

Identification of Surface Defects on Solar PV Panels and Wind Turbine Blades using Attention based Deep Learning Model

Divyanshi Dwivedi^{a,b}, K. Victor Sam Moses Babu^{a,c}, Pradeep Kumar Yemula^b, Pratyush Chakraborty^c, Mayukha Pal^{a,*}

^a*ABB Ability Innovation Center, Asea Brown Boveri Company, Hyderabad, 500084, Telangana, India*

^b*Department of Electrical Engineering, Indian Institute of Technology, Hyderabad, 502205, Telangana, India*

^c*Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani- Hyderabad Campus, Hyderabad, 500078, Telangana, India*

Abstract

According to Global Electricity Review 2022, electricity generation from renewable energy sources has increased by 20% worldwide primarily due to more installation of large green power plants. Monitoring the renewable energy assets in those large power plants is still challenging as the assets are highly impacted by several environmental factors, resulting in issues like less power generation, malfunctioning, and degradation of asset life. Therefore, detecting the surface defects on the renewable energy assets would facilitate the process to maintain the safety and efficiency of the green power plants. An innovative detection framework is proposed to achieve an economical renewable energy asset surface monitoring system. First capture the asset's high-resolution images on a regular basis and inspect them to detect the damages. For inspection this paper presents a unified deep learning-based image inspection model which analyzes the captured images to identify the surface or structural damages on the various renewable energy assets in large power plants. We use the Vision Transformer (ViT), the latest developed deep-learning model in computer vision, to detect the damages on solar panels and wind turbine blades and classify the type of de-

*Corresponding author

Email address: mayukha.pal@in.abb.com (Mayukha Pal)

fect to suggest the preventive measures. With the ViT model, we have achieved above 97% accuracy for both the assets, which outperforms the benchmark classification models for the input images of varied modalities taken from publicly available sources.

Keywords: Solar PV panels, Wind turbines, Structural health monitoring, Vision Transformer, Damage detection, Drone inspection, Deep Learning, Renewable Energy Sources.

1. Introduction

1.1. Motivation

Electricity generation by renewable energy sources is set to influence the global power generation capacity, accounting for 75-80% of new installed capacity by 2050 and mainly led by solar and wind [1]. The transition is rapidly undergoing because of the ongoing policies and public demand to shift from fossil fuels to green energy sources for electricity generation. Governments of various countries are in high-priority planning and taking measures to achieve the target of net zero carbon emissions, sustainable economic and social growth [2]. According to International Energy Agency (IEA), wind power and solar power generation are the essential sources responsible for achieving the target for sustainable development of countries all over the world. Governments are coming up with huge investments for commissioning large power plants based on renewable. On May 2022, energy ministers from Germany, Belgium, the Netherlands and Denmark signed an agreement to create the Green Power Plant of Europe in the North Sea as the continent steps up its efforts to cut emissions, reduce reliance on Russian gas imports and could produce 65 GW by 2030 and 150 GW by 2050 [3]. Núñez de Balboa plant in Europe, a largest solar plant has an installed capacity of 500 megawatts with 1.4 million solar panels cover nearly 10 square kilometers [4]. Adani Green Energy (AGEL) has recently installed a 600 MW Solar-Wind Hybrid power plant at Jaisalmer in Rajasthan, India [5]. These massive projects contribute to green energy; however, such large

installations need proper management, monitoring, and maintenance. In context, United Nations (UN) has also announced that the objective of sustainable development is possible when energy management is considered as the primary object [6]. However, the maintenance and monitoring of renewable power plants are less explored and not emphasized, they are sell-pitched as low-maintenance energy sources, which is an erroneous fact. Thus, proper maintenance and monitoring of renewable power plants is important, and it adds on to achieve sustainable development targets, increase power generation and enhance the operating life span of renewable energy assets. In this work, we propose a technique that effectively check the assets and detect the damages or defects. For large renewable power plants, it recommended to perform drone image-based monitoring regularly, examine the collected images, find the defects, and take measures to harness more power because the large-scale installation. Wind turbines are almost of height 65 meters, and solar panels are spread around 60 acres of land; thus it is challenging to identify the defects in the plants [7]. In this work, we focused on detecting damages on the wind turbine and solar PV panels as they are deployed in the system on a large scale, using automated deep learning-based computer vision algorithm as shown in Figure 1. After the identification of defects, we could suggest precautionary measures to improve their performance.

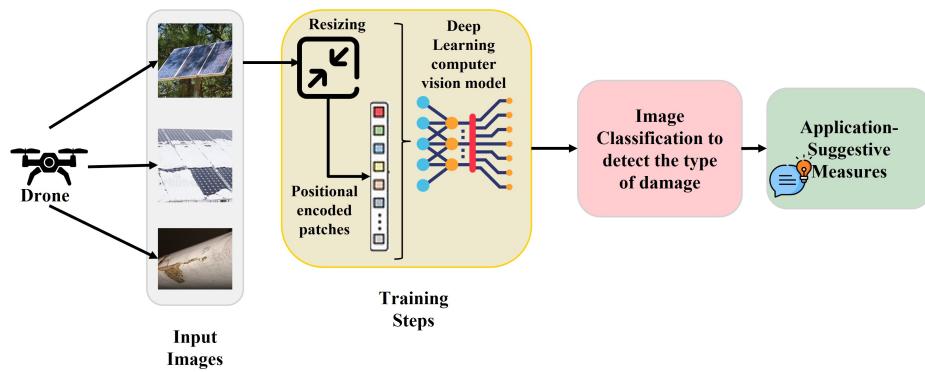


Figure 1: Proposed framework for monitoring and detection of damages in the solar panels and wind turbines.

