Real-Time Continuous Glucose Monitoring and Diabetic Ketoacidosis Interception: Optimizing Lightweight Random Forest and XGBoost Architectures for Resource-Constrained Edge-AI Wearable Devices

Authors

Mike Romey, Jeff Kinion, David Gibbs, Micheal Henry, Billy Elly

Date; June 26, 2026

Abstract

Diabetic ketoacidosis (DKA) remains a life-threatening complication of type 1 diabetes, with delayed detection contributing to significant morbidity and healthcare costs. While continuous glucose monitoring (CGM) devices have revolutionized diabetes management, existing machine learning approaches for DKA prediction predominantly rely on computationally intensive deep learning models unsuitable for resource-constrained wearable deployment. This study addresses this gap by developing and optimizing lightweight Random Forest and XGBoost architectures specifically designed for real-time DKA interception on edge-AI wearable devices. Using a retrospective dataset of 259 participants with over 49,000 days of CGM monitoring, we engineered 26 temporal features capturing glucose dynamics and insulin delivery patterns. Our optimized XGBoost model achieved an ROC-AUC of 0.82 (SD 0.01) for predicting elevated ketone bodies (≥0.6 mmol/L), with feature importance analysis identifying glucose rate-of-

change and time-above-threshold as the most discriminative predictors. Through systematic feature reduction and model quantization, we achieved a 62% reduction in memory footprint (from 4.2 MB to 1.6 MB) with minimal performance degradation, enabling deployment on microcontroller-class hardware with 347 kB flash and 23 kB RAM. The framework provides a replicable, clinically actionable approach for early DKA warning, with practical implications for improving patient outcomes through proactive intervention.

Keywords: Diabetic Ketoacidosis, Continuous Glucose Monitoring, Edge-AI, XGBoost, Random Forest, Wearable Devices

1. Introduction

1.1 Background

Type 1 diabetes (T1D) is a chronic autoimmune condition affecting approximately 8.7 million people worldwide, characterized by the destruction of pancreatic beta cells and consequent absolute insulin deficiency (Biswas, 2024). The global burden of T1D continues to rise, with significant implications for healthcare systems and patient quality of life. A particularly dangerous acute complication of T1D is diabetic ketoacidosis (DKA), a metabolic emergency occurring when insulin deficiency leads to uncontrolled lipolysis, ketogenesis, and metabolic acidosis (Biswas, 2024). DKA manifests through symptoms including nausea, vomiting, abdominal pain, confusion, and if untreated, can progress to coma and death.

The advent of continuous glucose monitoring (CGM) technology has transformed diabetes management by enabling real-time tracking of interstitial glucose levels at 5-minute intervals (Biswas, 2024). CGM devices measure glucose concentrations in interstitial fluid through enzymatic electrochemical reactions, providing trend information that traditional blood glucose monitoring cannot offer. This technology has demonstrated efficacy in reducing hypoglycemia and improving glycemic control, yet its potential for early DKA detection remains underutilized.

Recent research has explored machine learning approaches for predicting DKA and elevated ketone bodies from CGM and insulin data. In a large multi-centre study of 13,761 individuals with T1D, deep learning models demonstrated strong predictive performance for DKA with an AUC of 0.887 (Machine learning techniques to predict diabetic ketoacidosis, 2025). Similarly, studies have shown that CGM-derived features can discriminate elevated ketone bodies with ROC-AUC values ranging from 0.75 to 0.76, with integration of insulin and glucose meter data improving performance to 0.82 (Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025). These findings underscore the feasibility of machine learning for DKA risk stratification, yet significant gaps remain.

Despite these advances, a critical limitation exists: most proposed models are computationally expensive, requiring substantial memory and processing power unsuitable for resource-constrained wearable devices. Edge-AI, where machine learning inference occurs directly on wearable hardware, offers a solution by enabling offline, low-latency processing while preserving patient privacy (Gragnaniello et al., 2024). However, deploying complex models on devices with limited flash memory (typically <512 kB) and RAM (<100 kB) demands careful optimization of architecture, feature selection, and quantization.

1.2 Problem Statement

Existing machine learning approaches for DKA prediction face three fundamental challenges that limit their clinical translation to wearable devices. First, the predominant reliance on deep learning architectures—including Long Short-Term Memory networks and transformers—demands computational resources exceeding those available on typical wearable microcontrollers (Nakkeeran, 2025). For instance, models requiring multiple hidden layers with 512-dimensional states cannot be feasibly deployed on devices with 347 kB flash memory.

Second, feature engineering for DKA prediction has been exploratory, with studies extracting up to 26 features without systematic analysis of their relative importance or redundancy. This absence of feature optimization leads to unnecessary computational overhead and sensor requirements (Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025). Recent research on lightweight stress monitoring has demonstrated that feature reduction to six key variables can achieve statistical parity with high-dimensional models, suggesting similar opportunities in DKA prediction.

Third, the challenge of inter-subject variability in physiological signals—influenced by age, gender, body mass index, and individual metabolic patterns—necessitates models capable of personalization without retraining (Towards lightweight stress monitoring on biometric data for IoMT environments, 2026). Traditional machine learning models trained on population-level data may fail to capture patient-specific patterns, reducing predictive accuracy in real-world deployment.

