

# Improved Estimation of Soil-to-Sugar Beet Ratio Using Deep Learning and Domain Adaptation for Precision Agriculture

Elham Alfuqara  
ORCID: 0009-0002-1971-7231  
Technische Hochschule Deggendorf  
Cham, Germany  
Email: elham.al-fuqara@th-deg.de

Prof. Dr. Matthias Hien  
ORCID: 0009-0009-1664-8948  
Technische Hochschule Deggendorf  
Cham, Germany  
Email: matthias.hien@th-deg.de

Johann Brunner  
ORCID: 0009-0000-4184-9797  
Technische Hochschule Deggendorf  
Cham, Germany  
Email: johann.brunner@th-deg.de

Tobias Dobmeier  
ORCID: 0009-0003-1864-5122  
Technische Hochschule Deggendorf  
Cham, Germany  
Email: tobias.dobmeier@th-deg.de

Raphael Roeser-Mueller  
ORCID: 0009-0006-9027-8998  
Technische Hochschule Deggendorf  
Cham, Germany  
Email: raphael.roeser-mueller@th-deg.de

**Abstract**—Accurate estimation of soil adhering to sugar beets is important for both sugar factories and farmers. Since weighing each truckload before and after cleaning is impractical, soil tare is typically estimated visually by trained evaluators. However, Gage R&R and accuracy analyses revealed high variability in these manual assessments, highlighting the need for a more reliable and objective solution. This study proposes an automated pipeline for soil-to-beet ratio estimation while addressing two major challenges: the absence of image-level ground truth field data and the difficulty of manually annotating soil regions. The pipeline combines sugar beet segmentation using Detectron2 with soil segmentation based on intensity and texture analysis. Extracted features are then used to train a Semi-Supervised Domain-Adversarial Neural Network (SS-DANN) to handle domain shifts between laboratory and field datasets. The proposed approach achieved a deviation of 0.39% from ground truth, outperforming human estimators with an average deviation of 1.91%.

## I. INTRODUCTION

**A** FAIR and objective payment system for farmers is essential to ensure that the assessments accurately reflect both the quantity and quality of sugar beets delivered. One key factor in this process is the relative amount of soil adhering to sugar beet after harvest, commonly referred to as **soil tare** [1]. Currently, soil tare is estimated visually by human estimators during the loading process at sugar factories. However, this method relies heavily on subjective judgment, which introduces variability due to differences in individual assessment practices. As a result, there is growing interest in developing more reliable and standardized approaches to soil tare evaluation.

Beyond its relevance to quality control and operational logistics, soil tare presents substantial economic and environmental challenges throughout the sugar beet processing chain. Excess soil adhering to harvested beets accelerates mechanical

wear on processing equipment, increases water consumption during washing procedures, and contributes to elevated costs associated with the handling, transport, and disposal of residual soil. Environmentally, the impact extends beyond the processing facility: the unintentional removal of millions of tons of topsoil annually leads to long-term degradation of soil health in agricultural fields and increases greenhouse gas emissions due to the added weight during transportation. Between 2000 and 2016, the Evolution of Soil Loss by Crop Harvesting (SLCH) for sugar beet reached 1,663.3 thousand tons in Germany and 9,589.2 thousand tons across Europe [2], emphasizing the scale of the problem at a continental level.

This large-scale impact is further illustrated by regional figures that reveal substantial variability in soil tare depending on local conditions and practices. In Konya, Turkey, the average soil tare was estimated at 8% in 2000, corresponding to over 256,000 tons of soil transported annually with the harvested crop [3]. In the Italia Zuccheri districts, reported values ranged from 2% to 10% between 1995 and 2008 [4]. In France, current levels remain around 10%, despite a 65% reduction achieved over the past 25 years [4]. The amount of tare soil typically ranges from 2% to over 8% [5], depending on many variations.

Soil tare is influenced by several interdependent factors, including soil type, weather conditions during harvest, harvester type, harvesting method, and loading techniques. These variations highlight the need for standardized and automated assessment methods to improve accuracy, fairness, and sustainability.

The sugar beet handling process from field to factory is illustrated in Fig. 1. During harvesting, specialized machines extract the beets from the soil and partially clean them using rollers, turbines, or vibratory systems. However, a considerable

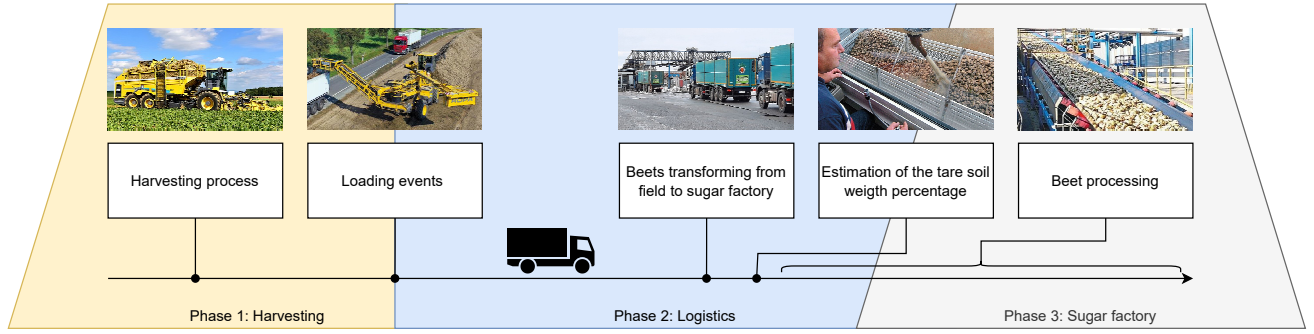


Fig. 1. An overview of the sugar beet supply chain process from harvesting to processing. The figure illustrates the key phases including beet harvesting, logistics, and beet processing at the sugar factory.

amount of soil and debris often remains attached. The beets are then transported to the factory, where trucks are weighed and visually inspected by human estimators to assess soil tare before the beets undergo further cleaning and sugar production.

