

Standing Tall: Sim to Real Fall Classification and Lead Time Prediction for Bipedal Robots

Gokul Prabhakaran, Jessy W. Grizzle, and M. Eva Mungai

Abstract—This paper extends a previously proposed fall prediction algorithm to a real-time (online) setting, with implementations in both hardware and simulation. The system is validated on the full-sized bipedal robot Digit, where the real-time version achieves performance comparable to the offline implementation while maintaining a zero false positive rate, an average lead time (defined as the difference between the true and predicted fall time) of 1.1s (well above the required minimum of 0.2s), and a maximum lead time error of just 0.03s. It also achieves a high recovery rate of 0.97, demonstrating its effectiveness in real-world deployment. In addition to the real-time implementation, this work identifies key limitations of the original algorithm, particularly under omnidirectional faults, and introduces a fine-tuned strategy to improve robustness. The enhanced algorithm shows measurable improvements across all evaluated metrics, including a 0.05 reduction in average false positive rate and a 1.19s decrease in the maximum error of the average predicted lead time.

I. INTRODUCTION

Humanoid robots have attracted significant attention over the past year, with a surge in companies developing general-purpose humanoids for home and factory applications. Although some humanoids have already been deployed in home and factory environments [1] and their performance in the recent World Humanoid Robot Games in Beijing showcased their increased ability to handle complex situations [2], their widespread adaptation, especially in unstructured or unknown environments, remains limited. A key challenge in broadly deploying humanoid robots in these environments is their high-dimensional, hybrid nature, which, paired with their smaller support polygons, makes it particularly difficult to maintain stable motion, especially in unstructured terrains in the presence of disturbances and uncertainties. Although a well-designed controller can mitigate the effects of disturbances and uncertainties, the unpredictable nature of real-world operations makes falls inevitable. This is especially critical in applications involving human-robot interaction where safety is paramount. As a result, it is imperative to implement fall prediction algorithms alongside robust controllers. Therefore, this paper extends the previously proposed fall prediction framework [3] to real-time operations, analyzes the algorithm’s robustness to omnidirectional disturbances during the task of standing, and offers pathways to expand its capabilities. Standing was chosen in [3] as a precursor to more dynamic behaviors.

{Robotics, Robotics, Robotics} Department, University of Michigan, Ann Arbor, MI, USA. {gok,grizzle, mungam}@umich.edu

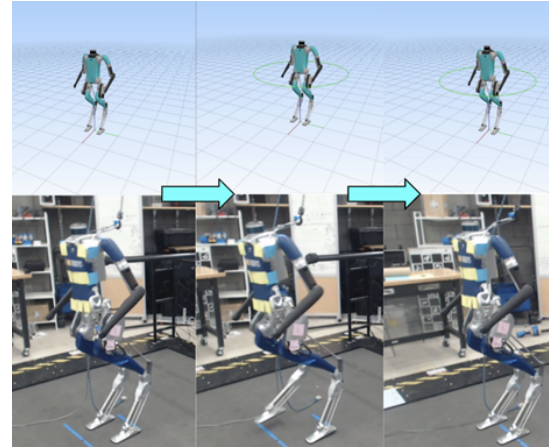


Fig. 1: Real-time detection of a critical fault introduced between the first and second panels.

A. Background

The overarching objective of a fall prediction algorithm is to accurately predict falls with sufficient time to implement a recovery strategy. As discussed in [3] and [4], faults can be a precursor to a bipedal robot falling. Faults are defined as unforeseen deviations in one or more operational variables. Based on their temporal behavior, they can be categorized as either **abrupt**, **incipient**, or **intermittent**. Abrupt faults are sudden, incipient faults develop gradually, and intermittent faults occur sporadically. Intermittent faults are comprised of abrupt and incipient faults [5]. In this paper, we refer to faults that lead to a fall as **critical faults**. The term **lead time** is used to describe the difference in time between the prediction of a fall and its actual occurrence, while **response time** is the difference in time between when a critical fault is introduced and when a fall is predicted. A trajectory or state is considered **unsafe** if it leads to a fall, and **safe** if it does not.

B. Literature Review

While numerous fall prediction algorithms have been proposed for bipedal robots—ranging from static threshold-based methods to machine learning approaches [6], [7]—the implementation of algorithms capable of real-time prediction of falls caused by abrupt, incipient, or intermittent faults in full-sized humanoid robots remains limited. Moreover, the prediction of lead time is also limited. Given the trade-off between false positive rate and lead time, and the dependence of the minimum required lead time on both the recovery

algorithm and hardware platform, lead time prediction can significantly inform the selection and timing of an appropriate recovery strategy.

Fall prediction algorithms in the literature can be categorized based on several factors: the type of faults they address (abrupt, incipient, or intermittent), the size of the robot used, the experimental setting (offline with simulation data, offline with hardware data, online in simulation, or online in hardware), and whether they estimate lead time. In this section, we briefly go over the fall prediction algorithms in the literature based on these categorizations.

The fall prediction algorithm proposed in [8] uses sensor fusion to predict falls in real-time, across various terrains and during dynamic motions such as squatting, in simulation and hardware for the 1.92m WALK-MAN humanoid [9]. However, the analysis is limited to falls caused by abrupt faults. Similarly, the SVM-based algorithm from [10] is tested online in hardware and simulation for the 1.6m BHR-6 robot during walking and standing tasks, and only considers abrupt faults.

