

A Wide-and-Deep-Based Time Sequence Model for Predicting Power Outages Caused by Extreme Winter Storms

Jikun Liu^a, Zhe Zhang^{a*}, Yuhan Cheng^b, Jangjae Lee^a, Stephanie Paal^a, Diya Li^a

^a Texas A&M University, Jikun Liu - jikun@tamu.edu, Zhe Zhang - zhezhang@tamu.edu, Jangjae Lee - jangjlee7@tamu.edu, Stephanie Paal - sgpaal@tamu.edu, Diya Li - diya.li@tamu.edu

^b Meta USA, Yuhan Cheng - yuhancheng@meta.com

* Corresponding Author

Abstract: In February 2021, Winter Storm Uri caused widespread power outages across Texas, affecting over 5 million people and resulting in an estimated \$190 billion in damages. To support extreme weather outage resilience, this study introduces the Wide-and-Deep-Based Time Sequence Algorithm (WD TSA) for predicting power outage severity. The model combines a deep bidirectional LSTM for time-lagged weather and outage history with a wide pathway for weakly temporal features, enabling synergistic integration of heterogeneous inputs. This dual pathway design significantly outperforms standard baselines, achieving 0.99 accuracy at coarse resolution ($K=3$) and 0.84 at fine granularity ($K=15$), utilizing fewer parameters than expanded LSTM alternatives. Ablation and comparative analyses confirm that the performance gains arise from specialized feature routing and non-additive synergy between input groups, growingly so under complex classification tasks. County-level visualizations during Winter Storm Uri are provided to illustrate the model's ability to anticipate outage progression, offering actionable forecasts for emergency planning. While current validation focuses on extreme events and does not offer spatial dependency modeling, the framework provides a compact and flexible foundation for resilient grid operations and targeted response in potential weather-induced disruptions.

Keywords: Power outage prediction, Wide-and-Deep neural network, Bidirectional LSTM, Multi-class classification, Extreme weather resilience, Spatiotemporal modeling, Feature routing, Geospatial Artificial Intelligence, GeoAI

1. Introduction

In February 2021, Winter Storm Uri severely impacted the southern United States, affecting states such as Texas, Mississippi, and Louisiana (City of Austin & Travis, 2021). The storm brought snow, ice, and ultralow temperatures, resulting in large-scale power outages and an estimated \$190 billion in damages (Austin Water, 2021). Over five million individuals experienced prolonged blackouts, with many relying on rudimentary heating sources for survival (City of Austin & Travis, 2021). According to the U.S. Energy Information Administration, approximately 45% of the System Average Interruption Duration Index (SAIDI) is attributed to major events, with this percentage rising to 76% in 2020 and 72% in 2021 (EIA, 2022). These statistics highlight the unyielding need for predictive systems to anticipate and mitigate the impact of extreme weather-induced outages.

Accurate, timely, and high-resolution power outage forecasting supports emergency preparedness by providing prioritization for grid stabilization and resource deployment (Zhang et al., 2020, Cheng et al., 2024, Li et al., 2024). A growing body of Machine Learning (ML) work has pursued this objective: random forest frameworks for hurricane impacts across the Eastern United States (Taylor et al., 2023), support vector and recurrent models for extratropical storms in Northern Europe (Tervo et al., 2021),

and comprehensive surveys of hurricane studies (Fatima et al., 2024). Yet, as these reviews and case studies repeatedly note, accuracy degrades when forecasts are required at finer spatial granularity or longer lead times, and computational costs often exceed what is deployable in real time.

Several knowledge gaps motivate the present study. First, most existing models channel heterogeneous inputs through a single pathway, despite their markedly different temporal signatures, while an observable accuracy drop occurs at high class granularity (Fatima et al., 2024, Taylor et al., 2023, Tervo et al., 2021). Secondly, early spatial generalized linear mixed models demonstrated that neighboring assets experience correlated fragility during storms, but contemporary deep learners do not integrate spatial and temporal dependencies, jeopardizing their reliability when storm tracks deviate from the training climatology (Liu et al., 2008). Finally, recent reviews focused on uncertainty quantification highlight overconfident long-range forecasts and the absence of probabilistic severity bounds (Arora and Ceferino, 2023). Some ML studies indicate that operational adoption demands lightweight predictors to support rapid conditional analyses (Hughes et al., 2024).

To address these gaps, this study proposes the Wide-and-Deep Time-Sequence Algorithm (WD TSA), a hybrid architecture that routes strongly temporal features through a bidirectional long-short-term-memory deep

branch while static or weakly temporal inputs flow through a parallel wide layer. The fusion network contains roughly 15,000 parameters and delivers county-level, 10-hour severity forecasts that outperform deeper recurrent models with simple designs by up to 57 percentage points. With improved spatial granularity, temporal depth, and computational efficiency, WDTSA provides a practical tool for real-time resilient energy management.

