacting COVID-19 hospitalizations. These transportation hub cities were the major ports of entry for international travelers coming to the country. The DSA-BEATS model was tested on a 55-day test set with a 12-day horizon, resulting in a Mean Absolute Percentage Error (MAPE) of 14.23%, which implies an accuracy of 85.77%. These results demonstrate substantial improvements over state-of-the-art models such as N-BEATS and Informer, validating the efficacy of the DSA-BEATS model in accurately predicting COVID-19 hospitalizations. The study provides a significant contribution to the ongoing development of enhanced timeseries forecasting methods, particularly in the context of public health crises. The DSA-BEATS model's ability to capture complex temporal relationships and effectively handle multivariate data inputs underscores its potential in a wide range of forecasting tasks beyond the COVID-19 pandemic.

INDEX TERMS N-BEATS, self-attention, forecasting, wastewater, transportation hubs, COVID-19.

I. INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the Coronavirus disease 2019 or COVID-19, has spread worldwide causing about 650 million reported cases and more than 6 million deaths [1], [2]. Throughout the pandemic, healthcare services were frequently interrupted by overwhelming numbers of additional COVID-19 patients, indicating the need for models to forecast hospital admissions [3]. The forecasting of COVID-19 hospitalization cases is a critical task that enables healthcare providers and policy-makers to effectively plan and allocate resources. Accurate and timely predictions can assist in managing hospital capacities, optimizing staffing, and informing mitigation strategies, all of which contribute significantly to controlling the disease's impact on health systems and societies. Traditionally, timeseries forecasting models such as Autoregressive Integrated Moving Average (ARIMA) [4] and its variants have been used for this task. Machine learning and compartmental models are the most popular baseline forecasting approaches for hospitalization predictions, according to the Centers for Disease Control and Prevention (CDC) [5]. For example, Rodríguez et al. [6] incorporated COVID-19 data as timeseries and built a neural network architecture for forecasting. Along with the COVID-19 timeseries data, they also used mobility information as inputs to their network. In a

The associate editor coordinating the review of this manuscript and approving it for publication was Akin Tascikaraoglu.

model developed by Jin et al. [7] attention and transformer methods were used for forecasting and the inputs were past COVID-19 trends. On the other side, researchers such as Park [8] modified a standard Susceptible-Exposed-Infective-Recovered (SEIR) model to aggregate different variables including mobility, transportation, and spatial-temporal to capture COVID-19 dynamics and eventually forecast hospitalizations. Wang et al. [5] used statistical and data-driven models for hospitalization predictions that use COVID-19 public search (internet search) information as inputs to the model.

A. PROBLEM DEFINITION

More recently, machine learning-based methods have started to gain traction due to their ability to model complex patterns and dependencies in the data. Among these, the Neural Basis Expansion Analysis for Timeseries (N-BEATS) [9] model, a feed-forward neural network specifically designed for timeseries forecasting, has shown impressive performance. However, despite its strengths, N-BEATS is not without its limitations. Specifically, the model does not inherently account for interdependencies within the timeseries data, making it potentially less effective at capturing complex temporal dynamics. This is a notable challenge in the context of COVID-19 hospitalization forecasting, where temporal dependencies (e.g., the effects of transportation and variant spread) play a crucial role. On the other hand, attention-based models, such as the Transformer, effectively capture these dependencies via self-attention mechanisms, thereby providing context-aware predictions. However, they often struggle with capturing long-term dependencies and trends, an area where N-BEATS excels.

To address these issues, this study proposes to augment the N-BEATS model with a Dual-Self Attention mechanism. The intent is to harness the interpretability and trend-capturing capabilities of N-BEATS while leveraging the context-aware forecasting ability of attention mechanisms. By merging these complementary strengths, we aim to enhance the accuracy and robustness of COVID-19 hospitalization forecasts. In this study, we operationalize our problem as follows: Given a sequence of historical COVID-19 hospitalization data from a region and the major transportation hub cities, our goal is to forecast future hospitalization cases of the region using an N-BEATS model augmented with a Dual-Self Attention mechanism. The model's performance will be evaluated using common error metrics and compared with traditional and state-of-the-art forecasting models. Furthermore, to prove the efficiency and effectiveness of our proposed approach, we will conduct an ablation study. In this analysis, the predictive performance of the original N-BEATS model, the original Transformer model, and our proposed N-BEATS with Dual-Self Attention will be compared. This comparative study will shed light on the relative contribution of each model component to the predictive performance, thereby quantifying the added value of the proposed dual-self-attention mechanism. Ultimately, we seek to develop a tool that provides more reliable, effective, and context-aware predictions to support decision-making in healthcare during the ongoing pandemic.

B. MOTIVATION

Accurate and timely forecasting of hospitalization cases is crucial for optimizing resource allocation, planning patient care, and informing public health policies. Traditional time-series forecasting models, although effective in certain scenarios, may not fully capture the complexities inherent in pandemic data, which includes temporal dependencies, non-linear trends, and effects of interventions.

Recent advances in machine learning, specifically deep learning, have shown promise in timeseries forecasting tasks, including N-BEATS [9] and Transformer [10] models. N-BEATS, with its trend-capturing capabilities and interpretability, and Transformers, with their powerful self-attention mechanism, each bring unique advantages to the table. However, they also have their own limitations. N-BEATS does not inherently account for complex temporal interdependencies, and Transformers often struggle with long-term dependencies. In the context of COVID-19 hospitalization forecasting, where both long-term trends and complex temporal dependencies are vital, a model that combines these strengths could provide a more robust and accurate solution.

C. CONTRIBUTIONS

This paper makes several significant contributions toward the goal of improved COVID-19 hospitalization forecasting:

- **Introducing and using a new variable for forecasting COVID-19 hospitalization:** The hospitalization rates of the transportation hub cities have been used as a new variable for forecasting the hospitalization rates of a non-transportation hub city for the first time.
- **Proposing a new model:** We proposed a novel integration of N-BEATS and Transformer models, augmenting N-BEATS with a Dual-Self Attention mechanism. This unique model combines the trend-capturing learning abilities of N-BEATS with the context-aware capabilities of Transformers, enhancing both the accuracy and robustness of forecasts.
- **Real-world Dataset and Applicability:** Our work contributed to the real-world challenge of COVID-19 hospitalization forecasting, and by extension, to healthcare resource management during the pandemic. The proposed model has been applied to a real-world dataset that is uniquely collected by us and can be utilized by healthcare providers and policymakers for informed decision-making, ultimately contributing to better pandemic management.
- **Ablation Study:** We carried out an ablation study comparing the original N-BEATS model, the original Transformer model, and our proposed N-BEATS with Dual-Self Attention. This study illuminates the specific

contribution of each component and quantifies the added value of the proposed dual-self attention mechanism.
- **Benchmarking Performance:** We evaluated the performance of our proposed model against other state-of-the-art models such as the informer using suitable error metrics, thereby providing a comprehensive understanding of its efficacy in the task of COVID-19 hospitalization forecasting.

Although this study's focus is on COVID-19 hospitalization rate forecasting, our study pushes the frontier of timeseries forecasting as a general problem, demonstrating the potential of combining the state-of-art model's strengths in tackling real-world forecasting problems. It opens up avenues for further research into the integration of various modeling approaches for improved predictive performance.

D. MANUSCRIPT OUTLINE

The rest of the paper is organized as follows: Section II provides a comprehensive literature review, discussing both traditional and state-of-the-art methods in timeseries forecasting, with a particular emphasis on the N-BEATS and Transformer models. We also cover existing research on COVID-19 hospitalization forecasting to give a context for our work. In Section III, we first delve into our data augmentation and analysis and then will describe our forecasting methodology, based on the N-BEATS model, the self-attention mechanism, and our proposed integration of these into a Dual-Self Attention N-BEATS model. Section IV outlines the model configuration and training details as well as our results, including the comparative performance of the models and the insights gained from the ablation study. In Section V, we discuss the implications of our results and the strengths and limitations of our study. Finally, in Section VI, we conclude our paper, summarizing our contributions and the potential impact of our work on COVID-19 hospitalization forecasting. The possible directions for future research are discussed in section VII.

