IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 February 2023, accepted 9 February 2023, date of publication 14 February 2023, date of current version 22 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3245085

RESEARCH ARTICLE

A Multi-Head Self-Attention Transformer-Based Model for Traffic Situation Prediction in Terminal Areas

ZHOU YU[1,2], XINGYU SHI[1], AND ZHAONING ZHANG[1]
[1]College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China
[2]School of Intelligent Manufacturing and Transportation, Urban Vocational College of Sichuan, Chengdu 610100, China

Corresponding author: Zhaoning Zhang (zzhaoning@263.net)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1600103, and in part by the Key Projects of the Civil Aviation Joint Fund of the National Natural Science Foundation of China under Grant 2233209.

ABSTRACT Terminal operations management is an important part of air traffic management. Accurately detecting and predicting the operational status of the terminal area can help formulate more appropriate and efficient management methods. To achieve more accurate results in predicting the traffic situation, a ConvTrans-TCN (Convolutional Transformer with Temporal Convolutional Network) model is proposed in this paper. The model first constructs the feature extraction part using the causal-convolution multi-head self-attention module. It can effectively model the long-term dependency in the sequence and match the local patterns of the sequence, and it enhances the performance of feature extraction. Then the TCN (Temporal Convolutional Network) module is used to build the information fusion part to complete the fusion of feature data. The TCN architecture can accurately learn long-term and short-term dependencies in time series, and it has sufficient memory. Finally, the situation prediction is obtained by a feedforward neural network. The experiment's results prove that this model is feasible and it performs better than the common models such as LSTM, BP, which can help air traffic managers to identify the operational status of the terminal area and provide decision support.

INDEX TERMS Intelligent transportation system, air transportation, traffic situation prediction, transformer, temporal convolutional network.

I. INTRODUCTION

Due to the rapid growth of air transportation, traditional ATM (Air Traffic Management) methods can no longer meet the demand. ITS (Intelligent Transportation System) is one of the hottest research fields nowadays, which aims to achieve super-efficient navigation and safer travel journey [1]. ATM is also on its way to becoming intelligent, with the combination of advanced AI technologies enabling efficient and accurate management.

Traffic situation awareness of terminal area is an important part of intelligent air traffic management, Accurate comprehension of the traffic situation is the basis for optimizing aircraft operations in the terminal area and ensures safe aircraft operations in the terminal area.

Situation awareness has two main components, situation recognition and situation prediction [2]. In the field of air traffic, A number of researchers have attempted to study the traffic situation on routes and terminal areas using methods such as complex network theory [3], [4], [5], [6]. These studies focus on the analysis and comprehension of traffic situations by analyzing the evolutionary patterns of air traffic situations combined with the corresponding theories. Du et al. proposed a novel spatio-temporal hybrid deep learning model for airspace complexity prediction to effectively capture the spatial correlation and temporal dependence associated with airspace complexity data [7]. Sui et al. abstracted the airspace containing multiple sectors as an undirected graph, and

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

developed the spatiotemporal graph convolutional network to describe the correlation between changes in the operational situation of each sector and to predict the operational situation of these sectors [8]. These researchers were more concerned with the traffic situation in the airspace when studying situation prediction. As the busiest part of the air traffic network, few researchers have systematically and comprehensively investigated how to predict its operational situation.

The transformer-based deep learning model is currently a hot research topic, and many researchers use it to solve the task of time-series prediction [9], [10]. Since the traffic situation data in the terminal area is time series data, in this paper, we propose a transformer-based traffic situation prediction model for the terminal area, which is called ConvTrans-TCN model. The model consists of three parts, which are the information encoding module, the information synthesis module, and the situational value calculation module. The model takes multidimensional situation data as model's input, and after layers of calculation, the final situation prediction value is derived. Accurate predicting results can provide ATM with supporting decision information when making traffic management decisions and implementing traffic management policies.

The rest of the paper is organized as follows: Section II reviews the related works about this paper. Section III describes transformer model and the ConvTrans-TCN model. Section IV represents the experiment and discusses the feasibility and predictability of the proposed model in traffic situation predicting. Section V draws the conclusion.

II. RELATED WORKS

A. TRAFFIC SITUATION PREDICTION

The process of traffic situation prediction is based on time series data $X_t \in \mathbb{R}^{N\times D}$, which are generated from situation features, where N denotes the number of nodes, D denotes the number of situation features in time steps t [8]. For the given historical situation data, the future situation can be calculated by Equation (1).

$$[X_{t-T'+1}, \ldots, X_t] \xrightarrow{f} [X_{t+1}, \ldots, X_{t+T'}] \tag{1}$$

where f is the mapping function.

B. ATTENTION MECHANISM

Attention is a fundamental process used to describe the relationship between a set of variables and the goal of a query. Natural language processing (NLP), picture recognition, protein identification, recommendation systems, and other fields have all made extensive use of the attention mechanism [11], [12].

Self-attention is a variation of the attention mechanism that stresses many-to-many relationships. It can extract complex irregular patterns by digging deeper into the hidden associations in individual feature data [13]. Many studies have combined this with other types of network results to predict traffic flows. Fang et al. incorporated the attention mechanism to solve the problem of LSTM's inability to focus on the long-term dependence of traffic flow, and built a model that yielded accurate short-term prediction results [14]. As current prediction methods struggle to perform both long-term and short-term prediction tasks, to address this problem we have built Long Short-term Graph Convolutional Network predict traffic, a network that combined attention mechanism with graph neural network to capture complex spatial features [15]. Kong et al. proposed a graph talking-heads attention layer for capturing spatial dependencies [16]. As the terminal area traffic situation data is similar to the traffic flow data, multi-head self-attention can be used as a feature extraction module in the prediction model to capture the complex patterns in the situation data.

C. TRANSFORMER

Transformer has been widely used in NLP and Visual Representation, its core building block is the attention mechanism [17]. Due to transformer's parallelization-in-time mechanism, it can better model long sequence than RNNs.

