A Multi-Layer Data-Driven Security Constrained Unit Commitment Approach with Feasibility Compliance

Ali Feliachi *, Talha Iqbal, Muhammad Choudhry and Hasan Ul Banna

1. Introduction

In a power system, the daily load profile exhibits significant fluctuations between off-peak and peak demand hours. If generators are scheduled according to the peak demand of a day, then several generators will operate at their lowest power levels during off-peak hours. It will increase the cost of power production, which can be avoided using efficient scheduling. Hence, it is the job of a grid operator to determine cost-efficient generation schedules, known as a Unit Commitment (UC) problem. It is a mathematical optimization problem where the goal is to coordinate the production of a set of generators to meet the anticipated energy requirements, as illustrated in Figure 1.

Figure 1. Unit Commitment Problem Illustration (Colored boxes indicate the committed generators in the corresponding-colored time intervals).

Abstract: Security constrained unit commitment is an essential part of the day-ahead energy markets. The presence of discrete and continuous variables makes it a complex, mixed-integer, and time-hungry optimization problem. Grid operators solve unit commitment problems multiple times daily with only minor changes in the operating conditions. Solving a large-scale unit commitment problem requires considerable computational effort and a reasonable time. However, the solution time can be improved by exploiting the fact that the operating conditions do not change significantly in the day-ahead market clearing. Therefore, in this paper, a novel multi-layer data-driven approach is proposed, which significantly improves the solution time (90% time-reduction on average for the three studied systems). The proposed approach not only provides a near-optimal solution (<1% optimality gap) but also ensures that it is feasible for the stable operation of the system (0% infeasible predicted solutions). The efficacy of the developed algorithm is demonstrated through numerical simulations on three test systems, namely a 4-bus system and the IEEE 39-bus and 118-bus systems, and promising results are obtained.

Keywords: artificial intelligence; security constrained unit commitment; predictive modeling; mixed-integer optimization; machine learning; data-driven scheduling
Following deregulation in the USA in the 1990s, Independent System Operators (ISOs) and Regional Transmission Organizations (RTOs) replaced utilities as grid operators. Figure 2 illustrates the areas operated by RTOs in the USA. They typically run two bid-based markets to determine economic dispatch: real-time and day-ahead markets.

ISOs and RTOs solve security-constrained unit commitment (SCUC) problems multiple times daily for various probable operating scenarios to clear an annual energy market of USD 400 billion [2]. The goal is to set not only the least-cost schedules, which ensure balanced generation and demand but also to make sure that the system does not violate any security constraints. These constraints not only include systems security constraints, e.g., thermal limits of transmission lines to carry maximum power, but also include physical and operational constraints. The involvement of such security and operational constraints makes it a very complicated large-scale mixed-integer energy optimization problem. A classical mixed-integer-programming (MIP) solver takes about twenty minutes to reach a viable optimal generation strategy for a large power system with a duality gap of 0.1%. Improvement in this solution time and computational speed can result in multiple improvements, which can lead to a more efficient energy market and bring several additional rewards, such as the implementation of cost curves having higher granularity, within-interval dispatch, etc. [3].

This paper proposes a novel three-layer data-based algorithm to solve a SCUC problem, as described in Figure 3. The first layer involves training machine learning (ML) models to predict the binary commitment status of the generators based on the cost coefficients and the nodal demands. In the second layer, another set of ML models predicts the dispatch level of each of the generators using predictions from the first layer and cost and nodal demand data. Finally, the third layer implements a post-processing strategy to satisfy any violated constraints in the obtained predicted solutions from the first two layers. The contribution of this research work is twofold:

- A multi-layer modeling framework is proposed using predictive modeling techniques to model and solve a SCUC problem.
- A third feasibility compliance layer is proposed to ensure that the predicted optimal solution is also a feasible solution for the secure operation of the system.

![Figure 2. Electric Power Markets in the USA [1].](image)

![Figure 3. The Proposed Multi-Layer SCUC Algorithm.](image)
The rest of the article is set up as follows: Section 2 reviews commonly used approaches for solving a UC problem. Section 3 discusses problem formulation for a single time interval SCUC problem, followed by the proposed approach in Section 4. Case studies and simulation results are examined in Section 5. Section 6 presents a performance comparison of the CPLEX mixed-integer quadratic programming (MIQP) solver with the proposed multi-layer approach. Finally, Section 7 summarizes the conclusions and future work.

2. Literature Survey

In the literature, a lot of approaches have been explained for solving a UC problem which can be categorized in one of the categories below. Some of the most discussed optimization techniques are discussed in detail here.

- Exhaustive Enumeration (EE)
- Priority List (PL)
- Dynamic Programming (DP)
- Lagrangian Relaxation (LR)
- Mixed-Integer-Programming (MIP)
- Evolutionary Programming (EP)

In EE techniques, UC problems are solved by listing all combinations of binary variables representing the commitment status of generators. The sequences that result in the least operational cost are selected as optimal solutions for commitment decisions. Earlier solutions to UC problems were obtained using EE techniques [4,5]. Although EE-based approaches provide accurate optimal solutions, they do not apply to large-scale systems because of the dimensionality curse problem.

In PL methods, generators are sorted in increasing production costs. This prespecified assortment is then utilized to solve the UC problem so that the power demand is fulfilled. Refs. [6–9] solved the UC problem using PL-based approaches. Ref. [6] proposed a three-stage PL-based approach to solve a UC problem involving ramp rate constraints. Refs. [7,8] solved single-area and multi-area UC problems using PL approaches based on a classical index. Ref. [9] presented a PL-based approach to solving a UC problem which also includes import/export constraints. The advantage of the PL-based methods is the fast computational time; however, it suffers from producing sub-optimal generation schedules.

