OPTIMIZING TOMATO STORAGE-TIME USING SUPPORT VECTOR MACHINE ALGORITHM TO IMPROVE QUALITY AND REDUCE WASTE

Rahmat[1*]; Sunardi[2]; Fitriani[3]; Andi Saenong[4]; Muhammad Rusdi Rahman[5]; Herman Heriadi[6]; Hernawati[7]

Software Engineering[1*], Computer Science[2,3,6], Information Management[4,5]
Universitas Dipa Makassar. Makassar, Indonesia[1,2,3,4,5,6]
https://undipa.ac.id/ [1,2,3,4,5,6]
rahmat2024@undipa.ac.id[1*], sunardi@undipa.ac.id[2], fitriani@undipa.ac.id[3], andi.saenong@undipa.ac.id[4], rusdi.rahman@undipa.ac.id[5]

Information system security[7]
Politeknik Nusantara Makassar. Makassar, Indonesia[7]
https://polinus.ac.id/[7]
hernawatizainal90@gmail.com[7]

(*) Corresponding Author

Abstract—*Tomatoes are an agricultural commodity that is susceptible to spoilage, with a limited shelf life if not stored under optimal conditions. Optimizing tomato storage time is very important for improving product quality and reducing waste in distribution. This study aims to implement the Support Vector Machine (SVM) algorithm in predicting the optimal storage time for tomatoes, taking into account environmental factors such as temperature and humidity, as well as tomato ripeness. The dataset used consists of tomato images taken at various ripeness levels, as well as environmental data during storage. The SVM model was trained to classify tomato ripeness conditions and predict the optimal storage duration before significant quality deterioration occurs. The results of the study show that the SVM model has high accuracy in classifying tomato ripeness and can be used to predict the optimal storage time, which in turn can extend the shelf life of tomatoes and reduce crop waste. This research contributes to more efficient and sustainable tomato post-harvest management.*

Keywords: *Food Waste, Machine Learning Algorithm, Optimization, Ripeness, Support Vector Machine.*

Intisari— *Tomat merupakan komoditas pertanian yang rentan terhadap amandemen, dengan masa simpan yang terbatas jika tidak disimpan dalam kondisi yang optimal. Optimasi waktu penyimpanan tomat sangat penting untuk meningkatkan kualitas produk dan mengurangi pemborosan dalam distribusi. Penelitian ini bertujuan untuk mengimplementasikan algoritma Support Vector Machine (SVM) dalam memprediksi waktu penyimpanan tomat yang optimal, dengan mempertimbangkan faktor-faktor lingkungan seperti suhu dan kelembaban, serta kematangan tomat. Dataset yang digunakan terdiri dari citra tomat yang diambil pada berbagai tingkat kematangan, serta data lingkungan selama penyimpanan. Model SVM dibor untuk mengklasifikasikan kondisi kematangan tomat dan memprediksi durasi penyimpanan yang optimal sebelum terjadi penurunan kualitas yang signifikan. Hasil penelitian menunjukkan bahwa model SVM memiliki akurasi tinggi dalam mengklasifikasikan kematangan tomat dan dapat digunakan untuk memprediksi waktu penyimpanan yang optimal, yang pada gilirannya dapat memperpanjang umur simpan tomat dan mengurangi pemborosan hasil panen. Penelitian ini memberikan kontribusi dalam pengelolaan pascapanen tomat dengan cara yang lebih efisien dan berkelanjutan.*

Kata Kunci: *Algoritma Machine Learning, Limbah Makanan, Optimalisasi, Kematangan, Support Vector Machine*

INTRODUCTION

Tomatoes are one of the horticultural commodities with high production and consumption rates, but they are also classified as perishable products. The physiological characteristics of tomatoes make them very sensitive to environmental changes during storage and distribution. Factors such as initial ripeness, temperature, and storage humidity significantly affect the respiration rate and heating of tomatoes, which ultimately impact product quality and shelf life (Abdipour et al., 2025; Ekinci et al., 2025).

The issue of food waste in horticultural commodities, particularly tomatoes, remains a serious problem in the agricultural supply chain. Inaccuracies in determining the optimal storage time often cause tomatoes to spoil before reaching consumers. Several studies show that post-harvest losses of tomatoes can occur due to the lack of a prediction system capable of adjusting the storage duration to the actual condition of the product and the storage environment (Goyal et al., 2024; Shankaraswamy & Radhika, 2024). Therefore, a data-driven approach is needed to help make more accurate decisions in tomato storage management.

