Continual Anomaly Detection for Evolving Time Series in Steel Industry

Shreyas Gaddam*
gshreyas@iitbhilai.ac.in
IIT Bhilai
Bhilai, India

Himanshu Kurrey*
himanshukurrey@iitbhilai.ac.in
IIT Bhilai
Bhilai, India

Gagan Raj Gupta
gagan@iitbhilai.ac.in
IIT Bhilai
Bhilai, India

Vishwesh Jatala
vishwesh@iitbhilai.ac.in
IIT Bhilai
Bhilai, India

Naman Agarwal
namanagarwal097@gmail.com
Radcom
Delhi, India

Abstract

In this paper, we present a long-term study of deploying a system for predicting and reducing failures through Time Series Anomaly Detection in a rolling mill steel plant. We implemented two pipelines: a Machine Learning pipeline for detecting collective anomalies to predict potential failures, and an Event-Flow Analysis pipeline to identify event-based anomalies. A key challenge for inference was frequent concept drift caused by physical, mechanical, and configuration changes in the mill operation, often influenced by hidden or inaccessible parameters, leading to a multimodal distribution of data. Additionally, the mill operates in a closed-loop control system, which increases the likelihood of False Positives. We discuss the incremental features added to the pipeline, including a custom data normalization technique, and continual learning, all of which significantly reduced false alerts and improved overall performance. Notably, these improvements extended the lead time, which is the interval between the predicted failure and its actual occurrence, giving operators more time to react.

CCS Concepts

• **Computing methodologies → Machine learning**.

Keywords

Time Series Anomaly Detection, Event Flow Analysis

ACM Reference Format:
Shreyas Gaddam, Himanshu Kurrey, Gagan Raj Gupta, Vishwesh Jatala, and Naman Agarwal. 2025. Continual Anomaly Detection for Evolving Time Series in Steel Industry. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

*Both authors contributed equally to this research.

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

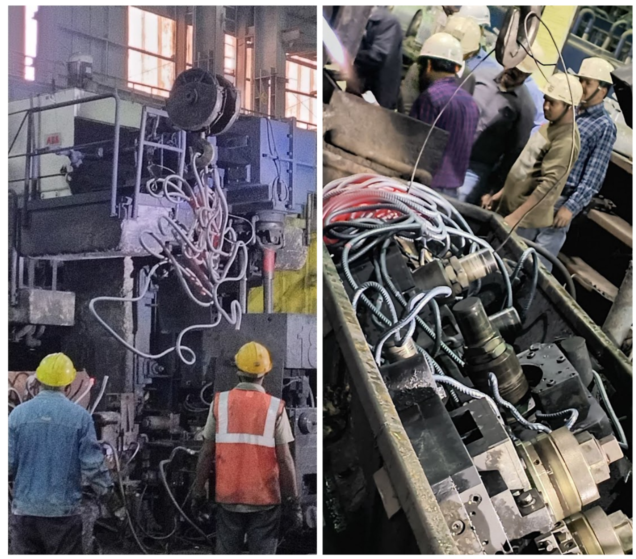


Figure 1: Cobble Failure in Intermediate and Fast Finishing Areas

1 Introduction

Many industrial environments deploy several sensors to monitor the equipment and the manufacturing process. These sensors generate a huge amount of data that can be processed to detect failures during production environment. Predicting failures precisely and ahead of time has the potential to improve production, reduce maintenance overhead, and reduce cost. To predict failures, many previous anomaly detection techniques [7][8][27] were proposed in the literature; however, most of them are evaluated on synthetic datasets.

In this paper, we present a system[1] for predicting failures in a steel rolling mill through Time Series Anomaly Detection (TSAD). The rolling mill has a capacity of producing 900,000 metric tons of product per annum[2]. It encounters many failures during the production, an example has been depicted in Figure 1 with high safety risks. The failures encountered during rolling process can incur huge loss (approx $18,000 for every hour of inactivity) to the

[1]Source Code: https://github.com/shreydan/KDD25_ADS_Continual_Anomaly_Detection_Steel_Industry
[2]The details of the mill are described in Section 2.

steel plant. Therefore it is crucial to predict and prevent potential failures early enough for them to take any preventive action. We were tasked to design a pre-failure alerting system with high recall. Our system has been deployed at the rolling mill for 6 months.

1.1 Challenges

The rolling mill is operated using an automated closed-loop control feedback system. The control system sends timed commands and reference values to each equipment as the metal progresses down the manufacturing line. Based on the feedback, the control system recalculates the timing and reference values. Most of the abnormalities are occurred due to the inability of the automation system to auto-correct the process, posing several challenges to predict failures precisely during the deployment.

Cascading effect due to self-correcting nature: The control system self-corrects the rolling process based on the feedback, which leads to change in the signal behavior. For example, increasing looper height (explained in Section 2) to reduce tension in one section can inadvertently change in other signals, i.e., torque and revolution per minute (RPM) signals. Adjusting multiple signal values can result in an higher reconstruction error leading to a failure prediction. Most of the times such predictions are *false positives* as the self-correcting nature is an expected behavior of the mill instead of an anomaly.

Indeterminable state of the mill: While we train our initial baseline models on the offline data, the model may not perform well during deployment due to frequent data drifts, as shown in Figure 2(a) for the RPM signal. The drifts are difficult to estimate as it depends on several hidden control parameters, the quality of the billet, and the current status of the equipment.

Multimodal distribution of data: As the mill configurations are dynamic, many of the signals exhibit multimodal distributions, as depicted in Figure 2(c). Hence, it becomes difficult to apply the traditional data normalization techniques. Moreover, the data can exhibit different trends as shown in Figure 2(b), which depends on the automation while processing for consecutive billets.

Transient States: When the billet enters or leaves the equipment rapidly, there is a sudden change in the state of the signals. These changes introduce peaks and troughs in the response signals. These are difficult are to be trained in the model as the transient states have short duration.

1.2 Contributions

To address these challenges, we develop a system that takes a multi-pronged approach to detect and prevent failures. The system implementation has two pipelines, i.e., a machine learning pipeline for detecting collective anomalies to predict potential failures, and an event-flow analysis pipeline to identify event-based anomalies. The system is also provisioned with an alerting mechanism to facilitate the mill operators to act on the failures.

In the machine learning pipeline, we perform data normalization and apply the multivariate time series analysis and anomaly detection using continual machine learning techniques. We train our machine learning model on each of the identified pain regions and perform inference in parallel. The event flow analysis pipeline uses specific signals to detect the occurrence of micro-events in the mill, which substantiates the ML predictions. The alerting system produces alerts to the mill operator to act on the failure. In addition, it keeps track of the feedback received from the operators, which is verified by domain experts.

To summarize, this paper makes the following contributions.

(1) We deployed a multi-variate time series anomaly detection for failure prediction in the industrial environment. We observed that with the help of our continual learning techniques, the traditional TSAD algorithms, such as PCA and LSTM based autoencoders are effective in timely predicting anomalous patterns along with the relevant signals.
(2) We deployed an Event Flow Analysis system to detect any event-based failures which helped in targeted and quick maintenance. It also helped detect sensor faults and equipment condition.
(3) The proposed system has been deployed in the rolling mill of the steel plant for six months. With the help of our alerting dashboard, mill operators have taken corrective actions while mill is operating, or they performed short maintenance checks, which resulted in reducing the number of unplanned downtimes.
(4) Our system was effective in preventing 15 failures (Approx. 20% reduction in the downtime of the mill in a month) during January 2025.

