

Article

A Surface Defect Detection System for Industrial Conveyor Belt Inspection Using Apple's TrueDepth Camera Technology

Mohammad Siami ¹, Przemysław Dąbek ^{2,*}, Hamid Shiri ³, Tomasz Barszcz ¹ and Radosław Zimroz ²

¹ AGH University of Krakow, Faculty of Mechanical Engineering and Robotics, Department of Robotics and Mechatronics, al. A. Mickiewicza 30, 30-059 Krakow, Poland; msiami@agh.edu.pl (M.S.); tbarszcz@agh.edu.pl (T.B.)

² Faculty of Geoengineering, Mining and Geology, Wrocław University of Science and Technology, Na Grobli 15, 50-421 Wrocław, Poland; radoslaw.zimroz@pwr.edu.pl

³ National Oceanography Centre, European Way, Southampton SO14 3ZH, UK; hamid.shiri@noc.ac.uk

* Correspondence: przemyslaw.dabek@pwr.edu.pl

Abstract

Maintaining the structural integrity of conveyor belts is essential for safe and reliable mining operations. However, these belts are susceptible to longitudinal tearing and surface degradation from material impact, fatigue, and deformation. Many computer vision-based inspection methods are inefficient and unreliable in harsh mining environments characterized by dust and variable lighting. This study introduces a smartphone-driven defect detection system for the cost-effective, geometric inspection of conveyor belt surfaces. Using Apple's iPhone 12 Pro Max (Apple Inc., Cupertino, CA, USA), the system captures 3D point cloud data from a moving belt with induced damage via the integrated TrueDepth camera. A key innovation is a 3D-to-2D projection pipeline that converts point cloud data into structured representations compatible with standard 2D Convolutional Neural Networks (CNNs). We then propose a hybrid deep learning and machine learning model, where features extracted by pre-trained CNNs (VGG16, ResNet50, InceptionV3, Xception) are classified by ensemble methods (Random Forest, XGBoost, LightGBM). The proposed system achieves high detection accuracy exceeding 0.97 F1 score in the case of all proposed model implementations with TrueDepth F1 score over 0.05 higher than RGB approach. Applied cost-effective smartphone-based sensing platform proved to support near-real-time maintenance decisions. Laboratory results demonstrate the method's reliability, with measurement errors for defect dimensions within 3 mm. This approach shows significant potential to improve conveyor belt management, reduce maintenance costs, and enhance operational safety.



Academic Editors: Vinko Škrlec, Muhamed Suceska and Hajime Ikeda

Received: 26 November 2025

Revised: 22 December 2025

Accepted: 24 December 2025

Published: 7 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The operational continuity and safety of mining transport systems depend heavily on the integrity of their conveyor belts [1–5]. These belts are particularly prone to longitudinal tearing from impacts with sharp objects and from material fatigue, which can cause costly unplanned downtime and resource losses [6]. Phenomena such as belt mistracking can lead to rubbing against the conveyor structure, significantly shortening belt life cycle [7,8]. Given the limitations of manual visual inspection, developing reliable automated methods for detecting longitudinal tears has become an important research challenge [9,10].

Monitoring conveyor belt condition is challenging because damage to the top cover, such as cuts and gouges caused by sharp, falling material, can propagate and cause core degradation [11]. In practice, surface and geometric assessments at many mining sites still depend largely on manual visual inspection by supervisors. Prior efforts to automate inspection with image-based methods have met limited success, largely due to adverse field conditions, including poor illumination and airborne dust, that degrade data quality and complicate reliable analysis. These limitations highlight the need for more robust, environment-resilient inspection techniques [6,12].

Automatic methods for detecting longitudinal tears in industrial conveyor belts fall into two main categories: contact and non-contact. Contact techniques use hardware that physically interacts with the belt, such as linear detectors, swing rollers, and pressure sensors [13]. Although these approaches can be fast and conceptually simple, they tend to be costly and can collide with conveyed material, producing false alarms. Non-contact approaches, including electromagnetic induction and X-ray fluoroscopy, generally yield fewer false detections but depend on precise sensor-belt coupling, an arrangement that is difficult to maintain in the dynamic conditions of mining operations [14].

Surface reconstruction through 3D scanning offers a valuable approach to evaluating conveyor belt condition, enabling the generation of precise digital representations of belt geometry. Consequently, this perspective suggests that affordable 3D scanning could be a beneficial method for inspecting conveyor belts, despite the ongoing challenge of adapting these readily available technologies for reliable performance in demanding industrial environments [15]. Recent developments have expanded the availability of these technologies across a spectrum of devices, ranging from sophisticated industrial setups to more affordable scanners [16]. Although high-end systems offer greater precision, lower-cost devices frequently prove adequate and more practical when the objective is to identify significant geometric anomalies rather than minute imperfections [15].

Modern consumer-grade smart devices, including mobile phones and tablets, now incorporate sophisticated scanning technologies that extend beyond traditional photogrammetry. Among these, LiDAR (Light Detection and Ranging) functions on a Time-of-Flight (ToF) principle, determining distance by measuring the delay between emitting a light signal and receiving its reflection [17]. In contrast, Apple's proprietary TrueDepth camera utilizes a vertical-cavity surface-emitting laser (VCSEL), a dot projector, a flood illuminator, and an infrared camera. Its operational core involves projecting a pattern of over 30,000 infrared dots onto a scene; the distortion of this pattern, captured by the infrared camera, is then analyzed to construct a depth map [18]. This map is subsequently processed by machine learning algorithms to generate a precise mathematical model of the environment [19].

The TrueDepth camera built into Apple devices functions as a low-cost 3D scanner, using infrared illumination to produce depth maps for applications such as facial recognition and augmented reality. Its use in scientific contexts, particularly for anthropometric data collection, has been investigated in recent studies [15,20,21]. Nevertheless, the sensor's capacity to reliably reconstruct fine surface geometry remains an area of ongoing research.

In this paper, we present a low-cost, smartphone-driven maintenance system for intelligent condition monitoring of conveyor belt surfaces that leverages the iPhone 12 Pro Max as an integrated sensing platform. The system uses the device's TrueDepth camera to create accurate 3D point cloud models of the moving belt. This allows for a quantitative assessment and identification of surface issues based on their geometric characteristics. This smartphone-only approach supports cost-effective, near-real-time monitoring to aid maintenance decision-making.

Deep learning models have exhibited strong efficacy in image classification, detection, and segmentation, largely due to the application of 2D convolutional neural networks (CNNs). These models effectively capture both global and spatial features, demonstrating considerable generalization capabilities, which renders them particularly well-suited for the analysis of RGB imagery. Recent developments in 3D deep learning have expanded the application of these techniques to include classification, object detection, and semantic segmentation of point cloud data. Despite this progress, 3D CNNs continue to be computationally demanding and exhibit reduced scalability compared to their 2D counterparts [22]. To address this limitation, we can use dimension reduction methods, such as projection-based techniques. This allows us to use pre-trained and efficient 2D convolutional neural network (CNN) architectures to analyze 3D data.

The contributions of the proposed surface defect detection system are summarized as follows:

- A smartphone-driven 3D inspection pipeline: A cost-effective system has been developed to capture point clouds, utilizing the iPhone 12 Pro Max's TrueDepth camera, and subsequently processes these point clouds through a novel 3D-to-2D projection method. This pipeline capitalizes on pre-trained 2D convolutional neural networks (CNNs) to extract deep features, and it incorporates efficient tree-based classifiers to facilitate robust defect detection and classification, relying exclusively on geometric data.
- An industrial benchmark and empirical assessment: a specialized dataset comprising TrueDepth point clouds, encompassing various induced fault types and detailed annotations, serves as the foundation for evaluating the generalizability of the proposed method in the context of conveyor-component condition monitoring.
- A lightweight, deployable detection and quantification pipeline: a computationally efficient approach suitable for edge and mobile deployment that identifies topographic defects and provides practical geometric measurements (e.g., depth, volume) to support maintenance decisions.

The remainder of this paper is structured as follows. Section 2 reviews related work on conveyor belt damage modes and point cloud based condition monitoring methods. Section 3 delineates the proposed methodology for geometric defect detection. Section 4 describes the experimental setup, including the dataset, data collection process, and evaluation metrics. Section 5 reviews the model training and validation process. Section 6 presents and discusses the experimental results. Finally, Section 7 concludes the study and suggests directions for future work.

2. Literature Review

A wide range of techniques are used to diagnose and inspect conveyor belts, each aimed at identifying various types of faults and maintaining reliable operations. These approaches span from traditional visual and manual inspections to more advanced technologies, such as ultrasonic testing [23,24] or magnetic belt inspection [25]. Additional methods include the use of RGB and infrared cameras, commonly applied to monitor idlers [26–28], detect material blockages [29], or detect misalignment of the track belt [30], as well as to evaluate the belt's own condition [31,32], alongside acoustic analysis [33,34] and X-ray imaging [35]. Multisensor systems designed for belt conveyor monitoring, such as DiagBelt+ can incorporate mentioned sensors to great results [36–38].

The approach using LiDAR data can be found in [6] where the belt has been scanned with the TLS system (terrestrial laser scanner), obtaining the elevation data of the belt and consequently allowing the detection of local defects. In [39] the authors used a binocular line laser stereo vision camera mounted between the upper and lower belts to obtain the

data for the detection of longitudinal rip. As in the previous case, the suspected points come from fluctuations of the point position in selected directions. An example of the damage detection in a multi-wedge belt can be found in [40], where the authors detected the most common pits, scratches, and cracks with a detection rate of 96%. The proposed methodology consisted of point cloud extraction, clustering of separate tooth top surfaces with use of DBSCAN, and final defect detection through adaptive moving window.

Some other uses of surface inspection with the use of scanning technology can be found in the inspection of buildings, road damage detection, and quality control of various materials. In [41], the authors used TLS in the bridge structural health monitoring task. The accuracy of the scanning equipment (Faro S350) proved to be high in comparison with manual measurements. As expected, the differences increased with the increasing angle of the scanning axis perpendicular to the surface. In [42] the colored point cloud of the ship hull was utilized in the detection of corroded regions, allowing better estimation and optimization of maintenance routines. The authors used threshold-based detection similar to image segmentation methodologies. On a larger scale, scanning has been used in cases such as quality inspection of large prefabricated housing units [43], where the geometric dimensions (together with parameters such as straightness or flatness) of different elements have been measured. In this case, the preserved accuracy remained below 2.3 mm. A few different scanners, including the iPhone Lidar, have been tested in the task of damage estimation of forest road surfaces. The quality and accuracy of iPhone Lidar proved sufficient for the task, although with error tied strictly to the distance from the scanned surface [44].

