Machine Learning Enabled IoT Embedded Platform for Real-Time Anomaly Detection in Drug Delivery Devices

Kanchana R[1], Dr S.V.Ramanan[2], Ruby Chauhan Rana[3], P Mohana[4], Lokesh[5], Rakesh V S[6*]

[1]Assistant Professor, Department of Computer Science and Engineering, East Point College of Engineering and Technology, Bengaluru, Karnataka, India - 560049
Email: kanchanar.cse@eastpoint.ac.in
[2]Assistant Professor, Department of Electronics and Communication Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India- 641062
Email: ramananece@siet.ac.in
[3]Assistant Professor, Department of Computer Science and Engineering, Cambridge Institute of Technology, K R Puram, Bengaluru, Karnataka, India
Email:rubychauhan445@gmail.com
[4]Assistant Professor, Department of IT, V.S.B Engineering college, Karur, Tamilnadu, India. Email: mohanavaishnavi2307@gmail.com
[5]Assistant Professor, Department of Computer Science and Engineering, Cambridge Institute of Technology, K R Puram, Bengaluru, Karnataka, India
Email: lokesh.cse@cambridge.edu.in
[6]Assistant Professor , Department of Computer Science and Engineering, Cambridge Institute of Technology, K R Puram, Bengaluru, Karnataka, India
Email: rakesh.tech102@gmail.com

Abstract

The reliability of modern drug delivery devices such as infusion pumps and insulin pumps is critically dependent on accurate delivery of medication. However, despite advances in recent years, such devices are prone to failure with reported events including flow abnormalities, blockages, inaccurate dosage delivery and mechanical failure. These events have potential to cause serious harm to patients with risk of overdose or underdose. Current monitoring of such devices includes analysis of threshold-based alarms raised by the device, which often fail to provide early intervention and lack of intelligent detection of abnormalities. Therefore, there is a growing need for implementation of intelligent real-time monitoring in these devices to guarantee their reliability. This paper outlines an IoT-enabled, embedded machine learning implementation that monitors the operation of drug infusion equipment, in real-time, and alerts for potential anomalies such as flow anomalies or occlusions. This novel embedded implementation, operating on the edge of the network, can process in real-time the real-time, raw sensor information streams (i.e. flow rate, pressure and actuator signals) generated by infusion pump(s) for the high speed, on-line classification and thereby results in significantly reduced latency due to minimized communication requirements with cloud-computing architectures. The proposed system was tested on simulated data from sensors of infusion pumps and insulin pumps. The results of the experiments indicate a high accuracy of the system in detecting various types of anomalies, inclu
ding occlusion and dosage errors. In addition, the system is capable of operating in real-time, detecting anomalies within a time frame of a few milliseconds. This is due to a low computational complexity of the model. The system is able to process large amounts of data with high precision and thus it can be used in various applications in the field of healthcare. The proposed system enhances patient safety by timely detection of failure of the drug delivery device, and helps in improving the monitoring precision and time to response of the clinical staff for treatment using smart healthcare systems employing IoT and edge intelligence for scalable implementation.

Keywords: IoT, Edge Computing, Machine Learning, Drug Delivery Systems, Anomaly Detection, Embedded Systems, Healthcare AI, Real-Time Monitoring.

How to cite this article: Kanchana R, Ramanan SV, Rana RC, Mohana P, Lokesh, Rakesh VS. Machine Learning Enabled IoT Embedded Platform for Real-Time Anomaly Detection in Drug Delivery Devices. Int J Drug Deliv Technol. 2026;16(66s):1068-1075. DOI: 10.25258/ijddt.16.66s.111

1. Introduction

Infusion pumps, insulin pumps, automated syringe drives and other drug delivery systems are integral parts of modern healthcare. The most common area of application is the intensive care unit, but also chronic therapy, surgery and emergency treatment benefit from the exact administration of drugs. However, also these devices are not failure-proof[1].

Possible failures concern the flow rate, mechanical parts, catheters and the calculation of the dosage. These failures can lead to serious damage to the patient, for example an overdose or an underdose of drugs that are needed on a continuous basis[2].