For its advanced sequence modeling ability, transformer has been used to perform the traffic predicting task. And many researchers have improved on Transformer's architecture by adopting other network structures to achieve better performance, such as GMAN [18] and NAST [19]. These methods have combined transformer's encoder-decoder structure with neural networks to make the model more capable of capturing dependencies.

Although existing methods have achieved better performance, these methods have limitations on the input sequence and the models are not capable of modeling local features in the data. Therefore, this paper combines the temporal convolutional network [20] with transformer, and introduces a self-attention module based on causal convolution, to build a prediction model. This model has better performance when dealing with the problem of predicting traffic situation in the terminal area.

III. METHODOLOGY

A. TRANSFORMER

Google first introduced Transformer in 2017, a model that solved the problem of RNNs' failure to model long-range dependencies when dealing with long sequences. And the Transformer model was first used in the field of machine translation with good effects.

The basic structure of the Transformer model contains two parts, Encoder and Decoder, as shown in Fig. 1. Both Encoder and Decoder contain 6 blocks.

The Encoder part contains the Self-Attention layer and the Feed Forward Neural Network layer. This part is used to encode sequences and add them as input to the model. The decoder part is similar to the Encoder part, but the difference is that Decoder adds a Mask mechanism to the Attention mechanism of Encoder. The reason is that the input

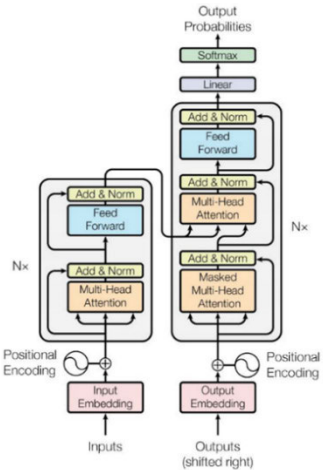

FIGURE 1. Framework diagram of transformer model [21].

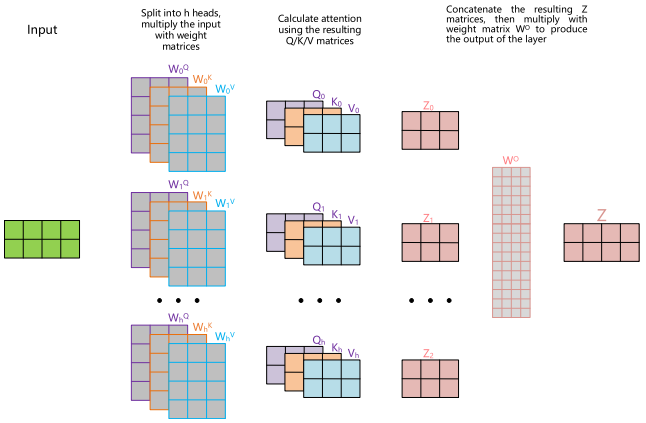

FIGURE 2. The calculation process of the multi-head self-attention mechanism.

of Decoder needs to be predicted, and the Mask hides the part that is not predicted.

The data processing module in the prediction model proposed in this paper is the Encoder part, so the self-attention layer and the feedforward neural network layer in Encoder are introduced here first.

1) MULTI-HEAD SELF-ATTENTION MECHANISM

Self-attention is one of the popular attention mechanisms. To obtain the queries matrix $Q \in R^{T\times Dq}$, and the keys matrix $K \in R^{T\times D_k}$ and values matrix $V \in R^{T\times D_v}$, the input matrix $X \in R^{T\times D_x}$ is linearly mapped. Then, the attention matrix $M_A \in R^{T\times T}$ is obtained from the attention-scoring function, which is normalized by the softmax funtion to obtain the attention weight matrix $W_A \in R^{T\times T}$. Finally, the matrix multiplication operation is performed on the attention weight matrix and the values matrix to obtain the output $H \in R^{T\times D_v}$. And the attention-scoring function uses the scaled dotproduct function. These functions are as follows:

$$Q = XW_q \tag{2}$$

$$K = XW_k \tag{3}$$

$$V = XW_v \tag{4}$$

$$M_A = \frac{QK^T}{\sqrt{D_K}} \tag{5}$$

$$H = \text{softmax}(M_A)V = W_A V \tag{6}$$

where, the W_q, W_k, and W_v are parameters matrix. The softmax$(\cdot)$ function normalizes the row vector. Equation (2) is the scaled dot-product function, where D_k represents the dimensionality of a query vector in the queries matrix K, its open square as the denominator is mainly to solve the problem that when the vector's dimensionality is large, the value of the numerator has too much variance and leads to the gradient is too small and the model is difficult to train.

The Multi-Head Self-Attention Mechanism is a linear transformation after combining the computational results of multi attentions, which enables the model to have the ability to use different feature information from different locations. The calculation method is to map the vectors of Q, K, and V into multiple $Q/K/V$ matrices with different linear projection matrices, and then perform the attention function calculation for each group of $Q/K/V$ matrices at the same time, and finally joint the results of each group of attention calculation and multiply them with the parameters matrix W^O to get the output value, whose process is shown in Fig. 2.

Supposing that there are h projection spaces, the multihead self-attention is calculated as

$$\text{Multihead}(H) = W^O[h_1; h_2; \ldots; h_h] \tag{7}$$

where each head component is calculated according to (6).

2) POSITION-BY-POSITION FEED-FORWARD NETWORKS

The feedforward neural network part contains two layers of full connect neural network with a ReLU function between the two layers, and the whole part is used to synthesize all the coding information, as in

$$F_{\text{FFN}}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{8}$$

where W_1 and W_2 are weight matrix. b_1 and b_2 are bias. x is the input.

3) IMPROVED TRANSFORMER MODEL

The terminal traffic situation data belongs to time sequence, and regarding this kind of data, it is necessary to accurately grasp the long-term and short-term patterns of the time sequence and dig out the laws of the complex patterns hidden in the sequences to make highly accurate predictions. And the situation data contains emergency data, which leads to some abnormal changes in the data. That is, there are outliers in the data set, and whether a time point is an outlier is judged according to its neighboring data points. Therefore, to extract the complex information contained in the data, a more appropriate model for the local features of the data is necessary.