In contrast, while [11], [12], [13], and [14] do not explicitly address all three faults, their analyses are based on scenarios that may encompass all temporal fault categories—for instance, varying the robot’s initial state [11], robot-to-robot interactions [12], walking on uneven terrain [13], and motor failures [14]. Despite the potential inclusion of all three faults, the analyses in [11] and [13] are limited to offline simulations of the 1.43m planar bipedal robot Rabbit [15] and the 90cm NUGUS humanoid robot, respectively. While [12] conduct their analysis online in both hardware and simulation, it is limited to the 0.57m tall Nao H25 robot [16], [17]. On the other hand, [14] uses both hardware and simulation data to evaluate their algorithm on the 1.82m Q1 humanoid robot and a Cassie bipedal robot—originally 1m tall [18]—fitted with a torso, however, their analysis does not include online hardware experiments.

[3] proposes a 1D CNN model capable of predicting falls caused by abrupt, incipient, and intermittent faults, both in simulation and hardware, using the 1.6m tall bipedal robot, Digit [19]. While comparable to [14], the algorithm in [3] is, to the best of our knowledge, the only one that also estimates lead time on a full-sized humanoid. Additionally, the authors provide an open-source dataset for reproducibility [20]. However, the study is limited to offline prediction in a planar standing task, where faults are simulated by applying external forces to the robot’s back.

C. Objective

Given the high offline accuracy in critical fault detection and lead time prediction of the fall prediction algorithm proposed by [3], along with the availability of its dataset, the objective of this paper is to extend the algorithm to a real-time (online) setting and validate its performance in both hardware and simulation on the full-sized bipedal robot Digit. Furthermore, as the algorithm is only trained and evaluated on faults introduced by horizontal forces applied to the center back of the robot’s torso, this work seeks to

identify its limitations under omnidirectional faults during a standing task and propose a fine-tuned strategy to improve robustness with minimal additional data. Note that the task of standing is chosen as a precursor to more dynamic motions. The algorithm proposed in [3] is referred to as the **nominal algorithm**.

Specifically, this paper:

- Maintains the nominal algorithm’s goal of reliably detecting imminent falls caused by abrupt, incipient, and intermittent faults, with sufficient lead time for a full-sized bipedal robot, while also accurately predicting lead time.
- Adopts the nominal fall prediction algorithm to an online setting, both in hardware and in simulation.
- Evaluates the performance of the nominal algorithm on critical faults applied in the sagittal and frontal plane at various angles and locations on the robot, using the same metrics: false positive rates, lead time, and response time.
- Proposes a strategy for robustifying the nominal algorithm against a broader set of faults.

As in [3], the robot is considered to have fallen if its center of mass is less than 10 percent of its initial height, and the minimum desired lead time is set to 0.2s.

Achieving the objective outlined above is challenging due to several factors: the masking effects of the controller, the sporadic nature of intermittent faults, the trade-off between lead time and false positive rates [4], and the crowding effect of incipient faults [21]. Furthermore, the decreasing number of available data points as the lead time increases [3], along with the diverse range of faults, complicates the prediction of falls. Implementing the algorithm online introduces additional challenges, including the need for fast inference times and the management of communication delays between different components of the fall prediction algorithm and other processes, such as the controller.

D. Contributions

The main contributions of this paper are as follows:

- Implementation of the nominal fall prediction algorithm online, both in hardware and simulation
- Extension of the open-source dataset [20] to include critical and non-critical faults, introduced by applying angled and non-angled forces at various locations in the sagittal and frontal planes of the robot
- Analysis and robustification of the nominal fall prediction algorithm against a broader range of faults.

While our study focuses on the nominal fall prediction algorithm for the bipedal robot Digit during standing, the techniques used are applicable to other algorithms, platforms, and tasks.

II. METHOD: NOMINAL FALL PREDICTION ALGORITHM, HARDWARE DESCRIPTION

This section provides a brief overview of the nominal fall prediction algorithm, along with a description of the robot hardware used.

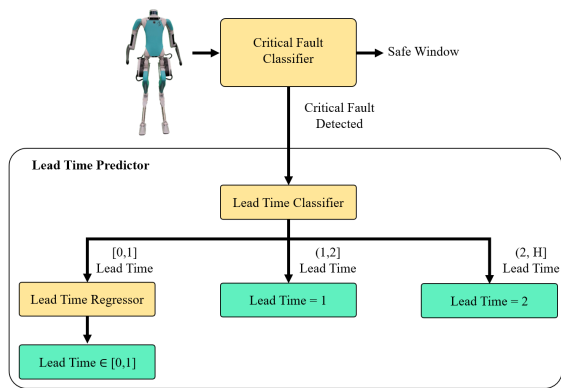


Fig. 2: The nominal fall prediction algorithm is comprised of three components: a critical fault classifier, a lead time classifier, and a lead time regressor. The green boxes depict the algorithm’s predicted lead time.

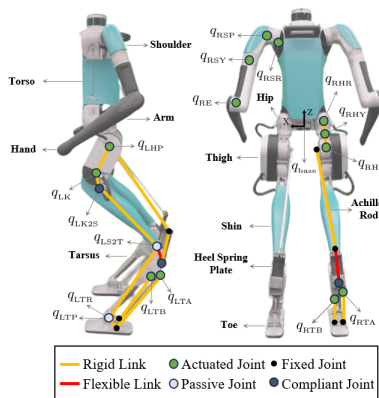


Fig. 3: Kinematics architecture of the Digit robot by Agility Robotics [3], [19].