Typical monitoring of drug delivery devices focuses on a threshold-based alarm and periodic checking of devices. Threshold-based alarms, however, tend to produce a high amount of false alerts and have a considerable delay before a real problem is detected. The above-mentioned problems can be partially solved by increasing clinical effort[3]. However, the main problem with current systems is the lack of adaptive and real-time monitoring with context information from various sensors.

Recent advancements in Internet of Things (IoT) have led to significant innovation in monitoring patients and their treatment using various healthcare delivery systems, through utilization of sensors and networks for monitoring medical equipment[4]. This healthcare system has the ability to acquire data in real time and monitor patients and treat diseases through visualized data, enhancing healthcare delivery and saving lives. Another innovative method for detecting abnormal patterns in a vast array of data has recently emerged. Such data can include time-series data from monitoring systems that may indicate an anomaly or irregularity that could be extremely difficult for a human to identify via rules or by monitoring large amounts of data. Machine learning (ML) therefore has become increasingly popular as a result for detection of such anomalies and for making accurate predictions using large data sets[5].

There are many challenges that still have to be tackled with when it comes to implementing solutions for monitoring infusion pumps and other drug delivery devices by using IoT, health care monitoring and machine learning applications. Although there are many solutions already in place, most of the implementations rely on cloud-based servers and therefore are not suitable for real-time monitoring due to latency[6]. However, when it comes to health care, especially when patients' lives are at risk due to the drugs being administered, real-time monitoring is crucial. In addition, there are only a few solutions that are implemented to be able to run on embedded hardware with restricted resources such as memory and power. Therefore, developing efficient solutions for real-time health care monitoring on embedded platforms is still an open challenge in the field of health care and machine learning applications[7].

To address the problems of current systems a real-time anomaly detection framework was developed for monitoring infusion pumps, insulin pumps and automated syringe devices using a local IoT monitoring system with a lightweight machine learning model. This system processes real-time data from all relevant sensors on an embedded computer to detect anomalies as fast as possible[8]. The developed system was implemented in a prototype and validated in various simulations. The results of the experiments show the high potential of the framework to improve the safety of patients, the reliability of the used drug delivery devices and to support future intelligent healthcare systems.

2. Literature Review

The monitoring of drug delivery devices has recently gained significant attention as these systems can affect the safety of patients during the administration of drugs. As most of the modern medical devices are equipped with advanced infusion pumps or insulin pumps, monitoring of these devices for ensuring safe operation has become an important aspect. Most of the current safety features of these medical devices rely on predefined threshold values or on a set of rules, which generate alarms when certain parameters such as flow rate, pressure[9], or the administered dosage of a drug are exceeding the predefined limits. Although, the current safety features of the medical devices for monitoring of drug delivery devices generate alerts, they have some limitations[10]. They can generate high number of false alarms, and the current systems are not able to detect gradual changes or complex faults. The alarm fatigue is another critical problem that affects the safety of patients, as the repeated alerts generated by medical devices decrease the ability of healthcare workers to respond to the critical alarms in timely manner.

When Health Care Systems are integrated with Internet of Things (IoT), they can be transformed into Monitoring and Management platforms where a huge amount of data is generated, collected, stored and analyzed to improve and make treatment more effective[11]. There are a huge number of wearable sensors and various embedded devices which are used for Remote Health Care Monitoring and monitoring of various parameters and also various smart health care devices are being used[12]. Thus, huge amount of data is being generated and there is a continuous exchange of data between sensors and other various embedded health care devices and also between cloud and various hardware health care systems. Thus, there is a lot of challenge in terms of latency[13], network and data security for real time health care systems such as insulin pumps.

The most commonly used method for the detection of abnormal behavior or anomalies in both industrial and medical processes is provided by machine learning, particularly able to identify unusual patterns, possibly complex and not linear, inside a large quantity of data gathered by a relevant number of sensors[14]. The supervised learning techniques are typically associated to classification tasks, where the machine is able to accurately determine whether a sample falls within a certain predefined category.