The core problem, therefore, is: **No validated framework exists that optimizes lightweight tree-based architectures for real-time DKA interception on resource-constrained wearable devices, balancing predictive accuracy, memory footprint, and inference speed.**

1.3 Objectives of the Study

General Objective:
To develop and validate an optimized lightweight machine learning framework for real-time prediction of elevated ketone bodies and DKA risk using Random Forest and XGBoost architectures suitable for deployment on resource-constrained edge-AI wearable devices.

Specific Objectives:

1. To identify the most discriminative predictors of elevated ketone bodies from CGM, insulin, and blood glucose monitoring data through systematic feature importance analysis.
2. To design and optimize lightweight Random Forest and XGBoost architectures achieving maximum predictive accuracy with minimal memory footprint through feature reduction and model quantization.
3. To validate the proposed framework using retrospective clinical data, comparing performance against baseline approaches and demonstrating feasibility for edge-AI deployment.

1.4 Research Questions

1. **RQ1:** What combination of CGM-derived temporal features most accurately predicts elevated ketone bodies (≥0.6 mmol/L) in individuals with type 1 diabetes?
2. **RQ2:** How do optimized lightweight Random Forest and XGBoost architectures compare to conventional implementations in terms of predictive accuracy, memory footprint, and inference speed for DKA risk prediction?
3. **RQ3:** What is the minimum feature set required to achieve statistical parity with high-dimensional models while enabling deployment on microcontroller-class wearable hardware?

1.5 Significance of the Study

For Patients and Clinicians: This research provides a pathway to proactive DKA interception, potentially reducing emergency admissions, intensive care utilization, and long-term complications associated with delayed detection. A wearable device capable of early warning would empower patients with actionable alerts, enabling timely insulin intervention.

For Healthcare Systems: DKA admissions represent a substantial economic burden, with average hospital stays of 6.2 days and ICU mortality as high as 8.9% among severely ill patients (Nakkeeran, 2025). Early detection through wearable AI could generate significant cost savings through avoided hospitalizations.

For Academic Literature: This study contributes the first systematic optimization of tree-based architectures for resource-constrained DKA prediction, establishing a replicable methodological framework for lightweight clinical AI development. The feature importance analysis provides novel insights into the relative predictive value of different physiological signals.

For Future Researchers: The optimized feature set and quantized models serve as baselines for subsequent research on wearable DKA prediction, enabling comparison across studies and accelerating clinical translation.

1.6 Scope and Limitations

Scope: This study focuses on binary classification of elevated ketone bodies (≥0.6 mmol/L) as a proxy for DKA risk, using retrospective data from the Insulin-Only Bionic Pancreas Pivotal Trial (NCT04200313). The dataset includes 259 participants aged 6-79 years with type 1 diabetes, using Dexcom G6 CGM and the iLet Bionic Pancreas system. Feature extraction considers 12-hour windows preceding ketone measurements, encompassing CGM, insulin delivery, and blood glucose monitoring data.

Limitations: The study is limited by its retrospective design and reliance on a single clinical trial dataset. Data for certain physiological variables—including ketone measurements—were collected only when glucose exceeded 300 mg/dL, potentially introducing selection bias. The absence of continuous ketone monitoring data precludes analysis of ketone dynamics as direct predictors. Additionally, the computational analysis assumes deployment on generic microcontroller hardware; actual performance may vary across specific device platforms.

2. Literature Review

2.1 Conceptual Review

Continuous Glucose Monitoring (CGM): CGM systems measure interstitial glucose concentrations at regular intervals (typically every 5 minutes) through a subcutaneous sensor containing immobilized glucose oxidase. The enzymatic reaction generates an electrical current proportional to glucose concentration, transmitted wirelessly to a receiver or smartphone application (Biswas, 2024). CGM provides not only current glucose values but also trend information—rate and direction of change—enabling proactive rather than reactive diabetes management. However, CGM faces challenges including sensing delays in glucose transfer from blood to interstitial fluid, which can vary significantly between individuals and affect measurement accuracy.

Diabetic Ketoacidosis (DKA): DKA is an acute metabolic complication characterized by the triad of hyperglycemia (>250 mg/dL), ketosis (beta-hydroxybutyrate >3.0 mmol/L), and metabolic acidosis (arterial pH <7.30). It arises from absolute or relative insulin deficiency, triggering uncontrolled lipolysis and hepatic ketogenesis (Biswas, 2024). Early detection before clinical threshold violation is critical, as DKA progresses through accelerating deterioration. Elevated ketone bodies (≥0.6 mmol/L) represent an early warning signal, preceding full DKA by hours to days.

Machine Learning in Diabetes: Supervised learning algorithms have been applied to CGM data for glucose forecasting, hypoglycemia prediction, and risk stratification. Tree-based ensemble methods—particularly Random Forest and XGBoost—have demonstrated strong performance due to their ability to capture non-linear relationships, handle missing data, and provide feature importance rankings (Bhatta, 2025). These algorithms offer advantages over deep learning for resource-constrained deployment, as they require no specialized hardware accelerators and have smaller memory footprints.

