AN EARLY WARNING
PIPELINE FAULT DETECTION SYSTEM
USING MACHINE LEARNING

An Early-Warning Pipeline Fault Detection System Using
LSTM Autoencoder with Multi-Parameter Sensor Monitoring
for Anomaly Detection and Incident Management

August 2026

Prepared by:

Dulminí Wijerathna | Niyas Abdulla | Dinasi Hirunima | Nishanth Nadarasa | Ruzayn Ahmed

Content

1.0 INTRODUCTION

Fluid pipelines are widely used in industrial processes to transport liquids and gases between different process units. Reliable operation of these pipelines is essential for maintaining process continuity, minimizing material losses, protecting equipment, and ensuring operational safety. However, pipeline systems can experience abnormal operating conditions due to leakage, blockage, pressure disturbances, mechanical degradation, and other forms of deterioration. If these conditions are not identified at an early stage, a developing abnormality can progress into a major pipeline failure.

Conventional pipeline monitoring approaches commonly rely on direct inspection, threshold-based alarms, pressure monitoring, flow measurements, and other established leak detection techniques. These methods can be effective for identifying an existing abnormal condition; however, their effectiveness may be limited when the objective is to recognize the early stages of a developing fault. Existing research has investigated mass or volume balance, negative pressure waves, pressure point analysis, statistical techniques, transient modelling, acoustic sensing, and vibration-based monitoring for pipeline leak detection [1], [2].

The main objective of the proposed project is therefore to move beyond detecting a fault only after it has occurred. The proposed system aims to identify early changes in measurable pipeline parameters that indicate a developing fault before the fault becomes critical. This requires continuous monitoring of pipeline operating conditions and analysis of the relationships between multiple parameters rather than relying on a single measurement.

Parameters such as flow rate, pressure, temperature, and vibration can provide complementary information about pipeline behavior. A developing leak, for example, can cause changes in the hydraulic conditions of the pipeline, while associated disturbances may also be reflected in pressure waves or vibration signals [3], [4]. The combination of multiple sensor measurements can therefore provide a more informative representation of the pipeline's operating condition.

Recent research has demonstrated increasing interest in machine-learning approaches for pipeline leaks and anomaly detection. Machine-learning models can identify complex patterns within sensor data and distinguish abnormal behavior from normal operating conditions [5], [6]. This provides an opportunity to develop monitoring systems that can recognize subtle deviations before they develop severe failures.

The proposed system integrates sensors, a data acquisition unit, and a machine-learning-based analysis stage. Sensors continuously measure selected pipeline parameters, while a microcontroller-based data acquisition system collects and transfers the measurements. The acquired data are then processed and analyzed to identify abnormal patterns associated with an emerging fault. When a potentially developing fault is identified, an early warning can be generated to support preventive maintenance.

Thus, the proposed system is intended to shift pipeline maintenance from a predominantly reactive approach toward a proactive early-warning approach, where developing faults can be recognized before they progress into critical failures.

1.1 Objectives

The main objective of this project is to develop a sensor-based pipeline monitoring system capable of identifying early indications of developing faults before the occurrence of a critical pipeline failure. The specific objectives are as follows:

1. To continuously monitor relevant pipeline operating parameters using suitable sensors.
2. To investigate changes in flow rate, pressure, temperature, and vibration associated with abnormal pipeline conditions.
3. To develop a data acquisition system for collecting and transmitting real-time sensor measurements.
4. To establish characteristic patterns of normal pipeline operation and identify deviations that may indicate an emerging fault.
5. To investigate the application of machine-learning techniques for identifying abnormal patterns in the collected sensor data.
6. To develop an early-warning mechanism capable of indicating a potential fault before it develops into a major failure.
7. To provide information that can support preventive maintenance and reduce the consequences of unexpected pipeline failures.

1.2 Significance of the Project

Early identification of pipeline faults is important because a developing fault can result in material loss, process interruption, equipment damage, environmental impacts, and safety hazards if it is allowed to progress. Prompt identification of abnormal pipeline behavior can provide operators with an opportunity to investigate the condition and perform preventive maintenance before a critical failure occurs [1], [5].

The proposed system is significant because it combines multi-parameter sensing and data-driven analysis rather than depending solely on a single threshold or measurement. Pressure and flow measurements can provide information about hydraulic changes, while temperature and vibration measurements can provide additional information about changes in the physical operating condition of the pipeline. Research on vibration-based leak detection has demonstrated the potential of accelerometer-based measurements for identifying leakage-related disturbances [4]. The use of machine learning further provides the possibility of identifying relationships and patterns within multiple sensor signals. Recent reviews have shown increasing application of machine-learning techniques to pipeline leakage detection, including approaches based on pressure, flow, acoustic, and other sensor measurements [5], [6].

The significance of the project can therefore be summarized as follows:

- Early warning: Provides an indication of developing abnormal behavior before a critical failure.
- Preventive maintenance: Allows maintenance activities to be planned before major damage occurs.
- Reduced downtime: Early intervention can reduce unexpected process interruptions.
- Improved safety: Early recognition of abnormal pipeline conditions can reduce the consequences of hazardous failures.
- Reduced losses: Early identification can minimize fluid losses and associated economic impacts.
- Data-driven monitoring: Continuous sensor measurements provide quantitative information about pipeline conditions.
- Scalability: The proposed sensor and data acquisition concept can potentially be adapted to different pipeline monitoring applications.

