# Automated Landslide Impact Detection Through Machine Learning Analysis of Pipeline IMU Bending-Strain Data

**Aron Zahradka[1], Sarah Newton[1], Owen Bunce[2], Joel Van Hove[2], Jared Kowis[3]**

[1]Cambio Earth, [2]BGC Engineering, [3]Enbridge

PPIM 2026
PIPELINE PIGGING & INTEGRITY MANAGEMENT
JANUARY 19-22

Organized by

CLARION
TECHNICAL CONFERENCES

## Abstract

Inertial Measurement Unit (IMU) bending-strain data play a critical role in detecting and assessing landslide impacts on pipelines. In line inspection (ILI) vendors often flag numerous pipeline bending-strain features based on IMU data; however, most identified features (>90–95%) are not associated with geohazards such as landslides. Manually reviewing these strain indications is labour-intensive and relies heavily on the experience and expertise of the analyst. This study describes a machine learning (ML) method developed to automatically screen and prioritize IMU bending-strain anomalies linked to landslide-related deformation.

A Convolutional Neural Network (CNN) deep-learning classifier model was trained using IMU data from North American transmission pipelines, primarily in the Appalachian region of the USA. The dataset included over 30,000 strain feature training examples segmented into 200-meter windows, with 478 labelled as landslide impacted by pipeline geohazard experts. Model performance was evaluated using a hold-out test dataset consisting of 25% of the available dataset, including 112 landslide impacted strains. The model achieved 90% recall, 95% specificity, 95% accuracy, and an ROC-AUC of 0.96. In other words, it filtered out approximately 95% of benign features while capturing over 90% of landslide-related deformations.

This research demonstrates the potential of ML models applied to IMU data which already has the potential to reduce analysis time per ILI run and improve consistency in identifying high-risk features. By rapidly identifying landslide related pipeline deformation, it helps operators allocate resources more effectively, respond faster to emerging threats, and maximize the value gained from their IMU data. The future integration of additional contextual data such as lidar and landslide mapping, along with increases in the training dataset size and diversity, are expected to yield substantial future improvements to model performance.

## Introduction

Inertial measurement unit (IMU) data from in line inspections (ILI) have become a critical tool for managing landslide threats to pipeline integrity and have been discussed in many previous publications (Theriault et al., 2019, Hart et al., 2019, Dowling et al., 2024, Van Hove et al., 2024). Integrating IMU-based bending strain assessments into geohazard management programs (GMPs) has had a measurable effect in reducing failure rates from slow-moving landslides. Over the past decade, pipelines managed under programs that incorporate IMU have achieved failure rates up to ten times lower than historic averages, with more than half of newly identified critical landslide sites discovered through IMU data review (Van Hove et al., 2024).

While the industry widely recognizes the value of IMU data, its use remains limited by the high volume and complexity of the datasets and the specialized expertise required for their interpretation.

A subject matter expert (SME) must integrate knowledge of landslide mechanics, soil–pipe interaction, and IMU signal processing to evaluate potential landslide impacts. Reviewing the thousands of bending strain features generated in a single inspection run is impractical and cost-prohibitive; thus, operators rely on screening thresholds to reduce the review burden. Although this approach improves efficiency, it is imperfect. Benign sites are often flagged unnecessarily, while subtle true positives can be missed. Alternatively, screening processes have been developed to focus the effort of manual assessment on short sections of the pipeline where landslide impacts are most likely to be discovered. These screening processes have enabled the industry to leverage IMU data for landslide risk management, but there is wasted effort when sites are flagged unnecessarily and missed threats from sites that slip through the screening criteria.

This study presents a proof-of-concept machine learning (ML) classifier that automates the screening of IMU bending strain features for potential landslide impacts. The data used to train the model consisted of a subset of bending strain features classified by subject matter expert (SME) analysts stored in the Cambio database. Cambio is a cloud-based platform used to manage geohazard integrity programs for >35 pipeline operators. The goal is to improve landslide detection rates (i.e., increased recall) while reducing the volume of features requiring manual review (i.e., increased precision), thereby freeing engineers to focus on high-probability sites.

It is important to note that this initial study focuses primarily on Appalachian transmission pipelines, where specific geological conditions (predominantly shallow rotational slides) and construction practices create distinctive IMU signatures. The generalizability to other physiographic regions with different landslide mechanisms (e.g., deep-seated failures in glaciated terrain) remains to be validated. We present this work as a regional proof-of-concept that demonstrates the feasibility of ML-based screening while acknowledging the need for geographic expansion.

## Background

### Geohazards and Integrity Management

Geohazards, geotechnical and hydrotechnical processes that impose loads on buried pipelines, are complex to manage within modern integrity programs. Many types of geohazard require management within mature integrity programs, including slow and rapidly moving landslides, surface and groundwater erosion issues, watercourse crossings, encroachments, and permanent ground displacement triggered by seismic events. The spatial distribution of hazard sites is typically extensive, requiring management of thousands of locations across large pipeline networks (Newton et al., 2019).

For the past decade or more, the pipeline industry has progressively adopted risk-informed approaches for geohazard management. A persistent insight from these programs is that a small fraction of sites (on the order of 2%) accounts for most expected failures, underscoring the need for reliable triage methods to identify the highest risk locations (Newton et al., 2022). Many geotechnical hazards evolve slowly, and visible indications are often absent or very subtle until significant ground movement has occurred (Van Hove et al., 2024). IMU aids with early identification of landslide movement and adds the quantitative evidence needed to identify which lengths of pipeline have been impacted by landslide activity, and to what extent (Van Hove et al., 2024).