DP-based techniques are one of the earliest and the most widely used optimization methods for solving a UC problem [10–14]. The advantage of DP-based approaches is that they are easily modifiable to include specific features of utilities, and it is rather easier to include restrictions that affect hourly operations. However, the drawback is that it is more complicated to add time-dependent constraints and their substandard treatment. Moreover, solving a large-scale UC problem using DP can be time-consuming. Ref. [10] presented an adaptive DP-based algorithm to solve a UC problem involving coupled power and gas networks. In [11], Lowery et al. studied the feasibility of using DP for solving UC problems. They implemented dynamic programming for solving the UC problem for a 14-machine test system. Ref. [12] solved a multiple-area UC problem by employing a truncated DP and implementing it in a two-area Indian Power System. In [13], Forootani et al. proposed an approximate DP-based methodology using Markov decision processes to solve a stochastic UC problem. Ref. [14] discusses quadratic approximate DP-based approach for resource scheduling using Uruguayan power system.

LR method is another widely used optimization technique used by the utilities for UC problems. Its use in the utilities for solving the UC is more modern than the DP methods. Like DP, it is easily modifiable to include specific characteristics of utilities. However, it also has the disadvantage of having inherent suboptimality. In LR-based methods, the UC problem is solved by ignoring the coupling constraints as if they do not exist. In [15], Merlin et al. developed a novel technique for the UC problem using the LR method and tested it on Electricité De France. Ref. [16] proposed a new three-stage LR algorithm for UC. The first involves maximization of the Lagrangian dual of the UC problem. The second stage consists of finding a reserve feasible dual solution, and economic dispatch is performed.
in the third stage. In [17], Takriti et al. developed a refined LR algorithm using integer programming. There is a lot of research being conducted on methods using LR methods and its refinements using hybrid approaches for solving UC problems.

EP-based approaches are established on the rules of evolution [18–21]. In such approaches, a population of potential solutions is maintained based on the fitness score of individuals. Ref. [18] implemented the genetic algorithm (GA) to the UC problem of time span involving a day up to a week. The feasibility of GAs for solving the UC problems has been explored for smaller-size as well as bigger-size systems [19]. Ref. [20] developed a new UC approach using a GA with mutation operators specific to a domain. The robustness of the presented approach is illustrated by comparison with an LR-based UC approach. In [21], Chen et al. presented a hybrid PL-GA algorithm to solve a UC problem involving wind farms and studied it performance on two test systems having 10 and 20 units.

In the 90s and early 2000s, there was some interest among researchers in using neural networks (NNET) for solving the UC problem, but because of the processing constraints at the time, it proved very challenging to reach high-mark solutions [22–24]. Lately, several research studies have been conducted focusing on the use of machine learning (ML) for optimization problems; however, their focus has been on the reduction of the number of constraints to enhance the speed of the MIP engines rather than switching them with the learning-based approaches [25].

There has also been a lot of research work on solving the unit commitment problem using decentralized approaches. Ref. [26] presented a decentralized SCUC approach to accelerate the scheduling of large-size systems. In this method, a power system is decoupled into multiple inter-connected areas using tie-lines, and each area is responsible to solve its reduced SCUC problem in a parallel fashion. In [27–29], a decentralized approach is proposed for solving a reliability SCUC which consists of two stages. A power system is broken down into multiple areas, and every area solves its SCUC problem.

3. Unit Commitment Problem Formulation

Even though the UC problem involves multiple intervals, for simplicity and to emphasize the efficacy of the developed approach, it is defined here as a single time interval MIQP problem as discussed in several other research studies [30,31]. The fundamental distinction is the presence of time-dependent constraints; however, the developed method is identically applicable to an SCUC problem involving multiple time periods.

Let \(N_B, N_G, N_L\) represent respectively the number of buses, generators and transmission lines of the power network. The sets \(B = \{1, 2, 3, \ldots , N_B\}\), \(G = \{1, 2, 3, \ldots , N_G\}\) and \(L = \{1, 2, 3, \ldots , N_L\}\) are the sets of buses, generators and lines respectively. For every generator \(g \in G\), let \(I_g \in \{0, 1\}\) represent the commitment decision variable and \(P_g\), a continuous variable, represent the output power being generated by \(g\). A single-period SCUC problem can be formulated as an MIQP problem as presented below.

3.1. Objective Function

The goal of a UC problem is to minimize the total operating expense of all the generators. Let \(F_g\) indicate a generator’s cost function. Assuming a quadratic cost of power production (i.e., \(F_g = a_g I_g + b_g P_g + c_g P_g^2\)) and neglecting the start-up costs, the objective function can be formulated as:

\[
\min_{I_g, P_g} \sum_{g \in G} \left( a_g I_g + b_g P_g + c_g P_g^2 \right)
\]  

(1)

Here \(I_g, P_g, a_g, b_g\) and \(c_g\) represent respectively the commitment status, the production levels, and the cost coefficients of generator \(g\). If a generator \(g\) is committed, its no-load cost \((a_g I_g)\) and power production cost \((b_g P_g + c_g P_g^2)\) will be included in the objective function to be minimized.
3.2. The Constraints

There are several operational as well as security requirements that need to be met while minimizing operational costs. Some of the commonly used constraints in formulating a single-period SCUC are discussed here.

3.2.1. Nodal Power Balance

At each bus \( b \in B \), the nodal power balance (i.e., \( \sum_{g \in G_b} P_g - \sum_{l \in L_b} P_l - d_b = 0 \)) must be satisfied. While formulating a UC problem as a mixed-integer programming (MIP) problem, the line flows are usually approximated by DC power flow assumptions in the literature. Hence, the nodal power balance constraints are expressed as follows.

\[
\sum_{g \in G_b} P_g = d_b + \sum_{l \in L_b} P_l \quad \forall b \in B
\]  

Here \( d_b \) represents the estimated demand at node \( b \), \( P_l \) represents the active power being injected into line \( l \) from node \( b \), and \( L_b \) and \( G_b \) respectively represent the sets of transmission lines and generators connected to node \( b \). There will be \( N_B \) power balance constraints, each representing the DC power balance at node \( b \).

3.2.2. Generator Power Limits

These constraints ensure that, if committed, the power generated by a generator \( g \), does not exceed its maximum (\( P_g^\text{max} \)) and minimum (\( P_g^\text{min} \)) power generation limits.

\[
P_g^\text{min} I_g \leq P_g \leq P_g^\text{max} I_g \quad \forall g \in G
\]  

\[
P_g + R_g \leq \left( P_g^\text{max} - P_g \right) I_g \quad \forall g \in G
\]

Here \( R_g \) is the spinning reserve power provided by a generator \( g \).