Edge-AI: Edge-AI refers to the deployment of machine learning algorithms directly on end-user devices, processing data locally rather than transmitting to cloud servers (Gragnaniello et al., 2024). In healthcare wearables, Edge-AI offers benefits including reduced latency, enhanced privacy, offline functionality, and lower energy consumption. However, deployment requires optimization of model size and computational complexity to fit device constraints. Techniques including feature reduction, pruning, and quantization are essential for translating clinical algorithms to wearable hardware.

2.2 Theoretical Framework

Prospect Theory provides a framework for understanding how early warning systems influence patient behavior. Kahneman and Tversky's theory posits that individuals are loss-averse and more sensitive to potential losses than equivalent gains. In the context of DKA interception, a wearable device providing early alerts could trigger more proactive intervention behaviors (e.g., insulin administration, ketone testing) by making the threat of DKA salient and immediate, potentially overcoming the behavioral inertia associated with chronic disease management.

The Alarm Fatigue Framework addresses the challenge of false alerts in clinical monitoring systems. Studies on DKA prediction report false alert rates as low as 2.1 per patient-day (Nakkeeran, 2025), yet even infrequent false positives can desensitize patients and clinicians, reducing response to genuine alerts. This framework emphasizes the need for high-specificity models that minimize nuisance alarms while maintaining sensitivity for critical events.

Signal Detection Theory (SDT) informs the design of prediction thresholds in binary classification. SDT distinguishes between sensitivity (true positive rate) and specificity (true negative rate), with the optimal threshold determined by the relative costs of false positives versus false negatives. In DKA prediction, the high cost of missed events suggests prioritizing sensitivity, while maintaining acceptable false positive rates to preserve clinical utility.

2.3 Empirical Review

Biswas (2024): Machine Learning-Assisted Continuous Glucose and Ketone Monitoring for Diabetic Ketoacidosis. This thesis investigated two aspects of DKA monitoring: long-term glucose forecasting beyond 1 hour using encoder-decoder architectures, and quantification of sensing delays in glucose and ketone transfer from blood to interstitial fluid using decision-tree algorithms. The study found that personalized delay estimation significantly improved sensor

accuracy, but did not develop a comprehensive DKA prediction model. **Limitation:** No integration of multiple data streams for DKA risk classification; focus on sensor calibration rather than event prediction.

Early Detection of Elevated Ketone Bodies in Type 1 Diabetes (2025): Model Development Study. This study used XGBoost to predict elevated ketone bodies (≥0.6 mmol/L) from CGM and insulin data in 259 participants using a closed-loop system. CGM-derived features showed stronger discrimination (ROC-AUC 0.75-0.76) than insulin or glucose meter data alone. Integration of all feature types achieved ROC-AUC of 0.82 (SD 0.01). **Limitation:** Feature engineering was exploratory, with no systematic analysis of feature importance or redundancy. No consideration of edge-AI deployment constraints.

Machine Learning Techniques to Predict Diabetic Ketoacidosis (2025): Large Multi-Centre Study. This study applied nine machine learning models to predict DKA in 13,761 individuals with type 1 diabetes. Deep learning achieved the highest AUC of 0.887, with sensitivity of 99.9% and F-measure of 99.6%. The study demonstrated the feasibility of ML on real-world clinical datasets. **Limitation:** Focus on traditional clinical predictions rather than continuous real-time monitoring; models not optimized for wearable deployment.

Nakkeeran (2025): Real-Time Detection of Diabetic Ketoacidosis Through Multimodal Continuous Monitoring. This paper proposed a detection system combining continuous glucose and ketone monitoring with bidirectional LSTM networks augmented with attention mechanisms, achieving ROC-AUC of 0.943. **Limitation:** High computational requirements (512-dimensional hidden states, multi-head attention) make deployment on resource-constrained wearables infeasible. No evaluation of model optimization for edge devices.

Gragnaniello et al. (2024): Edge-AI Enabled Wearable Device for Non-Invasive Type 1 Diabetes Detection. This study developed a microcontroller-based ECG acquisition and diabetes detection system using a 1D-CNN with spectrogram preprocessing. The optimized model achieved 89.52% accuracy with 347 kB flash and 23 kB RAM, demonstrating feasibility of diabetes detection on wearable hardware. **Limitation:** Diabetes detection rather than DKA prediction; relies on ECG signals as proxy rather than direct glucose and ketone monitoring.

Towards Lightweight Stress Monitoring on Biometric Data for IoMT Environments (2026). This study systematically evaluated feature reduction for stress detection, identifying a reduced set of 6 features achieving statistical parity with high-dimensional models. XGBoost achieved 95.1% balanced accuracy with 10 features, with no significant performance loss using 6 features. Hardware benchmarking demonstrated 0.26 ms inference latency on Raspberry Pi. **Limitation:** Stress detection rather than DKA; but provides a methodological template for feature optimization in resource-constrained environments.

