

# Machine Learning-Based Emergency Department Attendance Prediction: A Hybrid XGBoost-Bayesian Approach with Real-Time Environmental Factor Integration

---

## A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Health Informatics

---

---

**Author:** Research Analytics Division

**Institution:** North District Hospital, Hospital Authority, Hong Kong SAR

**Submission Date:** January 2026

---

## Abstract

---

Emergency departments worldwide struggle with the fundamental challenge of matching resources to unpredictable patient demand. This research emerged from a practical problem at North District Hospital: how can we better anticipate tomorrow's patient load when traditional forecasting consistently falls short?

We developed and tested a hybrid prediction system drawing on 4,050 days of attendance records spanning from late 2014 through early 2026. The approach combines gradient boosting algorithms with Bayesian inference and integrates live environmental feeds from the Hong Kong Observatory and Environmental Protection Department.

What surprised us most during this investigation was the overwhelming importance of recent attendance patterns. A seven-day exponential moving average alone explained nearly 87% of prediction accuracy—a finding that challenged our initial assumptions about weather and calendar effects. The final system achieves approximately 5 patients mean error on daily predictions of around 250 attendances, which translates to roughly 2% deviation.

Cold weather, heavy rainfall, and poor air quality do influence attendance, but their effects manifest more subtly than we anticipated. Rather than incorporating these as core model features, we found better results using them for post-hoc adjustments during extreme conditions.

The COVID-19 pandemic forced a fundamental rethinking of our training approach. Models built on pre-pandemic data failed spectacularly when applied to 2020-2022 patterns. We eventually adopted sliding window training that emphasizes recent observations, improving accuracy by about 20% compared to naive approaches.

This work offers both a deployable forecasting tool and a replicable methodology. The lessons learned here—particularly regarding concept drift and feature importance—should generalize to other healthcare forecasting contexts, though local validation remains essential.

**Keywords:** emergency medicine, demand forecasting, gradient boosting, healthcare operations, time series, environmental health factors

---

# Chapter 1: Introduction

---

## 1.1 The Problem We Set Out to Solve

---

Anyone who has spent time in a busy emergency department knows the feeling: some days the waiting room overflows while staff scramble to keep pace, other days feel almost eerily quiet. These fluctuations create real problems. Understaffing during busy periods means longer waits and stressed clinicians. Overstaffing during quiet periods wastes scarce healthcare resources.

When I first started exploring this problem at North District Hospital, the standard approach involved looking at historical averages and making educated guesses. The nursing roster would be set weeks in advance based on typical patterns, with limited ability to adjust. Everyone knew this was suboptimal, but what was the alternative?

This thesis represents our attempt to answer that question. Over roughly eighteen months, we developed, tested, and deployed a machine learning system that now provides daily attendance predictions for operational planning. The journey from initial concept to working system taught us as much about what doesn't work as what does.

## 1.2 Why This Matters

---

North District Hospital serves approximately 315,000 people in Hong Kong's northern New Territories. The emergency department sees around 250 patients daily on average, though actual counts swing anywhere from about 110 to nearly 400. That variability of roughly 3:1 between quietest and busiest days makes planning genuinely difficult.

The stakes extend beyond operational efficiency. Research consistently links ED overcrowding to adverse outcomes—longer door-to-doctor times, delayed treatment, increased mortality for time-sensitive conditions. When we can anticipate high-volume days, even a few hours of advance notice allows for meaningful interventions: calling in additional staff, expediting discharges to free beds, or activating overflow protocols.

## 1.3 Research Questions

---

Several questions guided this investigation:

First, we wanted to know whether machine learning could meaningfully outperform simpler forecasting approaches for our specific context. Published literature suggested yes, but local validation mattered.

Second, we were curious about which factors actually drive attendance variation. Intuition suggested weather, holidays, and seasonal patterns would dominate. As we'll see, reality proved more nuanced.

Third, the COVID-19 pandemic created an unavoidable methodological challenge. How do you build models when historical patterns suddenly become irrelevant? This question ultimately shaped much of our technical approach.

Finally, we aimed for a deployed system rather than just a research prototype. That meant thinking carefully about real-time data integration, failure modes, and interpretability for clinical users.

