

IPC2026-185622

USE OF MACHINE LEARNING TO IDENTIFY LANDSLIDE SIGNATURES IN IMU BENDING STRAIN DATA

Aron Zahradka
Cambio Earth
Victoria, BC

Sarah Newton
Cambio Earth
Calgary, AB

Corey Scheip
BGC Engineering
Nashville, TN

Owen Bunce
BGC Engineering
Calgary, AB

Caio Stringari
BGC Engineering
Vancouver, BC

Jared Kowis
Enbridge
Houston, TX

ABSTRACT

Inertial Measurement Unit (IMU) bending strain data are critical to pipeline geohazard management, but identifying landslide signatures among thousands of strain features remains labor-intensive and reliant on manual expert interpretation. This paper presents progress on three machine learning approaches for automating landslide detection in IMU data, each operating at a different level of the analysis workflow. A bending strain classifier identifies landslide-impacted bending strain features, achieving 87% recall and 0.96 ROC-AUC on a spatially independent test set of over 9,000 features. At a matched recall to conventional magnitude-based screening criteria, the classifier reduces false positives by a factor of six, substantially lowering engineering review burden. A rolling window scanner extends detection to full IMU runs independent of strain feature reporting; it was tested on approximately 10,200 km of IMU across 224 runs and achieved 80.4% recall, including detection of landslide signatures below standard strain feature reporting thresholds. An early-stage segmentation framework moves toward localizing the spatial extent of landslide deformation within the IMU signal. Models are trained and evaluated using a spatial splitting strategy that enforces geographic independence between data partitions, with test data primarily drawn from Appalachian transmission lines. Together, these capabilities outline a progression from inefficient and inconsistent expert-driven screening toward scalable, pattern-driven geohazard assessment.

Keywords: Geohazards, Machine Learning, Pipeline Integrity, In-line Inspection, Inertial Measurement Unit, Bending Strain Assessment

1. INTRODUCTION

Very slow to extremely slow-moving landslides [1] are among the greatest geohazard-related threats affecting transmission pipeline integrity [2] and are capable of applying sustained external loads that lead to strain accumulation and potentially rupture. In-line inspection (ILI) tools equipped with inertial measurement units (IMUs) are the primary means of providing high-resolution measurements of pipeline pitch and azimuth from which landslide-induced bending strain can be detected in buried pipelines. Despite the precision of modern IMU inspection, analysis of these data is challenging and relies primarily on manual expert interpretation. IMU vendors or sub-consultants can provide services to identify and prioritize features for further review; however, the volume of strain features vastly exceeds those attributable to active ground movement, imposing a substantial review burden on geohazard management programs. Machine learning (ML) offers a path toward automating and improving this screening process. This paper describes the continued development of an ML framework for IMU bending strain interpretation, building on two prior studies that established proof-of-concept classification performance and explored different ML architectures [3,4]. By training a model to recognize patterns specifically associated with landslide deformation, we hypothesize that movement signatures can be detected at low strain magnitudes, potentially providing years of additional lead time within an operator's typical inspection cycle [5] or averting failures that might otherwise occur before the next IMU run.

1.1. State of Practice in IMU for Geohazard Management Programs

IMU-equipped ILI tools are now widely deployed across transmission pipeline networks on regular inspection intervals. These tools continuously record pitch and azimuth through the pipeline, from which horizontal and vertical bending strains are calculated [6]. As part of a bending strain analysis, ILI vendors or consultants may delineate strain features that exceed a minimum reporting threshold and deliver them to operators as structured strain feature reports. This threshold is commonly set at 0.125% total bending strain, outside of formed bends; bending strain below this threshold is generally not considered to pose an immediate integrity concern [7]. Different vendors and third parties apply various data processing techniques and reporting criteria, which can produce variable feature counts from the same source data [7]. Comparison of multiple IMU runs, when available, enables detection of changes in strain at very low levels (generally in the range of 0.04% or less).

Bending strain assessments identify potentially anomalous bending strain features or movement features regardless of source. The majority of strain features (approximately 94%) are not associated with geohazard activity but rather reflect strain features as a result of construction, pipeline replacements, maintenance excavations, or other third-party impacts [8,9]. To reduce the volume of features requiring detailed expert review, the industry has developed magnitude-based screening criteria that exploit the observed correlation between strain orientation and geohazard association. Dewar [5] presented a systematic screening criterion which has been widely adopted for identifying features requiring further inspection: all features with total strain $\geq 0.2\%$ or a horizontal strain component $\geq 0.125\%$, and all features intersecting known geohazard extents or interacting ILI anomalies. The emphasis on horizontal strain reflects its strong empirical association with landslide-induced deformation. Theriault et al. [9] found that the correlation between horizontal strain and likely geohazards exceeded 50% at magnitudes above 0.14% in a dataset of 4,618 features; McKenzie-Johnson et al. [8] subsequently confirmed horizontal strain as the strongest single predictor of geohazard association across a broader dataset of over 17,000 features, though the correlation at higher magnitudes was weaker than originally reported, likely reflecting greater geographic and construction diversity in the expanded dataset.

Geohazard specialists interpret characteristic indicators in the IMU, including specific sinusoidal bend patterns, bend lengths spanning girth welds, and deformation extents and patterns consistent with mapped landslide morphology [10]. This expert interpretation is effective but labor-intensive, dependent on specialist availability, and difficult to apply consistently at the scale of modern ILI programs where a single inspection run may generate hundreds of strain features for review. High-quality IMU screening is achievable but demands an often-unjustifiable investment, with the result that landslide signatures may already be present in existing data but remain unidentified.

1.2. Machine Learning for IMU Analysis

Human pattern recognition in complex, high-volume signal data is inherently limited, particularly at the scales encountered in modern ILI programs. ML is well-suited to this type of problem: recognizing spatial patterns in high-volume signal data, learning from labeled examples, and applying those learned patterns consistently across large datasets without fatigue or analyst variability. The objective is not to replace expert interpretation but to reduce the volume of features requiring it. A well-performing ML model can screen strain features automatically, allowing specialists to focus their time on the uncertain and borderline cases where detailed review is needed. This is the motivation for the work described in this paper.

Beyond landslide screening, ML has been applied to other aspects of IMU bending strain interpretation. Li et al. [11] used a wavelet neural network to improve the measurement precision of IMU-derived bending strain calculations by compensating for vibration and system noise during inspection. Liu et al. [12] developed a deep belief network to automatically classify pipeline features such as bends and welds from bending strain signals, addressing inconsistency and inefficiency in manual feature identification. Zhang et al. [13] applied a hybrid deep learning model to a large IMU dataset from the China–Russia crude oil pipeline to classify local deformation types including dents, elbows, and abnormal girth welds. These studies demonstrate growing interest in applying ML to IMU data but address different objectives than the present work.