## IMU Data in Geohazard Programs

IMU ILI tools record continuous data used to calculate pipeline curvature and is combined with pipeline diameter to calculate vertical, horizontal and total bending strain (Czyz & Adams, 1994, Hart et al., 2019). If a bending strain assessment is requested by the pipeline operator, the ILI tool vendor will calculate bending strains continuously along the entire tool run. ILI tool vendors and other specialists screen through the data to identify bending strain features (bends that have characteristics which are atypical of construction and may indicate threats to the pipeline) (Dowling et al., 2024). Different data processing techniques and reporting criteria are in use, which can lead to significantly different numbers of bending strain features with the same source data (Dotson et al., 2024). For landslide assessment, different rules of thumb have been published (Dewar 2020), but 0.125% total bending strain is the most common reporting threshold (Dotson et al., 2024).

Despite the benefit of bending strain features highlighting potential problems, the number of bending strain features is often high while the percentage caused by landslide interaction tends to be low, around 5% (Mackenzie-Johnson et al., 2024). Previous studies report frequencies on the order of two bending strain features per mile (Dotson et al., 2024) with one large-scale study of 12,000 miles of pipeline producing 38,400 vendor bending strain features (three per mile) (Scheevel et al., 2022). Manual evaluation of every bending strain feature by a subject matter expert (SME) is labour-intensive, which prompted the development of additional screening methods that broadly rely on correlations between bending strain orientation (horizontal versus vertical) and magnitude, and landslide interactions (Dewar, 2020, Theriault et al., 2019, Mackenzie-Johnston et al., 2024). These screening methods significantly reduce the overall number of sites to assess, while identifying a significant portion of the landslide interactions (Theriault et al., 2019, Mackenzie-Johnston et al., 2024).

Screening criteria for single-run IMU are variable, but >0.1% horizontal bending strain and >0.2% total strain are common. Considering a criterion of >0.1% horizontal bending strain, the data presented in a 2019 study by Theriault et al., would reduce the total number of bending strain features by around 90%, and would still identify 50% of the landslide interactions. Applying strain magnitude-based screening methods has allowed operators to efficiently prioritize bending strain features for landslide identification, but there is still room for improvement. In the above example only 20% of the bending strain features reviewed would be due to landslides, and only 50% of the bending strain features caused by landslides would be identified. The screening method could be improved by both reducing the number of false positives (bending strain features reviewed unnecessarily) and decreasing the number of false negatives (landslides not identified). Improved methods are likely to become even more important as operators across the industry catch up with the early adopters who have mature geohazard management programs. Every year tens of thousands of miles of pipeline are inspected with IMU in the US alone, and operators need more powerful ways to organize, manage and derive value from these data.

The goal of this work was to develop and test a ML classifier to detect landslide-related bending strain features from single run IMU data, thereby reducing manual effort while increasing feature identification accuracy. Improved single-run screening will be most valuable for initial identification of pipeline segments impacted by ground movement since installation. The classifier also adds context to multi-run comparisons, helping distinguish ongoing landslide loading from construction-related or operational changes.

## Data and Labelling

### Cambio IMU Strain Feature Training Dataset

Robust data management is imperative to manage geohazard program data, which includes IMU data, lidar based change detection, InSAR, field inspections and geotechnical instrumentation data. Cambio centralizes these data in a unified schema. The Cambio platform has been used to manage pipeline geohazard integrity threats on over 370,000 miles of pipelines, representing the combined networks of more than 35 pipeline operators.

Cambio's IMU data model normalizes the bending strain data provided by any IMU vendor so they can be visualized and analyzed in the context of the geohazard program data. Within Cambio, IMU datasets are stored in a relational database and as spatial features linked to the pipeline centerlines. Vendor-specific formatting differences (such as coordinate systems, naming conventions, and units) are standardized into a consistent schema. Bending-strain results are stored alongside metadata describing gauge length, pipeline diameter, wall thickness, steel grade, and run year. Cambio's IMU strain feature management workflows (Newton & Scheevel, 2025) integrate these data with terrain models, lidar change detection, and landslide inventories, enabling contextual interpretation and providing a structured foundation for model training.

This architecture allows automation and reproducibility, critical for applying ML to IMU data. The ML classifier described in this study was trained directly on a subset of the Cambio-managed IMU datasets, ensuring consistent preprocessing and metadata tracking. The dataset used for model development consisted of IMU runs from transmission gas pipelines, with known landslide activity verified through historical field programs. Most training and testing data originated from the Appalachian region (Ohio, Kentucky, Pennsylvania), where landslides are common. Additional test sites were selected in Mississippi, Texas, Pennsylvania, and British Columbia (Canada) to assess model performance terrain with different deformation mechanisms.

The combined dataset consisted of 2,409 IMU strain features labelled as landslide-related or unrelated. Labels were derived from BGC Engineering SME inspections where IMU bending strain features were assessed against lidar imagery, historical ground movement mapping and geotechnical instrumentation data. An additional 28,168 strain features that were not assessed in detail by a SME were included in the dataset with an assumed label of unrelated to ground movement. Typically, these features would not have been flagged for review by a conservative screening level criterion, a prior review of bending strains at the same location, or subject to a rapid visual screening by a SME. Confidence in this labelling approach is supported by the comprehensive geohazard management programs from which the data were sourced, which included thorough corridor-wide lidar mapping, regular lidar change detection (LCD) monitoring to identify active ground movement, and extensive field validation campaigns at identified geohazard sites. These multi-layered assessment protocols provide reasonable assurance that landslide-impacted pipeline segments would have been identified through at least one pathway (geomorphic mapping, active movement detection, or strain screening), making the probability of unlabelled landslide impacts in the remaining dataset low.

