

Data-model Hybrid-driven Approach for Unit Commitment Considering Conditional Transmission Section Limits

Hong Yu, Yong Zhao, Manli Yan, Yuanzheng Li, and Yaowen Yu

Abstract—Conditional transmission section limits (C-TSLs) depend on the actual operational state of the power grid, such as commitment status and reserve capacity, exacerbating the computational complexity of the unit commitment problem. To overcome this complexity, this paper proposes a data-model hybrid-driven approach for solving unit commitment problems. Furthermore, a unit commitment prediction algorithm is proposed to reduce the problem scale by fixing a subset of unit commitment variables, thereby eliminating inactive C-TSL intervals. Specifically, the proposed algorithm incorporates a convolutional attention module to perform deep feature extraction and enhancement on time series data, complemented by a multi-head cross-attention mechanism designed to synthesize temporal and non-temporal features. The introduction of attention mechanisms enhances the prediction performance of the proposed algorithm. Numerical results demonstrate that the proposed approach significantly reduces the number of variables and constraints in the unit commitment model, thereby expediting the solution process while maintaining high solution accuracy.

Index Terms—Data-model hybrid-driven, prediction, deep learning, feature extraction, conditional transmission section limit (C-TSL), attention mechanism unit commitment.

NOMENCLATURE

A. Sets

Ω_s^+, Ω_s^-	Sets of buses with positive and negative power transfer distribution factors for transmission section s
$\mathcal{B}$	Set of transmission lines
$\mathcal{G}$	Set of conventional thermal units

$\mathcal{G}_t$	Set of thermal units in entire system that are predicted to be online at time t
$\mathcal{G}_t^a$	Set of thermal units in group a that are predicted to be online at time t
$\tilde{\mathcal{G}}_t$	The maximum set of thermal units in entire system that can be online at time t
$\tilde{\mathcal{G}}_t^a$	The maximum set of thermal units in group a that can be online at time t
$\mathcal{L}$	Set of linear segments of cost function
$\mathcal{M}$	Set of renewable energy units
$\mathcal{N}$	Set of buses
$\mathcal{S}$	Set of transmission sections
$\mathcal{T}$	Set of time periods
$\mathcal{Z}^{\text{train}}$	Set of training samples
$\mathcal{Z}^{\text{val}}$	Set of validation samples
$\mathcal{Z}^{\text{test}}$	Set of test samples

B. Indices

a	Unit group index
b	Transmission line index
g	Conventional thermal unit index
k	Threshold iteration index
l	Segment index of piecewise linear cost
m	Renewable energy unit index
n	Bus index
p	Conditional variable index
s	Transmission section index
t	Time index
z	Sample index of a training batch

C. Parameters

α, β	Confidence thresholds
η	Threshold adjustment coefficient
$\lambda_{n,b}^L, \chi_{n,s}^L$	Sensitivity factors from bus n to line b and section s
$\lambda_{g,b}^P, \chi_{g,s}^P$	Sensitivity factors from conventional thermal unit g to line b and section s
$\lambda_{m,b}^S, \chi_{m,s}^S$	Sensitivity factors from renewable energy unit m to line b and section s

Manuscript received: July 18, 2025; revised: November 13, 2025; accepted: January 23, 2026. Date of CrossCheck: January 23, 2026. Date of online publication: April 2, 2026.

This work was supported by the State Grid Corporation Science and Technology Project (No. 5108-202455034A-1-1-ZN).

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

H. Yu, Y. Zhao, M. Yan, Y. Li, and Y. Yu (corresponding author) are with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China, and Y. Zhao is also with the Key Laboratory of Image Processing and Intelligent Control (Huazhong University of Science and Technology), Ministry of Education, Wuhan, China (e-mail: m202373572@hust.edu.cn; zhiwei98530@hust.edu.cn; d202481469@hust.edu.cn; Yuanzheng_Li@hust.edu.cn; yaowen_yu@hust.edu.cn).

DOI: 10.35833/MPCE.2025.000658



$\mathbf{b}^F, \mathbf{b}^I, \mathbf{b}^C$	Bias vectors associated with forget gate, input gate, candidate cell state, and output gate	$\mathbf{W}^F, \mathbf{W}^I$	Weight matrices associated with forget gate, input gate, candidate cell state, and output gate
$\mathbf{c}_t$	Cell state vector at time t	$\mathbf{W}^Q, \mathbf{W}^K$	Query, key, value projection matrices and linear transformation output weight matrix
$\tilde{\mathbf{c}}_t$	Candidate cell state vector at time t	$\mathbf{W}^V, \mathbf{W}^T$	
$C_g^{\min}$	The minimum output cost of thermal unit g	$\mathbf{x}_t$	Input vector at time t fed into long short-term memory (LSTM) network
C_g^U, C_g^D	Start-up and shut-down costs for thermal unit g	$y_{z,g,t}$	Actual commitment status of thermal unit g at time t for sample z
$C_{g,l}$	Slope of segment l of piecewise linear cost function for thermal unit g	$y'_{z,g,t}$	Predicted commitment status of thermal unit g at time t for sample z
d^K	Dimension of key matrix $\mathbf{K}$	<i>D. Variables and Functions</i>	
$\mathbf{f}_t$	Forget gate vector at time t	$\sigma(\cdot)$	Sigmoid activation function
$F_b^{\max}$	The maximum capacity limit of line b	$C_{a,t}^{\min}, C_{a,t}^{\max}$	Lower and upper bounds of committed generation capacities in group a at time t
$\mathbf{h}_t$	Hidden state vector at time t	$P_{g,t,l}$	Output of segment l of thermal unit g at time t
$\mathbf{i}_t$	Input gate vector at time t	$P_{m,t}^s$	Scheduled output of renewable energy unit m at time t
I_g^U, I_g^D	Initial consecutive up and down time of thermal unit g	r_t^+, r_t^-	Upward and downward reserve capacities at time t
L_g^U, L_g^D	Numbers of time periods for which thermal unit g must be online and offline	$R_t^{\min}, R_t^{\max}$	Lower and upper bounds of upward reserve capacities at time t
M_i^p, M_{i+1}^p	Lower and upper limits of conditional variable x_p in interval i	$S_{a,t}^{\min}, S_{a,t}^{\max}$	Lower and upper bounds of numbers of committed units in group a at time t
N	Number of intervals for conditional variables	$u_{g,t}$	Commitment status of thermal unit g at time t
N_g	Number of conventional thermal units	$u_{z,g,t}$	Commitment status of thermal unit g at time t for sample z in test set
N_t	Number of time periods	$u_{z,g,t}^{\text{pre}}$	Predicted probability that thermal unit g is online at time t for sample z in test set
N_z	Number of samples in a training set	x_p	Conditional variable affecting C-TSLs
$\mathbf{o}_t$	Output gate vector at time t	y_i^p	Binary auxiliary indicator variable of conditional variable x_p in interval i
P	Number of conditional variables	$y_{g,t}, z_{g,t}$	Start-up and shut-down variables of thermal unit g at time t
$P_{g,l}^{\max}$	The maximum output of segment l of thermal unit g	<i>I. INTRODUCTION</i>	
$P_g^{\min}$	The minimum output of thermal unit g		
$P_{m,t}^F$	Forecasted maximum generation of renewable energy unit m at time t		
$P_{n,t}^L$	Demand level at bus n at time t		
P_t^L	System demand level at time t		
$\mathbf{Q}$	Query matrix		
R_g^U, R_g^D	Ramp-up and ramp-down limits of thermal unit g		
R_t^+, R_t^-	Upward and downward reserve requirements at time t		
S_g^U, S_g^D	Start-up and shut-down ramp limits of thermal unit g		
$T_{s,t}^{\max}$	Conditional transmission section limit (C-TSL) of section s at time t		
T_i^p	C-TSL when conditional variable x_p falls within interval i		
$\tilde{T}^p$	C-TSL corresponding to conditional variable x_p		
T_g^U, T_g^D	The minimum up and down time of thermal unit g		
$u_{g,0}$	Initial status of conventional thermal unit g		
$\mathbf{V}$	Value matrix		

THE formulation and solution of the security-constrained unit commitment (SCUC) problem constitute a core component of market clearing for modern power systems. SCUC is typically modeled as a large-scale mixed-integer programming (MIP) problem [1]. The security constraints refer to not only transmission line constraints but also transmission section limits for key corridors such as ultra-high-voltage direct current (UHVDC) transmission corridors, inter-provincial tie-lines, and export channels for renewable energy bases.

