Vision-Informed Neural Decomposition of Motor-Current Signatures for Predictive Belt Maintenance in Conveyor-Driven Sortation Systems

Amar Sharma
Independent researcher
Systems Engineer
amar.sharma93@outlook.com

Nikhil Mallala
Independent researcher
Hardware Support Engineer
raviraja1199@gmail.com

***Abstract*—Predictive maintenance for belt-driven conveyor systems often relies on motor-current signature analysis to detect degradation. In practice, however, changes in transported item mass can produce current variations that look much like wear-induced signatures, which leads to false alarms and premature maintenance actions. This paper presents VPD-Net (Vision-Physics Decomposition Network), a method that uses a depth camera mounted above the conveyor to estimate the mass of each passing item from its bounding volume and then passes that estimate to a physics-informed neural network that separates the measured motor current into a load-explained component and a wear residual. A gated recurrent unit tracks the wear residual over time to predict remaining useful belt life. The depth estimate provides something the model otherwise does not have: a per-item load reading that helps it distinguish normal operating variation from actual degradation without depending on manifest data that may be incomplete or delayed. We evaluate VPD-Net on an anonymized industrial conveyor testbed dataset whose degradation characteristics are benchmarked against published motor-current fault studies. Across five random seeds, VPD-Net achieves a remaining-useful-life prediction RMSE of 41.3 ± 3.2 cycles, compared with 78.6 ± 5.1 for current-only baselines, 62.4 ± 4.3 for manifest-informed models, and 45.8 ± 3.6 for a physics-constrained manifest baseline that isolates the contribution of vision-based load estimation ($p < 0.05$, one-sided Wilcoxon signed-rank). False-positive maintenance alerts fall by 54% relative to the current-only approach, and the vision-assisted mass estimator reaches 94.1% accuracy within $\pm 10\%$ of true mass. These results suggest that combining direct load observation with current-signal analysis can make conveyor predictive maintenance substantially more reliable.**

***Index Terms*—predictive maintenance, motor current signature analysis, conveyor systems, depth estimation, physics-informed neural networks, remaining useful life, vision-based inspection**

I. Introduction

Belt-driven conveyors are widely used in logistics, manufacturing, and sortation. A single high-throughput sortation facility may run dozens of conveyor segments continuously at 2–3 m/s while handling thousands of items per hour. In these systems, belt wear is a primary maintenance concern. As the belt surface degrades, friction characteristics shift, tracking accuracy drops, and the chance of a sudden failure—which can stop an entire sortation line—increases [1], [2].

Motor current signature analysis (MCSA) is the standard non-invasive method for monitoring belt-driven systems. When the driven load degrades mechanically, the torque profile changes, and that change modulates the motor current waveform in detectable ways [3], [4]. MCSA has been used for conveyor belt mis-tracking [5], gearbox faults [6], and general bearing degradation [7]. In sortation applications, however, it has a well-known weakness: the transported load itself also changes current draw. A heavy parcel can produce a current spike that looks very similar to the extra drag caused by belt-surface wear. In facilities where item mass spans an order of magnitude—from 0.1 kg envelopes to 25 kg boxes—this load-driven variation can dominate the signal and bury the slower wear-driven drift underneath.

A common workaround is to use manifest data—the expected weight of each item from the warehouse management system—as a covariate. In practice, though, manifest data has limits. Weight fields may be approximate or missing for some item types, updates can lag behind physical handling by minutes, and re-inducted or exception items may not have a manifest entry at all. On a high-speed sortation line processing more than 3,000 items per hour, even a 5% manifest-miss rate would leave 150 items per hour without load correction.

This paper uses a different source for that missing load information: a depth camera mounted above the conveyor that estimates each item's mass from its observed bounding volume in real time. The estimated mass is then fed into a physics-informed neural decomposition network (VPD-Net), which separates the measured motor current into a load-explained component and a wear residual. A gated recurrent unit (GRU) tracks that residual over time to predict remaining useful belt life (RUL).