2.0 LITERATURE REVIEW

Pipeline fault detection has been investigated using a wide range of techniques, from conventional inspection and hydraulic monitoring to advanced sensor systems and data-driven machine-learning approaches. Existing methods differ in their operating principles, detection speed, sensitivity, cost, and ability to identify small or developing faults.

The literature relevant to this project can be broadly categorized into conventional pipeline fault detection methods, sensor-based multi-parameter monitoring, and data-driven or machine-learning approaches.

2.1 Conventional Pipeline Fault Detection Methods

Traditional pipeline fault detection methods include direct visual inspection, mass or volume balance, pressure monitoring, negative pressure wave analysis, and mathematical or hydraulic modelling. Computational reviews have classified pipeline leakage detection methods into categories such as mass/volume balance, negative pressure wave, pressure point analysis, statistical methods, and real-time transient modelling [1].

Mass or volume balance methods compare the amount of fluid entering and leaving a pipeline section. A significant difference between the inlet and outlet measurements can indicate a possible leak. Although this approach is relatively straightforward, measurement uncertainty and changes in normal operating conditions can result in false alarms or reduced sensitivity to small leaks [1].

Negative pressure wave methods identify pressure disturbances generated by sudden leakage from a pressurized pipeline. The resulting pressure wave can propagate along the pipeline and be detected using pressure sensors. Research has demonstrated the use of negative pressure wave techniques together with advanced signal-processing methods for pipeline leak detection [3].

Direct inspection methods, such as manual observation and periodic inspection, can also identify visible signs of pipeline deterioration. However, such approaches are generally dependent on inspection frequency and human observation and may not provide continuous monitoring of the pipeline condition.

A major limitation of conventional approaches in relation to the present project is that many methods are designed primarily to identify a leak or abnormality after a measurable disturbance has become established. The proposed project instead focuses on identifying early deviations from normal behavior that may precede a significant fault.

2.2 Sensor-Based Multi-Parameter Monitoring

Sensor-based monitoring provides a method for continuously observing the physical condition of a pipeline. Pressure, flow rate, temperature, acoustic signals, and vibration can be measured using appropriate sensors and used to characterize pipeline operating behavior.

Pressure and flow measurements are particularly important for hydraulic pipeline monitoring. A review of data analytics for water distribution networks highlighted the use of pressure and flow smart sensors for timely leakage detection and analysis [7]. Changes in these parameters can provide information about deviations from expected pipeline behavior.

Vibration-based monitoring provides another source of information. Leakage can produce pressure fluctuations and mechanical disturbances that propagate through the pipeline structure. A review published in IEEE Access investigated vibration detection techniques using accelerometer sensors for water pipeline leakage and identified vibration sensing as a useful approach for pipeline monitoring [4].

Acoustic and pressure-wave methods can also be used to detect disturbances associated with leakage. Pressure-wave-based research has investigated different acoustic and pressure-wave sensing techniques, together with signal processing and machine-learning approaches [2].

The use of multiple parameters provides an advantage over single-parameter monitoring because different faults may produce different combinations of changes. For example, an emerging leak may be associated with changes in flow rate and pressure, while physical disturbance may also produce a change in vibration. Combining these measurements can therefore provide a more comprehensive representation of pipeline conditions.

2.3 Data-Driven and Machine Learning Approaches

Data-driven approaches have become increasingly important in pipeline monitoring because modern sensor systems can continuously generate large quantities of operating data. Machine-learning methods can be used to learn relationships within these measurements and identify patterns that differ from normal operating behavior.

Machine-learning techniques have been increasingly investigated for pipeline leakage detection. A recent review of machine-learning applications in fluid pipeline leakage detection identified a growing range of data-driven approaches applied to oil, gas, and water pipelines [5]. These approaches can use parameters such as pressure and flow measurements to identify leakage-related patterns.

Different machine-learning methods can be used depending on the available data and the objective of the system. Supervised learning methods can classify known normal and fault conditions, while unsupervised or semi-supervised approaches can identify deviations from normal behavior when labelled fault data are limited. The selection of the appropriate model depends on factors such as dataset size, sensor characteristics, fault types, computational requirements, and the availability of labelled data.

Machine-learning approaches can also support the analysis of multiple sensor signals simultaneously. This is particularly relevant to the present project because the early stage of a fault may produce relatively small changes in individual measurements, while the combined changes across several parameters may provide a stronger indication of abnormal behavior.

Recent work has also investigated the use of machine learning together with pressure, flow, vibration, acoustic signals, and time-dependent analysis for pipeline leakage detection [2], [6]. These developments demonstrate the potential of combining physical measurements with data-driven methods to improve pipeline monitoring.

However, the objective of the proposed project is more specific than simply classifying an already-developed leak. The desired system should identify the transition from normal operating behavior toward a developing fault, thereby providing an opportunity for early intervention.

2.4 Comparison of Existing Approaches

Different pipeline fault detection approaches provide different advantages and limitations. A comparison of commonly used approaches is presented in Table1.