The general multi-head self-attention module performs linear variation processing by considering only the features of each time point when making projections on the input data, which leads to little focus on the information surrounding the

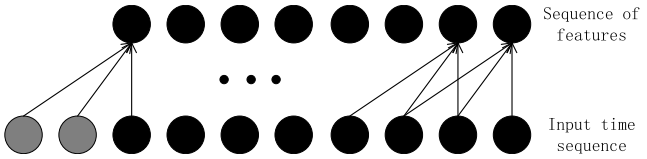

FIGURE 3. Causal convolution.

data. As a result, when calculating the attention's weights, there is a decrease in the accuracy of prediction due to incorrect attention caused by outliers' interference. To solve this problem, Zhou et al. [22] proposed a causal convolutional-based self-attention module, which implements the extraction of local features of sequence data using a convolution network, thus improving the multi-head self-attention part of the Transformer model and improving the matching of the module with the prediction of temporal data. Experimental results demonstrate that the module using causal convolutional yields more accurate prediction results. Therefore, this paper uses the causal convolutional self-attention module as the core component of the information encoding part of the terminal traffic dynamics prediction model. The following is a brief description of the module structure.

The multi-head self-attention module of the Transformer model uses a parameter matrix to linearly map the input sequence into a query matrix and a key matrix. While the causal-convolution self-attention module chooses a Conv1D layer that can extract local features of the sequence to map the input sequence, generating the query matrix Q and the key matrix K as in

$$Q = \text{Conv1D}(X) \tag{9}$$
$$K = \text{Conv1D}(X) \tag{10}$$

After obtaining Q and K, more accurate attention scores can be obtained to match the local position information in the sequence. The value vector matrix is obtained by linear mapping according to the original method. To facilitate the subsequent processing, a padding operation is added to the Conv1D layer in the module here to keep the shape of the input matrix consistent before and after encoding. Since the time sequences is directional, the causal padding method is chosen here to fill the Conv1D layer by doing forward complementary zeros to the sequence so that the future information data will not be leaked, and it also ensures that the length of the sequence does not change before and after the convolution. The diagram of causal convolutional is shown in Fig. 3.

In order that each vector in the self-attention output matrix also does not contain future information, the model turns the attention-scoring matrix into a lower triangular matrix using a mask matrix Mask (only the elements above the diagonal are not zero and have the value -1e9). The causal convolutional self-attention output is shown in (11), and its structure is shown in Fig. 4.

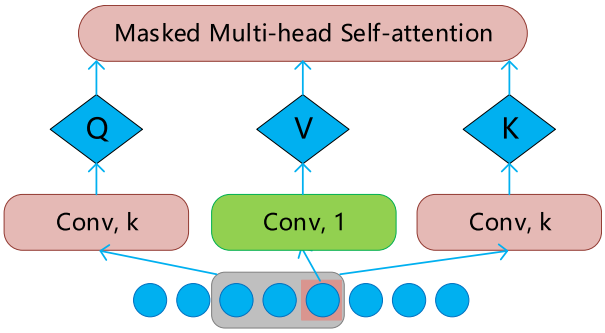

FIGURE 4. Causal convolutional multi-head self-attention.

$$H = \text{softmax}(QK^T \Big/ \sqrt{D_K} + \text{Mask})V \tag{11}$$

B. TCN-BASED INFORMATION FUSION MODULE

To implement the prediction of traffic situation in the terminal area, the causal convolutional self-attention module is used to encode the information data to obtain a sequence of encoded features with the same dimension as the input, and then further fuse the feature data to improve the accuracy of the prediction results. Since Transformer was designed to complete the work related to the NLP field, its architecture is not fully suitable for prediction, so only the causal convolutional multi-head self-attention module of it is used here as the feature extraction module. And the TCN (Temporal Convolutional Network) architecture is chosen for the information fusion processing module to complete the final prediction [23].

The TCN architecture is mainly composed of stacked 1D convolutional layers, and causal convolution is used as the basic convolutional layer to maintain temporal directionality and avoid information-leaking problems. However, to obtain a larger receptive field, a deeper network is required. To solve this problem, the TCN architecture uses the kernel of Dilated Causal Convolution [24] as the computational unit of each layer and introduces Residual Connection between layers to achieve better performance.

Overall, the TCN architecture is able to accurately learn the long-term and short-term dependencies in time sequences, and for traffic situation prediction, it has sufficient memory, and its architecture is simple. The following is a brief description of the basic computation of the Dilated Causal Convolution kernel and the Residual Connection of the TCN.

1) DILATED CAUSAL CONVOLUTION

Dilated convolution allows the model to expand the receptive field at an exponential level. Specifically, for the model's input $X \in R^n$, let $f \in R^k$ denote the one-dimensional causal dilated convolution kernel, then it is calculated as in

$$F(X) = \sum_{i=0}^{k-1} f(i) \cdot X_{s-d \cdot i} \tag{12}$$

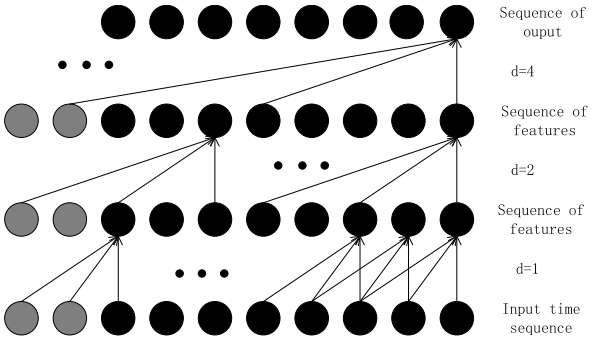

FIGURE 5. Dilated causal convolution.

where, d is the dilation rate, k is the size of the convolution kernel, and $s - d \cdot i$ denotes the corresponding position in the input sequence.