This study focuses on three main objectives. First, a custom data acquisition system is developed to capture truckload images and related information during loading. Second, the reliability of manual soil tare estimation is evaluated. Third, deep learning models are trained and validated to predict soil tare at the image level. The proposed approach aims to provide an objective and scalable alternative to subjective manual estimation methods.

## II. RELATED WORK

In recent years, computer vision has become increasingly important in precision agriculture, particularly for tasks such as plant detection, weed classification, disease identification, and post-harvest crop assessment. A wide range of image analysis techniques (from classical thresholding to modern deep learning models) have been explored to support automation and quality control in crop production.

Early vision-based systems for sugar beet fields relied on classical techniques. For instance, Bakker et al. implemented a Hough-transform-based method to detect plant rows using grayscale imagery [6]. Other approaches combined vegetation indices with color and low near-infrared (NIR) imagery for plant and soil discrimination [7]. More advanced methods used pixel-based discriminant analysis in different color spaces to handle variations in lighting when segmenting weeds from sugar beets [8].

The rise of deep learning brought significant improvements in both detection and classification. Gao et al. trained a YOLOv3-based model to detect *Convolvulus sepium* weeds and sugar beet plants, showing high detection accuracy, especially when augmented with synthetic data [9]. Similarly, CNN-based models have been used for identifying sugar beet leaf diseases [10] and estimating sucrose content from beet slices using NIR spectroscopy [11].

In terms of semantic segmentation, recent works employed deep neural networks like U-Net with ResNet encoders to

distinguish sugar beets, weeds, and soil under varying conditions [12]. Nasiri et al. applied a U-Net variant to perform pixel-level weed segmentation in sugar beet fields, using over 1,300 annotated images [13]. These models demonstrated strong performance but were mostly limited to plant vs weed discrimination, with limited focus on soil or post-harvest conditions.

There has also been a growing interest in object detection for mechanical damage assessment in sugar beets. Nasirahmadi et al. used YOLOv4 and CSPDarknet53 to detect cracks, abrasions, and breakages incurred during harvesting, treating all damage types as a unified class [14]. However, these studies generally lack fine-grained annotations for issues like soil adhesion, rot, or topping defects.

The SemanticSugarBeets framework introduced by Croonen et al. addresses many of these gaps by proposing a comprehensive dataset and a multi-task pipeline for inspecting sugar beets after harvest and storage [15]. Their approach uses YOLOv11 for instance segmentation, followed by fine-grained semantic segmentation using U-Net, MANet, and PSPNet architectures with various encoders. The dataset includes over 2,900 annotated beets across three stages (Sample, Harvest, Storage), with labels for soil adhesion, damage, rot, and leaf remnants. The segmentation model achieved a mean IoU of 64.0% and highlighted the difficulty of detecting soil in wet or low light conditions which are challenges also relevant to this study. In comparison to this study, our approach pursues a different objective which is estimating the soil tare with providing detailed semantic segmentation for both beets and adhering soil. While SemanticSugarBeets focuses on visual quality assessment through fine-grained labeling, our method targets the direct prediction of a measurable, task-specific metric.

## III. MATERIAL AND METHODS

To develop an objective and robust method for soil tare estimation, two complementary datasets and a structured processing pipeline were established to address both field variability and controlled laboratory conditions.

A custom data acquisition system was integrated into the conveyor belt of a sugar beet loading machine to collect

operational field data. The system includes a JAI FS-3200T-10GE-NNC multi-spectral prism camera [16], capable of simultaneously capturing visible, low near-infrared (NIR), and high near-infrared spectral channels. Additional CAN bus modules recorded machine parameters such as conveyor speed, acceleration, and inclination. To ensure reliable outdoor operation, the camera was installed in a protective housing together with a dedicated illumination system. All image and sensor data were synchronized and stored on an industrial computer.

Although RGB, low NIR, and high NIR images were captured, only the low NIR channel was used for the **field dataset**, as it provided the clearest contrast between soil and beet surfaces. Approximately 250,000 low NIR images were collected from 540 loading events. While this dataset reflects real operational conditions, it did not contain image-level ground truth labels. Instead, only truck-level soil tare estimates provided by human estimators were available.

To evaluate the reliability of manual soil tare estimation, a Gage Repeatability and Reproducibility (Gage R&R) study [17], [18] was conducted using 50 field images. Four trained estimators evaluated each image three times, resulting in a Gage R&R value of 77.3%. According to industrial standards, values above 30% indicate that the measurement system is not sufficiently reliable for consistent decision-making [19].

To further assess estimator accuracy, a **ground truth dataset** was created under controlled laboratory conditions. Twenty sugar beets were halved, resulting in 40 individual samples. Each sample was imaged before and after cleaning under consistent lighting conditions while being rotated to capture different orientations. The true soil tare values were calculated from the weight difference before and after cleaning. In total, 3,824 dirty and 478 clean images were collected. A subset of 80 dirty images was then evaluated by the same four estimators and compared against the true values. The analysis showed an average deviation of 1.91%, with relative errors exceeding 30% in some cases.

