# TFT-PPO: Proactive Power Quality Control for EV Fast-Charging Clusters via Transformer-Based Demand Forecasting and Reinforcement Learning

Kollatu Sravanthi
*Sr. Assistant Professor*
Vignan's Institute of Information Technology, Duvvada, Visakhapatnam, Andhra Pradesh-530049.
Kollatusravanthi2@gmail.com

Sandhya N
*Professor and HoD, Department of AIML*
Dayananda Sagar Academy of Technology and Management
Bangalore, Karnataka, India
dr.sandhya1909@gmail.com

Kevin Patel
*Independent Researcher*
Chicago, USA
ORCID - 0009-0008-7607-6913

R Rajeswari
*Department of Electronics & Communication Engineering (ECE)*
PERI Institute of Technology,
Chennai – 600048, Tamil Nadu
rajivarsha3@gmail.com

Ramakrishna
*Associate Professor*
Department of Electrical and Electronics Engineering,
NMAM Institute of Technology,
NITTE (Deemed to be University),
Karnataka, India

Bachina Harish Babu
*Assistant Professor*
Department of Automobile Engineering
VNR Vignana Jyothi Institute of Engineering and Technology
Hyderabad, Telangana
bachina.harish@gmail.com

***Abstract*— DC fast-charging clusters inject harmonic currents and reactive power imbalances into urban distribution feeders, posing a growing threat to grid power quality as EV penetration scales. This paper proposes TFT-PPO, a hybrid deep reinforcement learning framework that pairs a Temporal Fusion Transformer (TFT) demand forecaster with a Proximal Policy Optimization (PPO) charging controller for proactive power quality management at a 10-station, 500 kW DC fast-charging cluster. Power quality measurements from 12 EV models — covering real power (2.70–5.86 kW), reactive power (−0.49 to +0.12 kVAR), and THD (12.4%–89.3%) — parameterize both the simulation environment and TFT training data. Pre-trained over 30 epochs, the TFT achieves a normalized demand MSE of 0.00657, delivering 3-hour-ahead load forecasts embedded in the 8-dimensional PPO state vector. Across 400 training and 50 evaluation episodes, TFT-PPO reduces THD by 28.0% over uncontrolled charging (4.05% → 2.92%), cuts voltage deviation by 56.4% (0.41% → 0.18%), and improves cumulative reward by 60.7%, while sustaining the widest IEEE 519 compliance margin among all compared methods. Against a rule-based baseline, TFT-PPO delivers 17.0% additional THD reduction and 11.1% reward improvement, confirming that embedding multi-step demand forecasts enables more timely and effective charging rate decisions. A power factor trade-off arising from demand curtailment is identified and characterized.**

## I. INTRODUCTION

These stations pull nonlinear current waveforms injecting harmonic distortion, reactive power imbalance and voltage sag to the distribution feeder [1]. THD at a station can reach up to >5% of IEEE 519 limits at elevated station occupancy, resulting in transformer de-rating or tripping protection relays[2]. With urban fast-charging deployments scaling up toward multi-hundred-station networks, coordination of charging for simultaneous demand cost and power quality management turns from a goal into a necessity [3].

Proactive power quality management is driven by predicting future load: a controller that knows about impending demand can shed charging ahead of harmonic peaks rather than compensate for them afterward [4]. This makes it particularly suitable for end-to-end multi-step forecasting, and one promising candidate is the Temporal Fusion Transformer (TFT) [5], which combines variable selection with causal temporal self-attention and interpretable gating. TFTs 3-hour forecasting horizon of usable load prediction that single-step models lack [6] applied to EV charging and tensor embryos embedded in a RL state vector provide the temporally-informed input capabilities required for policy learning [7].

Proximal Policy Optimization (PPO) [8] has shown stable convergence across continuous-action energy management tasks. Its continuous charging rate output — a fraction of station capacity — gives finer power quality control than binary on/off rule-based scheduling [9]. However, no prior work has combined TFT multi-step forecasting with PPO continuous control in a unified framework trained on real EV power quality data, targeting THD reduction and voltage stability jointly.