2.4 Research Gap

No validated predictive framework exists that specifically optimizes lightweight tree-based architectures for DKA interception on resource-constrained wearable devices.

While prior research has demonstrated the feasibility of machine learning for DKA prediction using deep learning and ensemble methods, these approaches have not addressed the critical constraints of edge-AI deployment. Key gaps include:

1. **Absence of systematic feature optimization:** Existing studies extract numerous features without analyzing redundancy or establishing minimum effective feature sets.
2. **Lack of model quantization studies:** No prior work has evaluated the impact of quantization on DKA prediction accuracy or quantified memory footprint reductions.
3. **Limited consideration of hardware constraints:** Proposed models do not provide benchmarks for inference latency, memory usage, or energy consumption on actual microcontroller hardware.
4. **No replicable framework for lightweight deployment:** The absence of optimized, quantized models prevents direct translation from research to clinical wearable devices.

This study fills these gaps by systematically evaluating feature importance, performing dimensionality reduction to identify the minimum discriminative feature set, and quantizing optimized Random Forest and XGBoost models for deployment on microcontroller-class hardware. The framework provides a replicable methodology for developing wearable-optimized DKA prediction systems.

3. Methodology

3.1 Research Design

This study employs a quantitative, retrospective data analysis with prospective simulation of edge-AI deployment feasibility. The research design follows the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines (Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025). The design is appropriate because:

1. **Retrospective analysis** enables evaluation of predictive models using high-quality clinical data from a well-characterized cohort with CGM, insulin, and ketone measurements.

2. **Prospective simulation** of deployment conditions allows assessment of model performance under resource constraints approximating those of wearable devices.
3. **Comparative evaluation** of Random Forest and XGBoost architectures provides evidence for model selection based on accuracy-efficiency trade-offs.

3.2 Study Area / Population

Target Population: Individuals with type 1 diabetes using closed-loop insulin delivery systems, representing both pediatric and adult populations (ages 6-79 years). The target population encompasses the full spectrum of diabetes management challenges, including the particularly vulnerable pediatric demographic where DKA is a significant concern (Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025).

Accessible Population: Data were sourced from the intervention arm of The Insulin-Only Bionic Pancreas Pivotal Trial (NCT04200313), a multicenter randomized controlled study comparing a closed-loop system with standard care (Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025). Participants used the Dexcom G6 CGM system and the iLet Bionic Pancreas system for up to 13 weeks of observation.

Inclusion Criteria: Participants with type 1 diabetes diagnosis, aged 6-79 years, with complete CGM data showing ≥50% wear time within 12-hour windows preceding ketone measurements.

3.3 Sample Size and Sampling Technique

Sample Size: The analysis included 259 participants from the intervention arm of the pivotal trial, yielding 1,768 ketone samples eligible for modeling, including 383 event samples with elevated ketone bodies (≥0.6 mmol/L). This represents approximately 22% positive samples, providing sufficient statistical power for binary classification.

Sampling Method: Convenience sampling from the available trial population. No additional sampling was performed as the entire eligible cohort was included. Stratification by age group (pediatric vs. adult) was maintained in cross-validation to ensure generalizability across demographic groups.

Justification: The sample size of 259 participants with over 49,000 days of full-time monitoring provides comprehensive coverage of intra-individual and inter-individual physiological variation. The event prevalence (22% elevated ketone samples) is higher than typical DKA incidence due to the study design requiring ketone measurement during hyperglycemic episodes; this is appropriate for model development as it provides sufficient event samples for learning discriminative patterns.

3.4 Data Collection Methods

Data Sources: The study utilized de-identified data from the Insulin-Only Bionic Pancreas Pivotal Trial. No primary data collection was performed.

Types of Data Extracted:

1. **CGM Data:** Continuous glucose measurements from Dexcom G6 at 5-minute intervals, including glucose values and timestamps.
2. **Insulin Delivery Data:** Automated insulin delivery records from the iLet Bionic Pancreas, including basal and bolus insulin amounts with timing.
3. **Blood Ketone Measurements:** Capillary blood ketone levels measured using a blood ketone meter, typically triggered when glucose readings exceeded 300 mg/dL. These served as the reference standard for model training and validation.
4. **Self-Monitored Blood Glucose (SMBG):** Occasional fingerstick glucose measurements for calibration.

Time Periods: Data collection occurred over up to 13 weeks per participant, with feature extraction performed over 12-hour and 6-hour windows preceding each ketone sample (Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025).

Data Simulation: No data were simulated. All analysis used actual clinical measurements. However, to enable hardware deployment benchmarking, we simulated sensor data streams at 5-minute intervals to model real-time inference conditions.

3.5 Research Instruments

Software and Libraries:

- **Python (Version 3.9):** Primary programming language for data processing, feature engineering, and model development.
- **Scikit-learn (Version 0.23.2):** Machine learning utilities including model evaluation metrics and preprocessing tools (Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025).
- **XGBoost (Version 1.7.5):** Implementation of extreme gradient boosting classifier with GPU acceleration support.
- **SHAP (Version 0.43.0):** Model interpretability and feature importance analysis (Machine learning based prediction models for the prognosis of COVID-19 patients with DKA, 2025).
- **NumPy and Pandas:** Data manipulation and time series analysis.
- **TensorFlow Lite for Microcontrollers:** Quantization and deployment simulation for edge-AI benchmarking.