II. RELATED WORK

The task of timeseries forecasting, specifically in the context of epidemiology and health informatics, has been a vibrant area of research for many years. This literature review will provide a synopsis of significant traditional and state-of-the-art models used for this task, with a special focus on N-BEATS and Transformer models. We will also cover existing work on COVID-19 hospitalization forecasting to contextualize our proposed approach.

A. TRADITIONAL TIMESERIES FORECASTING METHODS

Traditional forecasting methods refer to statistical algorithms that are based on mathematical equations and are commonly utilized for timeseries forecasting [4]. These methods are relatively simple to implement and interpret, and they can effectively forecast any timeseries data that does not change its statistical properties such as its mean and variance over time (stationary timeseries). However, they are not well-suited for non-stationary data, and they may struggle to capture complex patterns or sudden changes in the data.

The most commonly used traditional timeseries forecasting models are ARIMA [11] (AutoRegressive Integrated Moving Average) and its variants. ARIMA is a traditional statistical model and is a combination of an Auto-regression part and a moving average. ARIMA models work best when there are clear trends and seasonal orders in the dataset. ARIMA assumes that the timeseries data is stationary, which may not be the case in many real-world applications.

The parameters of an ARIMA model are selected based on the characteristics of the timeseries, such as stationarity, trend, and seasonality. The selection of the parameters is often done through a grid search or a maximum likelihood estimation. SARIMA (Seasonal AutoRegressive Integrated Moving Average) [12] is a variant of ARIMA that includes additional parameters that account for seasonality in the data.

Regarding the COVID-19 forecasts, studies have used traditional models to forecast COVID-19 data. For example, García et al. [13] used ARIMA-based models for forecasting the number of COVID-19-related deaths in Spain. They tested their model for a short period of time (10 days in total) and reported an average accuracy of 95% over those 10 days. However, it's worth noting that their forecast covered a period where the trend of deaths was smoothly rising without any sharp changes or even a decline. Also, the number of COVID-19 death cases tends to be more stable and predictable compared to variables such as COVID-19 hospitalizations, thereby rendering the application of traditional models somewhat less challenging.

B. MACHINE LEARNING METHODS

Deep learning technologies, such as neural networks, have been vital in various fields, including the tracking and control of COVID-19. For COVID-19 forecasting, neural networks, including RNNs, CNNs, and Transformer Models, have been essential [14]. LSTM networks, a variant of RNNs, were identified as the most used method for COVID-19 forecasting, with 35% of related papers employing them or their variants [15]. Chandra et al. [16] specifically implemented multiple LSTM networks to forecast COVID-19 in India using a sliding window approach as well as a sequence-to-sequence approach.

Transformer models, known for revolutionizing natural language processing, are also suitable for forecasting by treating time-series data as sequences, capturing complex patterns [17]. Different variants of the transformer model have been proposed for providing timeseries forecasting and one of the most widely adopted ones is the Informer [18]. The Informer models utilize sparse attention transformers as a key feature in their architecture. This particular design allows them to efficiently extract and understand long-term temporal dependencies within the data, making them adept at recognizing patterns and trends that span over extended periods. Multiple variants of the Informer model have been proposed including the tightly coupled convolutional transformer (TCCT) algorithm, Autoformer algorithm, FEDformer algorithm, Pyraformer algorithm, and Triformer algorithm [19], each of them designed for a specific problem.

C. INFORMER: A TRANSFORMER-BASED MODEL

The Informer [18] model is a notable advancement in time-series forecasting, innovating on the original Transformer [10] model's structure. While Transformers, known for their self-attention mechanisms, were initially used for language translation and text summarization, Informer refines these principles for time-series forecasting. It introduces a novel ProbSparse Self-Attention, which activates only key time steps in the sequences, reducing computational complexity and memory usage. Furthermore, the Informer model features a Distilling-Informer mechanism, which condenses long-term historical data into a compact form. This aids in accurate forecasting by efficiently utilizing crucial historical information, making the model more adept at handling tasks requiring extensive historical context.

The Informer model's merits have been substantiated through several empirical studies. In the original informer paper [18] the Informer model is tested against several other models, including the Transformer model and traditional models such as ARIMA, across multiple public datasets. In these comparative analyses, the Informer model consistently demonstrated superior performance. In the context of COVID-19 forecasting, self-attention mechanisms enable the model to allocate varying attention levels across different steps in the input sequence, which can be particularly beneficial for identifying crucial past observations impacting future predictions. Previous studies have used similar architectures inspired by the informer or transformer models to forecast COVID-19. For example, [20] proposed an Interpretable attention network based on the informer model to forecast COVID-19 with promising results. Authors in [21] proposed a method based on transformers to capture short-term and long-term dependencies within the COVID-19 timeseries. The model used publicly available COVID-19 data such as confirmed cases, deaths, community mobility trends, and demographic information to produce predictions. Similarly, [22] and [23] developed a deep learning model based on the transformers for forecasting COVID-19 and their results outperformed state-of-the-art models. Therefore, deep learning models based on the transformers have shown their potential in capturing the short-term and long-term dependencies in timeseries data which results in accurate forecasting ability.

D. N-BEATS MODEL

N-BEATS is a deep learning-based forecasting model introduced in 2019 by Oreshkin et al. [9], designed to overcome the limitations of traditional models, offering a more flexible and interpretable framework. Unlike RNNs, it uses stack-propagated feedforward neural networks for timeseries forecasting and has been shown to outperform several state-of-the-art models in various datasets, including energy consumption and temperature data [9], [24].

The N-BEATS network is composed of two sub-networks: the Backcast and Forecast networks. The Backcast network is used to generate the history representation and capture the past information to be projected on the future forecasts, while the Forecast network takes the history representation and generates the forecast for a certain number of steps into the future. Notably, the architecture allows the model to be trained using standard backpropagation techniques due to its full differentiability.

N-BEATS architecture consists of multiple identical blocks, with each block having a backcast and forecast path. Each block in the architecture is constructed of fully connected layers, where each layer is equipped with normalization and activation functions. The backcast and forecast paths are separate in each block but are built from the same set of functions learned during the training process. The input to each block is the residual error between the original input series and the backcast of the previous block. This residual input mechanism allows N-BEATS to generate multiple additive components of the forecast. These blocks are stacked together, allowing the model to extract and learn complex temporal patterns at each level. The final forecast of the N-BEATS model is the sum of the forecasts from all blocks, allowing for a highly composite and nuanced prediction output [9], [24].

Previous studies have used the N-BEATS network for forecasting COVID-19 variables. For example, Papastefanopoulos et al. [25] implemented multiple forecasting models such as N-BEATS to forecast the number of confirmed COVID-19-positive cases, recovered cases, and death cases in 10 different countries. As they stated in their results, "a one-size-fits-all approach does not exist" but statistical models prevail over the deep learning counterparts such as N-BEATS in their experiments. However, the researchers noted that the lack of large volumes of data for training deep learning models might have contributed to this finding. Additionally, it is important to consider the temporal context in which this study was conducted. The data examined was from 2020, a period when COVID-19 data was limited. Moreover, the models relied solely on historical data as input, without any additional features. The absence of additional input features may limit the generalizability of the findings. It is thus crucial to replicate this study on more recent and extensive datasets that incorporate additional input features to strengthen the generalizability of these findings.

On the other hand, in a study done by Jin et al. [26] COVID-19 positive cases and death cases in India, the US, and the UK were forecasted with multiple single variate deep learning algorithms as well as traditional statistical models. The authors used only historical data and provided a comparison of the results in which the deep learning models such as N-BEATS outperform statistical models by far in all of their experiments. The authors have used the most updated data from 2022, but they didn't include any additional input features into their models to provide a multivariate forecast. Therefore, deep learning networks particularly the N-BEATS can provide more accurate results than statistical models in forecasting in case the input data is sufficient and suitable.