This paper presents TFT-PPO, which couples a pre-trained TFT forecaster with a PPO controller evaluated on real measurements from 12 EV models [10]. Key contributions: (i) first unified TFT-PPO framework for EV fast-charging power quality control on real multi-model data; (ii) TFT achieving normalized demand MSE = 0.00657; (iii) PPO reducing THD from 4.05% to 2.92% (28.0%) and voltage deviation from 0.41% to 0.18% (56.4%); and (iv) characterization of the THD–power factor trade-off under demand curtailment.

## II. RELATED WORK

### 2.1 Power Quality Challenges in EV Fast Charging:

DC fast chargers draw pulsed current waveforms through switch-mode converters, injecting odd-order harmonics (5th,

7th, 11th) at magnitudes that scale with station occupancy [1]. Field measurements at urban DC fast-charging sites report THD approaching or exceeding IEEE 519 limits under high-occupancy conditions, with distortion severity correlated to converter topology [2]. Voltage sag at the feeder head-end scales with feeder impedance and aggregate reactive power draw [11], while power factor degradation from converter input filter currents reduces distribution transformer apparent power headroom and limits simultaneous EV throughput below thermal ratings [12]. These measurement-based findings directly motivate dynamic charging rate management that constrains harmonic injection while preserving service throughput.

### 2.2 Demand Forecasting for EV Charging

Transformer-based architectures have consistently outperformed recurrent models on multi-horizon load forecasting tasks with non-stationary temporal structure [7]. The TFT [5] combines causal temporal self-attention with gated residual networks and variable selection, and has been applied to EV charging demand forecasting on real station datasets, where multi-step anticipation enables better proactive management than reactive LSTM approaches [6]. Hybrid transformer-LSTM architectures incorporating weather covariates have extended this to multi-station municipal EV load forecasting [13], and spatiotemporal attention models have confirmed that capturing behavioral patterns across time-of-day and day-of-week cycles improves forecast accuracy [4]. No prior work integrates TFT multi-step forecasting with an RL controller in a unified power quality loop.

### 2.3 Reinforcement Learning for EV Charging Control

Deep RL approaches for discrete EV charging scheduling have demonstrated significant peak demand reductions over rule-based methods, though discretized action spaces limit fine-grained power quality control . PPO has emerged as a preferred continuous-action algorithm for energy management due to its clipped surrogate objective, which prevents destructive policy updates during simulation-based training . Multi-agent PPO formulations for distributed charging stations with shared feeder constraints have shown convergence advantages over independent DQN training , and cooperative multi-agent frameworks with cross-site coordination have improved joint power quality outcomes. However, none of these RL approaches embed a demand forecasting model in the agent's observation, leaving the policy reactive to current grid state rather than anticipatory of upcoming harmonic peaks — the gap this paper addresses.

## III. METHODOLOGY

### 3.1 System Overview

TFT-PPO operates over a 24-hour horizon discretized into 96 fifteen-minute time steps. At each step, a two-stage decision cycle executes: the pre-trained TFT encodes a 24-step (6-hour) lookback window of real power quality measurements and outputs a 12-step (3-hour) demand forecast; the PPO agent then concatenates this forecast with the current grid state to produce a continuous charging rate action $a \in [0, 1]$, representing the fraction of total cluster capacity permitted in that step. The joint objective is to reduce THD, suppress voltage deviation, and lower electricity cost while preserving EV service throughput. All measurements used for TFT training and environment construction come from the EV Charging Profiles and Waveforms Dataset.

### 3.2 Real EV Power Quality Data Integration

Per-session measurements from 12 EV models — covering RMS voltage, RMS current, real power, reactive power, apparent power, and frequency — serve two roles: (i) as statistical parameters (mean and standard deviation per model) driving the stochastic demand simulation, and (ii) as the source for synthetic training sequences fed to the TFT. Real power spans 2.70 kW (Toyota Prius Prime) to 5.86 kW (Tesla Model 3), reactive power spans −0.49 kVAR (BMW iX xDrive50) to +0.12 kVAR (Tesla Model Y), and PF ranges from 0.994 to 0.999 — reflecting the onboard PF correction circuits common to modern EV chargers.