2. Background

Power outages represent a growing threat to public safety and economic stability in the United States. Electricity consumption has increased significantly, rising from 0.3 trillion kWh in 1950 to over 4.1 trillion kWh in 2022 (EIA, 2023). While indirect economic losses are considered difficult to quantify, estimates suggest that outages cost the U.S. economy \$150 billion annually, with the 2021 Texas blackout alone accounting for \$664 million in direct losses (Bhattacharyya and Hastak, 2022, Energy, 2018, Shuai et al., 2018). From 2000 to 2021, 83% of major outages were caused by weather-related events, including winter weather (22%), tropical cyclones (15%), and severe weather such as high winds and thunderstorms (58%) (Central Climate, 2022, USDOE, 2023). The largest outage reports came from states including Texas, Michigan, California, and North Carolina. During Uri, the Electric Reliability Council of Texas (ERCOT) reported that generation deficits reached 34,000 MW from February 15–17, constituting nearly half the winter peak grid load (FERC, 2021).

To forecast outages, researchers have employed both traditional statistical models and ML approaches. Generalized additive models (GAMs) have been applied to grid-level outage prediction using variables including wind speed, precipitation, infrastructure, and land cover, outperforming generalized linear models (GLMs) in hurricane contexts (Han et al., 2009a,b). Subsequent work introduced random forests that avoided the need for detailed grid component inventories and emphasized the predictive value of historical outage data (Roshanak Nateghi, 2013, Raicharoen et al., 2003).

Given the spatiotemporal nature of outages, time-series techniques such as ARIMA have been used to model temporal dynamics (Ho and Xie, 1998, Chen et al., 2008, Box et al., 2015). However, Autoregressive Integrated Moving Average (ARIMA) assumes stationarity and often requires differencing. In contrast, Long Short-Term Memory (LSTM) networks support nonlinear modeling of non-stationary sequences without preprocessing and are suited for outage prediction (Sherstinsky, 2020, Gonzalez and Yu, 2018).

Recent research further motivates this approach by highlighting the diverse datasets used in outage prediction, including weather but also considering infrastructure, vegetation, and socioeconomic indicators. For example, Yang et al. (2020) proposed an Event Severity Classification task to manage data imbalance in outage events, achieving high F1 scores. Xu et al. (2023) used nighttime satellite imagery and socioeconomic variables to detect outages and

assess environmental injustice. D'Amico et al. (2019) incorporated tree species into a hurricane outage prediction model, identifying specific species associated with more prominent outage risks.

Deep learning models, especially LSTMs, have been shown to perform well in forecasting electric loads (Zheng et al., 2017) and geomagnetic storm patterns (Tang et al., 2020). The adaptability of LSTM to non-seasonal, non-stationary sequences makes it a strong candidate for short-term outage prediction. LSTM was introduced by Hochreiter and Schmidhuber (1997) to address the problem of vanishing gradients in recurrent neural networks (RNNs) and is widely used in time series modeling due to its gated memory architecture. The original LSTM remains a common baseline (Greff et al., 2017, Van Houdt et al., 2020). Deep bidirectional LSTM (DB-LSTM) architectures, in which inputs are processed in both temporal directions across multiple layers, have been shown to capture time dependencies at different resolutions more effectively (Sak et al., 2014). Still, LSTMs are susceptible to overfitting in data-sparse environments (Gal and Ghahramani, 2016).

In the proposed architecture, the deep LSTM branch processes temporally prominent features, while the wide branch handles relatively static inputs. This dual-input structure enables flexible modeling of diverse feature types while supporting integration at the fusion layer. The final model accommodates inputs with differing not only shapes and scales but also distributional contexts, seeking improvements in overall performance.

3. Methodology

3.1 Problem Formulation

The power outage prediction task is formulated as a multi-class classification problem (e.g., based on different severity levels) as it better reflects the need for prioritization. The objective is to predict the severity of future outages using historical outage data and meteorological measurements. Given a sequence of observations $\mathbf{X} = \{x_{t-9}, \dots, x_{t-1}\}$, where each x_t is a vector of lagged features at time t , the model learns a mapping $f: \mathbf{X} \rightarrow \mathcal{Y}$, with $\mathcal{Y} = \{0, 1, \dots, K-1\}$ representing K severity classes. The Prediction Target is the percentage (pctg) of customers without power at time t_0 , transformed as:

$$\log_outage = \log(1 + \text{pctg_outage}_{t_0})$$

Then discretized into K classes with equal-size quantiles:

$$q_i = \frac{i}{K}, \quad b_i = Q(q_i, \log_outage)$$

thus $\text{class}(y) = i$ such that $b_i \leq \log(1 + y) < b_{i+1}$.

Classification accuracy is used as the primary metric. The models are compared with several baselines to be introduced in more detail in section 3.4. Granularity level K is varied to reflect different planning needs: low K for strategic planning, medium K for the deployment of tactical crews, and high K for real-time management. Synergistic learning is used to hypothesize that routing features

through specialized wide and deep pathways improves performance beyond their summed contribution. This procedure motivates the architectural fusion between temporal and non-temporal features.