In [45] the authors utilized density histograms, Euclidean clustering, and a dimension-based classifier for the detection of idle position for further diagnosis. Machine vision and artificial intelligence are often used in modern methods to make fault detection more accurate and to better predict when maintenance will be needed. This category includes approaches such as support vector machines [46–48], neural networks [49,50], or DBSCAN [51].

In the case of belt damage detection with the use of machine learning, there can be distinguished two types of damage that the methodologies focus on separately. Both of them usually rely on the image data (such as RGB or X-ray images), with exceptions such as magnetic detection [50]. First branch of the trained networks revolve around the belt deviation problem [28,52–54] that often incorporate various edge and line detection algorithms into the algorithm pipeline. The second one focuses on the surface damage [55,56]. In both cases, a robust region of interest reduction is usually very beneficial for the final results - for this purpose detectors such as MobileNet SSD have been implemented [28].

The TrueDepth camera utilized has been thoroughly tested in [57] where authors proved its usability in the millimeter range applications, obtaining 0.1 mm details detection on a working distance of 150 to 170 mm. For a stable measurement of less intensive textures it was recommended to stay in range of maximum 300 mm distance from the surface, 500 mm in case of more textured material. Similar conclusions have been reached by authors in [58], where they obtained results of point-to-plane deviation from 0.291 to 0.739 mm for distances from surface increasing from 175 to 450 mm. This highlights a significant importance of the distance from the measured object that might render it not useful in many industrial applications. A direct comparison with existing solutions have been provided in [59], where authors evaluated the scanning accuracy of iPad Pro TrueDepth with Artec Space Spider high resolution industrial scanner. In these tests, TrueDepth was outperformed by industrial solution, although the differences were in the range of one millimeter differences. Additionally, a high impact of the scanner movement

and measurement technique in the case of TrueDepth accuracy have been noticed, which may lead to a significant improvement in the scanner performance if implemented properly.

3. Material and Methods

This section delineates the proposed methodology for the intelligent inspection of conveyor belt surfaces based on geometric data. The overall procedure, illustrated in Figure 1, begins with the acquisition of 3D point cloud data using the integrated TrueDepth camera of an iPhone 12 Pro Max. A critical preprocessing step involves transforming the raw 3D point clouds into 2D feature projections. This transformation is essential to reduce computational complexity and to leverage pre-trained 2D CNNs for effective feature extraction. Subsequently, a hybrid framework is introduced, where deep features extracted from the 2D point cloud projections are classified using traditional machine learning models. The procedures for training and evaluating these models are detailed to identify the optimal CNN and classifier pair for defect detection.

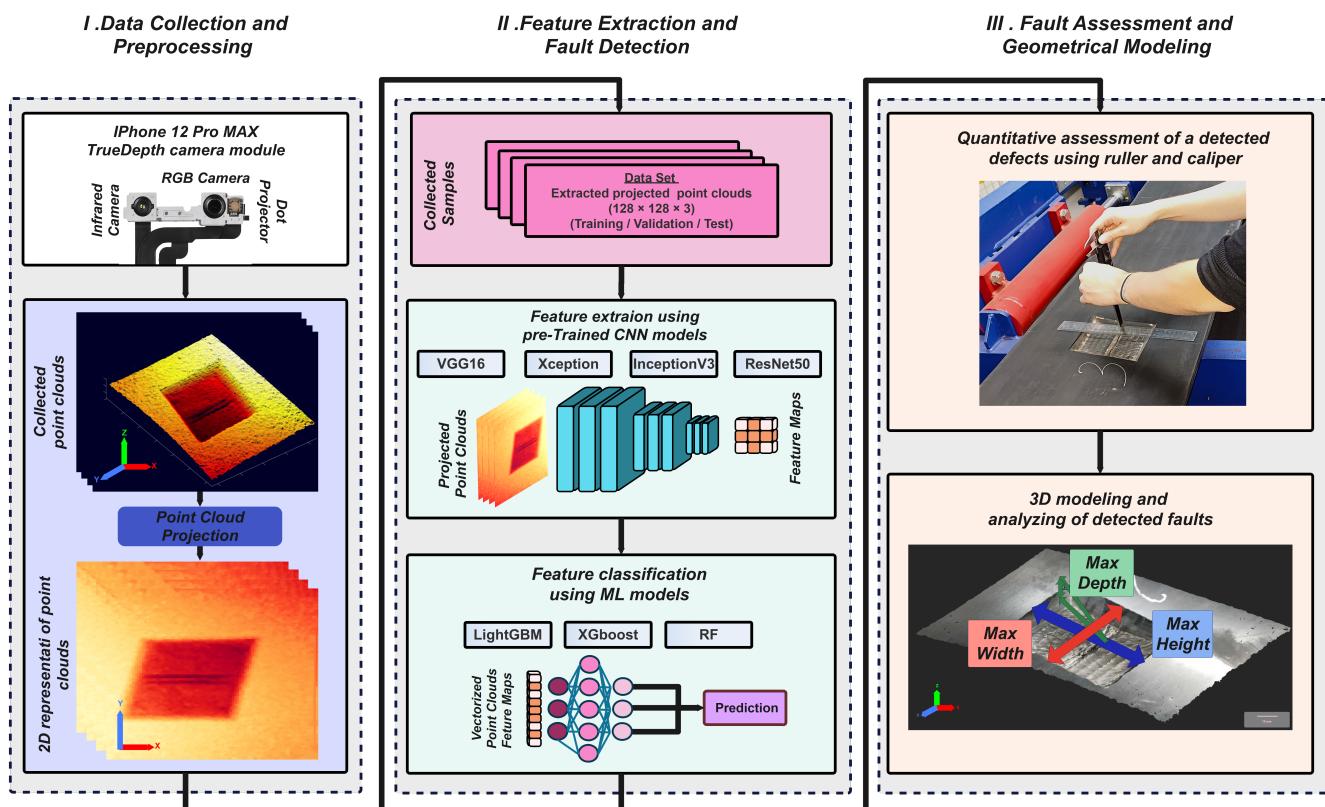


Figure 1. Flowchart of the hybrid CNN-ML pipeline for surface defect detection via 3D-to-2D projection of point cloud data captured by the TrueDepth camera.

3.1. Point Cloud Reconstruction

To leverage the TrueDepth camera for accurate geometric modeling of the conveyor belt surface, the raw sensor data must first be preprocessed. This section outlines the core computer vision principles for this processing, beginning with the camera's intrinsic parameters. These parameters govern the transformation between 3D world coordinates and 2D image pixels, forming the foundation for converting depth maps into 3D point clouds.

The projective transform maps a 3D point $\mathbf{p} = [x/z, y/z, 1]^T$ from the camera coordinate system to a point $\mathbf{u} = [u, v, 1]^T$ in the image plane $\mathbf{u} = \mathbf{K}\mathbf{p}$. Here \mathbf{K} is defined as:

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0.0 & f_y \cdot a & c_y \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (1)$$

The matrix \mathbf{K} contains the focal length f in pixels, an aspect ratio a , a shear factor, s and the principal point c_x, c_y .

The authors Urban et al. [60] performed the series of experiments to find the calibration factors. In their work for the iPhone 12 Pro Max the factory-calibrated intrinsics returned: $a = 1$ and $s = 0$. In addition, the principal point always coincides with the lens distortion center that can also be requested from the APIs provided by the manufacturer (Apple Inc., Cupertino, CA, USA) [60,61].

In the next step the acquired depth image needs to be used to reconstruct the point clouds. To convert a depth image $D \in \mathbb{R}^2$ to a point cloud, $\mathbf{X}_{i,j} \in \mathbb{R}^3$ the following mapping can be used:

$$\mathbf{X}_{i,j} = D(i, j)\mathbf{K}^{-1}\mathbf{u}_{i,j} \quad (2)$$

where $i = 1 \dots W$ and $j = 1 \dots H$ and $\mathbf{u}_{i,j} = [i, j, 1.0]^T$. In the test smartphone, the depth image has a resolution of 640×480 pixels.

3.2. 3D-to-2D Point Clouds Feature Projection

The proposed architecture processes point clouds using 2D convolutional layers, necessitating the transformation of 3D point data into 2D feature maps compatible with regular grid-based processing. Since point clouds inhabit continuous 3D space, they cannot be directly processed by 2D CNNs without prior conversion to a structured 2D representation [22].

For a point \mathbf{p} with coordinates (x, y, z) and associated feature f , the projection process involves normalizing the point cloud to a specified range relative to each projection plane. For the XY-plane with dimension $l \times l$, the x and y coordinates are normalized to the interval $[-\frac{l}{2}, \frac{l}{2}]$. Feature projection onto the grid is accomplished through bilinear interpolation, selected for its favorable balance of computational efficiency and memory requirements. When multiple features map to the same grid cell, they are aggregated through summation. This process generates a 2D feature map of dimension $l \times l$ from the 3D point cloud, formally defined as:

$$I_{XY}(x', y') = \sum_x \sum_y G(x', y', x, y) \cdot f(x, y, z) \quad (3)$$

$$G(A, B, a, b) = g(A, a) \cdot g(B, b) \quad (4)$$

$$g(N, n) = \max(0, 1 - |N - n|) \quad (5)$$

where $I_{XY}(x', y')$ denotes the 2D feature at grid position (x', y') on the XY-plane, $f(x, y, z)$ represents the 3D feature of point \mathbf{p} , and $G(\cdot)$ is the 2D bilinear interpolation kernel composed of one-dimensional linear kernels as defined in Equation (4). Figure 2 illustrates this projection mechanism.

For points that share identical (x, y) coordinates but differ in their z values, such as two points \mathbf{p} and \mathbf{p}' with coordinates (x, y, z) and (x, y, z') respectively, projecting into only the XY-plane results in identical 2D representations. This characteristic ensures that the surface geometry of the conveyor belt is effectively captured while maintaining computational efficiency.