The observed variability and limitations of manual estimation motivated the development of an automated processing pipeline (Fig. 2). The pipeline consists of four stages. First, an instance segmentation model generates a beet mask to isolate the sugar beet region. Second, a soil mask is created within this region using image processing methods based on intensity and texture information. Third, texture and intensity features are extracted from the detected soil regions to characterize soil thickness variations. Finally, these features are used to train an SS-DANN model for soil tare estimation. Further details of each stage are presented in the following sections.

#### IV. SOIL TARE ESTIMATION PIPELINE

##### A. Step 1: Sugar Beet Segmentation

The first step of the processing pipeline focuses on isolating sugar beets from the surrounding background to enable accurate analysis of the soil tare. Reliable segmentation is essential to eliminate interference from non-target elements

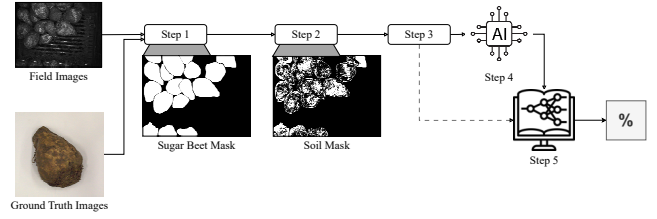


Fig. 2. The diagram illustrates the soil tare prediction pipeline, summarizing the sequence from images to final soil tare prediction

such as loose soil, conveyor structures, soil attached to the conveyor, and shadows.

To achieve this, instance segmentation was performed using the Detectron2 framework [20], configured with the Mask R-CNN model architecture incorporating a ResNet-50 backbone and Feature Pyramid Network (FPN) for multi-scale feature extraction. For model training, a representative subset of 1,095 low NIR images was selected from the field dataset. These images were semi-automatically annotated, resulting in 35,365 individual sugar beet instances. The annotated subset was then divided into 70% for training and 30% for testing. Training was carried out using a batch size of 2, a learning rate of 0.00025, and a maximum of 6,000 iterations.

The trained model demonstrated robust performance, achieving an average precision (AP) of 84.37% on the test set. This confirmed its effectiveness in separating sugar beets from complex backgrounds under real-world loading conditions.

Afterwards, the trained model was applied to generate beet masks for two additional image sets:

- 1) **Field Subset:** This set included 5,000 low NIR images sampled from 10 distinct truck loadings (500 images per loading event). Each loading was labeled by human estimators with an overall soil tare ranging from 4% to 13%, reflecting load-level estimations rather than per-image labels.
- 2) **Ground Truth Dataset:** This set consisted of 4,302 images of individual sugar beets with precisely measured soil tare, ranging from 0% to 23.33%, providing fine-grained, per-image ground truth.

These sets together with the beet masks generated in this step served as the foundation for the subsequent analysis. They were first used to restrict soil evaluation to the beet surface, ensuring that only soil tare was considered in later steps. In addition, the masks were used to compute the total beet pixel area per image by summing the number of pixels identified as sugar beet. This value became the first feature added to the Dual Domain Regression Dataset (DDRD) which is a structured set of features that would be used to train the SS-DANN model introduced in Step 4.

The DDRD dataset is constructed incrementally, with new features extracted and appended at each stage of the pipeline.

### B. Step 2: Soil Segmentation

Following sugar beet segmentation, the next step focuses on isolating soil adhered to the beet surface. Accurate soil segmentation is essential for reliable feature extraction and soil tare prediction. However, manual annotation was impractical due to the complexity of the dataset, where multiple overlapping beets make consistent labeling difficult and time-consuming. Therefore, a fully automated image processing approach was adopted.

Soil segmentation is performed adaptively within the beet regions identified in Step 1 by combining intensity- and texture-based information. First, a **Gaussian Mixture Model (GMM)** is applied to contrast-enhanced intensity images to detect darker regions typically associated with soil. Second, a **texture-based segmentation** is performed using Otsu's thresholding on Local Binary Pattern (LBP) outputs [21], [22]. This enables automatic threshold selection and effectively distinguishes the granular texture of soil from the smoother beet surface under varying lighting and soil conditions.

The final soil mask is obtained by intersecting the GMM and LBP segmentation results, ensuring that only regions fulfilling both intensity and texture criteria are classified as soil (see Fig. 3). Morphological operations are then applied to remove noise and improve mask continuity.

The generated soil masks are used to calculate the total soil pixel area for each image, which is stored as a feature in the DDRD. These masks also provide the basis for the following feature extraction stage.

### C. Step 3: Soil Feature Extraction

After segmenting the beet and soil regions in the previous steps, this step focuses on extracting meaningful features that describe the properties of the soil tare. These features are essential for distinguishing between different soil conditions, such as thin dust layers versus thick clumps, and for enabling accurate estimation of soil tare values.

In this step, **rotation-invariant soil features** [21] are extracted from the images in the DDRD, specifically from the **soil areas** identified by the binary soil masks. Using rotation-invariant features ensures that measurements remain consistent even when sugar beets appear in different orientations. This is particularly important for the ground truth dataset, where

individual beets are captured multiple times from various angles.

The selected features include granulometry [23], intensity-based statistics, texture descriptors based on LBP and Gray Level Co-occurrence Matrix (GLCM) [24], edge magnitude, and frequency-based features such as wavelet energy and Gabor responses [25]. Together, these features provide a comprehensive representation of the soil's appearance, structure, and texture. A detailed summary of the extracted features, their definitions, and the corresponding soil properties they reflect is presented in Table 1.