3.2.3. Thermal Limits of Transmission Lines

Because of the thermal limits of lines, there is a power limit that each line can carry. So, to ensure that the line flow (\( P_l \)) of line \( l \) does not surpass its maximum power carrying capacity (\( P_l^\text{max} \)), the constraints in (5) must be satisfied. The line flows are approximated using DC power flow assumptions. \( x_l \) is the reactance of line \( l \) and \( \delta_l \) is the voltage angle difference of the two busses between which line \( l \) is connected.

\[
- P_l^\text{max} \leq P_l = \frac{\delta_l}{x_l} \leq P_l^\text{max} \quad \forall l \in L
\]

3.2.4. Spinning Reserve Requirements

The spinning reserves are described as the online generation capacity minus the total generated power. The main purpose of the spinning reserve is to cope with generation losses or unexpected load changes. The spinning reserves must be higher than or equal to the minimum reserve requirements (\( R \)), also called system-wide reserves. \( R \) is usually described as a percentage of the total demand.

\[
\sum_{g \in G} R_g \geq R
\]

3.2.5. Binary and Non-Negative Variables

The UC problem is a mixed-integer non-convex optimization problem. It has both binary (\( I_g \)) and continuous (\( P_g \)) decision variables which are usually non-negative.

\[
I_g \in \{0, 1\} \quad \forall g \in G
\]

\[
P_g \geq 0 \quad \forall g \in G
\]
4. The Proposed Approach

The proposed data-driven approach for solving a SCUC problem contains three layers which are described below.

4.1. Layer-1: Prediction of Commitment Decision Variables

In the first layer of the proposed multi-layer approach, several machine learning classification models are trained by employing supervised machine learning approaches for predicting the commitment status of the units. During the training phase, either historic or synthetic data can be used. The data points in the training dataset represent the solved cases of the UC problem. There are a lot of machine learning approaches that can be employed for model training and the selection of an approach depends on its prediction performance. An approach working well for one application may not perform well for another. Hence, the common practice is to train multiple models and choose the best-performing one, as shown in Figure 4.

For the first layer prediction, multiple models are trained, such as “K-Nearest Neighbor” (KNN), “Generalized Linear Model” (GLM), “Linear Discriminant Analysis” (LDA), “Support Vector Machine” (SVM), “Decision Tree” (DT), “Random Forest” (RF), “Extreme Gradient Boosting” (XGB) and “Neural Network” (NNET). For model evaluation, the original dataset is divided into disjoint train and test datasets. For initial selection during the training process for hyperparameter tuning, 10-fold repeated cross-validation is utilized. The final selection is performed on the basis of the test set classification accuracy, which is defined in (9).

\[
\text{Accuracy} = \frac{\text{no. of correct predictions}}{\text{total no. of predictions}} = \frac{TP + TN}{P + N} \tag{9}
\]

Here TP, TN, FP and FN respectively indicate “true-positives”, “true-negatives”, “false-positives” and “false-negatives” as defined in Figure 5.

\[
\begin{array}{ccc|cc}
\text{Actual Class} & & \text{Predicted Class} & 1 & 0 \\
1 & True Positives (TP) & False Negatives (FN) \\
0 & False Positives (FP) & True Negatives (TN) \\
\end{array}
\]

4.2. Layer-2: Prediction of Production Decision Variables

The predicted commitment decision variables obtained from the layer-1 prediction are then used to predict the production levels \( P_g \) of the generators using trained ML regression
models. Similar to the layer-1 approach, several models are trained, such as KNN, GLM, DT, RF, XGB and NNET, and the best-performing model is used for the final prediction. The performance evaluation of layer-2 models is similar to layer-1, but the indices used to quantify the efficiency are different. The commonly used metrics are “mean-absolute-error” (MAE), “root-mean-square-error” (RMSE), “mean-absolute-percentage-error” (MAPE) and “coefficient-of-determination” (R-squared). We used MAE for the best model selection based on the test set performance, which is defined in (10).

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |P_{Gi} - \hat{P}_{Gi}|
\]  

(10)

Here \(P_{Gi}\) and \(\hat{P}_{Gi}\) are respectively the actual and the predicted production of a generator \(g\) for test set instance \(i\) and \(n\) is the size of the test dataset.

4.3. Layer-3: Feasibility Compliance

Since layer-1 and layer-2 models are trained on the data which consists of solved feasible solutions, they implicitly predict solutions that would somewhat fulfill the constraints requirements. However, because of the inherent uncertainty in predictive modeling techniques, it is required from the proposed approach to ensure that every constraint is realized for the safe functioning of the power system. Therefore, a mandatory third layer for feasibility compliance is added to attain this objective.

4.3.1. Spinning Reserve Requirements

To satisfy the spinning reserve constraints, it is confirmed if the total committed generation capacity \(\left(\bar{P}_{G}^{c}\right)\) is higher than total demand \(D\) plus spinning reserve requirement \(R\) as given by (11)–(13). If it is not satisfied, the nodal demands are assumed to have increased by a small percentage, and the commitment decisions are predicted again to increase the predicted committed capacity. Here \(G^c\) represents the set of committed generators.

\[
\bar{P}_{G}^{c} = \sum_{g \in G^c} P_{g}
\]  

(11)

\[
D = \sum_{b \in B} d_{b}
\]  

(12)

\[
\bar{P}_{G}^{c} \geq D + R
\]  

(13)

4.3.2. Generator Power Limits

The predicted production decisions \((P_{g})\), obtained from layer-2 prediction, are truncated by \(\bar{P}_{g}\) and \(\underline{P}_{g}\) to satisfy the upper and lower power limits of generators as follows.

\[
P_{g} \in [\underline{P}_{g}, \bar{P}_{g}] \quad \forall g \in G^c
\]  

(14)

4.3.3. Total Power Balance

To balance the total generation \((P_{G})\) with the total demand \((D)\), the generation load imbalance \((GLI)\) is calculated, and the production decisions are updated using (16) while satisfying the upper and lower power limits of generators.

\[
GLI = P_{G} - D = \sum_{g \in G^c} P_{g} - \sum_{b \in B} d_{b}
\]  