The renewable energy sources exhibits significant geographical distribution disparity and output uncertainty, which imposes high requirements on the coordinated scheduling of transmission sections across regional power systems [2]. Traditional SCUC models construct static transmission capacity constraints with conservative limits, which guarantee operational security but tend to result in economic inefficiency. Given the increasing complexity of grid operations [3] and

the growing penetration of renewable energy generation, it has become imperative to maximize transmission capacity utilization while maintaining operational security. For example, dynamic line rating (DLR) focuses on the thermal behavior of transmission lines and dynamically determines the maximum allowable power based on real-time environmental conditions, including wind speed, ambient temperature, and solar radiation. Transmission section limits can be generally classified into two categories [4]. The first one is influenced by factors independent of generation schedules, such as load levels, breaker statuses, and automatic control device settings, and can thus be determined prior to the scheduling process. The second one, the conditional transmission section limit (C-TSL), depends on factors directly coupled with generation scheduling decisions, such as the number of committed units in a unit group, total committed generation capacity, and sectional power flows. Compared with the DLR that depends on environmental conditions, the C-TSL depends on internal system conditions, and each C-TSL value is only active when its associated condition is met.

To delineate the coupling relationship between the C-TSLs and generation scheduling, several existing studies have modeled C-TSLs through multi-dimensional criteria such as the number of committed units, committed capacity, and reserve capacity. Based on these criteria, an integrated optimization approach has been implemented to enable automatic matching between generation schedules and section limit boundary conditions [4]-[6]. This approach reduced the need for manual intervention after unit commitment by ensuring that the transmission capacity accurately reflected the system states.

Existing research on unit commitment considering C-TSLs generally adopts a common formulation strategy wherein a large number of binary auxiliary variables are introduced to represent conditional limit intervals, thereby transforming the piecewise C-TSL expressions into MIP-compatible constraints. These constraints are then integrated into the SCUC model and solved by mathematical optimization solvers, which is essentially a model-based method. From a mathematical perspective, MIP problems are classified as non-deterministic polynomial hard (NP-hard) [7], meaning that as the problem size increases, the solution time grows exponentially [8]. The consideration of C-TSLs incorporates a large number of binary auxiliary variables, which places a severe computational burden on the optimization process for large-scale power systems. Therefore, improving the computational efficiency of this modeling paradigm remains an important topic for future research.

The efficient solution of the SCUC problem has long been a key research challenge [9]. Advancements in artificial intelligence have made data-driven algorithms an effective complement to traditional model-based scheduling methods [10]. The data-driven algorithms refer to algorithms that leverage historical operational data to learn patterns and make decisions (e.g., unit commitment) without relying on explicit models. A substantial amount of historical operation data of the power system provides valuable information about the potential operational patterns of power grid [11]. Applying

machine learning to SCUC data reveals systematic couplings between unit commitment and system states, leading to four types of data-driven algorithms that improve computational efficiency:

1) Replacing model-based optimization methods with machine learning surrogates [12], [13]. For instance, [13] employed classification and regression trees (CARTs) and random forest models to directly predict unit commitment solutions, alleviating the computational burden of optimization solvers.

2) Reducing the number of constraints in the model [14]-[18]. For example, [18] leveraged historical data with the k -nearest neighbors (KNN) algorithm to predict and filter out network constraints that were unlikely to be binding, significantly accelerating the computational process.

3) Reducing the number of variables [19]-[21]. For example, [21] developed a data-driven algorithm based on K -means clustering to fix part of the unit on/off states through a combined offline and online scheme.

4) Reducing the number of variables and constraints simultaneously [22] - [25]. For instance, [26] employed a graph neural network (GNN) to capture spatial correlations, while temporal dependencies were modeled using long short-term memory (LSTM) networks, facilitating the simultaneous prediction of generator commitments and active transmission constraints [22]. Reference [24] introduced a multi-scale spatiotemporal transformer to reduce variables and constraints through forecasting, leveraging its attention mechanism that excelled in sequence modeling [27], [28].

In [29], a data-driven algorithm based on probabilistic classification was proposed to identify inactive C-TSL intervals, thereby improving the computational efficiency of SCUC. Considering the coupling between unit status and C-TSLs, we advance this methodology by proposing a deep learning algorithm to predict unit status variables, which facilitates the identification of inactive C-TSL intervals and consequently reduces the complexity of the SCUC model.

The main contributions of this study are presented as follows.

1) A novel data-model hybrid-driven approach is proposed for solving unit commitment problems. Furthermore, a unit commitment prediction algorithm is developed to reduce the problem scale by fixing a subset of unit commitment variables, thereby eliminating inactive C-TSL intervals. Consequently, the number of variables and constraints in the SCUC model is significantly reduced, which helps the optimization solver to shorten the solution time. An iterative adjustment mechanism is also developed to guarantee the feasibility of prediction results.

2) The proposed algorithm incorporates a convolutional attention module to perform deep feature extraction and enhancement on time series data. By integrating convolutional operations with channel attention, it effectively captures the cross-temporal coupling characteristics embedded in the time series, providing enhanced feature representations for the subsequent unit commitment prediction.

3) The proposed algorithm also leverages a multi-head cross-attention mechanism that fuses temporal and non-tem-

poral features to improve prediction performance. Parallel attention heads capture latent correlations between temporal and non-temporal features, generating complementary representations that effectively fuse generator-related information.

II. PROBLEM FORMULATION AND SOLUTION FRAMEWORK

A. Formulation of SCUC Problem Considering C-TSLs

The objective function of the SCUC problem considering C-TSLs is shown in (1), which aims to minimize the total production cost. The nonlinear generation cost in (1) is represented by a piecewise linear function [30], [31] to enable efficient solution within the mixed-integer linear programming (MILP) framework [32]. Formulas (2) - (11) impose constraints on the units. The start-up and shut-down logics are enforced by (2) and (3), respectively. The minimum up and down time constraints are imposed via (4)-(6). The minimum and maximum output limits for thermal units are modeled by (7) and (8), respectively. The scheduled output of renewable energy units cannot exceed their forecasted generation levels as enforced by (9). The upward and downward ramping capabilities of thermal units are bounded by (10) and (11), respectively. The supply-demand balance is enforced through (12). The upward and downward reserve requirements for system security are formulated as (13) and (14), respectively. The branch power flow limit and C-TSL are enforced via (15) and (16), respectively.

$$\min_{u_{g,t}, y_{g,t}, z_{g,t}, P_{g,t,l}} \left\{ \sum_{t \in \mathcal{T}} \sum_{g \in \mathcal{G}} (C_g^{\min} u_{g,t} + C_g^U y_{g,t} + C_g^D z_{g,t}) + \sum_{t \in \mathcal{T}} \sum_{g \in \mathcal{G}} \sum_{l \in \mathcal{L}} C_{g,l} P_{g,t,l} \right\} \quad (1)$$

s.t.

$$y_{g,t} - z_{g,t} = u_{g,t} - u_{g,t-1} \quad g \in \mathcal{G}, t \in \mathcal{T} \quad (2)$$

$$y_{g,t} + z_{g,t} \leq 1 \quad g \in \mathcal{G}, t \in \mathcal{T} \quad (3)$$

$$u_{g,t} = u_{g,0} \quad g \in \mathcal{G}, t \in \{t | t \leq L_g^U + L_g^D, t \in \mathcal{T}\} \quad (4)$$

$$\sum_{\tau=t-T_g^U+1}^t y_{g,\tau} \leq u_{g,t} \quad g \in \mathcal{G}, t \in \{t | t > L_g^U + L_g^D, t \in \mathcal{T}\} \quad (5)$$

$$\sum_{\tau=t-T_g^D+1}^t z_{g,\tau} \leq 1 - u_{g,t} \quad g \in \mathcal{G}, t \in \{t | t > L_g^U + L_g^D, t \in \mathcal{T}\} \quad (6)$$

$$0 \leq P_{g,t,l} \leq u_{g,t} P_{g,l}^{\max} \quad g \in \mathcal{G}, t \in \mathcal{T}, l \in \mathcal{L} \quad (7)$$

$$P_{g,t} = P_g^{\min} u_{g,t} + \sum_{l \in \mathcal{L}} P_{g,t,l} \quad g \in \mathcal{G}, t \in \mathcal{T} \quad (8)$$

$$0 \leq P_{m,t}^s \leq P_{m,t}^F \quad m \in \mathcal{M}, t \in \mathcal{T} \quad (9)$$

$$P_{g,t} - P_{g,t-1} \leq R_g^U u_{g,t-1} + S_g^U y_{g,t} \quad g \in \mathcal{G}, t \in \mathcal{T} \quad (10)$$

$$P_{g,t-1} - P_{g,t} \leq R_g^D u_{g,t} + S_g^D z_{g,t} \quad g \in \mathcal{G}, t \in \mathcal{T} \quad (11)$$