The rest of the paper is organized as follows. Section 2 gives the required background for our approach, and Section 3 describes the related work. Section 4 presents the details of our approach. Section 5 and 6 discuss the experimental results. Finally, we conclude the paper in Section 7.

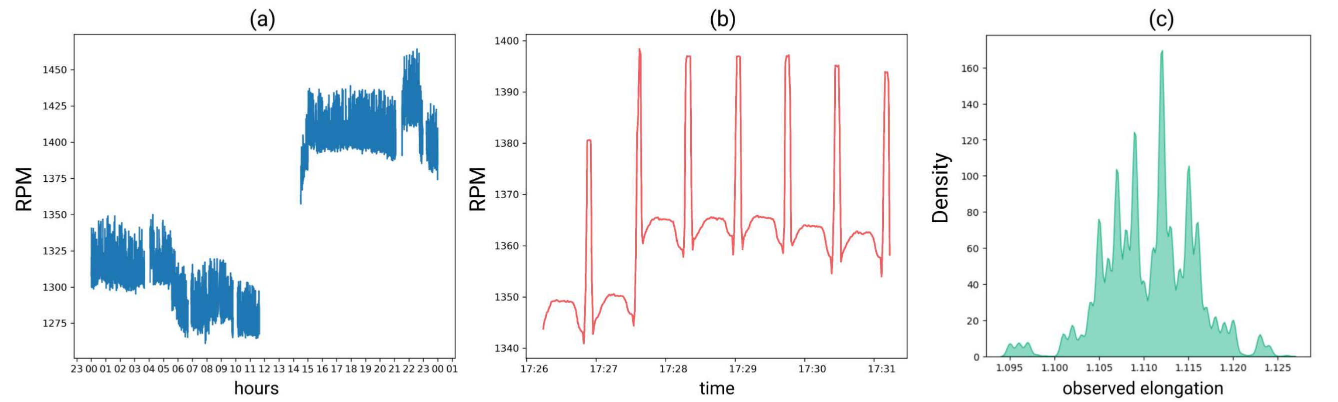
2 Background

In this section we give a brief background on the rolling process used in steel plant and describe the multi-variate time series anomaly detection method.

2.1 Rolling Process

The rolling mill studied in this paper consists of 18 rolling stands and two parallel lateral lines. The stands are located in three regions, viz., roughing area, intermediate area, and finish areas, as shown in Figure 3. The stands are responsible for applying the compressive forces leading to rise of tension in the metal. The regions also consists of loopers to reduce this tension. The loopers are placed between every two stands which lift the metal physically forming a loop to reduce the built-up tension.

The rolling process starts from an alloy known as billet, which enters through the furnace entrance at stand-1. During the rolling process, the billet is heated to very high temperature, the stands 1 to 18 apply compressive forces on heated billet. This reduces the cross-sectional area of the billet, increases the length of the billet, and the speed of the flowing metal increases rapidly along the mill. The initial stands in the roughing area carry out maximum work to elongate the metal which increases the tension rapidly. The intermediate area contains of stands and loopers to both elongate the metal and to control the internal tension. The finishing areas also consists of stands and loopers. But a slitter is used after stand 16 to split the metal in two parts and is sent to the parallel lateral



(a) shows change in RPM range over 24 hours; (b) shows the trend in the signal which is expected and not an anomaly; (c) displays multimodal distribution of a signal in a single day

Figure 2: Challenges in the Industrial Data at the rolling mill where our system is deployed

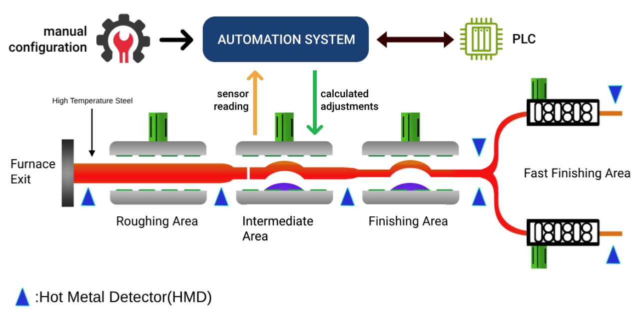


Figure 3: Mill Overview

lines. Due to the intricate arrangement of the equipment there is a high likelihood of failures in this region.

The fast finishing area shown in the figure is used to cut the metal in appropriate lengths to reduce deformities and for trade purposes. Each of the parallel laterlal line in the fast finishing area consists of fast finishing blocks (FFB), which are the twin equipment arrangements in both the high-speed lines which give the final shape and structure to the rod. The speed here is much higher than the previous regions and any failure here causes the most damage.

2.2 Time Series Anomaly Detection

The steel plant industries are equipped with multiple sensors in various of part of the mill to monitor the status of the mill. These sensors generate a huge amount time series data. Analyzing the multivariate time series (MTS) data is a crucial to predict the failures. An MTS, which is collected from n sensors, can be represented as matrix T, $(s_1, s_2 \ldots s_n)^T$. The length of the time series, $|T|$, corresponds to the number of time points captured. Each sensor s_i generates a $|T|$-dimensional vector $s_i = (x_{i,1},\ x_{i,2}, \ldots,\ x_{i,|T|}\ )$where $x_{i,j}$ represents the measurement from sensor i at time point j.

In the context of a rolling mill, these sensors monitor a diverse range of operational signals, such as torque, rotational speed (RPM), elongation, and temperature. Anomalies detected in these signals can indicate abnormal equipment behavior, which, if left unaddressed, may escalate into equipment failures or lead to suboptimal operational performance.

3 Related Work

3.1 Datasets

Researchers have used many benchmark datasets, such as SWaT [23], WADI [2], and SMAP [1] for Time Series Anomaly Detection (TSAD). However, most of these synthetically generated; they do not depict the real-world industrial scenarios, and hence have limitations in several dimensions, such as (1) model trained on the synthetic data will over-fit and may not work efficiently on the real-world data, (2) the real-world data can have many arbitrary nuances, which are not projected in the synthetic data. However, in this study we perform anomaly detection in the the real-world industry data collected from the rolling mill of a steel industry.

3.2 Anomaly Detection Techniques

Many different algorithms have been developed for detecting anomalies in time series. These generally fall into forecasting methods, reconstruction methods, and distance techniques. Forecasting methods, such as ARIMA [19], SARIMA [18], DeepAnT [24], and DeepNAP [10] predict the behavior the future time series data based on the past data. Reconstruction methods such as PCA [9], LSTM-VAE [13, 26], and USAD [6] apply the auto-encoder techniques to encode the high dimensional data into a lower dimension and subsequently decode the data to obtain the original time series data. If the reconstruction error, which is difference between the original and reconstructed data exceeds a certain a threshold, anomaly is detected.