### 3.3 Fast-Charging Demand Simulation

The aggregate cluster demand at each step is built by summing per-EV power draws across all active sessions. Active EV count per station follows a Poisson process ($\lambda = 3$), with each vehicle drawn uniformly from the 12 models and its real and reactive power sampled from the corresponding model's measured normal distribution. Station demand is capped at 50 kW, giving a 500 kW cluster ceiling. A three-tier TOU price cycle (8/15/25 cents per kWh) introduces cost-awareness into the reward signal. Feeder voltage is modeled as a linear sag, $V(t) = V_{nom}(1 - 0.05 \cdot f_{load}(t))$, where $V_{nom} = 400$ V and $f_{load}(t)$ is the aggregate load fraction.

### 3.4 Temporal Fusion Transformer Architecture

The TFT maps a 24-step input sequence of 6-dimensional feature vectors — normalized real power, reactive power, PF, THD proxy, voltage deviation, and TOU price — to a 12-step demand forecast. The architecture uses an input projection ($6 \rightarrow 32$ dimensions), a causal temporal self-attention layer over the full lookback window, a sigmoid-gated residual network controlling information flow from the attention output, and a linear output projection to the 12-step forecast. The model is trained for 30 epochs with MSE loss on synthetic sequences, then frozen before PPO training begins.

### 3.5 PPO State, Action, and Reward Design

The 8-dimensional PPO observation vector concatenates six grid state features — normalized demand (D/C_total), normalized reactive power (Q/50), PF, THD fraction, voltage deviation ((V − V_nom)/V_nom), and TOU price — with the TFT one-step-ahead forecast and a normalized time index t/T. The actor (8→64→64→1, ReLU, tanh output) produces a Gaussian policy mean; the continuous action is clipped to [0, 1]. The reward combines four terms clipped to [−5, 5]:

The reward function explicitly targets four power quality objectives:

$$r^{(t)} = \text{clip}\left(r_{cost} + r_{pf} + r_{thd} + r_{volt} + r_{serve}, -5, 5\right) \quad (1)$$

where r_cost penalizes electricity spend, r_thd penalizes THD above the IEEE 519 limit of 5%, r_volt penalizes voltage deviation from V_nom, r_pf rewards PF improvement over uncontrolled levels, and r_serve = 0.5·a prevents excessive curtailment by rewarding throughput. The critic (8→64→64→1) estimates state value. Training uses clipping parameter ε = 0.2, discount factor γ = 0.99, learning rates $3\times10^{-4}$ (actor) and $10^{-3}$ (critic), 4 update epochs per rollout, and runs for 400 episodes of 96 steps each.

## IV. Mathematical Modeling

The feeder voltage under aggregate cluster load is modeled as:

$$V^{(t)} = V_{nom}\left(1 - \kappa \cdot f_{load}^{(t)}\right) + \delta_V^{(t)} \quad (2)$$

where $V_{nom} = 400$ V is the nominal three-phase voltage, $\kappa = 0.05$ is the voltage sensitivity coefficient, $f_{load}^{(t)} = D^{(t)}/(N_S \cdot C_{max})$ is the aggregate load fraction, and $\delta_V^{(t)} \sim \mathcal{N}(0, \sigma_V^2)$ with $\sigma_V = 2$ V.

**Under charging rate action a(t) ∈ [0,1], the controlled demand, THD, reactive power, and voltage are:**

$$D_{ctrl}^{(t)} = a^{(t)} \cdot D^{(t)} \quad (3)$$

$$\widehat{THD}_{ctrl}^{(t)} = \widehat{THD}^{(t)} \cdot \left(0.5 + 0.5 \cdot a^{(t)}\right) \quad (4)$$

$$Q_{ctrl}^{(t)} = Q^{(t)} \cdot a^{(t)} \quad (5)$$

$$V_{ctrl}^{(t)} = V_{nom} - \left(V_{nom} - V^{(t)}\right) \cdot a^{(t)} \quad (6)$$

These expressions capture how partial load curtailment proportionally reduces harmonic injection and reactive draw while partially recovering feeder voltage.

