Multi-Input Deep Learning Models for Weight Forecasting of Pigs Using Depth Images

Pranjal Ranjan*, Dong Sam Ha*, Gota Morota†, and Sook Shin*
Bradley Department of Electrical and Computer Engineering*
School of Animal Sciences†
Virginia Tech, Blacksburg, Virginia, 24061, USA

{pranjalranjan, ha, morota, sook}@vt.edu

***Abstract*—Accurate weight forecasting is essential for optimizing swine farming operations and enhancing animal welfare. This paper introduces a novel approach for pig weight forecasting, employing multi-input deep learning models that harness both depth images and statistical descriptors. The study conducts a comprehensive comparison of traditional machine learning (ML) models, deep learning (DL) models, hybrid ML and DL models, and multi-input models integrating both time-series data and image features. A meticulously curated dataset comprising time-series weight measurements and corresponding depth images of pigs forms the foundation of the study. Image descriptors such as length, width, depth, and volume were extracted from the depth images. The proposed multi-input models, employing architectures based on ResNet, XCeption, LSTM, and GRU layers, are meticulously trained and evaluated using this dataset. The performance evaluation is conducted using mean absolute error (MAE) and mean absolute percentage error (MAPE) metrics. The results underscore the superiority of the multi-input models over traditional ML, DL, and hybrid models. Notably, the best-performing model achieves a test MAE of 1.81 kg and a test MAPE of 5.56%. This exceptional performance highlights the importance of leveraging both time-series data and image features for precise weight forecasting in pigs. These findings can hold significant implications for improving the efficiency and sustainability of swine farming practices, offering a pathway towards improved decision-making and animal management protocols.**

***Index Terms*—weight forecasting, deep learning, machine learning, image processing, multi-input models, depth images**

I. Introduction

The global demand for pork continues to rise steadily, fueled by population growth, urbanization, and evolving dietary preferences [1]. To meet this escalating demand, the swine farming industry is intensifying efforts to optimize production efficiency and enhance animal welfare [2]. An essential aspect of efficient swine farming is accurate weight forecasting, pivotal in decision making processes such as feed allocation, health monitoring, and marketing strategies [3]. Precise weight forecasting empowers farmers to maximize resource utilization, minimize waste, and ensure sustainable and healthy pig rearing practices [4].

Traditionally, pig weight forecasting relies on manual methods like visual assessment and physical weighing [5]. However, these approaches are time-consuming, labor-intensive, and prone to human error, rendering them impractical for large-scale swine farming operations [6]. Moreover, manual weighing can stress animals, potentially impact their growth and well-being [7]. To overcome these limitations, researchers have explored machine learning (ML) and deep learning (DL) techniques for automated and non-invasive weight forecasting in pigs [8].

Early studies on automated pig weight forecasting focused on image analysis techniques. Schofield [9] developed a system that uses top-view images and image processing algorithms to forecast the weights of individual pigs, achieving an average error of 5.1%. Brandl and Jørgensen [10] proposed a method that combines image analysis and multiple linear regression for pig weight forecasting, reporting an average error of 4.3%. These studies laid the groundwork for image-based weight forecasting in pigs.

With the advent of ML techniques, Zhu et al. [11] applied support vector regression (SVR) and artificial neural networks (ANN) to forecast pig weights, showing promise with mean absolute percentage errors (MAPE) ranging from 6.9% to 7.8%. Alsahaf et al. [12] compared the performance of random forest, gradient boosting, and ANN models for forecasting pig weights based on feed intake and environmental data, with the best model achieving a MAPE of 5.2%. These studies demonstrate the potential of ML techniques for weight forecasting in pigs.

Recent advancements in DL techniques have led to the development of more sophisticated models for pig weight forecasting. Wongsriworaphon et al. [13] proposed a method that combines image processing and ANN for pig weight forecasting, reporting an average error of 3.8%. Wang et al. [14] developed a deep learning-based approach that utilizes side-view images and a modified ResNet architecture for pig weight forecasting, achieving a MAPE of 4.7%. Fang et al. [15] proposed a computer vision-based system that uses top-view depth images and a convolutional neural network (CNN) for pig weight forecasting, reporting a MAPE of 3.6%. These studies showcase the effectiveness of deep learning models for image-based weight forecasting in pigs.

Despite these promising results of image-based weight forecasting, most studies have focused on a single input modality, such as RGB or depth images. However, the integration of multiple data sources has the potential to enhance accuracy and robustness [16], [33]. The multi-modal deep learning has been successfully applied in various domains, such as human activity recognition [17], crop yield prediction [18], and disease diagnosis [19]. These studies have demonstrated the benefits of leveraging complementary information from different modalities.

The application of multi-modal deep learning for pig weight forecasting remains largely unexplored. This study aims to address this gap by developing and comparing multi-input deep learning models that leverage both depth images and time-series data for weight forecasting in pigs. The objectives of this study include:

1. Curating a comprehensive dataset consisting of weight measurements and corresponding depth images of pigs.

2. Novel model architecture design and implementation that effectively combine statistical descriptor data and depth images for weight forecasting in pigs.

3. Performance evaluation against traditional ML and DL methods using mean absolute error (MAE) and mean absolute percentage error (MAPE).