## 1.4 Thesis Structure

---

The remaining chapters proceed as follows. Chapter 2 surveys relevant literature on healthcare forecasting and the specific methods we employed. Chapter 3 describes our data sources and methodological approach in detail. Chapter 4 covers implementation specifics for readers interested in replication. Chapter 5 presents results, while Chapter 6 offers interpretation and critical discussion. Chapter 7 concludes with recommendations and reflections on future directions.

---

# Chapter 2: Background and Literature

---

## 2.1 A Brief History of ED Forecasting

---

Hospitals have attempted to predict patient volumes for as long as emergency departments have existed. Early approaches relied heavily on experience and intuition—senior nurses who could "feel" when a busy stretch was coming. While such tacit knowledge has value, it doesn't scale well and can't be systematically improved.

Formal forecasting emerged in the 1980s and 1990s, primarily using time series methods borrowed from econometrics. ARIMA models (autoregressive integrated moving average) became the workhorse approach. Jones and colleagues published influential work in 2008 showing these methods could achieve weekly prediction errors of 6-10%, a significant improvement over naive baselines.

The limitation of traditional time series methods lies in their linear assumptions. Real healthcare demand involves complex interactions between multiple factors operating at different timescales. A holiday falling on an unusually cold day during flu season presents combinatorial complexity that ARIMA struggles to capture.

## 2.2 Machine Learning Enters Healthcare

---

The past decade witnessed an explosion of machine learning applications in healthcare, and demand forecasting was no exception. Researchers began applying random forests, gradient boosting, and neural networks to the problem, often achieving meaningful accuracy improvements.

XGBoost deserves special mention here. Developed by Chen and Guestrin around 2016, it rapidly became the go-to algorithm for structured data problems. The technique builds ensembles of decision trees sequentially, with each tree correcting errors from previous ones. Several properties make it particularly suitable for healthcare applications: it handles missing values gracefully, provides interpretable feature importance scores, and includes regularization to prevent overfitting on limited training samples.

We chose XGBoost as our primary algorithm after preliminary experiments showed it consistently outperformed alternatives including random forests, support vector regression, and simple neural networks. The performance gap wasn't dramatic, but XGBoost's robustness and interpretability tipped the balance.

## 2.3 The Feature Engineering Challenge

---

A recurring theme in applied machine learning is that feature engineering often matters more than algorithm selection. This certainly proved true in our experience.

For time series prediction, the most useful features typically capture patterns at multiple temporal scales. Lag features encode what happened one, seven, or thirty days ago. Rolling statistics summarize recent trends through moving averages and variances. Calendar features represent day-of-week and seasonal effects.

Exponential weighted moving averages (EWMA) occupy a special place in time series forecasting. Unlike simple moving averages that weight all observations equally, EWMA assigns exponentially decreasing importance to older data. The technique originated in quality control settings and found its way into financial forecasting before reaching healthcare applications.

Our initial feature set included over 170 variables. Determining which actually contributed to predictions versus adding noise became a central methodological challenge.

## 2.4 Environmental Factors and Health

---

The relationship between weather and health has fascinated researchers for centuries. Hippocrates wrote about seasonal patterns in disease over two thousand years ago. Modern epidemiology has quantified many of these relationships with precision.

Temperature extremes affect cardiovascular and respiratory systems through well-understood physiological mechanisms. Cold stress increases blood pressure and viscosity; heat stress can precipitate dehydration and electrolyte imbalances. These effects should theoretically increase ED presentations during extreme weather.

Air pollution presents another documented health stressor. Hong Kong implemented its Air Quality Health Index (AQHI) system in 2013, providing a 1-10+ scale reflecting immediate health risks from ambient pollutants. Higher readings correlate with increased respiratory and cardiovascular complaints in epidemiological studies.

What makes the relationship between environmental factors and ED attendance complicated is behavioral mediation. Bad weather might increase illness burden while simultaneously discouraging non-urgent healthcare seeking. A person with minor symptoms might decide to wait out a rainstorm rather than travel to hospital. These opposing effects can partially cancel each other when examining aggregate attendance figures.