A. Overview of Nominal Fall Prediction Algorithm Framework

The 1D CNN-based nominal fall prediction algorithm consists of three parts: a critical fault classifier, a lead time classifier, and a lead time regressor. The critical fault classifier identifies critical faults, while the lead time classifier and the regressor work together to predict the lead time. The lead time classifier divides the lead time into three intervals: $[0, 1]$, $(1, 2]$, and $(2, H]$. If the lead time is classified into the $[0, 1]$ interval, the lead time regressor calculates the lead time. Otherwise, the lead time is considered as the infimum of the interval. To prioritize the most recent data points, sliding windows with 10 data points are used. The algorithm is trained with trajectories sampled at approximately 40 Hz. Fig. 2 depicts the algorithm’s framework.

B. Hardware Description

We utilize the bipedal robot Digit [19] developed by Agility Robotics. Digit is approximately 1.6m, weighs 48 kg and has 30 degrees of freedom, 20 actuated joints, and an integrated perception system. Fig. 3 illustrates Digit’s kinematic architecture.

III. ONLINE IMPLEMENTATION AND EVALUATION OF NOMINAL FALL PREDICTION ALGORITHM

This section presents the implementation of the nominal fall prediction algorithm online in both simulation and hardware and assesses its performance.

A. Online Implementation

As in the offline implementation, we employ the use of Agility’s MuJoCo-based simulator with a standing controller designed to maintain the center of mass and zero moment point inside the support polygon [22]. This standing controller will be referred to as the **nominal controller**. To run the fall prediction algorithm online, the robot is first initialized in a standing position with the nominal controller. The robot’s state and the torques commanded by the nominal controller are then queried via ROS [23] at a rate of 400Hz. The nominal fall prediction algorithm, however, runs at 40 Hz. Therefore, features for the algorithm are computed using newly queried data every 40 Hz. Once enough features are available to form a window, model inference is performed continuously until a fall is predicted. Once a fall is predicted, the recovery controller, if one is available, is activated. The average runtime of the nominal fall prediction algorithm is approximately 0.0045s, with the total inference time of the 1D CNN models being 0.004s and the computation of the features taking 0.0005s.

B. Evaluation Metrics

The algorithm’s performance online, following the offline analysis in [3], is evaluated using lead time, false positive rate, and response time. In addition to these metrics, we introduce a new metric to assess the algorithm’s effectiveness: the **recovery rate**. The recovery rate measures the number of unsafe trajectories that are stabilized by switching to a recovery controller when a critical fault is detected. We report the recovery rate only for simulation tests due to the difficulty of replicating the same fault multiple times on the hardware. Since implementing a recovery controller falls outside the scope of this work, we utilize Agility’s base controller, which already incorporates a fall prevention strategy. We note that only abrupt and incipient faults are considered in our evaluation, as the nominal fall prediction algorithm’s effectiveness in detecting intermittent faults, as demonstrated in [3], depends on accurately identifying critical abrupt and incipient faults.

C. Experimental Setup

Abrupt and incipient faults are emulated online in simulation by applying impulsive and trapezoidal pushing forces, respectively, to regions F in Fig. 6. To assess the fall prediction algorithm’s performance online in simulation and compare it to the offline results presented in [3], we replicate the 200 abrupt and incipient tests used in the offline evaluation and apply the algorithm to these tests. The simulation is run on a machine with an Intel i9-7900x processor.

In the hardware setup, we do not have direct access to the main computer, so we interact with the robot by sending

commands and receiving state information over UDP at approximately 1 kHz. This information is processed by the nominal controller, which publishes topics via ROS also at roughly 1 kHz. The robot is connected to an Alienware m15 computer with an Intel i9-12900H processor via Ethernet. This computer runs both the fall prediction algorithm and the nominal controller.

We execute the algorithm online in hardware 14 times, emulating 8 abrupt faults and 6 incipient faults. In contrast to the simulation setup, faults are introduced by applying pushing forces using a pole to region C in Fig. 6, as region F houses the on-board computer. To capture the gradual nature of incipient faults, the pole is first positioned on the robot before applying force. To prevent potential damage during experiments, the robot is secured to a gantry using a slack cable. Note that the nominal fall prediction algorithm is transferred from hardware to simulation in a zero-shot manner - no additional data is used for training, and no modifications are made to the algorithm.

D. Performance Analysis

1) *Simulation — Offline vs Online*: As shown in Table I, the true average lead time for the online execution in simulation is 0.01s higher than the offline execution. This minor discrepancy in fault detection is primarily due to communication delays, which cause the online algorithm to occasionally run slower, averaging 40 Hz but sometimes dipping to lower rates. As a result, windows with features are sampled at intervals slightly greater or less than 40 Hz, leading to differences in the timing of critical fault detection. These timing variations also contribute to small changes in the average response time and predicted lead time, which decrease by 0.02s and increase by 0.04s, respectively, compared to the offline results. The difference between the average true lead time and predicted lead time is 0.03s for the online execution and 0.02s for the offline execution. Both online and offline evaluations report an average false positive rate of 0. Given the small differences between the online and offline results, we can conclude that the algorithm’s online performance in simulation is closely aligned with its offline results. Furthermore, the 0.97 recovery rate for the unsafe trajectories demonstrates the effectiveness of running the fall prediction algorithm online. Fig. 4 showcases the results of switching vs not switching to a recovery controller when the nominal fall prediction algorithm detects a critical fault.

