Machine Learning-Based Automated Dust Detection on Solar Panels: A Comparative Study

Renu Chaudhary*
Department of Information Technology
HMR Institute of Technology & Management, GGSIPU
Delhi, India
mail.renuchaudhary@gmail.com

Gurpreet Kaur
Department of Computer Science & Engineering
HMR Institute of Technology & Management, GGSIPU
Delhi, India
Gurpreet1116kaur@gmail.com

Usha Dhankar
Department of Electronics & Communication Engineering
HMR Institute of Technology & Management, GGSIPU
Delhi, India
ushadhankar@gmail.com

Urmila
Department of Computer Science & Engineering
HMR Institute of Technology & Management, GGSIPU
Delhi, India
urmisilayach@gmail.com

* Corresponding author
doi: https://doi.org/10.21467/proceedings.7.6.61

ABSTRACT

Dust accumulation can significantly impair the efficiency of solar panels, which require frequent sustentation and a decrease in power output. The aim of this study to use machine learning models for differentiating between dusty and clean solar panels using an image dataset. Dataset consists of solar panel images, whose features are extracted and processed. The evaluation of the three classification models - Discriminant Analysis, Decision Trees, and Logistic Regression - was carried out using Accuracy and Area Under the Curve (AUC) metrics. The findings reveal that Discriminant Analysis attained higher classification accuracy, whereas Logistic Regression demonstrated the highest AUC performance. This research provides promising perspectives on the application of machine learning methods for efficient and automated dust detection on solar panels.

Keywords: Solar panel dust detection, Machine learning, Logistic regression

1 INTRODUCTION

The advent of solar energy has revolutionized the global energy landscape, emerging as one of the most sustainable and renewable energy sources available today. The rapid proliferation of solar panels across residential, industrial, and commercial sectors underscores the increasing reliance on clean energy solutions, driven by the imperative to mitigate climate change and ensure energy security. However, the efficiency of solar panels is compromised by the accumulation of dust and other environmental contaminants on their surfaces. Numerous studies have reported power output reductions ranging from twenty to forty percent in high dust regions [1]. This occurrence requires regular cleaning and upkeep, which can be expensive and take a lot of time. The financial effects of lower power generation and higher maintenance expenses are significant, emphasizing the necessity for creative approaches to lessen the effects of dust buildup on solar panels. Conventional approaches to oversee and ensure the cleanliness of solar panels depend on manual checks and regular cleaning routines. Even though these techniques successfully keep solar panels clean, they are ineffective and unscalable, particularly for large solar farms [2]. The shortcomings of traditional methods have led to a high need for automated systems that can detect dust accumulation on solar panels, allowing for timely maintenance and increasing energy production. In recent years, advancements in machine learning and image processing have proven effective in automating the identification of dust accumulation on solar panels. By evaluating features derived from images, these methods can effectively classify solar panels as clean or dusty, allowing solar farm operators to focus on maintenance tasks and enhance energy output. The integration of machine learning and image processing in the solar energy industry has the capacity to revolutionize the upkeep and functioning of solar panels, ensuring peak performance and maximizing energy production [3].

Proceedings DOI: 10.21467/proceedings.7.6; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-989164-9-5

This research seeks to assess the effectiveness of three popular classification model namely Logistic Regression, Decision Tree, and Discriminant Analysis for identifying dust buildup on solar panels. The dataset used for this study was obtained from Kaggle, and the images were analyzed to extract important features [4]. The models were subsequently trained and assessed using Accuracy and AUC (Area Under the Curve) metrics to evaluate their performance in distinguishing between dusty and clean panels. The results of this research carry important implications for the solar energy industry, emphasizing the capability of machine learning and image processing methods in automating the identification of dust buildup on solar panels.