Generally, d changes with the number of network layers i as in

$$d = O(2^i) \tag{13}$$

Such an exponential change increases the receptive field of the TCN more rapidly than adjusting the size of convolution kernel. Thus, ensuring that the receptive field of the higher-level convolution kernels in the network can cover all valid inputs of the input sequences, leading to better fusion of information and effective modeling of long-term patterns in the sequences, as shown in Fig. 5.

2) RESIDUAL CONNECTION

The Residual Connection structure consists of two branches, one of which is the original network branch, which is assumed to transform the input as a function F, and the other branch is the residual connection branch, which is responsible for directly transmitting the network input sequence to the network output and adding it to the network output to obtain the final output result, as in

$$o = \text{Activation}(x + F(x)) \tag{14}$$

This structure can allow multi-layer networks to efficiently learn equal mappings to inputs, instead of complex transformations of multi-layer networks. Thus, allowing the deeper network to adaptively adjust its depth according to the data's distribution, ensuring stable performance of the deeper network and enhancing the learning ability. It also ensures that deeper networks can be trained properly without being plagued by problems such as gradient disappearance. In addition, to ensure that the input and output tensor shapes are consistent, a 1D convolution with a convolution kernel size of 1 is added to the equivalent input branch of the residual connection to make appropriate adjustments to the input.

C. ConvTrans-TCN-BASED SITUATION PREDICTION MODEL

Prediction models based on neural networks or machine learning can effectively use multi-dimension traffic situation data to jointly complete the prediction, and the performance

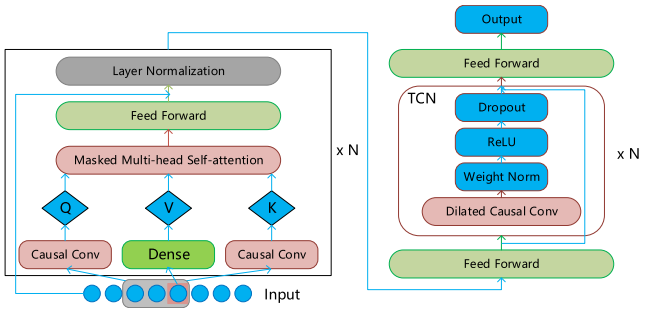

FIGURE 6. ConvTrans-TCN model.

is significantly improved compared with common methods. However, most of the prediction models based on machine learning or neural networks are superficial models, such as single-layer LSTM, GRU and other recurrent neural network models, and there is still some room for improvement in the accuracy.

To improve the accuracy, a situation prediction model named ConvTrans-TCN is proposed. The experiments prove that the model effectively improves the accuracy and achieves better results. In this section, the structure of the ConvTrans-TCN model is described in detail. And the overall architecture is shown in Fig. 6.

The ConvTrans-TCN model consists of three parts, namely the feature extraction and information encoding part, the information fusion part, and the calculation part of situation-predicting value.

The feature extraction and information coding part consists of a causal convolutional self-attention module. The input to the prediction model is the historical traffic situation data $X = [x_1; x_2; \cdots; x_T]$ from the current time point to the previous T time steps. The model encodes the input information data with the causal convolutional self-attention module, which is calculated as follows:

$$\text{MH} = \text{ConvSA}(X) \tag{15}$$

where the shape of $MH \in R^{T \times D}$ is kept consistent with the input data X. ConvSA() is the computational process in the attention module. After this, the model has two layers of FNN for further processing of the feature information, and the ReLU function is chosen for the activation function. As in

$$\text{FM} = \text{ReLU}(\text{MH} \cdot W_1 + b_1) \cdot W_2 + b_2 \tag{16}$$

where $\text{FM} \in R^{T \times D}$.

Finally, the model utilizes a residual connection structure to ensure that no performance degradation occurs throughout the network, and at the same time, summing it with the input data and performing layer normalization to enhance backpropagation. As in

$$\text{FX} = \text{LayerNorm}(X + \text{FM}) \tag{17}$$

where $\text{FX} \in R^{T \times D}$. In this paper, based on the experimental simulation results, the above-mentioned coding network structure with two stacked layers is finally adopted for the initial processing of the input information.

The second part of the model, the information fusion part, uses the TCN architecture as the main component to fuse the information of the feature matrix in the time dimension in preparation for the next step of situation-predicting values.

First, the feature information is further extracted by FCNN (Fully Connected Neural Network), as in

$$\mathrm{TX} = \mathrm{ReLU}(\mathrm{FX} \cdot W + b) \tag{18}$$

where $\mathrm{TX} \in R^{T \times D}$.

Second, the TCN architecture does further processing of the information, and its fundamental component units are described in the previous section. First is a causal dilated convolution layer, which extracts and fuses the information. This is followed by a weight normalization operation, which is beneficial for the training of the model and has a low overhead. Then a ReLU function is used to obtain the network activation values. Finally, the dropout method is used to improve the generalization ability of the network and avoid overfitting. The calculation process is

$$\mathrm{TY}^1 = \mathrm{Dropout}(\mathrm{ReLU}(\mathrm{WN}(\mathrm{DConv}(\mathrm{TX})))) \tag{19}$$

where TY^1 stands for the output matrix with a single layer. The proposed model uses a 12-layer convolutional structure with residual connection structures introduced every two levels and repeated six times. Each layer of convolution employs a dilated convolutional kernel, which can completely guarantee that the output information of the network has a sufficient receptive field for the useful part of the input information and the various-length time sequences pattern features in the input information are fully extracted and fused. The final output of the situation prediction component of the ConvTrans-TCN model is the last vector $\mathrm{ty}_T \in R^c$ (c is the number of convolutional kernels) in the output information sequence $\mathrm{TY} = [\mathrm{ty}_1; \mathrm{ty}_2; \cdots; \mathrm{ty}_T]$, which contains a prediction of future situation.