1.3. Prior ML Work

Zahradka et al. presented a proof-of-concept ML classifier for screening IMU bending strain features for landslide impact at the 2026 Pipeline Pigging and Integrity Management Conference [3] and it was extended at the 2026 Pipeline Technology Conference [4]. The model was trained on over 30,000 strain features on North American transmission pipelines. Of these, 478 were labelled as landslide-impacted through subject matter expert (SME) interpretation supported by lidar imagery, historical ground movement mapping, and field validation. The remaining features carried either confirmed or assumed non-landslide labels. The dataset was heavily concentrated in the Appalachian region, which accounted for 94% of positive examples.

The test set was defined using geographically distinct polygons ensuring complete spatial independence from training data. The resulting holdout set contained 7,136 strain features including 112 confirmed landslide impacts. The model achieved 90% recall, 23% precision, and an area under the receiver operating characteristic curve (ROC-AUC) of 0.963. Compared to conventional magnitude-based screening criteria at a matched recall rate, the ML model reduced false positives by over 80%, cutting the engineering review burden from approximately 21 non-landslide features per landslide discovered to approximately four: a five-fold improvement in screening efficiency.

1.4. Objectives and Scope

This research extends the previously published proof-of-concept model [3,4] in two main directions: introducing a new full-run scanning capability that operates directly on raw IMU data, independent of vendor-reported strain features, and improving the strain feature classifier with expanded and cleaned training data.

The strain feature classifier (Section 2.5) screens features that have already been identified through other assessment methods. Section 2.6 introduces a methodology to adapt this classifier to run along a full ILI dataset and is a detection tool that identifies landslide-impacted pipeline segments that may not have been previously identified at all. Most existing strain feature analysis is predicated on a minimum total bending strain threshold, typically between 0.10% and 0.20%, below which strain features are not typically reported [5,7]. The full-run approach also provides consistent, automated analysis independent of vendor-specific feature identification methods and thresholds [7], and can be run alongside conventional strain analysis workflows or applied where vendor strain feature reports are not available.

The strain feature classifier previously presented demonstrated strong performance but was trained predominantly on Appalachian pipelines, with limited positive examples from other regions. The expanded dataset incorporates more data from operators with assets outside Appalachia, providing a somewhat improved basis for assessing model generalization, though further data collection beyond Appalachia is still required. Data quality has also been improved through targeted label review informed by the patterns of misclassification identified in the earlier work.

Both the classifier and the scanner remain focused on single-run IMU analysis. Existing run-to-run comparison methods are effective at detecting strain change between inspections but do not interpret its cause; conversely, the spatial patterns that distinguish landslide deformation from construction or operational sources are present in individual runs and do not require multi-run data. Single-run classification therefore provides baseline geohazard context that complements run-to-run change analysis. Integration of contextual data and multi-class strain classification (e.g. integrity digs, third-party activity, sinkholes, or construction) are reserved for future work.

2. MATERIALS AND METHODS

2.1. Dataset and Geographic Coverage

This study uses the same underlying IMU bending-strain dataset developed for the previous work with expansion, particularly in geographic coverage, relative to the original proof of concept model [3].

The dataset was sourced from Cambio, a software platform that stores IMU data from pipelines throughout North America within a standardized data framework. Vendor-specific IMU outputs are normalized into a common schema [14] and bending strains are calculated from the raw (generally 0.1 m) pitch and heading data to ensure consistent processing and gauge length

selection. Bending-strain features and the raw IMU data are stored as spatially referenced objects linked to pipeline centerlines and associated metadata. This standardized structure allows automation and reproducibility, critical for the work described in this paper.

The dataset remains geographically dominated by large-diameter, steel, gas transmission pipelines in the Appalachian region of the United States, primarily Ohio, Kentucky, and Pennsylvania, where landslide-pipeline interactions are common. Pipelines in the dataset are predominantly 30" and 36" nominal pipe size, with installation dates ranging from the 1940s to the 2020s. Landslides affecting these pipelines are nearly all classified as earth slides [15], predominantly extremely slow-moving [15], and with dominantly transverse loading in Appalachia. However, the dataset does include IMU data collected in other regions including Western Canada and the American Midwest, which are evaluated separately. These regions introduce variability in terrain, geology, landslide mechanisms, and pipeline installation histories, which influence the characteristics of IMU-derived bending strain signatures. For example, landslides in Western Canada tend to be deeper seated and with movement generally parallel to the pipeline (axial loading), producing distinct strain patterns from those in Appalachia.

2.2. Data Labeling

All strain features were screened by geohazard subject matter experts (SME) and features with potential ground movement signatures visible in the IMU underwent detailed review and were labelled using a five-tier classification system reflecting confidence in landslide association, as documented in Scheevel et al. [16]. SME interpretation draws on terrain analysis, historical ground movement mapping, lidar change detection (LCD) data, repeat IMU comparisons, and field-based site knowledge where available.

Tiers 1 and 2 indicate features considered likely landslide-related based on available evidence; tiers 4 and 5 indicate features considered unlikely to be associated with ground movement; and tier 3 denotes features where landslide association is uncertain. For model training and evaluation, labels were collapsed to a binary scheme: tiers 1 and 2 were assigned a positive (landslide) label, tiers 4 and 5 were assigned a negative label, and tier 3 were excluded from the analysis since the appropriate label cannot be determined.

In an advancement from the prior work, labels from coincident strain features from prior ILI runs were reviewed for consistency in interpretation. To iteratively improve label quality, SMEs conducted targeted review of FP and FN cases, correcting mislabeled features identified through misclassification analysis. This iterative label refinement was motivated by previous findings that several misclassified features reflected labeling ambiguity rather than model error [3]. This process has substantially improved the labelling quality since previous publications, but due to the volume of data it has not been exhaustive and some labelling quality issues likely remain. A summary of the number of strain features available for

modelling following data collection and cleaning, broken down by the labels drawn from SME inspections, is provided in TABLE 1.

TABLE 1: COUNT OF STRAIN FEATURES AVAILABLE FOR MODEL TRAINING BY SME INSPECTION STATUS LABEL.

Inspection Status	Count	Fraction
Inspected: Landslide	595	1.4%
Inspected: Non-landslide	2,472	5.8%
Not Inspected	39,941	92.8%
Total	43,008	

2.3. Spatial Splitting Strategy

Training, validation, and test sets were defined using geographically delineated polygons developed by SMEs. Polygon boundaries were drawn to promote complete spatial independence between splits, preventing information leakage that could arise from the correlation of strain patterns between adjacent pipeline segments within the same corridor. The target split was approximately 60% training, 15% validation, and 25% test by feature count, with class balance maintained. Polygons were distributed across all regions represented in the dataset to ensure each split contained examples from Appalachia, the American Midwest, American Gulf states, and Western Canada.