(15)

\[
P'_{g} = P_{g} \left(1 - \frac{GLI}{P_{G}}\right) = P_{g} \left(\frac{D}{P_{G}}\right) \in [\underline{P}_{g}, \bar{P}_{g}] \quad \forall g \in G^c
\]  

(16)

The decision variables \((P_{g})\) which violate \(\underline{P}_{g}\) or \(\bar{P}_{g}\) limits are truncated at their limits, and the remaining \(P_{g}\) variables are updated again using (15)–(16).
4.3.4. Thermal Limits of Transmission Lines

To ensure that none of the line capacity limits are violated, the generator powers \( P_g \) are re-dispatched to ensure that all the line flows are within their capacity limits. For rescheduling the generator powers, small variations \( \Delta P_g \) are introduced in \( P_g \) variables using Generation Shift Factors (GSF) and Eigen Value Decomposition (EVD) discussed below.

Generation Shift Factors (GSF)

The effect of nodal/bus power injections on the line flows can be studied using linear sensitivity factors called Generation Shift Factors. They indicate how much power being injected at bus \( b \) will appear on line \( l \). Using the DC load flow assumptions, the effect of varying the nodal injections \( \Delta P \) on the node voltage angles \( \Delta \delta \) is given by (17).

\[
\Delta \delta = \left[ B' \right]^{-1} \Delta P \tag{17}
\]

\( B' \) is the DC load flow admittance matrix. In a similar manner, the effect of varying the node voltage angles \( \Delta \delta \) on the line flows \( \Delta P_{line} \) is given by (18).

\[
\Delta P_{line} = (D_X A_N) \Delta \delta \tag{18}
\]

\( D_X \) is a square diagonal matrix with a diagonal element \( D_X(i,i) = -1/X_l \). Here \( X_l \) is the inductive reactance of the \( l \)th transmission line, and matrix \( A_N \) is the node-arc-incidence matrix which indicates how lines are connected in the network. Using (17) and (18), we can find the effect of varying nodal injections on the power flow in lines.

\[
\Delta P_{line} = T \Delta P \tag{19}
\]

Here \( T \) is the matrix of GSFs given by (20), and \( \Delta P_{line}, \Delta \delta \) and \( \Delta P \) are vectors of change in line flows, node voltage angles and nodal injections, respectively.

\[
T = (D_X A_N) \left[ B' \right]^{-1} \tag{20}
\]

\[
\Delta \delta = [\Delta \delta_1 \ldots \Delta \delta_{NB}]^T \tag{21}
\]

\[
\Delta P = [\Delta P_1 \ldots \Delta P_{NB}]^T \tag{22}
\]

\[
\Delta P_{line} = [\Delta P_{line1} \ldots \Delta P_{lineNL}]^T \tag{23}
\]

Eigenvalue Decomposition (EVD)

A vector \( x \) varies its magnitude and its direction when operated on a system matrix \( A \) as \( y = Ax \). However, when eigenvectors \( (v_{\lambda_i}) \) of the matrix \( A \) are multiplied with it, they only change their magnitudes while keeping the same direction, as given by (24).

\[
A v_{\lambda_i} = \lambda_i v_{\lambda_i} \tag{24}
\]

Figure 6 shows the effect of multiplying all vectors \( \{x_i\} \) of equal lengths with \( A \). It can be seen that when \( x \) lies in the span of the eigenvector \( v_m \), an eigenvector of the highest eigen-value \( \lambda_m \), then \( y \) is maximized, i.e., if \( x = \alpha v_m \), then \( y = \alpha A v_m = \alpha \lambda_m v_m \). Therefore, to have maximum effect of \( \Delta P \) on a specific line \( l \), weighted L2 norm of \( \Delta P_{line} \) can be utilized. Using (19), it can be deduced that:

\[
\Delta P_l = \Delta P_{line}^T W \Delta P_{line} = \Delta P^T N \Delta P \tag{25}
\]

Here, \( N = T^T W \) is a real symmetric matrix, and \( W \) is a selection matrix to select the lines whose flows are to be controlled. As mentioned, \( \Delta P_l \) would have maximum change...
if $\Delta P$ lies in the span of the eigen-vector $v_m$ of $N$, i.e., if $\Delta P = \alpha v_m$, then $\Delta P_{l}^{\text{max}} = \alpha^2 \lambda_m$. Scalar $\alpha$ is a design parameter. It gives the desired variation in a line flow and can be calculated as $\alpha = \pm \sqrt{\Delta P_{\text{req}} / \lambda_m}$ [32,33]. Since the system’s load and total generation must be balanced in the obtained solution, the condition $\sum \Delta P_b = 0$ must also be satisfied. Therefore, only the signs of the elements of $v_m$ are used to get a sense of the required change in generators power (i.e., increase or decrease $P_g$) to maximize the effect $\Delta P$ on $\Delta P_l$. The flowchart of the proposed three-layer algorithm is presented in Figure 7.

![Figure 6. Manipulation of $\{x_i\}$ with $A$.](image)

![Figure 7. Three Layer Proposed SCUC Approach.](image)
5. Case Studies and Simulation Results

To evaluate the efficacy of the developed algorithm, we will study three power systems, i.e., a 4-bus 3-machine system and the IEEE 39-bus 10-machine and 118-bus 54-machine systems. Application of the developed approach on a large-size system is under study. All the simulations are performed using MATLAB and R-Studio on an AMD 16-cores machine having 64 GB of DDR4 RAM.

5.1. Case Study: 4-Bus System

This small system is utilized as an explanatory example to investigate the working of the presented method. The system has four buses, three generators, two loads and five transmission lines. The system diagram is shown in Figure 8. Further information on the system is given in Table 1.