Distance methods, such as DAMP [22], KNN [29][11], STAMP [32], MADRID [21] computes z-normalized Euclidean distance between any subsequence within a time series data. The position at which the subsequence having a maximum distance, known as discord, is treated as a anomaly point. While most of the these techniques are developed for uni-variate time series data they are not evaluated

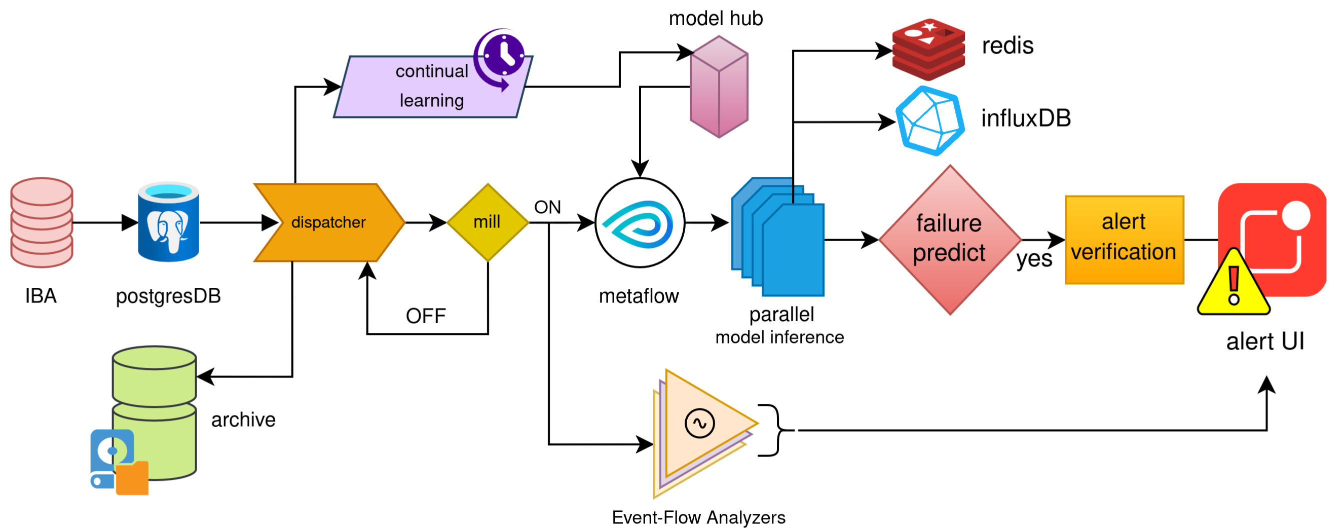


Figure 4: System Pipeline depicting data collection, pre-processing, parallel inference, continual learning, event-flow analyzers and alert UI

for multi-variate time series. However, few distance methods techniques, such as DAMP [22] are limited for uni-variate time series data. In this study we compare the performance of PCA, LSTM, TSG, and USAD techniques for the multi-variate time series data to understand the effectiveness of their approach on multi-variate series and choose the efficient approach for our study in the rolling mill.

Most of the anomaly detection techniques are applied on the synthetically generated data. However, real world applications comes with additional complexities. For instance, real-world data streams often do not conform to the same distribution over time to address this, MCD-DD [30] propose a drift detection technique and adaptively identify varying forms of concept drift to improve the accuracy of the model. Similarly, real-world application need to understand root-cause analysis [28, 33] of the failure. Recently, Agarwal et al [3] propose root-cause analysis technique to not only predict the failures, but also identify the root cause of the anomaly. REASON [31] uses hierarchical graph neural network to identify the root-cause anomalies. In addition, several ad hoc methods [12][5] have been been proposed to mitigate the false positives in the both synthetic and real-word datasets [4].

Most of the anomaly detection uses precision and recall as the key metrics for measuring the effectiveness of their approaches. Recently, PATE [14] proposed Proximity-Aware Time series anomaly Evaluation (PATE), a novel evaluation metric that incorporates the temporal relationship between prediction and anomaly intervals to improve the accuracy of the prediction. However, in the real-world scenarios, it is crucial detect the failures early. As a result, Agarwal et al [3] introduced a new metric, lead time, which is time difference between the actual event anomaly and the predicted anomaly. When the lead time is sufficient enough, the operators in the industry will be able to act on the predicted failures. Several surveys are conducted for the time series data [27][8][7][15], although some studies have compared unsupervised TSAD algorithms, there is a notable gap in research comparing these algorithms on real-life industrial data.

4 Methodology

Our end-to-end pipeline is shown in Figure 4. The pipeline starts with the acquisition of data from the Programmable Logic Controllers (PLCs) followed by several stages of pre-processing. The pre-processed data goes through two different pipelines, i.e., one machine learning pipeline and even flow analysis pipeline. The ML pipeline performs the continual learning using multi-variate anomaly detection techniques, performs inference, and stores the raw data along with anomaly prediction scores in the persistent data storage. Event flow analysis pipeline tracks different micro-events in the mill to substantiate ML model predictions.

Along with the failure prediction, we provide a user-friendly alerting system that sends alerts as shown in Figure 14 to the mill operators. For every failure prediction, our alerting system sends the information about the failure, such as the region of failure, the type, severity of alert, the most relevant signals associated with that particular alert. Along with alerting mechanism, we provide a live-monitoring dashboard, which provides comprehensive view of all signals of any region.

4.1 Data Acquisition and Pre-Processing

We collect the real-time data using IBA data acquisition server [12], which reads the data from the relevant PLC signals deployed at several locations in the mill. The data is stored in PostgreSQL database [16], which is subsequently used by both event flow analysis and another and the model inference

Once the data is collected, we perform the data preprocessing, which is a crucial step to ensure data quality and the accuracy of the anomaly detection models. Due to the periodic nature of the signals and the different characteristics of the signal, we've updated data preprocessing steps with our own custom pipeline based on the activity in the equipment. We have grouped the mill into 11

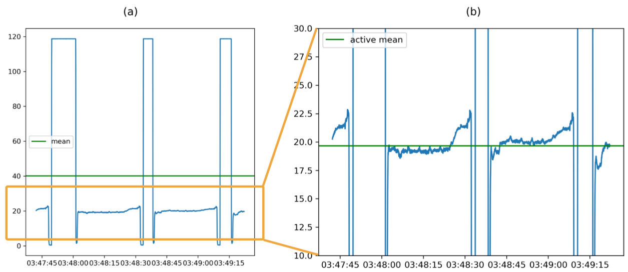


Figure 5: Custom Data Normalization: Active State

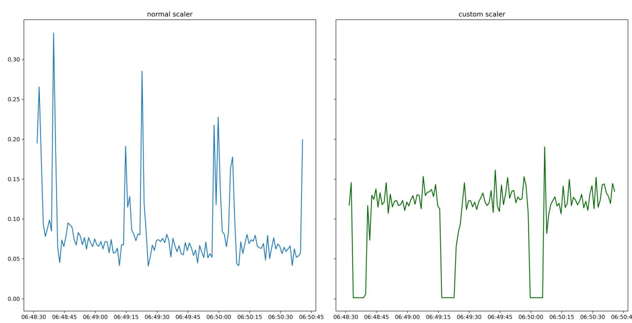


Figure 6: Custom Data Normalization: Anomaly Scores

different regions for targeted detection and root cause analysis. We further branch each region into its constituting equipments. This information is used by our custom data normalization pipeline, which we discuss further.