3.2 Data Description and Preprocessing

The dataset consists of timestamp-aligned outage and weather data across Texas counties during the 2021 Winterstorm Uri period (Feb 13-23, 2021) from 207 unique counties. Specifically, the weather data include the following features (1) temperature, (2) dew points, (3) humidity, (4) wind speed, (5) wind gust, and (6) pressure and are matched with raw outage percentage data at a 3-hour interval at the county level in Texas. From here, each sample is constructed by extracting a 10-hour sliding window of lagged outage and weather features, forming input vectors such as:

$$\text{row}_t = \left[t_{-9}^{f_1}, \dots, t_0^{f_1}, t_{-9}^{f_2}, \dots, t_0^{f_2}, \dots \right]$$

where $t_i^{f_j}$ represents the feature value at timestamp i of feature j .

The outage percentage at t_0 is log-transformed and universally binned into K quantile-based classes using thresholds from the entire dataset, thus ensuring highly balanced class distributions across classification granularity. The dataset is partitioned on a 4:1 stratified split with class balance preservation but no temporal awareness. All preprocessing steps are equally applied to training and validation data without leakage. There are no missing values.

3.3 Wide-and-Deep Architecture

3.3.1 Architecture Design

Illustrated by Figure 1, the model follows a two-branch architecture:

1. The **deep path** processes sequential features through a bidirectional LSTM, followed by a dense layer, batch normalization, and dropout.
2. The **wide path** handles static features through a dense layer and batch normalization.

The two outputs are concatenated in a fusion layer and passed through one additional dense layer to produce final-class logits. The FEATURE GROUP element represents a specific variable that directs how each feature is directed into deep or wide channels, e.g. pressure measurement into wide path, historical outage percentages into deep path, etc.

3.3.2 Feature Routing Strategy

Features are routed to the deep or wide pathway based on empirical evaluations. Those exhibiting strong temporal dependencies are assigned to the LSTM path, while the weakly time-correlated ones are routed to the wide path. This strategy seeks to align data characteristics with processing pathways and thus enhance representation. The features are finally collected in a fusion layer in the final

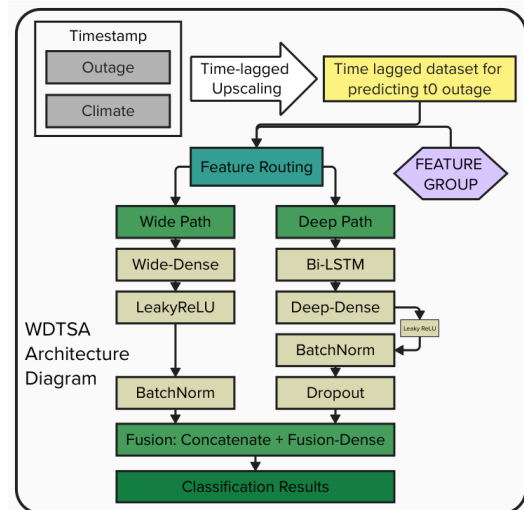


Figure 1. WDTSA Architecture Diagram

stage to generate synergistic gains. In particular, the evaluation provides a routing preference that directs:

wind gust, pressure → wide path

temperature, dew point, humidity, wind speed,

outage pctg → deep path

This routing scheme is thereon held constant throughout the study.

3.4 Comparative Models

To evaluate the contribution of architectural components and input modalities, this study compares the Wide-and-Deep model with a set of alternative models. The model-reduction baselines are derived by restricting the input feature set, and there are also comparisons drawn between architecturally different models.

- **Pure LSTM:** This model takes all features through a bidirectional LSTM without a wide pathway, subsequently connected to a dense layer for predictions.
- **Extended LSTM:** This model builds upon the Pure LSTM architecture by incorporating additional LSTM units and an extra dense layer, resulting in a significantly larger number of parameters.
- **Linear Model:** This model implements a simple Logistic Regression classifier. Rationales and analyses are given in Section 4.4.
- **Outage Model:** This simplified variant uses only lagged outage percentage values as input, routed through the LSTM path.
- **Weather Model:** This simplified variant excludes all outage-related features, relying solely on weather variables routed through their respective paths. The model evaluates the contribution of the meteorological input to the decision.
- **T1 Model:** This simplified variant uses only the t_{-1} value of outage percentage as input, routed through the deep path. It tests how much short-term persistence contributes to prediction performance.

3.5 Experimental Framework

Additional analyses, including a temporal ablation study and spatiotemporal visualization, are presented in Section 4.7 and Section 4.9, respectively. These use consistent architecture and training protocols unless otherwise noted. Specifically, all models are trained using the Adam optimizer with a learning rate of 10^{-3} and a mini-batch size of 32 (Kingma and Ba, 2015). Early stopping is applied based on validation accuracy to prevent overfitting. Cross-entropy loss is used for multi-class classification. All models use a 10-hour input window (t_{-9} to t_0) for each feature. The LSTM components have 16 hidden units per direction, and all dense layers use 64 units with ReLU or Leaky ReLU activations, followed by batch normalization and dropout.

Model performance is assessed using classification accuracy on the validation set. To evaluate robustness across different levels of task granularity, experiments are conducted with multiple values of K (the number of severity classes), selected from the set $\{3, 5, 7, 10, 15\}$. Random seeds are set to ensure reproducibility. Additionally, the data split, quantile thresholds, and feature routing configurations are kept consistent across all model comparisons.

4. Results

All models follow the shared training and evaluation protocol described in Section 3.5. Validation accuracy is reported using the best model checkpoint per run with a 30-epoch cap.