Table 1 Comparison of pipeline fault detection approaches

Detection approach	**Main measurement/information**	**Advantages**	**Limitations**
Pressure-based	Pressure	Simple and relatively easy to implement	Can be affected by normal pressure fluctuations
Flow-based	Flow rate	Directly reflects changes in fluid transport	Sensitive to measurement uncertainty and operating conditions

Mass/volume balance	Inlet and outlet flow	Useful for identifying discrepancies	Small leaks can be difficult to detect
Acoustic/vibration-based	Acoustic or vibration signals	Can provide information about leak-related disturbances	Sensitive to noise and sensor placement
Multi-parameter sensing	Pressure, flow, temperature, vibration, etc.	Provides a more comprehensive representation of system behavior	Requires multiple sensors and data processing
Machine learning-based	Multiple sensor/data features	Can learn complex patterns and classify conditions	Requires suitable training data and model validation

Existing research therefore demonstrates that no single detection approach is universally suitable for every pipeline application. The effectiveness of a method depends on the pipeline configuration, fault characteristics, measurement quality, and operating environment [1], [2], [4].

2.5 Research Gap

The reviewed literature demonstrates that pipeline faults can be detected using hydraulic measurements, pressure waves, acoustic signals, vibration sensors, statistical approaches, and machine-learning methods [1], [7]. However, several challenges remain.

First, conventional methods may be more suitable for detecting established leakage conditions than for identifying the earliest stages of fault development. Second, single-parameter monitoring may not provide sufficient information when the change associated with an emerging fault is small. Third, although machine-learning approaches have increasingly been applied to pipeline leakage detection, their effectiveness depends strongly on the quality, quantity, and representativeness of the sensor data [5].

Another important challenge is distinguishing actual developing faults from normal changes in operating conditions. Changes in flow demand, pressure, temperature, and other process variables can occur during normal operation and may resemble abnormal conditions. Therefore, a reliable system must consider the relationships between multiple parameters and their changes over time.

Based on these limitations, the research gap addressed by this project is the development of an integrated sensor-based monitoring concept that focuses on recognizing early abnormal patterns associated with developing pipeline faults before a critical failure occurs.

The proposed approach combines multi-parameter sensing, real-time data acquisition, and machine-learning-based analysis. This provides a foundation for moving from conventional fault detection toward early anomaly detection, early warning, and preventive intervention.

3.0 PROPOSED SYSTEM

The proposed system consists of a pipeline monitoring section, multiple sensors, a data acquisition unit, a data-processing stage, a machine-learning-based analysis stage, and an early-warning mechanism.

The sensors continuously measure selected pipeline parameters and transfer the measurements to the data acquisition unit. A microcontroller-based platform such as the ESP32 can be used to collect and transmit sensor measurements. The collected data are then processed and supplied to the machine-learning stage.

The machine-learning stage analyzes the measured parameters to identify deviations from normal pipeline behavior. If the observed behavior deviates sufficiently from the learned healthy baseline and satisfies the defined alert criteria, the system generates an early warning.

The overall objective is therefore:

Normal operation → Parameter changes → Early abnormal pattern → Developing fault indication → Warning → Preventive action

The main advantage of this framework is that it does not treat fault detection as a single event occurring after failure. Instead, it considers pipeline condition as a continuously changing process and aims to recognize abnormal behavior during the early stages of fault development.

3.1 Operating Principle

During normal operation, the pipeline produces characteristic values and patterns of flow rate, pressure, temperature, and vibration. These measurements are continuously collected to establish the normal operating behavior of the system.

When a fault begins to develop, the physical conditions within and around the pipeline may change. These changes can affect the measured parameters. For example, a developing leak may alter the flow distribution and pressure conditions, while the leakage flow can generate additional acoustic or vibration-related disturbances [2], [3].

The proposed system continuously compares the incoming sensor information with the learned characteristics of normal operation. Rather than waiting for a large deviation, the system seeks to identify persistent or combined changes that may indicate the beginning of an abnormal condition.

If the detected pattern reaches the defined criteria for a potential developing fault, an early warning is generated. This warning can allow an operator or maintenance team to inspect the system before the condition progresses to a critical failure.

3.2 Fault Detection Approach

The proposed fault detection approach consists of three main stages: establishing normal behavior, identifying deviations, and generating an early warning.

First, sensor data corresponding to normal pipeline operations are collected. These data represent the expected operating behavior of the system under different normal conditions.

Second, real-time measurements are compared with the learned or established normal patterns. Changes in individual parameters and relationships between multiple parameters are considered.

Third, if the observed behavior deviates sufficiently from the learned healthy baseline and satisfies the defined alert criteria, the system generates an alert.

The approach can therefore be represented as:

Sensor measurements → Data processing → Analysis → Anomaly Detection → Early Warning

The implemented fault-detection approach uses a healthy-only LSTM autoencoder for unsupervised anomaly detection. The model is trained using approved healthy operating data and learns the temporal relationships among pressure, temperature, flow rate, and vibration over a 60-second window. During monitoring, the reconstruction error between the measured sequence and the reconstructed sequence is calculated as an anomaly score. A high reconstruction error indicates that the current operating behavior differs from the learned healthy baseline. The model therefore provides evidence of abnormal behavior rather than directly predicting or classifying a specific physical fault.

The objective of the implemented approach is to identify abnormal behavior early enough to provide an opportunity for preventive action.

3.3 Data Acquisition and Processing

The collected sensor measurements are organized as time-series data and passed through a defined preprocessing and data-quality pipeline before being used by the machine-learning model. Incoming readings are first checked for missing values, abnormal gaps, frozen measurements, repeated values, and physically invalid readings. Short gaps of approximately 1–3 seconds may be filled only in the processed model stream, while longer gaps make the affected window unavailable for scoring. Isolated single-sensor spikes may also be cleaned when they are identified as likely sensor glitches, whereas persistent or multi-sensor changes are preserved because they may represent genuine operating disturbances.