4.1.1 Custom Data Normalization. For each equipment we have in a particular region, we define its corresponding signals and their normalization technique as follows:

- **Control Signal:** It is a binary signal indicates if the metal is present in the equipment, this allows us to bifurcate normalization into two states: active and inactive as the behaviour of the signals varies as per the state.
- **Affected Signals:** These are the signals which have different behaviour in active and inactive state. For these signals we employ two different Mean-Variance scalers to normalize as they have different mean and variance in the two equipment states.
- **Unaffected Signals:** These are the remaining signals of an equipment whose behaviour does not depend on the state of the equipment. In this case we only have one Mean-Variance scaler to normalize them.

For the affected signals, we further apply mask the signals in their inactive state, since the domain experts in the mill have suggested us to mainly focus on the signal characteristics when the metal is present in a particular equipment. This allowed us to focus on the active state in fig.5(b) in the signals instead of the unwanted sensor readings in fig.5(a) disturbing the mean and variance of important timespans.

Figure 6 illustrates how this custom scaler resulted in the better representation of the final anomaly scores as we can clearly see the behaviour of individual billets with minimized transients in contrast to the normal scaler where the transients completely suppresses the stable state without the custom normalization pipeline.

4.2 Machine Learning Pipeline

The deployment pipeline as seen in Figure 4 seamlessly integrates data processing, model execution, and continual learning strategies. It monitors operations, and mill configurations to automate model deployment as per the configuration. By incorporating continual learning, the system refines models over time, adapting to evolving patterns in the data. It also enables dynamic threshold updates, ensuring predictions remain accurate based on changes in the mill configuration. Parallel processing of models enhances efficiency, delivering reliable insights for real-time decision-making.

4.2.1 Dispatcher. The dispatcher unit constantly monitors the mill's operational status. If the mill is inactive, it resets the necessary buffers and statistics to ensure a clean slate for when operations resume. When the mill is running, the system processes incoming data in batches, prepares it for inference, and triggers the downstream model workflow.

4.2.2 Model Inference. For each region, we require a framework to run inferencing in parallel for each model to avoid delay in our predictions. Hence, we chose to use Metaflow[25] which is a framework that helps manage machine learning workflows. Each Metaflow branch corresponds to one region, each step selects the appropriate signals for the region, scales the data using our custom data normalization technique, runs model inference, and makes prediction of a failure as per the model outputs, all while using our in-memory and persistent databases as required. The databases we use during inferencing are as follows:

- **InfluxDB:** It is a time series database which is used to persistently store our time-series data from PLC, and the scores which we can use for live signal monitoring in our dashboard or look at data historically.
- **Redis:** It is an in-memory database which we use to store inference-time data such as profile configurations, hyperparameters such as the current on-going thresholds and other information that needs to be shared across multiple batches.

AutoEncoder Based Approach: We mainly deployed PCA and LSTM autoencoders as per their performance per region. The absolute reconstruction error for each signal s_j at timestep t_i is defined as:

$$E_{i,j} = \left|X_{i,j} - \hat{X}_{i,j}\right|$$

The anomaly score for every region is defined as the mean absolute reconstruction error across m signals at each timestep:

$$\text{anomaly\_score}_i = \frac{1}{m}\sum_{j=1}^{m} E_{i,j}$$

We then apply the Grouping Algorithm [20] to find collective anomalies from the anomaly scores of a region.A collective anomaly is a collection of point anomalies that cross the threshold where each point is in the fixed proximity of neighbouring anomalous

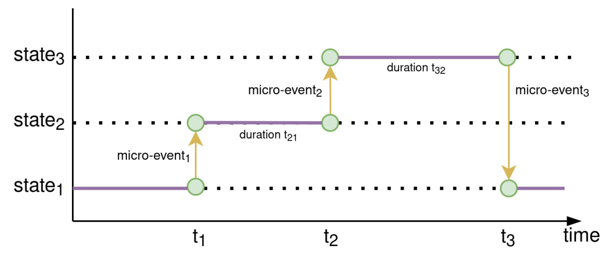


Figure 7: Micro-Event Tracker

points defined as *max_td* or maximum time difference between two points. If the amount of point anomalies becomes equal to another hyperparameter *count*, we predict that there is a possibility of failure. The top reconstruction errors E_j affectively give us the *root-cause signals* for any alert which makes it easy for the operator to verify.

4.2.3 Continual Learning. To address the challenges of concept drift (as discussed in 1.1) in real-time data, our pipeline periodically utilizes the latest 3 hours of data to update the scalers, and retrains the models. We preprocess the data by removing mill-off timespans and scale using our custom approach as detailed in Section 4.1. This approach minimizes the impact of concept drift over time and ensures that the model remains adaptive to changing conditions in the rolling mill environment. For PCA, we update the entire model, and for LSTM we load the existing weights and train it with a lower learning rate.

4.3 Event Flow Analysis Pipeline

There are different micro-events in the mill that need to be tracked alongside our Machine Learning pipeline, these events can then be correlated with our model predictions to support our root cause analysis further. These micro-events can be sequential and sometimes cyclic due to the periodic nature of the mill. These micro-events can also affect the signal behaviour causing our ML system to predict failures. Hence, to further improve our inference pipeline, we use specific signals to detect the occurrence of these micro-events, the duration between these contiguous events, and their order as observed in Figure 7. Any change in the order of specific events may be deemed catastrophic which are termed as "continuity alarms" in the mill. The duration of a micro-event also needs to be consistent across upcoming metals as a failure can occur if the duration falls short or goes beyond the tolerable range. Cyclic micro-events correspond to the position of movable equipment which also needs to follow a certain repetitive order at the right intervals.

An example we can discuss here is a sensor known as a Hot Metal Detector (HMD). HMDs play a vital role in the functioning and the feedback loop of the mill. The mill has an array of these HMDs strategically placed at key intervals along the manufacturing line to help detect the position, speed, and length of the metal used by the control system to generate calculated commands along the line. Some HMD signals are binary indicating if metal is present or not, and some also have analog capabilities to sense the height of the metal. Sometimes these HMDs can have faulty sensing due to external factors such as water, dust, or accumulation of residual high-temperature metal nearby. Any imperfections in the sensing by these detectors may result in misinterpretation of the position of the metal leading to control commands being generated at the wrong time culminating in an avoidable failure, or a preventive action is taken by their control system to protect the expensive equipment but results in metal wastage. Our events watch system monitors all 35+ HMD signals at a higher sampling rate, quickly alerting the operators on micro-events such as the duration of metal in a region, faulty sensing, incorrect metal height measurements, etc. As we are aware of the physical locations of these detectors, it becomes easier for the mill operators to trace the origination of a failure based on alerts raised due to HMD and our inference pipeline.

5 Deployment Setup

We performed our experiments on the industrial dataset collected from the roll mill of a steel plant. The data is collected from through the IBA data acquisition server [17], which reads the data from the PLC signals deployed at various locations in the mill. The dataset consists of 230+ signals. We use 1 second sampling rate for ML pipeline and 50ms for Event-Flow Analysis pipeline

We trained the multi-variate anomaly detection model with the popular approaches: PCA, LSTM, USAD on the top-5 regions of the mill. We configured LSTM with around 250,000 trainable parameters. We set the upper bound of lead time as 7 minutes, so anything beyond is a False Positive.

Our system was deployed on a High-Performance Computing (HPC) platform featuring an Intel(R) Xeon(R) Gold 5218 CPU at 2.30 GHz and an NVIDIA RTX A2000 GPU. We used Python v3.11.4, PostgreSQL v15.6, Metaflow v2.11.10, InfluxDB v2.7.0, Redis v2.54.0, in the pipeline, and Grafana v10.0.2 for the dashboard.