$$\sum_{m \in \mathcal{M}} P_{m,t}^s + \sum_{g \in \mathcal{G}} P_{g,t} = P_t^L \quad t \in \mathcal{T} \quad (12)$$

$$r_t^+ = \sum_{g \in \mathcal{G}} (P_g^{\max} - P_{g,t}) u_{g,t} \geq R_t^+ \quad t \in \mathcal{T} \quad (13)$$

$$r_t^- = \sum_{g \in \mathcal{G}} (P_{g,t} - P_g^{\min}) u_{g,t} \geq R_t^- \quad t \in \mathcal{T} \quad (14)$$

$$\left| \sum_{g \in \mathcal{G}} \lambda_{g,b}^P P_{g,t} + \sum_{m \in \mathcal{M}} \lambda_{m,b}^S P_{m,t}^s - \sum_{n \in \mathcal{N}} \lambda_{n,b}^L P_{n,t}^L \right| \leq F_b^{\max} \quad b \in \mathcal{B}, t \in \mathcal{T} \quad (15)$$

$$\sum_{g \in \mathcal{G}} \chi_{g,s}^P P_{g,t} + \sum_{m \in \mathcal{M}} \chi_{m,s}^S P_{m,t}^s - \sum_{n \in \mathcal{N}} \chi_{n,s}^L P_{n,t}^L \leq T_{s,t}^{\max} \quad s \in \mathcal{S}, t \in \mathcal{T} \quad (16)$$

$T_{s,t}^{\max}$ is determined by taking the intersection of all conditional limits, which corresponds to selecting the minimum value among all computed results from different condition variables, as expressed in (17) and (18). The SCUC model specifically considers the number of committed units, the committed generation capacity, and the reserve capacity as condition variables.

$$T_{s,t}^{\max} = \min \{\tilde{T}^1, \tilde{T}^2, \dots, \tilde{T}^P\} \quad s \in \mathcal{S}, t \in \mathcal{T} \quad (17)$$

$$\tilde{T}^p = \begin{cases} T_1^p & x_p \in [M_1^p, M_2^p) \\ T_2^p & x_p \in [M_2^p, M_3^p) \\ \vdots & \\ T_N^p & x_p \in [M_N^p, M_{N+1}^p) \end{cases} \quad p = 1, 2, \dots, P \quad (18)$$

Since x_p is difficult to determine prior to the optimization of the SCUC model and $\tilde{T}^p$ is a conditional piecewise function, (18) is reformulated as (19)-(21) [4].

$$\sum_{i=1}^N y_i^p = 1 \quad p = 1, 2, \dots, P \quad (19)$$

$$\sum_{i=1}^N y_i^p M_i^p \leq x_p < \sum_{i=1}^N y_i^p M_{i+1}^p \quad p = 1, 2, \dots, P \quad (20)$$

$$\tilde{T}^p = \sum_{i=1}^N y_i^p T_i^p \quad p = 1, 2, \dots, P \quad (21)$$

y_i^p is introduced to represent the active interval of the C-TSLs. Specifically, $y_i^p = 1$ indicates that the conditional variable falls within interval i , and $y_i^p = 0$ otherwise. The constraint (19) guarantees that only one segment is activated at a time. When a specific interval is activated, the associated conditional variables must satisfy predefined interval boundaries as expressed in (20). Thus, $\tilde{T}^p$ is converted from a conditional piecewise function to a linear function in (21).

In sum, the SCUC problem is formulated as an MILP problem. Since the solution time of the MILP problem typically grows rapidly with increasing integer variables, the large number of commitment status variables $u_{g,t}$ is a key factor affecting the solution time. Furthermore, by modeling the transmission section limits as C-TSLs, $N \times P$ additional integer variables are introduced during each time period. Over a scheduling horizon of N_t time periods, this formulation entails $N_t \times N \times P$ additional integer variables, thereby significantly exacerbating the computational complexity of the SCUC problem.

B. Solution Framework

To solve the SCUC problem considering C-TSLs, a data-model hybrid-driven approach is proposed. The corresponding solution framework is shown in Fig. 1. The framework

employs a unit commitment prediction algorithm to reduce the number of integer variables, followed by solving the reduced problem with an optimization solver, aiming to shorten the overall computational time.

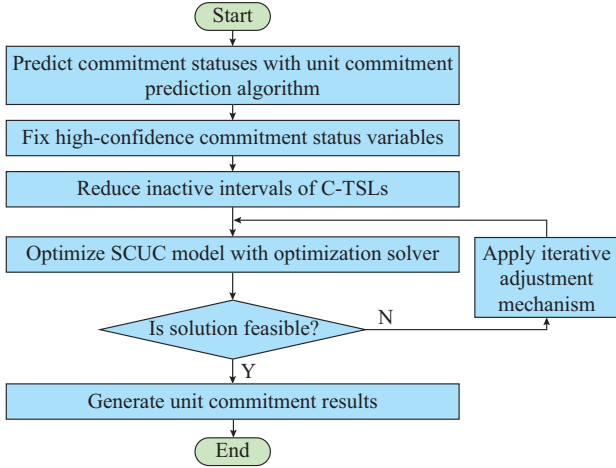


Fig. 1. Solution framework of proposed data-model hybrid-driven approach.

C. Reduction of Inactive C-TSL Intervals

The number of committed units, committed capacity, and upward reserve capacity within unit groups are critical conditional variables influencing C-TSLs. Here, different unit groups are classified according to the sending and receiving ends of the conditional transmission section: $\mathcal{G}_t^a = \{g | u_{g,t} = 1, g \in \mathcal{G}^a\}$, $\tilde{\mathcal{G}}_t^a = \{g | u_{g,t} \neq 0, g \in \mathcal{G}^a\}$, $\hat{\mathcal{G}}_t^a = \{g | u_{g,t} \neq 0, g \in \mathcal{G}\}$. The logic to set $u_{g,t}$ will be presented in Section III-E. The inclusion relationships among these sets are: $\mathcal{G}_t^a \subseteq \tilde{\mathcal{G}}_t^a \subseteq \hat{\mathcal{G}}_t^a \subseteq \mathcal{G}$, and $\tilde{\mathcal{G}}_t^a \subseteq \hat{\mathcal{G}}_t^a \subseteq \mathcal{G}$.

The unit commitment prediction results directly impact the upper and lower bounds of the number of committed units. Specifically, in the final unit commitment decision, the minimum number of online units in unit group a at time t corresponds to the number of units predicted as “on” in unit group a , while the maximum equals the total number of units in unit group a minus the number of units predicted as “off”, i.e.:

$$\begin{cases} S_{a,t}^{\min} = |\mathcal{G}_t^a| \\ S_{a,t}^{\max} = |\hat{\mathcal{G}}_t^a| \end{cases} \quad (22)$$

A similar principle applies to the committed capacity as:

$$\begin{cases} C_{a,t}^{\min} = \sum_{g \in \mathcal{G}_t^a} P_g^{\max} \\ C_{a,t}^{\max} = \sum_{g \in \hat{\mathcal{G}}_t^a} P_g^{\max} \end{cases} \quad (23)$$

Regarding the system-wide upward reserve capacity, its strong correlation with the actual power output of online units renders it significantly more challenging to predict accurately. The minimum reserve capacity must satisfy the system reserve requirement at time t , whereas the maximum reserve capacity cannot exceed the aggregate capacity of all possible online units minus the system demand, i.e.:

$$\begin{cases} R_t^{\min} = R_t^+ \\ R_t^{\max} = \sum_{g \in \tilde{\mathcal{G}}_t} P_g^{\max} + \sum_{m \in \mathcal{M}} P_{m,t}^F - P_t^L \end{cases} \quad (24)$$

Other conditional variables such as transmission section power flows can also be bounded according to the maximum and minimum numbers of committed units within a group. Within the SCUC model, inactive C-TSL intervals are pruned according to the tightened bounds of conditional variables. Subsequently, the associated auxiliary integer variables are fixed to zero, and the redundant constraints are eliminated. This reduction method effectively reduces the number of binary auxiliary variables and constraints required for identifying the active intervals.

III. UNIT COMMITMENT PREDICTION ALGORITHM

Owing to the complex temporal dependencies inherent in unit commitment time series, conventional deep learning algorithms may not adequately capture the critical features and long-term patterns. To overcome this limitation, a unit commitment prediction algorithm is proposed based on attention mechanisms to enhance feature representation and improve decision accuracy. The modular structure of the unit commitment prediction algorithm is shown in Fig. 2, where CNN is short for convolutional neural network.