![Figure 8. System Diagram (4-Bus System).](image)

| Parameter                              | Value   |
|----------------------------------------|---------|
| Common Voltage Base (kV)              | 138     |
| Common Power Base (MVA)               | 100     |
| Lines Reactances (PU)                 | 0.25    |
| Lines Capacities (MW)                 | 200     |
| Generators Power Limits (MW)          | 0–500   |
| mean MW Loads (bus2/bus3)             | 300/200 |
| Spinning Reserve Requirements (R)     | 30% of Total Power Demand |

For model training, the dataset is generated by solving the MIQP problem presented in Section 4 using the IBM CPLEX toolbox for MATLAB. A dataset of 1000 samples, representing different scenarios of power demands and price signals, is generated assuming normal distributions for the nodal demand \((d_2, d_4)\) and the cost coefficients \((a_g, b_g, c_g)\). The spread of the cost curve resulting from these cost coefficients is given in Figure 9. Figure 10 displays the histograms of the nodal demands and the cost coefficients in the generated synthetic dataset.

The generated dataset is divided into an 80–20% train-test split using stratified random sampling to preserve the distributions of the classes in the train and the test sets. The initial selection of the top-performing models is conducted based on their performance using 10-fold cross-validation (10-CV) repeated five times. This 10-CV is also used for hyperparameter tuning. The selected models are further evaluated using the test dataset, and the best-performing model is selected.
Table 1. System Data (4-Bus System).

| Parameter          | Value |
|--------------------|-------|
| Common Voltage Base (kV) | 138   |
| Common Power Base (MVA)    | 100   |
| Lines Reactances (PU)   | 0.25  |
| Lines Capacities (MW)   | 200   |

For model training, the dataset is generated by solving the MIQP problem presented in Section 4 using the IBM CPLEX toolbox for MATLAB. A dataset of 1000 samples, representing different scenarios of power demands and price signals, is generated assuming normal distributions for the nodal demand ($d_2$, $d_3$) and the cost coefficients ($a_1$, $b_1$, $c_1$). The spread of the cost curve resulting from these cost coefficients is given in Figure 9.

Figure 9. Spread of Generators’ Cost Curves (4-Bus System).

Figure 10. Distributions of nodal demands and cost coefficients (4-Bus System).

5.1.1. Layer-1: Prediction of Commitment Decision Variables

Table 2 shows the models’ performance during training. TPR and TNR represent the “true positive” and “true negative” rates, respectively. Figure 11 shows the boxplot for the training accuracy score to show the differences among these trained models. The test set performance of the initially selected models is given in Table 3. Based on the test set performance, NNET was selected for layer-1 prediction.

Table 2. Layer-1 Training Performance (4-Bus System).

| Model               | Mean TPR | Mean TNR | Accuracy (95% CI) |
|---------------------|----------|----------|-------------------|
| LDA                 | 0.70     | 0.97     | 0.86 (0.84, 0.89) |
| KNN                 | 0.48     | 0.97     | 0.82 (0.79, 0.85) |
| DT                  | 0.33     | 0.94     | 0.69 (0.66, 0.71) |
| SVM-Radial          | 0.55     | 0.92     | 0.58 (0.55, 0.61) |
| SVM-Poly            | 0.59     | 0.93     | 0.59 (0.55, 0.62) |
| RF                  | 0.50     | 0.97     | 0.83 (0.81, 0.84) |
| NNET                | 0.76     | 0.98     | 0.89 (0.86, 0.91) |
| XGB                 | 0.58     | 0.97     | 0.83 (0.80, 0.86) |

Figure 10. Distributions of nodal demands and cost coefficients (4-Bus System).

5.1.1. Layer-1: Prediction of Commitment Decision Variables

Table 2 shows the models’ performance during training. TPR and TNR represent the “true positive” and “true negative” rates, respectively. Figure 11 shows the boxplot for the training accuracy score to show the differences among these trained models. The test set performance of the initially selected models is given in Table 3. Based on the test set performance, NNET was selected for layer-1 prediction.
Table 2. Layer-1 Training Performance (4-Bus System).

| Model       | Mean TPR | Mean TNR | Accuracy (95% CI) |
|-------------|----------|----------|-------------------|
| LDA         | 0.70     | 0.97     | 0.86 (0.84, 0.89) |
| KNN         | 0.48     | 0.97     | 0.82 (0.79, 0.85) |
| DT          | 0.33     | 0.94     | 0.69 (0.66, 0.71) |
| SVM-Radial  | 0.55     | 0.92     | 0.58 (0.55, 0.61) |
| SVM-Poly    | 0.59     | 0.93     | 0.59 (0.55, 0.62) |
| RF          | 0.50     | 0.97     | 0.83 (0.81, 0.84) |
| NNET        | 0.76     | 0.98     | 0.89 (0.86, 0.91) |
| XGB         | 0.58     | 0.97     | 0.83 (0.80, 0.86) |

Figure 11. Layer-1 Training Performance Comparison (4-Bus System).

Table 3. Layer-1 Models Performance (4-Bus System).

| Model | Accuracy |
|-------|----------|
| Train | Test     |
| LDA   | 0.86     | 0.85     |
| KNN   | 0.82     | 0.83     |
| RF    | 0.83     | 0.83     |
| NNET  | 0.89     | 0.89     |
| XGB   | 0.83     | 0.83     |

5.1.2. Layer-2: Prediction of Production Decision Variables

Similar to layer 1, several models are trained for predicting the production levels of the committed generators. These models include GLM, KNN, RF, SVM, NNET and XGB. Table 4 shows the performance of the models during the training and testing phases.
Table 4. Layer-2 Models Performance (4-Bus System).

| Model       | MAE | R²  |
|-------------|-----|-----|
| GLM         | 37.3| 0.81|
| KNN         | 56.5| 0.60|
| DT          | 58.9| 0.54|
| SVM-Radial  | 17.2| 0.95|
| SVM-Poly    | 81.9| 0.40|
| RF          | 39.4| 0.80|
| NNET        | 15.3| 0.96|
| XGB-Linear  | 37.5| 0.80|

| Model       | MAE  | R²  |
|-------------|------|-----|
| SVM-Radial  | 11.3 | 0.97|
| NNET        | 10.1 | 0.98|

5.1.3. Layer-3: Feasibility Compliance

In this layer, the solutions obtained from the first two layers are further processed to ensure all constraints are satisfied. Table 5 shows the percentage of the infeasible instances (out of 200 total instances in the test set) with and without layer-3.