The system retains both the raw sensor stream and the processed stream so that any modification made during preprocessing remains traceable. After quality checking, the processed measurements are arranged into a chronological 60-second × 4-sensor window. The four sensor features are then robustly scaled using the median and IQR obtained from approved healthy training data before the resulting window is supplied to the LSTM autoencoder.

3.4 Overall System Workflow

The complete workflow of the proposed system can be summarized in the following stages:

1. Pipeline operation: The pipeline operates under normal or abnormal conditions.
2. Sensor measurement: Relevant operating parameters are continuously measured.
3. Data acquisition: Sensor outputs are collected by the data acquisition unit.
4. Data preprocessing: Noise, erroneous values, and unsuitable measurements are processed.
5. Feature preparation: Relevant characteristics of the sensor signals are extracted or organized.
6. Machine-learning analysis: The processed measurements are analyzed to identify abnormal patterns.
7. Anomaly detection: An anomaly is identified when the observed pattern differs sufficiently from the learned healthy behaviour.
8. Warning generation: An incident or early warning is generated according to the defined sudden-event or slow-degradation rules.
9. Operator interpretation: The operator reviews the anomaly evidence and determines the suspected physical cause.
10. Preventive action: Inspection or maintenance can be initiated where appropriate**.**

Pipeline Sensors → Sensor Data Acquisition → Data-Quality Checking → 60-Second × 4-Sensor Window → Robust Scaling → LSTM Autoencoder → Reconstruction Error (MSE) → Sudden-Event / Slow-Degradation Rules → Incident or Warning → Operator Interpretation → Preventive Action

4.0 FLOW CHANGES DUE TO FAULTS

The behavior of fluid flow through a pipeline is directly related to the pressure, resistance, geometry, and operating conditions of the system. When a fault develops, the hydraulic conditions may change, resulting in measurable variations in flow rate and pressure. These changes can be used as important indicators for identifying abnormal pipeline behavior.

For the proposed system, flow-related measurements are particularly important because the objective is to recognize changes that occur before the fault becomes critical.

4.1 Normal Pipeline Operation

Under normal operating conditions, the pipeline maintains a relatively stable relationship between flow rate, pressure, temperature, and other operating parameters, subject to expected process variations.

For a given operating condition, the measured flow rate and pressure are expected to remain within a characteristic range. Small fluctuations may occur because of pump operation, valve adjustments, temperature changes, or other normal process disturbances.

The normal operating dataset therefore provides a baseline against which subsequent measurements can be compared.

4.2 Effect of a Leak on Pipeline Flow

When a leak develops in a pipeline, a portion of the transported fluid leaves the intended flow path. This changes the flow distribution and can result in measurable changes in flow rate and pressure. For a developing leak, the changes may initially be small and may not exceed conventional alarm thresholds. However, continuous monitoring can capture the gradual change in the measured parameters.

The difference between expected and measured flow behavior can therefore provide an indication of a developing abnormality. Combining flow measurements with pressure and other sensor information can improve the ability to distinguish a developing leak from normal operating fluctuations [1], [7].

4.3 Pressure Variation

Pressure shows a significant decrease when the illustrated leak occurs. Before the fault, the pressure signal remains approximately stable around its normal operating level. Following the occurrence of the leak, the pressure drops to a lower level.

This pressure deviation provides an important indication of abnormal pipeline behavior. Continuous pressure monitoring therefore allows the system to identify changes that may be associated with a leak.

4.4 Flow Rate Variation

Flow rate is one of the primary parameters considered in the proposed monitoring system. Under normal conditions, the flow rate follows the operating requirements of the process. A developing fault can disturb this behavior and produce a deviation from the expected flow pattern.

The magnitude and rate of change of the flow deviation may provide information about the severity and development of the abnormal condition. However, flow rates should not be considered independently because normal process changes can also produce flow variations.

For this reason, the proposed system considers flow rate together with other sensor measurements. Multi-parameter analysis can provide a more reliable indication of whether a flow change is associated with a developing fault.

4.5 Temperature Variation

Temperature can influence fluid properties and pipeline operating conditions. Changes in temperature may affect viscosity, density, and consequently the relationship between pressure and flow.

A temperature change may also occur as a consequence of process conditions rather than a pipeline fault. Therefore, temperature measurements should be interpreted together with other parameters rather than being used as an independent fault indicator.

Continuous temperature monitoring can nevertheless provide useful contextual information for the machine-learning model. By incorporating temperature with pressure, flow, and vibration measurements, the system can distinguish normal operating changes from patterns that are more consistent with abnormal pipeline behavior.

4.6 Vibration Variation

Vibration provides an additional measurement for identifying physical disturbances in pipeline systems. Fluid leakage and other abnormal flow conditions can generate pressure fluctuations and mechanical disturbances that may be detected using vibration sensors or accelerometers.
Previous research has investigated accelerometer-based vibration detection for water pipeline leakage and demonstrated the relevance of vibration measurements for leak monitoring [4].
In the proposed system, vibration measurements can therefore complement flow and pressure measurements. A developing fault may produce a combination of small flow, pressure, and vibration changes that are more informative when analyzed together than when each parameter is considered separately.