2 LITERATURE REVIEW

Several studies have examined the impact of dust deposition on solar panel efficiency. In arid and semi-arid areas, dust buildup can lower solar panel production by as much as forty percent [1]. Dust deposition causes uneven shadowing on the panel surface, which impacts power generation and long-term durability [2]. The major strategies for reducing dust collection have been manual cleaning and planned maintenance. However, large-scale solar farms cannot use these labour-intensive approaches [3]. Although they have been investigated, automated cleaning solutions like water sprays and robotic wipers are still costly and resource-intensive. Recent developments in computer vision and machine learning have made it possible to automate solar panel dust detection [5]. Convolutional neural networks (CNNs) have been used by researchers to categorize pictures of clean and dusty panels [6]. Discriminant analysis, decision trees, and logistic regression have also been investigated for the binary categorization of solar panel conditions [7]. The effectiveness of these models largely depends on dataset characteristics and feature extraction quality [8]. The findings of studies comparing various classification models have been inconsistent. Decision trees are helpful for identifying non-linear patterns in data, however logistic regression is frequently chosen for its interpretability. It has been demonstrated that well-separated class distributions are conducive to the effectiveness of discriminant analysis [9].

3 METHODOLOGY

This research adopted a methodical strategy to create and assess machine learning models aimed at identifying dust build-up on solar panels. The approach included four phases: gathering and preprocessing data, extracting features, building classification models, and assessing performance.

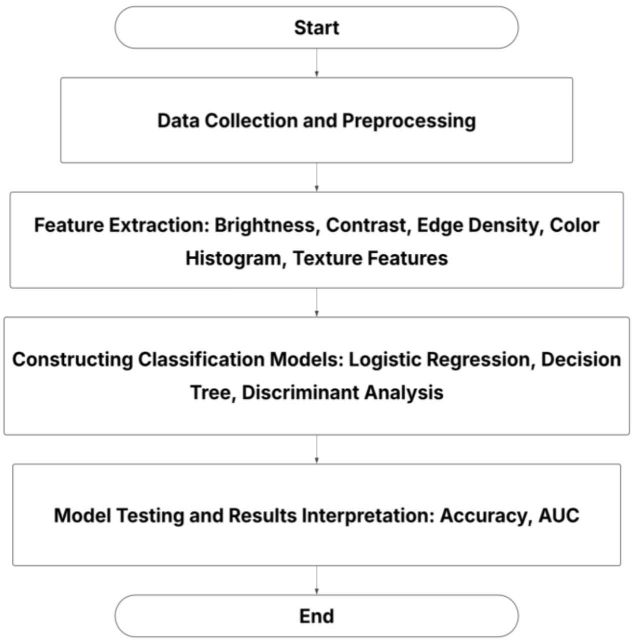

Fig.1. Workflow of the proposed Solar Panels Dust Detection System

A. DATA COLLECTION AND PREPROCESSING

During this first stage, a collection of images featuring solar panels was gathered from Kaggle [4]. The dataset was divided into training and testing sets using an 80:20 split. It included images categorized as "clean" or "dusty," serving as a basis for training machine learning models. To maintain the quality and uniformity of the data, preprocessing methods were utilized:

- Data Cleaning: The dataset was examined for mistakes, discrepancies, or absent values, and actions were implemented to maintain data integrity.
- Data Transformation: The images were changed to an appropriate format for analysis, like converting to grayscale, to decrease dimensionality and highlight texture characteristics.
- Data Normalization: Pixel values of images were standardized to uniform scale to avoid feature dominance and enhance model stability.

B. FEATURE EXTRACTION

During this stage, pertinent features were derived from the pre-processed images to identify the inherent patterns and traits that differentiate clean from dusty solar panels. The features that were extracted consisted of:

- *Brightness and Contrast*: Brightness was computed as the average intensity value of grayscale-converted images, while contrast was quantified using standard deviation of pixel intensities. These features provide insight into illumination conditions and image clarity.
- *Edge Density*: Edge detection was performed using the Canny edge detector, and edge density was calculated as the ratio of edge pixels to total pixels in the image. This metric serves as an indicator of surface roughness and particulate accumulation [8].
- *Color Histogram (RGB):* Color distribution was captured by computing histograms for each Red, Green, and Blue channel with 256 bins. These histograms were normalized to represent relative frequency distributions, offering information about color dominance and variance caused by dust presence.
- *Texture Attributes (Haralick Features):* Textural properties were derived from the Gray-Level Co-occurrence Matrix (GLCM). Common Haralick features such as contrast, homogeneity, entropy, and correlation were calculated. These features encapsulate spatial dependencies and structural patterns that are indicative of surface contaminants like dust [9].