For each image in the DDRD dataset, the extracted features were combined with the previously computed soil pixel area and sugar beet pixel area. These were compiled into a structured Excel sheet, forming a complete feature vector for each image. This dataset served as the input for training the SS-DANN model described in Step 4.

### D. Step 4: SS-DANN model training

The final step in the pipeline focuses on learning a Deep learning model to estimate soil tare values from the extracted image features. To address the challenge of domain shift between controlled laboratory images and field data collected under operational conditions, a SS-DANN was employed.

As described in Steps 1 through 3, each image was transformed into a structured feature vector, forming the DDRD dataset. This dataset aggregates key features including total beet Pixel area, total soil Pixel area, and multiple soil texture descriptors. To finalize the dataset and ensure comparability across different acquisition setups, a normalization step was applied. Specifically, pixel-based measurements (beet and soil areas) were divided by the resolution of their corresponding images. This adjustment compensates for sensor differences and ensures that the model learns content-based representations rather than being biased by image resolution or size.

The resulting normalized and structured DDRD dataset was used to train the SS-DANN model. Training was performed for up to 300 epochs with early stopping based on the mean absolute error (MAE) on a validation subset of the Ground Truth data. The architecture of the proposed SS-DANN model, along with the domain adaptation strategy, is detailed in the following section.

## V. DOMAIN ADAPTATION AND SS-DANN

Conventional machine and deep learning models assume that training and testing data follow the same distribution. In practice, however, this assumption is often violated due to domain shift, where differences in imaging conditions, sensors, resolution, or environmental factors lead to changes between datasets [26], [27]. Such variations can significantly reduce model performance when applied to unseen data.

In this work, two different domains are considered: a **source domain** containing laboratory-acquired high-resolution RGB images with accurate soil tare measurements, and a **target domain** consisting of field-acquired low NIR images labeled through human visual estimation. The substantial differences

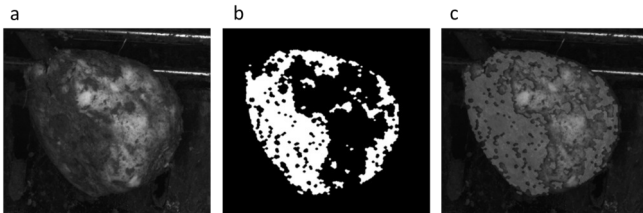


Fig. 3. (a) Sample sugar beet from a field-captured infrared image. (b) Corresponding binary soil mask generated using the proposed algorithm (white regions indicate soil). (c) Overlay of the soil mask on the original image to visually illustrate the segmentation outcome.

TABLE I  
SOIL FEATURE ANALYSIS AND IMPLICATIONS

| Feature            | Measured Aspect                    | Soil Indicator                                                             |
|--------------------|------------------------------------|----------------------------------------------------------------------------|
| Granulometry       | Multiscale structural distribution | Describes the prevalence of soil structures at different spatial scales.   |
| Mean Intensity     | Average pixel intensity            | Higher: Brighter soil region. Lower: Darker soil region.                   |
| Median Intensity   | Median pixel intensity             | Higher: Generally brighter pixels. Lower: Generally darker pixels.         |
| IQR Intensity      | Intensity variability              | Higher: Greater intensity variation. Lower: More uniform intensity.        |
| LBP Histogram      | Local texture patterns             | Distribution indicates the prevalence of different local texture patterns. |
| GLCM Contrast      | Local intensity differences        | Higher: Stronger local intensity variation. Lower: More uniform texture.   |
| Edge Magnitude     | Local intensity-gradient strength  | Higher: Stronger or more frequent edges. Lower: Smoother regions.          |
| Wavelet Energy     | Multiscale intensity variation     | Higher: Stronger multiscale variations. Lower: More uniform structure.     |
| Gabor Response Max | Oriented texture response          | Higher: Stronger directional texture. Lower: Weaker directional texture.   |

in image characteristics and label quality introduce a domain shift that must be addressed to achieve reliable field predictions.

To overcome this challenge, a Semi-Supervised Domain-Adversarial Neural Network (SS-DANN) is proposed. Unlike conventional DANN approaches that assume unlabeled target data, the proposed method utilizes both accurate source labels and noisy target labels. Through adversarial training, the model learns domain-invariant features while simultaneously leveraging the available target information to improve adaptation to field conditions.

#### A. Proposed SS-DANN Model Architecture for Regression

The proposed architecture is based on the DANN framework introduced by Ganin et al. [28], adapted here for regression instead of classification. The model consists of three main components: a **feature extractor**, a **regression head**, and a **domain classifier** (Fig. 4). The feature extractor learns shared representations from both domains, the regression head predicts soil tare values, and the domain classifier is trained adversarially using gradient reversal to encourage domain-invariant feature learning.

1) *Regression Head*: The regression head is designed to estimate the soil tare based on the domain-invariant features extracted by the feature extractor. It consists of a sequence of three fully connected layers with intermediate Batch Normalization, Mish activation functions, and Dropout for regularization. The final layer employs a Sigmoid activation to constrain predictions within the range [0, 1], corresponding to 0% to 100% soil tare. Initially, the regression head was trained using only the source-domain data, which contained high-quality ground-truth labels, while the target-domain data remained unlabeled, resulting in an unsupervised DANN setup. However, the model consistently overestimated the expected soil tare values. To address this issue, noisy target-domain labels provided by human estimators were incorporated during training in a semi-supervised manner, guiding the regression head toward more realistic predictions. Additionally, to handle the overestimation more severely than underestimations, an *asymmetric Mean Squared Error (MSE)* loss function was introduced. This loss function assigns a higher weight to positive residuals (i.e., over-predictions), which are considered