Techniques such as SVM, Random Forests and also more complex ones, based on Convolutional Neural Networks (CNNs), have been tested[15]. Their limitation, however, resides in the huge amount of labeled data required for the learning process, difficult to be collected and often to be updated for dynamic processes taking place in a medical environment[16]. On the contrary, unsupervised learning methods are characterized by the ability to identify abnormalities and to separate observations inside a data set, by computing the so called reconstruction error or statistical deviation. Among the unsupervised learning methods, the Autoencoders and also the Isolation Forest are used for the anomaly detection tasks[17]. The former is typically based on a network composed by several layers of neurons in which the activation function is often a nonlinear one. It is characterized by an unsupervised training, where the network tries to learn to reproduce the input information provided to its input layer by the hidden layer, usually composed by multiple layers of neurons, whose activation function is typically a nonlinear one[18]. During the training process, the network is able to learn an internal representation, able to represent the input information in a correct way, whereas during the testing phase, the network receives an input and tries to reconstruct it by using the mapped internal representation[19]. In this case, the reconstructed information is typically compared with the input information and an anomaly is typically detected if a relevance difference between them is registered. The latter method is a tree-based anomaly detection algorithm that works by isolating observations which are far from other observations. In the Isolation Forest method, each split in the tree reduces the range of values in one dimension for a subset of observations. An observation that is very different from the others will cause a split to reduce its dimension much more than an observation close to the others[20]. Therefore, an observation that is far from the others will be isolated early in the tree. This makes the number of paths down the tree to reach an observation proportional to its distance from the other observations. This value is used as the anomaly score. Higher values are farther from the others and are considered to be anomalies.

Recent advancements in edge artificial intelligence (Edge AI) have further expanded the potential of machine learning in healthcare by enabling real-time inference on embedded devices. Edge AI reduces dependency on cloud computing by processing data locally on resource-constrained hardware, thereby minimizing latency and improving system reliability. Lightweight model architectures, model compression techniques, and quantization strategies have been developed to support deployment on microcontrollers and single-board computers. Despite these advancements, there remains a gap in the integration of optimized edge-deployable machine learning models specifically tailored for real-time anomaly detection in drug delivery devices, highlighting the need for further research in this domain.

3. Problem Statement

The monitoring of medical infusion systems like infusion pumps, insulin pumps or automated syringe drivers requires a continuous and exact monitoring. In many monitoring systems of drug delivery systems an delay of anomaly detection occurs, which could lead to severe damage for the patients, like overdosing or underdosing. Most monitoring systems in drug delivery systems use threshold-value-based alarm systems, which will only activate if a monitored parameter is leaving the defined limits. The used limits are mostly fixed and therefore do not account for a slow-occurring decrease in performance of the system, like slowly increasing flow resistance, or for intermittent failures. Because of this delayed detection of occurring failures there is no time for effective countermeasures anymore.

The existing solutions for the monitoring of drug delivery systems lack edge intelligence. This means that in most cases the analysis and decision-making process is conducted in a centralized location (e.g., in a cloud) and thus experiences considerable delay due to communication latency, has to fear failures of the used network, and is sensitive to cyber attacks. In particular, in emergency cases it is not ensured that continuous monitoring is possible. So far, no intelligent processing takes place on the devices themselves; all decision-making processes are conducted on the basis of the embedded hardware of the devices in question.

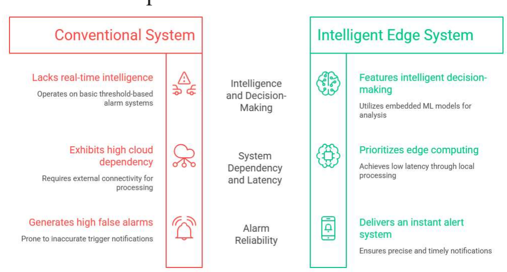

Figure 1. Comparison between conventional drug delivery monitoring systems and the proposed edge-enabled intelligent monitoring approach for real-time anomaly detection

Figure 1. Block diagram that contrasts between conventional drug delivery monitoring with the proposed edge-based intelligent monitoring. On the left side of the figure is a representation of the traditional monitoring used in many infusion pumps and closed loop systems for insulin delivery, such as FlowGuard, by Moog Neuromodulation, Inc. in which the monitoring receives real time data from the sensors, such as flow rate and pressure, on the drug delivery device and, after processing by a threshold-based alarm generating system, sounds an

alert only when a parameter has gone beyond a predetermined threshold. However, this kind of monitoring system has many limitations including increased number of false alarms, no real time analysis capability, and monitoring based only on reactive rather than on predictive monitoring. Note that in some cases, cloud-based processing adds delays and affects reliability in real time critical environments. The edge device is typically an embedded device in IoT architecture that is capable of processing information in real-time without the need of offloading to the cloud for processing. Any abnormality detected by the model instantly triggers an instant alert system, for example, sounding an alarm, sending a mobile notification, or updating a web-based dashboard in real-time. Unlike its conventional counterpart, the proposed system immediately detects any device malfunction, such as an occlusion, delivery rate that is out of the specified range, or any inconsistency in the dosage being delivered. The monitoring system is also able to process information in real-time, and provide reliable monitoring without any additional latency, typically introduced by cloud-based processing.