To the best of our knowledge, VPD-Net is the first method to combine depth-based load estimation with physics-constrained current decomposition for belt predictive maintenance, jointly addressing the load–wear confounding problem in conveyor

MCSA. A physics-informed loss term ties the load-explained current component to the motor torque–current relationship so that the network stays grounded in electrical machine fundamentals rather than simply fitting whichever correlations happen to reduce training loss. We evaluate the method using anonymized industrial conveyor testbed data benchmarked against published MCSA fault characteristics [5], [6], and include ablations across the vision, physics, and recurrent components.

II. Related Work

A. Motor Current Signature Analysis for Conveyors

Farhat et al. [5] showed that MCSA can diagnose conveyor belt mis-tracking using higher-order spectral features such as bicoherence and tricoherence, demonstrating that belt faults leave measurable signatures in motor current. Gelman et al. [6] extended this line of work to broader conveyor monitoring through current signals and confirmed that mechanical load conditions shape the motor current spectrum. Benbouzid [3] provides a foundational review of MCSA for induction motors, while Nandi et al. [4] survey condition monitoring and fault diagnosis for electrical motors more broadly. Together, these studies establish MCSA as a viable monitoring tool, but they do not address the decomposition problem that matters here: separating load-induced current variation from wear-induced drift when both appear at the same time.

B. Predictive Maintenance and RUL Estimation

Lee et al. [7] developed prognostics and health management frameworks for rotary machinery and helped establish the degradation-modeling paradigm used in many RUL methods. Li et al. [8] proposed deep learning approaches for remaining useful life estimation based on convolutional and recurrent architectures. Wang et al. [9] combined graph-based multi-sensor fusion with attention mechanisms for fault diagnosis. These methods generally assume that the degradation signal is directly observable, which is reasonable under constant load. In sortation systems, however, load changes from item to item, so the degradation signal is mixed with normal operating variation. That is the gap these approaches leave open.

C. Physics-Informed Neural Networks

Raissi et al. [10] introduced physics-informed neural networks (PINNs), which embed governing equations in the loss function to encourage physically consistent predictions even when data are limited. Karniadakis et al. [11] later surveyed the broader PINN landscape. In maintenance applications, Yucesan and Viana [12] used physics-informed methods for wind turbine fatigue and showed that physics constraints can reduce data requirements and improve extrapolation. We use the same general idea for a different task: not to predict degradation directly, but to split the observed current signal into load-related and wear-related components.

CONVEYOR LINE
[item on belt] $\rightarrow$ [depth camera above]

S1: DEPTH-BASED LOAD ESTIMATION
depth frame $\rightarrow$ SegNet mask $\rightarrow$ volume (Eq. 1)
volume $\times$ density_lookup(SKU) $\rightarrow \hat{m}_t$ (Eq. 2)

MOTOR CURRENT SENSOR
$i_{\text{meas}}(t)$

S2: PHYSICS-INFORMED CURRENT DECOMPOSITION (MLP, Eq. 3–8)
$[\hat{m}_t, v_{\text{belt}}, \theta, c_t] \rightarrow f_\theta \rightarrow \hat{i}_{\text{load}}(t)$
$i_{\text{wear}}(t) = i_{\text{meas}}(t) - \hat{i}_{\text{load}}(t)$
Loss: $L_{\text{decomp}} = L_{\text{phys}} + \lambda_{\text{smooth}} L_{\text{smooth}}$

S3: RECURRENT WEAR TRACKER (GRU, Eq. 9–10)
$\{i_{\text{wear}}(1), \ldots, i_{\text{wear}}(t)\} \rightarrow \widehat{\text{RUL}}_t$
$\rightarrow$ maintenance scheduling recommendation

Fig. 1. VPD-Net pipeline. A top-down depth camera estimates item mass, a physics-informed network separates load-explained current from wear residual, and a GRU tracks the residual sequence to predict remaining useful belt life.

D. Depth-Based Object Property Estimation

Depth cameras are now widely used for parcel dimensioning in logistics. Liang et al. [13] demonstrated RGB-D-based volume estimation for parcels in warehouse environments, and Ranftl et al. [14] advanced dense depth prediction with vision transformers. For our problem, the key capability is estimating item bounding volume from a top-down depth frame at conveyor speed. This is a constrained version of the broader 3D estimation problem and benefits from fixed camera geometry and a flat belt surface that serves as a known reference plane.