The structure of the third part is a two-layer FCNN, whose role is to make computational judgments on the previous output data, output the situation prediction values, and obtain the prediction results, as in

$$y = \mathrm{ReLU}(\mathrm{ty}_T \cdot W_1 + b_1) \cdot W_2 + b_2 \tag{20}$$

The overall architecture of the model uses components including the causal convolutional multi-head self-attention module, causal dilated convolution and FCNN. And it uses such techniques as residual connectivity, normalization, and regularization to optimize the model, which makes the model as a whole highly parallel and has a faster convergence speed.

Simultaneously, the model's backpropagation path is not in the same direction as the time dimension of the input sequence, which effectively avoids the problem of too small or too large gradients caused by too long sequences when dealing with multi-time-step situation sequences, and allows the model to retain longer information and achieve better generalization performance and higher prediction accuracy.

The model's loss function employs the mean square error (MSE) function, the formula for which is presented in (21), the smaller its value, the greater the model's accuracy. And the backpropagation technique completes the whole training of the model. The model is empirically validated to have superior prediction accuracy when compared to other models. The next section will go through specific experimental results, data preparation and pre-processing model parameter design, and so on.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{observed}_i - \mathrm{predicted}_i)^2 \tag{21}$$

IV. EXPERIMENTS

This section first verifies the feasibility of the situation prediction model, then designs an experiment to confirm the effect of the information encoding part and the causality of the situation data. A comparison experiment is also designed to compare the performance of this paper's prediction model with several frequently-used prediction models to reflect the advancement of this paper's model.

A. PREPARATION TRANSFORMER

1) DATASET

The traffic situation data of the terminal area used in this paper is the data of ZBTJ (Tianjin Binhai International Airport) from June 3-16, 2019. ZBTJ is one of the busiest airports in China, it is a typical two-runway airport and the airport reaches its maximum handling capacity from June onwards. And each sample corresponds to a 10-min air traffic situation, all these samples are grouped in chronological sequence. Each sample has 13 dimensions, of which the first 12 characteristics are situation elements, and the last one is the situation level (smooth/normal/congested/standstill) provided by ATM experts, these experts include tower controllers with over ten years of work experience and researchers in civil aviation.

Then, the traffic situation data in the dataset are normalized and 70% of the data are used as training samples, and 10% and 20% of the data are used as validation and testing samples, respectively. While the training samples are interpolated and extended.

Finally, a sliding window (of size 10 timesteps) is introduced to process the entire dataset to obtain the multi-step time sequences data, which is used as the input to the prediction model, and the situation value of next time point is used as the output to obtain the prediction's target value.

2) MODEL'S PARAMETERS

The specific parameters of the prediction model proposed in this paper are as follows:

The causal convolutional layer of the data encoding part: the size of the convolutional kernel is 3 and the number of kernels is 28; the number of heads of the multi-head self-attention is 4.

TCN part: the number of double-layer 1D causal dilated convolutional layers is 6, the size of each convolutional layer is 2, the number of convolutional kernels is 64, and the dilated factors are set to 1, 2, 4, 8, 16, 32, respectively.

The parameters of the model are optimized by Adam algorithm [25] and finally set as follows: $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$. And the model was trained for a total of 500 rounds with a training batch of 32.

3) EVALUATION INDICATORS

The output of the terminal area traffic situation prediction model proposed in this paper is a quantified traffic situation value, which is a regression problem. Therefore, to make an effective comparison of accuracy with other models, two index data are introduced in this paper to evaluate the prediction effect of the model. The first one is the Root Mean Square Error (RMSE), as shown in (22). The other one is the Mean Absolute Error (MAE), as shown in (23).

$$\mathrm{RMSE}(X, f) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f(x_i) - y_i)^2} \tag{22}$$

$$\mathrm{MAE}(X, f) = \frac{1}{N}\sum_{i=1}^{N}|f(x_i) - y_i| \tag{23}$$

where, X is the model's input. f denotes the operation process. $f(x_i)$ is the output of the model. y_i is the actual situation value. N is the size of the data's volume.

RMSE and MAE are independent of the sample size and can be used to evaluate the model's accuracy. Therefore, the smaller the value of RMSE, the less likely the model has a prediction error. The smaller the value of MAE, the smaller the average difference between the prediction and the actual value, and the higher the accuracy of the model's output.

B. PREDICTABILITY ANALYSIS OF TRAFFIC SITUATION IN THE TERMINAL AREA

The shorter the sequence, the less information it contains, thus the information's uncertainty is difficult to eliminate, which means the predictability of short sequences does not fully represent the predictability of traffic situation. However, the longer the sequence, the longer the computation time. In summary, a suitable length of sequence needs to be determined to ensure that the prediction algorithm contains the required information and can complete the prediction in a short time. Based on Liu et al. [26], the entropy estimation algorithm is used here to calculate the upper bound and lower bound of predictability of the traffic situation elements.

First, the entropy for various sequence lengths is determined, and the results are displayed in Fig. 7. As seen in the image, the information entropy and the actual entropy tend to be steady as the duration of the data sequence rises. The actual entropy rises as the length of the sequence shrinks. The actual entropy is larger than the information entropy when the sequence's length is less than a specific threshold (205) since the method for calculating the actual entropy cannot determine the proper value when the sequence is too short. The information entropy is always stable around 1. This is because the number of values of the two state values in the data series is approximately equal, so it can be judged that the

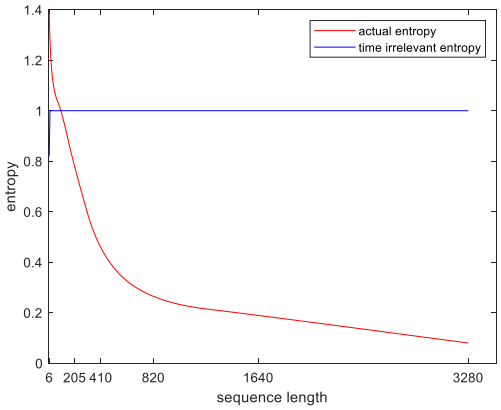

FIGURE 7. The relationship between sequence length and sequence entropy.