E. MULTI-VARIATE COVID-19 FORECASTING

Machine learning excels in forecasting disease incidence by utilizing multiple variables from external sources, unlike traditional methods that rely solely on historical data. This approach enables more robust predictions, though the selection of the optimal input variables continues to be a challenging area of research interest. For example, Neural networks, a deep learning algorithm, have been applied to forecast the number of COVID-19 positive cases. Wieczorek et al. [27] in their first study developed a complex network of artificial neurons (ANNs) containing multiple layers and the model was used to precisely forecast the COVID-19 case count and spread [28]. Their model considered the geographical conditions, i.e., location, latitude, and longitude, as input data and used a dataset extracted from the actual number of cases of the last two weeks obtained from each region as a training dataset

Pinter et al. [29] studied a dataset from Hungary to forecast the number of infected people and the mortality rate. They used a hybrid deep learning approach including a Multi-layered Perceptron (MLP) and Imperialist Competitive Calculation (MLP-ICA). The MLP was used as the predictor in the model while ICA was used as the optimizer. The model was optimized to forecast for a horizon of nine days and the results were compared against an adaptive network-based fuzzy inference system (ANFIS) [30].

In previous studies, wastewater-based epidemiology (WBE) as a method of tracking the presence of SARS-CoV-2 RNA in wastewater samples, has shown the ability to provide an early warning of viral outbreaks in a community [10]. Therefore, many researchers considered the wastewater data in their COVID-19 forecasting models to increase the forecasting accuracy. For example, Ai et al. [31] used viral (SARS-CoV-2) load and location identity, domain-specific features like biochemical parameters of wastewater, geographical parameters of the sewer sheds, and some socioeconomic parameters of the community data in multiple deep learning architectures to provide a forecast for COVID-19 positive cases. Their results showed that those deep learning models that considered WBE outperformed other forecasting models.

III. METHOD

In this research, we propose a novel model that combines the N-BEATS [9] forecasting model with a dual self-attention mechanism based on the Transformer model inspired by [32] to forecast the number of COVID-19 hospitalizations. Our methodology is rooted in leveraging the unique strengths of each model: the ability of N-BEATS to extract and learn complex temporal patterns of the timeseries, and the capability of the Transformer model with its self-attention mechanism to capture intricate long-range dependencies within the data. The following sections provide a comprehensive explanation of our unique real-world dataset, the proposed model's architecture, the training process, and the performance evaluation metrics. This innovative integration aims to harness the combined predictive power of these models, offering a robust and efficient solution to the complexities inherent in forecasting the volatile nature of COVID-19 hospitalization data.

A. DATASET COLLECTION

During the pandemic situation, the Government of Saskatchewan, Canada [33] gathered epidemiologic data of the COVID-19 pandemic in different Saskatchewan (SK) territories. The hospitalization counts for the City of Regina, the capital city of Saskatchewan, are recorded and publicly published daily. Therefore, the dataset includes dally hospitalized cases related to COVID-19 from 2020/08/04 to 2022/02/06. Our goal is to forecast the number of hospitalization cases in the city of Regina.

To increase the accuracy of the forecast, in addition to the hospitalization counts for Regina, the hospitalization counts of cities such as Saskatoon, Calgary, Ottawa, Toronto, Vancouver, and Montreal are also collected in the dataset. To gather data from transportation hub cities, we used a combination of web scraping and publicly available data sources. Table 1 summarizes all data that is present in the dataset regarding the daily number of hospitalization data from the mentioned cities and their description.

Wastewater has been used to study COVID-19 epidemiology because the virus or its RNA is expelled into the water drain system via feces of infected individuals [39], [40]. The University of Regina is gathering the wastewater samples of Regina and analyses the samples to report the SARS-COV-2 RNA load in Regina's wastewater. Raw untreated sewage is picked up from the wastewater plant and then concentrated in the lab environment. The solids of the samples get broken up and the genetic material is extracted from them. The SARS-CoV-2 RNA load in wastewater (RNA copies/mL) is then reported [41].

B. DATASET ANALYSIS

The epidemic status of Regina is supposed to be related to the other cities of Canada, especially the ones that have international airports. Studies have shown that transportation hub cities, such as airports and train stations, have been identified as potential hotspots for the spread of the contagious virus. The introduction of COVID-19 to regions has been through international airports, and the virus spread around the region is more attributed to local trips between cities [42]. In this research, we consider Regina as a non-international transportation hub as it has a small airport, and most international travelers with the intention to visit Saskatchewan have to first arrive at one of the major Canadian international airports and then travel locally to SK. By incorporating data from these

TABLE 1. Input Hospitalization data from all cities and their correlation amount with Regina.

Description	Relation	Maximum Correlation with Regina	Lead time (Days)
Regina [33]	Non-transportation hub city – to be forecasted	-	
Calgary [34]	Canadian Transportation hub	0.60	24
Toronto [35]	Canadian Transportation hub	0.52	12
Ottawa [36]	Canadian Transportation hub	0.44	3
Montreal [37]	Canadian Transportation hub	0.43	24
Vancouver [38]	Canadian Transportation hub	0.48	32
Saskatoon [33]	SK Largest City	0.75	13

transportation hub cities, we can better capture the regional variations in the spread of the virus. Therefore, the pandemic status of the transportation hub cities is an important external data source that can help increase the accuracy of COVID-19.

A Time lagged Pearson Correlation analysis [43] is done to assess and confirm the correlation between Regina's hospitalization and those transportation hub cities, and the maximum amount of the correlation at the reported time lag. The correlation results suggest that the hospitalization data from all of those transportation hub cities are correlated with Regina's hospitalization and all of those cities' hospitalization counts lead Regina's hospitalization. Although the calculated amount of correlation and the lead time are not directly used in the proposed forecasting model, these results confirm that the additional input data would help the model provide a more accurate forecast.

Mathematically, the time-lagged correlation for two timeseries x_t and y_t, for delay (time displacement or lag) L, represented as $r(L)$, is calculated as follows [43]:

$$r(L) = \frac{\sum_t [(x_t - \overline{x}) * (y_{t-L} - \overline{y})]}{\sqrt{\sum_t (x_t - \overline{x})^2}\sqrt{\sum_t (y_{t-L} - \overline{y})^2}} \tag{1}$$

where:

- r is the correlation coefficient.
- x_i and y_i are values of the x-variable and y-variable in a sample respectively.
- $\overline{x}$ and $\overline{y}$ are mean of the values of the x-variable and y-variable respectively.

Therefore, the maximum correlation is calculated with the following algorithm:

Algorithm 1 Calculating the Maximum Correlation

```
1: Input: timeseries x_t, y_t, MaxLag
2: Output: Maximum Correlation, OptimalLag
3: Maximum Correlation = NEGATIVE_INFINITY
4: OptimalLag = 0
5: While lag <= MaxLag do
6:    x'_t = x_t
7:    y'_t = y_{t-lag}
8:    correlation = Compute r(L) (x'_t, y'_t) # From Eq. (1)
9:    if correlation > Maximum Correlation then
10:       Maximum Correlation ← correlation
11:       OptimalLag = lag
12:   end if
13:   lag = lag+1
14: end While
15: return Maximum Correlation, OptimalLag
```

In addition to collecting data from hospitalization cases, we have collected and analyzed Regina's wastewater data. Given the things we know about viral dynamics in individuals and fecal shedding, wastewater surveillance is expected to lead hospitalization by a few days. The presence of the virus's RNA in a patient's shedding occurs shortly after an infection, while an infected person must develop symptoms and seek treatment to be hospitalized through the health system [40], [44].

The same time-lagged correlation analysis shows that the SARS-CoV-2 RNA load in Regina's wastewater has a correlation amount of 0.62 with Regina's hospitalization and leads Regina's hospitalization by 9 days. Therefore, the SARS-CoV-2 RNA load in wastewater is also included as the input data to the forecasting model.

C. MODEL SUMMARY AND CONFIGURATION

This section provides an overview of the proposed DSA-BEATS model and its modules. In this paper, we have proposed a multistep multivariate approach to provide a robust accurate forecast of the number of hospitalized cases of COVID-19. The approach uses multiple variables such as the rate of presence of the SARS-CoV-2 virus in wastewater of the city and the epidemiological status of the adjacent cities and provinces to create a multivariate timeseries problem. Then a deep learning neural network is specifically designed to learn the relationships between the timeseries.