6 Results and Discussion

In this section, we evaluate our deployment approach on real-world operational data in the rolling mill as it progressed and verify it with our experimental results. We aim to answer the following key research questions:

RQ1: Pre-Deployment (Offline) Studies: Can existing Time Series Anomaly Detection Algorithms be effective in real-world industrial settings?

RQ2: Improvements in Deployment: How effective have custom data normalization and continual learning been, in improving recall and lead time?

RQ3: Deployed System Efficacy: Was our system helpful in detecting potential failures and what advantage did the operators gain?

6.1 RQ1: Pre-Deployment (Offline) Studies

Before our deployment began, we evaluated multiple models: PCA, LSTM, and USAD[6] and were able to predict failures in different regions with their case studies in the Appendix8.2.1. Although the performance of USAD was good as seen in Table 3, we did not have any interpretability for root-cause analysis. Hence, we proceeded with PCA and LSTM for the deployment. For ease of deployment,

Table 1: Performance Metrics for 12mm Profile

Region	Scaler	PCA				LSTM			
		Continual		Base		**Continual**		Base	
		Recall	Lead Time (s)	Recall	Lead Time (s)	Recall	Lead Time (s)	Recall	Lead Time (s)
Roughing	**Custom**	**0.981**	**197**	0.830	191	**0.792**	186	0.760	189
	Normal	0.776	189	0.630	191	**0.802**	207	0.400	251
Intermediate	**Custom**	**0.983**	**185**	0.675	163	**0.823**	182	0.343	127
	Normal	0.817	170	0.190	252	**0.854**	140	0.121	157
Finishing	**Custom**	**0.991**	**192**	0.670	181	**0.930**	175	0.286	167
	Normal	0.951	189	0.190	152	**0.984**	185	0.250	159
FFB Line 1	**Custom**	**0.913**	**203**	0.203	157	**0.629**	175	0.265	92
	Normal	0.816	198	0.345	178	**0.764**	170	0.318	105
FFB Line 2	**Custom**	**0.827**	**217**	0.660	181	**0.687**	167	0.301	185
	Normal	0.586	198	0.155	119	**0.752**	179	0.278	129

Table 2: Performance Metrics for 16mm Profile

Region	Scaler	PCA				LSTM			
		Continual		Base		**Continual**		Base	
		Recall	Lead Time (s)	Recall	Lead Time (s)	Recall	Lead Time (s)	Recall	Lead Time (s)
Roughing	**Custom**	**1.0**	**159**	1.0	168	**0.812**	139	0.653	154
	Normal	0.625	167	0.742	124	**0.687**	136	0.333	166
Intermediate	**Custom**	**0.991**	**168**	0.675	123	**0.481**	103	0.130	157
	Normal	0.562	146	0.459	119	**0.812**	178	0.174	149
Finishing	**Custom**	**0.996**	155	0.940	147	**0.677**	164	0.286	120
	Normal	0.375	154	0.310	132	**0.625**	159	0.615	130
FFB Line 1	**Custom**	**0.916**	162	0.908	184	**0.623**	184	0.095	134
	Normal	0.558	173	0.231	126	**0.498**	153	0.174	137
FFB Line 2	**Custom**	**0.791**	**206**	0.658	178	**0.560**	142	0.578	134
	Normal	0.614	177	0.154	148	**0.686**	203	0.217	162

Finishing Area	PCA	LSTM	USAD
Recall	0.190	0.250	0.219
Lead Time (s)	152	159	156

Table 3: Performance comparison in the Finishing Area for 12mm profile

we started with PCA as it was consistent across multiple regions and profiles as we see in Table 1 and 2.

On deployment, we got too many alerts as seen in Table 4 without our custom scaler and continual training features. So many that we were unable to differentiate relevant and irrelevant alerts and made it difficult for us to assess what caused the alert to trigger. The mill operators wanted us to provide them with relevant alerts even if it was caused by the self-correcting nature of the mill.

On comparing the data distribution of online and offline data, we noticed significant concept drift within a day and across days for every profile resulting in multimodal data distribution depicted in Figure 2. We also noticed that the transients in the signals added lot of noise in our anomaly scores. This analysis confirmed that existing TSAD approaches are valid only if we effectively address the concept drift.

6.2 RQ2: Improvements in Deployment

Based on our detailed trend and distribution analysis, and requirements from the mill operators, we started with the development of custom data normalization pipeline as described along with the benefits in Section 4.1. To address the data drift issue, it was evident that we need to deploy continual learning. The continual learning approach is described in Section 4.2.3. Due to periodic updation of the scalers and models, we were able to reduce our alerts by an average of 70% (Table 4). This helped us to get more relevant alerts allowing us to take constant feedback from the mill incharges.

As we progressed, we added Event Flow Analysis to our system whose working is described in 4.3. HMD Event Flow Analyzer was able to detect HMD sensor faults which helped in maintenance and failure prevention. Elongation Event Flow Analyzer detected whenever elongation signals went out of acceptable range.

Table 4: False Positive Alerts Across 5 Regions

Method	Alerts
Without Continual & Custom Scaler	813
With Continual & Custom Scaler	220

Table 5: Anomaly Distribution Across Mill Regions in the Previous Month

Region	Root Cause			
	Loopers	DRM	HMD	Misc
Roughing	0	1	1	1
Intermediate	2	4	0	0
Finishing	2	1	0	0
Fast Finishing	0	1	1	1

We were further able to validate that when we address the data behaviour, simple TSAD approaches work well with good interpretability.

6.3 RQ3: Deployed System Efficacy

An analysis of anomaly alerts over a one-month period are categorized in Table 5 in the four operational regions of the rolling mill are shown in Figure 3. Anomalies are categorized into four root causes: Malfunctions in Looper, Drive Failures (DRM), Sensor faults in HMD, and Miscellaneous which are unexpected disturbances; all of which were detected by our system. We elaborated on the alerts our system gave which included the contributing signals, the reason for the alert and what action was taken by the operators based on our alerts in Appendix Table 6.

This has given us substantiate evidence that our system has been able to reduce potential failures and instabilities with both our pipelines.

6.4 Case Studies

The following case studies illustrate various failure types and demonstrate how the deployed system's alerts have contributed to minimizing downtime and improving mill productivity:

Figure 8: Case Study 1 - Looper Scanner Malfunction and HMD Flickering Impact on Stand 13-18

6.4.1 ***CASE STUDY 1:*** A warning was received for Stand 13-18 (fig.8), where Stand 15 elongation fluctuated, disturbing looper height (Stand 14-15). The CP2 operator verified signals, stopped the mill, and cleaned the faulty looper scanner. Additionally, Stand 11 downstream HMD flickering coincided with the scanner malfunction, confirming signal correlation.