Table 5. Layer-3 Performance (4-Bus System).

| Violation                  | Percentage of Infeasible Samples (Test Set) |
|----------------------------|---------------------------------------------|
|                            | Without Layer-3 | With Layer-3 |
| Generator Limits           | 0%             | 0%           |
| Reserve Requirement        | 1%             | 0%           |
| GLI > 1% of Total Demand   | 75.5%          | 0%           |
| Line Limits                | 7.5%           | 0%           |

5.2. Case Study: IEEE 39-Bus System

The system has thirty-nine buses, forty-six transmission lines, eighteen loads and ten generators. The spinning reserve requirements were taken to be 10% of the total demand. Further details of the system can be obtained from the MATPOWER. Figure 12 shows the system diagram.

A dataset of 1000 samples is created using normal distribution functions for the cost coefficients and the nodal loads. The distributions of the nodal demands in the generated dataset are shown in Figure 13. For training and evaluation, the data is divided with an 80–20% train-test split using stratified random sampling. The training and test set performances of the trained models for layer-1 and layer-2 predictions are shown in Tables 6 and 7, respectively. Based on the performance, GLM was selected for predicting the commitment decision variables and NNET was selected for layer 2 prediction. The effect of layer-3 on the percentage of infeasible test samples is evident in Table 8.

Table 6. Layer-1 Models Performance (IEEE 39-Bus System).

| Model       | Train Accuracy | Test Accuracy |
|-------------|----------------|---------------|
| GLM         | 92.6%          | 91.4%         |
| RF          | 92.8%          | 88.3%         |
| SVM-Poly    | 93.3%          | 89.2%         |
| SVM-Radial  | 93.1%          | 88.1%         |
| NNET        | 93.8%          | 90.1%         |
| XGB         | 94.1%          | 90.7%         |
Table 5. Layer-3 Performance (4-Bus System).

| Violation Percentage of Infeasible Samples (Test Set) | Without Layer-3 | With Layer-3 |
|------------------------------------------------------|-----------------|--------------|
| Generator Limits                                     | 0%              | 0%           |
| Reserve Requirement                                  | 1%              | 0%           |
| GLI > 1% of Total Demand                             | 75.5%           | 0%           |
| Line Limits                                          | 7.5%            | 0%           |

5.2. Case Study: IEEE 39-Bus System

The system has thirty-nine buses, forty-six transmission lines, eighteen loads and ten generators. The spinning reserve requirements were taken to be 10% of the total demand. Further details of the system can be obtained from the MATPOWER. Figure 12 shows the system diagram.

Figure 12. System Diagram (IEEE 39-Bus System) with Bus Number (black), Generator Number (black encircled), Line Number (red), Line MW Capacity (yellow) and Generators MW Capacity (gray).

A dataset of 1000 samples is created using normal distribution functions for the cost coefficients and the nodal loads. The distributions of the nodal demands in the generated dataset are shown in Figure 13. For training and evaluation, the data is divided with an 80–20% train-test split using stratified random sampling. The training and test set performances of the trained models for layer-1 and layer-2 predictions are shown in Tables 6 and 7, respectively. Based on the performance, GLM was selected for predicting the commitment decision variables and NNET was selected for layer 2 prediction. The effect of layer-3 on the percentage of infeasible test samples is evident in Table 8.

5.3. Case Study: IEEE 118-Bus System

The system has 118 buses, 99 loads, 54 generators and 186 transmission lines. The reserve requirements were taken to be 10% of the total demand. Further details of the system can be obtained from the MATPOWER. The system diagram is given in Figure 14.

Figure 13. Nodal Demands Distributions (IEEE 39-Bus System).

Table 7. Layer-2 Models Performance (IEEE 39-Bus System).

| Model    | Train MAE | Test MAE |
|----------|-----------|----------|
| GLM      | 42.20     | 53.00    |
| RF       | 49.70     | 56.98    |
| SVM-Poly | 41.25     | 51.90    |
| SVM-Radial | 43.48    | 51.66    |
| NNET     | 39.32     | 51.73    |
| XGB      | 48.98     | 56.68    |
Table 8. Layer-3 Performance (IEEE 39-Bus System).

| Violation                              | Percentage of Infeasible Samples (Test Set) |
|---------------------------------------|--------------------------------------------|
|                                       | Without Layer-3 | With Layer-3 |
| Generator Limits                      | 0%             | 0%           |
| Reserve Requirement                   | 13.5%          | 0%           |
| GLI > 1% of Total Demand              | 90.5%          | 0%           |
| Line Limits                           | 37.5%          | 0%           |

5.3. Case Study: IEEE 118-Bus System

The system has 118 buses, 99 loads, 54 generators and 186 transmission lines. The reserve requirements were taken to be 10% of the total demand. Further details of the system can be obtained from the MATPOWER. The system diagram is given in Figure 14.

A dataset of 3500 instances is generated using normal probability distribution functions for the cost coefficients and the nodal loads. Figure 15 shows the nodal demand distributions of some of the buses. For training and evaluation, the data is divided with an 80–20% train-test split using stratified random sampling. The train set and the test set performances of the trained models for layer-1 and layer-2 predictions are shown in Tables 9 and 10 respectively. Based on the performance, GLM was selected for layer-1 as well as layer-2 predictions. The effect of layer-3 on the percentage of infeasible test samples is evident in Table 11.
A dataset of 3500 instances is generated using normal probability distribution functions for the cost coefficients and the nodal loads. Figure 15 shows the nodal demand distributions of some of the buses. For training and evaluation, the data is divided with an 80–20% train-test split using stratified random sampling. The train set and the test set performances of the trained models for layer-1 and layer-2 predictions are shown in Tables 9 and 10 respectively. Based on the performance, GLM was selected for layer-1 as well as layer-2 predictions. The effect of layer-3 on the percentage of infeasible test samples is evident in Table 11.