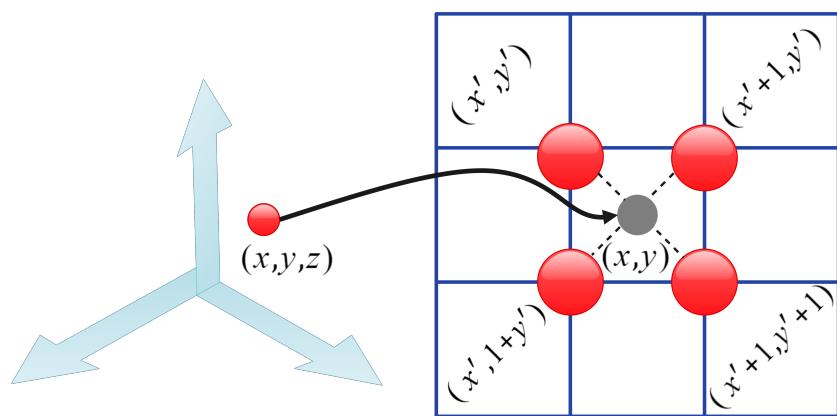


Figure 2. The operation of the point cloud transformation from 3D space into a 2D plane.

To ensure data continuity and the reliability of the geometric profile in the captured point cloud, missing data points resulting from infrared dots that were not correctly reflected and captured by the TrueDepth camera were reconstructed using linear interpolation. For each missing point at coordinates (x, y) , the algorithm estimates its depth value z by analyzing the values of its valid neighboring points within a defined spatial kernel. This process reconstructs a spatially consistent point cloud suitable for further analysis.

3.3. Deep CNN Models for Defect Feature Extraction

Convolutional Neural Networks (CNNs) are a distinct category of deep learning architectures designed for analyzing structured grid data, with a primary focus on images. They mirror the hierarchical pattern recognition processes observed in biological vision systems [62,63]. Their remarkable effectiveness in image processing, coupled with the ability to learn directly from raw pixel data, has established CNNs as the prevailing approach in computer vision applications. The fundamental CNN architecture employs a series of convolutional filters, activation functions, and pooling operations to autonomously derive hierarchical feature representations from input images. These models are typically optimized using gradient-based methods, such as backpropagation, for a variety of tasks, including image classification and feature extraction [64,65].

The standard structure of a convolutional neural network is a series of specific layers arranged in order. Each layer has a distinct computational role in a step-by-step process of extracting features. The process begins with the input layer, which receives the original image data and applies preprocessing steps, such as normalization, to standardize the data. Following this, the convolutional layers, which are the core of the architecture, use learnable filters to extract spatial features through convolution. The generated feature maps pass through activation layers, often utilizing rectified linear units (ReLU), to incorporate non-linear transformations crucial for the acquisition of intricate mappings. Pooling layers then execute spatial down-sampling, diminishing feature dimensionality while maintaining essential information, thus improving computational efficiency and offering translational invariance. The concluding phase entails flattening the extracted features and subjecting them to fully connected layers, which consolidate high-level representations to produce classification outputs.

Transfer Learning with Pre-Trained Architectures

Over the last ten years, the development of Convolutional Neural Network (CNN) architectures has produced numerous models that excel in large-scale visual recognition applications. This study utilizes transfer learning, employing four well-established CNN architectures—VGG16, ResNet50, InceptionV3, and Xception—to extract distinguishing

features from the 2D point cloud representation of the conveyor belt surface. Transfer learning facilitates the transfer of knowledge from models initially trained on extensive datasets, such as ImageNet, to our specific area of interest, thereby substantially improving performance, particularly when labeled data is scarce [66].

Each architecture presents unique benefits for feature extraction. VGG16, distinguished by its consistent architecture, features 13 convolutional layers utilizing 3×3 filters, thereby establishing a deep yet uncomplicated structure that effectively identifies hierarchical features [67]. Conversely, ResNet50 incorporates residual connections throughout its 48 convolutional layers to address the vanishing gradient issue, thus facilitating the effective training of considerably deeper models [68]. The InceptionV3 architecture, in contrast, utilizes parallel convolutional pathways with diverse receptive fields to efficiently capture multi-scale features while managing computational complexity [69]. Lastly, Xception, an advancement of the Inception concept, is predicated on depthwise separable convolutions within its 71-layer design, which improves parameter efficiency while preserving robust representational capabilities [70].

Initially trained on the ImageNet dataset, which comprises more than 13 million images spanning 20,000 categories, these pre-trained models offer strong feature extraction abilities. We leverage these capabilities to identify surface anomalies within point cloud representations of conveyor belts. The features extracted from this process are then utilized as input for conventional machine learning classifiers, ultimately facilitating the classification of defects.

To prepare the extracted point cloud data for feature extraction using convolutional filters, the proposed projection algorithm was applied to convert the 3D point clouds captured by the TrueDepth camera into 2D top-view representations on the XY plane (see Figure 3). The z-axis (depth parameter) of point clouds was normalized to a range of 0 to 255 and represented as grayscale images, enabling their direct use as input to the employed feature extraction models. To reduce computational cost while preserving the essential geometric information, all projected point clouds were resized to 128×128 pixels.

To use the pre-trained architectures without changing the point cloud matrices to RGB, two extra dimensions were created by copying the single-channel image across all three channels. This channel duplication is a common practice to adapt single-channel data for models pre-trained on 3-channel RGB images, preserving the learned filter structures from ImageNet.

3.4. Hybrid Deep Learning and Machine Learning Framework

To address the challenges of limited computational resources and small datasets in industrial applications, particularly in our case study, we propose a hybrid framework. This framework combines deep feature extraction, using pre-trained CNN architectures, with traditional machine learning models for classification tasks based on the extracted features.

The proposed framework functions via a two-step process. Initially, pre-trained convolutional neural network (CNN) models are employed to extract high-level features from two-dimensional point cloud representations, which are themselves derived from the original three-dimensional point cloud data. Subsequently, traditional machine learning classifiers, such as Random Forest, Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), conduct the final classification based on the features that have been extracted. This modular design facilitates the efficient optimization of each individual component while simultaneously mitigating computational requirements when contrasted with end-to-end deep learning methodologies. Consequently, the framework retains the lightweight attributes that are critical for practical implementation within industrial contexts.

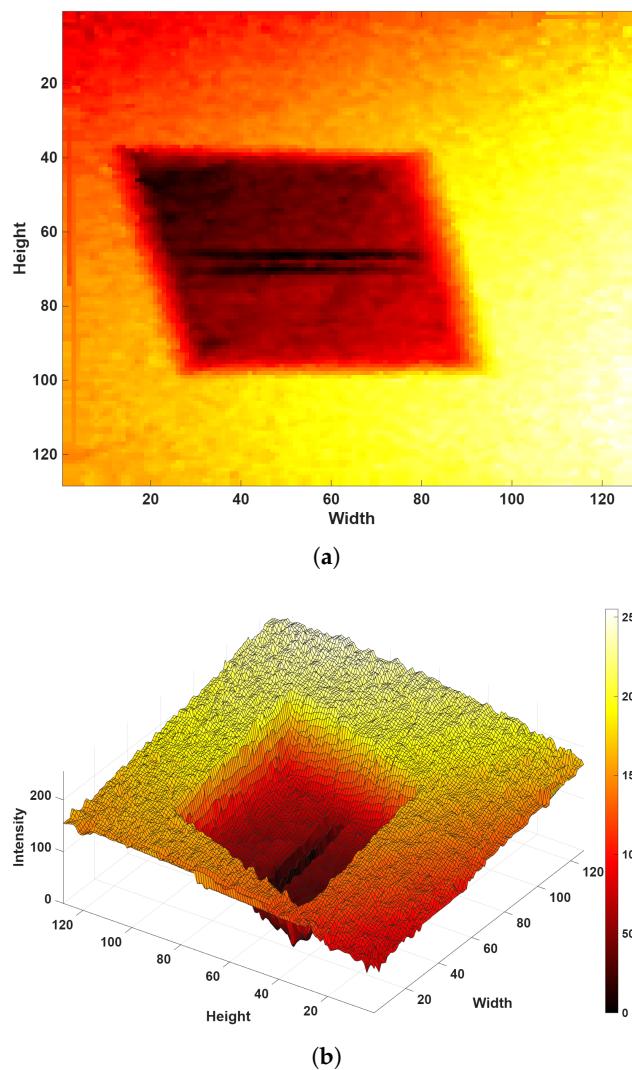


Figure 3. 2D representation of a 3D point cloud after normalization in XY and XYZ plane views. (a) 2D representation of a 3D point cloud for an exemplary defect. (b) 3D representation of normalized 2D point cloud.

Ensemble Machine Learning Classifiers

To augment the profound feature extraction capabilities inherent in pre-trained Convolutional Neural Networks (CNNs), we utilize three ensemble machine learning classifiers, each recognized for its strong performance in classification applications: Random Forest (RF), XGBoost, and LightGBM. These algorithms each present unique benefits when processing the features derived from 2D point cloud representations.

Random Forest (RF), as introduced by Breiman [71], constitutes an ensemble learning technique that employs bootstrap aggregation (bagging) to construct numerous decision trees during the training phase. RF operates by building trees in parallel; each tree is trained on randomly selected subsets of both data and features. This methodology serves to effectively reduce the overfitting tendencies often observed in individual decision trees. The ultimate predictions are derived from the aggregation of individual tree outputs via majority voting, thereby improving both accuracy and robustness.

XGBoost, as presented by Chen and Guestrin [72], constitutes an advanced version of Gradient Boosting Decision Trees (GBDT). In contrast to the bagging methodology of Random Forests (RF), XGBoost constructs trees sequentially through boosting, wherein each subsequent tree addresses the errors of its predecessors by minimizing a specified loss function using gradient descent. The algorithm's structure is characterized by level-wise

tree growth, which promotes balanced architectures, and it incorporates regularization techniques to manage model complexity. Consequently, the ultimate prediction is derived from a weighted aggregation of all tree outputs, thereby demonstrating considerable efficacy in handling intricate datasets.

LightGBM, a variant of GBDT, is engineered for both efficiency and rapid processing of extensive datasets [73]. This algorithm introduces two principal innovations: Gradient-based One-Side Sampling (GOSS), which emphasizes instances exhibiting substantial gradients while randomly sampling those with smaller gradients, and Exclusive Feature Bundling (EFB), which diminishes feature dimensionality by bundling mutually exclusive features. Furthermore, LightGBM utilizes leaf-wise tree growth, thereby facilitating faster convergence and enhanced performance relative to conventional level-wise methodologies.

4. Experimental Setup and Data Collection

The experimental investigation employed a steel cord conveyor belt with a rubber top cover as the test specimen. The belt was maintained in excellent condition with intact edges, attributable to its exclusive use in a controlled laboratory environment, as depicted in Figure 4. All data acquisition and testing were performed in the specialized belt conveyor laboratory at the Wrocław University of Science and Technology (WUST). This controlled facility enabled the precise induction of artificial defects and the subsequent collection of high-quality data necessary to validate the proposed inspection methodology.