Operation	Predicted Lead Time	True Lead Time	False Positive Rate	Response Time (s)	Recovery Rate
Offline [3]	1.18	1.16	0.0	0.44	N/A
Online	1.14	1.17	0.0	0.42	0.97

TABLE I: Online vs offline average simulation results of the nominal fall prediction algorithm evaluated on abrupt and incipient faults

2) *Hardware — Offline vs Online*: In the hardware tests online, 4 of the 8 abrupt faults are detected as critical faults, and 3 of the 6 incipient faults are identified as critical faults.

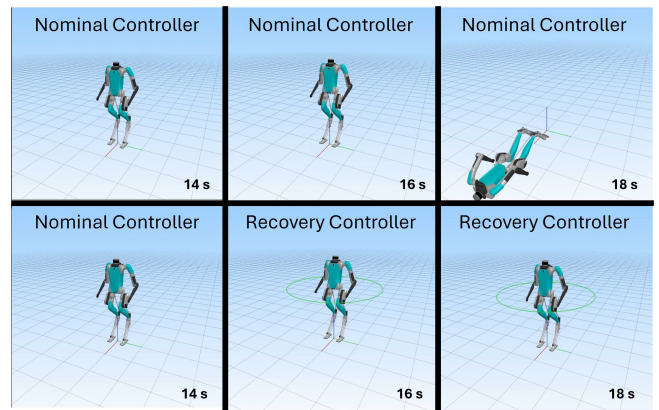


Fig. 4: Showcases the results of switching vs not switching to a recovery controller when the nominal fall prediction algorithm correctly detects a critical abrupt fault. The fault is introduced at 15s.

The average predicted lead time is 1.04s which is similar to the offline hardware results reported in [3]. Note that we only report the average predicted lead time, as measuring the force application time and hence the response time online without force sensors proved challenging. Additionally, false-positive rates are unknown since the recovery controller is triggered whenever falls are detected. However, the similarity in average predicted lead times between the online and offline implementations indicates that the nominal fall prediction algorithm’s performance on hardware remains consistent with its offline results and the simulation results. Fig. 5 demonstrates the fall prediction algorithm detecting a critical fault online in hardware.

3) *Simulation vs Hardware — Online*: The average predicted lead time achieved online in hardware is 0.1s lower than the results obtained in the online simulation. This deviation can be attributed to the smaller number of hardware online tests, which do not encompass as broad a range of critical faults. In contrast, the average runtime of the fall prediction algorithm for the hardware setup is 0.0061s, which is comparable to the algorithm’s runtime in simulation. The 0.0016s difference in runtime is negligible, given the algorithm’s sampling rate of 0.025s (40Hz). Furthermore, the nominal controller publishes at about 1kHz in both hardware and simulation. The negligible difference in runtime for the fall prediction algorithm between simulation and hardware, the small sim-to-real gap demonstrated offline in [3], and the similarity in performance when executed offline and online in simulation, as shown in Table I, suggest that the algorithm’s performance offline in simulation reliably translates online to hardware, further validating its real-world applicability. Fig. 1 offers a consolidated view of critical fault detection in both hardware and simulation, with the recovery algorithm in the loop, as illustrated in Figs. 4 and 5.

IV. ANALYSIS OF NOMINAL FALL PREDICTION ALGORITHM ON OMNIDIRECTIONAL FAULTS

Following the impressive online performance of the nominal fall prediction algorithm, we now evaluate its robust-

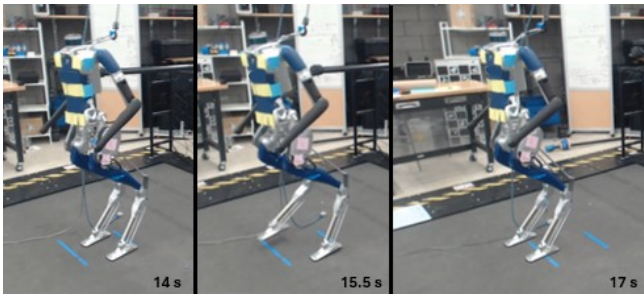


Fig. 5: The nominal fall prediction algorithm detects a critical incipient fault and switches to the recovery algorithm

ness to omnidirectional faults. The nominal fall prediction algorithm was trained on both abrupt and incipient faults introduced by applying pushing forces to region F at the back of the torso, as shown in Fig. 6, using features defined exclusively within the sagittal plane. As such, it is not expected to generalize well to faults outside of its training distribution, such as those introduced by pushing the robot on its sides [24]. However, since the training set included several trajectories in which the robot exhibited rocking motions before falling forward or backward within the sagittal plane, it remains an open question whether the algorithm can adapt to previously unseen fault conditions, such as those caused by forces applied at different locations and angles. To address this, we evaluate the performance of the three components of the fall prediction algorithm across a broad range of faults. We limit our analysis here to abrupt and incipient faults for reasons similar to those discussed in Section III. In addition, the evaluation is performed offline. However, given the negligible difference in performance between online and offline execution, as presented in Section III, we can reasonably conclude that the offline results presented here are indicative of the algorithm’s performance in real-time, online scenarios. Furthermore, the results in [3] show that the performance of the individual components directly translates to the overall performance of the algorithm. Note that the Digit open-source dataset, [20], will be extended to include the trajectories used here.