2.4. Machine Learning Model Architectures

Several different ML models were used throughout this paper. The main model is based on a CNN architecture called temporal convolutional network [17]. CNNs process data through a hierarchical learning framework, where learned filters (convolutions) detect features ranging from simple strain peaks in early layers to complex signatures in deeper layers.

Four IMU signals (pitch, azimuth, and the derived horizontal and vertical strain) are used as model inputs and are analyzed within approximately 200-meter windows to capture typical landslide deformation lengths (particularly within Appalachia). Horizontal and vertical strain were calculated from pitch and azimuth using a consistent gauge length of 3 m and were found to be useful engineered features that increased model performance over pitch and azimuth alone.

The second ML technique used in this work is XGBoost (Extreme Gradient Boosting) [18], a decision-tree-based model designed for structured, tabular inputs rather than spatial signal data. Where the CNN learns patterns directly from the shape of the IMU trace, XGBoost operates on scalar quantities such as peak strain magnitudes, combining multiple simple decision rules into a single predictor. This makes it well-suited to incorporating vendor-reported strain metrics that summarize a feature in a handful of numbers rather than a continuous signal.

The following three sections describe the application of these ML approaches: the two-stage CNN-XGBoost ensemble for classifying vendor-identified bending strain features, the rolling CNN applied to full ILI runs independent of vendor reporting, and an early-stage CNN segmentation framework that

moves toward spatially resolving landslide signatures within the IMU signal.

2.5. Bending Strain Classification

The bending strain feature classifier predicts a likelihood of each individual strain feature being landslide impacted. It uses a two-stage ensemble approach: in the first stage, a base CNN model processes the IMU window for a given strain feature, producing an output prediction score. In the second stage, this score is combined with vendor-reported static strain metrics (maximum horizontal, vertical, and total bending strain magnitudes for the feature and at girth welds) and passed to an XGBoost classifier trained on the combined input. The two-stage design allows the model to incorporate scalar magnitude information that is not directly available in the spatial IMU signals.

2.6. Rolling CNN Classification

The rolling CNN classifier extends the bending strain feature screener to operate directly on processed IMU runs, independent of vendor-reported strain features. Rather than classifying pre-identified features, this approach applies the base CNN model in a sliding window fashion along the full length of an IMU run, producing a continuous probability of landslide impact along the pipeline.

The CNN is applied to windows sampled at 10 m intervals along the IMU run. Each window produces a single scalar probability assigned to that span. The 10 m sampling interval means that any given point along the pipeline falls within 20 overlapping windows, each contributing a prediction. The contributions from all overlapping windows are then averaged to produce a single probability value at each evaluated point along the run.

This approach provides coverage of the full IMU run, including segments that fall below vendor reporting thresholds or were not prioritized for review for other reasons. However, this solution is not well-suited to precise spatial delineation.

2.7. Toward Full Landslide Segmentation

The ultimate objective of the segmentation framework is to identify the precise spatial extent of landslide-induced deformation within IMU runs. Landslides are rare events, representing a small fraction of the total pipeline length in any inspection dataset, and their strain signatures are spatially variable and difficult to distinguish from construction-associated deformation or other integrity concerns without broader contextual information. The class imbalance, combined with the limited number of fully spatially labeled landslide examples available for training, makes direct landslide segmentation a challenging problem. The work presented here represents an early experiment of this capability, and results should be interpreted accordingly.

3. RESULTS

All models were evaluated on the held-back, spatially independent, test set described in Section 2.3. The test set

comprises 9,108 strain features, including 131 labelled as landslide-impacted, drawn from polygons spanning Appalachia, Western Canada, and the American Midwest. These features were fully excluded from model training and represent pipelines and landslides not seen during the learning process. Results are presented first for the strain feature classifier across the full test dataset and regionally, then for the rolling CNN applied to full ILI runs. The model results are also compared to common industry screening criteria.

3.1. Bending Strain Classifier Performance

The two-stage strain feature classifier (Section 2.5) was applied to all test set features. FIGURE 1 presents a scatterplot of the model's prediction scores, grouped by SME-assigned label. Each point represents a single strain feature. The same results are presented in histogram form in FIGURE 2.

Features predicted as landslide-related cluster toward the right side of the plot (high confidence), while non-landslide features cluster toward the left. A classification threshold of 0.5 is shown as a dashed vertical line, dividing the plot into four quadrants corresponding to the standard outcomes of binary classification: true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP).

At this threshold, the model correctly identifies 114 of 131 (87.0%) landslide-impacted strain features as TP and correctly excludes 8,588 of 8,977 (95.7%) non-landslide features as TN. It produces 389 FP (non-landslide features incorrectly classified as landslides) and 17 FN (landslide-impacted features that the model did not identify). Standard classification performance metrics are summarized in TABLE 2.

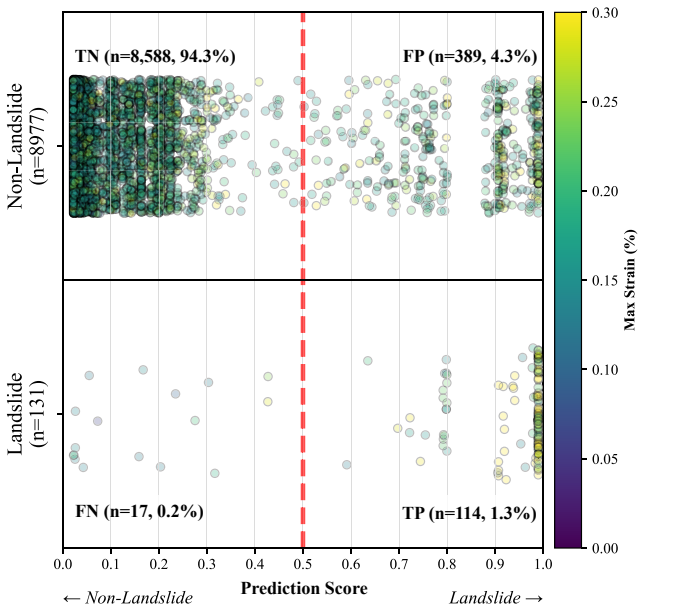


FIGURE 1: STRAIN FEATURE CLASSIFIER PREDICTION SCORES BY TARGET LABEL FOR 9,108 STRAIN FEATURES. POINTS ARE COLORED BY MAXIMUM TOTAL STRAIN REPORTED FOR THE BENDING STRAIN FEATURE; VERTICAL SPREAD IS RANDOM JITTER FOR BETTER DATA VISIBILITY.