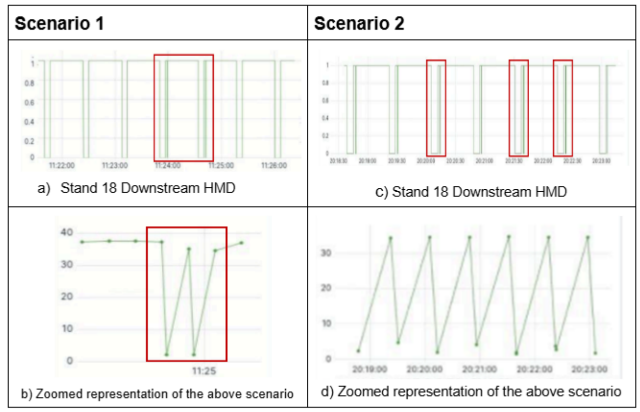


Figure 9: Case Study 2 - HMD Flickering & Hot Metal Deposition

6.4.2 ***CASE STUDY 2:*** HMD flickering (fig.9) was detected due to metal deposits, leading to false hot metal signals and potential mill failure. Scenario 1 (fig.9a,b) showed intermittent flickering with no impact, while Scenario 2 (fig.9c,d) showed continuous flickering, risking billet obstruction. The mill was halted for cleaning and realignment, preventing cobble failure and ensuring smooth operations.

7 Conclusion

This paper presents a long term study of our system for predicting failures in a rolling mill steel plant through Time Series Anomaly Detection. The system has two pipelines i.e., a machine learning pipeline for detecting collective anomalies to predict potential failures, and an event-flow analysis pipeline to identify event-based anomalies. The proposed system has been deployed at the industrial mill for six months and has been producing alerts to operators about the failure prediction. Our system was shown to be effective in preventing 15 failures, reducing 20% in the downtime of the mill in a month, during the January 2025.

References

[1] [n. d.]. SMAP Dataset Description, NASA Website; https://smap.jpl.nasa.gov/data/. https://smap.jpl.nasa.gov/data/
[2] [n. d.]. WADI Dataset, itrust-labs, SUTD; https://itrust.sutd.edu.sg. https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/
[3] Naman Agarwal, Gagan Raj Gupta, Vishwesh Jatala, and Aniket Saha. 2025. Multivariate Time Series Data Mining for Failure Prediction Root Cause Analysis. In *ICASSP*.
[4] Ignacio Aguilera-Martos, David López, Marta García-Barzana, Julián Luengo, and Francisco Herrera. 2022. Time-series Industrial Anomaly dataset; https://github.com/ari-dasci/OD-TINA. https://github.com/ari-dasci/OD-TINA

[5] Khaled Al Jallad et. al. 2020. Anomaly detection optimization using big data and deep learning to reduce false-positive. *Journal of Big Data* 7, 1 (2020), 68. doi:10.1186/s40537-020-00346-1
[6] Julien Audibert et. al. 2020. USAD: UnSupervised Anomaly Detection on Multivariate Time Series. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Virtual Event, CA, USA) *(KDD '20)*. Association for Computing Machinery, New York, NY, USA, 3395–3404. doi:10.1145/3394486.3403392
[7] Heraldo Borges et. al. 2021. Anomaly Detection in Time Series. *Transactions on Large-Scale Data- and Knowledge-Centered Systems* (2021). https://hal-lirmm.ccsd.cnrs.fr/lirmm-03359500
[8] Varun Chandola et. al. 2009. Anomaly Detection: A Survey. 41, 3, Article 15 (jul 2009), 58 pages. doi:10.1145/1541880.1541882
[9] Stéphane Crépey et. al. 2022. Anomaly detection in financial time series by principal component analysis and neural networks. *Algorithms* 15, 10 (2022), 385.
[10] Chunggyeom Kim et. al. 2018. DeepNAP: Deep neural anomaly pre-detection in a semiconductor fab. *Information Sciences* 457-458 (2018), 1–11. doi:10.1016/j.ins.2018.05.020
[11] Langfu CUI et. al. 2023. A method for satellite time series anomaly detection based on fast-DTW and improved-KNN. *Chinese Journal of Aeronautics* 36, 2 (2023), 149–159. doi:10.1016/j.cja.2022.05.001
[12] Martin Grill et. al. 2017. Reducing false positives of network anomaly detection by local adaptive multivariate smoothing. *J. Comput. System Sci.* 83, 1 (2017), 43–57. doi:10.1016/j.jcss.2016.03.007
[13] Pankaj Malhotra et. al. 2016. LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection. arXiv:1607.00148
[14] Ramin Ghorbani, Marcel J.T. Reinders, and David M.J. Tax. 2024. PATE: Proximity-Aware Time Series Anomaly Evaluation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Barcelona, Spain) *(KDD '24)*. Association for Computing Machinery, New York, NY, USA, 872–883. doi:10.1145/3637528.3671971
[15] Markus Goldstein and Seiichi Uchida. 2016. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PloS one* 11, 4 (2016), e0152173.
[16] PostgreSQL Global Development Group; https://www.postgresql.org/. 2024. PostgreSQL. https://www.postgresql.org/
[17] Integrated Business Architecture (IBA). 2024. Integrated Business Architecture (IBA); https://www.iba-ag.com/en/. https://www.iba-ag.com/en/
[18] Jian-Bin Kao et. al. 2019. Anomaly Detection for Univariate Time Series with Statistics and Deep Learning. In *2019 IEEE Eurasia Conference on IOT, Communication and Engineering*. 404–407. doi:10.1109/ECICE47484.2019.8942727
[19] Viacheslav Kozitsin et. al. 2021. Online Forecasting and Anomaly Detection Based on the ARIMA Model. *Applied Sciences* 11, 7 (2021). doi:10.3390/app11073194
[20] Suraj Kumar et. al. 2023. Fast Anomaly Detection For Multivariate Industrial Time Series Data. In *2023 14th International Conference on Computing Communication and Networking Technologies*. 1–7. doi:10.1109/ICCCNT56998.2023.10307775
[21] Yue Lu, Thirumalai Vinjamoor Akhil Srinivas, Takaaki Nakamura, Makoto Imamura, and Eamonn J. Keogh. 2023. Matrix Profile XXX: MADRID: A Hyper-Anytime and Parameter-Free Algorithm to Find Time Series Anomalies of all Lengths. In *IEEE International Conference on Data Mining, ICDM 2023, Shanghai, China, December 1-4, 2023*, Guihai Chen, Latifur Khan, Xiaofeng Gao, Meikang Qiu, Witold Pedrycz, and Xindong Wu (Eds.). IEEE, 1199–1204. doi:10.1109/ICDM58522.2023.00148
[22] Yue Lu et. al. 2023. DAMP: accurate time series anomaly detection on trillions of datapoints and ultra-fast arriving data streams. *Data Mining and Knowledge Discovery* 37 (01 2023), 1–43. doi:10.1007/s10618-022-00911-7
[23] Aditya P. Mathur et. al. 2016. SWaT: a water treatment testbed for research and training on ICS security. In *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*. IEEE, 31–36.
[24] Mohsin Munir et. al. 2019. DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series. *IEEE Access* 7 (2019), 1991–2005. doi:10.1109/ACCESS.2018.2886457
[25] Inc. Netflix. 2024. Metaflow; https://metaflow.org/. https://metaflow.org/
[26] Zijian Niu et. al. 2020. LSTM-based VAE-GAN for time-series anomaly detection. *Sensors* 20, 13 (2020), 3738.
[27] Hu sheng Wu. 2016. A survey of research on anomaly detection for time series. *2016 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)* (2016), 426–431.
[28] Florian Stoffel et. al. 2013. Finding anomalies in time-series using visual correlation for interactive root cause analysis. In *Proceedings of the Tenth Workshop on Visualization for Cyber Security* (Atlanta, Georgia, USA) *(VizSec '13)*. Association for Computing Machinery, New York, NY, USA, 65–72. doi:10.1145/2517957.2517966
[29] Mingyan Teng. 2010. Anomaly detection on time series. In *2010 IEEE International Conference on Progress in Informatics and Computing*, Vol. 1. 603–608. doi:10.1109/PIC.2010.5687485
[30] Ke Wan, Yi Liang, and Susik Yoon. 2024. Online Drift Detection with Maximum Concept Discrepancy. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Barcelona, Spain) *(KDD '24)*. Association for Computing Machinery, New York, NY, USA, 2924–2935. doi:10.1145/3637528.3672016
[31] Dongjie Wang, Zhengzhang Chen, Jingchao Ni, Liang Tong, Zheng Wang, Yanjie Fu, and Haifeng Chen. 2023. Interdependent Causal Networks for Root Cause Localization. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Long Beach, CA, USA) *(KDD '23)*. Association for Computing Machinery, New York, NY, USA, 5051–5060. doi:10.1145/3580305.3599849
[32] Chin-Chia Michael Yeh et. al. 2016. Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*. 1317–1322. doi:10.1109/ICDM.2016.0179
[33] Chuxu Zhang et. al. 2019. A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data. *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (Jul. 2019), 1409–1416. doi:10.1609/aaai.v33i01.33011409