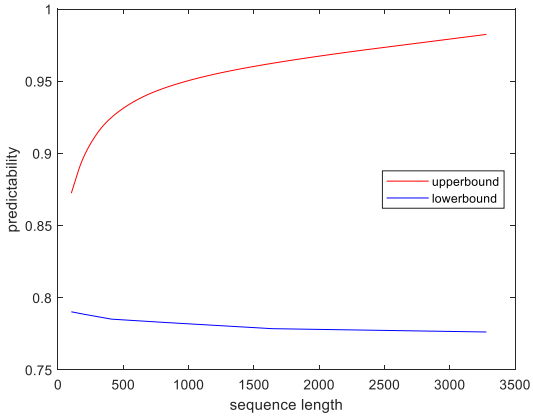

FIGURE 8. The relationship between sequence length and predictability.

lower bound of traffic situation predictability is stable. On the other hand, the value of the actual entropy is also essentially stable at a smaller value after the length of the sequence exceeds 2000. The figure shows that when the sequence length is 3280, the value of the actual entropy is 0.08, and the information containing regularity in the sequence has stabilized.

For different values of sequence entropy, the relationship between the upper and lower bound of predictability of traffic situation in the terminal area and the length of the sequence can be further analyzed, as shown in Fig. 8.

According to the above image, when the sequence length exceeds 2000, the upper bound of predictability tends to be stable, and when the sequence length is 3280, the upper bound is 0.9825, which proves that the dataset has enough information for situation prediction. And the lower bound has been relatively stable, and when the sequence length is 3280, the lower bound is 0.7763. Generally, the upper bound and lower bound of predictability change with the sequence length in the same pattern as the entropy value.

In summary, the upper bound of traffic predictability in the terminal area is 0.9825, while the lower bound is 0.7763, which sufficiently proves the feasibility of situation prediction and provides a theoretical reference for the accuracy comparison among models.

TABLE 1. Comparison of information encoding modules for different models.

Model	RMSE	MAE
TCN	0.0731	0.0415
Trans-TCN	0.0562	0.0386
ConvTrans-TCN	0.0405	0.0317

C. EFFECTIVENESS ANALYSIS OF THE PROPOSED ARCHITECTURE

This section conducts experiments related to the model architecture, starting with the TCN model, the Trans-TCN model, and the ConvTrans-TCN model to verify the validity of the proposed model's information encoding structure. This is followed by a causality experiment on the dynamical dataset, which focuses on comparing the prediction effects when different padding methods are used for the convolutional layers in these models, to select the applicable model architecture.

1) COMPARISON OF INFORMATION ENCODING MODULES

Based on the previously proposed evaluation indicators, here using experimental tests, the results obtained by those three models are shown in Table 1.

As shown in the above table, the RMSE value of the ConvTrans-TCN model is 0.0405 and the MAE value is 0.0317, both of which are smaller than the corresponding indicators of the other two models. Specifically, the indicators of the Trans-TCN model are better than those of the TCN model, which shows that using a multi-head self-attention module to pre-process the situation dataset with an information encoding before the TCN module can improve its learning ability and the prediction accuracy. Moreover, the RMSE value and the MAE value of the proposed ConvTrans-TCN model are further reduced compared with the other two models, which indicates that the ConvTrans-TCN model can indeed better model the long-term and short-term patterns in the time sequences through the causal convolutional self-attention module for the situation prediction data in this paper, accurately match the various local period information present in the sequences. And finally, through the TCN process, the proposed model can result in more accurate and stable prediction results.

2) CAUSALITY ANALYSIS OF THE DATASET

This section analyzes whether the used dataset has causal characteristics, and determines the padding method of the convolution part of the proposed model. The experiment mainly compares the accuracy indicators of the convolution layer in the ConvTrans-TCN model using the Same Padding and the Causal Padding, and the results are shown in Table 2.

From the table, the MAE value and RMSE value of the Causal Padding are smaller than the values of the Same Padding. This result indicates that the dataset in this paper has a certain degree of causality, so the use of causal convolution can avoid information interference, better model time sequences, and capture the dependencies in the data as well as the evolution pattern of traffic situation in terminal area, which leads to more accurate prediction results.

TABLE 2. Comparison of different padding methods.

Padding	RMSE	MAE
Same	0.0528	0.0461
Causal	0.0405	0.0317

TABLE 3. Effects of different parameters on the proposed model's performance.

Heads	Encoder layer	Decoder layer	MAE
8	8	8	0.0601
8	4	4	0.0449
4	8	8	0.0525
4	4	4	0.0317

TABLE 4. Comparison of the performance for different prediction models.

Prediction model	RMSE	MAE
ConvTrans-TCN	0.0405	0.0317
LSTM	0.0522	0.0497
BP	0.0674	0.0588
GA-GMNN [27]	0.0483	0.0366

Summarizing the results of the above experiments, this paper uses the ConvTrans-TCN as the structure of the traffic situation prediction model. Because it can better identify and represent the long-term and short-term patterns, local information, and data's causal features in the situation sequences, its results are also more accurate and stable.

3) EFFECTS OF DIFFERENT PARAMETERS TO THE PROPOSED MODEL

Table 3 shows the effects of different parameters on the performance of the proposed model.

As shown in Table 3, the model with 4 heads and 4 identical encoder-decoder layers has the best performance. The increased number of layers within the encoder-decoder structure worsens the performance of the model. Conversely, the increased number of heads in the attention mechanism does not provide evidence of improved performance.

D. COMPARISON OF THE PERFORMANCE FOR DIFFERENT PREDICTION MODELS

The comparative experiment selects from time sequences prediction models commonly used and models that have been used in the field of terminal's traffic situation prediction to verify the effectiveness and superiority of the ConvTrans-TCN model proposed in this paper.

The prediction accuracy values of each prediction model are shown in Table 4.