## Labelling Process

Each strain feature was labelled as landslide, non-landslide, or unknown based on SME interpretation following the Tier 1–5 framework described by Scheevel et al. 2022. Landslide cases typically exhibited broad, asymmetric horizontal-strain or vertical lobes consistent with soil movement. Non-landslide cases corresponded to construction bends, integrity digs, formed bends, or thermal alignment effects (Dowling et al., 2024). Strain features were labelled as unknown when SME interpretation could not confidently distinguish between landslide and non-landslide causes of bending strain based on available IMU data and contextual information. Given the conservative nature of geotechnical engineering practice, most unknown features are expected to be non-landslide related. However, to ensure high-quality training labels, strain features labelled as unknown were excluded from this study.

Labelling of the strain features as either landslide-related or landslide-unrelated was conducted through a structured workflow combining remote-sensing interpretation, IMU plotting, and field verification:

1. BGC geohazard inspections: Landslide inventories compiled through lidar change detection, lidar terrain mapping, and field inspections were reviewed to identify sites where the pipelines intersect known landslides.
2. IMU feature plotting: Horizontal and vertical bending-strain profiles were plotted and visually assessed to determine correspondence with mapped landslides.

3. Expert interpretation: Each feature was classified as landslide or non-landslide by at least two experienced analysts. Conflicts were resolved through joint review sessions to ensure consistency with previous screening campaigns (Scheevel et al., 2022).
4. Ground truthing: Nearly all the landslide-classified sites had subsequent field visit inspection records confirming movement after the IMU run date.

### Landslide and Pipeline Characteristics

Landslide attributes recorded in Cambio accompany each positive example. These include classification (according to Crudens & Varnes, 1996), basic slope geometry such as average angle, slope height, and length, movement direction as azimuth relative to the pipeline alignment, the length of pipeline impacted by the landslide, and an interpreted velocity class derived from instruments or remote sensing (categorized in seven classes described in Cruden & Varnes, 1996, from very slow (<0.63 in/yr) to extremely rapid (>16 ft/s)).

Landslides and the impacted pipelines in the modelling dataset were broadly characterized as follows:

- Nearly all classified (Crudens & Varnes, 1996) as earth slides, with only a handful classified as earth flow or earth creep
- Width predominantly <1000' and nearly all <2500', though with some outliers in B.C. up to 2 miles wide
- Length predominantly <600' and nearly all <2000', though with some outliers in B.C. up to 1.2 miles long
- Velocity (Crudens & Varnes, 1996) predominantly extremely slow (<0.63 in/yr) and all landslides characterized as slow (5.3 ft/yr) or slower
- Impacted length of pipeline predominantly <1300' and nearly all <2500', though with some outliers in B.C. up to 2 miles wide
- Landslide movement direction overwhelmingly perpendicular to the pipeline (transversely loaded) in Appalachia and generally parallel to the pipeline (axially loaded) in B.C.

Pipeline attributes were available for context and subsequent interpretation, including outside diameter, wall thickness, steel grade, transported product, and installation year. The pipelines in the modelling dataset were steel walled gas transmission lines broadly characterized as follows:

- Installation year ranged from 1940 to 2024, with the majority installed between 1950 and 1990
- Outer diameter predominantly 30" NPS and 36" NPS
- Wall thickness predominantly 0.39"

Although these parameters were not explicitly used as model inputs in this initial study, they provide valuable context for interpreting strain magnitudes and may be incorporated as auxiliary features in future model iterations.

# Methods

## IMU Data preparation

This work used the raw (heading and pitch) IMU data provided by ILI vendors to preserve spatial resolution and ensure consistent processing across vendors. Vendor files were first normalized so that column names, coordinate reference, and units align with a single standard for consistency. The bending strain was calculated from the derivative of heading and pitch using a linear regression method as outlined in Czyz & Adams, 1994. A nominal gauge length of three meters was used, and the channels were all resampled to uniform spacing of 0.1 m for consistency in the dataset. The result is five continuous channels of IMU data: odometer, pitch, heading, and the calculated vertical and horizontal bending strain. Processing was executed in a secure cloud environment so full ILI runs could be efficiently processed, stored and visualized.

## Imbalance and Data Distribution

Landslide-related IMU strain features are inherently rare compared to the vast number of benign curvature changes recorded in ILI runs. The training dataset reflects this imbalance, with only 1.6% of strain features labelled as showing landslide signatures. This was consistent with previously published industry datasets. For example, Scheevel et al., 2022 reported 0.3% Tier 1–2 features while Dewar, 2020 reported roughly 2-3 features per mile with <5% geohazard origin. The model was therefore trained using class weighting to prevent bias toward the majority (non-landslide) class.

## Test Set Definition

Rigorous evaluation of ML models requires careful definition of the testing dataset to ensure that performance metrics accurately reflect real-world deployment scenarios. The primary concern in this study was preventing information leakage between training and testing datasets — a situation where the model inadvertently learns patterns specific to the test data, leading to overly optimistic performance estimates that do not generalize to new pipeline inspections.

Two potential sources of information leakage were identified in the IMU dataset. Multiple ILI runs of the same pipeline segment at different times or parallel pipelines sharing the same right of way (RoW) and crossing the same landslide could allow the model to learn site-specific characteristics rather than generalizable landslide signatures. If features from the same landslide appeared in both training and test sets (even on different pipelines), the model might perform well on the test set simply because it had already seen that specific landslide's characteristics during training.

To ensure spatial independence, seven geographically distinct polygons were drawn across the study area to delineate the test set. These polygons were positioned to capture regional heterogeneity in both pipeline characteristics (diameter, installation year) and landslide characteristics (geology, movement rate, classification). All strain features geographically within a test polygon were assigned exclusively to the test set and not shown to the model during the training process. This ensures that the "true" model performance in these test areas is evaluated. It is important to note that the performance of the model on IMU from pipelines or in landslide regimes that differ substantially from the test set is cannot be known. The test set was designed to comprise 25% of the total dataset. Class balance (ratio of landslide to non-landslide features) was maintained, ensuring that the proportion of landslide-related features in the test set matched the overall dataset distribution.