more critical in the soil tare estimation context. Through extensive experimentation, an overestimation penalty weight of 3 and an underestimation weight of 1 were found to yield the most balanced results. The final regression loss  $\mathcal{L}_{\text{reg}}$  is defined as:

$$\mathcal{L}_{\text{reg}} = \frac{1}{N} \sum_{i=1}^N \left[ \mathbf{1}_{\hat{y}_i > y_i} \cdot (\hat{y}_i - y_i)^2 \cdot w_{\text{over}} + \mathbf{1}_{\hat{y}_i \leq y_i} \cdot (\hat{y}_i - y_i)^2 \cdot w_{\text{under}} \right] \quad (1)$$

where  $\hat{y}_i$  and  $y_i$  are the predicted and label values, respectively, and  $w_{\text{over}} = 3$ ,  $w_{\text{under}} = 1$ . This approach is inspired by cost-sensitive regression frameworks that emphasize asymmetric error penalties depending on the application context [29], [30].

2) *Domain Classifier*: The domain classifier in our model plays a critical role in distinguishing between source and target domains during the domain adaptation process. It operates by classifying the input features into two classes: source and target domains. The classifier is composed of three fully connected layers, each followed by Batch Normalization and LeakyReLU activation with a slope of 0.2, enabling effective learning despite domain shifts. The final output layer has two units, representing the two domains, with softmax activation applied to produce domain probabilities.

The domain classifier's goal is to maximize the adversarial loss, encouraging the model to learn domain-invariant features. To optimize the classifier, Focal Loss [31] was used, a modification of the standard cross-entropy loss, to address the class imbalance between source and target domains. Focal Loss places greater emphasis on harder-to-classify examples, particularly in imbalanced datasets, by down-weighting well-classified examples. The equation for Focal Loss is given by:

$$\mathcal{L}_{\text{focal}} = \text{CE}(\hat{y}, y) \cdot (1 - \hat{p})^\gamma \cdot \alpha \quad (2)$$

where  $\text{CE}(y_i, \hat{y}_i)$  is the standard cross-entropy loss defined as:

$$\text{CE}(y_i, \hat{y}_i) = -\log(p_t) \quad (3)$$

and the term  $(1 - p_t)^\gamma$  acts as a modulating factor that reduces the loss contribution from well-classified examples,  $\gamma$  is the focusing parameter, and  $\alpha$  is the weighting factor [28]. This

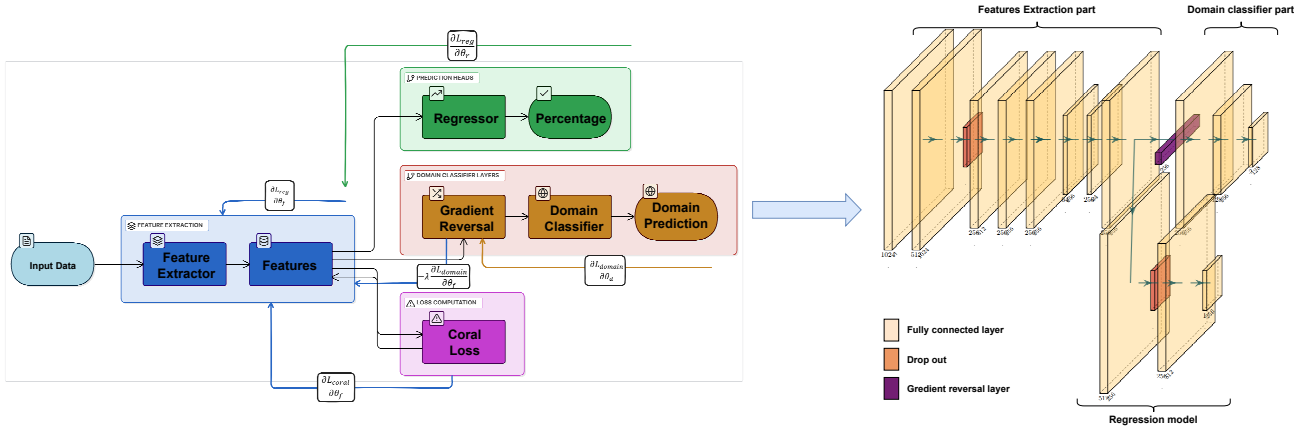


Fig. 4. SS-DANN architecture overview. The left side shows the model flow, including the feature extractor, regression head, domain classifier. It also highlights the associated loss functions and the direction of backpropagation for each component, including the use of CORAL loss for domain alignment. The right side provides a detailed view of the model architecture, including individual layers, dropout, and the gradient reversal layer.

loss is especially beneficial for domain adaptation, as it enables better handling of difficult samples and increases the focus on misclassified instances, which is essential when learning from imbalanced or noisy domains. The domain loss is calculated as follows:

$$\mathcal{L}_{\text{domain}} = \frac{1}{N} \sum_{i=1}^N [\text{CE}(y_i, \hat{y}_i) \cdot (1 - \hat{p}_i)^\gamma] \quad (4)$$

In our implementation, Focal Loss is applied with a  $\gamma$  value of 2.0 and label smoothing ( $\delta = 0.1$ ) to reduce the impact of noisy labels and ensure more robust learning, especially for the target domain. This helps the model better adapt to the target domain and improve its generalization capability in the presence of domain shifts.

