Table S1: Summary of Related Work and Research Gaps Identified from Literature.

Ref.	Paper Focus	Dataset Used	Key Methods	Key Findings	Relevance to This Study & Gap Addressed
[1]	Solar irradiance prediction in Saudi Arabia using ML & ANN	PVGIS, KACARE, NREL	Machine Learning, ANN	Weather parameters correlate with irradiance; data analysis improves prediction accuracy	Supports: Use of Saudi data; emphasizes multiple data sources; Gap: Limited to ANN, lacks comprehensive model benchmarking
[2]	Solar power prediction leveraging large-scale PV data & weather inputs	Large-scale PV farm in Saudi Arabia	ML models (multiple)	Weather data improves prediction accuracy; various models compared	Supports: Saudi context; weather inputs importance; Gap: No probabilistic forecasting or dust event analysis
[3]	Comparative analysis of ML vs DL for short-term solar forecasting	Not specified	Linear Reg., Lasso, SVR, RF, LSTM, BiLSTM	BiLSTM achieved lowest MAE (0.085) and highest R^2 (0.890)	Supports: ML vs DL comparison; Gap: No arid climate focus; no dust or uncertainty analysis
[4]	Multi-model evaluation of direct solar radiation forecasting	Not specified	RF, LR, ANN, k-NN, SVR, GBR, GPR, Deep Learning	Deep Learning lowest RMSE; Trigonometric Cyclical Encoding improved temporal features by 49-53%	Supports: Multi-model comparison; temporal feature engineering; Gap: Saudi context; dust impact not analyzed
[5]	Review of data-driven approaches for	Not applicable (review)	19 models reviewed from statistical to DL	Comprehensive review; actionable	Supports: Comprehensive benchmarking rationale; Gap:

Ref.	Paper Focus	Dataset Used	Key Methods	Key Findings	Relevance to This Study & Gap Addressed
	solar irradiance forecasting			insights for model selection	No experimental validation; no dust-specific analysis
[6]	Long-term solar irradiation forecasting in Riyadh, Saudi Arabia	Filtered solar irradiation dataset from Riyadh	SVM, Gaussian Process with DWT	Improved forecasting accuracy for Riyadh's challenging conditions	Supports: Riyadh, Saudi Arabia location; dust storm consideration; Gap: SVM and GP only; no dust quantification
[7]	ML comparison with correlation feature selection for solar irradiation	Not specified	ANN (various architectures), RF	Best ANN: RMSE 25.92 W/m², R^2 0.9926; outperformed RF (RMSE 32.83)	Supports: ANN effectiveness; feature selection; Gap: No Saudi or arid climate validation
[8]	RL-tuned cascaded regressors for irradiance forecasting	Not specified	GBDT + SVR with RL hyperparameter tuning	Superior RMSE, MAE, R^2 compared to state-of-the-art	Supports: Uncertainty quantification; ensemble methods; Gap: No arid climate validation; complex implementation
[9]	Gradient Boosting with Kneedle-based data splitting for GTI prediction	High-resolution GTI data	Gradient Boosting Regression	$R^2 > 0.996$ across configurations; data-driven split improves evaluation	Supports: Gradient Boosting; principled evaluation; Gap: Single model focus; no dusty vs clean comparison
[10]	Ensemble learning models for accurate solar	HI-SEAS weather station data	XGBoost, LightGBM, CatBoost, NGBoost,	Ensemble learning significantly improves	Supports: Ensemble methods; Gap:

Ref.	Paper Focus	Dataset Used	Key Methods	Key Findings	Relevance to This Study & Gap Addressed
	irradiance forecasting		AdaBoost, GBM, RF, SGD Boost	forecast accuracy	Hawaii data; no Saudi validation
[11]	Forecasting solar radiation using ML tools	Meteorological parameters + historical solar radiation	ANN, SVM, RF, GBM	Ensemble methods (GBM, RF) outperform traditional approaches	Supports: Ensemble methods superiority; feature selection importance; Gap: No arid climate or dust analysis
[16]	PV solar radiation prediction using advanced ML and DL models	HI-SEAS dataset (32,686 entries, 11 parameters)	XGBoost, MLP, RF, SVR, GBM, KNN, LightGBM, CatBoost, LSTM, BiLSTM	Gradient Boosting and LightGBM perform well; LSTM/BiLSTM capture non-linear correlations	Supports: Comprehensive model comparison; feature engineering; Gap: Hawaii dataset; not Saudi arid context
[18]	Short-term GHI forecasting using weather classified categorical boosting	NSRDB (2 geographical locations)	RF (feature selection), K-means (clustering), CatBoost	WC-CB reduces errors; 16% RMSE reduction; 6% MAE reduction	Supports: Weather classification; significant error reduction; Gap: No Saudi context; no dust
[19]	Time series forecasting based on irradiance classification	GHI and on-site power data (10-min resolution)	Irradiance classification + GEKKO optimization	GEKKO optimized model outperforms other ML and ensemble models	Supports: Classification approach; Gap: No Saudi context; no dust
[20]	Comparative analysis of solar power forecasting using ML	Historical solar generation data	KNN, GBRT, LSTM, MLP	Preprocessing crucial for accuracy; precise	Supports: Data quality emphasis; Gap: No probabilistic