4.7 Overall Fault Signature

A pipeline fault does not necessarily produce a single distinctive change in one parameter. Instead, a developing fault may be represented by a combination of changes in flow rate, pressure, temperature, and vibration.
The concept of the proposed system is therefore to identify an overall fault signature based on the combined sensor measurements. A simplified representation is:
Normal operating pattern → Small parameter deviations → Consistent abnormal pattern → Developing fault signature → Early warning
The machine-learning stage identifies deviations from learned normal operating behaviour across the combined sensor measurements.
This multi-parameter approach is particularly important for the project's main objective. The desired outcome is not merely to identify a pipeline after it has failed, but to recognize the early warning signs of a developing fault and provide sufficient time for preventive action.
The overall concept can therefore be summarized as:
Continuous sensing → Multi-parameter analysis → Anomaly detection → Early warning → Operator interpretation → Preventive maintenance
This approach provides the basis for developing a proactive pipeline monitoring system capable of identifying developing faults before they progress into critical failures.
It is important to distinguish these theoretical fault signatures from the operation of the implemented machine-learning model. The relationships described above are used to understand and interpret how physical pipeline faults may influence the monitored parameters; they are not predefined fault patterns programmed into the model. The implemented LSTM autoencoder is trained only on approved healthy operating data and learns the normal temporal relationships among pressure, flow rate, temperature, and vibration. During monitoring, it identifies deviations

from this learned healthy behaviour through reconstruction error but does not automatically classify the deviation as a leak, blockage, or other specific fault. The theoretical fault signatures therefore support the operator or engineer in interpreting the anomaly evidence generated by the model.

5.0 MACHINE LEARNING

Fluventa learns normal pipeline behavior from healthy pressure, temperature, flow, and vibration data. It detects deviations from this learned healthy behavior using an LSTM autoencoder. The machine-learning model does not independently identify whether a deviation represents a leak, blockage, or another specific physical fault. Instead, the anomaly evidence produced by the model is processed through separate rule-based sudden-event and slow-degradation pathways, which generate operator-facing incident or warning records. The operator or engineer then interprets the likely physical cause using the model evidence together with operational context.

5.1 Machine Learning Approach

A supervised classification approach, which requires labelled examples of both normal and faulty operation, was considered but not adopted as the primary detection mechanism. Reliable, representative fault data are often rare, incomplete, or specific to fault types that may never have occurred at a given checkpoint. An unsupervised anomaly-detection approach was therefore selected: the model is trained exclusively on approved examples of healthy operation and learns the normal relationship between the four monitored signals (pressure, temperature, flow, and vibration RMS) over a rolling one-minute window.

The implemented model is an autoencoder. During training, the network learns to compress and then reconstruct many overlapping one-minute examples of healthy operation. When presented with a new one-minute window during live monitoring, the model attempts to reconstruct that window using what it learned about healthy behavior. The reconstruction error — the mean squared error (MSE) between the input pattern and the model's reconstruction of it — is used as the anomaly score. A small reconstruction error indicates that the observed pattern resembles healthy operation; a large reconstruction error indicates that the pattern departs from what the model has learned, which is treated as evidence of unusual behavior rather than as an automatic diagnosis of a specific fault.

5.2 Data Collection

The four sensor signals are sampled once per second at the monitored checkpoint, producing a continuous time series of pressure, temperature, flow rate, and vibration velocity RMS. For model training, only data periods that have been explicitly reviewed and approved as healthy operation are used, so that the learned baseline reflects genuine normal behavior rather than unreviewed or ambiguous conditions. For the demonstration described in this report, deterministic synthetic data streams were used so that the same behaviour could be reproduced, inspected, and repeated during evaluation and presentation.

5.3 Data Preprocessing

Because the quality of the model output depends directly on the quality of its input, incoming readings are screened before being used for scoring. Table 2 summarizes how different data-quality conditions are handled.

Data condition	Implemented response	Reason
Gap of 1–3 seconds	May be filled only in the processed model stream	Preserves a short window without rewriting the original evidence
Longer gap	The affected window is not scored	A guess would be more misleading than an honest absence of evidence
Isolated spike on one sensor	May be cleaned in the model input and clearly labelled	Reduces sensitivity to a single-sample glitch
Several sensors change, or change persists	The change reaches the model	May be a real operating event rather than sensor noise
Frozen or invalid value	Quality status is shown and scoring can be suppressed	A plausible-looking number can still be untrustworthy

Table 2: Data-quality handling implemented in the preprocessing stage.

Two parallel records are retained throughout: the raw readings, which preserve exactly what arrived from the sensors, and the processed stream, which shows exactly what the model received after any gap-filling or spike cleaning. Keeping both makes every preprocessing decision reviewable rather than hidden.

5.4 Window Formation and Feature Preparation

The model does not evaluate individual readings in isolation. Instead, each usable comparison is built from a rolling window of the previous 60 one-second readings across all four sensors, forming a one-minute, four-channel pattern. Because a complete comparison requires 60 readings, the first 59 seconds after start-up are a deliberate context-building period during which live values are visible, but no model score is produced; the first usable score is generated only once the 60th valid reading completes the window. After that point, the window advances one second with each new valid reading, and a fresh comparison is produced continuously.

After the sensor-quality checks, the processed measurements are robustly scaled before being supplied to the machine-learning model. For each sensor feature, the scaled value is calculated using,

$$z_r = \frac{x - median}{IQR}$$

where the median and interquartile range (IQR) are calculated using only the approved healthy training data. Robust scaling places pressure, temperature, flow rate, and vibration on comparable numerical scales while reducing the influence of extreme values. This prevents features with larger numerical ranges from dominating the reconstruction-error calculation.