With the development of artificial intelligence technology, machine learning approaches are increasingly being applied in agriculture, particularly in the analysis of quality and shelf life prediction of horticultural products. Various methods such as Support Vector Machine (SVM), Artificial Neural Network (ANN), and deep learning models have been used to classify ripeness levels and predict the shelf life of tomatoes based on digital images and environmental parameters (Goyal et al., 2024; Affognon et al., 2025). Among these methods, the SVM algorithm is known to have advantages in handling high-dimensional data, being stable against overfitting, and being effective for use on limited-sized datasets (Edika & Hartati, 2026).

Several previous studies have proven that SVM is capable of providing high accuracy in digital image-based tomato ripeness classification and quality analysis during storage (Rajamoorthi et al., 2025; MDPI, 2025). However, most of these studies still focus on the aspects of maturity classification or tomato quality evaluation separately. The integration between maturity classification, storage environmental factors, and optimal storage time prediction is still relatively limited, especially in the context of directly reducing crop waste.

Based on these conditions, this study proposes an approach to optimize tomato storage time using the Support Vector Machine (SVM) algorithm, taking into account the level of tomato ripeness and storage environment parameters such as temperature and humidity. This approach is expected to produce more accurate and applicable storage time predictions, thereby improving tomato quality during storage, extending shelf life, and reducing waste in the agricultural supply chain. The results of this study are expected to contribute significantly to the development of an efficient and sustainable artificial intelligence-based post-harvest management system.

MATERIALS AND METHODS

Type and Design of Research

This study is a quantitative study with an experimental approach, which aims to optimize the storage time of tomatoes using the Support Vector Machine (SVM) algorithm. This approach is used to model the relationship between tomato ripeness, storage environmental factors, and optimal storage duration. The machine learning method was chosen for its ability to handle non-linear and complex patterns in post-harvest agricultural data (Goyal et al., 2024; Abdipour et al., 2025).

Dataset and Data Collection

The research dataset consists of:

1. Images of tomatoes at various ripeness levels (e.g., unripe, semi-ripe, ripe).
2. Storage environment data, including temperature (°C) and relative humidity (%).
3. Storage duration, which represents the time until the quality of tomatoes deteriorates.

Imaging was performed with uniform lighting and a fixed shooting distance to minimize visual noise. Environmental data was obtained through temperature and humidity sensors that were recorded periodically during the storage process. This approach is in line with previous studies that combined visual and environmental data in predicting the shelf life of tomatoes (Shankaraswamy & Radhika, 2024; Affognon et al., 2025).

Pre-Processing Data

Pre-processing was performed to improve data quality before inputting it into the SVM model, which included:

1. Image preprocessing, consisting of:
   a. Resizing images to standardize their size,
   b. Converting the RGB color space,
   c. Normalizing pixel values.
2. Numeric data normalization, performed using the min-max normalization method to equalize the scale between environmental variables, with the formula:

where:

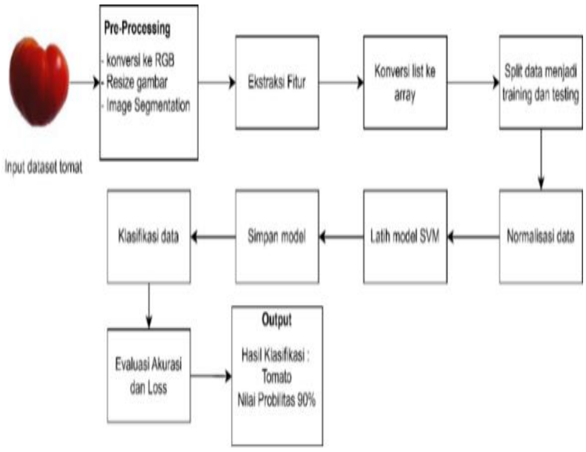
$$X' = \frac{X - X_{menit}}{X_{maksimum} - X_{menit}}$$

a. $X'$is the normalized value,
b. X is the original value,
c. $Xmenit$ andXmaksimum Xmax are the minimum and maximum values in the dataset, respectively.