Ref.	Paper Focus	Dataset Used	Key Methods	Key Findings	Relevance to This Study & Gap Addressed
				forecasts achieved	forecasting; no dust analysis
[21]	Hybrid numerical ensemble method for solar irradiance forecasting	GFS data, clear sky index	ML + RNN models, parametric model, ensemble	Hybrid ensemble outperforms GFS; hourly classification reduces errors	Supports: Hybrid ensemble; probabilistic evaluation; Gap: No Saudi context; no dust
[22]	Solar irradiance forecasting using satellite and limited ground data	Satellite data + ground measurements	Tree-based regression, LSTM	Tree-based models map satellite to ground GHI; LSTM outperforms statistical baseline	Supports: Data-sparse regions; satellite data integration; Gap: No Saudi-specific; no dust
[23]	Hybrid ML technique for solar radiation prediction	Not specified	Hybrid: Random Forest + XGBoost	Hybrid model outperforms traditional models in accuracy	Supports: Hybrid/ensemble approaches; Gap: No Saudi context; no dust analysis
[24]	Smart solar forecasting: predictive models for radiation and energy generation	Historical irradiance records; real-time meteorological inputs	Data-driven ML platform with IoT sensing	Improved forecasting performance and adaptability to dynamic weather	Supports: Data-driven integration; real-time inputs; Gap: No experimental comparison; no dust
[25]	Solar radiation forecasting in Saudi Arabia	Weather dataset from KACARE (Riyadh)	ANN, Random Forest, Linear Regression	Random Forest achieved 92.86% accuracy, lowest RMSE (10.3157)	Supports: Saudi Arabia context; Random Forest effectiveness; Gap: Single location; limited model comparison

Ref.	Paper Focus	Dataset Used	Key Methods	Key Findings	Relevance to This Study & Gap Addressed
[26]	Day-ahead forecasting using statistical ML methods	60-month dataset from Thuwal, Saudi Arabia	Autoregressive KNN-regression (TSFKNN)	10-25% accuracy improvement; computationally light models	Supports: Saudi Arabia context; accurate and efficient models; Gap: Single method focus; no probabilistic uncertainty
[27]	Optimizing solar energy forecasting and site adjustment with ML	Meteorological + solar radiation data	XGBT, RF	39% accuracy improvement over traditional statistical methods	Supports: XGBT and RF effectiveness; Gap: No Saudi validation; no probabilistic
[28]	Survey of ML algorithms for solar irradiation forecasting	Satellite data from Meteosat 5 & 7	Deep Learning, Decision Tree	ML more effective than traditional methods; cloud pictures with physical models	Supports: ML for solar forecasting; Gap: No systematic benchmarking; lacks Saudi-specific validation
[29]	ML-based solar power forecasting techniques analysis and comparison	Not specified	LSTM, GRU, persistence baseline	LSTM and GRU outperformed persistence; data split technique proposed	Supports: RNN models; benchmarking approach; Gap: No arid climate analysis; no dust event study
[30]	Solar radiation forecasting using robust ML techniques	National Solar Radiation Database	MLR, SVR, DT, RFR, GBR, Lasso, Ridge, LSTM	LSTM least error (0.0040); GBR and RFR high accuracy across seasons	Supports: LSTM effectiveness; seasonal analysis; Gap: No Saudi-specific validation
[31]	AI-enhanced solar irradiance forecasting models	Not specified	ANN, SVM, SVR, Light GBM	AI-enhanced models improve forecasting accuracy	Supports: AI for solar forecasting; Gap: No experimental