5.5 Machine Learning Model Training

Model training and deployment follow a controlled, three-stage lifecycle intended to keep the active model trustworthy and auditable:

- Teach: The autoencoder is trained using only data that has been reviewed and approved as a healthy operation.
- Check: The trained model's behavior is confirmed against a later, separate healthy period that was not used in training, to verify it generalizes sensibly rather than memorizing the training window.
- Freeze: The model is saved together with its matching data-scaling parameters and detection thresholds as a single, versioned package.

A newly trained model is always treated as a candidate rather than an automatic replacement for the active model. The candidate is trained without disturbing the currently deployed model; its evidence is compared against saved evaluation results and operational expectations, and it is deployed only after a person explicitly records an approval decision. Deployment is atomic: the model, its input-scaling parameters, and its detection thresholds move together as one matched package, because interpreting scores using a mismatched scale or threshold could produce convincing-looking but invalid results. The previously approved package is always retained as a controlled rollback point.

5.6 Alert and Notification

Sudden-event path: Every usable one-minute comparison is evaluated as it arrives. A single reconstruction-error score above the defined opening level starts an incident record; further abnormal scores update the same incident rather than creating duplicates, and the per-sensor contribution to the overall error is retained as evidence of which signal influenced the score most. An incident does not close as soon as one reading looks better — the score must remain below a separate, lower recovery level for 30 consecutive valid comparisons before the incident is closed, which prevents a brief calm moment from prematurely closing an event that may still be developing. If a score cannot be produced (for example, because of a data gap), the recovery count pauses rather than advancing, since absence of evidence is not treated as proof of recovery.

Slow-degradation path: In parallel, the same underlying evidence is summarized into five-minute medians to reduce the influence of any single isolated peak. A six-hour trend view is then re-evaluated every 30 minutes, and a low-priority warning is opened only when five of the latest six scheduled evaluations support a sustained upward trend. This two-path design exists because a single compromise rule would either react too slowly to a sudden event or too noisily to small, gradual drift; separating the two timescales makes both forms of behavior easier to detect, explain, and govern.

In both paths, the system explicitly reports the strength of the evidence and the sensors that contributed most to it, but it does not assign a physical cause. Classification of an incident (for example, as a suspected blockage or a suspected leak-like event) is recorded as an operator decision, informed by the model evidence together with field context, maintenance history, and other protection systems — the model supplies anomaly evidence; the person supplies the diagnosis. Alerts and lifecycle updates are surfaced on the operator dashboard described in Section 6, with a full audit trail of who acknowledged, classified, or annotated each record.

6.0 SYSTEM DESIGN

6.1 Overall System Architecture

The implemented architecture follows the layered structure proposed in Section 3, with each layer given a specific, auditable responsibility: a sensing layer, a data processing layer, a windowing and feature-preparation layer, a machine-learning layer, and an alert/notification layer that feeds an operator dashboard. Table 3 summarises the responsibility of each layer as implemented.

Layer	Responsibility
Sensing layer	Continuous measurement of pressure, temperature, flow, and vibration at 1 Hz
Data processing layer	Quality checking, gap handling, and preservation of raw and processed streams
Windowing and feature-preparation layer	Construction of the rolling 60-second, four-channel comparison window
Machine-learning layer	Autoencoder inference, reconstruction-error scoring, per-sensor contribution
Alert and notification layer	Fast (sudden-event) and slow (trend) rule engines, incident lifecycle
Operator dashboard	Live context, model evidence, lifecycle state, audit trail, model governance

Table 3: Layered system architecture implemented.

6.2 Sensing Layer

Four signals are measured at the monitored checkpoint once per second: pressure (bar), temperature (°C), flow rate (L/min), and vibration velocity RMS (mm/s). These four signals were selected because, individually, each can look ordinary while the combined pattern is unusual — the system is therefore designed to consider them together rather than as isolated threshold alarms.

6.3 Data Processing Layer

Readings from the sensing layer pass through the quality-checking rules summarised in Table 2. Missing, late, frozen, repeated, or physically impossible values are identified and flagged rather than silently accepted, and both the raw and processed streams are retained so that any cleaning decision remains reviewable.

6.4 Windowing and Feature Preparation Layer

During the first 59 seconds after start-up (or after a break in valid data), the buffer is shown as filling rather than producing a score, which is expected behaviour rather than a fault of the system: it protects against producing a comparison without sufficient context.

6.5 Machine Learning Layer

The machine-learning layer hosts the trained autoencoder and computes, for every valid one-minute window, an overall reconstruction-error (MSE) score together with a per-sensor breakdown of that error. This layer also hosts the candidate-model workflow (train, check, freeze, approve, deploy, rollback), so that model updates are versioned and auditable rather than silent.

6.6 Human Oversight and Governance

Although scoring, comparison, and record-keeping are automated, interpretation and action remain with people. Table 4 sets out the division of responsibility implemented in the system.

System responsibility	**Human responsibility**
Collect and preserve readings	Confirm sensors and operating context are trustworthy
Apply consistent data-quality rules	Investigate conditions the software cannot observe
Compare behaviour with the approved healthy baseline	Decide whether evidence represents a fault, maintenance need, or benign change
Open or update records when explicit rules are met	Acknowledge, classify, comment, and follow site procedures
Present candidate-model evidence	Approve deployment or choose rollback

Table 4: Division of responsibility between the system and its operators.