### B. Feature Extractor

The feature extractor is designed to capture rich, high-level representations from the input features while promoting domain invariance across source and target domains. It is implemented as a sequence of fully connected layers with Batch Normalization, Mish activation, and Dropout regularization to enhance generalization. To improve feature expressiveness, two *Residual Blocks* are incorporated, enabling better gradient flow and reducing the risk of vanishing gradients. The final feature vector passes through an *attention mechanism* that dynamically scales feature dimensions by learning a soft attention mask. This module is implemented as two linear layers with Tanh and Sigmoid activations respectively, allowing the model to focus on the most informative components.

To facilitate domain adaptation, the feature extractor is connected to both a regression head and a domain classifier. Domain confusion is achieved using a *Gradient Reversal Layer (GRL)*[28], defined as a custom function that reverses the gradient during backpropagation. The GRL passes input

$x$  unchanged during the forward pass and multiplies the gradients by  $-\alpha$  during the backward pass:

$$\text{GRL}(x) = x \quad (\text{forward pass}), \quad (5)$$

$$\text{gradient} = \frac{d\mathcal{L}}{dx} \rightarrow -\alpha \cdot \frac{d\mathcal{L}}{dx} \quad (\text{backward pass}). \quad (6)$$

The *CORAL loss* [32] was incorporated to align the second-order statistics of feature distributions between the source and target domains. This loss minimizes the distance between the covariance matrices of source and target features, encouraging domain-invariant representations. The total training loss combines three terms: the regression loss  $\mathcal{L}_{\text{reg}}$  (source-labeled data only), the domain classification loss  $\mathcal{L}_{\text{domain}}$  (all domains), and the CORAL loss  $\mathcal{L}_{\text{coral}}$ . Adaptive weights are applied to the domain and CORAL terms to balance their contributions during training. Specifically, the final loss is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{reg}} + \lambda_d \cdot \mathcal{L}_{\text{domain}} + \lambda_c \cdot \mathcal{L}_{\text{coral}}, \quad (7)$$

where  $\lambda_d$  is dynamically computed based on the relative magnitudes of the regression and domain losses, and  $\lambda_c$  is gradually increasing over time by:

$$\lambda_c = 0.5 \cdot \frac{\text{epoch}}{\text{NUM\_EPOCHS}} \quad (8)$$

Once the feature extraction process is completed and domain-invariant representations are obtained, the model's performance is validated on both the source and target domains, focusing on the regression task and domain adaptation. The validation step calculates key metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared ( $R^2$ ) for both domains, as well as evaluating the domain gap, which measures the difference in performance between source and target domains. These metrics provide insights into the model's ability to generalize and adapt to the target domain. In the next chapter, the results of these validation metrics will be presented and discussed in detail.

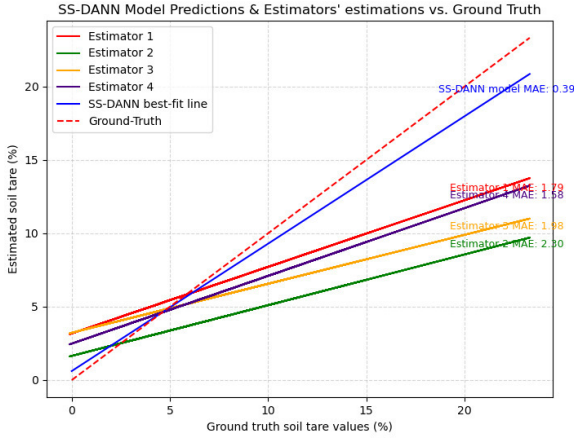


Fig. 5. Source domain evaluation: comparison of best-fit regression lines for SS-DANN predictions and human estimators against the ground truth. The plot highlights prediction accuracy, supported by corresponding MAE values for each line.

## VI. RESULTS

For evaluation, two types of assessments were performed: (1) **source domain evaluation**, using ground truth labels, and (2) **target domain evaluation**, representing real-world field conditions without accurate image-level labels.

### A. Source Domain Evaluation

The availability of accurate ground truth labels in the source domain enabled a direct evaluation of the model's prediction performance. The proposed SS-DANN model achieved strong results, with an MAE of 0.0102, RMSE of 0.0156, and an  $R^2$  score of 0.8840, demonstrating its ability to accurately model soil tare variations.

To compare the model with human performance, the estimations of four human evaluators were analyzed on a subset of the source dataset. As illustrated in Fig. 5, best-fit regression lines for both the estimators and the model predictions were compared against the ideal identity line ( $y = x$ ). The SS-DANN model achieved a substantially lower MAE of 0.39, compared to 1.79, 2.30, 1.98, and 1.58 for the four estimators, respectively. These results demonstrate that the proposed model not only generalizes effectively but also provides more consistent and accurate soil tare estimation than manual assessment.

### B. Target Domain Evaluation

While the source dataset enabled direct evaluation using ground truth labels, the main objective of this study is reliable deployment under real-world field conditions, where such labels are unavailable. Therefore, the target domain evaluation focused on assessing the effectiveness of the proposed domain adaptation approach.

During training, the domain discrepancy was successfully reduced, with the final domain loss converging to 0.0021

and the domain classifier maintaining an average accuracy of approximately 52%. This indicates that the classifier could no longer reliably distinguish between source and target features, suggesting successful learning of domain-invariant representations.

To evaluate model behavior in the target domain, the SS-DANN predictions were compared with human estimations. Kernel Density Estimation (KDE) plots were used to compare: (1) estimator errors relative to ground truth in the source domain, and (2) estimator errors relative to SS-DANN predictions in the target domain (see Fig. 6). The similar distribution patterns observed in both cases indicate that the model predictions reflect the statistical behavior of the true ground truth.