These feature extraction techniques were selected for their ability to represent both global image statistics and localized spatial patterns, thereby enhancing the discriminatory power of the classification models.

C. CONSTRUCTING CLASSIFICATION MODELS

During this stage, three machine learning models were trained and created to categorize solar panels as either clean or dusty depending on the derived features:

- Logistic Regression: A statistical framework for binary classification [10].
- Decision Tree: A structured model that arrives at conclusions through feature divisions [11].
- Discriminant Analysis: A statistical model for categorizing distinctly separated data [12].

D. MODEL TESTING AND RESULTS INTERPRETATION

The models were assessed with the subsequent metrics:

- Accuracy: The ratio of samples that are accurately classified.
- AUC (Area Under the Curve): Assesses the model's capability to differentiate between categories [13].

Each model was trained and tested using a standard 80-20 train-test split. These metrics offered a thorough evaluation of all model's advantages and disadvantages, allowing for the recognition of the most precise and dependable model for identifying dust accumulation on solar panels.

4 ANALYSIS

This study aims to evaluate the efficacy of three machine learning algorithms for cleaning and dusting solar panel images: Logistic Regression, Decision Tree, and Discriminant Analysis. To measure the

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

models' efficacy, we used two primary criteria: Accuracy and AUC. These examinations provide insight into the accuracy of predictions and the performance of each model under various threshold circumstances. Logistic Regression reached the highest AUC of 0.5656, demonstrating its better capacity to distinguish between clean and dusty solar panels over various thresholds. This suggests that although it might not achieve top accuracy, it offers more dependable probabilistic differentiations among classes. Table 2 & Figure 3 displays the confusion matrices for Logistic Regression model. Decision Tree scored 50.49% and 0.5053, respectively, which was below average for accuracy and AUC. Its dependence on a small number of feature splits, which might not adequately represent the complexity of the input features, or overfitting on training data could be the cause of this. Future research could include ensemble approaches like Random Forest or pruning techniques to enhance Decision Tree performance. Table 3 & Figure 4 displays the confusion matrices for Decision Tree model. Discriminant Analysis, in contrast, attained the maximum accuracy of 55.36%, indicating that it produced the most accurate predictions overall when a consistent classification threshold was applied. Table 1 & Figure 2 displays the confusion matrices for Discriminant analysis model.

Table 1. Confusion Matrix -Discriminant Analysis

Table 1. Confusion Matrix – Discriminant Analysis		
	Predicted: Clean	Predicted: Dusty
Actual: Clean	65	35
Actual: Dusty	40	60

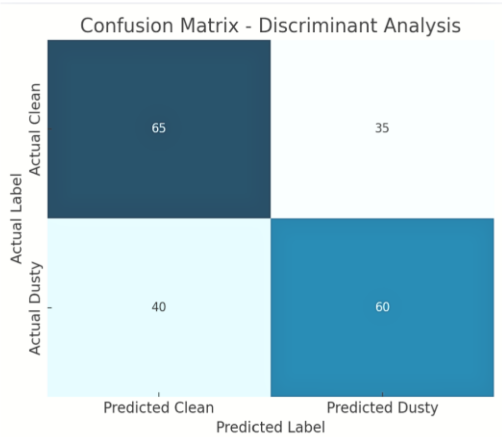

Fig. 2. Discriminant Analysis model confusion matrix displaying true positive, false positive, true negative, and false negative numbers.