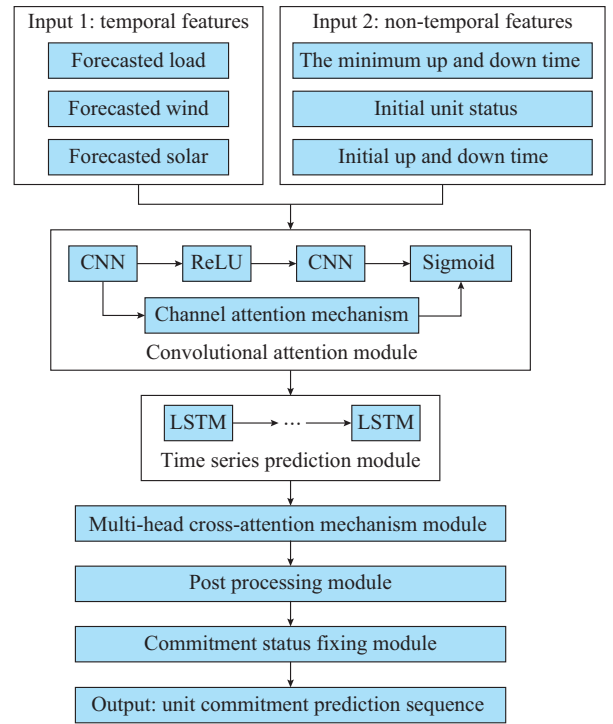


Fig. 2. Modular structure of unit commitment prediction algorithm.

Regarding unit commitment prediction, existing studies employed GNN to capture spatial correlations within the data, incorporating power grid topology and net load features [22], [25]. Although [22] utilized the LSTM network to capture temporal correlations, such approaches often fail to fully capture the intricate cross-temporal coupling characteristics embedded within the time series. In this study, we de-

sign a convolutional attention module to enhance temporal feature representations for the time series prediction module based on LSTM network, and further leverage a multi-head cross-attention mechanism module to fuse temporal and non-temporal features. A post processing module further ensures the feasibility of the predicted commitment statuses. High-confidence commitment statuses are then fixed to 0 or 1 in the commitment status fixing module. The pseudocode of the proposed algorithm is presented in Supplementary Material A Algorithm SAI.

A. Convolutional Attention Module

A convolutional attention module is employed to perform deep feature extraction and enhancement on the input time series. The structure of the convolutional attention module is illustrated in Fig. 3. This module first uses a CNN to extract temporal features from the temporal data, generating multi-channel three-dimensional features. It then executes the following operations in parallel to enhance features.

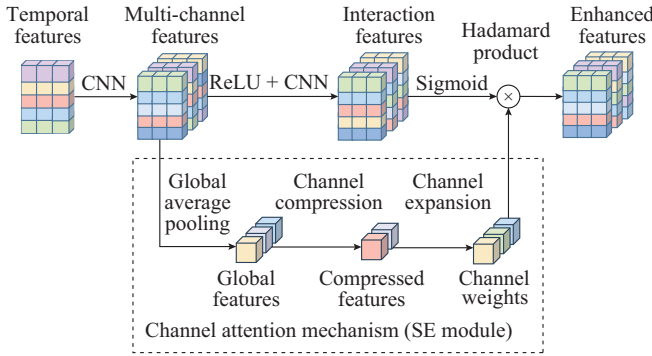


Fig. 3. Structure of convolutional attention module.

1) Channel-wise dynamic weighting: a squeeze and excitation (SE) module [33] is incorporated to facilitate the adaptive calibration of feature channels, functioning as a channel attention mechanism. This process involves global average pooling followed by fully connected layers to learn channel weights.

2) Feature interaction: convolution operations are performed on the feature dimensions, with both ReLU and Sigmoid activation functions employed to enhance the modeling of nonlinear features.

Finally, the interaction features are multiplied by the learned channel weights via a Hadamard product to emphasize informative features while suppressing less relevant ones. This structure not only accurately captures the intrinsic patterns across heterogeneous data types but also exploits the complex correlations inherent in load and renewable energy generation forecasts, thereby yielding enhanced feature representations for the subsequent unit commitment prediction task.

B. Time Series Prediction Module

While traditional recurrent neural networks (RNNs) are capable of capturing dynamic features via sequential propagation, they are susceptible to the vanishing gradient problem over long input sequences, which compromises their ability

to retain historical information and consequently degrades prediction accuracy. To address the need for modeling long-term historical dependencies in unit commitment prediction, this paper employs the LSTM network for time series prediction. The LSTM network addresses this issue by employing a gating mechanism within its sequential chain structure, thereby effectively preserving long-term dependencies. Consequently, it is extensively adopted for time series prediction [22], [26]. The structure of the LSTM network for a single step is shown in Fig. 4.

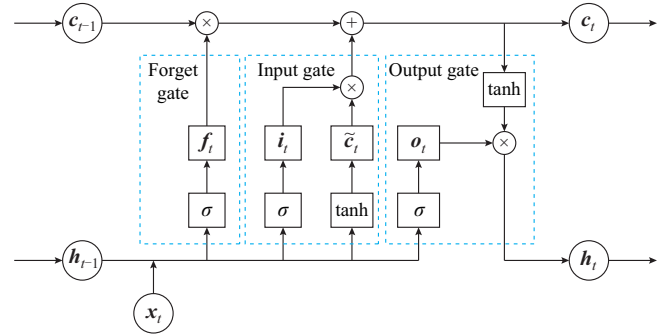


Fig. 4. Structure of LSTM network for a single step.

The LSTM network dynamically regulates the information flow through its gating mechanism. The calculation formulas are given by (25).

$$\begin{cases} f_t = \sigma(W^F[h_{t-1}, x_t] + b^F) \\ i_t = \sigma(W^I[h_{t-1}, x_t] + b^I) \\ \tilde{c}_t = \tanh(W^C[h_{t-1}, x_t] + b^C) \\ c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \\ o_t = \sigma(W^O[h_{t-1}, x_t] + b^O) \\ h_t = o_t * \tanh(c_t) \end{cases} \quad (25)$$

where the symbol “*” denotes the Hadamard product (element-wise multiplication).

The Sigmoid activation function $\sigma(\cdot)$ maps values into the range (0, 1), serving as a gating mechanism that determines how much information is retained or discarded. The $\tanh(\cdot)$ activation function produces outputs in the range (-1, 1), and is used to regulate the scale of $\tilde{c}_t$ and h_t . These two activation functions work together to maintain long-term dependencies and guarantee gradient stability in the information flow. At the final time, h_t is transformed via the Sigmoid activation function to produce the time series prediction output.

C. Multi-head Cross-attention Mechanism Module

A heterogeneous feature interaction strategy is implemented via a multi-head cross-attention mechanism to synthesize temporal and non-temporal features, which enables cross-modal interaction by aggregating contextual information across distinct modalities [27]. This strategy employs multi-head mapping and parallel computation scheme to achieve deep fusion of heterogeneous information. Specifically, the module takes the output of the time series prediction module as the query features, while the minimum up/down time and the initial unit status features are mapped into key-value

pairs. Through linear transformations and multi-head operations, these features are projected into multiple orthogonal subspaces where scaled dot product attention (SDPA) operation is performed independently. SDPA is formulated as follows:

$$SDPA = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d^k}} \right) \mathbf{V} \quad (26)$$

d^k is used to scale the dot product between $\mathbf{Q}$ and $\mathbf{K}$ to prevent numerical overflow. SDPA computes the similarity between $\mathbf{Q}$ and $\mathbf{K}$ to reweight $\mathbf{V}$. Finally, the outputs from multiple heads are concatenated and linearly transformed to generate the fusion feature.

The structure of the multi-head cross-attention mechanism module is illustrated in Fig. 5. By leveraging parallel computation across multiple attention heads, the model concurrently extracts latent correlations between temporal and non-temporal features within diverse representation subspaces. The outputs from all subspaces are concatenated and linearly transformed to form complementary features, which are then fused with the original temporal features through a residual connection.

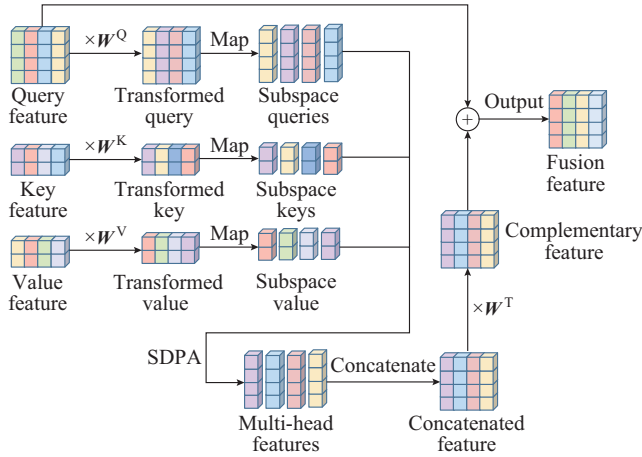


Fig. 5. Structure of multi-head cross-attention mechanism module.