III. Proposed Method: VPD-Net

A. System Overview

VPD-Net has three stages (Fig. 1): (S1) depth-based load estimation from a top-down depth camera, (S2) physics-informed neural decomposition of motor current into load and wear components, and (S3) recurrent wear tracking for RUL prediction. The stages run sequentially at each conveyor cycle, defined here as the passage of one item past the sensor station.

B. Stage 1: Depth-Based Load Estimation

A top-down D435i-class RGB-D camera (640×480, 30 fps) is mounted 1.2 m above the belt surface. The raw depth image is first processed by a lightweight segmentation network—a 4-layer encoder-decoder with skip connections (3×3 convolutions, 16–32–64–32 filters, ReLU, total 0.38M parameters)—that produces a binary item mask separating the item from the belt background. This learned segmentation handles cases that simple depth-thresholding often misses, including items whose top surface lies close to belt height, items touching the belt side rails, and depth shadows near item edges. The network is trained on 2,000 labeled depth-mask pairs sampled from anonymized testbed frames and manually checked for representative edge cases. In deployment, initial pseudo-labels could be bootstrapped from simple belt-plane subtraction with a fixed depth threshold and then manually refined for edge cases before fine-tuning the segmentation network.

Given the predicted mask, the bounding volume is computed by integrating per-pixel height above the belt reference plane:

$$\hat{V}_t = \sum_{(u,v)\in \text{mask}} d(u,v) \cdot A_{\text{px}} \tag{1}$$

where $d(u,v)$ is the per-pixel height above the belt plane and A_{px} is the physical area of one pixel at belt distance, calibrated from camera intrinsics. The mass estimate is

$$\hat{m}_t = \hat{V}_t \cdot \hat{\rho}(\text{cat}_t) \tag{2}$$

where $\hat{\rho}(\text{cat}_t)$ is a learned density lookup indexed by item category. The category is a coarse grouping with 16 classes, including envelopes, small boxes, medium boxes, large boxes, polybags, and tubes, derived from the upstream barcode scan. This is not the same as manifest weight. A category such as "medium box" can span a broad weight range (0.5–10 kg) while still maintaining a relatively stable average density (~ 180 kg/m^3 for mixed consumer goods). For items without a valid barcode scan, the system falls back to a default density of 250 kg/m^3, which is the median across a typical sortation mix. The density lookup is implemented as a lightweight embedding (16×1) trained jointly with the decomposition network.

C. Stage 2: Physics-Informed Current Decomposition

For a belt-driven conveyor operating at approximately constant speed, the load-related component of motor current can be approximated from the conveyor force balance and the motor torque–current relationship. A simplified steady-state model is

$$i_{\text{load}}(t) \approx \frac{(m_t g(\mu_{\text{belt}} \cos\theta + \sin\theta) + F_{\text{drag}})\, r_p}{K_T \eta} \tag{3}$$

where K_T is the motor torque constant (Nm/A), g is gravitational acceleration, μ_{belt} is the belt-surface friction coefficient, θ is the conveyor incline, F_{drag} is the no-load drag force, r_p is the drive pulley radius, and η is drive efficiency. In a healthy belt, μ_{belt} is approximately constant (~ 0.35 for typical PVC conveyor belting). As the belt wears, μ_{belt} can drift—upward with surface roughening or, in some cases, downward with surface glazing—while F_{drag} increases because of roller resistance buildup.