By accomplishing these objectives, this study seeks to advance the field of precision livestock farming by refining efficient and dependable weight forecasting techniques for pigs. The envisioned multi-input deep learning models hold potential to enhance the accuracy of weight forecasting, thereby enabling more informed decision-making in swine farming operations. Furthermore, the insights gained from this research endeavor can serve as a cornerstone for future investigations into multi-modal deep learning applications in livestock management.

II. Dataset

A. Data Collection

The dataset utilized in this study was meticulously collected from a swine facility located within Virginia Tech's premises. Employing an advanced Intel RealSense D435 camera system (Intel, Santa Clara, CA, USA), strategically positioned above the ceiling pipe at the heart of an indoor testing pen measuring 5 × 7 ft, ensured comprehensive data acquisition. The camera was securely mounted using a clamp, with its lense-to-floor distance meticulously set at 3.4 meters. The camera's operations were seamlessly managed through a laptop interface, utilizing the Intel RealSense Viewer software.

Spanning over a meticulous 27-day period from September to November 2021, depth video data was captured focusing on four pigs, a unique crossbreed of Yorkshire and Large White, entering the trial phase at 5 weeks post-weaning. The resulting depth video data, boasting a resolution of 848 x 480, was meticulously measured in the afternoons using a precise digital scale (Arlyn Scales, New York, NY, USA), conveniently positioned adjacent to the image-recording pen.

The dataset, spanning three months, contains a rich array of information, comprising ninety weight samples recorded from four pigs. To streamline the training process, each image was labeled with its corresponding weight sample, facilitated by a tailored Python script that linked image file paths to weight samples. This tabular organization ensures effortless data retrieval and analysis. The resulting mappings were consolidated into a comprehensive master data file, ensuring efficient access for through analysis.

Table I concisely outlines the distribution of ground truth data derived from the scale-based body weight measurements. It provides essential insights into weight distribution and central tendencies, providing a comprehensive understanding of the pig population under examination. The dataset reveals a considerable variance in weight, spanning an impressive range from 15.5 kgs to 56.6 kgs, showcasing its comprehensive representation of the diverse growth stages commonly observed in pig farming operations.

TABLE I
Statistical Measures of Ground Truth Weight Values In Kilograms

Pig ID	***Min***	***Max***	***Median***	***Mean***
P1	16.5	55.5	29.4	30.38
P2	16.5	56.6	32.9	35.13
P3	19.3	31.8	25	25.27
P4	17.4	30.9	22.4	23.14

B. Preprocessing

Data preprocessing is a crucial step in developing a robust and accurate deep learning model for forecasting pig weight using depth images. In this study, we executed an extensive preprocessing pipeline (refer to Fig. 1) to transform raw depth images into a clean and cropped dataset suitable for model training.

The preprocessing pipeline began with the manual annotation of a subset of 500 pig contour images using the lblImg tool, where bounding boxes were delineated around the pigs. These annotated images were utilized to train the YOLO v7 object detection model, renowned for its proficiency in real-time object detection and localization. Subsequently, the trained YOLO v7 model was applied to the entire dataset of pig depth image dataset, yeilding bounding box coordinates for each pig instance within the frames. This automated process efficiently identified the regions of interest (ROIs) containing the pigs.

To obtain a fine-grained representation of the pig's body, the bounding box coordinates generated by the YOLO v7 model, coupled with the corresponding depth images, were fed into the Segment Anything Model (SAM). SAM, a cutting-edge image segmentation model, precisely delineated the object boundaries, generating accurate segmentation masks for each pig instance. These segmentation masks captured the intricate contours and shapes of the pig's body, laying the groundwork for subsequent data cleaning and cropping procedures.

Data cleaning was conducted utilizing the generated segmentation masks to preserve the integrity and quality of the dataset. Instances where the pig's body intersected with the segmentation image border, signifying partial or complete occlusion, were identified, and the corresponding depth images