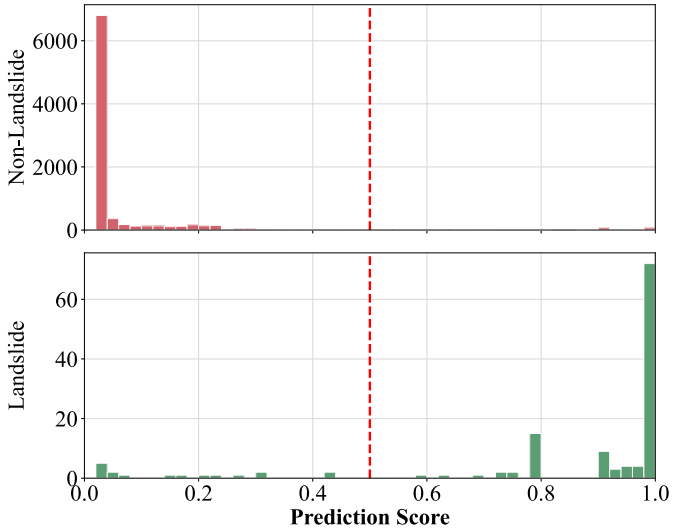


FIGURE 2: DISTRIBUTION OF STRAIN FEATURE CLASSIFIER PREDICTION SCORES BY TARGET LABEL FOR CLASSIFICATION OF 9,108 STRAIN FEATURES.

TABLE 2: PERFORMANCE METRICS FOR THE STRAIN FEATURE CLASSIFIER AT 0.5 PREDICTION SCORE THRESHOLD.

Metric	Value	Interpretation
Recall	87.0%	Proportion of landslides detected; missed landslides are undetected integrity threats
Precision	22.7%	Proportion of features classified as landslides that are actual landslides; reflects engineering review burden
Specificity	95.7%	Proportion of non-landslide features correctly excluded from review
F1 Score	0.36	Harmonic mean of precision and recall; single balanced performance measure
Accuracy	95.5%	Overall correct classifications; misleading here due to 69:1 class imbalance*
AUC	0.96	Threshold-independent metric of overall ability to differentiate between classes

*A trivial model classifying every feature as non-landslide would achieve 98.6% accuracy while detecting zero landslides. Accuracy is reported for completeness but should be interpreted alongside recall and precision.

The ROC curve provides a threshold-independent view of model performance by plotting recall against the FP rate (1 – specificity) across all confidence thresholds. The ROC-AUC summarizes discriminative performance as a single scalar, where 1.0 indicates perfect class separation and 0.5 indicates a random

guess. The strain feature classifier achieves an ROC-AUC of 0.96 as shown in FIGURE 3. The steep initial rise of the curve demonstrates that substantial recall can be achieved at very low FP rates, which is an important property for operational deployment where the cost of reviewing FPs is significant. The long tail before the TP rate reaches 1.0 (around FP rate of 0.6) indicates that, based solely on the model predictions, over half of the non-landslide strain features would need to be reviewed before no landslides are missed; in practice, there is generally a substantial amount of contextual data available to narrow the triage.

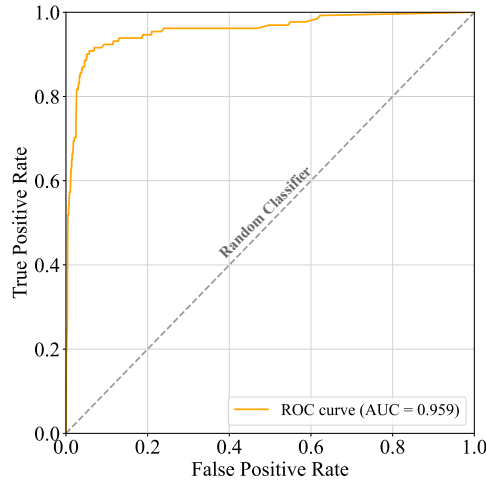


FIGURE 3: RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE FOR BINARY CLASSIFICATION OF VENDOR IDENTIFIED STRAINS WITH THE CNN-XGBOOST MODEL.

3.2. Bending Strain Classifier Regional Performance

To assess geographic variability in model performance, the test set was partitioned into three regions: Appalachia, Western Canada, and Other USA (comprising features from Pennsylvania, Mississippi, Texas, Arkansas, and Minnesota). Confusion matrices for each region are presented in FIGURE 4. Specificity is consistent across all three regions, ranging from 94.5% in Western Canada to 97.3% in Other USA, indicating that the model does not generate excessive FPs when applied to pipelines outside its primary training domain. This is an operationally important property: deploying the model in regions with limited training representation does not flood the screening workflow with spurious landslide predictions.

The non-Appalachian positive sample count has approximately doubled relative to the dataset previously used [3], but remains insufficient to characterize regional detection performance with any confidence. The test set contains 10 landslide-impacted features in Western Canada and two in Other USA, compared to 119 in Appalachia. Given these counts, computed recall rates would be statistically unreliable and are not reported as summary metrics here; the confusion matrices are presented for transparency.

An asymmetry between specificity and recall generalization is expected. Non-landslide strain patterns arising from construction techniques, road crossings, and utility crossings are broadly similar across regions, giving the model a consistent basis for excluding benign features. Landslide signatures, by contrast, vary with terrain, geology, and movement mechanism, and the model has substantially fewer labeled examples from non-Appalachian settings to learn from. Expanding the positive training set in underrepresented regions remains a priority for improving and validating regional recall and reducing the number of false negatives outside of Appalachia.

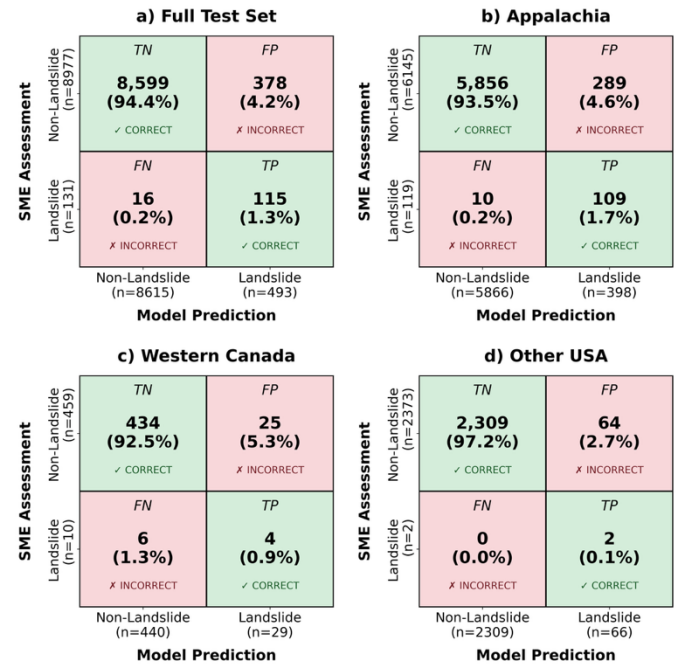


FIGURE 4: OVERALL AND REGIONAL CONFUSION MATRICES FOR THE STRAIN FEATURE CLASSIFIER AT 0.5 PROBABILITY THRESHOLD, EVALUATED ON SPATIALLY INDEPENDENT TEST SET PARTITIONS FOR A) THE FULL TEST DATASET (N=9,108); B) APPALACHIA (N=6,264); C) WESTERN CANADA (N=469); D) OTHER USA (N=2,375).

