A Metric-Driven Hyperparameter Optimization Framework for EV Charging Occupancy Forecasting Using LSTM and XGBoost

Juan Carlos Cortez[1,2], Cindy P. Guzmán[2], Lucas Pereira[3], Marcos J. Rider[1], and Hugo Morais[2]

[1]School of Electrical and Computer Engineering (FEEC), Universidade Estadual de Campinas (UNICAMP), Brazil
[2]INESC-ID, Instituto Superior Técnico, IST, Universidade de Lisboa, Lisboa, Portugal
[3]Interactive Technologies Institute, LARSyS, Universidade de Lisboa, Lisboa, Portugal
Emails: j265568@dac.unicamp.br, cindy.lascano@tecnico.ulisboa.pt, lucas.pereira@tecnico.ulisboa.pt, mjrider@unicamp.br, hugo.morais@tecnico.ulisboa.pt

***Abstract*—Forecasting the occupancy of electric vehicle (EV) charging stations is important for optimizing energy management, reducing user delays, and improving infrastructure planning. This paper presents a metric-driven hyperparameter optimization framework to enhance the performance of long short-term memory (LSTM) and eXtreme Gradient Boosting (XGBoost) models on three forecast horizons: 2 hours, 6 hours, and 24 hours. Using two optimization tools, Optuna and Hyperopt, models are tuned for three objective metrics: accuracy, F1 score, and balanced accuracy. Experiments are conducted on an open-access dataset sourced from the Dundee City Open Data Portal to facilitate reproducibility. The results show that using balanced accuracy as the objective function consistently led to better-performing models, especially in capturing occupied intervals under class imbalance. Compared to a state-of-the-art hybrid LSTM baseline, the proposed framework improved F1 score by up to 10.6% and balanced accuracy by 4.3%. These findings reinforce the importance of choosing appropriate metrics in hyperparameter optimization for EV occupancy forecasting.**

***Index Terms*—EV Charging Occupancy, Forecasting Metrics, Hyperparameter Optimization, LSTM, Public Datasets, XGBoost.**

I. Introduction

Electric vehicles (EVs) are rapidly transforming global transportation systems as part of initiatives to mitigate greenhouse gas emissions and address climate change. According to the International Energy Agency [1], EV sales are projected to reach 55% of total vehicle sales by 2035, reinforcing the urgent need for a robust public charging infrastructure, which is expected to exceed 15 million units by 2030. While EVs stand out as one of the few clean technologies currently on track with net-zero emission goals [2], their accelerated deployment brings operational challenges for distribution grids. In that context, forecasting EV charging station occupancy is important for efficient infrastructure use, improving grid reliability, and minimizing user waiting times through improved energy and mobility planning. This need is particularly relevant in shared or constrained environments like parking lots, where fairness, limited power capacity, and real-time decision-making must be considered [3].

Several studies have explored forecasting techniques to predict EV charging behaviors, focusing on different tasks, scopes, and data types. Some works address city-scale demand estimation using spatiotemporal frameworks that combine external urban features with deep learning architectures. For example, CityEVCP [4] predicts regional EV charging demand by integrating static and dynamic area characteristics with graph-based modules and Transformer encoders. Other studies, such as [5], focus on multi-task learning to simultaneously predict charging occupancy, volume, and duration using temporal GraphSAGE models. These methods are helpful for grid-level planning but do not offer charger-level occupancy granularity.

A second category includes studies that estimate occupancy across multiple chargers. In [6], a federated deep learning framework predicts charger occupancy over short-term horizons using Long Short-Term Memory (LSTM), BiLSTM, and ConvLSTM. Although the architecture improves accuracy and privacy, the authors rely on manual hyperparameter tuning. Similarly, [7] benchmarks 14 models for day-ahead EV load and occupancy forecasting across eight datasets from residential, workplace, and public charging stations. The authors use grid search to tune hyperparameters and highlight that the Aggregation of Experts, an ensemble approach, was best for occupancy forecasting.