Figure 1 demonstrates all of the modules of our proposed DSA-BEATS model with the connections and flow of information along it. In this research, we propose a network that exploits two convolutional structures to embed the timeseries data into two representations with temporal information of different scales, namely, the Global Temporal Convolution representation and the Local Temporal Convolution representation. These two representations are fed into self-attention modules to capture the dependencies between multiple series. The final forecast is generated by summing up the outputs of

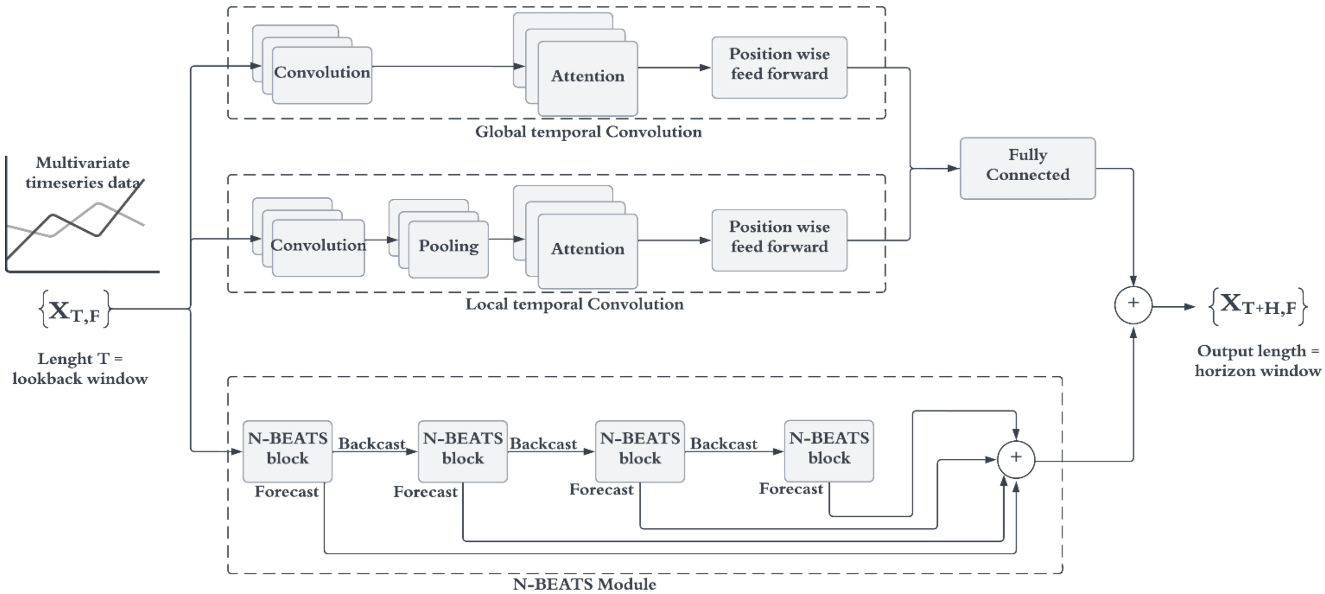

FIGURE 1. **DSA-BEATS Model structure. This figure shows different modules of the model.**

both attention modules ($\widehat{y_{Attention}}$) and the N-BEATS module ($\widehat{y_{N-BEATS}}$). The details of each block are discussed in the following sections.

All the mentioned data inputs form a multivariate timeseries represented in the Matrix $X_{T,F}$ where T is the timesteps and F is each feature (timeseries). Therefore, each $x_{t,f}$ represents the value of timeseries f at timestep t. In this network, the model input is a set of the observed values, and the timeframe of those observed values is named the lookback window with a length of $w \leq T$, which ends with the last observed value y_T. The forecasting problem is formulated as follows: a forecast horizon of length H ahead from the current timestep (t) is predicted from the past observations from $X_{t-w,F}$ until $X_{t,F}$. Therefore, we predict the future values of hospitalization as $[y_{t+1}, \ldots, y_{t+H}] \in R^T$ given the values of the lookback window. In this research the horizon is fixed at $H = 12$, and the lookback window size is fixed at $w = 32$ and we predict the hospitalization cases in a rolling fashion until the end of the test set. The model is introduced to a window of $W = 32$ days of input and generates 12 days of output. Then the window is lagged for 1 day and the process is repeated up to the end of the data.

D. GLOBAL TEMPORAL CONVOLUTION MODULE

Inspired by [32], the proposed network includes a temporal convolution module to learn the time-invariant patterns in the dataset. Convolutional neural networks have demonstrated their potential in capturing features and facilitating parallel computing. Hence, we propose utilizing a combination of the convolutional neural networks by using a kernel of size $1 \times T$, known as global temporal convolution [32], to extract time-invariant patterns for the timeseries.

Each filter in this module scans through the input matrix X, resulting in a $1 \times F$ vector using the Rectified Linear Unit (ReLU) activation function. The resulting vectors are merged ultimately to produce an output matrix M_g of size $n_g \times F$, where n_g is the number of filters in the global temporal convolution. Each column of the matrix M_g can be interpreted as a learned representation of each timeseries.

E. LOCAL TEMPORAL CONVOLUTION MODULE

In addition to the global convolution, a local temporal convolution is deployed in the model to capture the impact of each timeseries on one another, considering relative time steps. The global temporal convolution is designed to capture long-term and overall relations between time steps, and the local temporal convolution attempts to determine local temporal relations, which will be advantageous for forecasting the future. The length of kernels used in local temporal convolution is a hyperparameter l, where $l < T$. Each kernel moves along the time dimension and generates a matrix $P_{k,L}$. Then a $1 \times F$ max-pooling is applied over each row of the matrix $P_{k,L}$ to capture the most representative features on each row of $P_{k,L}$ to represent local temporal relations in timeseries. Consequently, an output matrix M_l of size $n_l \times F$ is created, where n_l is the number of filters in the local temporal convolution.

F. SELF-ATTENTION MODULE

The model utilizes a self-attention module inspired by the Transformer [10] due to its robust feature-extraction capability. The primary goal of the self-attention module is to capture the dependencies among various timeseries. For each learned representation of the timeseries provided by the convolution modules, the self-attention module learns the relationship with other learned representations, including itself. The self-attention module comprises a stack of N identical layers, with each layer consisting of two sub-layers: a self-attention layer and a position-wise feed-forward layer. The attention block structure has the same encoder-decoder configuration as discussed in [10].

The attention function is defined as mapping a query vector and a set of key-value pairs to an output vector. Specifically, given a query vector and a set of key-value pairs, the attention mechanism computes a weight for each key-value pair by taking the inner product between the query and the corresponding key. These weights are then normalized using a softmax function and used to compute a weighted sum of the value vectors, resulting in an output vector that is a function of the input query and key-value pairs at every other position in the timeseries. In the self-attention module of the proposed model, the output of the global temporal convolution, M_g, is projected onto a set of matrices Q_g, K_g, and V_g and representing the queries, keys, and values, respectively. Then, the attention mechanism computes the output matrix by computing the weighted sum of the value matrix, where the weight for each position is computed as the inner product between the corresponding query and key vectors. The resulting output matrix is used to represent the dependencies between different time steps in the input sequence, which can be utilized for forecasting. Mathematically, self-attention computes the scaled dot product for the global convolution by the following equation:

$$Z^G = softmax(\frac{Q^G(K^G)^T}{\sqrt{d}_k})V^G \tag{2}$$

In which, d_k is the dimension of the keys. To consider the information from different subspaces of representation at different positions, multi-head attention is applied. To achieve the final representation Z_O^G, we linearly project and concatenate the weighted representations.

ReLU activation is used between two linear transformations in the feed-forward layer which is mathematically represented as:

$$F^G = ReLU\left(Z_O^G w_1 + b_1\right) W_2 + b_2 \tag{3}$$

The self-attention is applied in the same way to the output Matrix of the local convolution, M_l and the final output of F^L is generated. Then a fully connected neural network layer is applied to combine both F^G and F^L and get a prediction $(\widehat{y_{Attention}})$ of the values for the length of forecasting horizon $H = 12$.