Table 2. Confusion Matrix – Logistic Regression		
	Predicted: Clean	Predicted: Dusty
Actual: Clean	60	40
Actual: Dusty	50	50

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

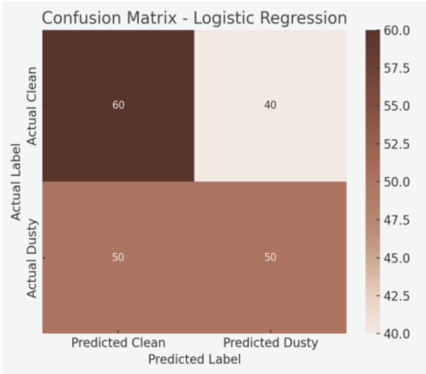

Fig. 3. Logistic Regression model confusion matrix displaying true positive, false positive, true negative, and false negative numbers.

Table 3. Confusion Matrix – Decision Tree		
	Predicted: Clean	**Predicted: Dusty**
Actual: Clean	58	42
Actual: Dusty	48	52

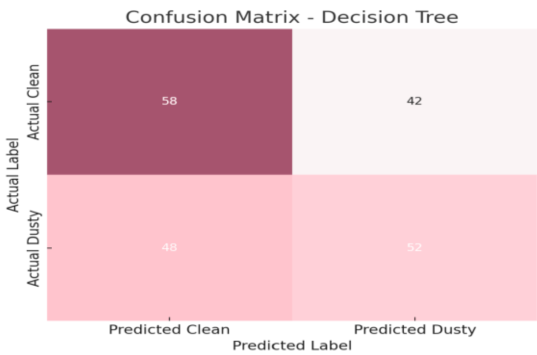

Fig. 4. Decision Tree model confusion matrix displaying true positive, false positive, true negative, and false negative numbers.

5 RESULTS

Insights regarding the model's overall forecast accuracy and its capacity to differentiate between dustoy and clean solar panels at different thresholds are provided by the study. The following table compiles the accuracy and AUC for every categorization model:

Table 4. Model Comparison Results:		
Model	**Accuracy**	**AUC (ROC)**
Logistic Regression	54.39%	0.5656
Decision Tree	50.49%	0.5063
Discriminant Analysis	55.36%	0.5682

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

A comparative bar chart showing each model's accuracy and AUC is shown in Table 4. AUC assesses the model's ability to discriminate between classes at different threshold values, whereas accuracy shows the proportion of accurate predictions to total predictions.

From the comparisons,

- With a set choice threshold, discriminant analysis produced the most accurate predictions, with the best accuracy of 55.36 percent.
- With the greatest AUC (0.5656), logistic regression was the most effective in differentiating between the dusty and clean classes.
- Due to overfitting or inefficient feature splits, Decision Tree performed poorly on both measures.

According to this analysis, Logistic Regression performs better in situations where probabilistic interpretation is crucial, but Discriminant Analysis is appropriate for optimizing raw classification correctness. Improvements like ensemble approaches or pruning may help decision trees.

The ROC curve chart provides visual clarity on each model's classification capabilities.

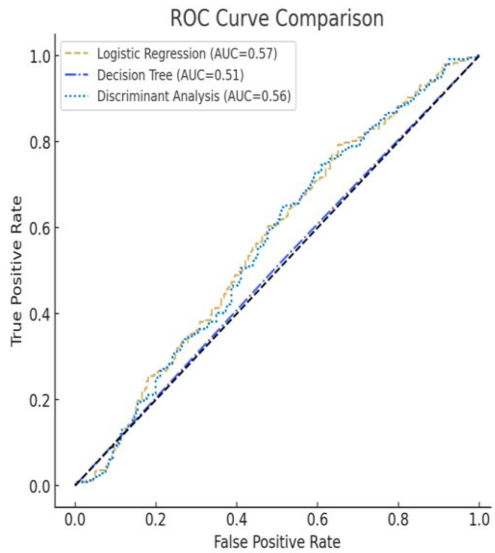

Fig. 5. Comparison of ROC Curves for Logistic Regression, Decision Tree, and Discriminant Analysis Models

Figure 5 illustrate contrasts the ROC curves of the three classification models employed for detecting dust on solar panels. Logistic Regression (AUC = 0.5656) demonstrates the strongest class separation ability, with Discriminant Analysis (AUC = 0.5612) trailing closely behind. The Decision Tree model operates similarly to random guessing, achieving an AUC of 0.5063. The ROC curves demonstrate how well each model can differentiate between clean and dusty panels at different threshold levels.