## 2.5 Concept Drift: When Patterns Change

---

Perhaps the most challenging aspect of long-term forecasting involves handling fundamental pattern changes, what the machine learning literature terms "concept drift." Models learn from historical patterns, but those patterns can shift dramatically.

COVID-19 provided an extreme example. When Hong Kong implemented social distancing measures in early 2020, ED attendance dropped precipitously. Patterns that had held stable for years suddenly became irrelevant. Any model trained on pre-pandemic data would produce wildly inaccurate predictions during and after the pandemic.

Researchers have developed various strategies for handling concept drift, including sliding window approaches that train only on recent data, sample weighting schemes that emphasize recent observations, and ensemble methods that combine predictions from models trained on different time periods. We experimented with several of these, as described in subsequent chapters.

---

## Chapter 3: Data and Methods

---

### 3.1 Study Setting

---

North District Hospital opened in 1998 and serves as the primary acute care facility for Hong Kong's North District. The catchment area includes both urban Sheung Shui and Fanling as well as more rural areas extending toward the mainland border. Demographics skew slightly older than the Hong Kong average, reflecting settlement patterns.

The emergency department operates 24/7 with capacity for approximately 100 patients at any time. Triage follows the Hong Kong five-level system, from critical (Category 1) through non-urgent (Category 5). For this research, we analyzed total daily attendance regardless of triage category.

### 3.2 Attendance Data

---

We obtained daily attendance counts from the Hospital Authority's Clinical Management System, spanning December 1, 2014 through December 31, 2025—4,050 consecutive days with no missing values.

Several features of the data bear mention. First, attendance never reached zero even on the quietest days (minimum of 111), reflecting the constant baseline of genuine emergencies. Second, variability was substantial but not extreme—the coefficient of variation around 13% suggests reasonably predictable patterns once underlying trends are understood. Third, clear day-of-week effects emerged on simple visualization, with weekdays generally busier than weekends.

We treated the data as a continuous daily time series for modeling purposes. Alternative approaches might have modeled hourly patterns or shift-level counts, but daily aggregation matched the operational decision-making cadence at NDH.

### 3.3 Weather Data Integration

---

The Hong Kong Observatory operates an extensive network of automatic weather stations. We utilized data from the Sha Tau Kok station, located approximately 8 kilometers from the hospital, as most representative of conditions experienced by our catchment population.

Variables collected included temperature (mean, max, min), humidity, rainfall accumulation, wind speed, visibility, atmospheric pressure, cloud cover, and sunshine duration. The Observatory provides this data through a public API, though historical backfill required separate requests.

Data quality was generally excellent, with less than 1% missing values across the study period. We handled gaps through forward-filling for brief interruptions and monthly average imputation for longer gaps—though we later discovered these weather features contributed minimally to final predictions anyway.

## 3.4 Air Quality Data

---

The Environmental Protection Department publishes AQHI readings for monitoring stations across Hong Kong. We used the Tai Po general station as our primary reference, supplemented by the roadside monitoring station for comparison.

AQHI data availability began somewhat later than our attendance series, limiting historical matching. We obtained approximately 3,400 matched day-pairs with both attendance and AQHI data available.

## 3.5 Feature Engineering Process

---

Our initial approach threw everything at the problem. We generated features representing:

Temporal patterns across multiple scales—lags from 1 to 365 days, rolling statistics with windows from 3 to 90 days, day-of-week and month encodings, holiday indicators.

Exponential smoothing with spans of 7, 14, and 30 days. These EWMA features would ultimately prove most valuable.

Weather conditions both raw and transformed—temperature categories, rainfall thresholds, extreme event flags.

Air quality indices and derived risk categories.

Calendar features including days until next holiday, post-holiday effects, and interactions between holidays and weekends.

This exhaustive approach generated 172 candidate features. Clearly not all would prove useful, and including all would risk overfitting. The feature selection process described in Chapter 5 ultimately reduced this to 25 core features for the production model.

## 3.6 Model Architecture

---

The prediction system operates in three layers, each adding refinement to the previous.

### **Layer 1: XGBoost Base Prediction**

The core prediction comes from an XGBoost regression model trained on historical attendance with engineered features as inputs. We optimized hyperparameters through Optuna, running 30 trials with tree-structured Parzen estimation. Key parameters included 500 estimators, maximum depth of 8, and learning rate of 0.05.