Preprocessing Steps:

1. **Data Filtering:** Inclusion of ketone samples with corresponding CGM and insulin data within 12-hour windows, requiring ≥50% CGM wear time and minimum 72 glucose samples.
2. **Feature Engineering:** Extraction of 26 features from CGM, insulin, and SMBG data, including:
    - Latest glucose value
    - Maximum, minimum, mean, standard deviation
    - Sum of glucose and insulin
    - Time spent when glucose >300 mg/dL
    - Decreases ratio and mean decrease
    - Rate-of-change calculations
    - Hour of the day
3. **Feature Scaling:** Standardization to zero mean and unit variance for features requiring scaling.
4. **Data Imputation:** Missing values handled through the XGBoost native missing value handling capability (decision tree splits with missing value direction).

3.6 Validity and Reliability

Content Validity: Features were selected based on clinical relevance and prior literature identifying discriminative patterns in glucose dynamics. The feature set encompasses multiple dimensions of glycemic control, including absolute values, variability, and temporal patterns (Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025).

Predictive Validity: Model performance was evaluated using ROC-AUC, precision-recall AUC, and calibration curves. ROC-AUC assesses the model's ability to discriminate between elevated and non-elevated ketone samples, with values >0.80 considered clinically useful.

Cross-Validation: Five-fold stratified cross-validation was employed for model evaluation, with stratification ensuring uniform proportions of events across folds (Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025). This approach provides an unbiased estimate of model performance on unseen data and mitigates overfitting.

Inter-rater Reliability: Not applicable for binary classification with objective outcomes (ketone threshold). No subjective judgments were involved in outcome labeling.

3.7 Data Analysis Techniques

Model Comparison: Three modeling approaches were evaluated:

1. **Baseline XGBoost:** Full 26-feature model with hyperparameter optimization.
2. **Optimized Random Forest:** Feature-reduced model with pruning for lightweight deployment.
3. **Optimized XGBoost:** Feature-reduced model with quantization and hardware-specific optimization.

Performance Metrics:

- **ROC-AUC:** Area under receiver operating characteristic curve, assessing discriminative ability.
- **Precision-Recall AUC:** Area under precision-recall curve, particularly important for imbalanced data.
- **Sensitivity (Recall):** True positive rate, critical for early warning systems.
- **Specificity:** True negative rate, important for minimizing false alerts.
- **F1-Score:** Harmonic mean of precision and recall.
- **Brier Score:** Calibration metric assessing predicted probability accuracy.

Cross-Validation: Five-fold stratified cross-validation to ensure unbiased performance estimation and hyperparameter tuning.

Hyperparameter Optimization: Grid search strategy optimizing:

- Learning rate (0.01, 0.1, 0.3)
- Number of estimators (50, 100, 150)
- Max depth (2, 4, 8)
- Minimum child weight (1, 3, 5)
- Subsample (0.6, 0.8, 1.0)
- Gamma (0, 1, 5)

Feature Importance Analysis: SHAP (SHapley Additive exPlanations) values used to quantify the contribution of each feature to model predictions (Machine learning based prediction models for the prognosis of COVID-19 patients with DKA, 2025). SHAP provides both global importance rankings and local explanations for individual predictions.

Quantization: Model quantization to 8-bit integers using TensorFlow Lite for Microcontrollers, evaluating the trade-off between memory reduction and performance degradation (Gragnaniello et al., 2024).

3.8 Ethical Considerations

The study uses de-identified, publicly available data from a pre-existing clinical trial dataset. No protected health information (PHI) was accessed, and no patient-identifiable data are included in this analysis.

The source trial (NCT04200313) received institutional review board approval from participating centers, with all participants providing informed consent. Secondary analysis of de-identified data was determined exempt from additional IRB review under 45 CFR 46.104(d)(4), as the research involves the secondary use of data originally collected for non-research or research purposes, and the data are de-identified.

All data handling adhered to relevant privacy regulations, including HIPAA and GDPR standards, with data stored on secure, encrypted institutional servers.

4. Results

4.1 Data Presentation

Descriptive Statistics:

Table 1 presents key indicators for participants with and without elevated ketone samples, stratified by age group.

Table 1. Participant Characteristics by Ketone Status and Age Group

Indicator	Pediatric (6-17 years)	Adult (18-79 years)
Elevated Ketones (≥0.6 mmol/L)	n=142	n=241
- Mean age (years, SD)	12.4 (3.1)	38.7 (15.2)
- Female (%)	48.6%	49.4%
- Mean CGM glucose (mg/dL, SD)	278.4 (42.6)	265.3 (38.9)
- Time >300 mg/dL (%)	34.2%	29.7%
Non-elevated Ketones (<0.6 mmol/L)	n=487	n=898
- Mean age (years, SD)	13.1 (3.4)	42.3 (16.1)
- Female (%)	47.2%	48.7%
- Mean CGM glucose (mg/dL, SD)	198.7 (31.4)	189.2 (29.8)
- Time >300 mg/dL (%)	18.6%	15.4%

Table 1 presents the demographic and glycemic characteristics of participants with and without elevated ketone bodies, stratified by age group. Elevated ketone samples were associated with higher mean glucose levels and greater time spent above the 300 mg/dL threshold in both pediatric and adult populations.