Figure 4. The studied conveyor belt system was within a controlled laboratory setting.

To assess the efficacy of the suggested non-destructive inspection (NDI) system, nine unique artificial defects were incorporated onto the surface of the test conveyor belt. These defects were engineered to simulate common failure mechanisms observed in industrial contexts, specifically the significant impact and abrasive wear typical of hard rock mining operations. As depicted in Figure 5, the induced damage exhibits variations in geometry, depth, and severity, thus creating a demanding and representative dataset for thorough system evaluation.

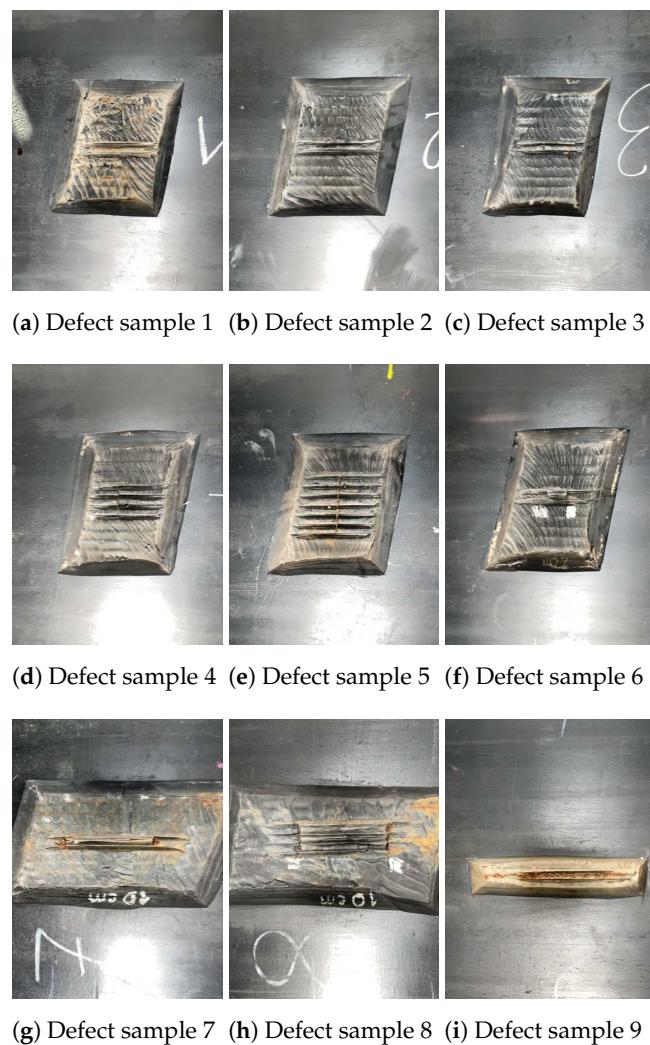


Figure 5. Representative collection of conveyor belt tearing defects analyzed in the study.

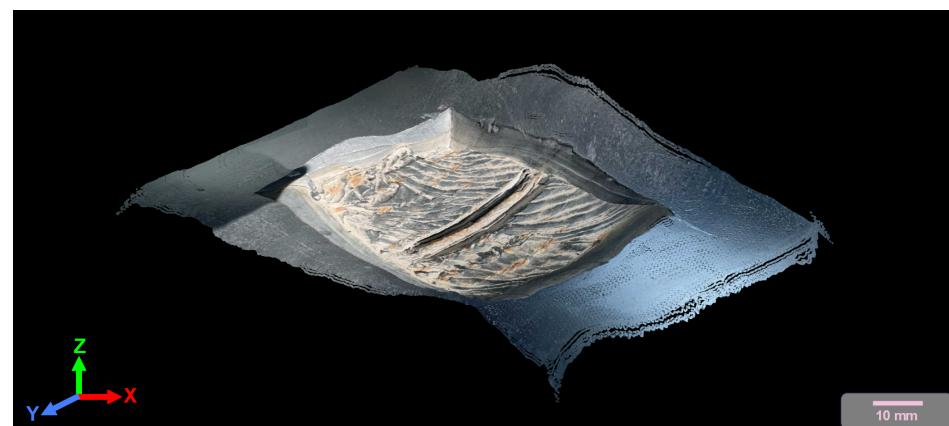
The simulated damage profile includes situations such as deep gouges and cuts that partially expose the underlying steel cord reinforcement. These defects are mainly characterized by localized geometric changes on the belt's surface. The proposed method uses these geometric indicators, specifically point clouds from a TrueDepth camera, to provide precise, quantitative measurements of 3D surface changes. This approach allows for reliable defect detection based solely on measurable geometric anomalies, maintaining accuracy regardless of lighting conditions or surface contamination.

4.1. Data Acquisition and Sensor Characteristics

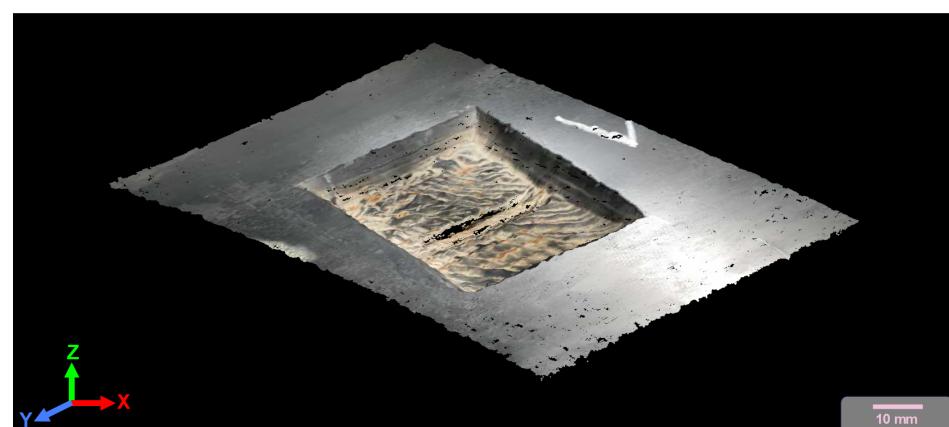
Data acquisition was performed using a smartphone-based sensing platform, using an iPhone 12 Pro Max, which features an integrated TrueDepth camera. This system employs structured light technology; it projects a pattern of over 30,000 infrared dots onto the surface, subsequently capturing the resultant deformation with an infrared camera. This procedure yields a dense 3D point cloud, with each point characterized by its spatial coordinates in relation to the sensor. Furthermore, the system integrates a flood illuminator to facilitate low-light operation and a CCD sensor for the concurrent capture of 2D RGB texture data.

As demonstrated in Figure 6, which compares the LiDAR and TrueDepth camera in capturing point clouds from an exemplary conveyor belt defect, the iPhone LiDAR shows limitations in accurately reproducing the defect's geometrical features. In contrast, the TrueDepth camera captures precise point clouds that enable accurate depth measure-

ment of the surface defects. The main reason behind the poor performance of the LiDAR sensor is the fact that the LiDAR module in the iPhone 12 Pro Max is specially designed for distance estimation rather than precise point cloud generation. Therefore, although it can work better with geometrical features at longer distances in comparison to the TrueDepth camera, it has a poor performance in capturing precise point cloud data at short distances, as was the intention of this case study.



(a) iPhone LiDAR sensor



(b) iPhone TrueDepth camera

Figure 6. Comparison of iPhone LiDAR and TrueDepth camera sensors in the correct capturing of the geometrical features of an exemplary defect.

The working distance of the TrueDepth camera was maintained within 200–300 mm, consistent with ranges validated in prior studies for reliable 3D data acquisition [57,74]. During experiments, the smartphone was positioned approximately 250 mm above the belt surface to capture samples. The TrueDepth camera operated at a frame rate of 30 frames per second, with each frame generating a corresponding point cloud. A rectangular region of interest (ROI) measuring 25 cm in width and 35 cm in height was continuously recorded from the central section of the belt. Each TrueDepth frame generated a point cloud containing 273,674 individual points (see Figure 7). To ensure smooth and consistent data acquisition, the conveyor belt was moving at a constant speed of 0.075 m/s. The test conveyor had a total length of 15 m and a belt thickness of 5 cm.

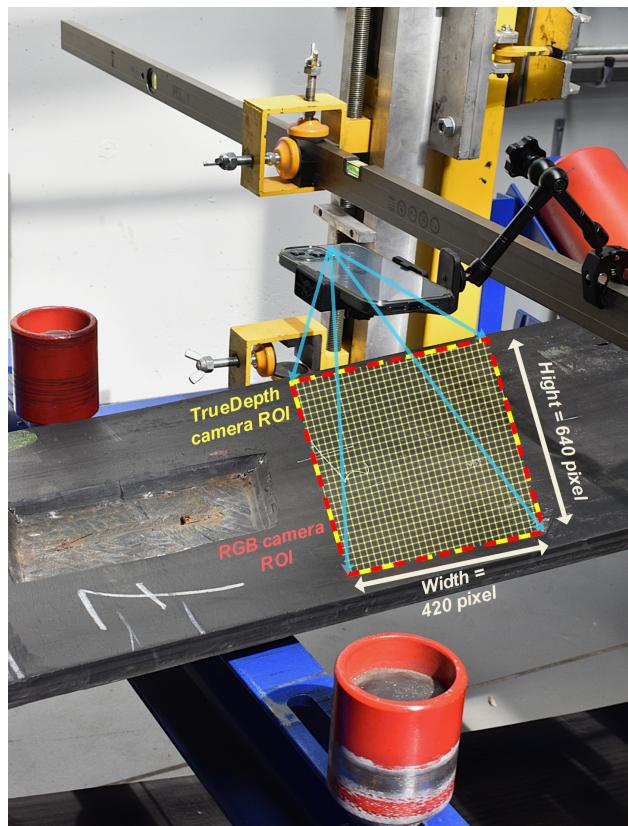


Figure 7. Smartphone mounting configuration for conveyor belt monitoring.

To ensure data integrity and avoid the opaque processing routines common in many 3D scanning applications, the “Record3D” application was utilized for data extraction. This application exports raw, unaltered point cloud data streams directly from Apple’s ARKit framework without applying proprietary post-processing or mesh refinement algorithms. Consequently, the dataset employed in this work consists exclusively of native ARKit outputs, establishing a transparent and reproducible foundation for analysis [58]. For validation purposes, manual measurements of maximum depth, width, and height for each defect were collected using precision rulers and calipers (see Figure 8) to enable comparative analysis with the camera-acquired results.