Finally, charger-level occupancy forecasting, treated as a binary classification task, is explored in [8] and [9]. The authors of [8] propose a hybrid LSTM architecture based on historical occupancy and time-related features. The model was tested on three months of data from nine rapid chargers from Dundee, UK, for forecast horizons from 10 minutes to 6 hours. The accuracy and F1 score of the model outperform four conventional machine learning and three deep learning methods on all forecast horizons. This model remains a baseline reference due to its performance and clear methodological design, although its hyperparameters were only manually tuned. In [9], authors evaluate data-driven models for day-ahead EV occupancy and power consumption forecasting. Hyperparameters are optimized via sequential model-based optimization; XGBoost performs best at a Boulder City charger point for charging occupancy.

Forecasting EV charger occupancy is particularly difficult due

This paper is supported by the European Union's Horizon Europe R&I programme under grant agreement no. 101160665, project AHEAD (AI-informed Holistic EVs integration Approaches for Distribution grids), by Brazilian institutions CAPES (process number: 8887.801350/2023-00) and São Paulo Research Foundation – FAPESP (Thematic project 2021/11380-5). This work was developed within the scope of the project no 56 - "Aliança para a Transição Energética", financed by European Funds, namely "Recovery and Resilience Plan (PRR) - Component 5: Agendas Mobilizadoras para a Inovação Empresarial", included in the NextGenerationEU funding program. This work also was funded by the Portuguese Foundation for Science and Technology (FCT) under grant UIDB/50021/2020.

979-8-3503-9042-1/24/$31.00

to low average usage rates, often below 40%, which results in highly imbalanced datasets [7], [8]. In such cases, models that merely predict non-occupancy can achieve high accuracy while failing to identify real charging events. Metrics like F1 score and balanced accuracy (bACC) are better suited to this task, as they account for false positives and false negatives [10]. However, few studies compare these metrics in the context of charger-level predictions. Moreover, machine learning and deep learning models depend heavily on hyperparameter optimization, but the effect of different optimization strategies on EV occupancy forecasting remains underexplored. Understanding which metrics and tools lead to better results is key to developing robust occupancy forecasting models under real-world constraints.

In response to those gaps, this paper proposes a metric-driven hyperparameter optimization framework tailored to EV charging station occupancy forecasting. Unlike prior studies that rely on manual or single-metric tuning, our approach systematically compares the performance of LSTM and XGBoost models optimized with Optuna [11] and Hyperopt [12] across multiple forecasting horizons (2-hour, 6-hour, and 24-hour ahead). Balanced accuracy, F1 score, and accuracy are used to guide optimization because they capture different aspects of model performance, reflecting distinct evaluation priorities such as class balance and event detection. The methodology is validated on a publicly available dataset from Dundee City to support reproducibility and benchmarked against a state-of-the-art hybrid LSTM model. The contributions include: i) a structured comparison of metric-guided hyperparameter tuning strategies, ii) the benchmarking of LSTM and XGBoost models under different objective functions and forecasting horizons, and iii) the provision of reproducible baselines for future research. The remainder of this paper is organized as follows: Section II describes the proposed framework and experimental setup; Section III presents and discusses the results; and Section IV concludes the paper.

II. Methodology

The proposed framework for metric-driven hyperparameter optimization for EV charging occupancy forecasting is shown in Fig. 1, with green boxes indicating the main steps, orange for internal tasks, and yellow for the final goal. The methodology begins with the search for a publicly accessible data set for the EV session to ensure the reproducibility of experiments and results. The data set is then preprocessed, starting by discarding sessions with incomplete data. Later, sessions with a duration time equal to zero or more than three standard deviations from the median are deleted. Then, for every session, all minutes from the start time of the connection to the final minute within its charging duration are assigned the value one (occupied). Once all sessions are processed, a minute-by-minute occupancy time series is obtained. This time series is then downsampled into 10-minute timestamps, marked as occupied if any minute in the segment is occupied. The data is split chronologically into training (60%), validation (20%), and testing (20%) subsets.