Additionally, regression analysis between the SS-DANN predictions and one human estimator in the target domain resulted in an MAE of 2.10, which is close to the 1.91 average MAE observed between the estimators and the source-domain ground truth (see Fig. 7).

Overall, the results demonstrate that the proposed SS-DANN model generalizes effectively to real-world field conditions, providing a reliable and scalable alternative to manual soil tare estimation.

## VII. CONCLUSION

This study presents an automated approach for soil tare estimation on sugar beets, addressing the variability and subjectivity of manual assessment methods. By combining Detectron2-based segmentation with a Semi-Supervised Domain-Adversarial Neural Network (SS-DANN), the proposed method successfully bridged the gap between controlled laboratory data and real-world field conditions. The model outperformed human estimators, achieving a MAE of 0.39% compared to an average deviation of 1.91% for manual estimation.

Through domain adaptation, the SS-DANN model effectively reduced discrepancies between the source (laboratory) and target (field) domains, enabling robust performance under

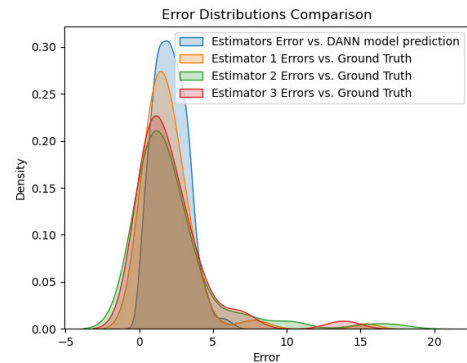


Fig. 6. KDE plots comparing estimator errors in the source domain (against ground truth) and in the target domain (against SS-DANN predictions). The similarity in distribution shapes and spreads supports the SS-DANN model's reliability as a proxy for ground truth in the target domain.

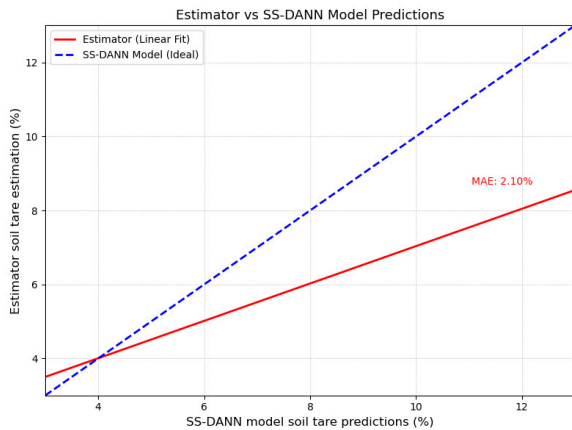


Fig. 7. Target domain evaluation: best-fit regression line for the estimator soil tare estimation against the model prediction. The plot highlights the estimator accuracy in the target domain, supporting the hypothesis that the model generalizes well to be considered as a ground truth for the field data.

practical operating conditions where accurate ground truth labels are unavailable. These results demonstrate the potential of the proposed approach as a scalable and objective alternative to manual soil tare estimation, improving operational efficiency and supporting fair compensation for both farmers and sugar companies.

Future work will focus on developing a laboratory setup that simulates field conditions to generate ground truth target-domain data and enable end-to-end soil tare estimation at both image and truck-load levels. In addition, the estimated soil tare values will be used to support the automation and optimization of sugar beet cleaning processes during harvesting.

#### ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the project partners ROPA and Südzucker for their invaluable collaboration, domain expertise, and ongoing support throughout this study. Special thanks are extended to the funding agency VDI/VDE-IT and the Bayerisches Verbundforschungsprogramm (BayVFP) of the Free State of Bavaria, funding line Digitalisierung whose support made this research possible.

The authors are also deeply thankful to Markus Zankl his continuous guidance and commitment. Finally, we gratefully acknowledge the estimators involved in the study, whose dedicated efforts and cooperation were essential to the success of this work; their contributions significantly facilitated the data annotation process and evaluation framework.