D. Post Processing Mechanism Module

In this subsection, a rule-guided sequence post processing mechanism is proposed to ensure consistency between predicted commitment status and historical initial conditions, thereby enhancing the operational feasibility of the generated schedules. The core idea of this mechanism is to dynamically correct the early-stage predictions based on the minimum up/down time constraints. Specifically, when the mechanism detects that the consecutive up/down time does not meet the minimum up/down time threshold in the early periods, it will locally adjust the prediction results, extending the continuous on/off duration to comply with the constraint requirements. This process can be expressed as:

$$u_{z,g,t}^{\text{pre}} = \begin{cases} 1 & I_g^U > 0, z \in \mathcal{Z}^{\text{test}}, g \in \mathcal{G}, t \in \{t | t \leq T_g^U - I_g^U, t \in \mathcal{T}\} \\ 0 & I_g^D > 0, z \in \mathcal{Z}^{\text{test}}, g \in \mathcal{G}, t \in \{t | t \leq T_g^D - I_g^D, t \in \mathcal{T}\} \end{cases} \quad (27)$$

E. Commitment Status Fixing Module

1) Setting Confidence Thresholds

In this part, commitment status variables are fixed by setting confidence thresholds $\alpha \in (0, 1)$ and $\beta \in (0, 1)$.

$$u_{z,g,t} = \begin{cases} 1 & u_{z,g,t}^{\text{pre}} > \alpha \\ 0 & u_{z,g,t}^{\text{pre}} < \beta \end{cases} \quad z \in \mathcal{Z}^{\text{test}} \quad (28)$$

When $u_{z,g,t}^{\text{pre}} > \alpha$ or $u_{z,g,t}^{\text{pre}} < \beta$, the prediction is considered to be of high confidence, and the corresponding commitment status variables are fixed to 1 or 0. Otherwise, the prediction is regarded as unreliable and the corresponding commitment status variables are left to be determined by the optimization solver.

2) Iterative Adjustment Mechanism

Given that unit commitment must satisfy multiple constraints, the prediction results carry a potential risk of constraint violations [34]. If the SCUC model becomes infeasible, the two confidence thresholds are adjusted to restore the feasibility of the optimization problem. Specifically, the confidence thresholds can be gradually adjusted with a certain step size. Let the initial confidence thresholds be α_0 and β_0 . If the SCUC model is infeasible, the confidence thresholds are updated as:

$$\begin{cases} \alpha_{k+1} = \min \{ \alpha_k + \eta(1 - \alpha_0), 1 \} \\ \beta_{k+1} = \max \{ \beta_k - \eta\beta_0, 0 \} \end{cases} \quad (29)$$

This process continues iteratively until a feasible solution is found. In the extreme case where no feasible threshold can be found, the iterative adjustment mechanism will revert to solving the full SCUC problem without any data-driven fixation, i.e., $\alpha_k = 1$ and $\beta_k = 0$. This indicates that the threshold boundaries have been reached, and no further adjustments are needed. This iterative adjustment mechanism effectively addresses the issue of model infeasibility caused by inappropriate selection of initial thresholds.

IV. RESULTS AND ANALYSIS

A. Extended IEEE 39-bus System Case

1) Case Description

In this part, we construct a test case to evaluate the proposed approach. Specifically, the IEEE 39-bus system is used as the receiving-end power grid. DC Section A is connected at Bus 16 to interconnect with sending-end power grid S1, while DC Section B is connected at Bus 4 to link with sending-end power grid S2. Both S1 and S2 consist of three thermal units and one equivalent renewable energy unit. Here, DC Sections A and B are both conditional sections. The DC transmission limits are constrained by the number of online thermal units at both sending and receiving ends and the upward reserve capacity of receiving ends. Accordingly, the conditional rules for DC Sections A and B are given in Tables I [6] and II [4], respectively. An extended IEEE 39-bus system is illustrated in Fig. 6.

All numerical experiments are conducted on a computer equipped with an Intel Core i7-10510U CPU (1.80 GHz), an NVIDIA GeForce MX250 GPU, and 8 GB of RAM. The software environment is based on Anaconda3 with Python 3.8.8. The deep learning framework is built using Tensor-

Flow 2.7.0, while the SCUC model is solved using Gurobi 11.0.1 solver. The solver employs a branch-and-cut algorithm to efficiently handle the MILP formulation. By solving relaxed linear subproblems, branching on binary variables, and iteratively tightening bounds, the branch-and-cut algorithm converges to a solution within a predefined optimality gap.

TABLE I
CONDITIONAL RULES FOR DC SECTION A

Transmission limit (MW)	Online capacity of receiving end (MW)	Upward reserve capacity of receiving end (MW)
150	≥ 600	≥ 75
200	≥ 700	≥ 100
250	≥ 900	≥ 125
300	≥ 1000	≥ 150
350	≥ 1100	≥ 160

TABLE II
CONDITIONAL RULES FOR DC SECTION B

Transmission limit (MW)	Number of committed units	
	S2	Receiving end
200	≥ 2	≥ 4
100		< 4
250	< 2	≥ 4
150		< 4

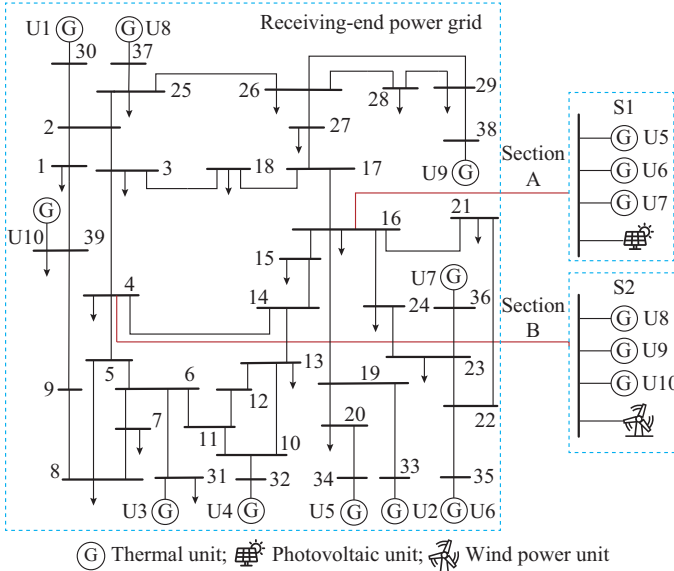


Fig. 6. An extended IEEE 39-bus system.

In SCUC research, it is common practice to leverage publicly available historical or forecast data from real-world power markets, e. g., Pennsylvania-New Jersey-Maryland (PJM) market, Midcontinent Independent System Operator (MISO), and California Independent System Operator (CAISO). These data are then scaled or preprocessed before being applied to standard test systems [25], [35]. In this study, the hourly data of load and renewable energy generation forecast are sourced from the PJM market for the years

2021-2023. The unit parameters are taken from [36]. The three thermal units in S1 have the same specifications as thermal units U5, U6, and U7 in the receiving-end power grid, while the three thermal units in S2 correspond to thermal units U8, U9, and U10 in the receiving-end power grid. An SCUC model is constructed and solved to generate a dataset containing 26280 samples, which are randomly divided into training, validation, and test sets with a ratio of 5:2:3. It is common to adopt a small but non-zero mixed-integer programming gap (MIPGap) as a stopping criterion in unit commitment, in order to balance solution accuracy and computational time [17], [21]. All test systems are solved with an MIPGap tolerance of 0.01% to obtain near-optimal solutions, except for the IEEE 73-bus system, which is also solved to optimality with an MIPGap tolerance of 0.

2) Training Process of Proposed Algorithm

The proposed algorithm is compared with three popular time series prediction algorithms: an algorithm based on GNN-LSTM model [22], an algorithm based on standard LSTM network, and an algorithm based on transformer architecture, which are denoted as algorithms 1, 2, and 3, respectively. All these algorithms are trained using the Adam optimizer, and the binary cross-entropy loss function, denoted as *Loss*, is used to evaluate the prediction loss. The binary accuracy metric *Acc* is employed to evaluate the prediction performance of unit commitment decisions. Specifically, *Acc* is defined as the proportion of correctly predicted on/off statuses across all thermal units and time. The formulas for calculating *Loss* and *Acc* are expressed as:

$$Loss = -\frac{1}{N_z N_g N_t} \sum_{z=1}^{N_z} \left\{ \sum_{g \in \mathcal{G}} \sum_{t \in T} [y_{z,g,t} \ln(u_{z,g,t}^{\text{pre}}) + (1 - y_{z,g,t}) \ln(1 - u_{z,g,t}^{\text{pre}})] \right\} \quad (30)$$

$$Acc = 1 - \frac{1}{N_z N_g N_t} \sum_{z=1}^{N_z} \left(\sum_{g \in \mathcal{G}} \sum_{t \in T} |y_{z,g,t} - y'_{z,g,t}| \right) \quad (31)$$

If $u_{z,g,t}^{\text{pre}} \geq 0.5$, $y'_{z,g,t} = 1$; if $u_{z,g,t}^{\text{pre}} < 0.5$, $y'_{z,g,t} = 0$. The hyperparameter configurations for these deep learning algorithms and the corresponding settings for grid search are summarized in Supplementary Material A Table SAI.