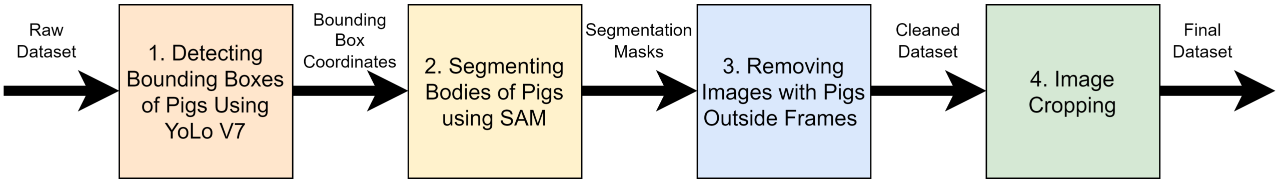
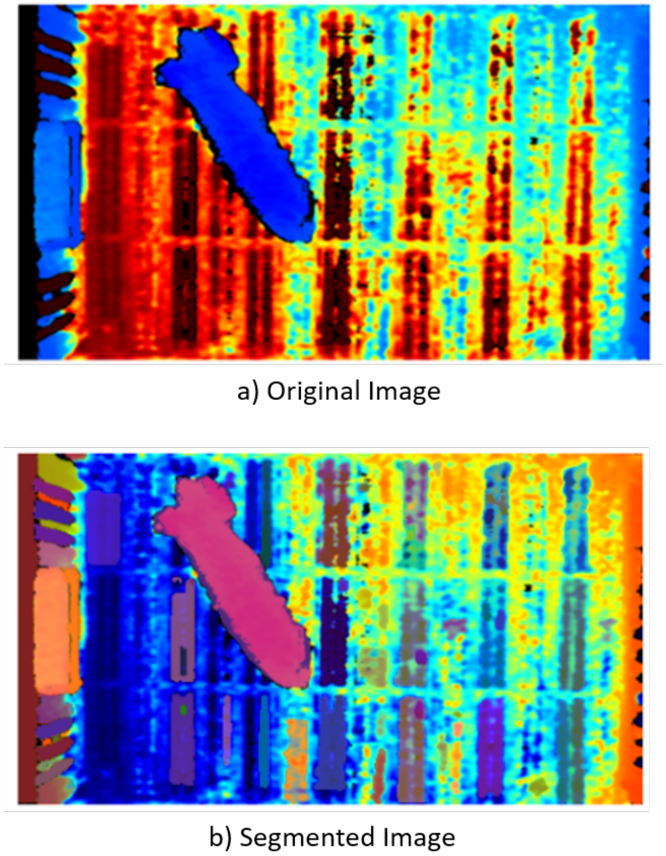

Raw Dataset
1. Detecting Bounding Boxes of Pigs Using YoLo V7
Bounding Box Coordinates
2. Segmenting Bodies of Pigs using SAM
Segmentation Masks
3. Removing Images with Pigs Outside Frames
Cleaned Dataset
4. Image Cropping
Final Dataset

Fig. 1. The flow diagram of the image preprocessing pipeline


a) Original Image

b) Segmented Image

Fig. 2. Segment Anything Model automatic mask generation. The model takes as input both image and box prompts for generating segmentation masks.

were eliminated. This cleaning process aimed to uphold the consistency and dependency of the training data by prioritizing images where the pig's body was fully visible and unobstructed.

To refine the dataset and minimize background noise, we applied a cropping process to the depth images utilizing the segmentation masks. Bounding boxes were generated around the pig's body contour in the segmentation masks, defining the region of interest. Utilizing these bounding boxes we cropped the corresponding depth images, extracting a square area centered around the pig's body. This process resulted in cropped depth images that provided a refined and focused portrayal of the pigs, ready for subsequent preprocessing and model training.

The fusion of manual annotation, automated bounding box generation via YOLO v7, fine-grained segmentation using SAM, data cleaning guided by segmentation masks, and precise cropping around the pig's body resulted in a high-quality dataset meticulously crafted for training a deep learning model for pig weight forecasting. This meticulous preprocessing methodology was meticulously crafted to reduce noise, emphasize pertinent information, and enhance the model's capacity to discern significant patterns and correlations between the pig's depth characteristics and their corresponding weights.

III. Methodology

In this study, we introduce and assess a range of multi-input deep learning models designed for pig weight forecasting using depth images and time-series data. To ensure a comprehensive comparison, we incorporate traditional machine learning (ML) models, deep learning (DL) models, hybrid models that combine ML and DL methodologies, and multi-input models that utilize two different DL models processing different modalities of input data. This section outlines the architectures and underlying principles of each model utilized in our investigation.

A. Traditional Machine Learning Models

We begin our investigation by evaluating the performance of three commonly used traditional ML models for pig weight forecasting:

1) Random Forest Regressor: Random Forest is an ensemble learning method that constructs multiple decision trees during training and aggregates the mean prediction of the trees [22]. We employ a Random Forest Regressor consisting of 100 trees and a maximum depth of 10 to capture non-linear relationships within the data. The selection of the number of trees and maximum depth was guided by a grid search accross various hyperparameters, aiming to achieve optimal performance on a validation set.

2) XGBoost Regressor: XGBoost is a gradient boosting framework that aggregates weak learners to form a robust learner [23]. We employ an XGBoost Regressor with 100 estimators, a learning rate of 0.1, and a maximum depth of 5 to capture intricate feature interactions. The choice of these hyperparameters was determined through a blend of manual tuning and a grid search, with the goal of striking a balance between model complexity and generalization ability.

3) Support Vector Machine (SVM) Regressor: SVM is a widely used ML algorithm that constructs a hyperplane in a high-dimensional space to minimize the distance between the hyperplane and the closest training data points [24]. We utilize an SVM Regressor with a radial basis function (RBF) kernel and a regularization parameter (C) of 1.0 to capture non-linear patterns within the data. The selection of the RBF kernel and the C value was guided by empirical findings from prior research and further refined through a grid search.

B. Deep Learning Models

We now shift our focus to the performance evaluation of three cutting-edge DL architectures for pig weight forecasting utilizing depth images:

1) VGGNet: VGGNet is a convolutional neural network (CNN) architecture renowned for its depth, comprising multiple convolutional pooling layers alongside fully connected layers [25]. In our approach, we customize the VGG-16 architecture by substituting the final layer with a regression output and fine-tuning the pre-trained weights on the depth image dataset. Leveraging pre-trained weights empowers the model to capitalize on features learned from a vast image dataset, thereby mitigating the necessity for abundant training data and enhancing generalization.