Normalization is necessary to prevent one feature from dominating other features in the SVM training process (Edika & Hartati, 2026).

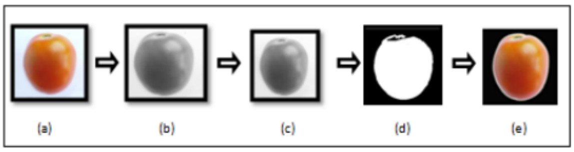
Implementation of the *Support Vector Machine (SVM)* Algorithm

Support Vector Machine (SVM) is used to build a tomato storage time prediction model. SVM works by finding the optimal hyperplane that maximizes the margin between classes in feature space (Goyal et al., 2024).



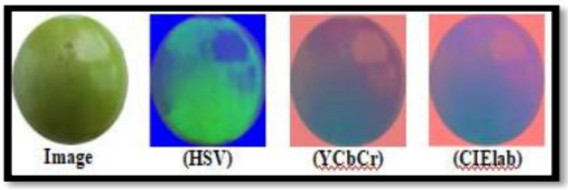
Source: (Research Results, 2026)
Figure 1. SVM Model Architecture

1. Tomato Input Dataset: The dataset consists of tomato images that are used as initial data for training and testing.
2. Pre-Processing:
   This stage includes several main steps:
   a. Conversion to RGB: Images are converted to RGB format to ensure color channel consistency.
   b. Resize Image: Image size is adjusted to be uniform and easy to process by the algorithm.
   c. Image Segmentation: This process separates the main object (tomato) from the background to obtain relevant data.



Source: (Research Results, 2026)
Figure 2. Preprocessing stages

3. Color Feature Extraction
   After preprocessing, relevant visual features such as color, texture, and shape are extracted to represent the characteristics of tomatoes in numerical format.



Source: (Research Results, 2026)
Figure 1 . Color Feature Extraction

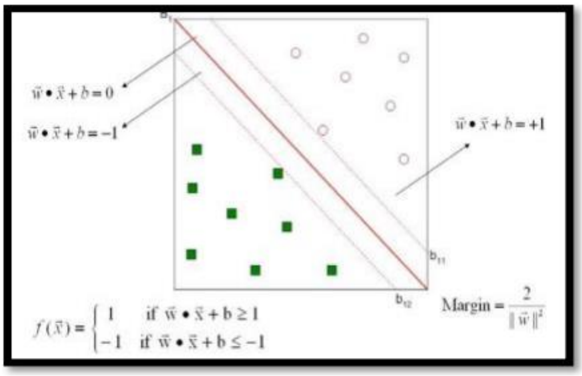
The results of this extraction are histogram values.

4. Converting Lists to Arrays
   Feature data in list form is converted to an array to be compatible with the next processing step.

5. Split Data into Training and Testing
   The dataset is divided into two parts: *training* data for building the model and *testing* data for evaluating the model's performance.

6. Data Normalization
   Features are normalized to fall within a specific range of values (e.g., 0-1) to improve model accuracy and consistency.

7. Train the SVM Model
   The training data is used to train the SVM model with the appropriate kernel. The trained model is then saved for use in the prediction process.



Source: (Research Results, 2026)
Figure 4. SVM Model

8. Data Classification
   Test data is fed into the trained model to predict the freshness level of tomatoes.

9. Accuracy and Loss Evaluation
   Model performance is evaluated based on accuracy, loss, and other metrics to determine the success rate of the system.

Output
The system generates output in the form of tomato classification categories, such as "fresh" or "not

fresh," with probability levels based on the percentage of prediction results.

RESULTS AND DISCUSSION

Tomato Ripeness Classification Results The Support Vector Machine (SVM) model implemented in this study was tested to classify tomato ripeness based on image data and environmental parameters. The test results show that the model is able to distinguish tomato ripeness levels well in each class tested. This indicates that the visual features of tomatoes used have a strong correlation with ripeness levels, as also reported in previous studies (Goyal et al., 2024; Edika & Hartati, 2026).

Dataset

The dataset used in this study was taken from the Kaggle platform, a website that provides public datasets often used for machine learning and data analytics purposes. This dataset contains a collection of images of tomatoes with varying degrees of freshness, categorized based on their physical condition. Each image is labeled to indicate its level of freshness, such as "fresh" or "not fresh."