Ref.	Paper Focus	Dataset Used	Key Methods	Key Findings	Relevance to This Study & Gap Addressed
				compared to traditional methods	comparison; no dust events
[32]	Multivariate optimal hybrid DL for day-ahead forecasting	Real-world solar irradiance + meteorological data	Data decomposition + deep learning with grid search	Outperforms traditional methods; lower RMSE and MAE	Supports: Hybrid DL; meteorological constraints; Gap: No Saudi context; no dust
[33]	ML-based forecasting of temperature and solar irradiance for PV systems	Temperature and solar irradiance data	DT, RF, XGBoost, SVM, etc.	ML improves accuracy over traditional meteorological models	Supports: ML for PV systems; Gap: No Saudi validation; no dust
[34]	Enhanced GHI forecasting using hybrid ML approach	Historical GHI data	Hybrid KNN-XGB	RMSE 24.18%, R^2 99.94%; outperforms individual algorithms	Supports: Hybrid methods; feature selection importance; Gap: No Saudi context; no dust
[35]	Meta-learning-based VotingRegressor for solar radiation prediction	NSRDB (3 years)	Meta-learning-optimized VotingRegressor	Significant RMSE/MAE reductions; high R^2 across diverse conditions	Supports: Meta-learning; ensemble optimization; Gap: No Saudi context; no dust
[36]	Ensemble voting for solar irradiation forecasting	Brazilian database; Nottingham Castle dataset	Ensemble voting: RF, CatBoost, etc.	Weighted average voting outperforms others; 6% MAE reduction	Supports: Ensemble voting; Gap: No Saudi context; no dust
[37]	Short-term PV power forecasting with advanced ML techniques	Not specified	SVM, XGBoost, DT, RF, AdaBoost	Insights on algorithm suitability; practical implications provided	Supports: Model suitability analysis; Gap: No Saudi validation; no dust

Ref.	Paper Focus	Dataset Used	Key Methods	Key Findings	Relevance to This Study & Gap Addressed
[38]	ML-based solar irradiance prediction using multi-site data	NELHA, ULL, HSU, RaZON+, UNLV, NWTC (6 PV stations)	GBR, MLP, CatBoost, stacking ensemble	$R^2 > 95\%$ at most stations; GBR best at 3 stations; MLP best at 2 stations	Supports: Multi-site data; feature identification; Gap: No Saudi context; no dust
[39]	Impact of weather variability on solar energy production forecasting	Not specified	Review of statistical, physical, ML, hybrid methods	Weather variability significantly impacts forecasting; need for ongoing research	Supports: Weather variability importance; Gap: No experimental contribution
[40]	Day-ahead solar irradiance forecasting with spatio-temporal context	Multi-modal dataset (satellite + time series)	Deep learning architecture (CrossViViT)	Accurate day-ahead forecasting; robust in cloudy conditions	Supports: Spatio-temporal context; uncertainty estimation; Gap: No Saudi validation; complex architecture

Table S2: Summary of Research Gaps and How This Study Addresses Them.

Research Gap Identified	Evidence from Literature	How This Study Addresses the Gap
Limited arid-climate benchmarking	Most studies use temperate climate data [3], [8], [13], [22]; Saudi studies focus on single models or specific locations [1], [6], [25], [26]	Comprehensive benchmarking of 6 ML models + persistence baseline using 4-year Riyadh dataset
Insufficient dust impact quantification	Few studies quantify dust effects [12], [13], [15]; no systematic multi-model analysis	Matched-condition comparison controlling for hour and season, showing no statistically significant difference between dusty and clean days

Research Gap Identified	Evidence from Literature	How This Study Addresses the Gap
Lack of probabilistic uncertainty for inverter applications	Probabilistic/uncertainty methods exist [21], [40], but not applied to grid-tied inverter control	Quantile regression and conformal prediction with coverage rates and interval widths for ramp-rate control
Seasonal analysis gap	Seasonal patterns noted [12], [15] but not systematically analyzed	Seasonal RMSE comparison across Winter, Spring, Summer, Autumn linked to dust cycle
Limited model diversity	Studies focus on 1-4 models [1], [2], [3], [10], [23]; reviews note limited comparisons [8], [17]	6 diverse models: Linear Regression, RF, XGBoost, GBM, SVR, LSTM + persistence baseline

Table S3. Final selected hyperparameters for the 3-feature track.