A. Data Generation of Omni-directional faults

To assess the algorithm’s robustness across a wider range of faults, abrupt and incipient faults are emulated using pushing forces as described in Section III. The test dataset is generated by applying both angled and non-angled forces at 16 distinct locations on the robot, distributed across the sagittal and frontal planes. These locations are grouped into four regions: torso back, torso front, right side, and left side, with five points on the back, five on the front, and three on each side. Fig. 6 illustrates the regions and points where the forces are applied.

For each of the 16 locations, 100 trajectories are generated for both abrupt and incipient faults, resulting in 200 trajectories per location, using both angled and non-angled pushes. For convenience, we term critical faults that are introduced by angled and non-angled forces as critical angled

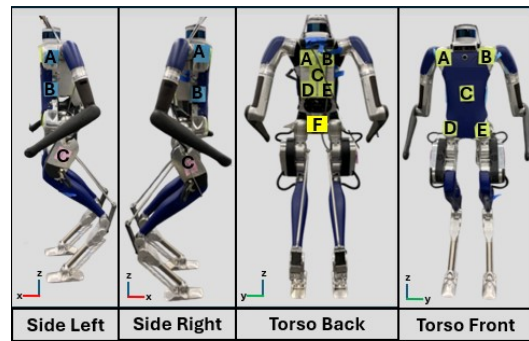


Fig. 6: Stickers in 16 positions indicating where faults were introduced. This includes (leftmost) 3 positions on the left side, (2nd from left) 3 positions on the right, (2nd from the right) 5 positions on the back of the torso, and (rightmost) 5 positions on the front of the torso.

and non-angled faults, respectively. Angled forces are applied by selecting the magnitudes of the x and y components of the force, as well as the angle of the force relative to the transverse plane of the robot. The x-component magnitude is chosen such that applying a non-angled x force causes a fall in approximately half of the trajectories (for both the torso back and front). Similarly, the y-component magnitude is set to produce a fall in about half of the trajectories when non-angled y forces are applied (for the torso sides). The angle of the force relative to the transverse plane is randomly selected from a uniform distribution within the range of $[-45^\circ, 45^\circ]$. These angles were determined through empirical testing, where the physical robot was pushed at various angles to identify the range within which it could be effectively disturbed. To ensure slight oscillations are present in the robot’s standing posture, random impulsive forces are introduced, but only windows during and after the fault introduction are used for fall prediction.

B. Nominal Critical Fault Classifier Results

In [3], the critical fault classifier demonstrated impressive results by detecting abrupt and incipient faults, achieving an average lead time of 1.52 seconds, an average response time of 0.44 seconds, and an average false positive rate of 0 when evaluated on unsafe trajectories derived in a similar manner to its training set. We refer to these results as **in-distribution results**.

When assessed on the broader range of abrupt and incipient critical faults, the classifier achieves an average response time comparable to the in-distribution results, as shown in Table II. While its average lead times decrease by 0.17s and 0.09s for non-angled and angled critical faults, respectively, the corresponding average false positive rates also rise to 0.08 for both cases. This indicates a decline in the nominal critical fault classifier’s performance. This decline in performance is particularly apparent for side pushes, as the false positive rates for critical faults applied to the sides of the robot are on average 0.13 for both angled and non-angled critical faults, whereas for critical faults introduced to the torso, the

rates are 0.02 and 0.04 for non-angled and angled forces, respectively. These results align with expectations, as the model’s training data primarily focuses on faults introduced along the sagittal plane and does not adequately capture the variability of side or angled faults.

Metrics	Nominal			Fine-Tuned		
	Non Ang	Ang.	In Dist	Non Ang	Ang	In Dist
FPR	0.08	0.08	0.0	0.0	0.003	0.0
Lead Time	1.35	1.43	1.52	1.18	1.26	1.18
Response Time	0.42	0.41	0.44	0.59	0.59	0.62

TABLE II: Average nominal [3] and fine-tuned critical fault classifier results evaluated on non-angled and angled abrupt and incipient faults introduced by forces applied at various locations on the robot. **Acronyms used:** FPR (false positive rate), Non ang (non-angled), ang (angled), and in dist (in-distribution or tests from [3]).

C. Nominal Lead Time Classifier

When evaluated on in-distribution critical faults, the lead time classifier achieved accuracies of 1.0, 0.95, and 0.74 for the $[0, 1]$, $(1, 2]$, and $(2, H]$ lead time intervals, respectively. However, similar to the critical fault classifier, the lead time classifier performs worse when evaluated on a broader range of critical faults. For non-angled critical faults, the classifier, as shown in Table III, achieves accuracies of 0.82, 0.57, and 0.61 for the $[0, 1]$, $(1, 2]$, and $(2, H]$ intervals, respectively. For angled critical faults, the classifier achieves accuracies of 0.84, 0.63, and 0.55 for the same intervals. As shown in Fig. 7, this decline in performance is particularly pronounced for side critical faults. These results align with those of the critical fault classifier and are consistent with expectations, especially considering that the lead time classifier achieves its highest accuracy within the $[0, 1]$ interval. Recall that, as demonstrated in [3], the number of available data points decreases exponentially as lead time increases. The asymmetry observed in Fig. 7 between the left and right sides can be attributed to randomness in data generation, leading to varying force magnitudes even within the same range. When critical faults from one side are emulated on the other, the lead time classifier produces similar results.