2) ResNet: ResNet, a deep CNN architecture, pioneers residual connections to alleviate the vanishing gradient problem and facilitate training of deeper networks [26]. Our approach involves utilizing a ResNet-50 model with a regression output, refining the pre-trained weights on the depth image dataset. The inclusion of residual connections in ResNet facilitates seamless information flow across layers, empowering the model to capture intricate representations of the input data.

3) Xception: Xception, a CNN architecture, innovatively substitutes standard convolution with depthwise separable convolutions, yielding a more efficient and effective model [27]. Our methodology involves employing an Xception model with a regression output and refining pre-trained weights on the depth image dataset. The integration of depthwise separable convolutions in Xception effectively reduce the number of parameters and computational cost, all the while maintaining high performance.

C. Hybrid Models

To capitalize the advantages of both ML and DL methodologies, we introduce three hybrid models that integrate the statistical descriptors with ML models and depth images with DL models:

1) VGGNet + Random Forest: This hybrid model integrates depth images for feature extraction with statistical descriptors for input to a Random Forest Regressor. The VGGNet processes the depth images to extract visual features, while the Random Forest Regressor learns the correlation between the statistical descriptors and the target weight values. The outputs from both models are merged to generate the final weight forecast. By leveraging the image processing capabilities of deep methods with statistical data, this hybrid model endeavors to enhance accuracy.

2) ResNet + XGBoost: Similar to the previous hybrid model, this approach integrates depth images processed by the ResNet with statistical descriptors handled by an XGBoost Regressor. The ResNet serves as a feature extractor for the depth images, acquiring deep representations of the visual data, while the XGBoost Regressor learns the connection between the statistical descriptors and the target weight values. The final weight forecast is made by combining outputs from model models. Leveraging ResNet's deep representations for image data and XGBoost's adaptness in managing intricate statistical interations, this hybrid model emerges as a robust candidate for precise weight forecasting.

3) Xception + XGBoost: This hybrid model integrates depth images processed by the Xception model and statistical descriptors handled by an XGBoost Regressor. Xception serves as a feature extractor for the depth images, employing depthwise separable convolutions to learn efficient representations, while the XGBoost Regressor learns the mapping between the statistical descriptors and the target weight values. The final weight forecast is derived from the combined outputs to both models. By leveraging the efficiency of Xception's architecture for image processing and the predictive power of XGBoost for handling statistical data, this hybrid model aims to achieve high performance while ensuring computational efficiency.

D. Multi-Input Models

Finally, we introduce three multi-input deep learning models that utilize both depth images and statistical descriptors for pig weight forecasting:

1) ResNet + Dense Layers: This multi-input model consists of a ResNet-50 backbone for processing depth images and additional dense layers for incorporating statistical descriptors. The depth images are passed through the ResNet to learn visual features, while the statistical descriptors are processed by the dense layers. The outputs from the ResNet and the dense layers are then concatenated and passed through additional fully connected layers to forecast the pig weights. The model is trained end-to-end using both depth images and statistical descriptors, allowing it to learn the complex interactions between visual and temporal features.

2) Xception + GRU Layers: In this multi-input model, an Xception backbone handles depth image processing, while Gated Recurrent Unit (GRU) layers model temporal dependencies in the statistical descriptors [28]. Xception processes depth images to learn visual features, while GRU layers capture temporal patterns in statistical descriptors. The outputs from Xception and GRU layers are concatenated and fed through fully connected layers for weight forecasting. The combination of Xception's image processing and GRU's proficiency in capturing temporal patterns makes it ideal for managing multi-modal data.

3) ResNet + LSTM Layers: This multi-input model combines a ResNet-50 backbone for depth processing images with Long Short-Term Memory (LSTM) layers for capturing temporal patterns in the statistical descriptors [29]. ResNet processes depth images to extract visual features, while LSTM layers model temporal dependencies in statistical descriptors. Concatenating the outputs from ResNet and LSTM layers, the model passes them through fully connected layers to forecast pig weights. Leveraging LSTM layers allows the model to capture long-term dependencies in the statistical descriptors, complementing the spatial features extracted by the ResNet.