Figure 7 illustrates the curves of *Loss* and *Acc* under different algorithms, which shows that the proposed algorithm not only exhibits a more rapid convergence rate in terms of loss but also attains a significantly superior accuracy compared to the other algorithms.

3) Predictive Performance

We evaluate the predictive performance of different algorithms from three aspects: the unit commitment prediction accuracy on the test set, the reduction rate of inactive C-TSL intervals, and the training time for each algorithm. Regarding the configuration of initial thresholds, priority is placed on guaranteeing solution feasibility rather than maximizing computational speed. We adopt a conservative and unified initial setting across all test cases: $\alpha = 0.99$ and $\beta = 0.01$. The average reduction rate of inactive C-TSL intervals is denoted as *Rec*. The prediction performance comparison of different algorithms on the extended IEEE 39-bus system is shown in Table III.

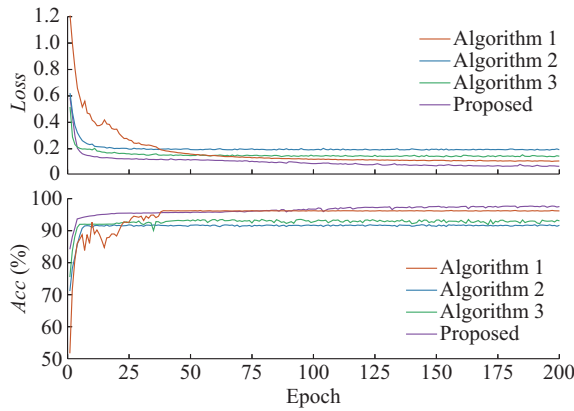


Fig. 7. Curves of *Loss* and *Acc* under different algorithms.

TABLE III
PREDICTION PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS
ON EXTENDED IEEE 39-BUS SYSTEM

Algorithm	<i>Acc</i> (%)	<i>Rec</i> (%)	Training time (s)
Algorithm 1	96.45	45.85	912
Algorithm 2	93.63	35.50	305
Algorithm 3	94.26	36.37	426
Proposed	98.05	51.96	590

The proposed algorithm achieves a higher *Acc* of 98.05%, which directly contributes to improved reduction rates for the number of committed units and committed capacity. Consequently, the proposed algorithm enables precise localization of the active intervals of conditional variables, achieving the highest *Rec* of 51.96%.

4) Solution Efficiency

We evaluate the solution efficiency of the proposed approach in solving the SCUC problem from the aspects of total solution time, objective function value (total production cost), and number of infeasible samples. Table IV presents a comparison of different approaches on the extended IEEE 39-bus system, where approaches 1, 2, and 3 denote data-model hybrid-driven approaches based on algorithms 1, 2, and 3, respectively.

TABLE IV
COMPARISON OF DIFFERENT APPROACHES ON EXTENDED IEEE 39-BUS
SYSTEM

Approach	Solution time (s)	Total production cost (\$)	Number of infeasible samples
Model-based	1.63	331515	0
Approach 1	0.45	331522	1
Approach 2	0.96	331521	3
Approach 3	0.74	331523	2
Proposed	0.27	331521	0

In terms of solution time, the proposed approach achieves an 83.44% reduction relative to the model-based approach, thereby demonstrating a significant acceleration in the solving process. Regarding the total production cost, the result of the proposed approach deviates by less than 0.01% from that of the model-based approach, indicating excellent model-

ing accuracy. Moreover, the proposed approach does not yield infeasible solutions.

5) Ablation Results

To verify the effectiveness of the two attention mechanisms in the proposed algorithm, we compare it with two algorithms with ablated models on the same dataset for the unit commitment prediction task. Specifically, RemoveCA refers to the algorithm with the removal of the multi-head cross-attention mechanism module. RemoveSE refers to the algorithm with the removal of the channel attention mechanism, i.e., the SE module. The ablation results are shown in Fig. 8.

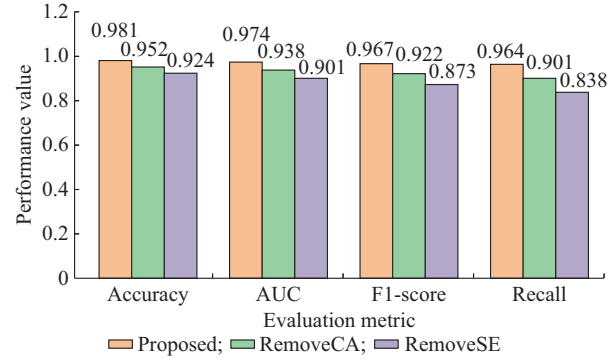


Fig. 8. Ablation results.

Several common evaluation metrics are used in binary classification tasks to assess module performance: accuracy, recall, F1-score, and area under the receiver operating characteristic curve (denoted as AUC). Accuracy represents the proportion of correctly predicted samples. Recall indicates the proportion of units that are actually online and are correctly predicted to be online. F1-score is the harmonic mean of precision and recall. Here, precision refers to the proportion of correctly predicted online units among all units that were predicted as online. A higher F1-score suggests that the corresponding algorithm achieves a better balance between precision and recall. AUC provides a comprehensive evaluation of the classification performance across different threshold values. The closer the AUC is to 1, the better the classification performance of the algorithm. The performance drops significantly after removing the SE module. Specifically, the accuracy decreases by 5.7%, and the recall drops significantly by 12.6%. This module captures global feature information through global average pooling and applies weights to each channel, effectively integrating local features with global information. In the absence of this module, the unit commitment prediction algorithm is restricted to local features extracted by the convolutional layers without the global context, thereby resulting in a marked decline in performance. The performance degradation after removing the multi-head cross-attention mechanism module also indicates its importance, with the accuracy decreasing by 2.9% and recall declining by 6.3%. The multi-head cross-attention mechanism module enables the unit commitment prediction algorithm to better understand the relationships between temporal and non-temporal features. By fusing heterogeneous features, the

overall prediction accuracy is improved. Therefore, the ablation results validate the effectiveness of the attention mechanisms in the proposed algorithm.

B. Extended IEEE 73-bus System Case

Furthermore, we construct a larger test case based on the extended IEEE 73-bus system, whose transmission sections are illustrated in Fig. 9.

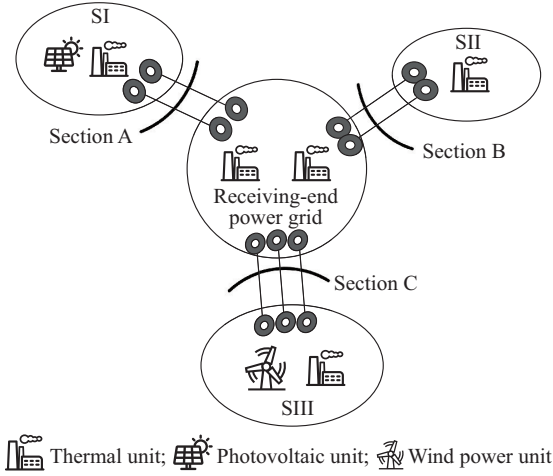


Fig. 9. Illustration of transmission sections in extended IEEE 73-bus system.

The entire system contains 67 conventional thermal units, with 42 in the receiving-end power grid, 7 in the sending-end power grid SI, 14 in the sending-end power grid SII, and 4 in sending-end power grid SIII. Furthermore, SI has one equivalent photovoltaic unit, while SIII has one equivalent wind power unit. Due to the intermittency and uncertainty of renewable energy sources, the system implements a bundled transmission scheme of renewable energy generation and thermal power generation. The parameters of the units are derived from [22]. Specifically, the limit constraint for Section A is conditional on the number of online units, while that for Section B depends on the upward reserve capacity. The capacity limit of Section C is determined by the transmission power of both Sections A and B. A total of 18 conditional section rules are contained in the entire system. The conditions for these sections are shown in Supplementary Material A Tables SAII-SAIV.