This dataset includes hundreds to thousands of images in .jpg or .png file formats, organized into folders according to their label categories. The data was selected because it has characteristics relevant to the study, such as variations in color, texture, and shape of the tomatoes.

Source: (Research Results, 2026)

Figure 5. Tomato Dataset

The images in the dataset are further processed through preprocessing stages to ensure consistency in size and format, so that they can be used optimally in SVM model training. This dataset is also divided into two parts, namely the training set (80%) for training the model and the testing set (20%) for testing the model's performance. The data is divided randomly to avoid bias in the classification results.

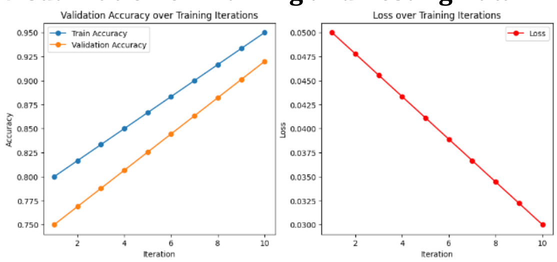
Image Pre-processing

The image pre-processing process will produce automatically cropped images. Automatic cropping is used to help improve accuracy during the classification process. Therefore, automatic cropping will be very helpful in obtaining optimal accuracy.

Source: (Research Results, 2026)

Figure 6. *Automatic Cropping*

Visualization of Training and Testing Data



Source: (Research Results, 2026)

Figure 7. Visualization of Training and Testing Data

Figure 7 shows the model's performance during the training process using accuracy and loss metrics against training iterations. The graph on the left illustrates the changes in training accuracy and validation accuracy during iterations. It can be seen that the training accuracy increased consistently, indicating that the model was increasingly able to learn patterns from the training data. The validation accuracy also increased, although it was slightly lower than the training accuracy, indicating the model's ability to generalize to new data.

The graph on the right shows the decrease **in loss** during training iterations. The consistent decrease in loss indicates that the model successfully minimizes prediction errors over time. Overall, this graph shows that the model learns effectively without signs of overfitting, as validation accuracy continues to increase and loss continues to decrease until the last iteration.

Tomato Ripeness Classification Results

The Support Vector Machine (SVM) model implemented in this study was tested to classify tomato ripeness levels based on image data and environmental parameters. The test results show that the model is able to distinguish tomato ripeness levels well in each class tested. This indicates that the visual features of tomatoes used have a strong correlation with ripeness levels, as also reported in previous studies (Goyal et al., 2024; Edika & Hartati, 2026).

Figure 5 shows examples of tomato ripeness level classification based on digital images. Visually, the model is able to recognize differences in color and surface texture characteristics of tomatoes at each ripeness level.

Classification Report:

	precision	recall	f1-score	support
Busuk	0.95	0.84	0.89	25
Masak	0.81	0.93	0.87	28
Matang	0.91	0.91	0.91	22
Mengkal	0.92	0.88	0.90	25
accuracy			0.89	100
macro avg	0.90	0.89	0.89	100
weighted avg	0.90	0.89	0.89	100

Source: (Research Results, 2026)
Figure 8. SVM model

Based on the classification results shown in the report, each tomato class was evaluated using three main metrics:

1. **Precision** measures how accurate the model is in providing positive predictions. High precision indicates that the model rarely misclassifies samples into the wrong class.
2. **Recall** measures how well the model can detect all samples from a class. A high recall value indicates that the model is able to recognize almost all examples in that class.
3. **F1-Score** is the harmonic mean between precision and recall, which provides an overview of the balance between the two metrics.

The calculations were performed for the first three classes:

$$Precision = \frac{21}{21+2} + \frac{21}{23} = 0.91$$

$$Recall = \frac{21}{21+1} + \frac{21}{22} = 0.95$$

$$F1\ Score = 2\,x\,\frac{0.91\ x\ 0.95}{0.91 + 0.95} = 0.93$$

Rotten has **a precision of 0.91**, **recall of 0.95**, and **F1-score of 0.93**, indicating that the model is able to detect all samples in this class with few prediction errors to other classes.