4.1 Multi-Class Performance Comparison

Validation accuracy across different numbers of classification bins is summarized in Table 1 and illustrated in Figure 2. As the number of classes increases, accuracy decreases across all models. However, the Wide-and-Deep model maintains substantially higher performance across all values of K , with accuracy remaining above 0.84 at the finest granularity of 15 classes.

Table 1. Validation accuracy by model and number of classes (K).

K	W&D	Outage	Weather	T1	P-LSTM
3	0.994	0.861	0.637	0.865	0.855
5	0.970	0.784	0.448	0.787	0.725
7	0.934	0.683	0.351	0.688	0.640
10	0.896	0.611	0.272	0.618	0.502
15	0.844	0.543	0.184	0.547	0.391

Several notable patterns emerge from these results. First, the Outage and T1 models perform remarkably similarly across all class counts, with differences of less than 0.015. This suggests that recent outage history (t-1) contains most of the predictive signal, with potentially minimal incremental value from earlier time points. Second, the Weather model shows the steepest accuracy degradation as K increases, dropping from 0.637 at $K = 3$ to 0.184 at $K = 15$. This indicates that while weather variables contribute

meaningful signal, they struggle in the classification task when employed alone without further modeling complexity.

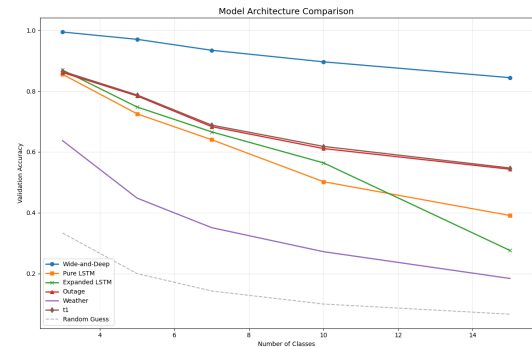


Figure 2. Validation accuracy across number of classes for each model variant.

Noticeably, despite access to the complete feature set, the Pure LSTM model substantially underperforms the Wide-and-Deep architecture across all classification granularity. The performance gap widens dramatically with task complexity, from 0.139 percentage points at $K = 3$ to 0.453 at $K = 15$. This suggests that routing all features through a single sequential pathway creates a representational bottleneck that becomes increasingly problematic as the classification task grows more complex.

4.2 Architecture Advantage Analysis

The relative performance gain of the Wide-and-Deep model over the Pure LSTM architecture is presented in Figure 3. The absolute accuracy difference increases consistently with task granularity, from 0.139 at $K = 3$ to 0.453 at $K = 15$. This widening margin suggests that architectural advantages become more pronounced as the classification task grows more complex.

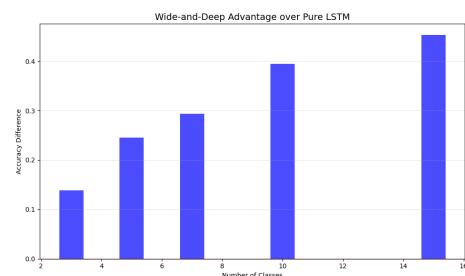


Figure 3. Accuracy difference between Wide-and-Deep and Pure LSTM across class granularity.

The advantage grows somewhat linearly with increasing class counts. While the Wide-and-Deep model maintains relatively stable performance across granularity (decreasing from 0.994 to 0.844), the Pure LSTM shows substantially steeper degradation (from 0.855 to 0.391). This divergent behavior highlights the structural limitation of processing heterogeneous features through a single pathway when attempting fine-grained discrimination tasks.

4.3 Model Capacity vs. Architectural Design

To investigate whether the Wide-and-Deep advantage stems from greater model capacity, this study compared it against an Expanded LSTM model with 43,139 parameters—nearly three times more than the Wide-and-Deep architecture’s 15,107 parameters. Table 2 presents the results of this comparison.

Table 2. Comparison between Wide-and-Deep and an expanded LSTM model with 2.9x more parameters.

K	W&D (15,107)	Expanded LSTM (43,139)	Advantage
3	0.994	0.869	0.125
5	0.970	0.748	0.223
7	0.934	0.666	0.268
10	0.896	0.564	0.332
15	0.844	0.276	0.568

Despite having only 35% of the parameters, the Wide-and-Deep model significantly outperforms the Expanded LSTM across all classification granularity. The performance gap widens dramatically from 0.125 at $K = 3$ to 0.568 at $K = 15$. This provides evidence that the architectural advantage stems from the dual-pathway design rather than model capacity, or equivalently, feature routing could bring more merits than raw parameter count for this prediction task.

4.4 Comparison Against Full Linear Model

Since the formulation of the problem is based on quantile borders, it is natural to investigate how a linear logistic classifier would work on seemingly clear borders. To evaluate the importance of nonlinear modeling capacity, this study compares the Wide-and-Deep model against a full linear baseline trained on the same input features. Table 3 reports classification accuracy across $K \in \{3, 5, 7, 10, 15\}$ and highlights the W&D advantage.

Table 3. Comparison of Wide-and-Deep model vs. full linear classifier.