The prediction performance comparison of different algorithms on the extended IEEE 73-bus system is shown in Table V. The comparison of different approaches on the extended IEEE 73-bus system is shown in Table VI. In the prediction phase, the proposed algorithm achieves an *Acc* of 96.39% and an *Rec* of 44.81%, significantly outperforming both algorithms 2 and 3. Although the training time of the proposed algorithm (954 s) is longer than those of algorithm 2 (314 s) and algorithm 3 (497 s), it is much shorter than that of algorithm 1 (2063 s). While algorithms 2 and 3 exhibit shorter training time, their prediction accuracy is relatively lower, leading to several infeasible solutions. The proposed algorithm, however, maintains high prediction accuracy with-

out producing infeasible solutions. In the solution phase, the proposed approach converges within 6.43 s, representing a 56.41% reduction in solution time against the model-based approach (14.75 s), and has a negligible total production cost deviation (<0.01%). The proposed algorithm and approach demonstrate superiority in prediction performance and solution efficiency, respectively.

TABLE V
PREDICTION PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS ON EXTENDED IEEE 73-BUS SYSTEM

Algorithm	<i>Acc</i> (%)	<i>Rec</i> (%)	Training time (s)
Algorithm 1	93.57	45.67	2063
Algorithm 2	90.82	42.18	314
Algorithm 3	92.95	48.72	497
Proposed	96.39	44.81	954

TABLE VI
COMPARISON OF DIFFERENT APPROACHES ON EXTENDED IEEE 73-BUS SYSTEM

Approach	Solution time (s)	Total production cost (\$)	Number of infeasible samples
Model-based	14.75	1544791	0
Approach 1	7.44	1545767	3
Approach 2	8.65	1544806	1
Approach 3	8.05	1544833	3
Proposed	6.43	1544863	0

Furthermore, we set MIPGap to be 0 to compare the performance of the proposed approach and the model-based approach in finding the global optimal solution, as shown in Table VII. The results indicate that the proposed approach reduces the solution time by 55.87% compared to the model-based approach. Moreover, the total production cost deviates by only 0.001% from that of the model-based approach, demonstrating the modeling accuracy of the proposed approach.

TABLE VII
COMPARISON OF MODEL-BASED APPROACH AND PROPOSED APPROACH WHEN MIPGAP IS 0

Approach	Solution time (s)	Total production cost (\$)	Number of infeasible samples
Model-based	16.43	1544782	0
Proposed	7.25	1544805	0

Table VIII presents the comparison of model size and solution efficiency between the model-based approach and the proposed approach based on the log analysis [37]-[39]. The solver logs of the model-based approach and the proposed approach are provided in Supplementary Material A Tables SAV and SAVI, respectively. These results demonstrate that the proposed approach significantly improves solution efficiency while preserving high solution quality.

TABLE VIII
COMPARISON OF MODEL SIZE AND SOLUTION EFFICIENCY BETWEEN
MODEL-BASED APPROACH AND PROPOSED APPROACH

Approach	Number of constraints	Number of integer variables	Number of iterations	Solution time (s)	Total production cost (\$)
Model-based	13670	4443	698769	74.71	1987854
Proposed	4615	1367	235601	21.83	1987854

C. Modified IEEE 118-bus System Case

To further assess the scalability of the proposed algorithm and approach, a case study is conducted on a modified IEEE 118-bus system, the topology of which is depicted in Supplementary Material A Fig. SA1. The modified IEEE 118-bus system comprises 54 thermal units, two wind farms located at Buses 15 and 30, and two photovoltaic plants located at Buses 82 and 100. Four transmission lines in the system are modeled as conditional transmission sections A, B, C, and D, corresponding to the lines connecting buses 15-33, 30-38, 82-77, and 100-99, respectively. Specific rules regarding the capacity limits of these sections are provided in Supplementary Material A Tables SAVII and SAVIII.

The test results on the modified IEEE 118-bus system further validate the superiority of the proposed algorithm and approach. As shown in Table IX, in the prediction phase, the proposed algorithm achieves an *Acc* of 97.15% and an *Rec* of 30.86%. As shown in Table X, in the solution phase, the proposed approach completes the solution of SCUC model within 3.69 s, which represents a substantial reduction of 72.84% relative to that of the model-based approach, thereby drastically enhancing decision-making efficiency. Meanwhile, the total production cost of the proposed approach is \$2747602, which is nearly identical to that of the model-based approach, with a deviation of less than 0.01%. Overall, the proposed approach achieves a balance among prediction accuracy, computational efficiency, and feasibility in large-scale systems.

TABLE IX
PREDICTION PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS ON
MODIFIED IEEE 118-BUS SYSTEM

Algorithm	<i>Acc</i> (%)	<i>Rec</i> (%)	Training time (s)
Algorithm 1	96.10	18.02	3809
Algorithm 2	94.45	25.10	435
Algorithm 3	95.67	29.92	582
Proposed	97.15	30.86	2055

TABLE X
COMPARISON OF DIFFERENT APPROACHES ON MODIFIED IEEE 118-BUS
SYSTEM

Approach	Solution time (s)	Total production cost (\$)	Number of infeasible samples
Model-based	13.58	2747584	0
Approach 1	6.88	2747711	0
Approach 2	7.65	2747592	1
Approach 3	6.08	2747597	1
Proposed	3.69	2747602	0

V. CONCLUSION

To efficiently solve the unit commitment problem considering C-TSLs, this paper proposes a data-model hybrid-driven approach, along with the design of a unit commitment prediction algorithm based on attention mechanisms. Numerical case studies demonstrate that the proposed approach substantially reduces the dimensionality of the optimization problem by fixing high-confidence commitment status variables and pruning inactive C-TSL intervals, thereby significantly alleviating the computational burden. The proposed approach reduces the solution time by 56.41% to 83.44% across the three test systems as compared to the model-based approach. The proposed algorithm employs a convolutional attention module to provide enhanced feature representations for unit commitment prediction. The proposed algorithm also leverages a multi-head cross-attention mechanism that fuses temporal and non-temporal features to capture the temporal coupling characteristics of unit commitment decisions effectively. The proposed algorithm achieves a prediction accuracy ranging from 96.39% to 98.05% and an average reduction rate of inactive C-TSL intervals ranging from 30.86% to 51.96%. The deviation in total production cost between the proposed approach and the model-based approach is less than 0.01%, confirming its high solution quality across different power grid scales.

From a practical perspective, this study provides two insights for practical power system operations.

1) By integrating data-driven prediction with model-driven optimization, the proposed approach leverages the strengths of both paradigms: machine learning excels in pattern recognition and feature extraction, while mathematical optimization ensures solution feasibility and optimality. This synergy enables significant improvements in computational efficiency without compromising solution quality.

2) Specifically, the proposed approach pre-determines a subset of commitment statuses while filtering out inactive C-TSL intervals. This significantly reduces the number of binary variables and constraints in the SCUC model, thereby improving computational efficiency. Given the prevalence of multiple C-TSLs in real-world power grids, the proposed approach demonstrates potential for practical use.

REFERENCES

- [1] X. Zheng, Y. Zhang, Z. Li *et al.*, "Day-ahead network-constrained unit commitment considering distributional robustness and intraday discreteness: a sparse solution approach," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 2, pp. 489-501, Apr. 2023.
- [2] L. Yi, M. Ding, J. Wang *et al.*, "A scenario-classification hybrid-based banding method for power transfer limits of critical inter-corridors," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 2, pp. 547-560, Mar. 2024.
- [3] N. Lu, T. Liu, and C. He, "Identification of transmission sections based on power grid partitioning," *International Transactions on Electrical Energy Systems*, vol. 29, no. 4, p. e2793, Apr. 2019.
- [4] Q. Ding, J. Yang, G. Wang *et al.*, "Integrated power generation plan optimization considering automatic matching of conditional transmission section limits," *Automation of Electric Power Systems*, vol. 44, no. 20, pp. 79-85, Oct. 2020.
- [5] F. Xu, R. Cao, M. Tu *et al.*, "Scheduling model for inter-regional HVDC tie-line generation and transmission system considering conditional sections," in *Proceedings of 2020 IEEE/IAS Industrial and Commercial Power System Asia*, Weihai, China, Jul. 2020, pp. 1285-1291.