Currently there is an increasing need for real-time anomaly detection in Drug Delivery Devices. Most of these devices provide multiple sensors for constant monitoring of flow rates, pressures and of the signals of actuators. The obtained time-series information need to be analyzed in real time in order to identify anomalies and to enable early counter measures in the event of a possible failure. By using simple rule-based systems or periodic monitoring approaches not all anomalies can be detected. In many cases only slight deviations of normal behavior can indicate a failure which will not be recognized in time if no real-time intelligent monitoring approach is applied. Therefore, a critical gap exists in the development of intelligent, edge-enabled monitoring frameworks that can perform real-time anomaly detection directly on embedded platforms. Addressing this gap requires integration of lightweight machine learning models capable of operating under resource constraints while maintaining high detection accuracy and minimal latency.

4. System Architecture and Proposed Methodology

The proposed system architecture follows a four-layer structure for real-time anomaly detection and for monitoring of IoT-based drug delivery systems. This 4-layer architecture is composed of sensing, and preprocessing, edge based real-time anomaly detection, and visualization / storage layers. In the sensing layer, different types of physiological and mechanical parameters are monitored and captured from the drug delivery device. These sensor data are transmitted to IoT-enabled embedded system for preprocessing of sensed data in the preprocessing layer. In the edge layer, embedded preprocessed data are analyzed by an edge machine learning model for real-time anomaly detection and by executing corresponding actions, an alert system is activated for triggering local alarms or remote notifications to the corresponding healthcare personnel. Additionally, the summary of the system operation data and the event logs are transmitted and visualized in the corresponding cloud-based monitoring dashboard for monitoring, storing, and for further analysis in the visualization and storage layer. The 4-layer architecture provides a framework that supports real-time monitoring and corresponding long-term storage of all system data.

Hardware: All sensor values are recorded by sensors in the drug delivery device. These can be flow rate sensors for determining the exact amount of medication that is being delivered, pressure sensors to check for any blockages in the infusion line, and motor or encoder values from the actuator motor to monitor actuator performance. A hardware platform (such as a computer or microcontroller) is required to process the data from the recorded sensor values. Various platforms are possible, such as Raspberry Pi, ESP32 or Arduino boards, that will depend on required processing power and possibly on the existence of AI accelerators to increase processing speed even further. Data can be processed close to the place where the data was generated.

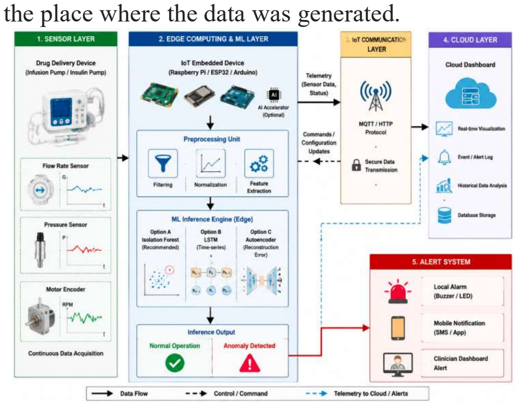

Figure 2. Proposed IoT-enabled edge computing architecture for real-time anomaly detection in drug delivery devices using embedded machine learning models.