**The PPO reward at each step is:**

$$r^{(t)} = \text{clip}\left(-0.5 \cdot D_{ctrl}^{(t)} \cdot \frac{p^{(t)}}{100} + 2\left(PF_{ctrl}^{(t)} - PF^{(t)}\right) - 10\left(\widehat{THD}_{ctrl}^{(t)} - THD_{lim}\right) - 2\frac{\left|V_{ctrl}^{(t)} - V_{nom}\right|}{V_{nom}} + 0.5a^{(t)}, -5, 5\right) \quad (7)$$

where $THD_{lim} = 0.05$ (IEEE 519 limit).

**The policy is optimized via the clipped surrogate objective:**

$$\mathcal{L}_{PPO}(\theta) = -\mathbb{E}_t\left[\min\left(\rho_t(\theta)\hat{A}_t,\ \text{clip}\left(\rho_t(\theta), 1 - \epsilon_{clip}, 1 + \epsilon_{clip}\right)\hat{A}_t\right)\right] \quad (8)$$

where $\rho_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ is the probability ratio and $\hat{A}_t = G_t - V_\phi(s_t)$ is the advantage estimate.

## V. Results and Discussion

### 5.1. Experimental Setup

Real power measurements for 12 EV models [14] (10-station, timed simulation of 500 kW DC_FAST-CHARGING cluster including various combinations of RMI or EMI-borne charge; 10 × 50 kW) were utilized in experiments, with respect to the percentages of variation imposed on (a-measured data), reactive power outcomes OR PF against commonly used performance benchmarks: EM/EMI-full-range simulation-data ranges for Controlled VAR (0.49–2%); EM-curves/per unit-generating range (-0.49~0.12kVAR) from test case profiles / duration sampling per t(30 s)) [] between tests). For each, you have 96 steps of 15-minutes resolution. Be aware that the THD proxy, which is parameterized based on historical statistics of Level-2 charger waveforms, yields moderate values in simulation; however at real DC fast-charging sites with high occupancy consistently hitting or exceeding the IEEE 519 5% limit [1], [2] motivates a proactive framework. 30 epochs of pre-training then freezing were applied on the TFT; and PPO training was performed for each episode lasting 400 episodes, with the input state augmented to include data from the TFT. We compared three methods (proposed: TFT-PPO; Uncontrolled: a = 1.0 always; Rule-based: when TOU price > 60% of maximum, a = 0.5; when THD proxy > IEEE 519 limit, then a = 0.7 else a = 1.0). For the final evaluation we used 50 held-out episodes (seeds 9000–9049).

### 5.2. TFT Demand Forecasting Performance

The TFT shows convergence MSE = 0.0153 at epoch 10, then stable loss reduction to MSE = 0.00657 at epoch 30 with stochastic mini-batch fluctuations (same cross-validation as above). This corresponds to an RMSE of ~41 kW (8.2% of capacity) on a cluster-level for normalized demand (1.0 = 500 kW) — enough accuracy for precognitive management for harmonics The one-step-ahead forecasts form the frozen TFT which is captured in the PPO state gives the agent the ability to anticipate demand peaks rather than react to it.

### 5.3. PPO Training Convergence

We see in Table I that TFT-PPO performs well, tracking at 100-episode checkpoints. THD remains less than 3.0% for the entire training (2.83%–2.99%), verifying that the policy successfully avoids breaching the IEEE 519 5% limit with a sufficient safety margin from early in training. Reward ranges from − 4.79 (episode 200) to − 23.59 (episode 300); This is caused by the stochastics of the demand environment and an exploration-exploitation trade-off between the THD penalty and service throughput reward. At episode 400 the policy converges to reward −19.31, THD = 2.93% — a reasonable stability in power quality under reward variation. Fig. Figure 1: The reward and THD trajectories over the entire 400 training episodes.