8 Appendix

8.1 Detailed Analysis and Discussion

In this section, we present some more case studies for a better understanding of failures predicted by our algorithms and the corresponding root cause analysis. For each case study, we first present the anomaly score plot for a specific region that is observed to have anomalies. In the anomaly score plot, we use a horizontal red dashed line to mark the threshold level; anomaly score above this line triggers an alert and is flagged as an anomaly. We use a blue dashed vertical line to indicate the timestamp of an actual cobble event in the mill.

8.2 Recent Prevented Failures

8.2.1 *Offline Positive Case Studies*.

CASE STUDY - 1

Figure 10 shows the analysis of the failure alert that started in the mill at 2024-04-15 13:25:09 for the **profile 12mm** involving CVAH L1 and Pinchroll L2 regions. The anomaly score plot for both regions indicates consecutive red spikes at small intervals leading to a failure alert. The following are the key signals in both regions for the anomalous behavior identified by our system:

The small continuous spikes in the E35_CVAH_Actual torque signal from timestamp 2024-04-15 13:25:09 almost completely subsided as highlighted by the red zone in the Figure 10b behaved as an anomalous signal but then got resolved on its own. The L1_PR4_Act Torque remains close to 0 for a longer period than expected, as shown in Figure 10c L1_PR4_Act Torque signal. This delay can be attributed to a delay in billet rolling and is not considered a failure. A downward and upward trend is evident in the L2_PR3_Act Speed signal, but the trend is brief and quickly returns to its original shape. Hence, the signal is identified as anomalous but not indicative of a failure.

Table 6: Our Anomaly Predictions and Actions in January 2025

Date	Time	Region	Signal	Event	Action Taken By Operator
11-01-2025	13:09:41	Fast Finishing Area	L1_PR2_Act_Speed	Cobble Fast Finishing Block	Alert received, but root cause was unknown at the time. Unable to prevent failure. Later found flapper broken in Line 1.
11-01-2025	13:09:41		L1_PR3_Act_Speed		
11-01-2025	13:09:41		FFB_L1 Motor Current		
11-01-2025	13:09:41		FFB_Line 1 Loop Height		
14-01-2025	11:24:00	Finishing Area	Stand 18 Downstream HMD	HMD Sensor Fault	**Failure prevented**: During mill checkup, operator found scale deposition on Stand 18, causing flickering. Cleaned during inspection.
14-01-2025	16:39:46	Intermediate Area	Stand 10 Act Elongation	Out of Range	**Failure prevented**: We observed Stand 10 elongation exceeding the highest range (1.295 → 1.302). Operator was informed and adjusted the metal, bringing it back within the acceptable range.
14-01-2025	16:39:46		Stand 7 Act Elongation		
14-01-2025	16:39:46		Stand 11 Act Elongation		
14-01-2025	16:39:46		Stand 11 RPM		
15-01-2025	11:45:00	Finishing Area	Stand 15 Downstream HMD	HMD Sensor Fault	**Failure prevented**: Flickering observed at Stand 15 downstream HMD. operator in-charge was informed and discussed with the electrical team.
18-01-2025	20:19:00	Finishing Area	Stand 18 Downstream HMD	HMD Sensor Fault	**Failure prevented**: We observed flickering. After physical inspection, scale deposition was found, which could lead to tracking issues or misinterpretation in the system.
21-01-2025	16:01:00	Finishing Area	Stand 14 Act Elongation	Out of Range	**Failure prevented**: We observed elongation reaching the lowest range and manually adjusted it after pressing metal in Stand 14 by the operator.
21-01-2025	16:01:00		Stand 15 RPM		
21-01-2025	16:01:00		Stand 16 RPM		
21-01-2025	16:01:00		Stand 13 Line Speed		
21-01-2025	16:24:12	Finishing Area	Loop Height Stand 13-14	Down Trend in Top Signals	**Failure prevented**: Due to metal pressing in Stand 14, looper height trended downward.
21-01-2025	16:24:12		Loop Height Stand 14-15		
21-01-2025	16:24:12		Stand 14 Act Elongation		
21-01-2025	16:24:12		Stand 15 Current		
22-01-2025	10:39:33	Fast Finishing Area	FFB_L1 Line Speed	Trend in Top Signals	**Failure prevented**: Plan Job taken and loop height adjustment.
22-01-2025	10:39:33		FFB_L1 RPM Act		
22-01-2025	10:39:33		FFB_L1 Loop Height		
22-01-2025	10:39:33		L1_PR3 Act Position		
30-01-2025	13:06:00	Finishing Area	Stand 18 Downstream HMD	HMD Sensor Fault	**Failure prevented**: Found flickering due to scale deposition at Stand 18 Exit Line 1 during mill checkup. Issue was addressed.
30-01-2025	13:06:00	Intermediate Area	Stand 11 Downstream HMD	HMD Sensor Fault	**Failure prevented**: Scale deposition caused flickering. Identified and cleaned during inspection.
31-01-2025	11:27:00	Intermediate Area	Stand 12 Act Elongation	Out of Range	**Failure prevented**: Operator observed elongation exceeding the highest range (1.122 vs. 1.115) and manually adjusted it to prevent further issues.
05-02-2025	12:50:00	Roughing Area	Stand 6 Act Elongation	Out of Range	**Failure prevented**: Reached lowest range (1.275). Operator identified front-end bending in Stand 5 and stock slightly opening but took no immediate action.
10-02-2025	13:54:40	Finishing Stand	Stand 15 Act Elongation	Out of Range	**Failure prevented**: Operator informed CP2, who verified signals, stopped the mill, and checked the looper. Found scanner malfunction and cleaned it.
10-02-2025	13:54:40		Loop Height Stand 14-15		
10-02-2025	13:54:40	Intermediate Stand	Stand 11 Downstream HMD	Out of Range and HMD Sensor Fault	**Failure prevented**: Identified looper scanner malfunction at the same time as HMD flickering. Scanner cleaned.