K	W&D	Linear	Advantage
3	0.994	0.990	0.004
5	0.970	0.958	0.012
7	0.934	0.890	0.044
10	0.896	0.772	0.124
15	0.844	0.687	0.157

The performance gap grows sharply with task complexity: from 0.004 at $K = 3$ to 0.157 at $K = 15$. These results suggest that the Wide-and-Deep model captures higher-order, nonlinear interactions that are increasingly important for accurate classification as granularity increases.

4.5 Feature Contribution Analysis

Figure 4 quantifies the contribution of features in the W&D model. The accuracy gain attributed to all features increases significantly with class granularity, reaching an

asymptotic bound after $K = 10$, while the contribution of extended outage history relative to a t_1 model remains nearly flat. Outage history contributes the largest share across all values of K .

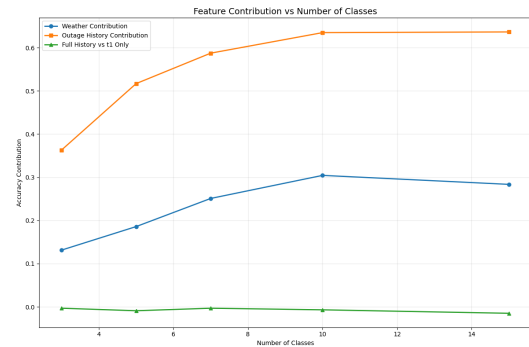


Figure 4. Estimated contribution of input components as a function of class granularity.

4.6 Synergistic Processing Evidence

The Wide-and-Deep model is evaluated against an adjusted additive baseline, calculated by summing the accuracies of the Outage-Only and Weather-Only models and subtracting the random guess rate ($1/K$), in accordance with the synergy hypothesis. Table 4 presents the measured accuracy of the Wide-and-Deep model and the corresponding baseline, along with the resulting synergy margin.

Table 4. Synergistic contribution of the Wide-and-Deep model compared to adjusted additive baseline.

K	W&D Acc	Outage + Weather - $1/K$	Synergy
3	0.994	1.165	-0.171
5	0.970	1.032	-0.062
7	0.934	0.891	0.043
10	0.896	0.783	0.113
15	0.844	0.661	0.183

The synergy margin becomes positive at $K = 7$ and increases with granularity, reaching a maximum of 0.183 at $K = 15$. This trend supports the hypothesis that specialized processing pathways facilitate non-additive feature interaction, particularly under higher task complexity. The negative evaluation when $K=3$ and 5 is not indicative of negative synergies but is due to high outage + weather model performance baselines. This interpretation is also supported by the feature contribution analysis (Figure 4). As K increases, the relative contribution of both outage history and weather variables increases despite the overall drop in classification accuracy, indicating predictive utility is only maintained when these inputs are fused. Therefore, the Wide-and-Deep architecture does not merely combine inputs but integrates them in a way that becomes increasingly important as the classification task becomes more difficult.

4.7 Temporal Ablation Analysis

To investigate the influence of historical window size on prediction accuracy, a temporal ablation study is conducted with a fixed granularity of $K = 10$ classes and a varying

maximum time lag from $t - 1$ to $t - 9$. Table 5 presents the validation accuracies across different temporal windows.

Table 5. Performance comparison across different temporal windows ($K=10$).

Max Lag	Accuracy	Deep Features	Wide Features
$t - 1$	0.929	9	4
$t - 3$	0.883	19	8
$t - 5$	0.891	29	12
$t - 7$	0.911	39	16
$t - 9$	0.888	49	20

The highest accuracy (0.929) was achieved using only the most recent time step ($t - 1$), potentially suggesting that the immediate past contains sufficient information for short-term outage prediction. The performance pattern is non-monotonic with a U-shape and a secondary peak at $t - 7$ (0.911). This secondary peak may suggest weekly patterns in power usage or grid conditions, though the relatively small differences between configurations, i.e. maximum spread of 0.046 points, suggest that the specific time window is considerably less influential than the architectural design itself. These results should be interpreted with appropriate caution, as the ablation experiments were limited to 30 training epochs per configuration and represent single runs rather than averages across multiple initializations.

4.8 Summary of Findings

Confusion matrices for $K = 3, 5, 7, 10$, and 15 are shown in Figures 5–9. For low values of K , misclassifications are mostly between adjacent severity classes. As K increases, the error becomes more diffuse, with confusion spanning multiple non-adjacent classes. Note that due to nondeterministic behavior, the validation accuracies and errors presented are slightly different from those previously reported without misrepresenting significant patterns.

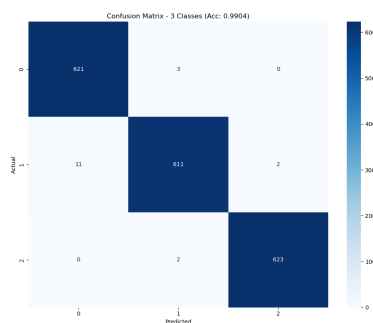


Figure 5. Confusion matrix for $K = 3$ (Accuracy: 0.9904).