G. N-BEATS MODULE

The input window from Matrix X first gets reshaped to be prepared for the N-BEATS network. Data reshaping is done by flattening the multi-variate timeseries data along the time axis which result in a one-dimensional representation of the input window. Therefore, the input Matrix of order $W \times F$ is converted into a vector with a length of $W \times F$. This reshaping is done through a flattening layer which is placed before the N-BEATS network. The flattened univariate input vector which includes inputs from all of the variables with the length of the window size (w) is then passed to the N-BEATS network. The output of the N-BEATS module is then summed up with the output of the attention module to produce the final forecast.

The fundamental component of the N-BEATS module architecture comprises a fully-connected non-linear regressor, which receives the historical data of a timeseries as input and generates multiple forecasts for the given horizon. Therefore, the N-BEATS module consists of stacks of blocks of fully connected (FC) neuron layers, which conceptually can be described as a multivariate linear regressor followed by a ReLu [45]. The first block is fed with the past values of the input timeseries of size $W \times F$ for the length of the lookback window and generates forecast (of size $H \times F$) and backcast outputs (of size $W \times F$). Backcast output is the block's contribution to decomposing the input and the subsequent blocks receive the backcast output (of size $W \times F$) of the previous block through residual connections. The network consists of multiple blocks and the process is repeated.

Taking a look into the mathematics, the N-BEATS architecture runs a residual recursion over the entire input window and sums block outputs to make its final forecast. Consider L hidden layers are included in each residual block and there are R residual blocks in the network. The input and residual block and layers are shown with r and ℓ, respectively. Take $W^{r,\ell}$ as the weight of the fully connected layer and take $b^{r,\ell}$ as biases, then the fully connected layer would be as follows:

$$C_{r,\ell}\left(h^{r,\ell-1}\right) \equiv \text{ReLu}\left(W_{r,\ell}h^{r,\ell-1} + b^{r,\ell}\right) \tag{4}$$

The N-BEATS operation is described as [9], [46]:

$$x^r = ReLu\left[x^{r-1} - \hat{x}^{r-1}\right], \tag{5}$$

$$h^{r,1} = C_{r,\ell}\left(x^r\right), \ldots, h^{r,l} = C_{r,\ell}\left(h^{r,\ell-1}\right), \tag{6}$$

$$\hat{x}^r = B^r h^{r,l}, \quad \hat{y}^r = F^r h^{r,l}. \tag{7}$$

Projection matrices have dimensions of $F^r \in R^{H\times d_h}$ and $B^r \in R^{w\times d_h}$ assuming that $\hat{x}^0 \equiv 0$ and $x^0 \equiv x$. Therefore, the final N-BEATS output is the sum of the output values from all of the residual blocks:

$$\widehat{y_{N-BEATS}} = \sum\nolimits_r \hat{y}^r \tag{8}$$

With all the details described around our proposed DSA-BEATS model, algorithm 2 summarizes the whole process of our proposed DSA-BEATS network.

IV. EXPERIMENTS AND RESULTS

In this section, the proposed DSA-BEATS model is applied to our specific dataset, and the results are compared with

Algorithm 2 DSA-BEATS Network

Input: Input Matrix $X_{T,F}$ with multivariate series for the size of the lookback window
Output: Forecasted output X_{T+H}
Initialize global and local temporal convolutions, self-attention module, and N-BEATS module
$i = 0$
while $i < D$ **do** # for each series in X
  Apply global temporal convolution to the series obtaining M_g
  Apply local temporal convolution to the series obtaining M_l
  Apply max-pooling to M_l
  $i = i + 1$
end while
$j = 0$
while $j < N$ **do** # for each layer in the self-attention module
  Apply the self-attention layer to M_l and M_g to generate key-value pairs Q_L, K_L, V_L and Q_G, K_G, V_G
  Compute attention scores for global module as Z^G using Q_G, K_G, V_G # Equation (2)
  Apply multi-head attention to Z^G and linearly project to obtain the final representation Z'^G
  Apply position-wise feed-forward layer on Z'^G to get F^G # Equation (3)
  Compute attention scores for local module s Z^L using Q_L, K_L, V_L # Equation (2)
  Apply multi-head attention to Z^L and linearly project to obtain the final representation Z'^L
  Apply position-wise feed-forward layer on Z'^L to get F^L # Equation (3)
  $j = j + 1$
end while
k = 0
while k < M **do** # for each block in N-BEATS module
Apply N-BEATS block to X to get backcast and forecast outputs
Subtract the backcast from X to get the remainder
Add the forecast to a cumulative forecast
Set X as the remainder
k = k + 1
end while
Set N-BEATS forecasted output as $\widehat{y_{N-BEATS}}$
Combine F^G and F^L using a dense layer to get self-attention based prediction $\widehat{y_{Attention}}$
Sum $\widehat{y_{N-BEATS}}$ and $\widehat{y_{Attention}}$ to get the final prediction X_{T+H}
return X_{T+H}

other state-of-the-art models such as the Informer and the N-BEATS. The Informer model has been chosen as the most widely adopted self-attention-based forecasting method. Its strong performance and widespread acceptance make it an essential benchmark for comparison. The N-BEATS model is also selected as a robust and representative example of cutting-edge neural network approaches in the forecasting field.

To report the performance of the model, two well-known metrics are selected and used. These two metrics, RMSE and MAPE values, are calculated and reported for the forecasted versus the actual values in the dataset for every model. The RMSE and MAPE values are calculated for the test set and reported as the result of the forecast.

A. MODEL CONFIGURATION

To train the model, we need to first split the dataset into train, validation, and test sets. The data was split into training, validation, and test sets, to ensure that the network was trained on the training set, optimized and checked on the validation set, and finally evaluated on the test set. Adam [47] optimizer has been used for training. 90% of the data was used for training and validation and the remaining 10% of the data was held out for the test set.

The test set includes 55 days and the model is going to forecast for a horizon of 12 days ahead. After training is done, the model is introduced to the last window of input ($w = 32$) days of the test set and will forecast for the next 12 days ($H = 12$). However, during the test phase, each time in the future is only forecasted once meaning the 32-day-lookback window is shifted by 12 days (as the model has already predicted those 12 days) to avoid any repeated forecasts. The train and validation subsets are also used for hyperparameter tuning of the model which is done by a grid search. The hyperparameter tuning was done to minimize the MAPE and then, the model was tested on the test set and the result on the test set was reported. Since the test set includes 55 days and the horizon is 12 days, the model generates forecasts 5 times to cover the whole test set and the reported

TABLE 2. RMSE and MAPE errors for the Proposed model vs. the state-of-the-art forecasting methods.

Description	RMSE	MAPE
N-BEATS [9] - Univariate	13.96	32.95
N-BEATS [9] - Multivariate	10.03	18.46
Informer [18] - Univariate	13.03	32.46
Informer [18] - Multivariate	9.75	21.98
DSA-BEATS	**6.11**	**14.23**

MAPE and RMSE results are the average results through the whole test set.

B. COMPARATIVE ANALYSIS

For comparative analysis, we have used the original N-BEATS model proposed in [9] and the Informer [18] model. In the univariate attempts, the COVID-19 hospitalization has been forecasted for a horizon of 12 days as a univariate timeseries problem considering the past number of hospitalizations from Regina. In the multivariate forecasting attempts, the same models have been used considering all the inputs mentioned in the dataset. The proposed model is trained with all the inputs from the dataset. Table 2 provides all the mentioned comparative results. In our evaluation, our proposed DSA-BEATS model consistently outperformed both the standalone N-BEATS and Informer [18] models in terms of forecasting accuracy when predicting COVID-19 hospitalizations. This performance superiority was apparent across both evaluation metrics, indicating the model's robust predictive capabilities. Moreover, the resilience of the DSA-BEATS model to overfitting was distinctly evident. Despite the test set comprising 55 unseen samples, and the model having to produce a forecast five times due to a 12-day horizon to cover the entire test set, the model maintained its high performance. This result suggests that our model can generalize well to unseen data. Detailed analyses of these findings will be provided in the subsequent sections, offering a comprehensive understanding of the strengths and potential limitations of our DSA-BEATS model. In addition, Table 2 shows that the inclusion of supplemental inputs, specifically information regarding transportation hub cities and RNA samples in the wastewater data, played a significant role in enhancing the accuracy of our forecasts.