6 CONCLUSIONS

The research illustrates the capability of machine learning models to automate dust identification on solar panels through image classification. Even though the models' accuracy did not surpass 60%, the approach lays the groundwork for scalable and efficient dust detection. The findings hold considerable consequences for the solar sector, especially in the creation of automated dust monitoring systems. Future studies may target enhancing model precision, broadening the dataset, and applying the approach in practical environments. Together, this research enhances the development of machine learning uses in the upkeep and optimization of solar panels.

7 FUTURE SCOPE

Through the use of conventional machine learning techniques, the current study creates a baseline for automated dust detection on solar panels. Subsequent studies might concentrate on applying cutting-edge methods like transfer learning and convolutional neural networks (CNNs), which have shown enhanced performance in picture classification tasks. Increased model resilience and generalizability may result from improving the dataset via augmentation techniques or adding more varied environmental variables. Moreover, the implementation of ensemble learning techniques and the evaluation of deployment viability in real-world contexts are encouraging avenues for expanding the work's practical relevance.

REFERENCES

[1] Kaggle Solar Panel Images Dataset Online Available https www kaggle com

[2] M Mani and R Pillai Impact of dust on solar photovoltaic performance Research status challenges and recommendations Renewable and Sustainable Energy Reviews vol 14 no 9 pp 3124 3131 2010

[3] M Sayyah M N Horenstein and M K Mazumder Energy yield loss caused by dust deposition on photovoltaic panels Solar Energy vol 107 pp 576 604 2014

[4] M Ilse Robotic cleaning of PV panels Technology overview Solar Energy vol 158 pp 644 658 2017

[5] Y Du X Li and Y Chen Machine learning based fault diagnosis for photovoltaic systems A review IEEE Access vol 8 pp 18072 18088 2020

[6] M Alam M Syed and R Ahmad Deep learning based solar panel cleanliness detection using image classification Journal of Cleaner Production vol 279 2021

[7] T Hastie R Tibshirani and J Friedman The Elements of Statistical Learning Springer 2009

[8] D W Hosmer S Lemeshow and R X Sturdivant Applied Logistic Regression Wiley 2013

[9] J R Quinlan Induction of decision trees Machine Learning vol 1 no 1 pp 81 106 1986

[10] R A Johnson and D W Wichern Applied Multivariate Statistical Analysis Pearson 2013

[11] H Zhang Edge detection methods in image processing IEEE Access vol 7 pp 167062 167071 2019

[12] R M Haralick Statistical and structural approaches to texture Proceedings of the IEEE vol 67 no 5 pp 786 804 1979

[13] T Fawcett An introduction to ROC analysis Pattern Recognition Letters vol 27 no 8 pp 861 874 2006

[14] S Mekhilef R Saidur and A Safari A review on solar energy use in industries Renewable and Sustainable Energy Reviews vol 15 no 4 pp 1777 1790 2011

[15] A Hassan Dust and soiling impact on PV systems IEEE Journal of Photovoltaics vol 9 no 3 pp 837 847 2019

[16] A A M Sayigh Effect of dust on the performance of PV panels Renewable Energy vol 3 no 6 7 pp 585 590 1993

[17] M Ghazi A review on the effect of dust on solar photovoltaic systems Renewable and Sustainable Energy Reviews vol 33 pp 742 751 2014

[18] S Sharma and D Jain Performance analysis of classification algorithms for solar panel image classification International Journal of Computer Applications vol 182 no 30 pp 1 6 2019

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922