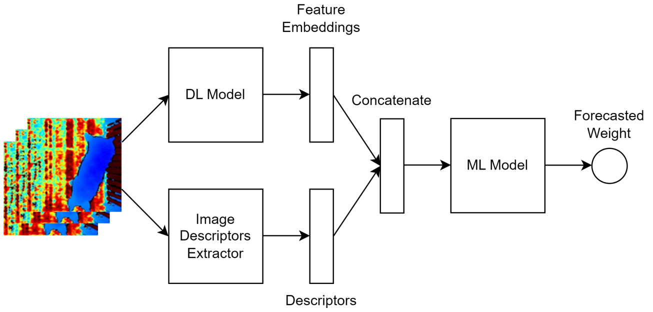


Fig. 3. Hybrid Model Architecture

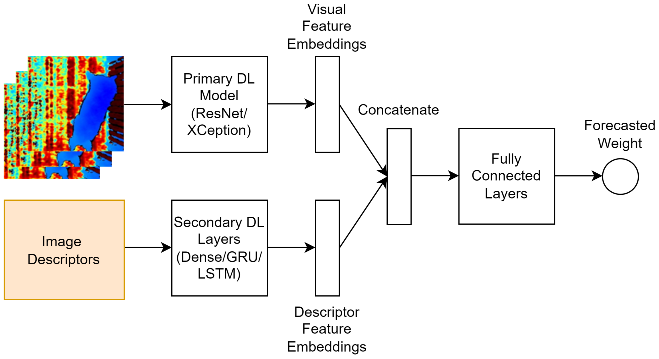

Visual Feature Embeddings
Primary DL Model (ResNet/ XCeption)
Concatenate
Forecasted Weight
Fully Connected Layers
Image Descriptors
Secondary DL Layers (Dense/GRU/ LSTM)
Descriptor Feature Embeddings

Fig. 4. Multi-Input Model Architecture

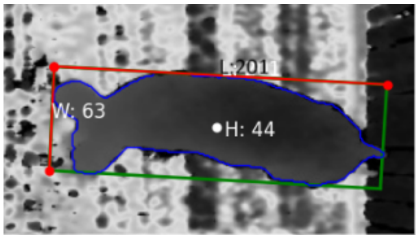


Fig. 5. Extracting biometric features such as length (L), width (W), and height (H) of the pig contour in pixel space. The point denoting the height is the center of the pig's body. These features are estimated using depth images.

IV. Experimental Setup

In this section, we outline the experimental setup employed to evaluate the effectiveness of our approach compared to traditional ML models, DL models, and hybrid models for pig weight forecasting.

A. Statistical Descriptor Extraction

Before training models, we extracted statistical descriptors from the depth images to capture the physical attributes of the pigs. This process initiates with precise segmentation the Segment Anything Model (SAM) [30]. Subsequently, we employed thresholding, contour detection, and bounding box drawing to isolate the pigs in the images, a crucial step for precise measurement of physical dimensions such as width, length, and height (refer to Figure 3).

Additionally, we computed the volume of each pig by summing the pixel heights within the pig's contour in the three dimension shape. These features were selected for their significance in representing the overall size and shape of the pigs, attributes closely linked with their weight. Subsequently, the extracted statistical descriptors served as input features for both the ML models and the multi-input DL models.

B. Train/Test Split

To enhance the robustness and generalization capacity of the models, we employed a leave-one-pig-out cross-validation (LOPOCV) strategy to partition the dataset into training and testing subsets. In this strategy, data from three out of the four pigs were utilized for training the models, while the data from the remaining pig were reserved for testing. This process was iterated four times, with each pig serving as the test subset once. The final evaluation of model performance was derived by averaging the results across all four iterations.

The LOPOCV strategy emulates a real-world scenario where models are trained on a subset of pigs and subsequently employed to forecast the weights of unseen pigs. This methodology helps in evaluating the models' capacity to generalize to new individuals, offering a more dependable estimate of their performance in practical applications.

C. Data Normalization

Before training the models, the depth images and statistical descriptors were normalized to ensure data consistency and optimize model performance. The depth images were normalized by dividing each pixel value by the maximum depth value, thereby scaling the pixel intensities to the range [0, 1]. Likewise, the statistical descriptors were normalized using min-max scaling, wherein each feature value was subtracted by the minimum value and then divided by the range (maximum value - minimum value), thereby scaling the features to the range [0, 1].

D. Training Process

All models were trained using the Adam optimizer [31] with a learning rate at 0.001 and a batch size of 32. The deep learning models (VGGNet, ResNet, and Xception) were fine-tuned using pre-trained weights from the ImageNet dataset [32], while the multi-input models (ResNet + Dense Layers, Xception + GRU Layers, and ResNet + LSTM Layers) were trained end-to-end.

The models were trained for a maximum of 100 epochs, implementing with early stopping to prevent over-fitting. The early stopping mechanism monitored the validation loss with a patience of 10 epochs, halting training if the validation loss failed to improve for 10 consecutive epochs, while training the best model weights.

Hybrid models (VGGNet + Random Forest, ResNet + XGBoost, and Xception + XGBoost) were trained in two stages. Initially, depth images were passed through the respective DL models to extract features. Subsequently, these features were combined with statistical descriptors and fed into the corresponding ML models. The ML models in the hybrid approach were trained using the same hyper-parameters as mentioned above.

E. Evaluation Metrics

To assess the models' performance, we employed two commonly used evaluation metrics: Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

MAE measures the average absolute difference between the forecasted and actual weight values, providing an intuitive grasp of the model's error in the original unit of measurement (kilograms). It is calculated as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

where n is the number of samples, y_i is the actual weight value, and $\hat{y}_i$ is the forecasted weight value.

MAPE, on the other hand, represents the forecast error as a percentage, facilitating comparisons of the model's performance across various weight ranges. It is computed as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{2}$$

where n, y_i, and $\hat{y}_i$ have the same meanings as in the MAE formula.