In summary, across all classification granularity, the Wide-and-Deep model consistently achieves the highest validation accuracy, with performance remaining strong at increasing values of K . The Outage-only and T1-only baselines yield similar results, highlighting the strong short-term temporal persistence in outage data. In contrast, the Weather-only baseline contributes less predictive power individually but complements outage history with synergistic behavior when combined. The synergy margin increases

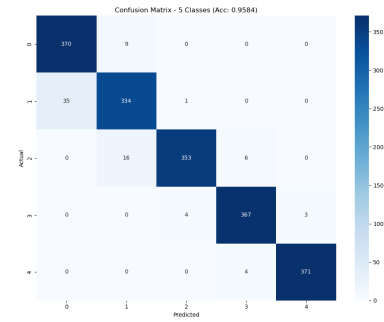


Figure 6. Confusion matrix for $K = 5$ (Accuracy: 0.9584).

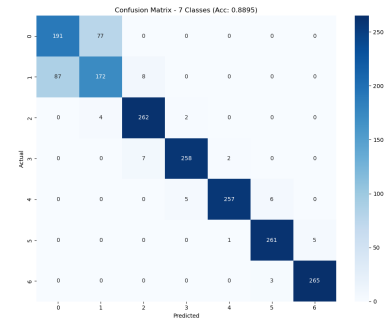


Figure 7. Confusion matrix for $K = 7$ (Accuracy: 0.8895).

with task complexity, supporting the architectural design hypothesis. Confusion matrices show tight class adherence at low K and diffuse errors at higher resolutions.

4.9 Spatiotemporal Visualization of Outage Forecasts

To provide references for the real-world applicability of the proposed model, map visualization of predicted outage severity across Texas counties during the days leading up to the peak of Winter Storm Uri is provided. These visualizations are generated using the best-performing Wide-and-Deep model checkpoint for $K = 10$, trained under the same architecture and hyperparameter configuration as the previous comparative experiments, with extended 100 epochs and multiple runs.

Figure 10 displays predicted outage class distributions for February 16 to 21, 2021, reflecting how the model forecasts evolved as the storm progressed. Higher class indices correspond to more severe outages, based on quantile-transformed outage percentages.

The spatial progression of predicted outages aligns with the timeline of storm escalation, showing increasing intensity across central and southeastern Texas counties. These visualizations highlight the utility of the Wide-and-Deep framework in producing interpretable geospatial forecasts that can support operational decision-making during extreme weather events.

5. Discussion

5.1 Architectural Impact and Synergistic Processing

The empirical findings demonstrate that model architecture plays a critical role in determining predictive performance. The Wide-and-Deep model, with approximately

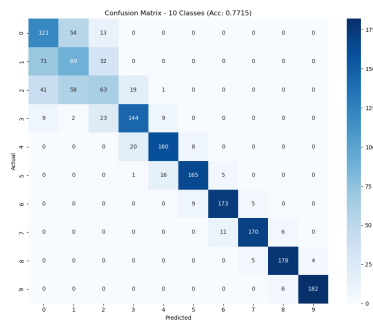


Figure 8. Confusion matrix for $K = 10$ (Accuracy: 0.7715).

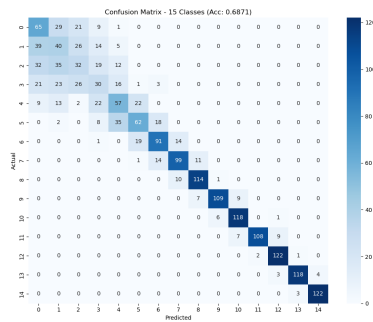


Figure 9. Confusion matrix for $K = 15$ (Accuracy: 0.6871).

15,000 parameters, significantly outperforms both a standard Pure LSTM (5,635 parameters) and an Expanded LSTM (43,139 parameters). The performance gap not only persists but grows with classification granularity, reaching a 0.568 difference at $K = 15$ when compared to the Expanded LSTM. This outcome indicates a representational limitation in single-pathway models when processing heterogeneous inputs. The Wide-and-Deep model achieves higher accuracy with fewer parameters by routing inputs through task-appropriate processing streams.

The design of the Wide-and-Deep framework is guided by the hypothesis that separating temporal and non-temporal feature processing facilitates richer representations. Empirical routing of input variables confirms that time-sensitive features, such as outage history, benefit from sequential modeling, whereas more static features, such as barometric pressure or wind gusts, perform better under non-temporal transformations. Evidence for synergistic processing is presented from comparisons with summed baselines. The synergy margin becomes positive at $K = 7$ and increases with task complexity. This pattern suggests that the Wide-and-Deep model captures non-trivial interactions between input types.