- [6] D. Xu, J. Zhou, and G. Huang, "An integrated day ahead scheduling model considering the dynamic adjustment of DC limits," *Transactions of China Electrotechnical Society*, vol. 36, no. 13, pp. 2844-2851, Jul. 2021.
- [7] B. Asfora, J. Banfi, and M. Campbell, "Mixed-integer linear programming models for multi-robot non-adversarial search," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6805-6812, Oct. 2020.
- [8] S. Lin, G. Fan, G. Jian *et al.*, "Stochastic economic dispatch of power system with multiple wind farms and pumped-storage hydro stations using approximate dynamic programming," *IET Renewable Power Generation*, vol. 14, no. 13, pp. 2507-2516, Sept. 2020.
- [9] X. Liu and A. J. Conejo, "Quasi-deterministic proxy for network-constrained stochastic unit commitment," *Journal of Modern Power Systems and Clean Energy*, vol. 13, no. 4, pp. 1167-1175, Jul. 2025.
- [10] C. Zhao and X. Li, "Microgrid optimal energy scheduling considering neural network based battery degradation," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 1594-1606, Jan. 2024.
- [11] A. Varghese, H. Shah, B. Azimian *et al.*, "Deep neural network-based state estimator for transmission system considering practical implementation challenges," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 6, pp. 1810-1822, Nov. 2024.
- [12] S. Schmitt, I. Harjunkoski, M. Giuntoli *et al.*, "Fast solution of unit commitment using machine learning approaches," in *Proceedings of the IEEE 7th International Energy Conference*, Berlin, Germany, Jun. 2022, pp. 1-6.
- [13] X. Lin, Z. Hou, H. Ren *et al.*, "Approximate mixed-integer programming solution with machine learning technique and linear programming relaxation," in *Proceedings of the IEEE 3rd International Conference on Smart Grid and Smart Cities*, Shanghai, China, Apr. 2019, pp. 101-107.
- [14] Y. Yang, Z. Yang, J. Yu *et al.*, "Fast economic dispatch in smart grids using deep learning: an active constraint screening approach," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11030-11040, Nov. 2020.
- [15] J. Qin, N. Yu, and Y. Gao, "Solving unit commitment problems with multi-step deep reinforcement learning," in *Proceedings of 2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids*, Aachen, Germany, Nov. 2021, pp. 140-145.
- [16] F. Mohammadi, A. Sahraei, D. Trakas *et al.*, "Machine learning assisted stochastic unit commitment: a feasibility study," in *Proceedings of 2020 52nd North American Power Symposium*, Tempe, USA, Apr. 2021, pp. 1-6.
- [17] S. Zhang, H. Ye, F. Wang *et al.*, "Data-aided offline and online screening for security constraint," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2614-2622, May 2020.
- [18] S. Pineda, J. Morales, and C. Jimenez, "Data-driven screening of network constraints for unit commitment," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3695-3705, Sept. 2020.
- [19] A. Xavier, F. Qiu, and S. Ahmed, "Learning to solve large-scale security constrained unit commitment problems," *INFORMS Journal on Computing*, vol. 33, no. 2, pp. 739-756, Apr. 2021.
- [20] A. Ramesh and X. Li, "Machine learning assisted model reduction for security-constrained unit commitment," in *Proceedings of 2022 North American Power Symposium*, Salt Lake City, USA, Oct. 2022, pp. 1-6.
- [21] Y. Zhou, Q. Zhai, L. Wu *et al.*, "A data-driven variable reduction approach for transmission-constrained unit commitment of large-scale systems," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 254-266, Jan. 2022.
- [22] A. Ramesh and X. Li, "Spatio-temporal deep learning-assisted reduced security-constrained unit commitment," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 4735-4746, Mar. 2024.
- [23] Y. Yang, X. Lu, and L. Wu, "Integrated data-driven framework for fast SCUC calculation," *IET Generation, Transmission & Distribution*, vol. 14, no. 24, pp. 5728-5738, Jan. 2020.
- [24] M. Liu, X. Kong, K. Xiong *et al.*, "Multi-scale spatio-temporal transformer: a novel model reduction approach for day-ahead security-constrained unit commitment," *Applied Energy*, vol. 380, p. 124963, Dec. 2025.
- [25] S. Park, W. Chen, D. Han *et al.*, "Confidence-aware graph neural networks for learning reliability assessment commitments," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 3839-3850, Jun. 2023.
- [26] C. Liu, J. Gu, and M. Yang, "A simplified LSTM neural networks for one day-ahead solar power forecasting," *IEEE Access*, vol. 9, pp. 17174-17195, Jan. 2021.
- [27] J. Zhang, X. He, Y. Liu *et al.*, "Multi-modal cross-attention network for Alzheimer's disease diagnosis with multi-modality data," *Computers in Biology and Medicine*, vol. 162, p. 107050, Aug. 2023.
- [28] A. Vaswani, N. Shazeer, N. Parmar *et al.*, "Attention is all you need," in *Proceedings of Advances in Neural Information Processing Systems*, Long Beach, USA, Dec. 2017, pp. 5998-6008.
- [29] Y. Wu, H. Yu, M. Pan *et al.*, "Data-driven identification of active conditional transmission section limits in SCUC," in *Proceedings of 2025 IEEE International Conference on Power and Integrated Energy Systems*, Haikou, China, Apr. 2025, pp. 272-277.
- [30] C. Koutschan, B. Moser, A. Ponomarchuk *et al.*, "Representing piecewise linear functions by functions with small arity," *Applicable Algebra in Engineering, Communication and Computing*, vol. 36, no. 4, pp. 595-610, Jan. 2025.
- [31] N. Theodorakatos and R. Babu. (2025, Jun.). Hybrid optimization strategies for global optimality in non-convex programming using SQP and IPMs: avoiding the Maratos effect in PMU placement – a case study. [Online]. Available: <https://iopscience.iop.org/article/10.1088/1742-6596/3027/1/012041>
- [32] A. Latify, A. Mokhtari, A. Alavi-Eshkaftaki *et al.*, "Security-constrained unit commitment: modeling, solutions and evaluations," *Applied Energy*, vol. 390, p. 125796, Apr. 2025.
- [33] X. Jin, Y. Xie, X. Wei *et al.*, "Delving deep into spatial pooling for squeeze-and-excitation networks," *Pattern Recognition*, vol. 121, p. 108159, Jan. 2022.
- [34] F. Mohammadi, A. Sahraei, D. Trakas *et al.*, "Machine learning assisted stochastic unit commitment during hurricanes with predictable line outages," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5131-5142, Nov. 2021.
- [35] J. Qin and N. Yu, "Solve large-scale unit commitment problems by physics-informed graph learning," *IEEE Transactions on Power Systems*, vol. 40, no. 6, pp. 4439-4451, Nov. 2025.
- [36] H. Qin and H. Wei, "A quantum-inspired approximate dynamic programming algorithm for large-scale unit commitment problems," *Proceedings of the CSEE*, vol. 35, no. 19, pp. 4918-4929, Oct. 2015.
- [37] N. Theodorakatos, R. Babu, C. Theodoridis *et al.* (2024, Apr.). Mathematical models for the single-channel and multi-channel PMU allocation problem and their solution algorithms. [Online]. Available: <https://www.mdpi.com/1999-4893/17/5/191>
- [38] N. Theodorakatos, M. Lytras, A. Moschoudis *et al.*, "Implementation of optimization-based algorithms for maximum power system observability using synchronized measurements," in *Proceedings of AIP Conference Proceedings*, Belgrade, Serbia, Aug. 2023, p. 120006.
- [39] N. Theodorakatos, R. Babu, and A. Moschoudis. (2023, Nov.). The branch-and-bound algorithm in optimizing mathematical programming models to achieve power grid observability. [Online]. Available: <https://www.mdpi.com/2075-1680/12/11/1040>

Hong Yu received the bachelor's degree from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2023. She is currently pursuing the master's degree in control science and engineering with the School of Artificial Intelligence and Automation, HUST. Her research interests include scheduling optimization in power systems and artificial intelligence method.

Yong Zhao received the B.S., M.S., and Ph.D. degrees from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1989, 1992, and 1996, respectively. He is currently working in the School of Artificial Intelligence and Automation, HUST. His research interests include machine learning, power market, power dispatching, and decision theory.

Manli Yan received the bachelor's and master's degrees from Tongji University, Shanghai, China, in 2019 and 2021, respectively. She is currently pursuing the Ph.D. degree in control engineering with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China. Her research interests include electricity market, artificial intelligence-based bidding, and multi-agent system.

Yuanzheng Li received the M.S. degree in electrical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2011, and the Ph.D. degree in electrical engineering from the South China University of Technology, Guangzhou, China, in 2015. He is currently a Full Professor with HUST. His research interests include deep learning, reinforcement learning, smart grid, optimal power system and microgrid scheduling, and decision making.

Yaowen Yu received the B.S. degree in automation from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2011, and the Ph.D. degree in electrical engineering from the University of Connecticut, Storrs, USA, in 2016. He is currently an Associate Professor with HUST. His research interests include power and energy system scheduling, machine learning, and mechanism design and strategic bidding in electricity markets.