C. ABLATION STUDY

In order to understand the contributions of the individual components of our DSA-BEATS model, we conducted an ablation study. This study involved the systematic removal of individual components, namely, the global temporal convolution part, the local temporal convolution part, and the N-BEATS component, from our DSA-BEATS model. The purpose of this exercise was to assess the impact of each component on the overall performance of the model Specifically, we focused on understanding the impact of integrating dual self-attention mechanism into the N-BEATS model and the performance of the standalone N-BEATS and Transformer models.

TABLE 3. RMSE and MAPE errors for the ablation study.

Description	RMSE	MAPE
DSA-BEATS (without Global temporal module)	8.36	17.07
DSA-BEATS (without Local temporal module)	8.12	16.90
DSA-BEATS (without N-BEATS module)	14.27	29.83
DSA-BEATS	**6.11**	**14.23**

According to the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) metrics in Table 3, removing the N-BEATS module from the DSA-BEATS model results in a significant decrease in performance, indicating the essential role played by the N-BEATS component. This is in line with previous research suggesting that the N-BEATS have a solid performance in forecasting the time-series data. Furthermore, removing any of both Global and Local temporal convolutions also exhibits a reduction in performance upon removal of their respective parts. However, the impact is relatively less severe than that observed upon removal of the N-BEATS component. This can be attributed to the overlap in the features learned by the two branches. In essence, when one part is removed, some of the lost features can be obtained from the other part.

Consequently, upon integrating the dual self-attention mechanism into the N-BEATS model to create the DSA-BEATS model, the resulting combination exhibited a significant improvement in performance according to both RMSE and MAPE results. This model leveraged the strengths of both individual models, effectively capturing complex temporal patterns resulting in the lowest error rates. The findings from this ablation study underscore the effectiveness of our approach in providing a robust method for forecasting COVID-19 hospitalizations, highlighting the unique benefits of integrating the N-BEATS and Transformer models.

V. DISCUSSION

Both the Informer [18] model and the N-BEATS model performed better when Regina's hospitalization data was augmented with the hospitalization data from the transportation hub cities and the amount of the virus RNA in wastewater data. Authors in [25] stated that COVID-19 data is not long enough for deep learning models such as the N-BEATS network to provide competitive performance. However, those proposed additional inputs, i.e., the hospitalization from the transportation hub cities and the daily amount of the virus RNA in wastewater samples helped to build a multivariate forecasting problem resulting in a more accurate forecast of Regina's hospitalization. These findings align with the previous research. For example, in research by Peccia et al, [40] a simple fully connected neural network model was developed that could forecast the number of COVID-19 hospitalization by using the RNA load data in the wastewater. By using a differential equation-based model, Kaplan et al. [48] demonstrated that hospitalizations could be anticipated from the

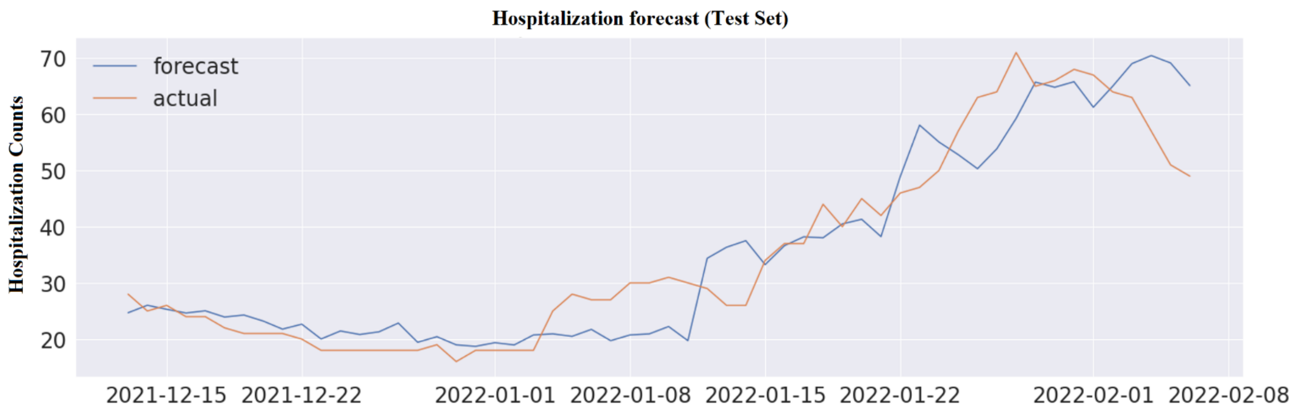

FIGURE 2. **Actual VS Forecasted values side-by-side in DSA-BEATS model.**

SARS-CoV-2 RNA load in wastewater. On the other hand, while some studies have used neighboring regions' data to build a forecasting model [16], we couldn't find any previous study that used the pandemic data from the transportation hub cities to build a model for forecasting other places.

While using the additional inputs from wastewater and transportation hub cities helped the N-BEATS and Informer models provide a more accurate forecast, the proposed DSA-BEATS model performed even better on the test set. One of the critical factors contributing to the model's superior performance was its ability to incorporate the dual self-attention mechanism effectively. This inclusion allowed the model to allocate varying levels of attention to different periods in the timeseries data, capturing significant past observations and their influence on future hospitalizations accurately. This attribute was especially crucial in effectively forecasting the volatile COVID-19 hospitalization data, which tends to fluctuate due to various external factors. As depicted in Figure 2, a side-by-side comparison of actual versus predicted hospitalization values on the test set clearly illustrates the success of the DSA-BEATS model. The forecasted values closely mimic the actual values, indicating that the model has captured the historical trends in the training set and was able to transfer this understanding to produce accurate forecasts for the test set. The similar shapes and matched trends of the forecasted and actual values affirm the model's capability to capture both the general trend and inherent volatility in the COVID-19 hospitalization data. Therefore, the outcomes of our study underscore the efficacy of this hybrid model in effectively capturing both the trend and volatility inherent in the COVID-19 hospitalization data.

VI. CONCLUSION

This study introduced the DSA-BEATS model, a novel hybrid forecasting approach that combines the dual self-attention mechanism of the transformer model with the forecasting abilities of the N-BEATS network. Our research underscored the effectiveness of this model in accurately predicting COVID-19 hospitalizations, a task characterized by volatility and a need to consider complex temporal relationships. To tackle this problem, multiple inputs were considered in a novel deep-learning model to forecast the hospitalization counts related to COVID-19 in Regina, Saskatchewan. We have used the epidemic data of the major transportation hub cities of the country and the virus RNA load in Regina's wastewater as auxiliary inputs to the model. The results analysis suggests these auxiliary inputs have proved to be beneficial when forecasting hospitalization in epidemic situations, and the proposed DSA-BEATS model was able to get the most out of this data and provide an accurate forecast.

The DSA-BEATS model performed exceptionally well in our experimental setup, outperforming both standalone models according to the RMSE and MAPE metrics. The incorporation of the dual self-attention mechanism allowed the model to allocate varying attention to different time periods in the data, leading to the accurate capture of significant past observations and their influence on future values. This enhanced ability to understand intricate temporal dependencies was particularly effective in dealing with the volatile nature of COVID-19 hospitalization data.

In conclusion, the DSA-BEATS model presents a significant step forward in timeseries forecasting, particularly in the context of pandemic response planning and management. We hope that our study will inspire further research in this direction, contributing to the ongoing global efforts to manage and mitigate the impacts of COVID-19 and future pandemics.