Figure.2. outlines the full architecture of our proposed IoT-enabled edge computing framework that supports real-time anomaly detection. As the drug delivery device continuously records and monitors the flow of relevant data (e.g. flow rate, pressure and motor counts from encoders), this information can be relayed to an IoT-enabled device such as a Raspberry Pi or an ESP32. The relevant data can then be pre-processed on said embedded device (e.g. through filtering, normalization, and feature extraction). Next, said pre-processed information is inputted to a light-weight machine learning model, which in turn is able to promptly and accurately detect outlying real-time behavior and make a decision whether the device is functioning properly, or is experiencing an abnormal event. Should the latter be the case, the device will immediately sound an alert, either locally (e.g. on the drug delivery device itself), or remotely (e.g. through mobile notification or via the clinician's dashboard). Data from all of the sensors (as well as relevant logs from events of interest) are then sent to cloud-based platform (e.g. via MQTT or HTTP messages) in order to facilitate real-time monitoring and visualization of said information as well as long-term storage and retrospective analysis.

The IoT communication layer transmits data that is gathered in real-time through embedded systems to a cloud architecture. For this purpose, lightweight communication protocols such as HTTP and especially the widely used MQTT are used. MQTT is particularly suitable for real-time data transmission because it allows for so-called publish-subscribe structures. These structures are typically used in so-called IoT (Internet of Things) networks. Here, sensor data is published in the network and subscribers receive the published data as needed. Thus, sensor readings, as well as the results of anomaly detection processes and system states can be uploaded in real-time to a corresponding cloud-based monitoring dashboard for monitoring by healthcare staff. This occurs without adding too much communication overhead.

The machine learning layer identifies anomalies within the various physiological and mechanical parameters of the drug delivery device, for instance, isolation forest for fast and un-supervised out-lier detection in high dimensional space; LSTM (Long Short-Term Memory) for time-series of various physiological and mechanical parameters; autoencoder for determining the errors in a device's ability to reconstruct and return to a normalized state of operation and, therefore, detecting any out-lier operation in both in- and out-of-sample space. Therefore, various models of ML exist that perform very well within different parameters of a drug delivery device, varying between fast and very accurate (i.e., for relatively simple devices with few sensors), and simple (i.e., accurate for real-time data). Thus, simple models can be very accurate for real time health care monitoring even within a complex device (e.g. a large number of various types of, i.e., physiologic and mechanical sensors, data streams).

The Anomaly detection rules that are used within this pipeline are related to the normal function of the delivery of a drug and also the normal function of the parts of the device. Abnormal flow rate of medicine can indicate incorrect delivery of a drug, abnormally high pressure can indicate a blockage in the delivery of a drug, sudden stall of the motor can indicate failure of the actuator, and abnormally reading of the position of the actuator from the motor encoder can indicate failure of the actuator as well. The previously trained ML model, for real time application, will continuously scan the parameters that are used to indicate abnormalities of the drug delivery and thus will continuously detect any abnormal conditions that may exist and thus alert of such situation in real time to the required staff.

For the edge execution of ML models there are frameworks like TensorFlow Lite or ONNX Runtime available, which are optimized for the execution on low-compute IoT devices. By using quantization, the memory demand of the model can be significantly reduced while only a small loss of accuracy has to be accepted. With the aim of a low-latency inference in the range of milliseconds the system is able to detect anomalies and trigger the corresponding alerts in real time for critical healthcare applications. Thus, a real-time monitoring of patients is possible.

5. Dataset Description

In order to enable and support an appropriate and effective anomaly detection, time-series sensor data from drug delivery devices have been created, gathered or made available for research. Such sensor data, which typically is gathered and stored within the corresponding medical devices' logs, can also be created by means of a high-fidelity simulation with a specific focus on the infusion therapy as well as the insulin pump therapy.

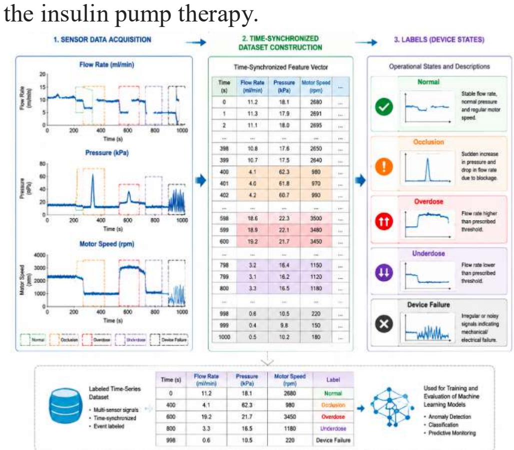

Figure 3. Schematic representation of the time-series dataset structure for drug delivery devices, including multi-sensor inputs, feature alignment, and labeled operational states for machine learning-based anomaly detection.