LSTM [13] and XGBoost [14] are chosen as forecasting models to be optimized. LSTM is a recurrent neural network architecture designed to model time-dependent patterns in sequential data, making it suitable for time series problems like EV occupancy. XGBoost, on the other hand, is a gradient boosting algorithm known for its strong performance on tabular data and its ability to handle non-linearities and feature interactions efficiently. Comparing these two modeling paradigms allows assessing their advantages in EV occupancy forecasting. Similar comparisons between LSTM and XGBoost have been successfully applied in other domains, including sales prediction [15] and short-term photovoltaic forecasting [16]

TABLE I
LSTM Hyperparameter Search Space

Hyperparameter	Search Space
history_steps	depends on forecasting horizon
num_lstm_blocks	{1, 2, 3}
lstm_units_i	[16, 128]
lstm_dropout	[0.0, 0.5] (step=0.05)
lstm_dense_units	[16, 64]
dense_dropout	[0.0, 0.5] (step=0.05)
learning_rate	[1e-5, 1e-3] (log)
batch_size	{16, 32, 64}

TABLE II
XGBoost Hyperparameter Search Space

Hyperparameter	Search Space
history_steps	depends on forecasting horizon
max_depth	[3, 10]
learning_rate	[1e-3, 1e-1] (log)
n_estimators	[50, 300]
subsample	[0.5, 1.0]
colsample_bytree	[0.5, 1.0]
gamma	[0.0, 5.0]
reg_lambda	[1e-3, 10.0] (log)
reg_alpha	[1e-3, 10.0] (log)

For the LSTM-based forecasting model, hyperparameter optimization defines its architecture and training dynamics. The model receives sequential occupancy data over a look-back window, whose length is tuned as part of the optimization process. One to three LSTM blocks can be stacked, each with between 16 and 128 units. A dropout layer follows these LSTM layers, preventing overfitting. A dense layer between 16 and 64 units is added, followed by another dropout layer. The output layer is a fully connected layer with sigmoid activation that produces a multi-step EV charging station occupancy adjusted according to the forecast horizon. The loss function used is binary cross-entropy, and the model is trained using the Adam optimizer, with batch size selected as part of the hyperparameter tuning. Models are trained for up to 30 epochs using early stopping to prevent overfitting. Table I summarizes the hyperparameters and search space of the LSTM-based model. For the XGBoost model, hyperparameter optimization is applied to control the complexity and regularization of the ensemble. The *history_steps* defines the number of look-back values for the input of the model. The search space includes key hyperparameters such as maximum tree depth, learning rate, number of estimators, subsample ratio, and column sampling ratio. Additionally, regularization parameters such as gamma, L1, and L2 penalties are optimized. Table II presents the complete hyperparameters used in the tuning process.

Three metrics are selected as objective functions for the hyperparameter optimization process: accuracy, F1 score, and balanced accuracy. These binary classification metrics evaluate various aspects of the performance of the EV charging occupancy forecast. Accuracy quantifies the overall proportion of correct predictions (1). However, it can be misleading in imbalanced datasets, where occupied time steps are much less frequent than unoccupied ones. The F1 score combines precision, minimizing false positives (FP), and recall, minimizing false negatives (FN), into one metric (2). A FP occurs when a charger is predicted to be occupied but is not, while a FN is when a charger is predicted