According to the data in the above table, each model is capable of making predictions. The RMSE and the MAE values of the ConvTrans-TCN model are smaller than the values of other three models. Compared with the GA-GMNN model, which has the smallest values of the three models, the

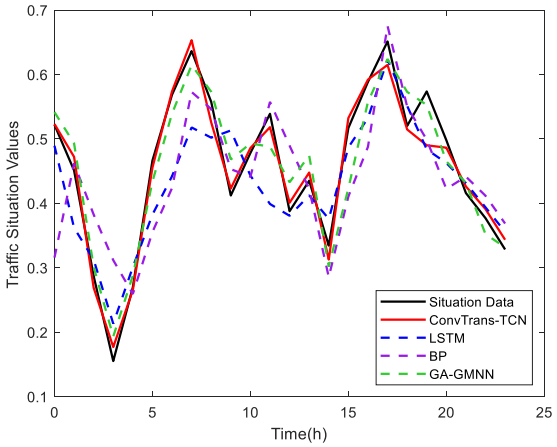

FIGURE 9. Comparison of the predicted situation values for different models.

ConvTrans-TCN model's RMSE value is reduced by 19.26%, and the MAE value is reduced by 15.25%. This shows that the ConvTrans-TCN model combined with the causal convolutional self-attention module and TCN can model the situation sequences more effectively, and make an accurate and stable prediction of future situation values, its prediction accuracy is superior to other prediction models.

To compare the performance of the models more visually, Fig. 9 shows a line graph comparing the predicted values of each model for each period of the coming day.

In Fig. 9, the solid black line is the actual situation data, the solid red line is the prediction data of the ConvTrans-TCN model, and the dashed lines in other colors indicate the prediction values of the comparison models. The errors between the predicted values of the proposed model and actual values are smaller than the errors of the other models in the whole test dataset. Moreover, the prediction results of the proposed model are closer to the actual situation than other models at some situation points, which fully demonstrates that the prediction performance of our model is very stable and its accuracy is improved compared with other models.

Additionally, the blue and green dashed lines in the figure show that the predicted values of LSTM and GA-GMNN can be generally fitted to the actual values, which indicates that neural networks can indeed model the time sequences data better. However, the fitted values at some specific points still have a gap compared with the ConvTrans-TCN model. This is because these neural networks are limited by the length of the sequence data when dealing with it, and cannot accurately represent the long-term dependencies in the sequence, which in turn leads to the limited accuracy of these models. Meanwhile, the fitting effect of the GMNN model based on the attention mechanism is better than that of the LSTM model, which illustrates the effectiveness of the attention mechanism in dealing with sequence problems. This paper also adopts a causal convolutional self-attention module in the information encoding part, which can accurately encode and fuse the information. And the module can also enhance the important local information features, which is beneficial for the model to learn the hidden patterns in the data.

Finally, the prediction accuracy of the BP network is significantly lower than several other models from the figure. Because the BP network simply takes a FCNN layer, which is hard to model the spatial and temporal relationships among terminal traffic situation data. What is worse, the BP model can overfit easily, and its representation is also limited, so it cannot achieve high accuracy even after adjusting the parameters.

Overall, the ConvTrans-TCN model proposed in this paper processes and analyzes the sequences by causal convolutional multi-head self-attention mechanism and multiple Conv1D layer. The model achieves high accuracy in the terminal traffic situation prediction, and it steadily predicts the situation value in line with the actual situation at all periods, which is advantageous in comparison with other models.

E. DISCUSSION

The terminal area is required to modify the future operation plan when the prediction results indicate that the situation level will reach the threshold for that terminal area at a certain period in the future. The commonly available adjustment method is flow adjustment, in which the terminal area seeks to change the departure time of aircraft by lengthening the interval since arriving aircraft has priority. The adjusted operating schedule can reduce delays and improve the safety of operations in the terminal area. Although for some passengers there are some changes to their flight departure times, by being able to predict and make adjustments in advance, passengers are also able to be informed in advance and adjust their travel plans.

Generally speaking, accurately predicting the terminal area traffic situation helps managers to grasp the evaluation trend of terminal operation and to learn more about the terminal's operational law. In addition, the results of terminal area situation prediction can be used to evaluate ATM decisions. When new management methods or optimal scheduling strategies are prepared, the prediction methods are used to obtain the results of the situation after the implementation of these method strategies, and if the results meet the expectations, they can be used for actual operation, and if not, further adjustments are needed.

V. CONCLUSION

Terminal area traffic situation prediction is the key to intelligent ATM. Since traffic situation is dynamic and sometimes unpredictable, traditional statistical models cannot provide effective performance. On the other hand, machine learning methods, especially deep neural network-based models, can handle these problems in a manageable way. Based on multi-head self-attention mechanism and temporal convolutional network, a terminal area traffic situation prediction transformer model is proposed. And this paper provides a comprehensive performance comparison with LSTM, BP, and GA-GMNN. The mean absolute error and root mean squared

error are used as evaluation metrics. The results confirm the superior performance of the proposed transformer model with an obvious improvement in MAE values over the comparison benchmark. In addition, the model allows for highly parallelized computation, which improves the efficiency of situational prediction.

In fact, external factors such as weather can also influence the traffic situation in the terminal area. In the future, we will take these external factors into account to further improve the accuracy of our prediction. Also, the prediction of traffic situation in multiple terminal areas is a priority due to the circulation relationship between the terminal areas.

REFERENCES

[1] K. Golestan, R. Soua, F. Karray, and M. S. Kamel, "Situation awareness within the context of connected cars: A comprehensive review and recent trends," *Inf. Fusion*, vol. 29, pp. 68–83, May 2016.

[2] M. R. Endsley, "Design and evaluation for situation awareness enhancement," in *Proc. Hum. Factors Soc. Annu. Meeting*, vol. 32. Los Angeles, CA, USA: Sage Publications, 1988, p. 2.

[3] M. Prandini, L. Piroddi, S. Puechmorel, and S. L. Brazdilova, "Toward air traffic complexity assessment in new generation air traffic management systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 809–818, Sep. 2011.