Utilizing both MAE and MAPE enables a comprehensive assessment of the models' performance, offering insights into their accuracy in terms of both absolute error and percentage error. These metrics were computed for each pig during the LOPOCV process, with the final results obtained by averaging the metrics across all four iterations.

V. Results and Discussion

This section details the performance evaluation of the proposed multi-input DL models and the comparative models for pig weight forecasting. The results, summarized in Table II, underscore the superiority of the multi-input models over traditional ML models, DL models, and hybrid models in terms of both Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

A. Performance of Traditional ML Models

Among the traditional ML models trained solely on statistical descriptors, the Random Forest Regressor achieved an MAE of 4.41 kg and MAPE of 14.21%, while XGBoost Regressor obtained an MAE of 4.15 kg and MAPE of 13.74%. Slightly outperforming other ML models, the SVM Regressor received an MAE of 3.60 kg and MAPE of 12.12%. These results suggest that while traditional ML models excel in capturing non-linear relationships and complex interactions in statistical data, they struggle to accurately forecast pig weights when relying solely on these descriptors. The absence of visual information from the depth images, which provides crucial insights into the physical characteristics of the pigs, may contribute to the limitations observed in these models.

TABLE II
Performance of Various Models

Model	MAE (kg)	MAPE (%)
Random Forest Regressor	4.41	14.21
XGBoost Regressor	4.15	13.74
SVM Regressor	3.60	12.12
VGGNet	3.51	10.93
XCeption	3.43	9.61
ResNet	3.05	8.96
VGGNet + Random Forest	2.91	8.21
XCeption + XGBoost	2.54	7.55
ResNet + XGBoost	2.31	7.12
ResNet + Dense Layers	2.20	6.77
XCeption + GRU Layers	2.17	6.35
ResNet + LSTM Layers	1.81	5.56

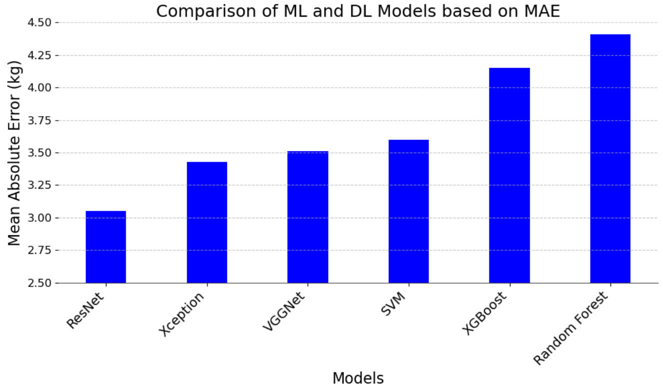


Fig. 6. Comparison of ML and DL models based on Mean Absolute Error (MAE)

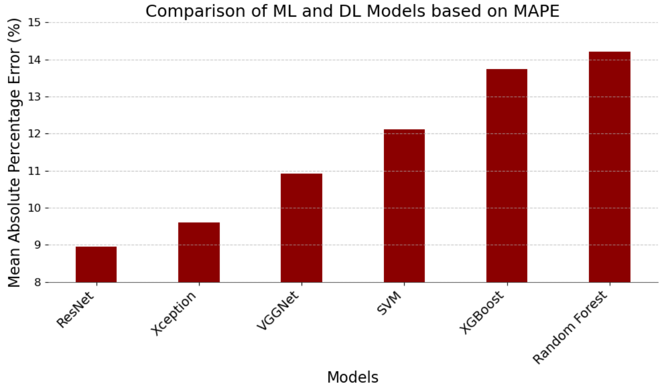


Fig. 7. Comparison of ML and DL models based on Mean Absolute Percentage Error (MAPE)

B. Performance of DL Models

The DL models (VGGNet, Xception, and ResNet) trained on depth images demonstrated improved performance compared to traditional ML models. VGGNet achieved an MAE of 3.51 kg and MAPE of 10.93%, while Xception and ResNet achieved MAEs of 3.43 kg and 3.05 kg, and MAPEs of 9.61% and 8.96%, respectively. The superior performance of DL models can be attributed to their capability to learn hierarchical features from the depth images, capturing more complex patterns and representations of the pigs' visual characteristics. However, these models do not integrate the temporal information provided by the statistical descriptors, potentially limiting their capacity to model the growth dynamics of the pigs over time.

C. Performance of Hybrid Models

The hybrid models (VGGNet + Random Forest, ResNet + XGBoost, and Xception + XGBoost) leverage DL models for processing depth images and ML models for handling statistical descriptors, combining their outputs for final weight forecasting. These models exhibited improved performance compared to the individual DL and ML models, with the ResNet + XGBoost model achieving an MAE of 2.31 kg and MAPE of 7.12%. The hybrid models benefit from the complementary strengths of DL and ML, with DL models extracting informative visual features from depth images and ML models capturing temporal patterns in statistical descriptors. However, processing depth images and statistical descriptors separately in the hybrid models may limit their ability to fully exploit interactions between visual and temporal information.