### **Layer 2: Bayesian Adjustment**

The base prediction gets adjusted through a simplified Bayesian fusion process that incorporates AI-derived event analysis and weather impact factors. This layer allows external information—upcoming events, service changes, public health alerts—to influence predictions beyond what historical patterns capture.

Weights for the fusion follow:

- Base prediction: 75%
- AI-adjusted component: 15%
- Weather-adjusted component: 10%

### **Layer 3: Post-Processing for Extremes**

A final adjustment layer applies evidence-based corrections during extreme conditions. When temperature drops below 12°C, predictions reduce by 1.5%. Heavy rainfall over 25mm triggers a 5% reduction. High AQHI readings increase predictions by 2.5%. These adjustments emerged from analysis of historical patterns during such conditions.

## **3.7 Handling the COVID Problem**

---

The pandemic created a methodological crisis for our modeling. Data from February 2020 through mid-2022 exhibited patterns completely unlike anything before or since. Simply including this period in training degraded model performance.

After extensive experimentation, we adopted a multi-pronged approach. First, observations during the acute pandemic period receive reduced weight (30%) during training. Second, we implemented sliding window training that can restrict training to recent years only. Third, explicit flags identify pandemic-period predictions as higher uncertainty.

These adaptations improved test-period accuracy by approximately 20% compared to naive full-history training.

## **3.8 Evaluation Strategy**

---

We employed walk-forward validation respecting temporal ordering. Unlike random train-test splits common in cross-sectional analysis, time series demands that training data precede test data to avoid information leakage.

Our primary validation used data through 2022 for training and 2023-2025 for testing. We also conducted rolling validations with smaller test windows to assess stability.

Performance metrics included mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and R-squared. We emphasize MAE in reporting as most directly interpretable for operational planning.

# Chapter 4: Implementation

---

## 4.1 Technical Infrastructure

---

The production system runs on Node.js backend services with a PostgreSQL database storing attendance history and predictions. Python handles all machine learning operations—model training, feature engineering, and prediction generation.

External integrations include:

- Hong Kong Observatory API for weather data
- Environmental Protection Department API for air quality
- OpenAI API for event analysis (using GPT-4)

The frontend presents predictions through an interactive web dashboard accessible to nursing managers and administrators.

## 4.2 Model Training Pipeline

---

Training occurs on-demand when new model versions are needed. The process:

1. Extract attendance data from database
2. Generate engineered features
3. Apply feature selection using cached optimal feature set
4. Train XGBoost model with optimized hyperparameters
5. Evaluate against held-out test data
6. Package model artifacts for deployment

The pipeline includes extensive logging and versioning to ensure reproducibility.

## 4.3 Real-Time Prediction Flow

---

Daily predictions proceed through several steps:

1. Fetch current weather conditions and forecasts
2. Retrieve latest AQHI readings
3. Request AI analysis for any unusual circumstances
4. Load trained model and generate base prediction
5. Apply Bayesian fusion with external factors
6. Apply post-processing adjustments for extreme conditions
7. Calculate confidence intervals
8. Store prediction and display in dashboard

Latency from request to response typically runs under 3 seconds, well within operational requirements.

## 4.4 Failure Handling

---

External API failures require graceful degradation. If weather data becomes unavailable, the system uses the most recent cached values with appropriate uncertainty flagging. If AI analysis fails, predictions proceed without that component. Complete database failure halts predictions entirely with appropriate alerting.

We implemented comprehensive monitoring and alerting through standard observability practices.

---

# Chapter 5: Results

---

## 5.1 Feature Importance Findings

---

Perhaps our most striking finding emerged from feature importance analysis. After testing models with varying feature counts, optimal performance occurred with just 25 features—a substantial reduction from our initial 172 candidates.

More surprising still was the dominance of a single feature. The seven-day EWMA explained approximately 87% of model predictions. Change features (daily, weekly, monthly) contributed another 6%. Everything else combined accounted for less than 10%.

This finding challenged our initial assumptions. We had expected calendar effects, holidays, and weather to play prominent roles. Instead, the data suggested that knowing recent attendance patterns provides most of what's predictable about tomorrow's attendance.