5.2 Operational Relevance and Spatial Generalization

The proposed model demonstrates significant practical value in outage forecasting tasks. Its performance remains robust across a wide range of classification granularity, supporting a flexible tiered-response framework for operational planning. At $K = 3$, accuracy exceeds 99%,

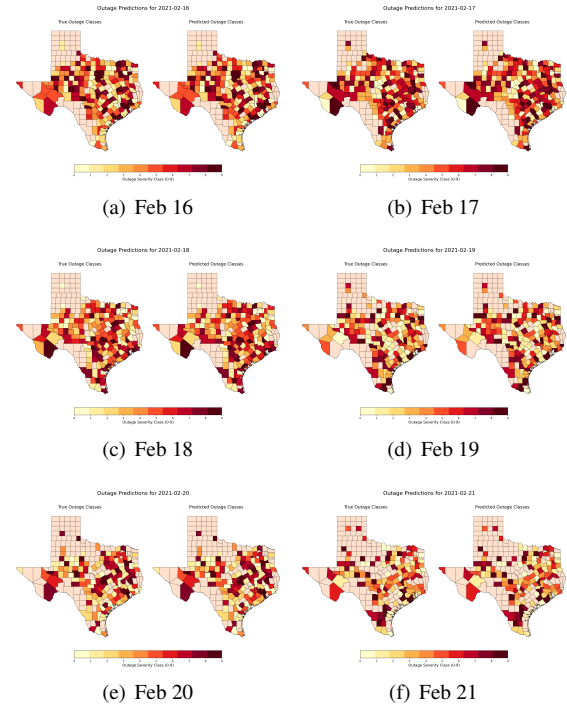


Figure 10. Predicted outage severity class across Texas counties during Winter Storm Uri, using Wide-and-Deep model with $K = 10$.

supporting strategic resource allocation, while at $K = 15$, the model maintains 84% accuracy, offering fine-grained resolution suitable for real-time triage and emergency response. High predictive fidelity indicates the model's potential utility in real-world settings, such as dispatching crews, managing grid load, and informing public advisories during weather-induced disruptions.

5.3 Limitations and Future Research Directions

Several avenues remain promising for future research. First, the current approach does not account for spatial dependencies; incorporating techniques such as graph neural networks or spatial autocorrelation constraints could improve predictive performance by leveraging geographic structure. Second, the model employs a fixed input window size across all features, but allowing dynamic window sizes may better capture temporal heterogeneity and enhance model synergy. Third, although the model demonstrates strong performance, it lacks explicit uncertainty quantification, an important consideration for applications in high-stakes decision-making contexts.

A limitation is that the current model evaluation focuses primarily on extreme weather scenarios. The model's false positive rate during normal weather conditions remains untested due to data acquisition challenges. The extensive computational and API resources required to collect historical weather data for all 254 Texas counties over extended non-event periods rendered comprehensive testing impractical within the study's scope. Establishing the model's specificity by verifying its ability to correctly identify non-emergency conditions and thus preventing alert fatigue and resource misallocation is an important next step.

Ethical considerations should also be addressed when deploying predictive systems in critical infrastructure contexts. Historical data may encode systemic disparities, and without appropriate oversight, model-driven decisions could perpetuate existing inequities. Future implementations should consider including transparent model governance, domain-informed validation, and explainability enhancement mechanisms for community feedback to ensure fair and responsible use.

6. Conclusion

This study presents a Wide-and-Deep neural network architecture for predicting power outage severity using inputs from lagged weather and outage history data on an hourly level. An empirically informed routing strategy assigns features to temporal and non-temporal processing paths and facilitates complementary, synergistic representation learning.

Through systematic evaluation across classification granularity, the proposed architecture demonstrates consistent improvements over baseline models, including those restricted to outage history, weather data, or short-term lag inputs. The architecture supports synergistic feature integration, proven by performance gains relative to additive input baselines.

The framework is modular and adaptable to different forecasting horizons and regional contexts. All code and data processing routines are implemented with reproducibility and support extension to broader grid resilience and risk management applications.

Data Availability

Outage data are publicly available from the ORNL EAGLE-I platform. Weather data were obtained via the commercial WeatherAPI service and cannot be redistributed. A processed dataset (`train_data_reconstructed.csv`) and associated code will be made available at <https://github.com/jikuntamu/WDTSA>.