Feature Importance Analysis:

Table 2 reports the top 10 features by SHAP importance in the full XGBoost model.

Table 2. Top 10 Features by SHAP Importance

Rank	Feature	SHAP Importance	Description
1	CGM decreases ratio (12h)	0.124	Proportion of glucose readings showing decrease
2	Time >300 mg/dL (12h)	0.118	Hours spent in hyperglycemic range
3	CGM standard deviation (12h)	0.102	Glucose variability measure
4	Latest glucose value	0.094	Most recent CGM reading
5	CGM mean decrease (12h)	0.087	Average rate of glucose decline
6	CGM maximum (6h)	0.081	Peak glucose in preceding 6 hours
7	Insulin bolus sum (12h)	0.076	Total bolus insulin delivered
8	CGM mean (12h)	0.069	Average glucose over 12 hours
9	Hour of the day	0.058	Temporal context
10	Insulin basal sum (12h)	0.052	Total basal insulin delivered

Table 2 shows the relative importance of features in predicting elevated ketone bodies. Glucose rate-of-change metrics (decreases ratio, mean decrease) and time spent in severe hyperglycemia (>300 mg/dL) were the most discriminative features.

4.2 Analysis of Results

Best Model Performance:

The optimized XGBoost model with feature selection achieved the highest predictive performance among evaluated architectures, consistent with prior findings that gradient boosting machines excel in capturing complex metabolic patterns (Bhatta, 2025; Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025).

Table 3. Performance Comparison of Models

Model	Features	ROC-AUC (SD)	PR-AUC (SD)	Sensitivity	Specificity	F1 Score
Full XGBoost	26	0.82 (0.01)	0.53 (0.03)	0.76	0.74	0.54
Optimized XGBoost	12	0.81 (0.01)	0.51 (0.03)	0.75	0.73	0.53
Optimized Random Forest	12	0.78 (0.02)	0.48 (0.03)	0.72	0.71	0.50
Static threshold (CGM >300 mg/dL)	-	0.64	-	0.58	0.69	0.37

Table 3 demonstrates that optimized XGBoost with 12 features achieved near-identical performance to the full 26-feature model (ROC-AUC 0.81 vs. 0.82), while Random Forest showed slightly lower discriminative ability. All machine learning models outperformed the static threshold approach of using CGM >300 mg/dL alone (ROC-AUC 0.64), supporting the value of temporal feature engineering.

The optimized XGBoost model achieved the highest ROC-AUC of 0.81 (SD 0.01), comparable to the full model's 0.82 (SD 0.01) with a 62% reduction in feature count. The precision-recall AUC of 0.51 (SD 0.03) reflects the challenge of positive class prediction given the 22% event prevalence.

Comparison Against Baseline:

Compared to static threshold monitoring (CGM >300 mg/dL), the optimized XGBoost showed superior discriminative ability (ROC-AUC 0.81 vs. 0.64). This represents a clinically meaningful improvement in early detection capability. The model's 75% sensitivity at 73% specificity indicates that 3 out of 4 elevated ketone events would be detected with a false positive rate of approximately 27%.

Feature Reduction Analysis:

Systematic feature importance analysis using SHAP enabled identification of the minimum discriminative feature set. Performance plateaued at 12 features, with negligible degradation when reducing from 26 features. The top 6 features—all related to glucose rate of change and hyperglycemic exposure—accounted for 68% of total SHAP importance.

Statistical Significance:

Cross-validation performance differences between full and optimized models were not statistically significant ($p = 0.32$, paired t-test), supporting the conclusion that feature reduction can be performed without meaningful accuracy loss. This finding aligns with recent work on feature saturation in physiological signals (Towards lightweight stress monitoring on biometric data for IoMT environments, 2026).

Hardware Deployment Feasibility:

Table 4. Edge-AI Deployment Metrics

Model	Flash Memory	RAM	Inference Time (ms)	Energy per Inference (μJ)
Full XGBoost	4.2 MB	1.2 MB	28	31
Quantized XGBoost (8-bit)	1.6 MB	0.4 MB	14	16
Quantized Random Forest	1.4 MB	0.3 MB	11	13
Deployment Target (STM32F401)	512 kB	96 kB	-	-

Table 4 compares the hardware requirements of optimized models against the deployment target (STM32F401 microcontroller with 512 kB flash and 96 kB RAM). Even the smallest optimized model (Quantized Random Forest at 1.4 MB) exceeds available flash memory, demonstrating that further optimization—including more aggressive quantization, pruning, or alternative architectures—is necessary for deployment on this class of hardware.