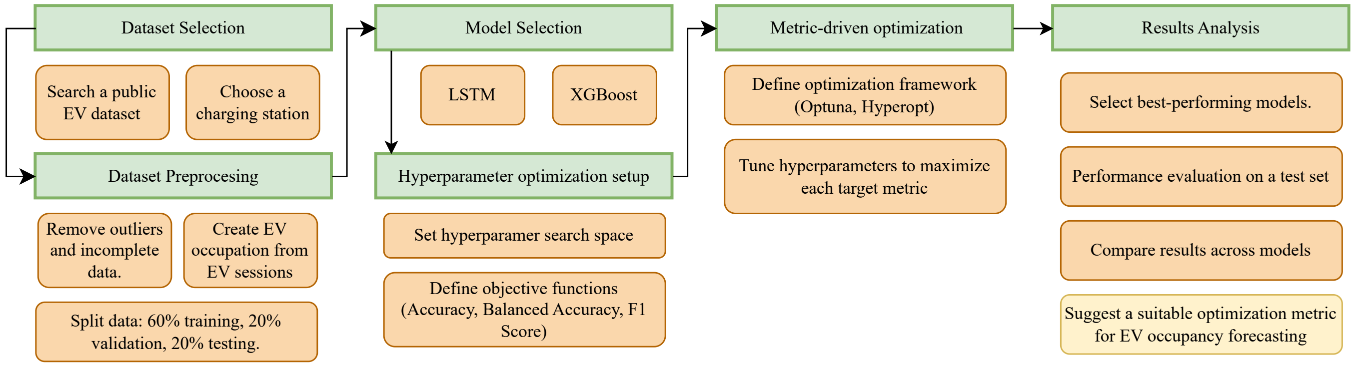

Fig. 1. Metric-driven hyperparameter optimization framework

to be free but is actually in use. Therefore, the F1 score evaluates the ability of the model to detect occupancy periods correctly while minimizing FP. Balanced accuracy effectively tackles the imbalance problem by equally weighing the correct predictions in the occupied and unoccupied states (3). It does this by averaging their true positive (TP) and true negative (TN) rates. The objective functions for hyperparameter optimization are maximized.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2\left(\frac{\text{TP}}{2\ \text{TP} + \text{FP} + \text{FN}}\right) \tag{2}$$

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}}\right) \tag{3}$$

Hyperparameter optimization leverages two frameworks: Optuna and Hyperopt. Both use Bayesian optimization with the Tree-structured Parzen Estimator for handling categorical and continuous hyperparameters. Although they share this approach, they differ in implementation and usability. These frameworks are renowned for their efficient exploration of large hyperparameter spaces. Each optimization framework performs 100 trials independently for every combination of objective function (accuracy, F1 score, and balanced accuracy), forecasting horizon (2h, 6h, and 24h), and forecasting model (LSTM and XGBoost), exploring the respective hyperparameter search spaces. After the trials are completed, a comparative analysis is performed to identify the most appropriate optimization metric for EV occupancy forecasting. The selected optimal models are evaluated on the test set and compared to a baseline model to demonstrate the advantages of metric-driven hyperparameter optimization.

III. Results and Discussion

A. Dataset Selection and Preprocessing

This paper uses a publicly available EV charging dataset from the Dundee City Open Data Portal [17]. It contains EV charger session data from January 2023 to December 2024 of 50 fast, 28 rapid, and 4 ultra-rapid chargers. The rapid charger with ID 51421 was selected for analysis due to a greater number of sessions with an occupancy rate of 31.05%, which reflects a typical class imbalance scenario. Initially, 8516 sessions were associated with this charger, but after preprocessing, 8137 sessions remained. It is 11.1 sessions per day on average, each lasting about 32.8 minutes on average. To contextualize the characteristics of the dataset, Fig. 2 shows the distribution of the start times of the