Rather than estimating μ_{belt} and F_{drag} directly, VPD-Net uses a neural network f_θ to learn a mapping from observable inputs to the load-explained current:

$$\hat{i}_{\text{load}}(t) = f_\theta(\hat{m}_t, v_{\text{belt}}, \theta, c_t) \tag{4}$$

where c_t denotes optional cycle-context features such as recent belt operating state or item spacing. The wear residual is then defined as the unexplained portion of the measured current:

$$i_{\text{wear}}(t) = i_{\text{meas}}(t) - \hat{i}_{\text{load}}(t) \tag{5}$$

The decomposition network f_θ is a 3-layer MLP (128–64–32 units, ReLU, batch normalization, total 0.03M parameters). It is trained with a two-term loss:

$$L_{\text{decomp}} = L_{\text{phys}} + \lambda_{\text{smooth}} L_{\text{smooth}} \tag{6}$$

The physics loss anchors the decomposition during an initial healthy operating window, defined here as the first 10% of each belt-health sequence. In a deployment setting, that window would be established from commissioning data or a known post-replacement baseline. The loss penalizes deviations from the approximate motor-torque relationship:

$$L_{\text{phys}} = \frac{1}{N_h} \sum_{t\in\text{healthy}} \left\| \hat{i}_{\text{load}}(t) - \frac{(\hat{m}_t g(\mu_0 \cos\theta + \sin\theta) + F_0)\, r_p}{K_T \eta} \right\|_2^2, \tag{7}$$

where $\mu_0 = 0.35$ is the nominal healthy-belt friction coefficient and F_0 is the initial no-load drag force, both taken from equipment-class specifications and commissioning baseline data. This term teaches the network the approximate energy-balance relationship between conveyed load and motor current, which reduces the chance that normal load variation will be absorbed into the wear term. As the belt degrades beyond the healthy window, the growing gap between measured current and load-explained current is captured by $i_{\text{wear}}(t)$.

The smoothness loss regularizes the wear residual so that high-frequency noise is not misattributed to wear:

$$L_{\text{smooth}} = \frac{1}{N} \sum_t \| i_{\text{wear}}(t) - i_{\text{wear}}(t-1) \|_2^2 \tag{8}$$

We set $\lambda_{\text{smooth}} = 0.1$ based on validation tuning. The physics loss weight is implicitly 1.0.

D. Stage 3: Recurrent Wear Tracking for RUL Prediction

The wear residual sequence $\{i_{\text{wear}}(1), \ldots, i_{\text{wear}}(t)\}$ is passed to a single-layer GRU (64 hidden units, 0.04M parameters), which maintains a running representation of the belt's degradation trajectory:

$$\widehat{\text{RUL}}_t = g_\psi(h_t), \quad h_t = \text{GRU}(i_{\text{wear}}(t), h_{t-1}) \tag{9}$$

where g_ψ is a linear projection to a scalar RUL prediction in cycles. The GRU is trained with an asymmetric loss that penalizes late predictions more heavily than early ones:

$$\begin{aligned} L_{\text{RUL}} &= \frac{1}{N} \sum_t w_t \cdot (\widehat{\text{RUL}}_t - \text{RUL}_t)^2, \\ w_t &= \begin{cases} 1.5, & \widehat{\text{RUL}}_t > \text{RUL}_t \\ 1.0, & \text{otherwise.} \end{cases} \end{aligned} \tag{10}$$

The asymmetry reflects the practical cost structure: replacing a belt too early wastes useful life, whereas a late prediction increases the risk of a line-stopping failure. We choose a GRU instead of a simple linear trend because belt degradation often accelerates nonlinearly near the end of life, and the recurrent structure can track the rate of change in the wear residual rather than only its level. Because the GRU is a general sequence model, it can also represent non-monotonic wear trajectories—for example, a temporary reduction in friction after a belt-cleaning event—without requiring architectural changes.

IV. Deployment Considerations

VPD-Net runs alongside the existing conveyor control system and does not modify belt speed or motor parameters. Its output is a maintenance recommendation delivered to the maintenance management system; the system is advisory and sits outside

TABLE I
ANONYMIZED INDUSTRIAL TESTBED PARAMETERS USED FOR VALIDATION.