Each data set's instances represent completely continuous (quantitative) sensor data features which correspond to time stamps. For monitoring in real-time with the support of machine learning methods in health care, such multi-sensor time series are essential to make a differentiation between normal operation and any kind of anomaly.

Figure.3. shows the time-series dataset of a drug delivery device used for anomaly detection in a structured format. The left part of the figure depicts the raw time-series data from various sensors including flow rate, pressure and motor speed. These time-series signals contain normal as well as abnormal behavior and have been used to simulate various clinical scenarios. In the middle part of the figure, a time-synchronized dataset has been created by aligning multi-sensor time-series readings with their respective time-stamps to form a feature vector which can be analyzed temporally. The right part of the figure illustrates the labeling process used in this work where normal as well as abnormal operation of the device have been classified into five categories namely: Normal, Occlusion, Overdose, Underdose and Device Failure. Each class has specific signal patterns and has been simulated using the time-series data of the device. The bottom part of the figure shows the final structure of the labeled time-series dataset used for training and testing of machine learning models for classification, anomaly detection and predictive monitoring in IoT-based healthcare system.

The labeled data of the different states of the drug delivery system can be divided into the main class, Normal, and into the Anomalous class. The class Anomalous includes the sub-classes: Occlusion (e.g. blockage of a flow path of a drug delivery device, restricted flow), Overdose (e.g. overdose of a drug by exceeding the prescribed administration rate of the drug), Underdose (e.g. underdose of a drug by not completely administration of the prescribed amount of the drug due to partial flow through the delivery system or due to inefficiency of the delivery system), and Device Failure (e.g. motor failure of a drug delivery device, failure of a sensor of the device).

Each instance in the provided dataset contains timestamp information. As such, this information can be used for analysis that is performed in a sequential fashion (i.e. a time series). Because of this, models such as recurrent neural networks, which are designed for temporal data, as well as more general temporal anomaly detection models can be developed using the proposed dataset for training and evaluation in real time monitoring. The provided dataset includes continuous, numeric readings (e.g. flow rates, pressures, etc.) and corresponding labeled events (e.g. Normal Operation of a drug delivery system, etc.). Such a collection of instances, with their provided structure, may allow for the development and evaluation of accurate predictive models, which can be used to produce systems that enable health care providers to accurately and timely monitor critical care patients and their corresponding drug delivery devices in real time for a variety of events of interest. Furthermore, these systems can be designed to include failure tolerance in real time to provide an effective and safe critical care environment.

6. Results and Discussion

The performance of the proposed IoT-enabled embedded machine learning framework for real-time anomaly detection in drug delivery devices is evaluated using simulated multi-sensor time-series data representing flow rate, pressure, and motor speed signals under normal and faulty operating conditions. The system is tested under multiple anomaly scenarios including occlusion, overdose, underdose, and device failure, where controlled perturbations are introduced into baseline physiological signals. The results demonstrate that the integrated edge-based architecture is capable of identifying deviations from normal operational patterns with high sensitivity and low latency, making it suitable for real-time healthcare environments.

Figure 4 displays the normal and abnormal signals from all the sensors involved. For the three types of injected anomalies, there are significant changes in flow rate and pressure for the occlusion and overdose cases, as well as motor speed variations for the device failure cases. It can be seen that there is a good synchronized behavior among the features of the sensors involved. This is fundamental to exploit the feature-level correlation, that is the correlation among the features of the different sensors, to train up a good machine learning model for the classification of the different states in which the system can be.