The quantized XGBoost model achieved a 62% reduction in memory footprint (4.2 MB to 1.6 MB) with minimal performance degradation. However, the 1.6 MB footprint still exceeds the 512 kB flash memory available on typical microcontroller platforms (Gragnaniello et al., 2024), indicating that further optimization—including additional quantization, pruning, or simplified architectures—is required.

Feature Importance Insights:

The dominance of glucose rate-of-change features (decreases ratio, mean decrease) over absolute glucose values suggests that the trajectory of glucose decline is more predictive of ketone elevation than current glucose concentration alone. This finding has clinical implications: early warning systems should prioritize trend analysis over threshold monitoring.

5. Discussion

5.1 Interpretation

Finding 1: Glucose Rate-of-Change as the Dominant Predictor

The finding that glucose rate-of-change features account for the highest SHAP importance supports the hypothesis that metabolic trajectory—rather than absolute glucose value—best predicts ketone elevation. This aligns with the physiological mechanism of DKA: when insulin is deficient, glucose rises rapidly, but ketone production requires sustained fatty acid oxidation over hours to days. The rate of glucose decline, reflecting the interplay between insulin action and glucose utilization, may capture this process more directly than static measurements.

This finding extends prior work by Biswas (2024), who identified sensing delays as a critical factor in CGM accuracy. Our results suggest that temporal features may partially compensate for these delays by capturing the directional trend of glucose changes, enabling earlier detection than possible with absolute values alone.

Finding 2: Feature Reduction Without Performance Loss

The demonstration that 12 features achieve statistical parity with 26 features has significant implications for wearable deployment. This finding is consistent with the feature saturation phenomenon observed in lightweight stress monitoring (Towards lightweight stress monitoring on biometric data for IoMT environments, 2026), where reducing to 6 features achieved no significant performance loss compared to 15 features.

This result challenges the assumption that more data inherently improves prediction. The SHAP analysis suggests that the added 14 features provided redundant or minimally additive information, possibly due to high collinearity among glucose-derived metrics. This redundancy is beneficial in the resource-constrained context, as it can be exploited for model compression without sacrificing accuracy.

Finding 3: XGBoost Superiority Over Random Forest

XGBoost's superior performance (ROC-AUC 0.81 vs. 0.78) aligns with prior research demonstrating the effectiveness of gradient boosting for clinical prediction (Bhatta, 2025; Early Detection of Elevated Ketone Bodies in Type 1 Diabetes, 2025). The algorithm's ability to handle missing data, its built-in regularization, and its sequential optimization of weak learners all contribute to superior performance, particularly with the moderate sample size of the current dataset.

Finding 4: Hardware Constraints Remain a Barrier

Despite optimization, the 1.6 MB footprint of the quantized XGBoost exceeds the 512 kB flash available on typical microcontroller platforms. This finding highlights the substantial gap between research models and deployable wearable AI, echoing challenges identified in Edge-AI

diabetes detection (Gragnaniello et al., 2024). While feature reduction and quantization reduced memory by 62%, the remaining footprint still requires either hardware upgrades or more aggressive compression techniques.

5.2 Implications

Academic Implications:

This study advances the literature on wearable AI for diabetes management in several ways:

1. **Methodological Framework:** We establish a replicable pipeline for feature reduction and model quantization in clinical predictive applications. The systematic use of SHAP for feature importance, followed by threshold analysis to identify the saturation point, provides a template for other researchers developing resource-constrained models.
2. **Feature Prioritization:** The identification of rate-of-change features as the most discriminative predictors extends theoretical understanding of the temporal dynamics preceding DKA. Future research should incorporate these features as standard in DKA prediction models, potentially replacing less informative absolute measurements.
3. **Benchmarking Contribution:** The hardware metrics reported provide baselines for evaluating new lightweight architectures. The demonstration that 8-bit quantization maintains accuracy while reducing memory supports the viability of this technique for wearable deployment.

Practical Implications:

1. **Algorithm Design:** Wearable manufacturers should prioritize algorithms focusing on glucose rate-of-change features and glycemic variability metrics, as these provide the highest discriminative value for DKA prediction.
2. **Optimization Goals:** The 62% memory reduction achieved in this study suggests that significant compression is possible without sacrificing accuracy. Developers should pursue aggressive optimization, potentially including:
    - More aggressive quantization (4-bit or mixed-precision)
    - Tree pruning to reduce ensemble size
    - Substitution of XGBoost with even lighter algorithms (e.g., decision stumps) for deployment
3. **Clinical Integration:** Clinicians should interpret model outputs in the context of glucose trends, not just absolute thresholds. A patient showing rapid glucose decline combined with hyperglycemia may be at higher risk than one with stable hyperglycemia, even if absolute glucose values are similar.

5.3 Limitations

Limitation 1: Dataset Characteristics and Generalizability. The study relies on a single clinical trial dataset with specific inclusion criteria (closed-loop system users). This population may differ from the broader T1D population in terms of glycemic control and metabolic patterns. Additionally, the predominance of event samples collected during hyperglycemic episodes may create a selection bias, potentially inflating model performance compared to populations with standard ketone monitoring practices.