Parameter	Reported value / class
Conveyor type	Belt-driven parcel conveyor
Belt speed / length	2.5 m/s / 12 m class
Drive pulley radius	0.08 m class
Motor type	Three-phase induction motor
Motor rating / rated current	2.2 kW class / 8.1 A class
Current sensing	Non-invasive AC current sensor
Current sampling rate	1 kHz, aggregated per item cycle
Motor torque constant, K_T	1.8 Nm/A class
Drive efficiency, η	0.85
Nominal belt friction, μ_0	0.35
Depth camera	D435i-class RGB-D camera
Depth resolution / frame rate	640×480 / 30 fps
Depth camera height / noise	1.2 m class / $\sim$3 mm

TABLE II
ANONYMIZED ITEM STREAM USED FOR VALIDATION.

Item Type	Share	Mass Range	Typical Volume
Flat envelopes	22%	0.1–0.5 kg	0.001–0.004 m^3
Polybags	18%	0.2–3.0 kg	0.002–0.018 m^3
Small boxes	26%	0.5–5.0 kg	0.004–0.035 m^3
Medium boxes	24%	2.0–12.0 kg	0.025–0.110 m^3
Large boxes	8%	8.0–25.0 kg	0.080–0.220 m^3
Tubes/irregular	2%	0.3–8.0 kg	0.005–0.060 m^3

the motor-drive safety loop. The depth camera is mounted on a passive bracket and connected through an isolated Ethernet segment, while model updates are deployed through a versioned pipeline with checksum verification, consistent with IEC 62443 least-privilege principles for OT-adjacent systems [15]. On an NVIDIA Jetson Orin Nano, the full pipeline completes in under 45 ms per item. If the camera is unavailable, the system falls back to current-only monitoring (equivalent to C2 performance) and flags the degraded mode on the maintenance dashboard, so predictions continue at reduced accuracy rather than halting entirely.

V. EXPERIMENTAL EVALUATION

A. Anonymized Industrial Testbed and Dataset

The validation data were collected from an anonymized industrial belt-conveyor testbed and benchmarked against published MCSA studies [5], [6]. Table I summarizes the conveyor, motor, and sensing parameters. The no-load drag term in Eq. (3) is taken as $F_0 \approx 0$ N for the level-conveyor configuration used here, with the no-load component absorbed into the measured healthy baseline; under this convention, a 25 kg item produces a load-current increase of approximately 4.5 A, while a 0.1 kg envelope adds less than 0.02 A, confirming that load variation alone spans nearly the full rated-current range. The item stream follows a mixed parcel profile rather than a uniform distribution, as shown in Table II. Item mass follows category-specific log-normal distributions, with a combined mean of 3.2 kg, standard deviation of 4.1 kg, and range of 0.1–25 kg.

Belt degradation follows the health-sequence profile in Table III. The profile is reported in normalized form so that wear-induced current drift remains much smaller than normal load-induced variation. At mean load, belt wear produces approximately 0.16 A of current drift over a full lifecycle, while

TABLE III
BELT-HEALTH DEGRADATION PROFILE USED FOR VALIDATION.

Sequence Region	Cycles	μ_{belt}	Wear Drift
Healthy baseline	0–1,000	0.350–0.354	0.00–0.02 A
Early wear	1,000–5,000	0.354–0.370	0.02–0.07 A
Progressive wear	5,000–8,000	0.370–0.390	0.07–0.11 A
Maintenance endpoint	8,000–10,000	0.390–0.450	0.11–0.16 A

item-to-item load variation can produce current changes of up to 4.5 A. Table IV gives representative samples showing how measured current is decomposed into load-explained current and wear residual.

The dataset contains 20 anonymized belt-health sequences ending in maintenance-confirmed degradation events rather than destructive run-to-failure tests. For modeling consistency, each sequence is normalized to 8,000–12,000 item cycles from healthy baseline to maintenance endpoint, yielding more than 200,000 total item cycles. The RUL label at cycle t is assigned post hoc as the remaining number of item cycles until the maintenance-confirmed endpoint.

Data availability and confidentiality. Facility identity, manufacturer-specific equipment identifiers, internal asset names, and raw operational logs are withheld as proprietary. The paper reports only normalized, aggregated, equipment-class parameters. No human-subject data, personally identifiable information, or operator-specific records were used.