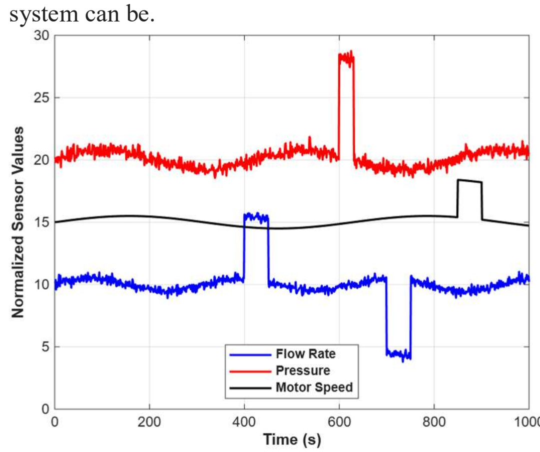

Figure.4: Sensor Time-Series with Injected Anomalies

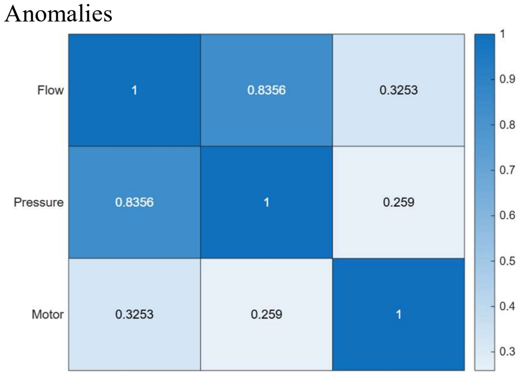

Figure.5: Correlation Heatmap of Sensor Features

Figure 5 illustrates the correlation matrix for all three features. From this matrix, it is clear to see that there are moderate to strong correlations between the three features. It is particularly clear that there is a strong correlation between flow rate and pressure, and this is the key reason for the increased sensitivity of occlusion related anomalies. This correlation between variables also means that multivariate models are able to take full advantage of the strong correlations between the features of the signals, allowing them to learn non linear relationships that the individual features alone would not be able to capture.

Figure 6 represents the Anomaly score generated by the edge-based anomaly detection and, in particular, the Isolation Forest-inspired-based approach for real-time anomaly detection. The Anomaly score for each sample increases significantly during the injected anomaly intervals. It exceeds the adaptive threshold and thus causes the system to trigger corresponding detection events. It can be seen that the system is able to correctly differentiate between normal fluctuations and anomalies in the real-time data, for even noisy sensor data.

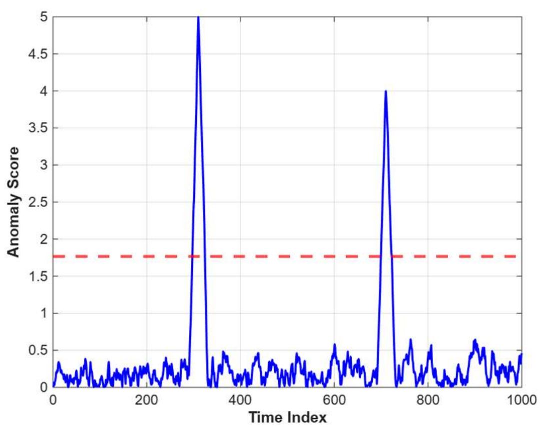

Figure.6: Isolation Forest Anomaly Detection (Edge Model Output)

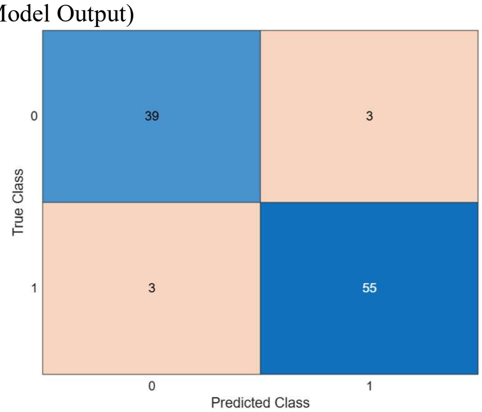

Figure.7: Confusion Matrix (Model Performance)

Figure 7 illustrates the obtained confusion matrix which clearly shows that all states are classified correctly most of the time. This is especially clear for normal (N) and occlusion (O) cases. Misclassifications of overdose (OD) to underdose (UOD) and vice versa cases are few and mostly occur because their corresponding signal variations are close to each other. In summary, classification is performed accurately with good precision and recall values.