3.3. Full ILI Run Performance

The rolling full run IMU scanner (Section 2.6) was evaluated on all IMU data within the spatially independent test region, comprising approximately 10,200 km across 224 ILI runs. The scanner operates using the base CNN model only, without access to vendor-reported scalar strain metrics or pre-identified strain features.

FIGURE 5 illustrates the scanner output for a representative 25 km section of IMU from the testing dataset. The model produces a continuous prediction score along the run and segments exceeding the threshold are highlighted as potential landslide features. In this example, the scanner correctly identifies one landslide at approximately 18 km chainage while producing four FP segments: three with no corresponding vendor

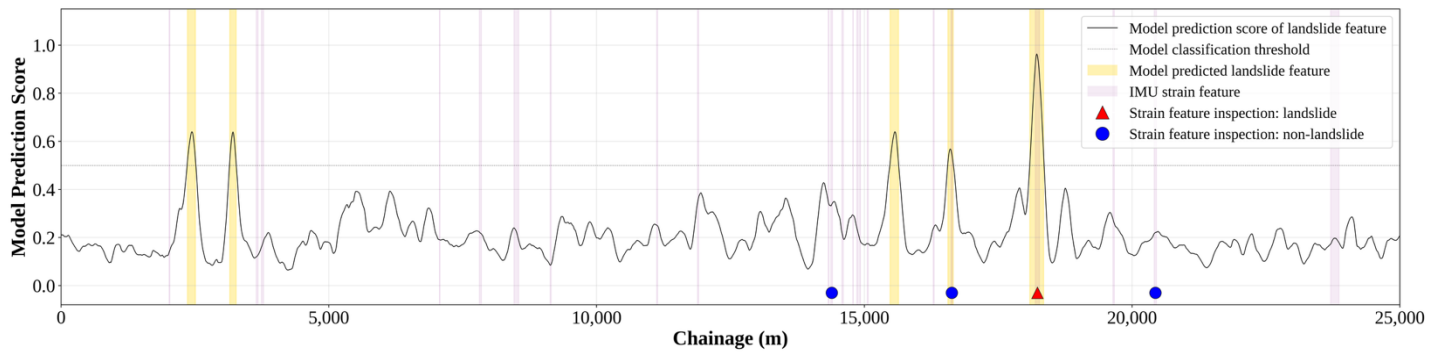


FIGURE 5: FULL IMU RUN SCANNER OUTPUT FOR A REPRESENTATIVE 25 KM SECTION OF IMU FROM THE TESTING DATASET. CONTINUOUS PREDICTION SCORE IS SHOWN AS BLACK LINE; SEGMENTS EXCEEDING THE 0.5 CLASSIFICATION THRESHOLD ARE HIGHLIGHTED IN YELLOW.

strain feature and one intersecting a reported (and SME inspected) feature. Several additional strain features along the run, including two with negative SME inspections, receive low prediction scores and are correctly excluded.

Because the scanner classifies full ILI runs rather than discrete pre-identified features, the evaluation framework differs from standard binary classification. A positive segment is classified as a TP if it intersects one or more strain features with a landslide-impacted SME inspection, and FP otherwise. A FN is a landslide-impacted strain feature not intersected by a predicted positive segment. TNs are not defined, as there are no discrete negative objects; instead, the fraction of each run classified as positive provides a measure of screening specificity.

Across the 224 test runs, the scanner produced 1,551 positive segments intersecting 115 of 143 landslide-impacted strain features, yielding 80.4% recall and 7.4% precision (FIGURE 6). This corresponds to approximately 12 non-landslide segments per landslide detected. Of the ~10,200 km of IMU evaluated, 249 km (2.4%) was classified as potentially landslide-impacted. Note that the slightly higher number of landslide-impacted strain features in this test relative to the bending strain classifier test is a result of some strain features missing the vendor-reported static strain metrics required for the strain classifier model to operate.

The composition of FPs provides insight into scanner behavior (FIGURE 7). Of 1,436 FP segments, 706 (49%) intersected vendor-reported strain features, indicating that the scanner is often responding to genuine anomalies in the IMU signal rather than noise. The remaining 730 FP segments (51%) occurred in areas with no vendor strain feature report; some of these could in principle represent unidentified ground movement, though the geohazard programs contributing data to this study are very comprehensive and it is unlikely that a substantial number represent never-identified landslides.

The scanner has also demonstrated the capability to detect landslide signatures at magnitudes below conventional vendor reporting thresholds. There were 83 bending strain features in the test dataset with maximum total unformed bending strain below 0.1%, a population that would receive no attention under

standard screening of single-run IMU; three of these were confirmed as landslide-impacted through independent evidence including run-to-run change and field observation; the scanner independently identified two of these three, operating solely on the spatial pattern of the IMU signal. The sample is small but demonstrates that pattern-based detection can identify meaningful signatures at magnitudes where conventional screening would not look.

		Full Run Scanner Confusion Matrix	
SME Assessment	Non-Landslide (n=3153)	TN N/A	FP 1,436 (43.6%) X INCORRECT
	Landslide (n=143)	FN 28 (0.8%) X INCORRECT	TP 115 (3.5%) ✓ CORRECT
		Non-Landslide (n=1745)	Landslide (n=1551)
		Model Prediction	

FIGURE 6: CONFUSION MATRIX FOR THE FULL IMU RUN SCANNER ACROSS 224 TEST RUNS (~10,200 KM). TRUE NEGATIVE CLASSIFICATIONS ARE NOT DEFINED AS THERE ARE NO DISCRETE NEGATIVE OBJECTS IN THE FULL-RUN EVALUATION FRAMEWORK.