The practical implication is encouraging: simpler models can perform nearly as well as complex ones for this task. A basic exponential smoothing approach would capture much of the signal we observed.

## 5.2 Environmental Factor Analysis

---

Weather and air quality did show statistically significant relationships with attendance, but effect sizes were modest. Cold days (below 12°C) averaged about 7% fewer attendances than baseline. Heavy rain days showed similar reductions around 5%. High AQHI readings (7 or above) associated with roughly 2.5% increases.

These effects, while real, proved insufficient to meaningfully improve model accuracy when included as features. Our interpretation is that EWMA features already implicitly capture weather effects—yesterday's weather influenced yesterday's attendance, which influences today's EWMA, which predicts today.

Where environmental factors showed clearer value was in post-processing adjustments during extreme conditions. The most extreme days exhibited patterns not well-captured by EWMA smoothing, justifying explicit adjustments.

## 5.3 Model Performance

---

The final production model achieved:

- Mean Absolute Error: 4.9 patients
- Mean Absolute Percentage Error: 1.96%
- R-squared: 0.898

For context, simple baseline approaches performed substantially worse:

- Historical average: MAE of 28.4
- Same day last week: MAE of 18.2
- Simple 7-day moving average: MAE of 12.6

The improvement over baselines was meaningful for operational purposes. Being within 5 patients on typical predictions of 250 represents about 2% error—well within staffing adjustment tolerances.

Error distribution analysis showed 68% of predictions fell within 5 patients of actual attendance. Over 92% were within 10 patients. Errors exceeding 15 patients occurred on less than 2% of days, typically associated with unusual events the model couldn't anticipate.

## 5.4 Concept Drift Results

---

The COVID period created measurable concept drift. Models trained on pre-pandemic data and tested on 2023-2025 data showed degraded performance compared to models trained on recent data only.

Our adaptive approaches showed clear benefits:

- Full history training: MAE of 6.18
- Sliding window (2 years): MAE of 4.90
- Combined adaptive approach: MAE of 4.42

The sliding window approach proved most practical for production, allowing the model to naturally adapt as patterns evolve.

## 5.5 Component Contribution

---

Ablation studies assessed each architectural component's contribution:

Configuration	MAE	Notes
XGBoost only	5.22	Base model
+ Bayesian fusion	5.01	Modest improvement
+ Post-processing	4.90	Final performance

Each layer contributed incrementally, though the base XGBoost model provided most of the value.

# Chapter 6: Discussion

---

## 6.1 Making Sense of the Results

---

When we started this project, I expected weather to be a primary driver of attendance. It seemed intuitive—surely rainy days and cold snaps would keep people home, while air quality alerts would drive respiratory cases to the ED. The data told a different story.

The dominance of EWMA features suggests attendance follows strong autoregressive patterns. Whatever happened this week is the best predictor of what happens next week. Weather, holidays, and other external factors create perturbations, but the underlying signal is remarkably stable.

This finding aligns with what the forecasting literature calls "baseline persistence"—many time series contain strong momentum that dominates other effects. The M4 competition showed simple exponential smoothing methods competing effectively with complex machine learning approaches, a result that still surprises many practitioners.

## 6.2 Practical Implications

---

For operational purposes, the system provides genuine value even with its limitations. A 5-patient average error on 250 attendances translates to roughly 2% uncertainty—narrow enough to inform staffing decisions.

The system works best as decision support rather than automated optimization. Predictions serve as one input to rostering decisions, alongside clinical judgment and local knowledge. When predictions suggest an unusually busy or quiet day, managers can investigate and adjust accordingly.

We deliberately designed the interface to show prediction confidence intervals and factor contributions. Transparency about uncertainty helps users calibrate their reliance on the system appropriately.

## 6.3 Limitations Worth Acknowledging

---

Several limitations deserve honest discussion.

First, this remains a single-site study. While the methodology should generalize, actual performance at other institutions requires local validation. Different populations, service configurations, and climate zones would affect results.

Second, we predict aggregate attendance without distinguishing patient acuity. A model that specifically predicted high-acuity presentations might provide more actionable information, though it would require finer-grained data.