### Comparison to Current Practice

To contextualize model performance, results were compared against a common industry screening criterion. As reported by Dewar (2020) and consistent with common current practice, many operators screen IMU data using magnitude-based thresholds: horizontal strain ≥0.125% or total strain ≥0.2%. These thresholds were applied to the ML model test set to establish a baseline screening performance, providing a reference point for evaluating whether the machine learning approach offers meaningful improvement over existing methods in terms of both detection accuracy and review workload reduction.

### Machine Learning Modelling

Landslide interactions produce characteristic spatial patterns in strain profiles – most commonly broad, asymmetric horizontal-strain lobes reflecting the lateral movement of soil mass. These patterns differ from the more symmetric, localized bends associated with construction, thermal effects, or integrity excavations. The goal of the machine learning approach was to automate the detection of these landslide signatures by learning to recognize their distinctive morphology directly from the IMU data.

Convolutional Neural Networks (CNNs) have become a standard approach for pattern recognition tasks because of their ability to automatically learn relevant features from raw data without manual feature engineering. While originally developed for image analysis, CNNs are well-suited to time-series style data classification because they can identify localized patterns (such as strain peaks or lobe shapes) regardless of where they appear in the sequence. This spatial invariance is critical for IMU data, where similar landslide signatures might occur at different positions within an inspection segment.

A CNN operates by applying learned filters to the input data, progressively extracting features at multiple scales. Early layers detect simple patterns (such as slopes or peaks in the strain curve), while deeper layers combine these into more complex features (such as the full signature of a landslide-induced bends). This hierarchical learning process mimics how an experienced analyst might evaluate a strain profile: first noting individual peaks or transitions, then recognizing the overall pattern as consistent with ground movement.

For this study, the IMU data were used as input, including horizontal strain, vertical strain, pitch angle, and heading angle. These were fed to the neural network in 200-meter windows, which provide sufficient context to capture the full spatial extent of most (though not all) landslide interactions in the dataset while including some baseline pipeline curvature on either side for contrast and in cases where the identified bending strain feature window didn't capture the entire extent of the affected pipeline. The existing vendor-provided strain measurements were also integrated into the model ensemble, including the maximum horizontal, vertical, and total bending strain magnitudes for the strain feature and at girth welds.

## Results

### Machine Learning Model Prediction Performance

The ML model was evaluated using the spatially independent test dataset containing 7,136 strain features, including 112 SME confirmed landslide impacts. The model outputs a probability score between 0 and 1, representing the likelihood that a given strain feature results from landslide impact.

Figure 1 illustrates the model predicted probabilities stratified by "true label" (i.e. the SME interpretation of the strain feature). The clear separation between landslide impacted strains and non-landslide related features demonstrates the model's discriminative capability. Most non-landslide features receive probabilities near zero, while landslide impacted features cluster toward higher probability values. The vertical scatter within each category is a random spread introduced to better visualize areas of the plot with high point densities.

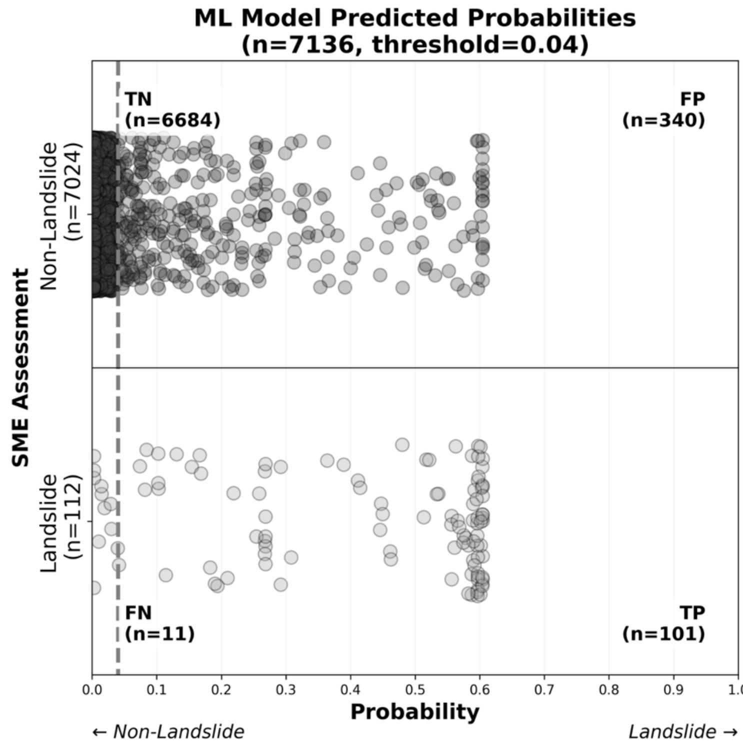


**Figure 1.** ML model predicted probabilities for test set strain features (n=7,136), stratified by SME Assessment label. Clear separation between landslide-impacted features (clustering toward high probabilities) and non-landslide features (clustering near zero) demonstrates the model's discriminative capability. The vertical dashed line represents the probability threshold selected for final classification of each strain feature; the classified counts of features based on this threshold are shown in each quadrant (TN and FN quadrants are left of the dashed threshold line).