References

- Arora, P. and Ceferino, L., 2023. Probabilistic and machine learning methods for uncertainty quantification in power outage prediction due to extreme events. *Natural Hazards and Earth System Sciences* 23(5), pp. 1665–1683.
- Austin Water, C., 2021. Winter storm uri after action report. Journal article.
- Bhattacharyya, A. and Hastak, M., 2022. Indirect cost estimation of winter storm-induced power outage in texas. *Journal of Management in Engineering* 38(6), pp. 04022057. doi: 10.1061/(ASCE)ME.1943-5479.0001084.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M., 2015. *Time Series Analysis: Forecasting and Control, 5th Edition*. 5th edition edn, John Wiley and Sons Inc.
- Central Climate, 2022. Surging power outages and climate change. Report.
- Chen, P., Yuan, H. and Shu, X., 2008. Forecasting crime using the arima model. In: *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, Vol. 5, pp. 627–630.
- Cheng, Y., Hu, B. and Du, J., 2024. Neural edge-cloud computing with information cascade attack. pp. 868–873.
- City of Austin & Travis, C., 2021. 2021 winter storm uri after-action review findings report. Report.
- D’Amico, D. F., Quiring, S. M., Maderia, C. M. and McRoberts, D. B., 2019. Improving the hurricane outage prediction model by including tree species. *Climate Risk Management*.
- EIA, 2022. Electric power annual 2022. Report, Energy Information Administration.
- EIA, 2023. Annual energy outlook. Report, US Energy Information Administration.
- Energy, O. o. N., 2018. Department of energy report explores u.s. advanced small modular reactors to boost grid resiliency.
- Fatima, K., Shareef, H., Costa, F. B., Bajwa, A. A. and Wong, L. A., 2024. Machine learning for power outage prediction during hurricanes: An extensive review. *Engineering Applications of Artificial Intelligence* 133, pp. 108056.
- FERC, 2021. The february 2021 cold weather outages in texas and the south central united states | ferc, nerc and regional entity staff report. Report, Federal Energy Regulatory Commission.
- Gal, Y. and Ghahramani, Z., 2016. A theoretically grounded application of dropout in recurrent neural networks. In: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon and R. Garnett (eds), *Advances in Neural Information Processing Systems*, Vol. 29, Curran Associates, Inc.
- Gonzalez, J. and Yu, W., 2018. Non-linear system modeling using lstm neural networks. *IFAC-PapersOnLine* 51(13), pp. 485–489.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R. and Schmidhuber, J., 2017. Lstm: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems* 28(10), pp. 2222–2232.
- Han, S.-R., Guikema, S. D. and Quiring, S. M., 2009a. Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models. *Risk Analysis*.
- Han, S.-R., Guikema, S. D., Quiring, S. M., Lee, K.-H., Rosowsky, D. and Davidson, R. A., 2009b. Estimating the spatial distribution of power outages during hurricanes in the gulf coast region. *Reliability Engineering & System Safety* 94(2), pp. 199–210.
- Ho, S. and Xie, M., 1998. The use of arima models for reliability forecasting and analysis. *Computers & Industrial Engineering* 35(1), pp. 213–216.
- Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*.

- Hughes, W., Watson, P. L., Cerrai, D., Zhang, X., Bagtzoglou, A., Zhang, W. and Anagnostou, E. N., 2024. Assessing grid hardening strategies to improve power system performance during storms using a hybrid mechanistic-machine learning outage prediction model. *Reliability Engineering & System Safety* 248, pp. 110169.
- Kingma, D. P. and Ba, J., 2015. Adam: A method for stochastic optimization. In: Y. Bengio and Y. LeCun (eds), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Li, D., Zhang, Z., Alizadeh, B., Zhang, Z., Nick Duffield, M. A. M., Thompson, C. M., Gao, H. and Behzadan, A. H., 2024. A reinforcement learning-based routing algorithm for large street networks. *International Journal of Geographical Information Science* 38(2), pp. 183–215.
- Liu, H., Davidson, R. A. and Apanasovich, T. V., 2008. Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms. *Reliability Engineering & System Safety* 93(6), pp. 897–912.
- Raicharoen, T., Lursinsap, C. and Sanguanbhokai, P., 2003. Application of critical support vector machine to time series prediction. In: *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS '03.*, Vol. 5, pp. V–V.
- Roshanak Nateghi, Seth Guikema, S. M. Q., 2013. Power outage estimation for tropical cyclones: Improved accuracy with simpler models. *Risk Analysis* 34(6), pp. 1069–1078.
- Sak, H., Senior, A. and Beaufays, F., 2014. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: *Interspeech 2014*, pp. 338–342.
- Sherstinsky, A., 2020. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena* 404, pp. 132306.
- Shuai, M., Chengzhi, W., Shiwen, Y., Hao, G., Jufang, Y. and Hui, H., 2018. Review on economic loss assessment of power outages. *Procedia Computer Science* 130, pp. 1158–1163.
- Tang, R., Zeng, F., Chen, Z., Wang, J.-S., Huang, C.-M. and Wu, Z., 2020. The comparison of predicting storm-time ionospheric tec by three methods: Arima, lstm, and seq2seq.
- Taylor, W. O., Cerrai, D., Wanik, D., Koukoulou, M. and Anagnostou, E. N., 2023. Community power outage prediction modeling for the eastern united states. *Energy Reports* 10, pp. 4148–4169.
- Tervo, R., Lång, I., Jung, A. and Mäkelä, A., 2021. Predicting power outages caused by extratropical storms. *Natural Hazards and Earth System Sciences* 21(2), pp. 607–627.
- USDOE, 2023. Electric disturbance events (oe-417) annual summaries. Report, U.S. Department of Energy.
- Van Houdt, G., Mosquera, C. and Nápoles, G., 2020. A review on the long short-term memory model. *Artificial Intelligence Review*.
- Xu, J., Qiang, Y., Cai, H. and Zou, L., 2023. Power outage and environmental justice in winter storm uri: an analytical workflow based on nighttime light remote sensing. *International Journal of Digital Earth* 16(1), pp. 2259–2278.
- Yang, F., Watson, P., Koukoulou, M. and Anagnostou, E. N., 2020. Enhancing weather-related power outage prediction by event severity classification. *IEEE Access* 8, pp. 60029–60042.
- Zhang, Z., Hu, H., Yin, D., Kashem, S., Li, R., Cai, H., Perkins, D. and Wang, S., 2020. A cybergis-enabled multi-criteria spatial decision support system: A case study on flood emergency management. In: *Social Sensing and Big Data Computing for Disaster Management*, Routledge.
- Zheng, J., Xu, C., Zhang, Z. and Li, X., 2017. Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network. In: *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6.