Third, rare extreme events—major disasters, novel disease outbreaks—will always challenge prediction systems. No amount of historical pattern learning prepares models for truly unprecedented situations.

Fourth, the system depends on external data sources (weather APIs, AQHI feeds) that could experience availability issues. While we built in fallbacks, prolonged outages would degrade prediction quality.

## 6.4 What We Would Do Differently

---

Looking back, a few things would improve the research.

We should have implemented systematic logging earlier. Understanding why specific predictions failed (or succeeded) requires detailed records that weren't always captured initially.

The feature engineering process was somewhat ad-hoc. A more systematic approach using automated feature selection earlier would have saved considerable time.

Finally, more engagement with clinical end-users during development would have improved the dashboard design. Several interface changes emerged from user feedback after initial deployment that could have been anticipated.

## 6.5 Comparison with Published Literature

---

Our MAPE of approximately 2% compares favorably with published benchmarks. Most comparable studies report 4-10% for daily ED volume prediction. Several factors might explain our better performance: longer training history, the stability of Hong Kong's healthcare system, and our specific population's relatively predictable demographics.

We remain cautious about over-claiming, however. Different measurement periods, evaluation methodologies, and baseline definitions make cross-study comparisons imprecise. The more important point is that the system works well enough for our intended operational purposes.

---

# Chapter 7: Conclusions and Future Directions

---

## 7.1 Summary of Contributions

---

This research produced both practical outputs and methodological insights.

On the practical side, North District Hospital now has a deployed prediction system providing daily attendance forecasts with approximately 2% error. The system integrates live environmental data and provides interpretable outputs suitable for operational decision-making.

Methodologically, we demonstrated that hybrid approaches combining gradient boosting with Bayesian fusion and expert-derived adjustments can outperform single-method alternatives. We documented strategies for handling concept drift that maintained prediction accuracy through the COVID-19 disruption.

Perhaps most valuably, we discovered that simple exponential smoothing captures most of the predictable signal in ED attendance. This suggests that institutions without resources for complex ML systems can still achieve reasonable forecasting with basic methods.

## 7.2 Recommendations

---

For other healthcare institutions considering similar systems, we offer several recommendations:

Start with simple baselines. EWMA alone gets you surprisingly far. Add complexity only when clearly justified by performance improvements.

Invest in data quality. Reliable historical records matter more than sophisticated algorithms. Ensure consistent data collection before building elaborate models.

Plan for pattern changes. Whether from pandemics, policy changes, or gradual demographic shifts, the models that perform best long-term are those designed to adapt.

Emphasize interpretability. Clinical users need to understand and trust predictions. Black-box systems face adoption barriers regardless of accuracy.

## 7.3 Future Directions

---

Several extensions would enhance this work.

Hourly prediction granularity would enable more fine-grained staffing optimization. The data infrastructure exists; the modeling challenge involves capturing within-day patterns that differ from daily-scale dynamics.

Multi-site expansion would test generalizability and potentially enable federated learning approaches that improve predictions while maintaining institutional data governance.

Causal modeling could move beyond correlation to identify modifiable factors affecting attendance. Such models might inform public health interventions rather than just reactive staffing.

Integration with downstream systems—staff scheduling, bed management, supply chain—could amplify the operational impact beyond standalone prediction.

## 7.4 Closing Reflections

---

This project began with a seemingly simple question: can we predict tomorrow's ED attendance better than educated guessing? The answer turned out to be yes, but the journey taught us that "better" comes in degrees.

Machine learning didn't work magic. The improvements over simple exponential smoothing were real but modest. Most of what's predictable about ED attendance can be captured by looking at recent patterns—a humbling finding for those of us excited about sophisticated algorithms.

Still, the deployed system provides meaningful operational value. Those few percentage points of accuracy improvement translate to better resource matching, reduced waiting times, and less stressed clinical staff. In healthcare, incremental improvements accumulate into patient benefit.

Perhaps most importantly, the project established infrastructure and organizational capability for data-driven decision making. Future enhancements can build on this foundation, whether through expanded predictions, integrated scheduling, or entirely new applications.