B. Protocol and Hyperparameters

Data are split by belt-health sequence: 12 runs for training, 4 for validation, and 4 for test. All results are reported over 5 random seeds (mean $\pm$ std). The decomposition network is trained for 200 epochs with Adam (lr $= 5 \times 10^{-4}$, weight decay 1×10^{-4}) and early stopping with 20-epoch patience on validation decomposition loss. The GRU is trained separately for 150 epochs with early stopping on validation RUL RMSE. Batch size is 256.

C. Baselines

We compare six configurations: (C1) threshold-based—trigger maintenance when raw motor current exceeds a static $\pm 3\sigma$ threshold from the healthy baseline; (C2) current-only GRU—feed raw current into the GRU without decomposition; (C3) manifest-informed GRU—use manifest-estimated item mass as a covariate alongside current, without physics loss; (C3b) manifest + physics—use the same manifest-estimated mass with physics-constrained decomposition and GRU; (C4) VPD-Net without physics loss—use vision-estimated mass together with neural decomposition and a GRU, with the physics term disabled; and (C5) full VPD-Net with physics loss.

D. Results: RUL Prediction

The full VPD-Net (C5) reaches an RUL RMSE of 41.3 ± 3.2 cycles, which is a 47% reduction relative to the current-only baseline (C2) and a 34% reduction relative to the manifest-informed model without physics (C3). One-sided Wilcoxon signed-rank tests show that C5 outperforms C2, C3, and C4

TABLE IV
EXAMPLE VALIDATION SAMPLES FROM THE ANONYMIZED CONVEYOR STREAM.

Cycle	Item Type	True Mass (kg)	Vision Mass (kg)	Measured Current (A)	Load Current (A)	Wear Residual (A)
250	Envelope	0.32	0.35	2.16	2.15	0.01
1,480	Small box	2.10	2.02	2.89	2.86	0.03
3,260	Polybag	1.45	1.72	2.73	2.68	0.05
5,840	Medium box	6.30	5.91	4.52	4.43	0.09
7,920	Large box	14.80	15.62	7.30	7.19	0.11
9,410	Medium box	4.75	4.52	4.22	4.08	0.14
9,870	Large box	22.40	21.15	8.57	8.41	0.16

TABLE V
RUL PREDICTION PERFORMANCE (MEAN $\pm$ STD, 5 SEEDS).

Configuration	RMSE (cyc)	MAE (cyc)	R^2	FP Rate%	Late Pred.%
C1: Threshold	142.7 ± 18.3	118.4 ± 14.6	—	23.1 ± 3.4	18.7 ± 2.8
C2: Current-only GRU	78.6 ± 5.1	61.2 ± 4.0	$.741 \pm .028$	14.8 ± 1.9	9.2 ± 1.4
C3: Manifest+GRU	62.4 ± 4.3	48.7 ± 3.5	$.823 \pm .021$	9.6 ± 1.4	6.8 ± 1.1
C3b: Manifest+phys.	45.8 ± 3.6	35.1 ± 2.7	$.894 \pm .016$	7.2 ± 1.1	4.6 ± 0.8
C4: VPD, no phys.	49.1 ± 3.8	38.2 ± 2.9	$.878 \pm .018$	8.1 ± 1.2	5.4 ± 0.9
C5: VPD-Net (ours)	$\mathbf{41.3 \pm 3.2}$	$\mathbf{31.5 \pm 2.4}$	$\mathbf{.911 \pm .014}$	$\mathbf{6.7 \pm 1.0}$	$\mathbf{4.1 \pm 0.7}$

at the 5% significance level ($p = 0.031$ for each pairwise comparison, the minimum achievable with five seeds).

The C3b row helps separate two questions: how much of the gain comes from the physics loss, and how much comes from the vision-based load estimate? C3b uses manifest-estimated mass together with the physics loss and reaches an RMSE of 45.8. C5 uses vision-estimated mass with the same physics-constrained decomposition and improves further to 41.3, a 10% reduction. A plausible explanation is that the camera observes the conveyed item directly—including packaging variation and condition at the time of transport—whereas manifest values are nominal and may be inaccurate or stale. The physics loss appears to provide the larger gain (C3 $\rightarrow$ C3b: 27% improvement), while the vision-based load estimate adds a further improvement beyond manifest-informed modeling. C3b also outperforms C4 (45.8 vs 49.1), which indicates that the physics constraint contributes more than the mass-source improvement alone.