Figure 8. Hand measurement of defect size using a ruler and caliper.

4.2. Performance Metrics

The performance of the proposed classifier was evaluated using standard metrics derived from the confusion matrix: accuracy, sensitivity (recall), precision, and F1 score. These metrics are formally defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (9)$$

In these formulations, TP (True Positive) represents correctly identified defective regions, TN (True Negative) denotes correctly classified non-defective areas, FP (False Positive) indicates non-defective areas misclassified as defective, and FN (False Negative) corresponds to defective areas incorrectly classified as non-defective. Sensitivity quantifies the model's capability to detect actual defects, while precision measures the accuracy of positive predictions. The F1 score provides a balanced metric through the harmonic mean of precision and sensitivity. All metrics range from 0 to 1, with 1 representing optimal performance, which served as the optimization objective in this study.

5. Model Training and Validation Process

The dataset for this study comprised a total of 7086 of samples captured from the conveyor belt surface, representing nine distinct fault conditions as represented in the Figure 5. To ensure a robust evaluation, a structured approach was employed for data partitioning into training, validation, and test sets.

To mitigate the risk of overfitting from sequential, highly correlated samples and to maximize feature diversity for enhanced model generalizability, a strategic sample selection process was employed, leveraging the ORB (Oriented FAST and Rotated BRIEF) algorithm [75]. Established by Rublee et al. as computationally efficient, ORB was ideal for rapidly quantifying visual dissimilarity. ORB features were extracted from an initial pool of 832 healthy and 832 faulty samples, and the Hamming distance between their binary descriptors was computed for all projected point cloud pairs to generate a dissimilarity score for each sample. The 832 faulty samples represent the defect numbers 1 to 5. The final balanced training dataset was constructed by selecting the 100 most dissimilar samples from each of the two categories, ensuring the selected data encapsulated the widest possible variation in surface conditions and promoting robust model performance from a limited number of samples.

Hyperparameter optimization was conducted using a separate validation set, which consisted of 274 healthy samples and 686 faulty samples encompassing the two defects (defect numbers 6 and 7) on validation dataset. The hyperparameter optimization is an integral part of tuning the employed ML-based classifier in this study. The optimization was performed using random search cross-validation, a technique that efficiently explores the hyperparameter space by evaluating a fixed number of parameter settings sampled from specified distributions in RF, XGBoost, and LightGBM classifiers. Unlike an exhaustive grid search, random search cross-validation offers a more computationally efficient approach to identifying a near-optimal configuration for the classifiers, providing a favorable trade-off between search time and model performance.

The final performance evaluation of the model was conducted on a held-out test set, which included samples from the two remaining fault types (Faults 8–9) not seen during training or validation. This test set contained 259 healthy samples and 814 faulty samples, providing a rigorous assessment of the model's ability to generalize to novel defect patterns.

The experimental setup utilized the following hardware configuration: a desktop computer equipped with an AMD Ryzen 7 5800H CPU (Advanced Micro Devices Inc. (AMD), Santa Clara, CA, US), an NVidia GeForce RTX 3060 graphics processing unit (GPU) (NVidia, Santa Clara, CA, USA), and 16 GB of RAM.

6. Results and Discussion

This section presents a comprehensive evaluation of the proposed defect detection models. The analysis begins by systematically evaluating the strengths and limitations of models trained exclusively on data from the TrueDepth camera. To quantitatively validate the system's measurement precision, we compare the physical dimensions (height, width, and depth) of identified faults against manual measurements obtained with laboratory-grade tools in Section 6.2. This comparative analysis confirms that the geometric data derived from the point clouds provides highly accurate quantitative assessments of surface damage, moving beyond mere detection to enable precise fault characterization.

6.1. Hybrid CNN-ML Model Performance Comparison in Performing Surface Defect Classification

The performance of the trained classification models was evaluated for their capability to identify surface defects on the conveyor belt using RGB and TrueDepth cameras separately. Classifiers' efficacy was quantified using standard performance metrics: accuracy and F1 score. An F1 score exceeding 0.9 indicates strong potential for real-world industrial deployment. The comprehensive performance results across both validation and test datasets are summarized in Table 1.

Table 1. Performance comparison of models for RGB image and TrueDepth camera classification on validation and test data sets.

Model	Depth	Parameters	Accuracy				F1 Score			
			RGB Camera		TrueDepth Camera		RGB Camera		TrueDepth Camera	
			Validation	Test	Validation	Test	Validation	Test	Validation	Test
VGG16-RF	16	138.4 M	0.9740	0.9644	0.9667	0.9691	0.9817	0.9769	0.9763	0.9793
VGG16-XGBoost			0.9375	0.9269	0.9271	0.9372	0.9573	0.9527	0.9465	0.9568
VGG16-LightGBM			0.9333	0.9316	0.9635	0.9709	0.9542	0.9559	0.9745	0.9805
InceptionV3-RF	189	23.9 M	0.8313	0.8988	0.9844	0.9878	0.8864	0.9363	0.9890	0.9919
InceptionV3-XGBoost			0.7969	0.8304	0.9521	0.9803	0.8637	0.8926	0.9654	0.9869
InceptionV3-LightGBM			0.8094	0.8332	0.9260	0.9419	0.8728	0.8948	0.9495	0.9624
ResNet50-RF	107	25.6 M	0.8302	0.8510	0.9115	0.9119	0.8848	0.9003	0.9403	0.9432
ResNet50-XGBoost			0.7521	0.6607	0.7479	0.7432	0.8250	0.7418	0.8210	0.8259
ResNet50-LightGBM			0.8542	0.6710	0.8281	0.8013	0.8988	0.7740	0.8821	0.8691
Xception-RF	81	22.9 M	0.9573	0.9719	0.9812	0.9841	0.9692	0.9813	0.9867	0.9894
Xception-XGBoost			0.9802	0.9513	0.9198	0.9157	0.9860	0.9674	0.9456	0.9458
Xception-LightGBM			0.9833	0.9653	0.9229	0.9185	0.9882	0.9771	0.9477	0.9476

For the RGB image modality, the Xception architecture paired with a Random Forest (Xception-RF) classifier achieved the highest F1 score (0.9813) on the test set, demonstrating its superior capability in detecting defects based on visual features. This model also showed strong consistency, with its performance on the validation set (0.9692) closely matching its test results, indicating robust generalizability. The VGG16-RF model also performed robustly, with an F1 score of 0.9769. Notably, while the Xception-XGBoost model achieved the highest validation F1 score (0.9882) for RGB, its test performance (0.9674) experienced a

more significant drop, suggesting a potential for overfitting compared to the more stable RF-based variants. In contrast, models based on the ResNet50 architecture showed markedly lower performance on RGB data, with F1 scores on the test set falling below 0.78 for its XGBoost and LightGBM implementations. This suggests that the ResNet50 architecture may be less suitable for extracting discriminative features from the surface texture and color variations present in the conveyor belt images under the studied conditions.

For the 2D representation of point cloud data modality, which captures 3D topographic information, the InceptionV3-RF model achieved the highest F1 score (0.9919) on the test set. This indicates the exceptional effectiveness of geometric features for defect detection, as surface deformations like gouges and cuts manifest clearly as anomalies in the 3D point cloud. The Xception-RF and VGG16-LightGBM models also performed exceptionally well on this modality, with test F1 scores of 0.9894 and 0.9805, respectively. The consistently high performance across multiple architectures for the TrueDepth modality—with six different model-classifier combinations achieving a test F1 score above 0.97—underscores the inherent robustness of geometric data. This data is less susceptible to the visual challenges, such as lighting variations and dust, that can adversely affect RGB image analysis, a fact highlighted by the performance gap between modalities for architectures like InceptionV3, where the TrueDepth F1 score was over 0.05 higher than its RGB counterpart.

The confusion matrix for the top-performing TrueDepth model including InceptionV3-RF, shown in Figure 9, reveals a conservative detection profile: no false negatives were observed on the test set (259 faulty instances), while the primary error mode consisted of 13 false positives where intact surface areas were flagged as defective. This behaviour reduces the risk of missed critical defects, which is an important advantage for maintenance in safety-critical operations, but increases the rate of unnecessary follow-up inspections. The false-positive burden can be managed in practice by adjusting detection thresholds, applying simple post-processing filters, or introducing a lightweight secondary verification step to improve precision without substantially compromising defect detection sensitivity.

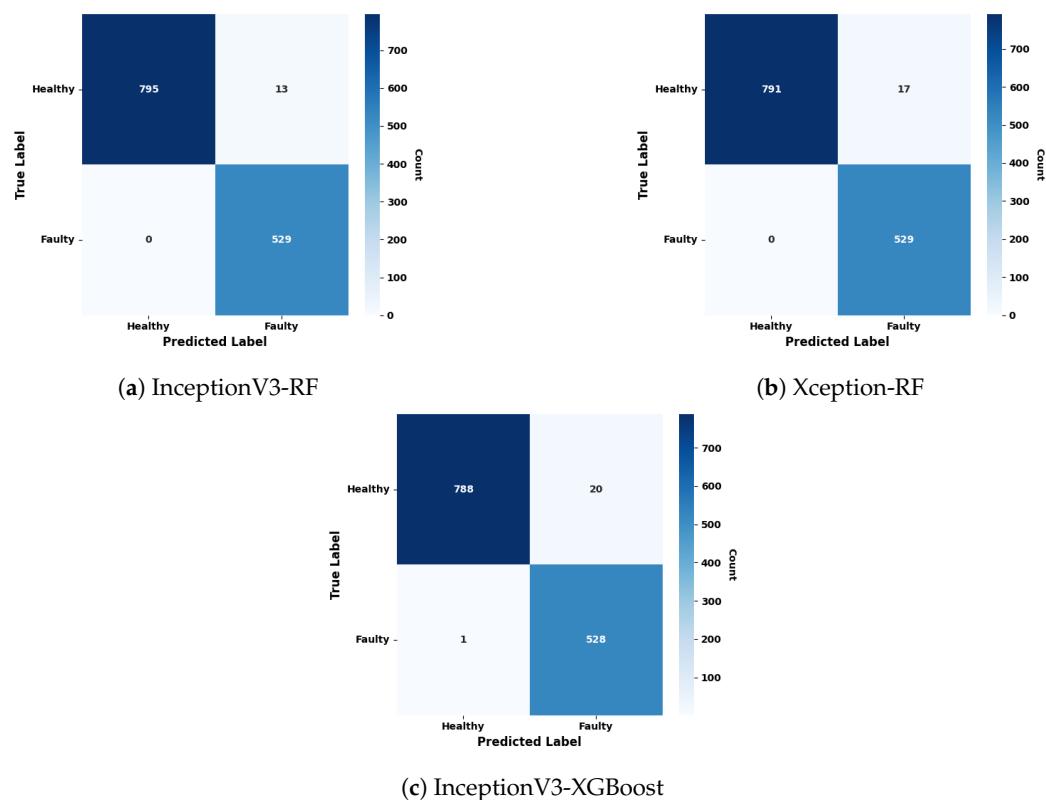


Figure 9. Confusion matrices for the three top-performing models from Table 1.