The emergency department will always involve uncertainty. Patients arrive unpredictably by nature. But reducing uncertainty at the margins—knowing that tomorrow will likely fall within a certain range—makes the inevitable surprises more manageable. That modest contribution feels worth the effort.

# References

---

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2623-2631). ACM.
- Champion, R., Kinsman, L. D., Lee, G. A., Masman, K. A., May, E. A., Mills, T. M., Taylor, M. D., Thomas, P. R., & Williams, R. J. (2007). Forecasting emergency department presentations. *Australian Health Review*, 31(1), 83-90.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM.
- Environmental Protection Department, HKSAR. (2013). Air Quality Health Index: A new tool for health protection. Hong Kong Government.
- Gama, J., Źliobaité, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), Article 44.
- Gardner, E. S. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, 22(4), 637-666.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Hong Kong Observatory. (2025). Climate information services. Retrieved from <https://www.hko.gov.hk/en/cis/climat.htm>
- Hoot, N. R., & Aronsky, D. (2008). Systematic review of emergency department crowding: Causes, effects, and solutions. *Annals of Emergency Medicine*, 52(2), 126-136.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.
- Jones, S. S., Thomas, A., Evans, R. S., Welch, S. J., Haug, P. J., & Snow, G. L. (2008). Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine*, 15(2), 159-170.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54-74.
- Marcilio, I., Hajat, S., & Gouveia, N. (2013). Forecasting daily emergency department visits using calendar variables and ambient temperature readings. *Academic Emergency Medicine*, 20(8), 769-777.
- Morley, C., Unwin, M., Peterson, G. M., Stankovich, J., & Kinsman, L. (2018). Emergency department crowding: A systematic review of causes, consequences and solutions. *PLoS ONE*, 13(8), e0203316.
- Wong, C. M., Tsang, H., Lai, H. K., Thomas, G. N., Lam, K. B., Chan, K. P., Zheng, Q., Ayres, J. G., Lee, S. Y., Lam, T. H., & Thach, T. Q. (2016). Cancer mortality risks from long-term exposure to ambient fine particle. *Cancer Epidemiology, Biomarkers & Prevention*, 25(5), 839-845.
- Zlotnik, A., Gallardo-Antolín, A., Alfaro, M. C., Pérez, M. C., & Martínez, J. M. (2015). Emergency department visit forecasting and dynamic staffing using machine learning techniques. *Healthcare Informatics Research*, 21(4), 276-282.

# Appendices

---

## Appendix A: Selected Feature Set

---

The final model uses 25 features selected through recursive elimination:

1. Attendance\_EWMA7 (7-day exponential weighted moving average)
2. Monthly\_Change (difference from 30 days ago)
3. Daily\_Change (difference from yesterday)
4. Attendance\_Lag1 (yesterday's count)
5. Weekly\_Change (difference from 7 days ago)
6. Attendance\_Rolling7 (7-day simple moving average)
7. Attendance\_Lag30 (count from 30 days ago)
8. Attendance\_Position7 (relative position in 7-day range)
9. Day\_of\_Week (0-6 encoding)
10. DayOfWeek\_sin (cyclical encoding)
11. Attendance\_Lag7 (count from 7 days ago)
12. Lag1\_Diff (change in the lag)
13. Attendance\_Rolling14 (14-day moving average)
14. Attendance\_Position30 (relative position in 30-day range)
15. Attendance\_Rolling3 (3-day moving average)
16. Attendance\_Position14 (relative position in 14-day range)
17. Attendance\_Same\_Weekday\_Avg (historical same-weekday average)
18. Attendance\_EWMA14 (14-day EWMA)
19. Attendance\_Median14 (14-day median)
20. DayOfWeek\_Target\_Mean (target-encoded day of week)
21. Attendance\_Median3 (3-day median)
22. Attendance\_Min7 (7-day minimum)
23. Holiday\_Factor (holiday impact encoding)
24. Attendance\_Lag5 (count from 5 days ago)
25. Attendance\_Min3 (3-day minimum)