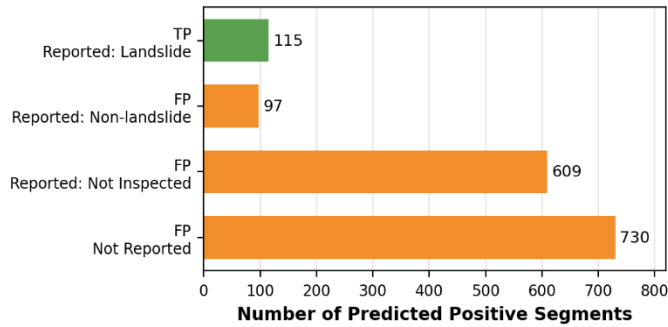


FIGURE 7: CLASSIFICATION BREAKDOWN OF 1,551 PREDICTED POSITIVE SEGMENTS IDENTIFIED BY THE FULL IMU RUN SCANNER ACROSS 224 IMU RUNS IN THE TEST DATASET, CATEGORIZED BY INTERSECTION WITH VENDOR-REPORTED STRAIN FEATURES AND SME INSPECTION OUTCOME.

3.4. Comparison to Current Practice

Assessing the value of an ML-based screening approach requires a controlled comparison against the alternatives available to operators. The industry screening criteria described in Section 1.1 (horizontal strain $\geq 0.125\%$ or total strain $\geq 0.2\%$ based on Dewar [5]) serves as the comparison baseline for the feature classifier. These criteria are deterministic thresholds applied to scalar strain magnitudes. They reduce review volume effectively but cannot account for the spatial structure of the strain signal, are binary in output (providing no indication of relative likelihood), and cannot detect landslide signatures below their magnitude thresholds.

The ML models presented in this paper operate on fundamentally different information by incorporating the IMU data itself (horizontal strain, vertical strain, pitch, and azimuth), learning to recognize the spatial patterns associated with landslide deformation directly from labeled examples. Because the ML model outputs a continuous probability for each strain feature rather than a binary classification, the classification threshold can be selected to match the recall of the industry screening criteria, enabling a controlled comparison of FP rates at equivalent detection performance.

FIGURE 8 presents confusion matrices for both approaches at this matched recall. The difference in screening efficiency is substantial: the industry criteria prioritize 2,468 non-landslide features as requiring review, achieving 4.5% precision, meaning that for every landslide identified, an engineer must review approximately 21 non-landslide features; the ML classifier reduces FPs to 408 at the same detection rate, achieving 22.1% precision and a non-landslide to landslide ratio of 3.5, a six-fold improvement in screening efficiency. Both approaches miss the same number of landslides but do not necessarily miss the same landslides.

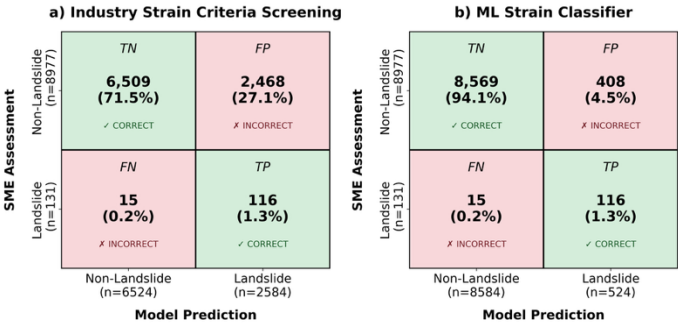


FIGURE 8: CONFUSION MATRICES COMPARING A) INDUSTRY STRAIN CRITERIA SCREENING AND B) THE ML STRAIN CLASSIFIER AGAINST SME ASSESSMENTS. THE TP COUNT IS INTENTIONALLY MATCHED FOR COMPARISON.

TABLE 3 summarizes performance metrics alongside the full-run scanner. The scanner operates without vendor-reported strain features and achieves 80.4% recall with a FP-to-TP ratio of 12.5. While less efficient than the strain feature classifier, it provides coverage of pipeline segments that would not be evaluated under either the industry criteria or the feature classifier, as both require pre-identified strain features to operate.

TABLE 3: COMPARISON OF SCREENING APPROACHES. THE ML STRAIN CLASSIFIER THRESHOLD (0.43) IS SELECTED TO MATCH THE RECALL OF INDUSTRY SCREENING CRITERIA.

Method	Recall	Precision	FP/TP
Industry Criteria	88.5%	4.5%	21.3
ML Strain Classifier	88.5%	22.1%	3.5
ML Full Run Scanner	80.4%	7.4%	12.5

3.5. Semantic Segmentation

The output of the segmentation framework (Section 2.7) for a representative 100 m segment from the spatially independent test set is shown in FIGURE 9. The top two panels display the raw IMU-derived (a) horizontal and (b) vertical bending strain and the locations for girth welds. Panel (c) shows the ML prediction score of formed bends (blue line) and the ground truth locations of the bends (shaded areas). The model can delineate the locations of the formed bends well, and for the selected area, does not miss any formed bends. The bottom panel (d) shows the landslide ML prediction score predicted by the cascaded model. The segmentation model correctly identifies the general location and extent of the landslide, although the segmentation is not spatially perfect. This reflects the early stage of this capability and the limited number and variability in fully annotated landslide examples available for training. Despite these limitations, the example highlights the key advantage of the segmentation approach over window-level classification by localizing deformation within the signal rather than assigning a single ML prediction score to an entire window.

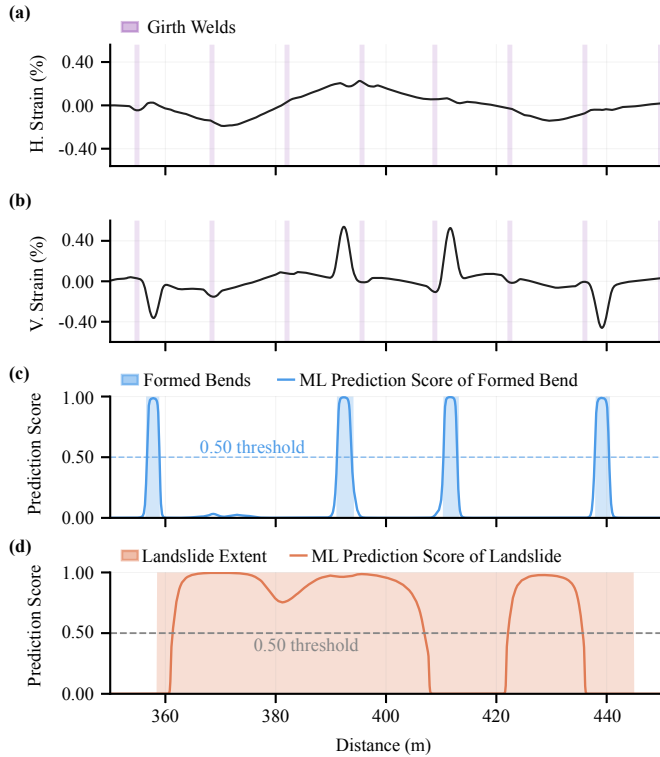


FIGURE 9: SELECTED EXAMPLE OF SEMATIC SEGMENTATION SHOWING A) HORIZONTAL STRAIN, B) VERTICAL STRAIN, C) ML PREDICTION SCORE AND GROUND TRUTH FOR FORMED BENDS, AND D) ML PREDICTION SCORE AND GROUND TRUTH FOR LANDSLIDES.