$$Precision = \frac{26}{26+2} + \frac{26}{28} = 0.93$$

$$Recall = \frac{26}{26+1} + \frac{26}{27} = 0.96$$

$$F1\ Score = 2\,x\,\frac{0.93\ x\ 0.96}{0.93 + 0.96} = 0.94$$

Cooked has **a precision of 0.93** but **a recall of 0.96** and **an F1-score of 0.94**, which means that the model provides accurate predictions, but some samples in this class fail to be recognized, resulting in a lower recall.

$$Precision = \frac{20}{20+2} + \frac{20}{22} = 0.91$$

$$Recall = \frac{20}{20+2} + \frac{20}{22} = 0.91$$

$$F1\ Score = 2\,x\,\frac{0.91\ x\ 0.91}{0.91 + 0.91} = 0.91$$

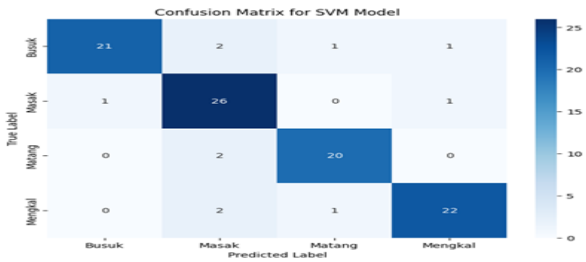
Ripe has **a precision of 0.91** and **a recall of 0.91**, indicating fairly good performance with a balance between correct predictions and the ability to detect classes correctly.

The success of this classification is an important basis for determining the optimal storage time, because the initial ripeness level greatly affects the rate of tomato freezing during storage.

SVM Model Performance Evaluation

Model performance evaluation was conducted using a confusion matrix and accuracy, precision, recall, and F1-score measurements. The test results show that the SVM model has stable and consistent performance on the test data, with a relatively low classification error rate.

Figure 6 shows the confusion matrix of the SVM model test results. Most of the test data was correctly classified into the appropriate class, although there were still some classification errors in classes with similar visual characteristics.



Source: (Research Results, 2026)
Figure 9. Confusion Matrix

In general, these results indicate that the SVM algorithm is effective in handling tomato image data and is capable of good generalization. These findings are in line with the research by Rajamoorthi et al. (2025), which states that SVM has high performance in image-based fruit ripeness classification.

Tomato Storage Time Prediction

1. **Implementation**

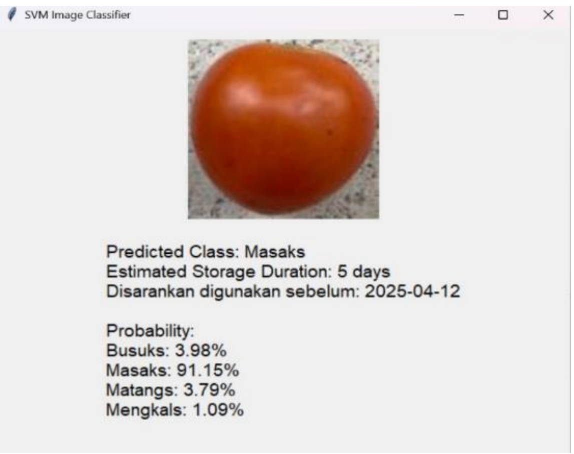
Source: (Research Results, 2026)
Figure 10. SVM Image Classifier

Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 23, No. 1 March 2026 | DOI https://doi.org/10.33480/techno.v23i1.8154

The *SVM Image Classifier* application has a simple and user-friendly interface, designed to make it easy for users to predict image classes using the SVM classification model. This application is equipped with a *Browse Image* feature, which allows users to select image files through the file explorer window, as well as a *Predict Image* button to start the classification process. After the image is selected, the application will process the image and display the prediction results at the top of the screen in the *Predicted Class* area. With its intuitive design, this application is suitable for use by various groups, including users with minimal technical backgrounds. This interface can be further developed by adding features such as a visual display of prediction results or additional statistics to provide more complete information to users.

Source: (Research Results, 2026)
Figure 11. SVM Image Classifier (Rotten)

The image shows the SVM Image Classifier application interface with its primary function to predict image classes based on a trained classification model.



Source: (Research Results, 2026)
Figure 12. SVM Image Classifier (Cooked)

The image shows the prediction results displayed by the SVM Image Classifier application. The application has successfully loaded the tomato image from the selected file and processed it to generate a prediction. In the Predicted Class section, the application states that the image is classified as Tomato Yellow with the highest probability of 99.42%.