The Receiver Operating Characteristic (ROC) curves in Figure 10 summarize the performance of models trained on TrueDepth-derived data. The models using 3D geometric features exhibit strong and consistent discriminative power, with curves rising sharply toward the top-left corner and area under the curve (AUC) values exceeding 0.98 across classifiers. These results indicate that features extracted from TrueDepth point clouds provide a robust and separable representation of defects, enabling high true positive rates with low false alarm rates across different model–classifier combinations. The consistency of ROC behaviour across architectures suggests that the projected point cloud representation delivers classifier-agnostic signal quality suitable for reliable defect detection in challenging environments.

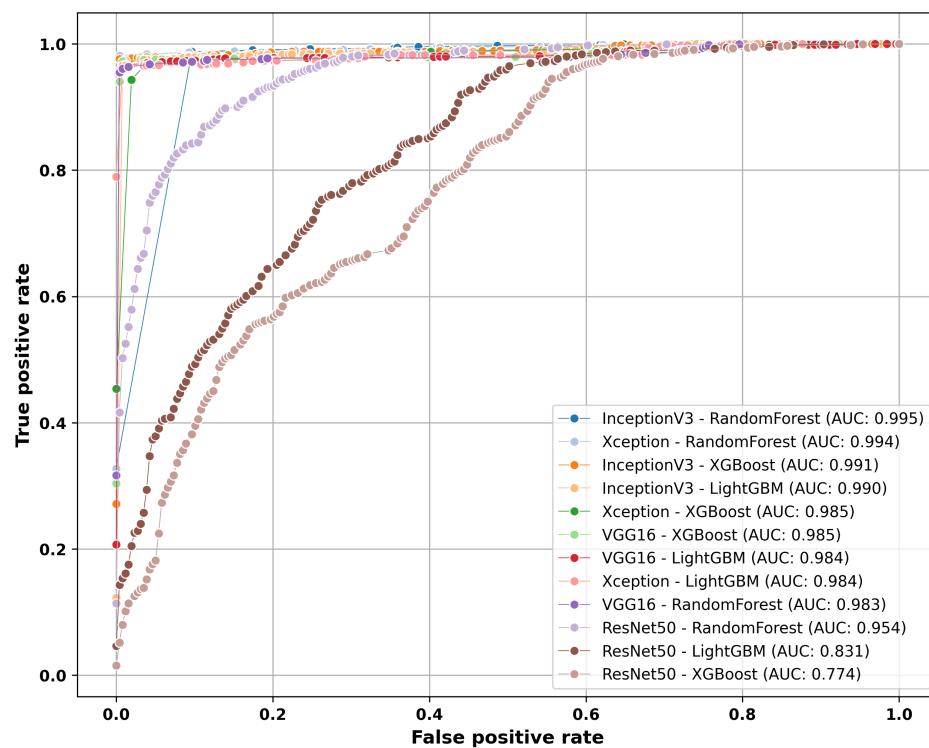


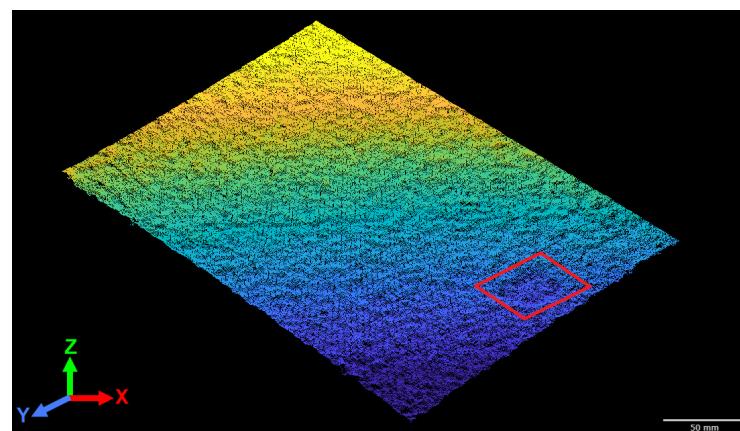
Figure 10. ROC curves for all studied models evaluated on TrueDepth camera data.

The results demonstrate that the TrueDepth camera captures superior data for identifying surface defects on conveyor belts. The underperformance of the model trained on RGB data is primarily due to false positives caused by surface textures such as permanent belt patches which were incorrectly classified as defects, as illustrated in Figure 11. In contrast, the point cloud data from the TrueDepth camera effectively excluded these patches from consideration, as their depth measured below 1 mm. Our model was specifically trained to recognize defects with a depth exceeding 2 mm, thereby ignoring superficial variations.

Furthermore, the 2D representation derived from the point cloud retains only depth-based anomalies, whereas RGB images contain all visual textures present on the belt surface. This additional complexity makes it significantly more challenging for a model to distinguish genuine defects from normal surface patterns. Consequently, using the 2D representation of point clouds substantially reduces the number of training samples required to achieve high performance. This is a crucial advantage in industrial environments, where acquiring large, accurately labeled datasets is often prohibitively expensive or logistically impractical.



(a) A belt patch (red box) captured by RGB camera



(b) A belt patch (red box) captured by TrueDepth camera

Figure 11. Comparison of RGB and TrueDepth camera sensors in the capturing of a belt patch (false positive case).

6.2. Real-World Accuracy Comparison for Defect Quantification

This section analyses the model outcomes and outlines how detected faults can be subjected to further quantitative evaluation using TrueDepth-derived geometry. By training on 2D projections of the point cloud, we substantially reduce computational complexity while retaining the geometric detail needed for reliable defect detection. Only samples classified as defective are forwarded for high-resolution, full-dimensional 3D reconstruction using the TrueDepth point clouds, producing detailed geometric models that support precise dimensional measurements and localization. Figure 12 illustrates the measured dimension of fault sample 2 obtained from the TrueDepth point cloud. This selective reconstruction strategy optimizes computational resources by avoiding expensive processing of non-defective data and delivers actionable, geometry-focused maintenance feedback to technicians.

To validate the quantitative accuracy of the proposed vision-based system, its measurements were benchmarked against a conventional manual method using a ruler and caliper as the baseline. A comprehensive comparison between the measured height, width, and depth of the faults obtained from TrueDepth 3D point clouds and the manual measurements is presented in Table 2.

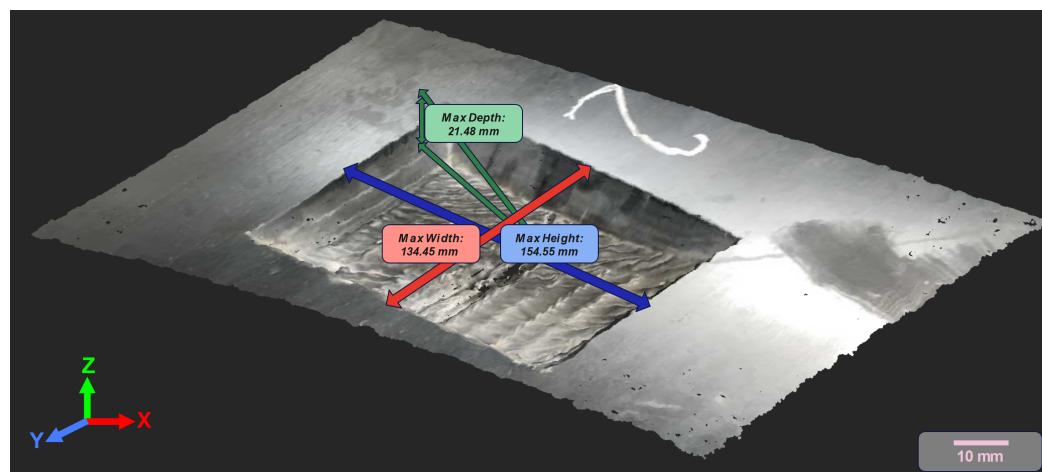


Figure 12. Quantitatively analyzing the defect sample 2 captured by the TrueDepth camera.

Table 2. Defect measurement results with absolute differences (mm) between Ruler & Caliper and TrueDepth camera measurements.

Defect Number	Ruler & Caliper			TrueDepth Camera			Absolute Difference		
	Height	Width	Depth	Height	Width	Depth	Height	Width	Depth
Defect 1	155	125	20	155	123	18	0	2	2
Defect 2	155	135	19	154	134	21	1	1	2
Defect 3	153	118	19	155	116	22	2	2	2
Defect 4	165	125	23	163	124	24	2	1	1
Defect 5	157	130	21	154	130	19	3	0	1
Defect 6	160	125	23	161	128	26	1	3	2
Defect 7	250	140	22	248	139	20	2	1	2
Defect 8	274	140	21	277	142	24	3	2	2
Defect 9	185	50	19	184	49	18	1	1	1

The comparative analysis revealed a high degree of concordance, with the measurement error for defect dimensions between the TrueDepth camera and the manual method remaining within 3 mm. This result confirms the system's reliability and precision in capturing key geometric parameters. Beyond replicating manual measurements, the TrueDepth camera offers a distinct advantage: the capability for near-real-time, quantitative assessment of complex defect morphology. While manual techniques struggle with the irregular, non-linear contours typical of surface damage, the system accurately determines the shape and calculates the actual surface area of a defect. This facilitates a more comprehensive damage assessment, including a preliminary estimation of material loss volume, derived from the product of the measured area and average depth. It is important to note that the accuracy of this volumetric estimate is contingent upon the slope and internal geometry of the surface defects.