Model	Horizon	Selected hyperparameters
SVR	1h / 3h / 6h	C = 100, gamma = 'scale', epsilon = 0.1 (all horizons)
Random Forest	1h	max_depth=10, max_features='log2', min_samples_leaf=2, min_samples_split=10, n_estimators=200
Random Forest	3h	max_depth=10, max_features='sqrt', min_samples_leaf=2, min_samples_split=10, n_estimators=300
Random Forest	6h	max_depth=5, max_features=None, min_samples_leaf=4, min_samples_split=10, n_estimators=200
XGBoost	1h	colsample_bytree=1.0, learning_rate=0.05, max_depth=3, n_estimators=200, reg_alpha=0, reg_lambda=1.0, subsample=1.0
XGBoost	3h	colsample_bytree=0.8, learning_rate=0.2, max_depth=3, n_estimators=300, reg_alpha=1.0, reg_lambda=0.1, subsample=0.6
XGBoost	6h	colsample_bytree=1.0, learning_rate=0.05, max_depth=3, n_estimators=200, reg_alpha=0.5, reg_lambda=1.0, subsample=0.6
Gradient Boosting	1h	learning_rate=0.05, max_depth=3, min_samples_leaf=4, min_samples_split=2, n_estimators=300, subsample=0.8

Model	Horizon	Selected hyperparameters
Gradient Boosting	3h	learning_rate=0.05, max_depth=3, min_samples_leaf=4, min_samples_split=10, n_estimators=200, subsample=0.6
Gradient Boosting	6h	learning_rate=0.05, max_depth=5, min_samples_leaf=4, min_samples_split=2, n_estimators=100, subsample=0.6

Source: console log from the clean pipeline run of 2026-10-27. Values for Random Forest, XGBoost, and Gradient Boosting are the ones that minimize the mean TimeSeriesSplit RMSE across the 3-fold inner loop.

Table S4. Recursive feature elimination stability (50 bootstrap iterations, training data only).

Feature	10-feature selection (%)	Rank-10 hits (of 50)	3-feature selection (%)	Rank-3 hits (of 50)
hour_sin	100.0	50	100.0	**50**
hour_cos	100.0	50	100.0	**50**
GHI_lag_1	100.0	50	100.0	**50**
GHI_lag_2	100.0	50	0.0	0
GHI_lag_3	100.0	50	0.0	0
Temp	100.0	50	0.0	0
RH	100.0	50	0.0	0
doy_sin	98.0	49	0.0	0
GHI_lag_6	98.0	49	0.0	0
doy_cos	52.0	26	0.0	0
Pressure	26.0	13	0.0	0
Temp_lag_1	20.0	10	0.0	0
Temp_lag_2	4.0	2	0.0	0
RH_lag_1	2.0	1	0.0	0
dust_flag	0.0	0	0.0	0

Feature	10-feature selection (%)	Rank-10 hits (of 50)	3-feature selection (%)	Rank-3 hits (of 50)
Wind_Speed	0.0	0	0.0	0
Wind_Speed_lag_1	0.0	0	0.0	0
Wind_Speed_lag_2	0.0	0	0.0	0
RH_lag_2	0.0	0	0.0	0
Wind_roll_mean_3	0.0	0	0.0	0
Wind_roll_mean_6	0.0	0	0.0	0
Wind_roll_mean_12	0.0	0	0.0	0
Wind_roll_std_3	0.0	0	0.0	0
Wind_roll_std_6	0.0	0	0.0	0
Wind_roll_std_12	0.0	0	0.0	0

Two RFE configurations were run: n_features_to_select = 10 (left block) and n_features_to_select = 3 (right block), each with 50 bootstrap iterations on the 2013–2015 training data. Only hour_sin, hour_cos, *and* GHI_lag_1 *are selected in the top-3 in every iteration. The three features therefore constitute an unambiguous, stability-validated minimum feature set. The 3-feature baseline used in Track B and in the ablation study is derived from this procedure, not chosen post hoc.*

Table S5. Rolling-origin validation across two temporal windows.

Train window	Test year	Horizon	N_train	N_test	Persistence RMSE	LR RMSE	SVR RMSE	LR skill (%)	SVR skill (%)
≤2014	2015	1h	8,008	4,028	168.04	89.19	66.37	46.92	60.50
≤2014	2015	3h	6,603	3,318	443.51	102.32	79.18	76.93	82.15
≤2014	2015	6h	4,442	2,223	608.66	96.84	83.57	84.09	86.27
≤2015	2016	1h	12,036	4,034	170.15	87.21	65.43	48.74	61.55
≤2015	2016	3h	9,921	3,330	449.31	101.64	77.16	77.38	82.83
≤2015	2016	6h	6,665	2,232	615.01	97.21	81.02	84.19	86.83

*All RMSE values in W/m². SVR uses the 4-feature configuration (*hour_sin, hour_cos, GHI_lag_1, doy_cos*). The main-text Table 15 reports the mean ± standard deviation across the two windows; this table provides the per-window breakdown that those means and standard deviations are computed from.*

Table S6. Complete feature importance ranking and ablation effect.