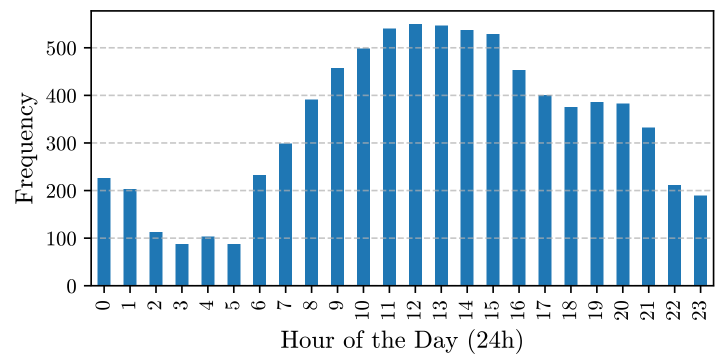

Fig. 2. Session Start Time Distribution for rapid station 51421

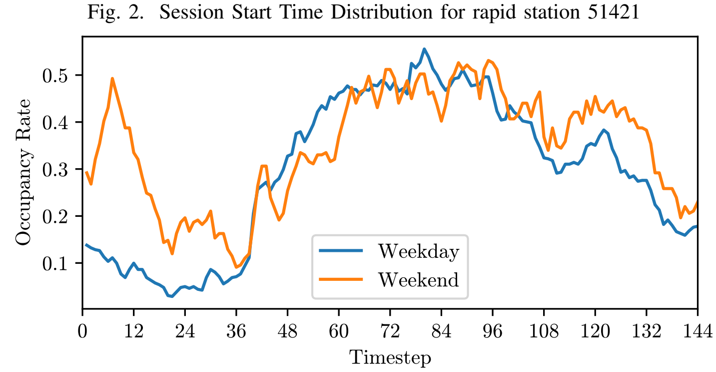

Fig. 3. Occupancy Rate Profile: Weekdays vs. Weekends for rapid station 51421.

sessions. Fig. 3 illustrates higher weekend occupancy rates than on weekdays, particularly in the early hours. These characteristics provide a realistic and challenging dataset to evaluate the effects of metric-driven optimization strategies.

B. Baseline Model and Input Features

The hybrid LSTM model proposed in [8] serves as a baseline to evaluate the performance of the forecast. Its architecture includes two input branches: one comprising a 12-step sequence of historical binary occupancy states, and another containing time-related features: time of day, day of the week, weekday/weekend classification, and a 144-dimensional occupancy rate profile. The latter is a fixed vector representing the typical charger occupancy rates, computed separately for weekdays and weekends using 60% of data. In this paper, all these features are concatenated for input into LSTM and XGBoost models. No additional feature engineering or selection is performed to ensure consistency with the baseline configuration. The processed dataset is chronologically split into 60% for training, 20% for validation, and 20% for testing. Unlike the fixed 12-step input used in the baseline, this study treats the number of historical steps as a tunable hyperparameter, allowing the models to adapt their temporal context based on forecasting horizon and optimization performance.

TABLE III
VALIDATION RESULTS FOR 2H FORECAST OF EV LOAD OCCUPANCY

exp	Objetive	Accuracy	F1	bACC	conv.	Time (min)
Baseline Model						
		0.749	0.400	0.612		
XGBoost - Optuna						
1	Accuracy	**0.773**	0.249	0.564	90	12.6
2	f1	0.744	0.443	0.633	96	13.2
3	bACC	0.744	0.443	0.633	84	11.1
XGBoost - Hyperopt						
4	Accuracy	0.772	0.215	0.554	99	16.5
5	f1	0.744	0.445	0.634	71	15.0
6	bACC	0.744	0.442	0.633	66	13.9
LSTM - Optuna						
7	Accuracy	0.771	0.300	0.579	52	372.3
8	f1	0.743	0.438	0.630	40	1051.1
9	bACC	0.737	**0.450**	**0.636**	98	412.6
LSTM - Hyperopt						
10	accuracy	0.769	0.335	0.590	89	420.6
11	f1	0.737	0.435	0.628	80	419.7
12	bACC	0.739	0.443	0.632	52	499.8

TABLE IV
VALIDATION RESULTS FOR 6H FORECAST OF EV LOAD OCCUPANCY

exp	Objetive	Accuracy	F1	bACC	conv.	Time (min)
Baseline Model						
		0.731	0.373	0.595		
XGBoost - Optuna						
13	Accuracy	**0.761**	0.072	0.517	41	24.3
14	f1	0.732	0.398	**0.607**	96	48.2
15	bACC	0.731	**0.399**	**0.607**	93	34.1
XGBoost - Hyperopt						
16	Accuracy	**0.761**	0.095	0.521	77	47.4
17	f1	0.731	0.397	0.606	74	36.2
18	bACC	0.731	0.396	0.606	84	40.2
LSTM - Optuna						
19	Accuracy	0.756	0.181	0.538	62	295.0
20	f1	0.724	0.385	0.599	14	461.8
21	bACC	0.725	0.397	0.605	76	216.4
LSTM - Hyperopt						
22	accuracy	0.753	0.215	0.545	73	452.8
23	f1	0.735	0.369	0.594	19	705.7
24	bACC	0.735	0.381	0.599	87	660.4