## Binary Classification Performance and Comparison to Industry Screening Criteria

Operators can use the predicted probabilities to prioritize the order of inspection of strain features. Alternatively, they can convert these continuous probability scores into actionable classifications by selecting a threshold probability value. Features with probabilities above the threshold are classified as landslide impacts requiring investigation, while those below are considered non-landslide related features that do not require further assessment. This classification produces four possible outcomes:

- True Positives (TP): Landslide impacts correctly identified by the model — these represent successfully detected geohazards
- False Negatives (FN): Landslide impacts the model missed — unidentified risks requiring alternative detection methods
- True Negatives (TN): Non-landslide related features correctly excluded — engineering resources saved through automated screening
- False Positives (FP): Non-landslide related features incorrectly flagged — inspection effort that could be redirected to higher-priority efforts to reduce risk

To benchmark the ML model's performance, results are compared against the industry-standard strain screening criteria previously introduced, which flags features exceeding specific total or horizontal strain magnitude thresholds. For direct comparison, a probability threshold of 0.04 (4%) for the ML model was selected to match the recall rate (90.2%) achieved by the industry screening criteria on the same test data as used for the ML model. This ensures both methods identify the same proportion of actual landslides, allowing meaningful comparison of their false positive rates.

Figure 2 presents confusion matrices for both approaches at matched recall of 90.2%. Both correctly identify 101 of 112 landslide impacted strains (though not the *same* 101 landslide impacted strains). However, their efficiency differs dramatically. The industry screening criteria generates 2,059 FP, achieving only 4.7% precision and requiring engineers to review approximately 21 non-landslide features for each actual landslide discovered. The ML model reduces FP by 83% (to 340 FP), achieving 22.9% precision and reducing the review burden to approximately four non-landslide features per landslide – a nearly five-fold improvement in screening efficiency.

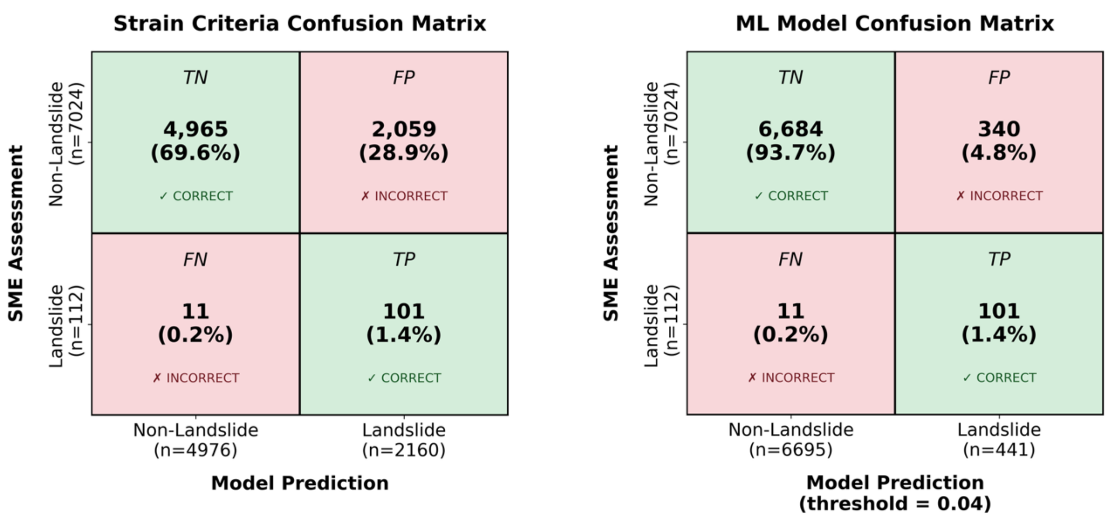


**Figure 2.** Confusion matrices comparing industry strain screening criteria (left) versus ML model (right) at matched 90.2% recall rate. The ML model reduces false positives by over 80% (from 2,059 to 340), cutting the review workload from ~21 to just ~4 non-landslide related features per landslide impacted strain, while maintaining equivalent landslide detection performance.

Binary classification also allows for the calculation and comparison of a range of standard performance metrics, summarized for both the industry screening strain criteria and ML model in Table 1. Beyond the matched recall rate, the ML model demonstrates superior performance across all other metrics, including:

- Recall: The proportion of actual landslides detected – critical for identifying the risk
- Precision: Indicates the proportion of flagged features that are actual landslides – important for maintaining efficiency
- F1 Score: provides a balanced measure by combining precision and recall
- Specificity: measures the model's ability to correctly exclude non-landslide related features

- Accuracy: measures overall correct classifications across both landslide and non-landslide features but can be a misleading metric when the dataset is heavily class imbalanced (many more non-landslide than landslide impacted strain features)

**Table 1.** Performance metrics for industry strain screening criteria and ML model at matched 90.2% recall threshold, demonstrating superior ML model performance across precision, F1 score, specificity, and accuracy.

| | **Strain Criteria** | **ML Model** |
|---|---|---|
| **Recall** | 90.2% | 90.2% |
| **Precision** | 4.7% | 22.9% |
| **F1** | 8.9% | 36.5% |
| **Specificity** | 70.7% | 95.1% |
| **Accuracy** | 71.0% | 95.1% |

### Receiver Operating Characteristic

A Receiver Operating Characteristic (ROC) curve plots the true positive rate (recall) against the false positive rate (1 – specificity) across all possible classification thresholds, with the area under the curve (AUC) providing a threshold-independent measure of discriminative capability. A perfect classifier achieves AUC = 1.0, while random guessing yields AUC = 0.5.

The ROC curve of the ML model is presented in Figure 3 and demonstrates strong performance with AUC = 0.963, indicating excellent discrimination between landslide impacts and benign features across the full range of operating thresholds. The industry screening strain criteria, being a fixed threshold, can only be plotted as a single point rather than a continuous curve. At the selected threshold (4% probability), the ML model achieves the same 90% recall as the strain criteria while reducing the false positive rate from 29.3% to just 4.8% - a six-fold improvement that translates directly to reduced engineering review time.