4. DISCUSSION

4.1. Misclassified Strain Features

As part of the review of the results, the FP and FN bending strain features from the strain feature classifier were investigated. The results were similar to those discussed in [3], highlighting the need for continued enhancement of the training data quality, despite substantial progress in cleaning the label quality issues previously identified, and addition of contextual information.

For most FPs (strains the ML model incorrectly identified), the SME generally would have relied on contextual information not provided to the ML model such as the ground morphology obtained from lidar, construction considerations (such as road or river crossings) or site history (previous digs or adjacent construction).

For the FNs (strains that the ML model missed), the incorrect determination was found to be a result of subtle bending strain signatures, incorrect data labeling, or contextual information aiding in the SME interpretation. IMU data is reviewed in more detail at known active geohazard sites, in cases of suspected or known pipeline impact, or where there are more obvious strain signatures on neighboring pipelines, resulting in

the SME identification of landslide impact from subtle strain signatures.

The FP and FN results from the full-run scanner were also reviewed. Many of the FNs were located at consistent locations, either from subsequent ILI runs on the same pipeline or from ILI runs on adjacent pipelines. This is a positive indication that the model is predicting consistent results on similar strain signatures. Other FNs were located at sites with complex site histories of mitigation and repair, where historical strain interpretations were applied to the bending strain features following the work on site, despite changes to the bending strain pattern. The lower recall of the full-run scanner relative to the strain feature classifier reflects the absence of the second-stage model which incorporates vendor-reported scalar strain metrics that are not available when scanning arbitrary segments of the IMU run and remain key components of the analysis.

Select FP strains were reviewed and revealed interesting insight into the model interpretation. One section of pipeline with a high model prediction score was found to be at a river crossing where the pipeline was installed via horizontal directional drill (HDD). HDDs have distinct signatures (and distinct landslide impact signatures), typically long wavelength, low magnitude bends. With no information or context of this change in installation method and strain signature, the ML model labeled this section with a high likelihood of landslide impact. This may be due to the abrupt change at the HDD tie-in location. Providing these models with additional information regarding the construction technique and potential sources of strain could increase the effectiveness of the models.

4.2. Program Implications

The three modeling capabilities presented in this paper, the bending strain feature classifier, the full-run scanner, and the segmentation framework, represent successive steps toward a more complete and autonomous approach to landslide detection from IMU data. Each step extends coverage and reduces reliance on bending strain screening criteria and often inconsistent human analysis; together, they outline a trajectory from expert-assisted triage toward scalable, signature-driven geohazard assessment.

The most immediate program implication of the classifier results is a meaningful reduction in manual review burden without a corresponding increase in missed landslides, freeing specialist time for features that genuinely warrant detailed attention. The model produces a continuous probability for each feature and the classification threshold can be adjusted to reflect operator-specific risk tolerance: lowering it increases recall at the cost of more features requiring review; while raising it reduces review burden but risks missing lower-confidence landslide impacts.

The feature classifier, however, operates on vendor-reported strain features, which are themselves subject to a minimum reporting threshold. In practice, features with low bending strain magnitudes are not typically reviewed [5] on the reasonable basis that not everything can be looked at and that the bulk of high-consequence deformation exceeds these levels. This is a pragmatic constraint, but not without cost. The full-run scanner

was able to successfully detect two of the three locations in the IMU testing dataset that had confirmed landslide impacts with total bending strains below 0.1% and would not have been flagged in single-run IMU through a traditional bending strain feature assessment. Whether due to low strain magnitude, provider-specific processing decisions, or the lack of a strain feature report altogether, the entire IMU run becomes accessible to detection with the full-run scanner and can provide earlier warning of landslide impact and more time to plan a response. In circumstances where landslides are moving fast enough to exceed strain capacity before the next IMU run, earlier detection may avert pipeline failures. For this capability, the relevant comparison is not only relative screening efficiency but whether the model identifies geohazard impacts that current practice would not detect at all.

4.3. Future Work

The immediate priority for future work is expanding the geographic and operational diversity of the training dataset. Expanding the number of landslide examples from other physiographic regions and pipeline types and diameters is essential for validating detection of signatures that differ in mechanism, depth, and size. This effort is actively underway with additional operators.

Integration of contextual data such as terrain morphology from lidar, construction records, and horizontal directional drill locations represents a high-impact avenue for improvement. As previously discussed, contextual information is the primary basis for SME interpretation that the model currently lacks and incorporating these data sources is expected to address a substantial fraction of the current FP population.

Validation of the full-run scanner requires targeted investigation of model predictions in segments below vendor reporting thresholds. The examples in Section 3.3 are encouraging but limited; systematic ground-truthing supported by lidar, LCD, and field observation is needed to establish whether pattern-based detection identifies signatures where conventional screening does not look at an operationally useful level.

The segmentation framework introduces spatial transparency to model output, identifying which portions of the IMU signal drew the model's attention rather than assigning a single probability to an entire window. This potential will require substantially more spatially annotated training data and extension to multi-class detection of non-landslide strain sources such as other geohazards (e.g. sinkholes), third-party excavation, or construction/maintenance.

The testing dataset contained several examples of landslide impacts overlapping field bends but the performance of the model in this situation needs to be further evaluated. Comparison of repeat IMU runs to detect movement areas becomes a critical analysis once multiple runs are available at a location. Integrating the ML classifier to identify the source of movement will be an important extension to this work.

Finally, these models will be integrated into the Cambio software platform's IMU data management module within the

existing strain plotting and inspection workflow, with continued development focused on operator-specific threshold tuning and confidence-based review prioritization to embed ML-assisted screening as a routine component of geohazard management programs.

5. CONCLUSION

This paper presents three machine learning capabilities for identifying landslide signatures in IMU bending strain data. On a spatially independent test set of over 9,000 features, the strain feature classifier achieves 87% recall and an AUC of 0.96, reducing the SME review burden roughly six-fold relative to conventional magnitude-based screening at matched recall. The full-run scanner extends detection to entire IMU runs independent of vendor-reported features and flagged confirmed landslide signatures below standard reporting thresholds where no screening currently exists. The early-stage segmentation framework begins to localize the spatial extent of deformation within the signal, pointing toward spatially resolved, interpretable output.