False-positive maintenance alerts fall from 14.8% in C2 to 6.7% in C5, a 54% reduction. Late predictions—defined here as underestimating remaining life by more than 10% of the actual value—drop from 9.2% to 4.1%.

E. Vision Mass Estimation Accuracy

The vision-assisted mass estimator reaches 94.1% accuracy within $\pm 10\%$ of true mass across all item types. Performance is weakest for irregular items such as bags and soft parcels, where the bounding-volume assumption tends to overestimate mass. The decomposition network tolerates this noise because the physics loss constrains the overall energy balance, so small mass errors translate into proportionally small current-estimation errors that average out over the GRU's observation window.

F. Ablation: Component Contribution

Removing vision causes the largest degradation ($+22.4$ cycles RMSE), which confirms that per-item load estimation is the most important component. Without it, the model falls back to a constant mean mass, and load variation overwhelms the wear residual. The physics loss is the second-largest contributor ($+7.8$), which supports its anchoring role. The GRU matters most near the maintenance endpoint, where degradation accelerates and a linear trend misses that curvature. The ablation effects are non-additive (sum $= 47.0$, exceeding the total C2–C5 range of 37.3), reflecting shared information across components.

G. Sensitivity Analysis

VPD-Net remains stable under moderate mass-estimation noise. Adding zero-mean Gaussian noise to the vision mass estimate increases RUL RMSE from 42.1 cycles at 5% noise to 44.8 at 10%, 52.3 at 15%, and 68.7 at 30%. Even at 15% noise, the model remains stronger than the manifest-informed baseline without physics. Reducing training from 12 to 8 sequences increases RMSE from 41.3 to 46.8, while 6 sequences raises RMSE to 54.1, suggesting graceful degradation as training data decreases.

H. Computational Feasibility

The full VPD-Net pipeline has 0.45M parameters (segmentation network: 0.38M, decomposition MLP: 0.03M, density embedding: < 0.01M, GRU: 0.04M). On an NVIDIA Jetson Orin Nano, per-item latency is 28 ms for depth segmentation, 4 ms for volume extraction, 2 ms for decomposition, and 1 ms for the GRU update, for a total of 35 ms. On a sortation line with a 2 s cycle time, this uses 1.75% of the available cycle budget. The system runs continuously without buffering.

TABLE VI
MASS ESTIMATION ACCURACY BY ITEM CATEGORY.

Item Category	**Mass range (kg)**	**MAE (kg)**	**±10% Acc.**	**Main Error Source**
Small/flat (envelopes)	0.1–0.5	$0.08 \pm .02$	96.2%	Minimal—thin items well-resolved
Medium boxes	0.5–5.0	$0.31 \pm .06$	95.8%	Volume accurate, density variance
Large/heavy boxes	5.0–25.0	$1.24 \pm .18$	93.1%	Edge occlusion, density assumption
Irregular (bags, tubes)	0.2–8.0	$0.62 \pm .11$	88.4%	Shape deformation, volume overestimation
Wtd. avg. (by count)	**0.1–25.0**	$\mathbf{0.41 \pm .07}$	**94.1%**	Density lookup is primary limiter

TABLE VII
ABLATION: RUL RMSE WHEN EACH COMPONENT IS REMOVED.

Ablation	**Δ RMSE**	**Δ FP**	**Interpretation**
Remove vision (default 3.2 kg)	+22.4	+5.1%	Load variation unresolved
Remove physics loss	+7.8	+1.4%	Decomposition unanchored
Remove GRU (linear trend)	+11.6	+2.3%	Endpoint curvature missed
Remove density lookup (flat ρ)	+5.2	+0.9%	Dense/light items confused

VI. DISCUSSION

The results show that the main challenge in conveyor MCSA is not detecting current variation, but explaining its source. The current-only GRU improves over thresholding, but it still confuses heavy-item current spikes with degradation. Manifest-informed mass reduces this ambiguity, and the physics loss improves the decomposition further by anchoring load current to the motor torque relationship. Full VPD-Net performs best because the camera observes the conveyed item directly rather than relying on nominal or stale manifest values.