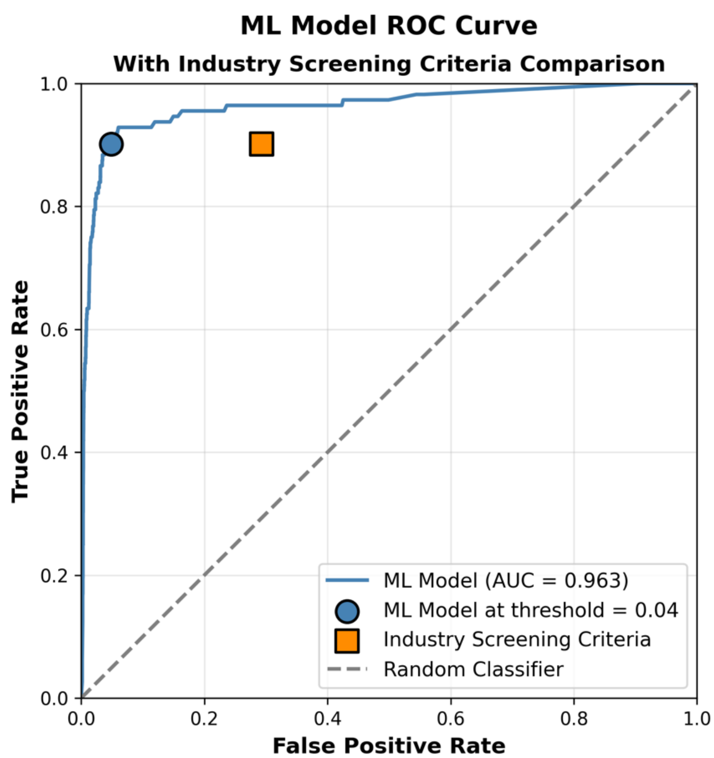


**Figure 3.** ROC curve for the ML model (AUC = 0.963) with industry screening criteria shown for comparison. At the selected threshold (4% probability), the ML model achieves 90% recall with only 4.8% false positive rate versus 29.3% for traditional magnitude-based screening – a six-fold improvement in false positive rate.

## Regional Results

Based on the location of the seven test polygons, the test dataset was grouped into four distinct geographic regions to attempt to capture the geographic variability in model performance due to differences in pipeline and landslide characteristics. Confusion matrices of the regional test set results are presented in Figure 4. The model demonstrates strong and consistent specificity across all regions (95–98%). Small sample sizes of SME Assessed landslide impacted strains in regions besides Appalachia (BC: n=5, Gulf Coast: n=2, Pennsylvania: n=0) limit any statistical confidence of regional recall estimates.