Table 9. Layer-1 Models Performance (IEEE 118-Bus System).

| Model    | Train Accuracy | Test Accuracy |
|----------|----------------|---------------|
| GLM      | 94.9%          | 94.9%         |
| RF       | 92.9%          | 92.7%         |
| SVM-Radial | 90.8%   | 91.0%         |
| LDA      | 93.8%          | 94.1%         |
| KNN      | 68.0%          | 68.1%         |
| NNET     | 92.6%          | 92.8%         |
| DT       | 92.1%;        | 91.9%         |

Table 10. Layer-2 Models Performance (IEEE 118-Bus System).

| Model    | Train MAE | Test MAE |
|----------|-----------|----------|
| GLM      | 14.97     | 18.03    |
| SVM-Radial | 16.28    | 19.42    |
| KNN      | 29.86     | 29.12    |
| NNET     | 27.36     | 26.89    |
| DT       | 29.86     | 29.12    |

Table 11. Layer-3 Performance (IEEE 118-Bus System).

| Violation | Percentage of Infeasible Samples (Test Set) |
|-----------|---------------------------------------------|
| Generator Limits | 0%            | 0%            |
| Reserve Requirement | 25.0%        | 0%            |
| GLI > 1% of Total Demand | 83.9%        | 0%            |
| Line Limits       | 51.0%            | 0%            |
6. Performance Comparison
6.1. Cost Comparison

First, the cost comparison is performed between the solutions obtained from the proposed approach and the conventional MIQP solver using IBM CPLEX MATLAB Toolbox. Table 12 presents the cost comparison for the test set for the three case studies. The table data shows that the developed algorithm obtained solutions with an optimality gap of less than 1% on average which can be considered a satisfactory performance in terms of cost. However, the primary advantage of the proposed approach is the fast computational speed which is evaluated next.

Table 12. Cost Comparison.

| Case Study | Approach       | Avg Cost ($/h) | Cost Difference (95% CI) |
|------------|----------------|----------------|--------------------------|
| 4-Bus      | Proposed       | 5424           | +0.35% (0.17, 0.49)      |
|            | CPLEX MIQP    | 5405           | -                        |
| IEEE 39-Bus| Proposed       | 44069          | +0.32% (0.23, 0.41)      |
|            | CPLEX MIQP    | 43931          | -                        |
| IEEE 118-Bus| Proposed      | 70613          | +0.68% (0.50, 0.89)      |
|            | CPLEX MIQP    | 70133          | -                        |

6.2. Time Comparison

Next, the computational time comparison is performed. From the test dataset, 100 instances are solved, and the solution time taken by an approach is observed. This is repeated several times and the average time to solve 1000 instances is calculated. Figure 16 shows the average time taken by each approach to solve 1000 instances. It is evident from the figure data that the developed algorithm provides significant time savings (96%, 94% and 81% reductions in time for the 4-bus and the IEEE 39-bus and 118-bus systems respectively) compared to the CPLEX MIQP solver.

![Figure 16. Computational Time Comparison.](image)

7. Conclusions and Future Work

This paper proposes a novel three-layer hierarchical learning-based algorithm for solving a security constrained UC problem. In the first layer, a set of data-driven ML models are utilized to predict the binary commitment variables. In the second layer, another set of predictive models predicts the production levels of the committed generators. Finally, in the third layer, a feasibility compliance algorithm post-processes the predicted schedule to ensure the feasibility of the obtained solution. The efficacy of the developed
method was demonstrated through numerical simulations using a 4-bus 3-machine system and two standard IEEE power systems, i.e., 39-bus 10-machine and 118-bus 54-machine power systems. Simulation results demonstrate that the proposed approach provided significant time savings (90% time-reduction on average for the three studied systems) in solving a single-period SCUC problem with minimal degradation (<1%) to the optimality index. Moreover, it also ensured that all the predicted solutions are feasible solutions (0% infeasible solutions) for the secure operation of the power system.

The significance of the proposed approach would be more eminent when we exploit its capability of learning patterns from the big data and study its performance for a realistic large-size SCUC problem. Future work will be to solve a UC problem with multiple time intervals and/or take into consideration renewable energy sources, e.g., battery storage, wind and solar energy.

Author Contributions: Conceptualization, A.F., T.I. and H.U.B.; methodology, A.F. and T.I.; software, A.F., T.I. and H.U.B.; validation, A.F. and T.I.; formal analysis, A.F., T.I. and H.U.B.; investigation, A.F., T.I. and M.C.; resources, A.F. and T.I.; data generation, A.F. and T.I.; writing-original draft preparation, A.F. and T.I.; writing-review and editing, A.F., T.I., H.U.B. and M.C.; visualization, A.F. and T.I.; supervision, A.F. and M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available because it was generated synthetically during the study using MATPOWER 7.1 and MATLAB R2021a.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Electric Power Markets. Available online: https://www.ferc.gov/electric-power-markets (accessed on 9 September 2022).
2. U.S. Department of Energy, Electric Power Annual 2020, Energy Information Administration (EIA), March 2022. Available online: https://www.eia.gov/electricity/annual/pdf/epa.pdf (accessed on 12 May 2022).
3. Xavier, A.S.; Qiu, F.; Ahmed, S. Learning to Solve Large-Scale Security-Constrained Unit Commitment Problems. INFORMS J. Comput. 2020, 33, 739–756. [CrossRef]
4. Hara, K.; Kimura, M.; Honda, N. A method for planning economic unit commitment and maintenance of thermal power systems. IEEE Trans. Power App. Syst. 1966, PAS-85, 427–436. [CrossRef]
5. Kerr, R.H.; Scheidt, J.L.; Fontana, A.J.; Wiley, J.K. Unit commitment. IEEE Trans. Power App. Syst. 1966, PAS-85, 417–421. [CrossRef]
6. Elsayed, A.M.; Maklad, A.M.; Farrag, S.M. Three-Stage Priority List Unit Commitment for Large-Scale Power Systems Considering Ramp Rate Constraints. IEEE Can. J. Electr. Comput. Eng. 2021, 44, 329–342. [CrossRef]
7. Lee, F.N. The application of commitment utilization factor (CUF) to thermal unit commitment. IEEE Trans. Power Syst. 1991, 6, 691–698. [CrossRef]
8. Lee, F.N.; Feng, Q. Multi-area unit commitment. IEEE Trans. Power Syst. 1992, 7, 591–599. [CrossRef]
9. Shoultz, R.R.; Chang, S.K.; Helmick, S.; Grady, W.M. A practical approach to unit commitment, economic dispatch and savings allocation for multiple-area pool operation with import/export constraints. IEEE Trans. Power App. Syst. 1980, PAS-99, 625–635. [CrossRef]
10. Xu, Y.; Ding, T.; Qu, M.; Du, P. Adaptive Dynamic Programming for Gas-Power Network Constrained Unit Commitment to Accommodate Renewable Energy with Combined-Cycle Units. IEEE Trans. Sustain. Energy 2020, 11, 2028–2039. [CrossRef]
11. Lowery, P.G. Generating unit commitment by dynamic programming. IEEE Trans. Power App. Syst. 1966; PAS-85, 422–426.
12. Sen, S.; Kothari, D.P. Evaluation of benefit of inter-area energy exchange of the Indian power systems based on multi-area unit commitment approach. Elect. Mach. Power Syst. 1998, 26, 801–813. [CrossRef]
13. Forootani, A.; Iervolino, R.; Tipaldi, M.; Neilson, J. Approximate dynamic programming for stochastic resource allocation problems. IEEE/CAA J. Auton. Sin. 2020, 7, 975–990. [CrossRef]
14. Castellano, A.; Martinez, C.; Monzón, P.; Bazerque, J.A.; Ferragut, A.; Paganini, F. Quadratic approximate dynamic programming for scheduling water resources: A case study. In Proceedings of the IEEE PES Transmission & Distribution Conference and Exhibition-Latin America (T&D LA), Montevideo, Uruguay, 28 September–2 October 2020; pp. 1–6.
15. Merlin, A.; Sandrin, P. A new method for unit commitment at Electricité De France. IEEE Trans. Power App. Syst. 1983, PAS-102, 1218–1225. [CrossRef]
16. Zhuang, F.; Galiana, F.D. Toward a more rigorous and practical unit commitment by Lagrangian Relaxation. IEEE Trans. Power Syst. 1988, 3, 763–773. [CrossRef]
17. Takriti, S.; Birge, J.R. Using integer programming to refine Lagrangian-based unit commitment solutions. *IEEE Trans. Power Syst.* 2000, 15, 151–156. [CrossRef]