7. Summary and Conclusions

The presented study successfully introduced and validated a novel smartphone-driven surface defect detection system framework for the inspection of industrial conveyor belt surfaces. Addressing the challenges of unreliable computer vision methods in harsh mining environments characterized by variable lighting and dust, the system uses the integrated TrueDepth camera of a commercial smartphone (iPhone 12 Pro Max) to simultaneously capture high-resolution visual data and precise 3D point clouds from a moving belt.

The presented methodology is based on a 3D-to-2D projection, which converts complex point cloud data into structured 2D representations. We employed a hybrid architecture where pre-trained CNNs (VGG16, ResNet50, InceptionV3, Xception) serve as deep feature extractors, followed by machine learning classifiers (Random Forest, XGBoost, LightGBM).

The InceptionV3-RF model, operating on geometric features, attained a high test F1 score of 0.9919 and maintained a near-perfect recall for the fault class. This capability is critical for operational safety as it prevents the occurrence of errors.

Additionally, the proposed methodology enables the quantitative assessment of surface damage. Comparative analysis against manual measurements confirmed the system's reliability, with measurement errors for defect dimensions remaining within 3 mm for point cloud-derived depth. Such accuracy allows the system to properly determine the complex morphology of defects and calculate the defect surface area and shape. This information allows for a robust system of maintenance with tracking of the belt condition through its full life cycle and, in turn, better prediction of the time of necessary intervention.

In conclusion, this research validates a reliable and cost-effective smartphone-based sensing platform that supports near-real-time maintenance decisions. We demonstrate the distinct advantage of the TrueDepth camera over conventional RGB imaging for capturing surface geometry, a capability that proves particularly suitable for low-light industrial conditions. This study successfully established the system's core effectiveness in a controlled laboratory setting, confirming its ability to detect geometric defects and generate accurate surface maps under simulated conditions.

Building on this validated foundation, the proposed methodology demonstrates significant potential to improve conveyor belt management, reduce maintenance costs, and enhance operational safety. In our future work, we plan to conduct field validation on an operational conveyor belt within a mining site to investigate the system's robustness against real-world challenges such as airborne particulates and mechanical vibration. Furthermore, we intend to explore data fusion techniques that integrate complementary information from both RGB and TrueDepth cameras. This multi-modal approach aims to generate a more comprehensive condition report, potentially providing supervisors with a richer, hybrid data stream for enhanced quality monitoring and decision-making in industrial settings.

Author Contributions: Conceptualisation, M.S.; methodology, M.S.; software, M.S.; validation, P.D., H.S.; formal analysis, P.D.; investigation, M.S., P.D., H.S. and R.Z.; resources, R.Z. and T.B.; data curation, M.S., P.D.; writing—original draft preparation, M.S., P.D. and H.S.; writing—review and editing, T.B. and R.Z.; visualisation, M.S.; supervision, R.Z. and T.B.; funding acquisition, R.Z. and T.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The measurement data presented in this study are not publicly available due to restrictions of privacy.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. He, D.; Pang, Y.; Lodewijks, G. Green operations of belt conveyors by means of speed control. *Appl. Energy* **2017**, *188*, 330–341. [\[CrossRef\]](#)
2. Siami, M.; Dąbek, P.; Shiri, H.; Michalak, A.; Wodecki, J.; Barszcz, T.; Zimroz, R. Heterogeneous Information Fusion for Robot-Based Automated Monitoring of Bearings in Harsh Environments via Ensemble of Classifiers with Dynamic Weighted Voting. *Sensors* **2025**, *25*, 5512. [\[CrossRef\]](#)
3. Siami, M.; Barszcz, T.; Wodecki, J.; Zimroz, R. Semantic segmentation of thermal defects in belt conveyor idlers using thermal image augmentation and U-Net-based convolutional neural networks. *Sci. Rep.* **2024**, *14*, 5748. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Siami, M.; Barszcz, T.; Wodecki, J.; Zimroz, R. Design of an Infrared Image Processing Pipeline for Robotic Inspection of Conveyor Systems in Opencast Mining Sites. *Energies* **2022**, *15*, 6771. [\[CrossRef\]](#)
5. Wodecki, J.; Shiri, H.; Siami, M.; Zimroz, R. Acoustic-based diagnostics of belt conveyor idlers in real-life mining conditions by mobile inspection robot. In Proceedings of the Conference on Noise and Vibration Engineering, ISMA 2022, Leuven, Belgium, 12–14 September 2022.

6. Trybała, P.; Blachowski, J.; Błażej, R.; Zimroz, R. Damage detection based on 3d point cloud data processing from laser scanning of conveyor belt surface. *Remote Sens.* **2021**, *13*, 55. [[CrossRef](#)]
7. Bajda, M.; Hardygóra, M. Examination and assessment of the impact of working conditions on operating parameters of selected conveyor belts. *Min. Sci.* **2022**, *29*, 165–178. [[CrossRef](#)]
8. Bajda, M. Predicting the working time of multi-ply conveyor belt splices in underground mines. *Min. Sci.* **2024**, *31*, 259–274. [[CrossRef](#)]
9. Andrejiova, M.; Grincova, A.; Marasova, D. Measurement and simulation of impact wear damage to industrial conveyor belts. *Wear* **2016**, *368–369*, 400–407. [[CrossRef](#)]
10. Hakami, F.; Pramanik, A.; Ridgway, N.; Basak, A. Developments of rubber material wear in conveyor belt system. *Tribol. Int.* **2017**, *111*, 148–158. [[CrossRef](#)]
11. Wang, R.; Li, Y.; Yang, F.; Wang, Z.; Dong, J.; Yuan, C.; Lu, X. PLC based laser scanning system for conveyor belt surface monitoring. *Sci. Rep.* **2024**, *14*, 27914. [[CrossRef](#)]
12. Huang, Q.; Pan, C.; Liu, H. A Multi-sensor Fusion Algorithm for Monitoring the Health Condition of Conveyor Belt in Process Industry. In Proceedings of the 2021 3rd International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 8–11 November 2021; pp. 1–6. [[CrossRef](#)]
13. Zhao, H.S.; Wu, Y.Z. A New Roller Conveyor Driven by Linear Motor. *Adv. Mater. Res.* **2011**, *201–203*, 1517–1520. [[CrossRef](#)]
14. Wang, J.; Miao, C.; Wang, W.; Lu, X. Research of X-ray nondestructive detector for high-speed running conveyor belt with steel wire ropes. In Proceedings of the Electronic Imaging and Multimedia Technology V, Beijing, China, 12–15 November 2007; SPIE: Bellingham, WA, USA, 2007; Volume 6833, pp. 482–490.
15. D’Ettorre, G.; Farronato, M.; Candida, E.; Quinzi, V.; Grippaudo, C. A comparison between stereophotogrammetry and smartphone structured light technology for three-dimensional face scanning. *Angle Orthod.* **2022**, *92*, 358–363. [[CrossRef](#)] [[PubMed](#)]
16. Catalucci, S.; Thompson, A.; Piano, S.; Branson, D.T., III; Leach, R. Optical metrology for digital manufacturing: A review. *Int. J. Adv. Manuf. Technol.* **2022**, *120*, 4271–4290. [[CrossRef](#)]
17. Nilneree, S.; Mizutani, T. An innovative framework for incorporating iPhone LiDAR point cloud in digitized documentation of road operations. *Results Eng.* **2025**, *25*, 103953. [[CrossRef](#)]
18. Alfaro-Santafé, J.; Gómez-Bernal, A.; Lanuza-Cerzóximo, C.; Alfaro-Santafé, J.V.; Pérez-Morcillo, A.; Almenar-Arasanz, A.J. Three-axis measurements with a novel system for 3D plantar foot scanning: iPhone X. *Footwear Sci.* **2020**, *12*, 123–131. [[CrossRef](#)]
19. Carey, N.; Werfel, J.; Nagpal, R. Fast, accurate, small-scale 3D scene capture using a low-cost depth sensor. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 27–29 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1268–1276.
20. Andrews, J.; Alwafi, A.; Bichu, Y.M.; Pliska, B.T.; Mostafa, N.; Zou, B. Validation of three-dimensional facial imaging captured with smartphone-based photogrammetry application in comparison to stereophotogrammetry system. *Helijon* **2023**, *9*, e15834. [[CrossRef](#)]
21. Pellitteri, F.; Scisciola, F.; Cremonini, F.; Baciliero, M.; Lombardo, L. Accuracy of 3D facial scans: A comparison of three different scanning system in an in vivo study. *Prog. Orthod.* **2023**, *24*, 44. [[CrossRef](#)] [[PubMed](#)]
22. Yang, J.; Lee, C.; Ahn, P.; Lee, H.; Yi, E.; Kim, J. PBP-Net: Point Projection and Back-Projection Network for 3D Point Cloud Segmentation. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 8469–8475. [[CrossRef](#)]
23. Ericeira, D.R.; Rocha, F.; Bianchi, A.G.; Pessin, G. Early failure detection of belt conveyor idlers by means of ultrasonic sensing. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–8.
24. Kirjanów-Błażej, A.; Błażej, R.; Jurdziak, L.; Kozłowski, T.; Rzeszowska, A. Innovative diagnostic device for thickness measurement of conveyor belts in horizontal transport. *Sci. Rep.* **2022**, *12*, 7212. [[CrossRef](#)]
25. Rzeszowska, A.; Jurdziak, L.; Błażej, R.; Kirjanów-Błażej, A. Application of Clustering and SOM Analysis for Identification of Conveyor Belt Damage Based on Data from the Diagbelt+ Magnetic System. In Proceedings of the International Conference on Intelligent Systems in Production Engineering and Maintenance, Wrocław, Poland, 13–15 September 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 461–475.
26. Siami, M.; Barszcz, T.; Wodecki, J.; Zimroz, R. Automated ir image segmentation for identification of overheated idlers in belt conveyor systems based on outlier-detection method. *SSRN* **2022**. [[CrossRef](#)]
27. Siami, M.; Barszcz, T.; Zimroz, R. Advanced Image Analytics for Mobile Robot-Based Condition Monitoring in Hazardous Environments: A Comprehensive Thermal Defect Processing Framework. *Sensors* **2024**, *24*, 3421. [[CrossRef](#)]
28. Liu, Y.; Miao, C.; Li, X.; Ji, J.; Meng, D. Research on the fault analysis method of belt conveyor idlers based on sound and thermal infrared image features. *Measurement* **2021**, *186*, 110177. [[CrossRef](#)]