Intervals	Nominal			Fine-Tuned		
	Non Ang	Ang	In Dist	Non Ang	Ang	In Dist
$[0, 1]$	0.82	0.84	1.0	0.96	0.95	0.98
$(1, 2]$	0.57	0.63	0.95	0.84	0.80	0.82
$(2, H]$	0.61	0.55	0.74	0.30	0.27	0.44

TABLE III: Average nominal [3] and fine-tuned lead time classifier results evaluated on non-angled and angled abrupt and incipient faults introduced by forces applied at various locations on the robot. **Acronyms used:** Non ang (non-angled), ang (angled), and in dist (in-distribution or tests from [3]).

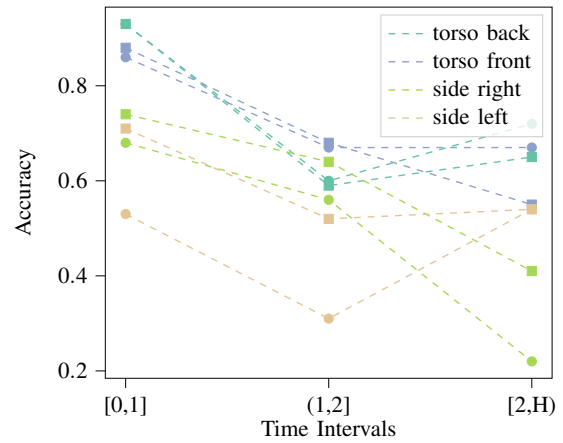


Fig. 7: Average lead time classifier accuracies when evaluated on non-angled and angled critical faults across various locations. The circles and squares indicate the non-angled and angled critical faults, respectively.

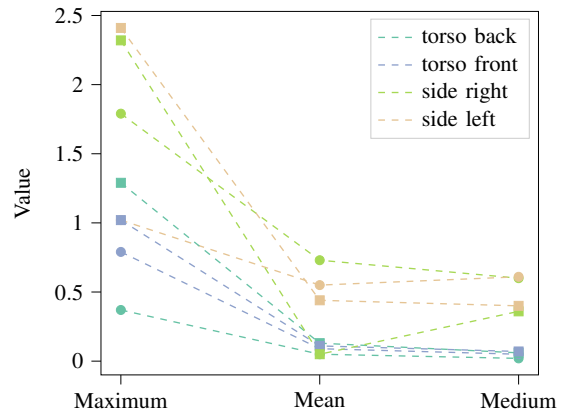


Fig. 8: Average lead time regressor results when evaluated on non-angled and angled critical faults across various locations. The circles and squares indicate the non-angled and angled critical faults, respectively.

D. Nominal Lead Time Regressor

When evaluated on in-distribution critical faults within the $[0, 1]$ lead time range, the lead time regressor shows minimal error, with a maximum difference of 0.09s between predicted and actual lead time, and a mean and median difference of just 0.01s [3]. However, similar to both the critical fault classifier and the lead time classifier, the regressor performs significantly worse on non-angled and angled critical faults. For non-angled critical faults, the maximum lead time prediction error increases to 1.79s, while the mean and median errors rise to 0.23s and 0.07s, respectively, as shown in Table IV. For critical angled faults, the maximum error increases to 2.41s, with mean and median errors of 0.25s and 0.11s, respectively. Given that the regressor’s objective is to predict lead time within the $[0, 1]$ interval, these error increases are unacceptable. Despite these significant errors, as shown in Fig. 8, the regressor performs better for the torso back non-angled critical faults compared to the other regions and faults.

Differences	Nominal			Fine-Tuned		
	Non Ang	Ang	In Dist	Non Ang	Ang	In Dist
Max	1.79	2.41	0.09	0.23	0.32	0.18
Mean	0.23	0.25	0.01	0.03	0.03	0.04
Median	0.07	0.11	0.01	0.02	0.02	0.03

TABLE IV: Nominal and fine-tuned lead time regressor results evaluated on non-angled and angled abrupt and incipient faults introduced by forces applied at various locations on the robot. **Acronyms used:** Non ang (non-angled), ang (angled), and in dist (in-distribution or tests from [3]).

V. ROBUSTIFYING NOMINAL ALGORITHM TO OMNIDIRECTIONAL FAULTS

In response to the nominal fall prediction algorithm's suboptimal performance under omnidirectional faults, particularly to faults introduced by forces applied to the frontal plane, this section explores strategies to enhance its robustness against such faults. A common challenge with learning-based methods is their reliance on large datasets. To mitigate this while leveraging the existing algorithm, we focus on fine-tuning the existing critical fault classifier, lead time classifier, and lead time regressor. Our goal is to improve performance while minimizing the need for additional training data, thus striking a balance between performance and data efficiency.

Fine-tuning is a widely used machine learning technique that adapts a pretrained model to new data by leveraging previously learned features while adjusting to the specific requirements of the new task. This approach is both efficient and effective, especially when the new data closely resembles the original data used for pretraining. Compared to training a model from scratch, fine-tuning enables rapid adaptation to new tasks with significantly less computational effort [25].

For our task of standing, it is reasonable to assume that detecting faults from omnidirectional pushes shares enough similarity with the nominal algorithm's training scenario, which involves detecting faults from pushes applied to the center-back of the robot's torso. In fact, when we compared the performance of the fine-tuned model to that of a model trained from scratch, we found their performance to be comparable.