TABLE V
VALIDATION RESULTS FOR 24H FORECAST OF EV LOAD OCCUPANCY

exp	Objetive	Accuracy	F1	bACC	conv.	Time (min)
Baseline Model						
		0.729	0.345	0.582		
XGBoost - Optuna						
25	Accuracy	**0.757**	0.078	0.515	99	186.3
26	f1	0.723	0.369	0.591	98	83.6
27	bACC	0.728	0.370	0.592	92	123.4
XGBoost - Hyperopt						
28	Accuracy	**0.757**	0.068	0.514	66	273.8
29	f1	0.720	0.368	0.589	90	94.2
30	bACC	0.728	0.369	0.592	79	104.5
LSTM - Optuna						
31	Accuracy	0.755	0.125	0.524	93	531.6
32	f1	0.718	0.384	0.597	93	620.6
33	bACC	0.718	**0.389**	**0.599**	91	304.6
LSTM - Hyperopt						
34	Accuracy	0.753	0.141	0.527	25	630.0
35	f1	0.721	0.372	0.591	76	575.0
36	bACC	0.721	0.381	0.596	25	1299.7

C. Hyperparameter Optimization

Hyperparameter optimization experiments were performed for each of the three forecasting horizons, 2, 6, and 24 hours, using LSTM and XGBoost models. Each model was tuned with Optuna and Hyperopt, using three objective metrics: accuracy, F1 score, and balanced accuracy, with 100 trials per run. This results in 12 optimization runs per horizon (2 models $\times$ 2 tools $\times$ 3 metrics), totaling 36 experiments. The results are shown in Tables III, IV, and V, and include validation performance of the baseline model, which was evaluated 10 times with different random seeds, as reported in the original study. For each metric, the highest-performing configuration is highlighted.

The metric-driven optimization framework consistently outperformed the baseline model across all forecasting horizons, especially considering the F1 score and balanced accuracy. While the baseline hybrid LSTM achieved slightly better accuracy in certain cases, it consistently showed lower capability to detect actual charging sessions, reflected by its notably lower F1 and balanced accuracy scores. This confirms that accuracy alone is misleading in EV occupancy forecasting. Forecasting performance decreased over longer horizons, which is expected due to higher uncertainty in longer-term predictions. The drop was more pronounced for balanced accuracy and F1 score, which emphasize the correct identification of occupied periods.

Between the evaluated models, XGBoost and LSTM exhibit very similar predictive performance. However, LSTM offers slightly better results at the cost of significantly higher training time, up to forty-five times longer than XGBoost, as seen when comparing experiments 3 (XGBoost) and 12 (LSTM). Given this context, XGBoost presents a valuable alternative for real-time operational environments where computational resources and speed are crucial. In terms of convergence, no consistent pattern favors one model across tools. Under Optuna, convergence rates are similar between XGBoost and LSTM. Furthermore, Optuna consistently outperformed Hyperopt, suggesting it better exploits the hyperparameter search space for these forecasting models.

Models optimized for accuracy often defaulted to predicting non-occupancy, inflating accuracy scores while failing to detect real charging events. Therefore, accuracy fails the real purpose of EV occupancy forecasting: to anticipate when a charger will be used. For instance, in the 2-hour horizon, experiment 1 reached 0.773 in accuracy but only 0.249 in F1 and 0.564 in balanced accuracy. In contrast, experiment 9, optimized for balanced accuracy, achieved 0.636 in bACC and 0.450 in F1, the highest for this horizon. The same pattern appears in the 6-hour horizon, where experiment 13 (accuracy) reached only 0.072 in F1, while experiment 15 (balanced accuracy) yielded 0.399 in F1 and 0.607 in bACC. For the 24-hour horizon, experiment 25 (accuracy) again scored low on F1 (0.078) compared to experiment 33 (balanced accuracy), which achieved the top F1 (0.389) and bACC (0.599). These results demonstrate that balanced accuracy consistently offers the best compromise between correctly identifying occupied and unoccupied states, making it the most appropriate metric for hyperparameter optimization in EV occupancy forecasting tasks.