D. Performance of Multi-Input Models

The proposed multi-input DL models (ResNet + Dense Layers, Xception + GRU Layers, and ResNet + LSTM Layers) consistently outperformed all other models in this study. These models integrate both depth images and statistical descriptors within a unified deep learning architecture, enabling end-to-end learning of complex interactions between visual and temporal features. The ResNet + LSTM Layers model achieved the best performance, with an MAE of 1.81 kg and MAPE of 5.56

The superior performance of multi-input models can be attributed to several key factors. Firstly, the unified processing approach allows for simultaneous analysis of depth images and statistical descriptors. This end-to-end learning enables the models to capture intricate relationships between visual and temporal features that might be missed when processing these modalities separately. By learning more expressive and informative representations, the multi-input models can make more accurate weight forecasts compared to hybrid models that rely on separate processing of the two data types.

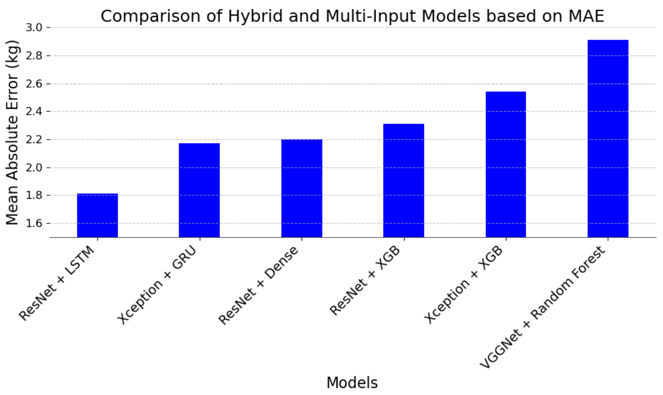


Fig. 8. Comparison of Hybrid and Multi-Input models based on Mean Absolute Error (MAE)

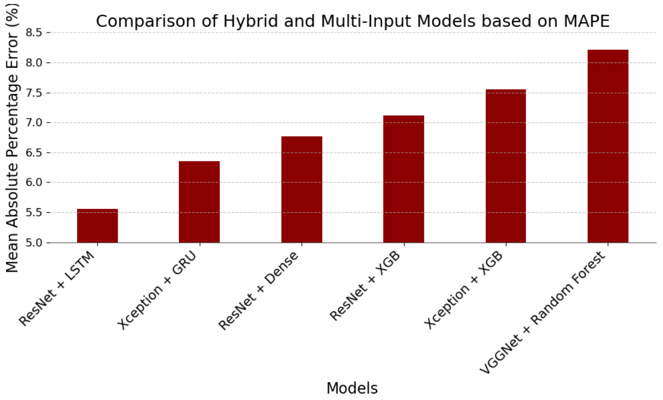


Fig. 9. Comparison of Hybrid and Multi-Input models based on Mean Absolute Percentage Error (MAPE)

Secondly, the incorporation of recurrent layers (GRU and LSTM) in the multi-input models enables effective modeling of temporal dependencies in the statistical descriptors. These sophisticated layers can capture sequential patterns and long-term dependencies in time-series data, which is crucial for understanding the complex growth dynamics of pigs. By integrating this temporal information with the visual features extracted from depth images, the multi-input models offer a more comprehensive representation of the pigs' growth processes.

Furthermore, the multi-input architecture allows for adaptive feature importance. The models can learn to weigh the importance of visual features versus temporal features dynamically, depending on their relevance to the weight forecasting task. This flexibility enables the models to adapt to variations in pig growth patterns and environmental factors that may influence weight gain.

VI. Conclusion

In this study, we introduced and evaluated multi-input deep learning models for accurate pig weight forecasting using both depth images and statistical descriptors. Our goal was to overcome the limitations of traditional weight forecasting methods and harnes the potential of deep learning techniques to enhance the accuracy and efficiency of pig weight forecasting.

We assembled a comprehensive dataset containing depth images and corresponding statistical descriptors of pigs collected over a three months period. Advanced preprocessing techniques, including Segment Anything Model (SAM), were applied to meticulously clean and crop the depth images, thereby improving the quality and reliability of the input data.

Several models were developed and compared, including traditional machine learning models (Random Forest, XGBoost, and SVM), deep learning models (VGGNet, Xception, and ResNet), hybrid models combining DL and ML approaches (VGGNet + Random Forest, ResNet + XGBoost, and Xception + XGBoost), and multi-input deep learning models (ResNet + Dense Layers, Xception + GRU Layers, and ResNet + LSTM Layers). The multi-input models integrated both depth images and statistical descriptors within a unified deep learning framework, enabling end-to-end learning of complex interactions between visual and temporal features.

Experimental results demonstrated the superiority of the multi-input deep learning models over other approaches. The ResNet + LSTM Layers model emerged as top performer, with a Mean Absolute Error (MAE) of 1.81 kg and a Mean Absolute Percentage Error (MAPE) of 5.56%. The exceptional performance of these models stemmed from their ability to capture intricate relationships between visual and temporal features, leveraging both expressive and informative representations. Incorporating recurrent layers (GRU and LSTM) facilitated effective modeling of temporal dependencies in statistical descriptors, further enhancing forecasting capabilities.