A. Fine-Tuned Critical Fault Classifier Results

Given the nominal critical fault classifier's poor performance on unsafe side trajectories, as discussed in Section IV-B, it is fine-tuned only with 200 non-angled abrupt and 200 non-angled incipient faults, each applied to region B of side left and right, as shown in Fig. 6. These regions are selected because they represent the "midpoint" of the possible pushing areas on the sides.

Despite being fine-tuned exclusively with non-angled data from region B, the fine-tuned critical fault classifier performs significantly better than the nominal critical fault classifier across all metrics, even on unseen angled-faults. Specifically, as shown in Table II, decreases significantly - from 0.08 to 0.0 for non-angled faults, and from 0.08 to 0.003 for

angled faults. As expected, due to the direct relationship between false positive rate and lead time, the average lead time decreases from 2.54s to 1.18s and from 2.41s to 1.26s for non-angled and angled critical faults, respectively. On the other hand, the average response time increases by 0.16s and 0.17s, respectively, for non-angled and angled critical faults.

When evaluated on in-distribution data, the fine-tuned model achieves a 0.0 average false positive rate, similar to the nominal model. However, the average lead time and response time decrease and increase by 0.34s and 0.18s, respectively. Despite this, the fine-tuned critical fault classifier model achieves lead times significantly above 0.2s and nearly 0 false positive rates for all critical faults, meeting the objective outlined in [3] of detecting critical faults with sufficient lead time.

B. Fine-Tuned Lead Time Classifier Results

The lead time classifier, fine-tuned using the limited set of non-angled side data as the critical fault classifier, also shows notable improvements for all faults. As shown in Table III, its accuracy increases from 0.82 to 0.96 and from 0.84 to 0.95 for the non-angled and angled critical faults in the $[0, 1]$ interval, respectively. The accuracy also increases from 0.57 to 0.84 and from 0.63 to 0.80 for the $(1, 2]$ interval. However, the accuracy decreases from 0.61 to 0.30 and from 0.55 to 0.27, respectively, for the $(2, H]$. This decline is expected and acceptable given that the number of data points decreases exponentially as lead time increases [3].

When evaluated on in-distribution data, the fine-tuned model's performance slightly decreases. The accuracies drop by 0.02, 0.13, and 0.3 for the $[0, 1]$, $(1, 2]$, and $(2, H]$ intervals, respectively. While the drop in performance for the $(2, H]$ interval might seem significant, it is justifiable due to the previously mentioned exponential relationship between lead time and data availability. Furthermore, given the long lead times in this interval, exceeding the 0.2s threshold, it is reasonable to disregard the classifier's output for this region, as it will not have severe consequences before transitioning to intervals with more accurate predictions.

C. Fine-Tuned Lead Time Regressor Results

Fine-tuning the lead time regressor using the same data as the critical fault and lead time classifier results in only acceptable performance for non-angled critical faults. Therefore, the lead time regressor is instead fine-tuned with 200 angled abrupt faults and 200 angled incipient faults, applied to the following regions: side left B and C, side right B and C, and torso front C. When evaluated on the in-distribution data, the maximum, mean, and median differences, as shown in Table IV, increase by 0.09s, 0.03s, and 0.02s, respectively, compared to the nominal model. However, when assessed on non-angled critical faults, the maximum, mean, and median differences decrease from 1.79s to 0.23s, from 0.23s to 0.03s, and from 0.07s to 0.02s, respectively, for non-angled critical faults. Similarly, for angled critical faults, the differences decrease from 2.41s to 0.32s, from 0.25s to 0.03s, and from 0.08s to 0.02s. Despite the substantial improvement in

maximum difference for both angled and non-angled critical faults, the maximum difference of 0.32s achieved for angled critical faults may appear significant within the $[0, 1]$ interval. However, the classifier attains this maximum difference in only 1 percent of the unsafe trajectories. This, along with the notable decrease in maximum difference, underscores the lead time regressor fine-tuned model's capability to perform well across a wide range of critical faults.

VI. CONCLUSION

In this paper, we extended the nominal fall prediction algorithm from [3] to a real-time setting, implementing it in both hardware and simulation. The online implementation yields results similar to the offline version when tested on comparable data. Additionally, in simulation, it achieves an average recovery rate of 0.97. The average runtime in simulation and hardware is approximately 0.0045s and 0.0061s, respectively. The difference of 0.0061s is deemed negligible given the algorithm's sampling rate of 0.025s (40Hz). Given the consistent performance across online and offline simulations, as well as between online and offline hardware tests, we conclude that offline results are reliable indicators of online performance. Furthermore, the similar results observed between offline simulation and hardware implementations [3], along with the comparable runtimes between online hardware and simulation implementations, suggest that the nominal fall prediction algorithm's simulation results can effectively predict its hardware outcomes.

To evaluate the nominal fall prediction algorithm's adaptability to new scenarios, we assessed its robustness against a broader range of faults. This evaluation was conducted for each of its three components: the critical fault classifier, the lead time classifier, and the lead time regressor. All three components exhibited significantly degraded performance under these conditions, particularly for faults in the frontal plane, highlighting key limitations of the nominal algorithm.

To address these limitations while minimizing the need for additional data, we fine-tuned the three components with minimal additional data. The fine-tuned components significantly outperformed the nominal components across all evaluation metrics. Furthermore, the fine-tuned algorithm achieved an average lead time greater than 0.2s, with nearly zero average false positive rate, and a maximum average mean difference of 0.03s for the lead time. These results highlight the fine-tuned fall prediction algorithm's suitability for real-time and offline fall and lead time prediction in bipedal robots, even across a wide variety of faults.