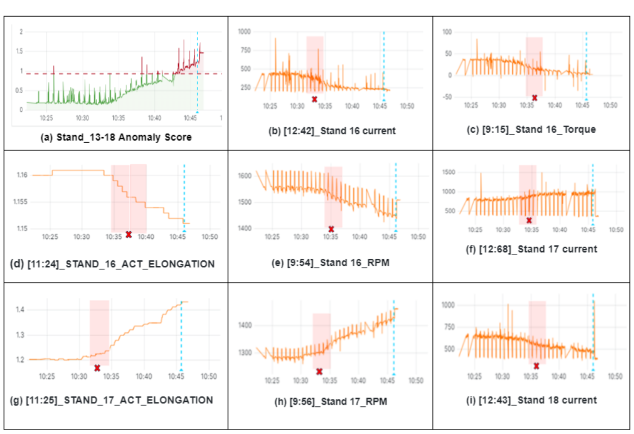


Figure 10: (a) Anomaly score plot for Stand_13-18 region.

CASE STUDY - 2

Figure 11 shows the analysis of the **overshoot** that occurred in the mill in between 2024-03-05 05:56:37 and 2024-03-05 06:05:12 for the **profile 16mm**. Our analysis reveals that the anomaly occurred at the CVR L1 and Pinchroll L1 region as shown in Figure 11.a,c. The plot indicates an anomaly providing an early warning with a **lead time of 1.22 minutes and 3.15 minutes** for CVR L1 and

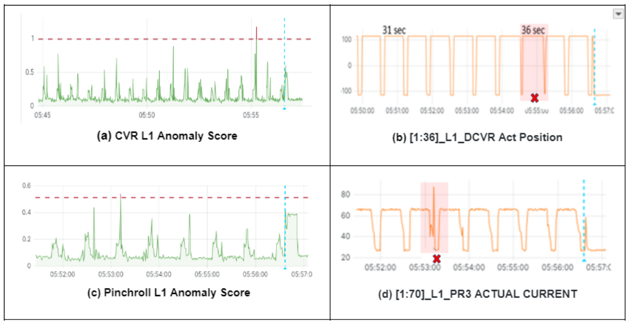


Figure 11: (a) Anomaly Score plot for CVR L1 region. (b) The highlighted red zone shows an increase in width confirming a true positive case. (c) Anomaly Score plot for Pinchroll L1 region. (d) Highlighted red spike shows a change in motif confirming a true positive case

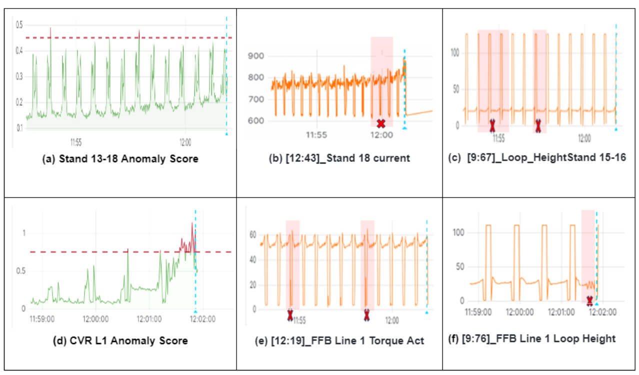


Figure 12: (a) Anomaly Score plot for Stand 13-18 region. Red spikes above threshold show anomalies. (b) Highlighted red zone and red cross mark indicates continous uptrend confirming true positive case. (c) Red cross marks show change in motifs confirming the true positive case. (d) Anomaly Score plot for CVR L1 region. Red spikes above threshold confirming true positive case. (e) Red cross marks showing change in motif leading to true positive case. (f) Red highlighted zone and cross mark show vibrations confirming failure.

Pinchroll L1 respectively. The following are the key signals for the abnormal behavior identified by our system .

The signal L1_DCVR Act Position is responsible for chopping the metal billets from head and tail, ensuring the metal is cut to the expected length. As illustrated in Figure 11b, the corresponding signal typically reaches a value of 100 for 31 seconds for each metal occurrence. However, for the highlighted section between 2024-03-05 05:54:40 and 2024-03-05 05:55:16, the signal remains at 100 for 36 seconds. This deviation in time span results in chopping the metal to an unwanted length, leading to an overshoot incident. Similarly, a huge abnormal spike is observed at 2024-03-05 05:53:10 in the L1_PR3 ACTUAL CURRENT signal contributing to the overshoot.

CASE STUDY - 3

Figure 12 shows the analysis of the **guide loose** at Stand 13-18 exit that occurred in the mill between 2024-03-06 12:01:52 and 2024-03-06 12:13:48 for the **profile 16mm** involving Stand 13-18 and CVR L1 regions. The anomaly score plot indicates an anomaly providing an early warning with a **lead time of 7.12 minutes and 1.16 minutes** for stand 13-18 and cvr l1 respectively. The following are the key signals in both regions for the abnormal behavior identified by our system:

Region Stand 13-18 Analysis: An **uptrend** is observed in Stand 18 Current shown in Figure 12b, indicating a continuous increase in current from 780 to 880 starting at 2024-03-06 11:58:32 resulting in a deviation of **12.8%**. Signal Loop Height Stand 15-16 is expected to lift the running metal from the lower end and return to its original position once the metal passes. However, at timestamps 11:54:12 and 11:57:32, the looper failed to return to its original position, resulting in a failure as shown in Figure 12c.

Region CVR Line-1 Analysis: Significant spikes were detected in Signal FFB L1 Torque Act at timestamps 2024-03-06 11:52:27 and 2024-03-06 11:54:27 as shown in Figure 13. **Vibrations** were observed in Signal FFB L1 Loop Height starting from 2024-03-06 12:01:40 until the point of failure.

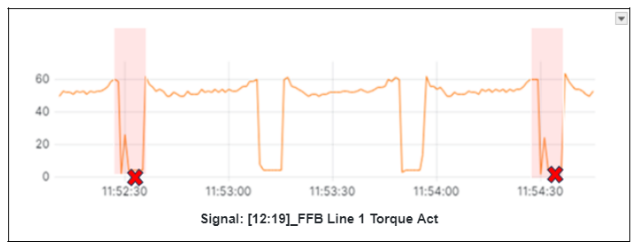


Figure 13: Zoomed Image for Signal: FFB L1 Torque Act

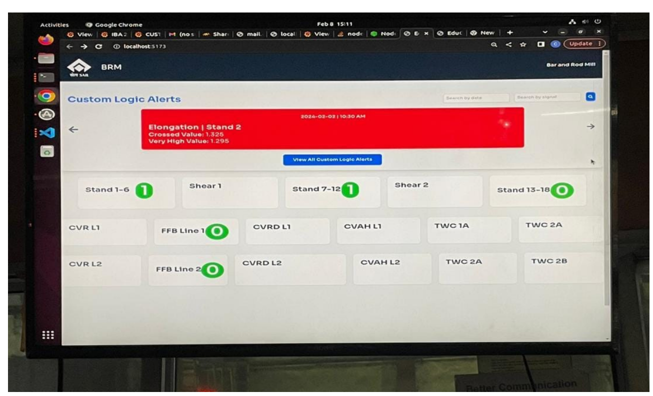


Figure 14: Counter Based Live Dashboard

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009