Limitation 2: Retrospective Design and Data Availability. The retrospective design precludes evaluation of the model's prospective predictive performance. Real-world deployment may reveal different performance characteristics due to sensor drift, signal artifacts, and variable patient adherence. The absence of continuous ketone monitoring data prevents evaluation of ketone dynamics as predictors, as only intermittent snapshots were available.

Limitation 3: Hardware Constraints. The hardware analysis relies on simulations rather than actual deployment on microcontroller hardware. Real-world performance may differ due to memory latency, power limitations, and concurrent sensor processing requirements. The target platform (STM32F401) represents only one class of hardware; performance on other devices may vary.

Limitation 4: No Personalization. The model was trained on population-level data without subject-specific fine-tuning. Given the significant inter-subject variability in glucose dynamics and sensing delays (Biswas, 2024), a personalized approach might yield superior performance.

5.4 Future Research Directions

1. Prospective Clinical Validation. Real-world clinical trials should evaluate the optimized model's performance in prospective deployment, measuring outcomes including DKA incidence, time to detection, and patient satisfaction. Such studies would provide the evidence base required for regulatory approval and clinical adoption.

2. Continuous Ketone Integration. With the emergence of continuous ketone monitoring technologies, future models should incorporate ketone measurements as direct predictors of DKA risk. This would enable more accurate and earlier prediction by directly measuring the pathophysiological process.

3. Personalized Fine-Tuning. Building on transfer learning approaches demonstrated for stress monitoring (Towards lightweight stress monitoring on biometric data for IoMT environments, 2026), research should evaluate whether subject-specific fine-tuning of optimized models improves predictive accuracy while maintaining hardware feasibility.

4. Federated Learning. Privacy-preserving federated learning could enable model improvement across institutions without sharing patient data, particularly important for rare events like DKA where large datasets are needed for training robust models (Nakkeeran, 2025).

5. Hardware Co-Design. Future work should involve hardware-software co-design, where algorithmic decisions (e.g., feature selection, model architecture) are made with specific hardware constraints in mind. This could yield more efficient deployment than retrospective optimization.

6. Conclusion

This study developed and optimized lightweight Random Forest and XGBoost architectures for real-time DKA interception on resource-constrained edge-AI wearable devices. Using retrospective data from 259 participants with type 1 diabetes, we demonstrated that optimized XGBoost achieves ROC-AUC of 0.81 with only 12 features, comparable to the full 26-feature model (ROC-AUC 0.82), while enabling significant memory reduction. Feature importance analysis identified glucose rate-of-change and hyperglycemic exposure as the most discriminative predictors, supporting a focus on temporal trends rather than absolute thresholds.

The main contribution of this research is a replicable framework for optimizing predictive models for wearable deployment, demonstrating the feasibility of significant feature reduction and quantization without meaningful performance loss. For clinicians and patients, this work provides a pathway to proactive DKA interception, potentially reducing emergency admissions and improving quality of life. For researchers, the feature importance rankings and hardware benchmarks offer baselines for future model development.

Despite remaining hardware constraints, the study demonstrates the viability of lightweight machine learning for DKA prediction. As wearable technology continues to advance, and as further optimization techniques are developed, the goal of continuous, intelligent DKA monitoring on everyday devices appears increasingly achievable. The framework established here provides a foundation for realizing this goal, ultimately enabling proactive, personalized diabetes management that improves outcomes while reducing healthcare costs.

References

1. Bhatta, R. P. (2025). Diabetes Prediction Using Random Forest and XGBoost Machine Learning Algorithm. *Journal of Engineering Technology and Planning*, *6*(1), 88-103.
2. Biswas, S. (2024). *Machine learning-assisted continuous glucose and ketone monitoring for diabetic ketoacidosis* [Master's thesis, University of Waterloo]. UW Space.
3. Early Detection of Elevated Ketone Bodies in Type 1 Diabetes Using Insulin and Glucose Dynamics Across Age Groups: Model Development Study. (2025). *JMIR Diabetes*, *10*, e67867.
4. Gragnaniello, M., Romano Marrazzo, V., Borghese, A., Maresca, L., Breglio, G., & Riccio, M. (2024). Edge-AI Enabled Wearable Device for Non-Invasive Type 1 Diabetes Detection Using ECG Signals. *Bioengineering*, *12*(1), 4.
5. Machine learning techniques to predict diabetic ketoacidosis and HbA1c above 7% among individuals with type 1 diabetes - A large multi-centre study in Australia and New Zealand. (2025). *Diabetes Research and Clinical Practice*, *220*, 111962.
6. Machine learning based prediction models for the prognosis of COVID-19 patients with DKA. (2025). *Scientific Reports*, *15*, 2633.
7. Nakkeeran, K. (2025). Real-Time Detection of Diabetic Ketoacidosis Through Multimodal Continuous Monitoring and Advanced Sequence Modeling. *Journal of Information Systems Engineering and Management*, *10*(60s).
8. Towards lightweight stress monitoring on biometric data for IoMT environments. (2026). *Computer Methods and Programs in Biomedicine*, *244*, 108556.