18. Sheblé, G.B.; Maifeld, T.T.; Brittig, K.; Fahd, G.; Fukurozaki-Coppinger, S. Unit commitment by genetic algorithm with penalty methods and a comparison of Lagrangian search and genetic algorithm economic dispatch example. *Int. J. Electr. Power Energy Syst.* 1996, 18, 339–346. [CrossRef]

19. Dasgupta, D.; McGregor, D.R. Thermal unit commitment using genetic algorithms. *Proc. Inst. Elect. Eng. Gen. Trans. Dist.* 1994, 141, 459–465. [CrossRef]

20. Maifeld, T.T.; Sheble, G.B. Genetic-based unit commitment algorithm. *IEEE Trans. Power Syst.* 1996, 11, 1359–1370. [CrossRef]

21. Chen, H.; Shi, L. Stochastic unit commitment incorporating wind power using improved genetic algorithm. In Proceedings of the 10th Renewable Power Generation Conference (RPG), Online Conference, 14–15 October 2021; pp. 910–917.

22. Walsh, M.P.; O’Malley, M.J. Augmented Hopfield network for unit commitment and economic dispatch. *IEEE Trans. Power Syst.* 1997, 12, 1765–1774. [CrossRef]

23. Liang, R.H.; Kang, F.C. Thermal generating unit commitment using an extended mean field annealing neural network. *IEE Proc. Gener. Transm. Distrib.* 2000, 147, 164–170. [CrossRef]

24. Bengio, Y.; Lodi, A.; Prouvost, A. Machine Learning for Combinatorial Optimization: A Methodological Tour d’Horizon. *Eur. J. Oper. Res.* 2021, 290, 405–421. [CrossRef]

25. Pineda, S.; Morales, J.M.; Jiménez-Cordero, A. Data-Driven Screening of Network Constraints for Unit Commitment. *IEEE Trans. Power Syst.* 2020, 35, 3695–3705. [CrossRef]

26. Kargarian, A.; Fu, Y.; Li, Z. Distributed Security-Constrained Unit Commitment for Large-Scale Power Systems. *IEEE Trans. Power Syst.* 2015, 30, 1925–1936. [CrossRef]

27. Malekshah, S.; Ansari, J. A novel decentralized method based on the system engineering concept for reliability-security constraint unit commitment in restructured power environment. *Int. J. Energy Res.* 2021, 45, 703–726. [CrossRef]

28. Malekshah, S.; Malekshah, Y.; Malekshah, A. A novel two-stage optimization method for the reliability-based security constraints unit commitment in presence of wind units. *Clean. Eng. Technol.* 2021, 4, 100237. [CrossRef]

29. Malekshah, S.; Banihashemi, F.; Daryabad, H.; Yavarishada, N.; Cuzner, R. A zonal optimization solution to reliability security constraint unit commitment with wind uncertainty. *Comput. Electr. Eng.* 2022, 99, 107750. [CrossRef]

30. Iqbal, T.; Banna, H.U.; Feliachi, A. AI-Driven Security Constrained Unit Commitment Using Eigen Decomposition and Linear Shift Factors. In Proceedings of the 2021 North American Power Symposium (NAPS), College Station, TX, USA, 14–16 November 2021.

31. Iqbal, T.; Banna, H.U.; Choudhry, M.; Feliachi, A. Solving Security Constrained Unit Commitment Problem Using Inductive Learning. In Proceedings of the 2022 IEEE Kansas Power and Energy Conference (KPEC), Manhattan, KS, USA, 25–26 April 2022.

32. Iqbal, T.; Feliachi, A. Discovering community structures in power system networks using voltage—Reactive power sensitivity. In Proceedings of the 2017 North American Power Symposium (NAPS), Morgantown, WV, USA, 17–19 September 2017; pp. 1–6.

33. Iqbal, T.; Feliachi, A. Decentralized Voltage Control Using Fast Community Detection Algorithm and Eigen Decomposition. In Proceedings of the 2019 IEEE Power & Energy Society General Meeting (PESGM), Atlanta, GA, USA, 4–8 August 2019; pp. 1–5.