D. Model Architecture Analysis

After identifying the best-performing configurations based on balanced accuracy, three models were selected for final evaluation: Experiment 9 (LSTM) for 2-hour forecasting, Experiment 15 (XGBoost) for 6-hour forecasting, and Experiment 33 (LSTM) for day-ahead forecasting. These models are detailed in Table VI. Notably, all three were obtained using Optuna.

Optuna's importance analysis highlights which parameters most influenced performance. For the 2-hour LSTM model, *lstm_dropout* (42.7%), *dense_dropout* (22.5%), and *learning_rate* (11.6%) were the most influential. This shows the critical role of regularization and stable training at very short horizons. For

TABLE VI
BEST MODEL ARCHITECTURES FOR EACH HORIZON

Exp	Hyperparameters	Horizon
9	history_steps=36, num_lstm_blocks=1, lstm_units=40, lstm_dropout=0.35, lstm_dense_units=57, dense_dropout=0.0, lr=3.27e-05, bs=32	2h
15	history_steps=12, max_depth=10, lr=0.08433, n_estimators=291, subsample=0.701, colsample_bytree=0.916, gamma=1.40, reg_lambda=0.282, reg_alpha=9.82	6h
33	history_steps=12, num_lstm_blocks=2, lstm_units=[87,100], lstm_dropout=0.25, lstm_dense_units=58, dense_dropout=0.0, lr=5.53e-04, bs=16	24h

TABLE VII
TEST PERFORMANCE: BEST MODEL VS. BASELINE

Horizon	**Best Model**			**Baseline**		
	Accuracy	F1	bACC	Accuracy	F1	bACC
2h	0.750	0.523	0.675	0.758	0.473	0.647
6h	0.730	0.439	0.624	0.732	0.425	0.618
24h	0.716	0.424	0.613	0.724	0.385	0.597

the 6-hour XGBoost model, performance was almost entirely determined by the *learning_rate* (90.1%), far exceeding the contributions of *n_estimators* (4.7%) and *max_depth* (2.2%). In contrast, the 24-hour LSTM model was mainly influenced by *history_steps* (81.2%), followed by *dense_dropout* (8.9%) and *lstm_dense_units* (3.8%). This suggests that capturing longer temporal patterns becomes more critical than regularization as the forecast horizon grows. Notably, *learning_rate* consistently remains moderately influential across all models.

E. Test Set Evaluation

The best trained model from each forecasting horizon (Table VI) was used for final evaluation on the independent test set. Table VII presents the corresponding test performance compared to the baseline hybrid LSTM model. In all horizons, the metric-driven optimization approach outperformed the baseline in both F1 score and balanced accuracy—metrics more appropriate for EV occupancy forecasting. However, accuracy decreased slightly in two of the three horizons.

At the 2-hour horizon, the optimized model improved the F1 score by 10.6% (from 0.473 to 0.523) and balanced accuracy by 4.3% (from 0.647 to 0.675), while accuracy dropped only 1.1%. For the 6-hour forecast, F1 score increased by 3.3%, and balanced accuracy by 1%, with nearly identical accuracy. At the 24-hour horizon, F1 score improved by 10.1% (from 0.385 to 0.424), and balanced accuracy by 2.7% (from 0.597 to 0.613) with a minimal 0.8% decrease in accuracy. These results show that the proposed metric-driven hyperparameter optimization framework outperforms the baseline model in key metrics. By optimizing balanced accuracy, the approach improves model generalization to new data while avoiding overfitting, thereby improving the reliability of EV occupancy forecasting.