#### REFERENCES

- [1] G. D. Vermeulen, J. J. Klooster, M. C. Sprong, and B. R. Verwijs, "Soil Tare and Relative Soil Adherence after Uprooting Sugar Beet by a Share Lifter, a Driven Rotary-shoe Lifter and a Grab Lifter," 2003, [https://doi.org/10.1016/S1537-5110\(03\)00117-X](https://doi.org/10.1016/S1537-5110(03)00117-X).
- [2] P. Panagos, P. Borrelli, and J. Poesen, "Soil loss due to crop harvesting in the European Union: A first estimation of an underrated geomorphic process," *Science of the Total Environment*, vol. 664, pp. 487–498, 2019, doi: 10.1016/j.scitotenv.2019.02.009.
- [3] M. Zengin, R. Uyanöz, and Ü. Çetin, "A Study on The Soil Tare of Sugar Beet in Konya", *Selcuk J Agr Food Sci*, vol. 17, no. 31, pp. 53–55, June 2003, [Online]. Available: <https://izlik.org/JA26LW26UB>
- [4] CIBE and CEFS, *The EU Beet and Sugar Sector: A Model of Environmental Sustainability*. Brussels, Belgium, Apr. 2010.
- [5] M. Asadi, *Beet-Sugar Handbook*. Hoboken, NJ, USA: Wiley-Interscience, 2006, doi: 10.1002/0471790990.
- [6] T. Bakker, H. Wouters, K. van Asselt, J. Bontsema, L. Tang, J. Müller, and G. van Straten, "A vision based row detection system for sugar beet," *Computers and Electronics in Agriculture*, vol. 60, no. 1, pp. 87–95, 2008, doi: 10.1016/j.compag.2007.07.006.
- [7] A. Milioto, P. Lottes, and C. Stachniss, "Real-Time Blob-Wise Sugar Beets vs Weeds Classification for Monitoring Fields Using Convolutional Neural Networks," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2/W3, pp. 41–48, 2017, doi: 10.5194/isprs-annals-IV-2-W3-41-2017.
- [8] A. Jafari, S. S. Mohtasebi, H. E. Jahromi, and M. Omid, "Weed Detection in Sugar Beet Fields Using Machine Vision," *International Journal of Agriculture and Biology*, vol. 8, no. 5, pp. 602–605, 2006.
- [9] J. Gao, A. P. French, M. P. Pound, Y. He, T. P. Pridmore, and M. J. Pieters, "Deep convolutional neural networks for image-based *Convolvulus sepium* detection in sugar beet fields," *Plant Methods*, vol. 16, Art. no. 29, 2020, doi: 10.1186/s13007-020-00570-z.
- [10] K. Adem, M. M. Ozguven, and Z. Altas, "A sugar beet leaf disease classification method based on image processing and deep learning," *Multimedia Tools and Applications*, vol. 82, pp. 12577–12594, 2023, doi: 10.1007/s11042-022-13925-6.
- [11] L. Pan, Q. Zhu, R. Lu, and J. M. McGrath, "Determination of sucrose content in sugar beet by portable visible and near-infrared spectroscopy," *Food Chemistry*, vol. 167, pp. 264–271, 2015, doi: 10.1016/j.foodchem.2014.06.117.
- [12] Z. Guo, H. H. Goh, X. Li, M. Zhang, and Y. Li, "WeedNet-R: A sugar beet field weed detection algorithm based on enhanced RetinaNet and context semantic fusion," *Frontiers in Plant Science*, vol. 14, Art. no. 1226329, 2023, doi: 10.3389/fpls.2023.1226329.
- [13] A. Nasiri, S. Taheri-Garavand, Y. Zhang, and Y.-D. Zhang, "Deep learning-based precision agriculture through weed recognition in sugar beet fields," *Sustainable Computing: Informatics and Systems*, vol. 35, Art. no. 100759, 2022, doi: 10.1016/j.suscom.2022.100759.
- [14] A. Nasirahmadi, U. Wilczek, and O. Hensel, "Sugar beet damage detection during harvesting using different convolutional neural network models," *Agriculture*, vol. 11, no. 11, Art. no. 1111, 2021, doi: 10.3390/agriculture11111111.
- [15] G. Croonen, A. Trondl, J. Simon, and D. Steininger, "SemanticSugarBeets: A Multi-Task Framework and Dataset for Inspecting Harvest and Storage Characteristics of Sugar Beets," *arXiv preprint arXiv:2504.16684*, 2025, doi: 10.48550/arXiv.2504.16684.
- [16] JAI, FS-3200T-10GE-NNC 3-CMOS area scan camera, JAI, [Online]. Available: <https://www.jai.com/products/fs-3200t-10ge-nnnc/> [Accessed: May 16, 2025].
- [17] W. O. S. Soares, R. S. Peruchi, R. A. V. Silva, and P. Rotella Junior, "Gage R&R studies in measurement system analysis: A systematic literature review," *Quality Engineering*, vol. 34, no. 3, pp. 382–403, 2022, doi: 10.1080/08982112.2022.2069505.
- [18] D. C. Montgomery, *Introduction to Statistical Quality Control*, 7th ed., Hoboken, NJ: John Wiley & Sons, 2012.
- [19] Automotive Industry Action Group (AIAG), *Measurement Systems Analysis (MSA) Reference Manual*, 4th ed., Southfield, MI: AIAG, 2010.
- [20] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He, "Detectron," Facebook AI Research, 2018. [Online]. Available: <https://github.com/facebookresearch/detectron>
- [21] T. Ojala, M. Pietikainen and T. Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002, doi: 10.1109/TPAMI.2002.1017623.
- [22] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979, doi: 10.1109/TSMC.1979.4310076.
- [23] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed. Springer, 2004, doi: 10.1007/978-3-662-05088-0.
- [24] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973, doi: 10.1109/TSMC.1973.4309314.

- [25] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, 1996, doi: 10.1109/34.531803.
- [26] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, Oct. 2018, doi: 10.1016/j.neucom.2018.05.083.
- [27] O. Kilim, A. Olar, T. Joó, T. Palicz, P. Pollner, and I. Csabai, "Physical imaging parameter variation drives domain shift," *Scientific Reports*, vol. 12, Art. no. 21302, 2022, doi: 10.1038/s41598-022-23990-4.
- [28] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., and Lempitsky, V. *Domain-Adversarial training of neural networks*. arXiv (Cornell University), 2015, <https://doi.org/10.48550/arxiv.1505.07818>
- [29] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," 2nd ed. New York, NY, USA: Springer, 2009, doi: 10.1007/978-0-387-84858-7.
- [30] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-Sensitive Learning of Deep Feature Representations From Imbalanced Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3573–3587, Aug. 2018, doi: 10.1109/TNNLS.2017.2732482.
- [31] T. Lin et al., "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020, doi:<https://doi.org/10.48550/arXiv.1708.02002>.
- [32] B. Sun and K. Saenko, "Deep CORAL: Correlation Alignment for Deep Domain Adaptation," in *Proc. ECCV Workshop*, 2016, pp. 443–450, doi: 10.48550/arXiv.1607.01719