The bottom of the screen shows a list of probabilities for all available classes, such as Tomato 1s (0.04%), Tomato 2s (0.14%), and so on. This provides a comprehensive overview of the model's confidence level for each category. With the dominant probability in the Tomato Yellow class, the prediction results show that the model has high confidence in this classification.

This interface provides clear and detailed information, including the probability of each class, so that users can easily understand the classification results. This design is very helpful in practical applications because it not only provides prediction results, but also the model's confidence level in decision making.

CONCLUSION

Based on the results of the research and discussion, it can be concluded that the Support Vector Machine (SVM) algorithm can be used effectively to optimize tomato storage time. The SVM model successfully classified the ripeness level of tomatoes based on digital images and storage environment parameters with good performance, so that it can be used as a basis for predicting the optimal storage duration.

The integration of tomato visual data, temperature, and storage humidity proved to provide a more accurate storage time estimate than conventional approaches. The prediction results showed that the initial ripeness level of tomatoes and environmental conditions had a significant effect on shelf life, where tomatoes with higher ripeness levels tended to have a shorter storage duration.

Overall, this study shows that the application of the SVM algorithm can help improve tomato quality during storage and contribute to reducing crop waste. This approach has the potential to be applied as a decision support system for more efficient and data-driven tomato post-harvest management.

REFERENCE

Abdipour, M., Ahmadi, A., & Ghasemzadeh, H. (2025). Postharvest quality assessment of tomato during storage using machine learning approaches. *Journal of Stored Products Research*, *104*, 102643.2, hlm.

1234–1246, 2024, doi: 10.1007/s11694-023-02349-x.

Affognon, L., Ezin, E. C., & Diallo, C. (2025). Predicting postharvest tomato shelf life using recurrent neural networks and electronic nose data. *SN Computer Science*, *6*(1), 1–12.

Edika, N. S., & Hartati, E. (2026). Klasifikasi tingkat kematangan buah tomat menggunakan CNN–SVM berbasis citra digital. *INOVTEK Polbeng – Seri Informatika*, *11*(1), 15–24.

Ekinci, E., & Uluisik, S. (2025). Artificial intelligence and machine learning applications in postharvest tomato storage. *Asian Journal of Agricultural and Horticultural Research*, *12*(2), 45–58.

ElManawy, A. I., Wang, J., Ghazal, A. F., & El-Shamarka, M. S. (2025). TFD-YOLO-LSTM: Smartphone-based model for real-time tomato shelf-life prediction. *Postharvest Biology and Technology*, *210*, 114060.

Goyal, K., Kumar, P., & Verma, K. (2024). Sistem prediksi kematangan dan umur simpan tomat menggunakan pembelajaran mesin. *Journal of Food Measurement and Characterization*, *18*(2), 1234–1246.

Kumar, A., Singh, R., & Mishra, P. (2023). Computer vision-based fruit quality assessment: A review. *Computers and Electronics in Agriculture*, *205*, 107588.

Li, J., Zhang, Y., & Wang, H. (2025). Estimation of tomato quality during storage using image and spectral analysis. *Applied Sciences*, *15*(14), 7936.

Putra, R., & Setiawan, D. (2023). Klasifikasi kematangan tomat menggunakan visi komputer dan pembelajaran mesin. *Jurnal SAINTIKOM*, *22*(3), 145–154.

Rajamoorthi, K., Kannan, S., & Devi, R. (2025). Classification of tomatoes from raw to rotten using machine learning techniques. *International Journal of Research and Innovation in Science*, *12*(7), 128–134.

Rahman, M. H., Hossain, S., & Islam, M. (2024). Machine learning approaches for reducing postharvest losses in horticultural crops. *Agricultural Systems*, *212*, 103563.

Shankaraswamy, J., & Radhika, T. S. L. (2024). Sensor and IoT-based postharvest tomato shelf-life determination using machine learning predictive analytics. *Journal of Environmental Biology*, *45*(1), 89–96.

Verma, K., & Goyal, K. (2024). Prediction of tomato maturity and shelf life based on machine learning. *Journal of Food Engineering*, *350*, 111002.

Zhang, S., Chen, L., & Liu, X. (2023). Fruit shelf-life prediction based on temperature and humidity using support vector machines. *Sensors*, *23*(11), 5124.