IV. CONCLUSIONS

This paper proposed a metric-driven hyperparameter optimization framework to improve electric vehicle (EV) charging station occupancy forecasting. The study systematically evaluated how different objective metrics affect predictive performance by tuning Long Short-Term Memory (LSTM) and eXtreme Gradient Boosting (XGBoost) models using two optimization tools: Optuna and Hyperopt, across forecasting horizons of 2, 6, and 24 hours. All experiments were carried out on a real-world dataset from the Dundee City Open Data Portal, ensuring reproducibility and allowing future benchmarking on publicly available data. The results showed that optimizing based on balanced accuracy produced the best overall models. In contrast, accuracy-based optimization consistently failed to capture charging events due to the imbalanced nature of occupancy data. Compared to a previously published hybrid LSTM baseline, the proposed models achieved significant improvements of up to 10.6% in F1 score and 4.3% in balanced accuracy for the test set. Optuna consistently provided the best results, while LSTM offered slightly higher performance at a significantly higher computational cost. Future work will explore additional architectures and incorporate richer temporal and contextual features.

REFERENCES

[1] International Energy Agency (IEA), “Global ev outlook 2024,” IEA, Paris, Tech. Rep., 2024, licence: CC BY 4.0. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2024

[2] P. IEA, “Tracking clean energy progress 2023,” in *IEA*, 2023. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2024

[3] H. Morais, “New approach for electric vehicles charging management in parking lots considering fairness rules,” *Electric Power Systems Research*, vol. 217, p. 109107, 2023.

[4] H. Kuang, K. Deng, L. You, and J. Li, “Citywide electric vehicle charging demand prediction approach considering urban region and dynamic influences,” *Energy*, vol. 320, p. 135170, 2025.

[5] Y. Shang, D. Li, Y. Li, and S. Li, “Explainable spatiotemporal multi-task learning for electric vehicle charging demand prediction,” *Applied Energy*, vol. 384, p. 125460, 2025.

[6] L. Douaidi, S.-M. Senouci, I. El Korbi, and F. Harrou, “Predicting electric vehicle charging stations occupancy: a federated deep learning framework,” in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. IEEE, 2023, pp. 1–5.

[7] Y. Amara-Ouali, Y. Goude, B. Hamrouche, and M. Bishara, “A benchmark of electric vehicle load and occupancy models for day-ahead forecasting on open charging session data,” in *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*, 2022, pp. 193–207.

[8] T.-Y. Ma and S. Faye, “Multistep electric vehicle charging station occupancy prediction using hybrid lstm neural networks,” *Energy*, vol. 244, p. 123217, 2022.

[9] H. Amezquita, C. P. Guzman, and H. Morais, “Forecasting electric vehicles’ charging behavior at charging stations: A data science-based approach.” *Energies (19961073)*, vol. 17, no. 14, 2024.

[10] P. Branco, L. Torgo, and R. P. Ribeiro, “A survey of predictive modeling on imbalanced domains,” *ACM computing surveys (CSUR)*, vol. 49, no. 2, pp. 1–50, 2016.

[11] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2623–2631.

[12] J. Bergstra, D. Yamins, and D. Cox, “Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures,” in *International conference on machine learning*. PMLR, 2013, pp. 115–123.

[13] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[15] A. Massaro, A. Panarese, D. Giannone, and A. Galiano, “Augmented data and xgboost improvement for sales forecasting in the large-scale retail sector,” *Applied Sciences*, vol. 11, no. 17, p. 7793, 2021.

[16] J. C. Cortez, L. Z. Terada, B. V. B. Bandeira, J. Soares, Z. Vale, and M. J. Rider, “Comparative analysis of arima, lstm, and xgboost for very short-term photovoltaic forecasting,” in *2023 15th Seminar on Power Electronics and Control (SEPOC)*. IEEE, 2023, pp. 1–6.

[17] Dundee City Council, “Ev charging sessions dataset,” https://data.dundeecity.gov.uk/search?tags=ev%2520charging, 2018, accessed: April 2025.