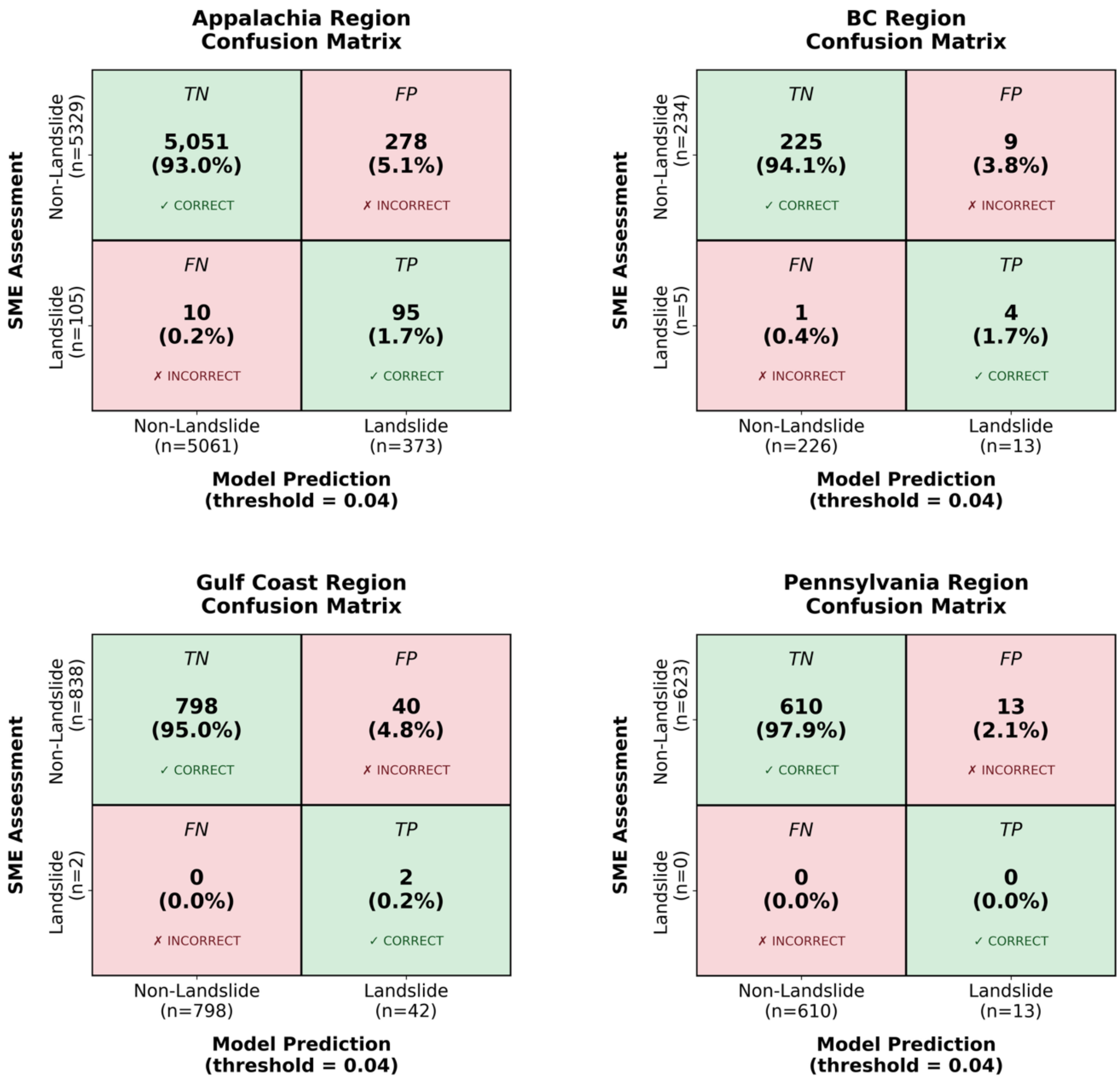


**Figure 4.** Regional performance breakdown of ML model at 4% probability threshold. The model demonstrates strong and consistent specificity across all regions (95–98%). Small sample sizes of SME Assessed landslide impacted strains in regions besides Appalachia limit the statistical confidence of regional recall estimates.

## Discussion

### Misclassified Strain Features

A selection of the strain features were reviewed to better understand why some strains were incorrectly labelled by the model. The model predicted a low probability (below the 4% probability threshold) for 11 strains that were landslide related (FN). These strains were located at seven unique locations: three locations in northern Appalachia, three locations in southern Appalachia and one in Northeastern BC. Two of the FN strains in the test set were below the typical vendor reporting limit (<0.125% total strain) and had been identified through run-to-run analysis.

The 11 FN strain features had the following characteristics:

- There were five strains that were unclear as to whether they were actual FN, as the SME had struggled to make a decisive conclusion. These strains had originally been flagged as ground movement related but subsequently labelled as unknown or non-landslide related as more contextual data was collected about the site (e.g. subsequent ILI runs, LCD, and ground inspections).
- There were three strains where the strain signatures were subtle, low magnitude horizontal strains and the SME relied largely on additional, complementary data for their interpretation. Critical complementary information included lidar, change between multiple runs of IMU (in this case spanning over 20 years), and more obvious landslide-impacted strain signatures in neighbouring pipelines.
- There was one strain where the FN conclusion was a result of a data entry error. The SME described uncertain expression of landslide impact in the IMU data but regardless labelled the strain landslide related instead of unknown. This demonstrates the potential usefulness of the model for providing QA/QC of SME interpretations.
- There was one strain where a mitigation had been completed following identification of the landslide impacted strain in the IMU data. The model predicted a low probability on a strain feature from an ILI completed following the mitigation and stress relief of the pipeline. More investigation is needed into why the model prediction was low in this instance as the model successfully predicted a TP for three other strain features at the same location from other post-mitigation ILI runs.
- There was one strain feature where it is unclear why the model predicted a low probability since the model successfully predicted a TP on strain features at the same location from different ILI runs.

The last two bullet points illustrate the limited interpretability of deep learning model results, which is a major drawback of the approach. More work is required to investigate the causes of these inconsistencies and to mitigate the likelihood of occurrence.

Of the 340 FP strain features predicted by the ML model (i.e. cases where the model incorrectly predicted a landslide impacted strain feature), there were 316 FP strain features that were not directly reviewed by a SME. These were assigned the assumed label of "non-landslide related", as they were not flagged for review by some screening criterion. A review of these features found that for 27 of these unlabelled strains, using the assumption of non-landslide related was incorrect. For these, a strain feature from a previous ILI run at that location had concluded that the pipeline *was* impacted by a landslide, so future strains at the location were not reviewed.

Generally, it was found that for false positives with high strain values, contextual information at the site location aided the SME in the interpretation that ground movement was not responsible for the strain signature. The strain features with typically high horizontal strains in this category were primarily found to be on flat ground or associated with river or road crossings where there are typically more construction induced bending strains, consistent with the previous findings of others (Theriault et al., 2019; Dowling et al., 2024).

These findings highlight two priorities for improvement. First, training data quality must be enhanced as inconsistent or erroneous labels degrade model performance and bias evaluation metrics. Second, incorporating more contextual data would strengthen model accuracy. SMEs leverage multiple information sources including terrain characteristics, adjacent pipeline behaviour, run-to-run IMU strain comparisons, and supplementary ILI data to distinguish landslide signatures from other strain sources. Integrating these data streams would provide much richer context for the model, particularly in ambiguous cases where IMU data alone is insufficient

## Regional Performance

The geographic distribution of landslide examples in the test dataset reveals an important limitation of the current model. Of the 112 landslide-impacted strains in the test set, 105 (94%) were located in Appalachia, with only five in British Columbia and two in the Gulf Coast region. Pennsylvania test polygons contained no landslide-impacted strains. This concentration means the model's reported performance metrics (90% recall, 95% specificity) primarily reflect its capability in detecting Appalachian landslide signatures.

While the model maintained high specificity across all regions (95-98%), demonstrating consistent ability to exclude non-landslide features, the small sample sizes outside Appalachia preclude meaningful regional recall estimates. The limited geographic and pipeline diversity in the training data suggests the model has primarily learned to recognize landslide signatures characteristic of Appalachian terrain, where smaller rotational and translational slides dominate.

Strain signatures from landslides in other physiographic regions may differ substantially from those in Appalachia. For example, landslides in glaciated terrain like British Columbia often exhibit more complex and deeper-seated deformation patterns. These regional variations in landslide mechanics, combined with differences in pipeline construction practices and terrain characteristics, likely present detection challenges that cannot be assessed with the current dataset. Future model development will prioritize expanding the training dataset to include more landslide examples from diverse geographic regions.