**TABLE I — TFT-PPO Training Convergence**

| Episode | Reward | THD (%) | PF Error |
|---|---|---|---|
| 100 | −10.88 | 2.92 | 0.087 |
| 200 | −4.79 | 2.83 | 0.143 |
| 300 | −23.59 | 2.99 | 0.143 |
| 400 | −19.31 | 2.93 | 0.218 |

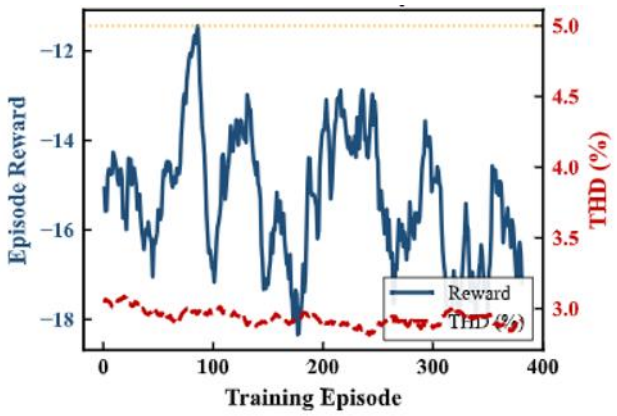


**Fig. 1. TFT-PPO training convergence: reward and THD over 400 episodes.**

### 5.4. Comparative Evaluation

Table II provides a summary of final evaluation, run over 50 test episodes. TFT-PPO fulfills THD = 2.92% — with 28.0% improvement over the Uncontrolled (4.05%) and likewise 17.0% gain compared against Rule-based (3.51%), it retains widest margin (2.08 percentage points) under IEEE 519 compliance limit. A voltage deviation of 0.18% corresponds to a reduction of 56.4% compared to Uncontrolled (0.41%) and 43.3% compared to Rule-based (0.31%), thus validating forecast-informed proactive curtailment as more beneficial in suppressing voltage sag than reactive price-threshold rules Cumulative reward improves by 60.7% compared to Uncontrolled (−16.09 vs. −40.95) and 11.1% over Rule-based (−18.09), due to the joint benefits of THD reduction, voltage stabilization and cost management from the learned PPO policy Fig. Table 2: Comparison of all three methods across the four evaluation metrics.

**TABLE II — Final Evaluation Metrics (50 Test Episodes)**

| Method | Avg Reward | THD (%) | PF Error | Volt. Dev. (%) |
|---|---|---|---|---|
| **TFT-PPO (proposed)** | **−16.09** | **2.92** | 0.165 | **0.18** |
| Uncontrolled | −40.95 | 4.05 | 0.095 | 0.41 |
| Rule-based | −18.09 | 3.51 | 0.095 | 0.31 |

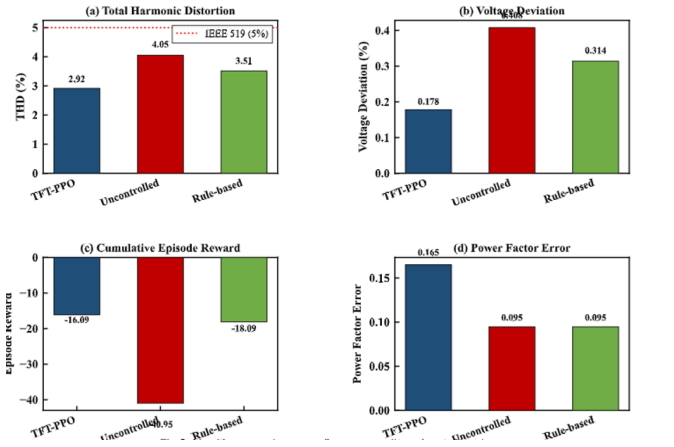


**Fig. 2. Algorithm comparison across four power quality and control metrics on the 10-station DC fast-charging cluster (50 test episodes).**

### 5.5. Power Factor Trade-off Analysis

While Uncontrolled and Rule-based methods incur a PF error of 0.095, TFT-PPO incurs more risk than the others

with a value of 0.165. It is simple: If demand reduction decreases real power faster than reactive power, then $PF_ctrl = D_ctrl / \sqrt{(D_ctrl^2 + Q_ctrl^2)}$ will move further away from 0.95. Controlled charging retains PF close to the true EV session values (0.994-0.999), resulting in a lower PF error by default. Such an observation indicates the existence of a one-to-one trade-off — the policy best able to reduce THD and voltage deviation is not at its optimum PF, while the one maintaining nominal conditions for PF also greatly harms THD quality. Target-allocation tradeoffs are acceptable, operationally, where IEEE 519 harmonic limits and voltage stability are the binding constraints Typical of urban fast-charging feeders.