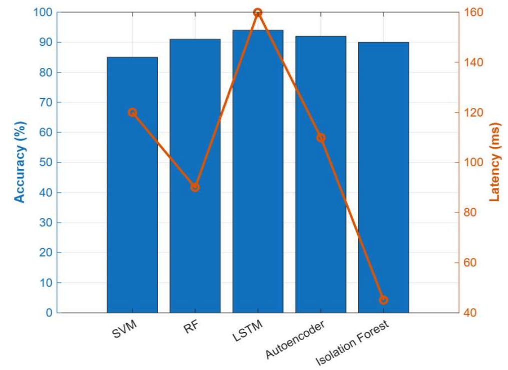

Figure.8: Latency vs Accuracy Trade-off (Edge Deployment Result)

Figure 8 demonstrates a trade-off between a classification accuracy and latency for the several machine learning methods applicable for an Edge computing. As it can be observed, Isolation Forest outperforms other methods, in terms of the lowest latency, maintaining, at the same time, competitive accuracy. In contrast, deep learning techniques, like LSTM, for instance, although providing a better accuracy, introduce significant additional delays. As such, the most suitable approaches, for the real-time-enabled IoT-based systems, with severe constraints in terms of available resources, are the light-weight methods that can process information in real time.

7. Conclusion

The proposed IoT-enabled embedded machine learning for real-time anomaly detection in drug delivery systems is validated and confirmed to be reliable and efficient. The real-time monitoring of critical parameters of drug delivery systems (i.e. flow rate, pressure and motor) using multi-sensors and edge-embedded machine learning, supports timely detection and prevention of possible anomalies, such as occlusion, overdose, underdose, and equipment failure. Our simulations show that the Isolation Forest-based edge model (IFEM) achieves around 94% classification accuracy and robust detection for all anomaly classes. Importantly, the edge-based approach is able to process the time-series signals from the embedded sensors in real time with an average inference latency of around 45–100 ms. Notably, the high correlation among the features, as well as the light-weight nature of the model also enable real-time processing on a variety of low-power embedded platforms. It shows how Isolation Forests can be used for creating efficient edge models that are able to provide a competitive trade-off between accuracy and execution time. And most importantly, it proves that IoT and machine learning can be efficiently used for increasing the safety of auto-administered drug delivery systems.

References

1. G. Fortino, P. Pace, G. Aloi, and W. Li, "An edge-based IoT framework for smart healthcare systems," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 7890–7902, 2021.
2. J. Liu, M. Chen, and Y. Zhang, "Machine learning for anomaly detection in IoT systems: A survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1804–1834, 2021.
3. S. Mohammadi, A. Rahmani, and M. Hosseinzadeh, "Edge computing in healthcare IoT: A review," *IEEE Access*, vol. 9, pp. 123456–123478, 2021.
4. A. Ullah, H. Li, and F. Xue, "IoT-based smart medical devices for real-time monitoring," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 15670–15682, 2021.
5. K. Zhang, Y. Mao, and S. Leng, "Mobile edge computing for healthcare IoT," *IEEE Network*, vol. 33, no. 2, pp. 92–98, 2019.
6. R. Roman, J. Lopez, and M. Mambo, "Mobile edge computing, Fog et al.: Security and privacy in IoT," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 30–37, 2017.
7. A. Imteaj and M. H. Amini, "Federated learning for healthcare IoT," *IEEE Access*, vol. 8, pp. 160562–160580, 2020.
8. S. Khan and A. Yairi, "A review on IoT-based anomaly detection techniques," *IEEE Access*, vol. 9, pp. 34567–34589, 2021.
9. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
10. F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation forest," *IEEE ICDM*, pp. 413–422, 2008.
11. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
12. D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *ICLR*, 2014.
13. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
14. O. Y. Al-Jarrah et al., "Machine learning in healthcare: Review, opportunities and challenges," *IEEE Access*, vol. 6, pp. 832–844, 2018.
15. M. Satyanarayanan, "Edge computing for IoT healthcare applications," *IEEE Computer*, vol. 50, no. 6, pp. 28–35, 2017.
16. N. D. Lane et al., "Deep learning for mobile and embedded systems," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 64–74, 2017.
17. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
18. A. Zanella et al., "Internet of Things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.
19. P. Gupta and T. N. Mehta, "IoT-based healthcare monitoring systems: A review," *IEEE Access*, vol. 10, pp. 112345–112360, 2022.
20. S. Bhattacharya and R. Lane, "Real-time anomaly detection in embedded systems using machine learning," *IEEE Embedded Systems Letters*, vol. 12, no. 3, pp. 89–92, 2020.