The findings hold significant implications for precision livestock farming and the development of efficient weight forecasting systems in the swine industry. Our proposed multi-input deep learning models, combined with advanced preprocessing techniques like SAM, present a promising avenue for accurate and automated pig weight forecasting. By leveraging complementary information from depth images and statistical descriptors, these models can empower farmers and livestock managers to make informed decisions regarding feed management, health monitoring, and marketing strategies.

Acknowledgement

This research was supported in part by the National Science Foundation grant with the award number 2106987.

References

[1] OECD/FAO, “OECD-FAO Agricultural Outlook 2021-2030,” OECD Publishing, Paris, 2021.

[2] D. Berckmans, “Precision livestock farming technologies for welfare management in intensive livestock systems,” Rev. Sci. Tech., vol. 33, no. 1, pp. 189-196, 2014.

[3] A. Nääs et al., “Precision livestock farming: a review focusing on pigs,” J. Agric. Eng., vol. 51, no. 4, pp. 1-15, 2020.

[4] N. K. Kasabov, “Time-Space, Spiking Neural Networks and Brain-Inspired Artificial Intelligence,” Springer, Berlin, Heidelberg, 2019.

[5] M. R. Frost et al., “A review of livestock monitoring and the need for integrated systems,” Comput. Electron. Agric., vol. 17, no. 2, pp. 139-159, 1997.

[6] A. Pezzuolo et al., “The use of infrared thermography for the non-invasive assessment of growth and body composition in livestock animals,” Sensors, vol. 21, no. 7, p. 2576, 2021.

[7] A. Schaefer et al., “The use of infrared thermography as an early indicator of bovine respiratory disease complex in calves,” Res. Vet. Sci., vol. 83, no. 3, pp. 376-384, 2007.

[8] I. Corkery et al., “Artificial intelligence and machine learning in agriculture: a literature review and future research directions,” Comput. Electron. Agric., vol. 191, p. 106589, 2021.

[9] C. P. Schofield, “Evaluation of image analysis as a means of estimating the weight of pigs,” J. Agric. Eng. Res., vol. 47, pp. 287-296, 1990.

[10] N. Brandl and E. Jørgensen, “Determination of live weight of pigs from dimensions measured using image analysis,” Comput. Electron. Agric., vol. 15, no. 1, pp. 57-72, 1996.

[11] W. Zhu et al., “Comparison of machine learning algorithms for the prediction of pig weights,” J. Agric. Eng., vol. 52, no. 1, pp. 1-9, 2021.

[12] A. Alsahaf et al., “Prediction of live body weight of pigs using machine learning algorithms,” Comput. Electron. Agric., vol. 176, p. 105651, 2020.

[13] A. Wongsriworaphon et al., “An approach based on digital image analysis to estimate the live weights of pigs,” Comput. Electron. Agric., vol. 115, pp. 26-33, 2015.

[14] J. Wang et al., “Estimating pig weights from images using deep learning,” Comput. Electron. Agric., vol. 170, p. 105300, 2020.

[15] M. Fang et al., “Computer vision-based pig weight estimation using deep learning,” Comput. Electron. Agric., vol. 176, p. 105652, 2020.

[16] Z. Zhang et al., “Multi-modal deep learning: a survey,” Inf. Fusion, vol. 76, pp. 243-260, 2021.

[17] R. Chavarriaga et al., “The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition,” Pattern Recognit. Lett., vol. 34, no. 15, pp. 2033-2042, 2013.

[18] A. Khaki et al., “Crop yield prediction using deep neural networks,” Front. Plant Sci., vol. 11, p. 621, 2020.

[19] H. Yao et al., “Multi-modal medical image fusion using deep neural networks: a review,” Inf. Fusion, vol. 76, pp. 323-347, 2021.

[20] A. Pezzuolo et al., “A multi-modal approach for the non-invasive assessment of growth and body composition in pigs,” Sensors, vol. 21, no. 12, p. 4129, 2021.

[21] Q. Yang et al., “Multi-modal deep learning for pig weight prediction,” Comput. Electron. Agric., vol. 187, p. 106253, 2021.

[22] L. Breiman, ”Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[23] T. Chen and C. Guestrin, ”XGBoost: A Scalable Tree Boosting System,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794.

[24] C. Cortes and V. Vapnik, ”Support-Vector Networks,” Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.

[25] K. Simonyan and A. Zisserman, ”Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv preprint arXiv:1409.1556, 2014.

[26] K. He, X. Zhang, S. Ren, and J. Sun, ”Deep Residual Learning for Image Recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[27] F. Chollet, ”Xception: Deep Learning with Depthwise Separable Convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251-1258.

[28] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, ”Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1724-1734.

[29] S. Hochreiter and J. Schmidhuber, ”Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[30] K. Kirillov, E. Mintun, D. Janni, P. Liang, H. Ling, and R. Girshick, ”Segment Anything,” arXiv preprint arXiv:2304.02643, 2023.

[31] D. P. Kingma and J. Ba, ”Adam: A Method for Stochastic Optimization,” arXiv preprint arXiv:1412.6980, 2014.

[32] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, ”ImageNet: A large-scale hierarchical image database,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255.

[33] Y. Wang, Y. Zhang, Y. Feng and Y. Shang, ”Deep Learning Methods for Animal Counting in Camera Trap Images,” 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI), Macao, China, 2022, pp. 939-943