REFERENCES

- [1] Morgan Stanley, "The humanoid 100: Mapping the humanoid robot value chain," https://advisor.morganstanley.com/john.howard/documents/field/jjo/john-howard/The_Humanoid_100_-_Mapping_the_Humanoid_Robot_Value_Chain.pdf, 2025, accessed: 09.13.2025.
- [2] Y. Zhuang, "The athletes at china's robot games fell down a lot," <https://www.nytimes.com/2025/08/18/world/asia/china-humanoid-robot-games.html>, 2025, accessed: 09.13.2025.
- [3] M. E. Mungai, G. Prabhakaran, and J. W. Grizzle, "Fall prediction for bipedal robots: The standing phase," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 13 135–13 141.
- [4] M. E. Mungai and J. Grizzle, "Optimizing lead time in fall detection for a planar bipedal robot," in *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2023, pp. 1–7.
- [5] R. Isermann, "Process fault detection based on modeling and estimation methods," *IFAC Proceedings Volumes*, vol. 15, no. 4, pp. 7–30, 1982, 6th IFAC Symposium on Identification and System Parameter Estimation, Washington USA, 7-11 June. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1474667017629598>
- [6] R. Subburaman, D. Kanoulas, N. Tsagarakis, and J. Lee, "A survey on control of humanoid fall over," *Robotics and Autonomous Systems*, vol. 166, p. 104443, 2023.
- [7] B. Zhang and M. Jia, "Fall recovery strategies in humanoid robots: A brief survey," in *2023 8th International Conference on Control, Robotics and Cybernetics (CRC)*. IEEE, 2024, pp. 68–72.
- [8] R. Subburaman, D. Kanoulas, L. Muratore, N. G. Tsagarakis, and J. Lee, "Human inspired fall prediction method for humanoid robots," *Robotics and Autonomous Systems*, vol. 121, p. 103257, 2019.
- [9] N. G. Tsagarakis, D. G. Caldwell, F. Negrello, W. Choi, L. Baccelliere, V.-G. Loc, J. Noorden, L. Muratore, A. Margan, A. Cardellino, *et al.*, "Walk-man: A high-performance humanoid platform for realistic environments," *Journal of Field Robotics*, vol. 34, no. 7, pp. 1225–1259, 2017.
- [10] T. Wu, Z. Yu, X. Chen, C. Dong, Z. Gao, and Q. Huang, "Falling prediction based on machine learning for biped robots," *Journal of Intelligent & Robotic Systems*, vol. 103, pp. 1–14, 2021.
- [11] J. Igual, P. Parik-Americano, E. C. Becman, and A. Forner-Cordero, "Time series classification for predicting biped robot step viability," *Sensors*, vol. 24, no. 22, p. 7107, 2024.
- [12] D. H. Tran, F. Hamker, and J. Nassour, "A humanoid robot learns to recover perturbation during swinging motion," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 10, pp. 3701–3712, 2020.
- [13] K. Saleh, T. O'Brien, Y. Sims, A. Mendes, and S. Chalup, "Efficient sequence model for early fall detection of humanoid robots," in *RoboCup 2024: Robot World Cup XXVII*, E. Barros, J. P. Hanna, H. Okada, and E. Torta, Eds. Cham: Springer Nature Switzerland, 2025, pp. 247–256.
- [14] C. Zhang, J. Gao, Z. Chen, S. Zhong, and H. Qiao, "Fall analysis and prediction for humanoids," *Robotics and Autonomous Systems*, vol. 190, p. 104995, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0921889025000818>
- [15] C. Chevallereau, G. Abba, Y. Aoustin, F. Plestan, E. Westervelt, C. C. De Wit, and J. Grizzle, "Rabbit: A testbed for advanced control theory," *IEEE Control Systems Magazine*, vol. 23, no. 5, pp. 57–79, 2003.
- [16] Aldebaran, "Nao," <https://corporate-internal-prod.aldebaran.com/en/nao>, accessed: 11.16.2024.
- [17] —, "H25 - construction," http://doc.aldebaran.com/2-1/family/nao_h25/dimensions_h25.html, accessed: 11.16.2024.
- [18] Y. Gong, R. Hartley, X. Da, A. Hereid, O. Harib, J.-K. Huang, and J. Grizzle, "Feedback control of a cassie bipedal robot: Walking, standing, and riding a segway," in *2019 American control conference (ACC)*. IEEE, 2019, pp. 4559–4566.
- [19] Agility Robotics, "Robots," <https://agilityrobotics.com/products>.
- [20] M. E. Mungai, G. Prabhakaran, and J. W. Grizzle, "Digit fall prediction dataset," https://github.com/UMich-BipedLab/Digit_Fall_Prediction_Dataset, 2023.
- [21] H. Safaeipour, M. Forouzanfar, and A. Casavola, "A survey and classification of incipient fault diagnosis approaches," *Journal of Process Control*, vol. 97, pp. 1–16, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0959152420303206>
- [22] G. Gibson, "Terrain-aware bipedal locomotion," Ph.D. dissertation, University of Michigan, 2023.
- [23] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng, *et al.*, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5.
- [24] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," *CoRR*, vol. abs/1610.02136, 2016. [Online]. Available: <http://arxiv.org/abs/1610.02136>
- [25] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," 2018. [Online]. Available: <https://arxiv.org/abs/1801.06146>