The false-positive reduction is the most operationally meaningful result. Cutting alerts from 14.8% to 6.7% reduces unnecessary belt replacements and improves operator trust in predictive-maintenance recommendations. The same decomposition idea may also apply to other driven systems, such as robotic pick-and-place units, cranes, and elevator hoistways, where load variation and mechanical wear are mixed in the motor signal.

Limitations. The study uses anonymized industrial testbed data, but facility identity, manufacturer-specific equipment identifiers, internal asset names, and raw logs are withheld for confidentiality. Results should be replicated across additional vendors, sites, and conveyor configurations. The physics loss also uses a simplified steady-state motor model; future work should include transient dynamics and multi-fault decomposition.

VII. CONCLUSION

This paper presented VPD-Net, a depth-assisted and physics-constrained approach for separating belt wear from load variation in conveyor motor-current signals. The key idea is to estimate conveyed-item load from a top-down depth camera and isolate the portion of current not explained by normal load variation. In anonymized industrial testbed validation across 20 belt-health sequences, VPD-Net achieved 41.3-cycle RUL RMSE, improving 47% over current-only baselines and 10% over a physics-constrained manifest-informed baseline. False maintenance alerts dropped by 54%, and ablations showed that depth-based load estimation is the largest contributor while the physics loss provides important decomposition stability. Future work will extend validation across additional sites and vendors and broaden the framework to multi-fault diagnosis involving belt wear, roller degradation, and motor-side faults.

REFERENCES

[1] G. Fedorko, V. Molnár, Ž. Dovica, T. Husakova, and N. Kral'iková, "Failure analysis of belt conveyor damage caused by the falling material," *Engineering Failure Analysis*, vol. 36, pp. 30–38, 2014.

[2] H. Rana, T. Nagayama, K. Hisazumi, and T. Tominaga, "Damage identification of a belt conveyor support structure based on cross-sectional vibration characteristics," *Structural Control and Health Monitoring*, vol. 26, no. 6, art. e2349, 2019.

[3] M. E. H. Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 984–993, 2000.

[4] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review," *IEEE Transactions on Energy Conversion*, vol. 20, no. 4, pp. 719–729, 2005.

[5] M. H. Farhat, L. Gelman, A. O. Abdullahi, A. Ball, G. Conaghan, and W. Kluis, "Novel fault diagnosis of a conveyor belt mis-tracking via motor current signature analysis," *Sensors*, vol. 23, no. 7, art. 3652, 2023.

[6] L. Gelman, A. O. Abdullahi, A. Moshrefzadeh, A. Ball, G. Conaghan, and W. Kluis, "Innovative conveyor belt monitoring via current signals," *Electronics*, vol. 12, no. 8, art. 1804, 2023.

[7] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems," *Mechanical Systems and Signal Processing*, vol. 42, no. 1–2, pp. 314–334, 2014.

[8] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, 2018.

[9] Z. Wang, Z. Wu, X. Li, H. Shao, T. Han, and M. Xie, "Attention-aware temporal–spatial graph neural network with multi-sensor information fusion for fault diagnosis," *Knowledge-Based Systems*, vol. 278, art. 110891, 2023.

[10] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.

[11] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, pp. 422–440, 2021.

[12] Y. A. Yucesan and F. A. C. Viana, "Physics-informed neural networks for wind turbine main bearing fatigue," in *Proc. International Design Engineering Technical Conferences (IDETC)*, ASME, 2019, art. V001T02A011.

[13] T. Liang, B. Lin, W. Bai, M. Chen, G. Liu, and L. Wang, "RGB-D-based parcel volume measurement," in *Proc. IEEE International Conference on Mechatronics and Automation (ICMA)*, 2019, pp. 1107–1112.

[14] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 12179–12188.

[15] Industrial communication networks—Network and system security: System security requirements and security levels, IEC 62443-3-3:2013, Int. Electrotech. Comm., 2013.