The prototype does not control valves, pumps, or emergency equipment; it supports observation, evidence-gathering, and investigation only. Pump-running status is displayed as interpretive context but does not automatically suppress an alert, so that unusual behaviour is not hidden simply because one status flag looks normal.

6.7 Alert and Notification

The operator dashboard brings together four categories of information in one place: live context (current sensor values and pump status), model evidence (overall reconstruction error and per-sensor contribution), lifecycle state (whether an incident or slow warning is open, recovering, acknowledged, classified, or closed), and an audit trail (who recorded a decision, what they selected, and what note they added). A guided demonstration mode reproduces six repeatable scenarios — healthy start-up, a sudden incident and its lifecycle, an unusual pattern requiring classification, degraded sensor data, a slow trend, and a full model-update cycle — so that the behaviour of every layer can be verified consistently.

7.0 RESULTS AND DISCUSSION

The results below are drawn from the code-verified prototype demonstration, using deterministic synthetic sensor streams processed through the real preprocessing, scoring, incident, and trend pipelines. This distinction is maintained deliberately throughout the discussion: the results demonstrate that the implemented software behaves as designed, but they do not, on their own, constitute field-validated performance on a live petroleum pipeline. That distinction and the corresponding validation work required before production use are discussed further in Section 8.

7.1 Prototype Detection Performance

Across the guided demonstration scenarios, the reconstruction-error score produced by the autoencoder remained low and stable during healthy operating periods and increased sharply at the point where a sudden fault-like event (for example, a pressure and flow deviation consistent with a developing leak or blockage) was introduced into the synthetic stream. This is consistent with the design intent set out in Section 3.3: the model correctly distinguishes the learned healthy pattern from an unfamiliar four-sensor pattern. Because the demonstration uses synthetic rather than field-collected fault data, the reported behaviour should be read as evidence that the detection logic operates correctly, not as a field-validated accuracy figure (for example, a false-alarm or missed-detection rate) for a real pipeline installation.

7.2 Fault Detection Time

The prototype requires an initial 60-second sensor window before the first anomaly score can be generated. Therefore, during start-up, the first 59 readings are used to build the required temporal context, and scoring begins when the 60th valid reading becomes available. Once continuous monitoring is established, a new anomaly score is generated every second using the latest 60-second window.

In the configured synthetic sudden-event evaluation, the introduced abnormal patterns crossed the incident-opening criterion with 0 s additional detection delay once a valid anomaly score was available. This indicates that the implemented sudden-event logic can respond immediately to a threshold-crossing anomaly in the synthetic demonstration. However, this result represents the response of the prototype under controlled synthetic conditions and should not be interpreted as evidence that real pipeline leaks or other physical faults can be detected with zero delay. Real-world detection time must be established through field validation using actual pipeline data.

7.3 REDUCTION IN DOWNTIME

A quantitative reduction in downtime could not be established from the prototype demonstration alone, since this requires operational data from a live installation over a representative period, including comparison against the pipeline's prior reactive maintenance history. Qualitatively, however, the sudden-event and slow-trend pathways are each designed to shorten the interval between the onset of an abnormal condition and the moment an operator becomes aware of it: the sudden path is designed to surface a developing event within seconds rather than after a conventional threshold alarm is crossed, and the slow path is designed to surface gradual deterioration during routine review rather than only once it has progressed far enough to trigger a hard alarm. Establishing an actual downtime-reduction figure is identified in Section 8.2 as a priority item for site validation.

7.4 Sensor Behaviour During a Simulated Leak-Like Condition

The sensor responses observed during the sudden-event demonstration scenarios are consistent with the flow-behavior analysis in Section 4: pressure showed a significant, sustained drop following the onset of the simulated leak-like event (Section 4.3); flow rate decreased correspondingly as fluid left the intended flow path (Section 4.4); temperature showed only a slight change, consistent with it being a secondary, contextual indicator rather than a primary fault signal (Section 4.5); and vibration increased noticeably around the affected region, consistent with the mechanical disturbance associated with leakage (Section 4.6). The per-sensor reconstruction-error contributions reflected which measured signals contributed most strongly to the overall anomaly score. These contributions support interpretation but do not establish the physical root cause of the abnormal condition.

7.5 Human-in-the-Loop Fault Classification

The implemented system deliberately separates automated anomaly detection from fault classification. In real time, the model produces an anomaly score and a per-sensor evidence breakdown; it does not independently assign a specific fault category such as a leak or blockage.

During the demonstration, an operator reviewing the incident record was able to acknowledge the event, select a suspected classification, such as **“suspected blockage”** or **“suspected leak-like**

condition,” and add a supporting note. These actions were retained within the auditable incident lifecycle described in Section 6.7.

This human-in-the-loop classification approach reflects the project’s design principle that the machine-learning model should provide **traceable evidence of abnormal behaviour rather than automatically diagnosing the physical root cause**. The final interpretation and operational decision therefore remain the responsibility of a qualified operator or engineer.

8.0 CONCLUSION AND FUTURE SCOPE

8.1 Conclusion

This project set out to move pipeline fault detection from a predominantly reactive activity toward a proactive, early-warning approach, using continuous multi-parameter sensing (pressure, flow, temperature, and vibration) combined with machine-learning-based pattern analysis. The literature review confirmed that conventional single-parameter methods are generally better suited to detecting an already-established fault than to recognizing its earliest stages, and that combining multiple sensor signals with a data-driven model can provide a more complete representation of pipeline condition (Sections 2 and 4).