[4] W. Hongyong, S. Ziqi, and W. Ruiying, "Study on evolution characteristics of air traffic situation complexity based on complex network theory," *Aerosp. Sci. Technol.*, vol. 58, pp. 518–528, Nov. 2016.

[5] H. Chen, "Air traffic complexity assessment based on ordered deep metric," *Aerosp. Sci. Technol.*, vol. 9, no. 12, pp. 518–528, Nov. 2022.

[6] Y. Kawagoe, R. Chino, S. Tsuzuki, E. Itoh, and T. Okabe, "Analyzing stochastic features in airport surface traffic flow using cellular automaton: Tokyo international airport," *IEEE Access*, vol. 10, pp. 95344–95355, 2022.

[7] W. Du, B. Li, J. Chen, Y. Lv, and Y. Li, "A spatiotemporal hybrid model for airspace complexity prediction," *IEEE Intell. Transp. Syst. Mag.*, early access, Sep. 28, 2022, doi: 10.1109/MITS.2022.3204099.

[8] D. Sui, K. Liu, and Q. Li, "Dynamic prediction of air traffic situation in large-scale airspace," *Aerospace*, vol. 9, no. 10, p. 568, Sep. 2022.

[9] H. J. Park, T. Kim, Y. S. Kim, J. Min, K. W. Sung, and S. W. Han, "CRFormer: Complementary reliability perspective transformer for automotive components reliability prediction based on claim data," *IEEE Access*, vol. 10, pp. 88457–88468, 2022.

[10] C. Wang, Y. Chen, S. Zhang, and Q. Zhang, "Stock market index prediction using deep transformer model," *Expert Syst. Appl.*, vol. 208, Dec. 2022, Art. no. 118128.

[11] L. Wu, Y. Wang, X. Li, and J. Gao, "Deep attention-based spatially recursive networks for fine-grained visual recognition," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1791–1802, May 2019.

[12] L. Wu, Y. Wang, J. Gao, M. Wang, Z.-J. Zha, and D. Tao, "Deep coattention-based comparator for relative representation learning in person re-identification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 722–735, Feb. 2021.

[13] S.-J. Bu and S.-B. Cho, "Time series forecasting with multi-headed attention-based deep learning for residential energy consumption," *Energies*, vol. 13, no. 18, p. 4722, Sep. 2020.

[14] W. Fang, W. Zhuo, J. Yan, Y. Song, D. Jiang, and T. Zhou, "Attention meets long short-term memory: A deep learning network for traffic flow forecasting," *Phys. A, Stat. Mech. Appl.*, vol. 587, Feb. 2022, Art. no. 126485.

[15] R. Huang, C. Huang, Y. Liu, G. Dai, and W. Kong, "LSGCN: Long short-term traffic prediction with graph convolutional networks," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, vol. 7, Jul. 2020, pp. 2355–2361.

[16] X. Kong, J. Zhang, X. Wei, W. Xing, and W. Lu, "Adaptive spatial–temporal graph attention networks for traffic flow forecasting," *Appl. Intell.*, vol. 52, no. 4, pp. 4300–4316, Mar. 2022.

[17] C. Chen, Y. Liu, L. Chen, and C. Zhang, "Bidirectional spatial–temporal adaptive transformer for urban traffic flow forecasting," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 30, 2022, doi: 10.1109/TNNLS.2022.3183903.

[18] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, Palo Alto, CA, USA, 2020, vol. 34, no. 1, pp. 1234–1241.

[19] K. Chen, G. Chen, D. Xu, L. Zhang, Y. Huang, and A. Knoll, "NAST: Non-autoregressive spatial–temporal transformer for time series forecasting," 2021, *arXiv:2102.05624*.

[20] T. Qi, G. Li, L. Chen, and Y. Xue, "ADGCN: An asynchronous dilation graph convolutional network for traffic flow prediction," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 4001–4014, Mar. 2022.

[21] A. Vaswani, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–15.

[22] Y. Zhou, K. Xu, and F. He, "Root cause diagnosis in multivariate time series based on modified temporal convolution and multi-head self-attention," *J. Process Control*, vol. 117, pp. 14–25, Sep. 2022.

[23] P. Hewage, M. Trovati, E. Pereira, and A. Behera, "Deep learning-based effective fine-grained weather forecasting model," *Pattern Anal. Appl.*, vol. 24, no. 1, pp. 343–366, 2021.

[24] A. Sasou, "Deep residual learning with dilated causal convolution extreme learning machine," *IEEE Access*, vol. 9, pp. 165708–165718, 2021.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[26] B. Liu and Y. Li, "Research on the predictability of network security situation," in *Proc. IEEE 20th Int. Conf. Commun. Technol. (ICCT)*, Nanning, China, Oct. 2020, pp. 1127–1133.

[27] C. Xu, "Research on traffic identification and prediction method in multi-terminal area," M.S. thesis, Dept. Comput. Technol., Sichuan Univ., Sichuan, China, 2021.

ZHOU YU was born in Chongqing, China, in 1998. He received the B.Eng. degree in communication engineering from China Agricultural University, Beijing, in 2020. He is currently pursuing the M.E. degree in traffic and transportation with the Civil Aviation University of China, Tianjin. His research interests include ITS and traffic security.

XINGYU SHI was born in Xingtang Country, Shijiazhuang, Hebei, China, in 1998. He received the B.S. degree in civil aviation from Shenyang Aerospace University, in 2020. His research interests include airspace management and UAV path planning.

ZHAONING ZHANG was born in Luannan County, Tangshan, Hebei, China, in 1964. He received the B.S. degree in mathematics from Hebei Normal University, in 1984, and the M.Sc. degree in mathematics and the Ph.D. degree in electrical engineering from Tianjin University, in 1989 and 1999, respectively. Since 2001, he has been a Professor with the College of Air Traffic Management, Civil Aviation University of China. He is the author of four books and more than 200 articles. His research interests include air traffic control, air traffic flow management, and airspace management.