Performance to date is best characterized in Appalachia, which dominates the training and test data; specificity generalizes well across regions, but recall elsewhere cannot yet be meaningfully assessed. Expanding the geographic diversity of labeled training data remains the primary constraint on broader deployment, supported by the iterative, misclassification-driven label refinement that has materially improved data quality in this work.

These capabilities are intended to complement, not replace, expert interpretation and existing screening practices. Together, they mark a progression from labor-intensive, threshold-driven screening toward prioritized, scalable, and pattern-driven geohazard assessment.

ACKNOWLEDGEMENTS

The report authors would like to acknowledge the contributions of the Cambio community of pipeline operators, BGC engineers, geoscientists, and data scientists, and Cambio Earth software developers, data scientists and product managers. Specific thanks to Pete Barlow, Michael Porter, Joel Van Hove, Andrew Johnson, Colin Roth, Colby Butcher, Devin Frioud, and Katherine Liu for their work on BGC and Cambio geohazard programs and the ML models described in this paper.

The following generative AI tools were used in the preparation of this manuscript: Cursor (v2.6.14 with Anthropic's Opus 4.6 and Sonnet 4.6, OpenAI's GPT-5.2, and Cursor's Composer 1.5) was used to assist with code implementation for data processing and machine learning; Anthropic's Claude Desktop (v1.1.5368 with Opus 4.6 and Sonnet 4.6) was used to identify relevant literature, critique proposed content, and edit proposed text for readability and style throughout this manuscript. All AI outputs were carefully reviewed and edited; the authors take full responsibility for the integrity of all content.

REFERENCES

- [1] Porter, M., Quinn, P., and Barlow, P., "Conceptual Landslide Velocity Transition Models for a Range of Landslide Behaviour Types."
- [2] Van Hove, J., and Newton, S., 2025, "Operational Experience with Monitoring Tools for Managing the Threat to Pipelines from Slow-Moving Landslides."
- [3] Zahradka, A., Newton, S., Bunce, O., Van Hove, J. V., and Kowis, J., 2026, "Automated Landslide Impact Detection Through Machine Learning Analysis of Pipeline IMU Bending-Strain Data."
- [4] Scheip, C., Stringari, C., Bunce, O., Zahradka, A., Newton, S., Perry, J., and Baumgard, A., 2026, "Machine Learning Classification of IMU Strain Anomalies for Detection of Landslide Impacts on Pipelines [Manuscript under Review]," Euro Institute for Information and Technology Transfer, Berlin, Germany.
- [5] Dewar, D., 2020, "Incorporating Inline Inspection Internal Measurement Unit Data Analysis Into Integrity Management Programs."
- [6] Czyz, D. J. A., and Adams, J. R., 1994, "Computation of Pipeline-Bending Strains Based on GEOPIG Measurements," Pipeline Pigging and Integrity Monitoring (PPIM), Houston, TX, USA.
- [7] Dotson, R., Jackson, B., Stevenson, M., and Newton, C., 2024, "Don't Be Strained: A Case Study and Lessons Learned on Comparing Multiple IMU Data Sets Over a Decade," *36th International Pipeline Pigging and Integrity Management Conference (PPIM 2024)*, Clarion Technical Conferences, Houston, TX, USA, pp. 561–576. <https://doi.org/10.52202/072781-0031>.
- [8] McKenzie-Johnson, A., Theriault, B., Dotson, R., Hart, J., and Varela, P., 2024, "Correlation of IMU Bending Strain Features to Geohazard Locations: An Update," *Volume 4: Geohazards Management and Strain-Based Design and Assessment*, American Society of Mechanical Engineers, Calgary, Alberta, Canada, p. V004T06A008. <https://doi.org/10.1115/IPC2024-133093>.
- [9] Theriault, B., Hart, J. D., McKenzie-Johnson, A., and Paulsen, S., 2020, "Correlation of Single-Run ILI IMU Bending Strain Features to Geohazard Locations," *Pipeline Integrity Management Under Geohazard Conditions (PIMG)*, ASME Press, pp. 293–301. https://doi.org/10.1115/1.861998_ch31.
- [10] Dowling, C., Barlow, P., Van Hove, J., Hanvi, T., and Hart, J., 2024, "Ground Movement Or Construction? How to Identify Ground Movement Signatures in Inertial Measurement Unit Data," *36th International Pipeline Pigging and Integrity Management Conference (PPIM 2024)*, Clarion Technical Conferences, Houston, TX, USA, pp. 1339–1360. <https://doi.org/10.52202/072781-0075>.
- [11] Li, R., Cai, M., Shi, Y., Feng, Q., Liu, S., and Zhao, X., 2016, "Pipeline Bending Strain Measurement and Compensation Technology Based on Wavelet Neural Network," *Journal of Sensors*, 2016, pp. 1–7. <https://doi.org/10.1155/2016/8363242>.
- [12] Liu, S., Wang, H., Li, R., and Ji, B., 2022, "A Novel Feature Identification Method of Pipeline In-Line Inspected Bending Strain Based on Optimized Deep Belief Network Model," *Energies*, 15(4), p. 1586. <https://doi.org/10.3390/en15041586>.
- [13] Zhang, D., Liu, X., Fu, M., Liu, S., Shao, J., Chen, P., Li, R., Fu, K., and Cheng, J., 2025, "A Novel Local Deformation Pipe Section Identification Method via IMU Detection Data and Hybrid Deep Learning Model," *Mechanical Systems and Signal Processing*, 224, p. 112091. <https://doi.org/10.1016/j.ymssp.2024.112091>.
- [14] Newton, S., and Scheevel, C., 2025, "A Novel Method for Geospatially Organizing and Integrating IMU Bending Strain Feature Data to Optimize Geohazard Threat Detection," *37th International Pipeline Pigging and Integrity Management Conference (PPIM 2025)*, Clarion Technical Conferences, Houston, TX, USA, pp. 1085–1100. <https://doi.org/10.52202/078572-0058>.
- [15] Cruden, D. M. and Varnes, D.J., 1996, "Landslide Types and Processes," Special Report - National Research Council, Transportation Research Board, 247, p. 76.
- [16] Scheevel, C., Dowling, C., Hart, J. D., and Cook, D., "IMU Bending Strain: Analysis as Geohazard Screening Tool."
- [17] Bai, S., Kolter, J. Z., and Koltun, V., 2018, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling." <https://doi.org/10.48550/arXiv.1803.01271>.
- [18] Chen, T., and Guestrin, C., 2016, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, San Francisco California USA, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.