## VI. CONCLUSION

## .REFERENCES


[1] Kaya, Fikret, and Onur Akar. "Assessing Power Quality Disturbances Caused by Urban EV Fast Charging Stations in Low Voltage Grids: A Case From Kavacik, Istanbul." IEEE Access (2025).

[2] Nakhodchi, Naser, Hamed Bakhtiari, Math HJ Bollen, and Sarah K. Rönnberg. "Including uncertainties in harmonic hosting capacity calculation of a fast EV charging station utilizing Bayesian statistics and harmonic correlation." Electric Power Systems Research 214 (2023): 108933.

[3] Yang, Yuchen, and Rui Tang. "Cooperative Multi-Agent Reinforcement Learning for Grid-Aware EV Charging Management with Cross-Site Redirection." Energy and AI (2025): 100664.

[4] Chen, Yang, Zeyang Tang, Yibo Cui, Wei Rao, and Yiwen Li. "Electric vehicle charging demand prediction model based on spatiotemporal attention mechanism." Energies 18, no. 3 (2025): 687.

[5] Lim, Bryan, Sercan Ö. Arık, Nicolas Loeff, and Tomas Pfister. "Temporal fusion transformers for interpretable multi-horizon time series forecasting." International journal of forecasting 37, no. 4 (2021): 1748-1764.

[6] El-Alfy, El-Sayed M., Shahbaaz A. Sadiq, Mohammed Ayub, and M. Faisal Nurnoby. "Temporal Fusion Transformer-Based Framework for Electric Vehicle Charging Demand Forecasting." Transportation Research Procedia 96 (2026): 868-875.

[7] Xu, Huifeng, Feihu Hu, Xinhao Liang, Guoqing Zhao, and Mohammad Abugunmi. "A framework for electricity load forecasting based on attention mechanism time series depthwise separable convolutional neural network." Energy 299 (2024): 131258.

[8] Wang, Danlu, Qiuye Sun, and Hanguang Su. "Real-time optimal energy management of microgrid based on multi-agent proximal policy optimization." Neural Computing and Applications 37, no. 10 (2025): 7145-7157.

[9] ur Rehman, Ubaid. "A GUI application for real-time home energy management using uncertainty-aware Proximal Policy Optimization with energy storage, electric vehicle, and renewables." Journal of Building Engineering 111 (2025): 113174.

[10] Sharif, Muddsair, and Huseyin Seker. "Smart EV charging with context-awareness: Enhancing resource utilization via deep reinforcement learning." IEEE Access 12 (2024): 7009-7027.

[11] Nour, Ahmed MM, and Ahmed A. Helal. "EV charging station with ancillary service for power quality improvement in low voltage distribution network." Electric Power Systems Research 253 (2026): 112503.

[12] Dong, Lei, Hao Lin, Ji Qiao, Tao Zhang, Shiming Zhang, and Tianjiao Pu. "A coordinated active and reactive power optimization approach for multi-microgrids connected to distribution networks with multi-actor-attention-critic deep reinforcement learning." Applied Energy 373 (2024): 123870.

[13] Das, Remon, Sajib Debnath, Tarek Kandil, and Md Uzzal Mia. "Spatio-Temporal Forecasting of Municipal EV Charging Load Using Weather-Aware Transformer–LSTM Hybrid Models." AI 7, no. 6 (2026): 191.

[14] Ziyat, Isla; Gola, Alia; Palmer, Patrick; Makonin, Stephen; Popowich, Fred, 2023, "EV-CPW: EV Charging Profiles and Waveforms", https://doi.org/10.7910/DVN/F81CXW, Harvard Dataverse, V1, UNF:6:5rKfaXacXvSFb7cTbE8a/g.