VII. FUTURE WORK

While our findings are promising, we believe there is room for further improvement and exploration. Future work could investigate the impact of other external factors on COVID-19 hospitalizations and the use of other hybrid models or ensemble techniques. Additionally, the application of the DSA-BEATS model to other complex timeseries forecasting tasks would further validate its effectiveness and flexibility.

While the DSA-BEATS was applied to the COVID-19 hospitalization dataset its potential extends beyond this specific context. It is equally suited for any other multivariate timeseries forecasting problem, particularly where auxiliary inputs are closely correlated with, and predictive of, the primary timeseries.

Although the primary focus of this study was not to delve deeply into the correlations between the epidemic statuses of Canadian cities, we utilized Pearson Correlation analysis as a means to validate the beneficial effects of our inputs on the model. However, exploring the intricate relationships between epidemic statuses across Canadian cities, particularly when coupled with transportation statistics such as flight and vehicular travel data, presents an intriguing avenue for future research. This line of investigation, while outside the scope of our current study, can potentially contribute valuable insights to enhance predictive models and strategic responses to epidemics.

VIII. LIMITATIONS

While the DSA-BEATS model has demonstrated promising results in the context of our study, it is important to recognize and articulate the limitations that may impact the interpretation and generalization of the findings. The fusion of Dual-self attention and N-BEATS adds a layer of complexity to the model, making it computationally expensive. This complexity may hinder its applicability in scenarios with limited computational resources. Moreover, While the model has shown to be effective in forecasting, it may lack interpretability in understanding the underlying relationships between variables. The complex architecture may make it difficult to explain why specific predictions are made, which could be a limitation in certain applications where clear interpretability is crucial.

A limitation of this study is the reliance on RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) for evaluating the forecasts. Although these metrics are the most common metrics being used in forecasting studies, they have their own limitations. Specifically, RMSE can be sensitive to outliers, leading to overestimation of general prediction error if extreme values are present. On the other hand, MAPE can be biased, especially with values close to zero, and may not equally penalize overestimations and underestimations. Thus, the interpretation of the results must be approached with an understanding of these inherent limitations, and future studies may consider incorporating supplementary metrics to provide a more comprehensive evaluation of forecast accuracy. However, the DSA-BEATS model has performed better than the studied state-of-the-art methods and has provided a better performance in both metrics ensuring the model's competitive performance.

In this study, we have introduced the utilization of hospitalization data from transportation hub cities as an exogenous input to forecast the situation in non-transportation hub areas. However, this methodology may present challenges in settings where transportation hub cities do not play a comparable role. Specifically, in regions such as many European countries where population density is higher and transportation dynamics are diverse, the concept of a central transportation hub may not apply in the same way as it does in areas like Canada. Under such circumstances, the model would necessitate the identification of alternative exogenous variables that can guide the hospitalization forecasts. The complexity and availability of these alternative inputs might vary significantly across different regions, introducing an additional layer of complexity to the model's implementation. The successful application of this approach in other geographical contexts may thus require careful consideration and possibly adaptation to local conditions, reflecting the heterogeneous nature of transportation systems and public health landscapes worldwide.

REFERENCES

[1] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, May 2020.

[2] M. Cascella, M. Rajnik, A. Aleem, S. Dulebohn, and R. Napoli, *Features, Evaluation, and Treatment of Coronavirus (COVID-19)*. Treasure Island, FL, USA: StatPearls, 2023.

[3] S. M. H. Bamakan, P. Malekinejad, M. Ziaeian, and A. Motavali, "Bullwhip effect reduction map for COVID-19 vaccine supply chain," *Sustain. Operations Comput.*, vol. 2, pp. 139–148, 2021.

[4] D. Freedman, R. Pisani, and R. Purves, *Statistics*. New York, NY, USA: WW Norton & Company, 2007.

[5] T. Wang, S. Ma, S. Baek, and S. Yang, "COVID-19 hospitalizations forecasts using Internet search data," *Sci. Rep.*, vol. 12, no. 1, p. 9661, Jun. 2022.

[6] A. Rodríguez, "DeepCOVID: An operational deep learning-driven framework for explainable real-time COVID-19 forecasting," in *Proc. AAAI*, vol. 35, no. 17, May 2021, pp. 15393–15400.

[7] X. Jin, Y.-X. Wang, and X. Yan, "Inter-series attention model for COVID-19 forecasting," 2020, *arXiv:2010.13006*.

[8] M.-B. Park, "The effect of advances in transportation on the spread of the coronavirus disease: The last is Africa and endemic," *J. Public Health Res.*, vol. 10, no. 3, p. 2058, Jun. 2021.

[9] B. Oreshkin, D. Carpo, N. Chapados, and Y. Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," in *Proc. ICLR*, 2019, pp. 1–31.

[10] A. Vaswani, N. Shazeer, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and L. Polosukhin, "Attention is all you need," in *Proc. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[11] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, Melbourne, VIC, Australia: OTexts, 2021.

[12] A. E. Permanasari, I. Hidayah, and I. A. Bustoni, "SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence," in *Proc. Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Oct. 2013, pp. 203–207.

[13] A. Carrión-García, J. Jabaloyes, and A. Grisales, "35—The growth of COVID-19 in Spain. A view based on time-series forecasting methods," in *Data Science for COVID-19*, U. Kose, D. Gupta, V. H. C. de Albuquerque, and A. Khanna, Eds. Academic Press, 2021, pp. 643–660.

[14] X. Zhang, C. Zhong, J. Zhang, T. Wang, and W. W. Y. Ng, "Robust recurrent neural networks for time series forecasting," *Neurocomputing*, vol. 526, pp. 143–157, Mar. 2023.

[15] C. Comito and C. Pizzuti, "Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review," *Artif. Intell. Med.*, vol. 128, Jun. 2022, Art. no. 102286.

[16] R. Chandra, A. Jain, and D. S. Chauhan, "Deep learning via LSTM models for COVID-19 infection forecasting in India," *PLoS One*, vol. 17, no. 1, Jan. 2022, Art. no. e0262708.

[17] M. A. A. Al-qaness, A. Dahou, A. A. Ewees, L. Abualigah, J. Huai, M. A. Elaziz, and A. M. Helmi, "ResInformer: Residual transformer-based artificial time-series forecasting model for PM2.5 concentration in three major Chinese cities," *Mathematics*, vol. 11, no. 2, p. 476, Jan. 2023.

[18] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. 35th AAAI Conf. Artif. Intell. (AAAI)*, 2021, pp. 11106–11115.