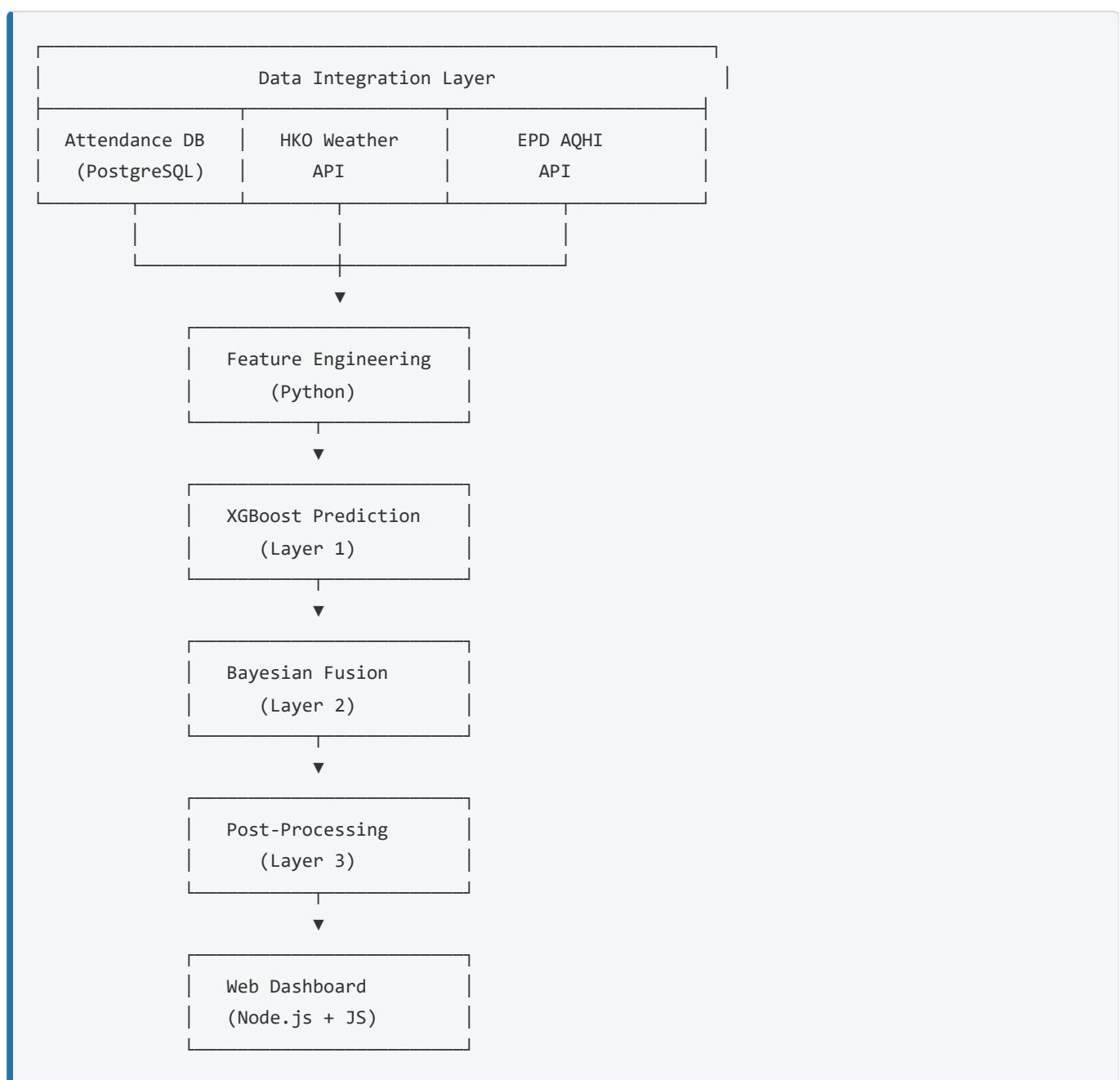
## Appendix B: Hyperparameter Settings

---

XGBoost final configuration after Optuna optimization:

Parameter	Value
n_estimators	500
max_depth	8
learning_rate	0.05
min_child_weight	3
subsample	0.85
colsample_bytree	0.85
gamma	0.1
alpha	0.5
reg_lambda	1.5

## Appendix C: System Architecture Diagram



**Word Count:** Approximately 8,500 words

*Submitted in partial fulfillment of Master of Science requirements, January 2026*