The proposed architecture includes physical sensors and a microcontroller-based data acquisition layer. The current working prototype implements and code-verifies downstream data-processing, machine-learning, anomaly detection, incident-management, and dashboard components using deterministic synthetic sensor streams. The prototype implements an unsupervised autoencoder trained only on approved healthy operation, using a rolling one-minute, four-sensor window and a reconstruction-error anomaly score, together with two complementary operating paths: a fast path for sudden events and a slower, persistence-based path for gradual degradation. Data-quality handling, an auditable incident lifecycle, a 30-reading recovery rule, and a controlled model-approval workflow were implemented specifically to keep the system's evidence trustworthy and its behavior explainable to a non-technical audience.

The synthetic evaluation demonstrated the operation of both monitoring pathways: the sudden-event pathway responded to configured abrupt deviations, while the slow-degradation pathway successfully executed the five-minute aggregation, six-hour trend evaluation, and persistence logic. However, not all injected long-horizon scenarios satisfied the configured 5-of-6 warning criterion, highlighting the need for further threshold calibration and field validation. The project therefore demonstrates a complete, functioning early-warning architecture and validates the software behavior described in the design chapters, while stopping short of claiming field-validated performance on an operating petroleum pipeline.

Overall, the project meets the specific objectives set out in Section 1.1: relevant parameters are continuously monitored, changes associated with abnormal conditions were investigated and characterized, a data acquisition and processing pipeline was developed, normal-operation patterns and their deviations were established, a machine-learning technique was investigated and implemented, an early-warning mechanism was built and demonstrated, and the resulting evidence is presented in a form intended to support preventive maintenance decisions.

8.2 Future Scope

A number of extensions and validation steps are identified for future work, in line with the honest boundaries of the current prototype:

- Field validation: Validate sensors, detection thresholds, and recovery behaviour against real historical operating conditions at a physical pipeline checkpoint, with input from site engineers.
- Operating envelopes: Define approved operating envelopes for start-up, shutdown, maintenance, and different flow regimes, so that expected process changes are not mistaken for developing faults.
- Quantitative performance evaluation: Measure false-alarm rate, missed-detection rate, detection time, and data-quality coverage over a representative operating period, and use these to establish an evidence-based estimate of downtime reduction.
- Expanded sensing: Investigate additional sensing modalities such as acoustic emission and negative-pressure-wave sensing, identified in the literature review as complementary to the four parameters currently monitored.
- Model development: Explore supervised or hybrid models as genuine fault-labelled datasets become available and compare their performance against the current unsupervised autoencoder baseline.
- Security and reliability: Complete security, access-control, audit-retention, reliability, and failover reviews appropriate to an industrial monitoring deployment.
- Controlled integration: Integrate the system with wider plant protection and control systems only through a formally approved operational and safety process; the current prototype does not, and is not intended to, control valves, pumps, or safety equipment.
- Regulatory compliance: Complete the regulatory and compliance review appropriate to the pipeline application and jurisdiction before any production deployment.

In summary, the project has produced a working, explainable, and governed early-warning prototype that fulfils its design objectives on demonstration data. The clearly scoped future work above is intended to carry the system from a code-verified prototype to a field-validated, production-ready pipeline monitoring solution.

9.0 REFERENCES

[1] J. Sekhavati, S. H. Hashemabadi, and M. Soroush, "Computational methods for pipeline leakage detection and localization: A review and comparative study," *J. Loss Prev. Process Ind.*, vol. 77, p. 104771, Jul. 2022, doi: 10.1016/J.JLP.2022.104771.

[2] L. Zhao, Z. Cao, and J. Deng, "A review of leak detection methods based on pressure waves in gas pipelines," *Measurement*, vol. 236, p. 115062, Aug. 2024, doi: 10.1016/J.MEASUREMENT.2024.115062.

[3] B. Liu, Z. Jiang, and W. Nie, "Application of VMD in Pipeline Leak Detection Based on Negative Pressure Wave," *J. Sens.*, vol. 2021, no. 1, p. 8699362, Jan. 2021, doi: 10.1155/2021/8699362.

[4] M. I. Mohd Ismail *et al.*, "A review of vibration detection methods using accelerometer sensors for water pipeline leakage," *IEEE Access*, vol. 7, pp. 51965–51981, 2019, doi: 10.1109/ACCESS.2019.2896302.

[5] J. Chen *et al.*, "Application of machine learning to leakage detection of fluid pipelines in recent years: A review and prospect," *Measurement*, vol. 248, p. 116857, May 2025, doi: 10.1016/J.MEASUREMENT.2025.116857.

[6] L. Chen *et al.*, "Research on an Oil Pipeline Anomaly Identification Method for Distinguishing True and False Anomalies," *Mobile Information Systems*, vol. 2022, no. 1, p. 9366897, Jan. 2022, doi: 10.1155/2022/9366897.

[7] X. Wan, P. K. Kuhanestani, R. Farmani, and E. Keedwell, "Literature Review of Data Analytics for Leak Detection in Water Distribution Networks: A Focus on Pressure and Flow Smart Sensors," *J. Water Resour. Plan. Manag.*, vol. 148, no. 10, p. 03122002, Jul. 2022, doi: 10.1061/(ASCE)WR.1943-5452.0001597.