[19] Q. Zhu, J. Han, K. Chai, and C. Zhao, "Time series analysis based on informer algorithms: A survey," *Symmetry*, vol. 15, no. 4, p. 951, Apr. 2023.
[20] B. Zhou, G. Yang, Z. Shi, and S. Ma, "Interpretable temporal attention network for COVID-19 forecasting," *Appl. Soft Comput.*, vol. 120, May 2022, Art. no. 108691.
[21] S. Er, S. Yang, and T. Zhao, "COUnty aggRegation mixup AuGmEntation (COURAGE) COVID-19 prediction," *Sci. Rep.*, vol. 11, no. 1, p. 14262, Jul. 2021.
[22] H. Abbasimehr and R. Paki, "Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization," *Chaos, Solitons, Fractals*, vol. 142, Jan. 2021, Art. no. 110511.
[23] C.-J. Huang, Y.-H. Chen, Y. Ma, and P.-H. Kuo, "Multiple-input deep convolutional neural network model for COVID-19 forecasting in China," *MedRxiv*, 2020. [Online]. Available: https://doi.org/10.1101/2020.03.23.20041608
[24] K. G. Olivares, C. Challu, G. Marcjasz, R. Weron, and A. Dubrawski, "Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx," *Int. J. Forecasting*, vol. 39, no. 2, pp. 884–900, Apr. 2023.
[25] V. Papastefanopoulos, P. Linardatos, and S. Kotsiantis, "COVID-19: A comparison of time series methods to forecast percentage of active cases per population," *Appl. Sci.*, vol. 10, no. 11, p. 3880, Jun. 2020.
[26] W. Jin, S. Dong, C. Yu, and Q. Luo, "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning," *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105560.
[27] M. Wieczorek, J. Siłka, and M. Woźniak, "Neural network powered COVID-19 spread forecasting model," *Chaos, Solitons Fractals*, vol. 140, Nov. 2020, Art. no. 110203.
[28] M. Wieczorek, J. Siłka, D. Połap, M. Woźniak, and R. Damaševičius, "Real-time neural network based predictor for cov19 virus spread," *PLoS One*, vol. 15, no. 12, Dec. 2020, Art. no. e0243189.
[29] G. Pinter, I. Felde, A. Mosavi, P. Ghamisi, and R. Gloaguen, "COVID-19 pandemic prediction for Hungary; A hybrid machine learning approach," *Mathematics*, vol. 8, no. 6, p. 890, Jun. 2020.
[30] N. Ayoobi, D. Sharifrazi, R. Alizadehsani, A. Shoeibi, J. M. Gorriz, H. Moosaei, A. Khosravi, S. Nahavandi, A. G. Chofreh, F. A. Goni, J. J. Klemeš, and A. Mosavi, "Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods," *Results Phys.*, vol. 27, Aug. 2021, Art. no. 104495.
[31] Y. Ai, F. He, E. Lancaster, and J. Lee, "Application of machine learning for multi-community COVID-19 outbreak predictions with wastewater surveillance," *PLoS One*, vol. 17, no. 11, Nov. 2022, Art. no. e0277154.
[32] S. Huang, D. Wang, X. Wu, and A. Tang, "DSANet: Dual self-attention network for multivariate time series forecasting," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 2129–2132.
[33] (2022). *'Health and Wellness,' Government of Saskatchewan*. [Online]. Available: https://dashboard.saskatchewan.ca/health-wellness
[34] Government of Alberta. *COVID-19 Alberta Statistics*. Accessed: Mar. 2023. [Online]. Available: https://www.alberta.ca/stats/covid-19-alberta-statistics.htm
[35] City of Toronto. *COVID-19: Monitoring Dashboard*. Accessed: Mar. 2023. [Online]. Available: https://www.toronto.ca/community-people/health-wellness-care/health-programs-advice/respiratory-viruses/covid-19/covid-19-pandemic-data/covid-19-monitoring-dashboard-data/
[36] Ottawa public health. *COVID-19 Dashboard*. Accessed: Mar. 2023. [Online]. Available: https://www.ottawapublichealth.ca/en/reports-research-and-statistics/covid19-dashboard.aspx
[37] Gouvernement du Québec. *Data on COVID-19 in Québec*. Accessed: Mar. 2023. [Online]. Available: https://www.quebec.ca/en/health/health-issues/a-z/2019-coronavirus/situation-coronavirus-in-quebec
[38] BC Center for disease control. *British Columbia COVID-19 Dashboard*. Accessed Mar. 2023. [Online]. Available: https://experience.arcgis.com/experience/a6f23959a8b14bfa989e3cda29297ded
[39] U. Anand, B. Adelodun, A. Pivato, S. Suresh, O. Indari, S. Jakhmola, H. C. Jha, P. K. Jha, V. Tripathi, and F. Di Maria, "A review of the presence of SARS-CoV-2 RNA in wastewater and airborne particulates and its use for virus spreading surveillance," *Environ. Res.*, vol. 196, May 2021, Art. no. 110929.
[40] J. Peccia, A. Zulli, D. E. Brackney, N. D. Grubaugh, E. H. Kaplan, A. Casanovas-Massana, A. I. Ko, A. A. Malik, D. Wang, M. Wang, J. L. Warren, D. M. Weinberger, W. Arnold, and S. B. Omer, "Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics," *Nature Biotechnol.*, vol. 38, no. 10, pp. 1164–1167, Oct. 2020.
[41] *University of Regina, COVID-19 Information*. University of Regina. Accessed: Jul. 3, 2023. [Online]. Available: https://www.uregina.ca/term-updates/
[42] M. W. Levin, M. Shang, and R. Stern, "Effects of short-term travel on COVID-19 spread: A novel SEIR model and case study in Minnesota," *PLoS One*, vol. 16, no. 1, Jan. 2021, Art. no. e0245919.
[43] C. Shen, "Analysis of detrended time-lagged cross-correlation between two nonstationary time series," *Phys. Lett. A*, vol. 379, no. 7, pp. 680–687, Mar. 2015.
[44] D. A. Larsen and K. R. Wigginton, "Tracking COVID-19 with wastewater," *Nature Biotechnol.*, vol. 38, no. 10, pp. 1151–1153, Oct. 2020.
[45] V. Nair and G. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 807–814.
[46] B. Oreshkin, G. Dudek, P. Pełka, and E. Turkina, "N-BEATS neural network for mid-term electricity load forecasting," *Appl. Energy*, vol. 293, Jul. 2021, Art. no. 116918.
[47] P. Diederik, K. Ba, and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.
[48] E. H. Kaplan, D. Wang, M. Wang, A. A. Malik, A. Zulli, and J. Peccia, "Aligning SARS-CoV-2 indicators via an epidemic model: Application to hospital admissions and RNA detection in sewage sludge," *Health Care Manage. Sci.*, vol. 24, no. 2, pp. 320–329, Jun. 2021.

AMIRHOSSEIN MOTAVALI received the B.Eng. degree in electrical engineering degree from Yazd University, in 2015. He is currently pursuing the M.A.Sc. degree in software systems engineering with the University of Regina. His research interest includes deep learning applications in forecasting and data analysis.

KIN-CHOONG YOW (Senior Member, IEEE) received the B.Eng. degree (Hons.) in electrical engineering from the National University of Singapore, in 1993, and the Ph.D. degree from the University of Cambridge, U.K., in 1998. In September 2018, he joined the University of Regina, where he is currently a Professor with the Faculty of Engineering and Applied Science. Prior to joining the University of Regina, he was an Associate Professor with the Gwangju Institute of Science and Technology (GIST), Republic of Korea, from 2013 to 2018; a Professor with the Shenzhen Institutes of Advanced Technology (SIAT), China, from 2012 to 2013; and an Associate Professor with Nanyang Technological University (NTU), Singapore, from 1998 to 2013, where he was the Sub-Dean of Computer Engineering, from 1999 to 2005, and the Associate Dean of Admissions, from 2006 to 2008. He has published more than 100 top-quality international journal articles and conference papers. His research interests include artificial general intelligence and smart environments. He is a member of APEGS and ACM. He has served as a Reviewer for a number of premier journals and conferences, including IEEE Wireless Communications and IEEE Transactions on Education. He has been invited to give presentations at various scientific meetings and workshops, such as ACIRS, from 2018 to 2019; ICSPIC, in 2018; and ICATME, in 2021. He is the Editor-in-Chief of the *Journal of Advances in Information Technology* (JAIT).

NICOLE HANSMEIER received the Diploma and Ph.D. degrees in biology from Bielefeld University, Germany. She is currently an Associate Professor in Biology with the Luther College, University of Regina. She leads a program using wastewater-based surveillance to detect and monitor infectious diseases (in the province) to inform public health responses. Her research interests include infectious diseases, disease tracking within communities, and identification of biomarkers to diagnose adverse health effects. To this end, the research applies systems biological multi-omics and high-end imaging techniques. Her research is supported by the Canadian Foundation for Innovation, Innovation Saskatchewan, Health Canada, SHRF, Natural Sciences and Engineering Research Council, and Mitacs.

TZU-CHIAO CHAO received the Diploma and Ph.D. degrees in molecular biology from Bielefeld University, Germany. He is currently the head of the Cellular Impacts Facility within the Institute of Environmental Change and Society (IECS) and an Adjunct Professor with the Department of Biology, University of Regina. His research interests include biomarker discovery, microbial physiology and disease surveillance. His laboratory utilizes a combination of high-throughput omics technologies (genomics, transcriptomics, and proteomics) and microscopic techniques. His research is supported by the Canadian Foundation for Innovation, Innovation Saskatchewan, Saskatchewan Health Research foundation, Natural Sciences and Engineering Research Council, Mitacs and Health Canada.