## Intended Use

This approach is meant to assist, not replace, expert judgement. Embedded in Cambio, the ML model can classify IMU runs with bending strain feature reports immediately after ingestion, and SMEs review the features exceeding an operator-chosen probability threshold. The probability can be tuned to match available review capacity, for example selecting the top ten to twenty percent of features for

SME interpretation. Outputs are displayed with the usual multi-panel strain plots and lidar or imagery context, so analysts can verify that the model is responding to physically meaningful portions of the topographic profile. When a feature is confirmed as landslide related, the geohazard likelihood term in the operator's PoF model can be updated. This closes the loop between quantitative screening, expert validation, and risk informed prioritization. Relative to the current practice of threshold driven triage followed by manual interpretation, the ML model reduces workload while improving consistency across vendors and projects.

### Limitations

As noted throughout this paper, there are some key limitations to the current version of this proof-of-concept model. Notably, this first application was weighted heavily toward larger diameter pipelines in Appalachian terrain, the present model focussed on single IMU runs rather than run to run change, and the study training set focused on vendor selected features. Each of these limitations is tractable with additional data and targeted model improvements.

## Future work

The next phase will expand the training dataset across additional physiographic regions and operators to improve generalization and increase the number of positive cases, especially in settings where axial or mixed-mode deformation dominates. Multi-run IMU change data will also be incorporated so that the model can recognize progression in strain between inspections. Contextual information such as topography, landslide polygons, lidar based- change, InSAR displacement, and horizontal directional drill (HDD) metadata will be introduced to the model to reduce false positives and improve detection of landslide impact signals. Data quality will continue to be improved through targeted reviews focussed on the use-case of training ML models. Gauge -length sensitivity will be evaluated, including adaptive schemes that scale with diameter.

## Acknowledgements

The authors thank the participating operator for permission to use anonymized IMU datasets in development of this model. We also acknowledge the many BGC Engineering SMEs who analyzed and documented the bending strain features in Cambio as well as the Cambio Earth team developing the database architecture and Cambio platform used to ingest and normalize the IMU data. The paper authors would like to specifically acknowledge the data preparation, modelling, and review work conducted by Katherine Liu, Caio Stringari, Colby Butcher, Colin Roth, Evelyn Moorhouse, Alex Mitchell, Chad Sutherland, Corey Scheip, Rodrigo Portela, and Andrew Johnson of Cambio Earth and BGC Engineering.

## References

Cruden, D.M., and Varnes, D.J., (1996, January). Landslide Types and Processes. Transportation Research Board, U.S. National Academy of Sciences, Special Report 247: 36-75.

Czyz, J.A., and Adams, J. R., (1994, February). Computation of Pipeline-Bending Strains Based on GEOPIG Measurements. Pipeline Pigging and Integrity Monitoring Conference. February 14-17, 1994. Houston, Texas. USA.

Dewar, D., (2020, September). Incorporating Inline Inspection Inertial Measurement Unit Data Analysis Into Integrity Management Programs. Proceedings of the International Pipeline Conference IPC2020. September 28-30, 2020. Calgary, Alberta, Canada.

Dotson, R., Brown, A., & Jackson, B. (2024, September). Consistently Inconsistent: Lessons from Operators Comparing Multiple IMU Data Sets. Proceedings of the International Pipeline Conference IPC2024. September 23-27, 2024. Calgary, Alberta, Canada.

Dowling, C., Barlow, P., Van Hove, J., Hanvi, T., Hart, J. (2024, February). Ground Movement or Construction? How to Identify Clear Ground Movement Signatures in Inertial Measurement. Proceedings of Pipeline Pigging Integrity Management Conference. Feb 12-16, 2024. Houston, Texas. USA.

Hart, J. D., Czyz, J. A., & Zulfiqar, N. (2019, March). Review Of Pipeline Inertial Surveying for Ground Movement-induced Deformations. Proceedings of the Conference on Asset Integrity Management-Pipeline Integrity Management Under Geohazard Conditions AIM-PIMG2019. March 25 - 28, 2019, Houston, TX, USA.

Mckenzie-Johnson, A., Theriault, B., Dotson, R., Hart, J. & Varela, P., (2024, September). Correlation of IMU Bending Strain Features to Geohazard Locations: An Update. Proceedings of the International Pipeline Conference IPC2024. September 23-27, 2024. Calgary, Alberta, Canada.

Newton, S., Zahradka, A., Ferris, G., Porter, M. (2019, March). Use of a geohazard management program to reduce pipeline failure rates. Proceedings of the Conference on Asset Integrity Management-Pipeline Integrity Management Under Geohazard Conditions AIM-PIMG2019. March 25 - 28, 2019, Houston, TX, USA.

Newton, S., Van Hove, J., Porter, M., Ferris, G. (2022, September). Assessing geohazard probability of pipeline failure: lessons and improvements from the last 10 years. 13th International Pipeline Conference. September 26 – 30, 2022. Calgary, Canada.

Newton, S., and Scheevel, C. (2025, January). A Novel Method for Geospatially Organizing and Integrating IMU Bending Strain Feature Data to Optimize Geohazard Threat Detection. Proceedings of Pipeline Pigging Integrity Management Conference. January 27-31, 2025. Houston, Texas. USA.

Scheevel C., Dowling C., Hart J.D., & Cook, D. (2022, March). IMU bending strain: analysis as geohazard screening tool. Pipeline Research Council International Research Exchange Meeting. Orlando, Florida, USA.

Theriault, B., Hart, J.D., Mckenzie-Johnson, A., & Paulsen, S. (2019, March). Correlation of single-run ILI IMU bending strain features to geohazard locations. Proceedings of the Conference of Asset Integrity Management-Pipeline Integrity Management Under Geohazard Conditions. Houston, Texas, USA.

Van Hove, J., Barlow, P., Dowling, C. (2024, February). The use of inertial measurement unit data for managing slow-moving landslides. Proceedings of Pipeline Pigging Integrity Management Conference. Feb 12-16, 2024. Houston, Texas. USA.