Feature	LR coefficient	RF importance	XGB importance	Ablation Δ at 1h (W/m²)	Ablation Δ at 3h (W/m²)	Ablation Δ at 6h (W/m²)
hour_cos	−71.23	0.6294	**0.9020**	+0.11 (removal)	−0.17 (removal)	−0.03 (removal)
GHI_lag_1	**276.68**	0.3080	0.0579	**+76.14** (removal)	**+58.72** (removal)	**+42.01** (removal)
hour_sin	59.71	0.0318	0.0185	**+50.34** (removal)	**+114.97** (removal)	**+14.44** (removal)
GHI_lag_6	14.76	0.0055	0.0028	—	—	—
RH	−35.61	0.0050	0.0034	—	—	—
GHI_lag_2	−21.01	0.0048	0.0023	—	—	—
GHI_lag_3	−53.26	0.0032	0.0011	—	—	—
Temp	230.13	0.0030	0.0017	**−2.30 (addition)**	**−2.81 (addition)**	**−2.72 (addition)**
Pressure	−2.49	0.0012	0.0007	—	—	—
doy_cos	−3.64	0.0010	0.0007	**−0.93 (addition)**	**−1.56 (addition)**	**−0.62 (addition)**
doy_sin	6.74	0.0010	0.0006	**−1.00 (addition)**	**−2.32 (addition)**	**−2.80 (addition)**
Wind_roll_std_3	1.39	0.0006	0.0006	—	—	—
Temp_lag_1	−272.33	0.0006	0.0010	—	—	—
Wind_roll_std_12	−1.57	0.0006	0.0004	—	—	—
RH_lag_2	−13.83	0.0006	0.0006	—	—	—
RH_lag_1	37.53	0.0006	0.0006	—	—	—

Feature	LR coefficient	RF importance	XGB importance	Ablation Δ at 1h (W/m²)	Ablation Δ at 3h (W/m²)	Ablation Δ at 6h (W/m²)
Wind_roll_std_6	−2.06	0.0006	0.0005	—	—	—
Temp_lag_2	53.54	0.0005	0.0005	—	—	—
Wind_Speed	12.39	0.0004	0.0007	—	—	—
Wind_roll_mean_12	−0.08	0.0004	0.0004	—	—	—
Wind_Speed_lag_2	2.04	0.0003	0.0006	—	—	—
Wind_Speed_lag_1	−7.74	0.0003	0.0005	—	—	—
Wind_roll_mean_6	−7.83	0.0003	0.0005	—	—	—
Wind_roll_mean_3	2.35	0.0003	0.0005	—	—	—
dust_flag	0.34	**0.0001**	0.0006	—	—	—

SVR base model with the 3-feature baseline (hour_sin, hour_cos, GHI_lag_1) at 66.36 / 78.72 / 81.64 W/m² at 1-, 3-, 6-hour horizons. Δ is the change in RMSE relative to the baseline when the feature is removed (positive Δ = removal worsens accuracy) or added (negative Δ = addition improves accuracy). Dash = not individually ablated; the three-horizon full ablation appears in main-text Table 12. The dust flag is the least important feature across all three importance measures — RF importance 0.0001, five orders of magnitude below hour_cos.

Operational note on the choice of fourth feature. The ablation study shows that adding Temp gives the best RMSE improvement at all three horizons (0.9–2.8 W/m²), followed by doy_sin, with doy_cos third. All three variants reduce RMSE by 1–3% relative to the 3-feature baseline and all three fall within the bootstrap 95% CI of the canonical doy_cos model ([61.50, 68.91] at 1h). The paper adopts doy_cos as the canonical fourth feature because it is **deterministic** — it depends only on the timestamp and requires no sensor — and therefore transfers to any site without additional instrumentation. The Temp variant would be preferable in a deployment that already maintains a temperature sensor; the doy_sin variant is marginally better at the 6-hour horizon. None of the three is a significantly worse choice than the others within the reported confidence band.