29. Bortnowski, P.; Gondek, H.; Król, R.; Marasova, D.; Ozdoba, M. Detection of Blockages of the Belt Conveyor Transfer Point Using an RGB Camera and CNN Autoencoder. *Energies* **2023**, *16*, 1666. [[CrossRef](#)]

30. Sun, L.; Sun, X. An Intelligent Detection Method for Conveyor Belt Deviation State Based on Machine Vision. *Math. Model. Eng. Probl.* **2024**, *11*, 1257–1264. [[CrossRef](#)]

31. Guo, X.; Liu, X.; Zhou, H.; Stanislawski, R.; Królczyk, G.; Li, Z. Belt tear detection for coal mining conveyors. *Micromachines* **2022**, *13*, 449. [[CrossRef](#)]

32. Banerjee, P.; Chakravarty, D.; Samanta, B. Design of a laboratory scale automatic optical inspection prototype system for scanning of conveyor belt surfaces—A case study. *Measurement* **2023**, *220*, 113342. [[CrossRef](#)]

33. Skoczylas, A.; Stefaniak, P.; Anufriev, S.; Jachnik, B. Belt conveyors rollers diagnostics based on acoustic signal collected using autonomous legged inspection robot. *Appl. Sci.* **2021**, *11*, 2299. [[CrossRef](#)]

34. Zhang, Y.; Li, S.; Li, A.; Zhang, G.; Wu, M. Fault diagnosis method of belt conveyor idler based on sound signal. *J. Mech. Sci. Technol.* **2023**, *37*, 69–79. [[CrossRef](#)]

35. Li, X.G.; Miao, C.Y.; Wang, J.; Zhang, Y. Automatic defect detection method for the steel cord conveyor belt based on its X-ray images. In Proceedings of the 2011 International Conference on Control, Automation and Systems Engineering (CASE), Jeju, Republic of Korea, 8–10 December 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1–4.

36. Błażej, R.; Jurdziak, L.; Kozłowski, T.; Kirjanów, A. The use of magnetic sensors in monitoring the condition of the core in steel cord conveyor belts—Tests of the measuring probe and the design of the DiagBelt system. *Measurement* **2018**, *123*, 48–53. [[CrossRef](#)]

37. Kirjanów-Błażej, A.; Jurdziak, L.; Burduk, R.; Błażej, R. Forecast of the remaining lifetime of steel cord conveyor belts based on regression methods in damage analysis identified by subsequent DiagBelt scans. *Eng. Fail. Anal.* **2019**, *100*, 119–126. [[CrossRef](#)]

38. Olchówka, D.; Błażej, R.; Jurdziak, L. Selection of measurement parameters using the DiagBelt magnetic system on the test conveyor. *J. Phys. Conf. Ser.* **2022**, *2198*, 012042. [[CrossRef](#)]

39. Xu, S.; Cheng, G.; Pang, Y.; Jin, Z.; Kang, B. Identifying and characterizing conveyor belt longitudinal rip by 3D point cloud processing. *Sensors* **2021**, *21*, 6650. [[CrossRef](#)] [[PubMed](#)]

40. Zheng, X.; Wang, J.; Liu, K.; Niu, F.; Zhang, D.; Lin, Y. Real-time detection of multi-wedge belt tooth surface defects based on 3D laser technology. *Eng. Res. Express* **2025**, *7*, 025286. [[CrossRef](#)]

41. Palanisamy, V.V.; Venkatachalam, S. LiDAR-based temporal surface damage assessment of bridge infrastructure using efficient scan area planning. *Autom. Constr.* **2025**, *180*, 106526. [[CrossRef](#)]

42. Aijazi, A.; Malaterre, L.; Tazir, M.; Trassoudaine, L.; Checchin, P. Detecting and analyzing corrosion spots on the hull of large marine vessels using colored 3D lidar point clouds. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 153–160.

43. Tan, Y.; Li, S.; Wang, Q. Automated geometric quality inspection of prefabricated housing units using BIM and LiDAR. *Remote Sens.* **2020**, *12*, 2492. [[CrossRef](#)]

44. Mikita, T.; Krausková, D.; Hrůza, P.; Cibulka, M.; Patočka, Z. Forest road wearing course damage assessment possibilities with different types of laser scanning methods including new iPhone LiDAR scanning apps. *Forests* **2022**, *13*, 1763. [[CrossRef](#)]

45. Jakubiak, J.; Delicat, J. Fast Detection of Idler Supports Using Density Histograms in Belt Conveyor Inspection with a Mobile Robot. *Appl. Sci.* **2024**, *14*, 10774. [[CrossRef](#)]

46. Hao, X.I.; Liang, H. A multi-class support vector machine real-time detection system for surface damage of conveyor belts based on visual saliency. *Measurement* **2019**, *146*, 125–132. [[CrossRef](#)]

47. Mao, Q.; Ma, H.; Zhang, X.; Zhang, G. An improved skewness decision tree svm algorithm for the classification of steel cord conveyor belt defects. *Appl. Sci.* **2018**, *8*, 2574. [[CrossRef](#)]

48. Li, X.; Li, Y.; Zhang, Y.; Liu, F.; Fang, Y. Fault diagnosis of belt conveyor based on support vector machine and grey wolf optimization. *Math. Probl. Eng.* **2020**, *2020*, 1367078. [[CrossRef](#)]

49. Zhang, M.; Zhang, Y.; Zhou, M.; Jiang, K.; Shi, H.; Yu, Y.; Hao, N. Application of lightweight convolutional neural network for damage detection of conveyor belt. *Appl. Sci.* **2021**, *11*, 7282. [[CrossRef](#)]

50. Kirjanów-Błażej, A.; Rzeszowska, A. Conveyor belt damage detection with the use of a two-layer neural network. *Appl. Sci.* **2021**, *11*, 5480. [[CrossRef](#)]

51. Netto, G.G.; Coelho, B.N.; Delabrida, S.E.; Sinatra, A.; Azpúrua, H.; Pessin, G.; Oliveira, R.A.; Bianchi, A.G. Early Defect Detection in Conveyor Belts using Machine Vision. In Proceedings of the VISIGRAPP (4: VISAPP), Virtual, 8–10 February 2021; SCITEPRESS—Science and Technology Publications, Lda.: Setúbal, Portugal, 2021; pp. 303–310.

52. Wu, X.; Wang, C.; Tian, Z.; Huang, X.; Wang, Q. Research on belt deviation fault detection technology of belt conveyors based on machine vision. *Machines* **2023**, *11*, 1039. [[CrossRef](#)]

53. Zhang, P.; Xu, S.; Wang, W. Belt deviation detection system based on deep learning under complex working conditions. *IAENG Int. J. Appl. Math.* **2023**, *53*, 92–97.

54. Zhang, M.; Jiang, K.; Cao, Y.; Li, M.; Hao, N.; Zhang, Y. A deep learning-based method for deviation status detection in intelligent conveyor belt system. *J. Clean. Prod.* **2022**, *363*, 132575. [[CrossRef](#)]

55. Guo, X.; Liu, X.; Królczyk, G.; Sulowicz, M.; Glowacz, A.; Gardoni, P.; Li, Z. Damage detection for conveyor belt surface based on conditional cycle generative adversarial network. *Sensors* **2022**, *22*, 3485. [\[CrossRef\]](#)

56. Liu, M.; Zhu, Q.; Yin, Y.; Fan, Y.; Su, Z.; Zhang, S. Damage detection method of mining conveyor belt based on deep learning. *IEEE Sensors J.* **2022**, *22*, 10870–10879. [\[CrossRef\]](#)

57. Breitbarth, A.; Schardt, T.; Kind, C.; Brinkmann, J.; Dittrich, P.G.; Notni, G. Measurement accuracy and dependence on external influences of the iPhone X TrueDepth sensor. In Proceedings of the Photonics and Education in Measurement Science, Jena, Germany, 17–19 September 2019; International Society for Optics and Photonics, SPIE: Bellingham, WA, USA, 2019; Volume 11144, p. 1114407. [\[CrossRef\]](#)

58. Beltrán, N.; Álvarez, J.; Fernández, A.; Álvarez, B.J.; Blanco, D. Performance analysis of Apple’s TrueDepth sensor for surface reconstruction of regular geometries across varying stand-off distances. *Int. J. Adv. Manuf. Technol.* **2025**, *1–9*. [\[CrossRef\]](#)

59. Vogt, M.; Rips, A.; Emmelmann, C. Comparison of iPad Pro®’s LiDAR and TrueDepth capabilities with an industrial 3D scanning solution. *Technologies* **2021**, *9*, 25. [\[CrossRef\]](#)

60. Urban, S.; Lindemeier, T.; Dobbelstein, D.; Haenel, M. On the Issues of TrueDepth Sensor Data for Computer Vision Tasks Across Different iPad Generations. *arXiv* **2022**, arXiv:2201.10865.

61. Tang, R.; Deng, B.; Li, J.; Yan, Y. An Efficient and Flexible Solution for Camera Autocalibration from $N \geq 3$ Views. *Math. Probl. Eng.* **2018**, *2018*, 8348340. [\[CrossRef\]](#)

62. Lindsay, G.W. Convolutional neural networks as a model of the visual system: Past, present, and future. *J. Cogn. Neurosci.* **2021**, *33*, 2017–2031. [\[CrossRef\]](#)

63. Siami, M.; Barszcz, T.; Wodecki, J.; Zimroz, R. Automated Identification of Overheated Belt Conveyor Idlers in Thermal Images with Complex Backgrounds Using Binary Classification with CNN. *Sensors* **2022**, *22*, 4. [\[CrossRef\]](#)

64. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)

65. Mostafavi, A.; Siami, M.; Friedmann, A.; Barszcz, T.; Zimroz, R. Probabilistic uncertainty-aware decision fusion of neural network for bearing fault diagnosis. *PHM Soc. Eur. Conf.* **2024**, *8*, 10.

66. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 248–255. [\[CrossRef\]](#)

67. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [\[CrossRef\]](#)

68. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778. [\[CrossRef\]](#)

69. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–9. [\[CrossRef\]](#)

70. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1251–1258. [\[CrossRef\]](#)

71. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)

72. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.

73. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3149–3157.

74. Nissen, L.; Hübner, J.; Klinker, J.; Kapsecker, M.; Leube, A.; Schneckenburger, M.; Jonas, S.M. Towards Preventing Gaps in Health Care Systems through Smartphone Use: Analysis of ARKit for Accurate Measurement of Facial Distances in Different Angles. *Sensors* **2023**, *23*, 4486. [\[CrossRef\]](#)

75. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 2564–2571. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.