1.2. Related works

Wind turbines generate power, and their blades are the important component that directly affects the performance and quality of power generated [8]-[9]. Wind turbines are installed in remote locations and exposed to environmental factors such as rain, sun, and wind gust, thus they are highly susceptible to corrosion, damages, and defects [10]. Manual detection is not feasible as it requires a large workforce and time [11]. However, drone images are captured and further analyzed to detect the defects as it would be more reliable, efficient, and cost-effective. In state-of-art, with the help of drone images, features are extracted, and defects are identified. Deep convolutional neural network (DCNN) is built and is trained on the ImageNet Large Scale Visual Recognition to extract the feature, but suitable for small samples of dataset [11]. Histogram of Oriented Gradient (HoG) detects the object locally by gradient distribution of intensity, illumination, and geometric gradients [12]. Other feature extraction techniques used are primitive-based method [13], statistical method [14], spectral method [15], Local binary patterns (LBP) [16], Image visibility graph as feature extractor [17], [18] and Gray-Level Co-occurrence Matrix (GLCM) [19]. These typical image processing techniques extricate the images' low-level features, which are not sufficient to identify and classify the type of defect in the wind turbine blades. Generally, wind turbine blades face issues such as cracks, scrapes, and erosion, which are non-uniform and not easy to differentiate.

Another essential renewable energy source is solar PV panels which are widely used for power generation at small-scale as well as large scale [20]. They are also significantly affected by the environmental factors that degrade their performance and life span. Usually, the lifespan of solar PV panels is 20-25 years, and their proper maintenance could result in harnessing maximum power output throughout their life span [21]. The defects which solar PV panels faces are as follows: accumulation of soil, dust, snow, birds nest or drop, cracks, construction cement deposit, and shadow of overgrown plants or grass, which hamper the performance of Solar PV panels. Usually, monitoring is done based on tracking the amount of energy produced by solar panels, but that is insuffi-

cient to find the root cause behind the dip in power generation and unable to suggest preventive measures. Another technique, convolution neural network (CNN), is applied for visual recognition tasks of the defects, but its classification accuracy is only 83.22%; thus, we cannot rely on the model [22]. Other researchers also used a modified version of CNN to detect the damages in the solar PV panels; those models are also not capable of achieving accuracy above 90% and require more computational time to evaluate the images [23], [24], and [25]. In the literature, several image classification algorithms have been proposed with the increasing availability of labeled datasets, several supervised learning approaches were applied ranging from deep convolutional neural networks [26], [27], attention bottleneck residual network [28], and discriminative and representation learning [29], [30].

1.3. Contribution

In this work, we used an attention-based model called Vision Transformer (ViT) to detect the damages in solar PV panels and wind turbines. Detection of damages from high-resolution drone images of varied modalities needs an effective model which can automatically extract the high-level features from the image in fewer computational hours with good accuracy. In 2017, the introduction of attention in natural language processing is appreciated because of its high performance. Further, using the concept, researchers from Google proposed a transformer-based model for computer vision classification, i.e., Vision Transformer [31]. In various domains, ViT has shown promising performance, and high accuracy for learning tasks [32],[33],[34],[35],[36] and [37]. Thus, in this work, we have used the transformer model to characterize and detect the damages on solar PV panels and wind turbine blades using the drone-inspection technique.

The key contribution of this work is as follows:

- Vision transformer for the first time implemented on electrical power system's problem.

- The framework is suitable to integrate on large-scale renewable power plants for inspecting the damage to assets at a very low cost, with less human intervention and less processing time.
- In comparison to other models, we can achieve better testing accuracy for the proposed model.

The paper is organized as follows: a description of the dataset is provided in Section 2 detailed explanation of the Vision transformer model is presented in Section 3. Section 4 provides the experimental result on a real dataset of the wind turbine blades and solar PV panels. Finally, Section 5 concludes the paper.

2. Description of Datasets

2.1. Wind Turbine Blades Dataset

For the identification of defects on the blades of the wind turbine, the dataset contemplated is taken from Mendeley- Drone inspection images of a wind turbine [38]. The images of wind turbine blades are captured using a Canon 5Ds DSLR camera with a resolution of 8688×5792 . The dataset comprises of images of wind turbine blades without any defects which are considered as reference images, and with defects including damaged area, damaged-edge area, erosion area, and space of rough area as shown in Figure 2. In total, there are 299 images which are further labeled into five classes for training the image processing model as shown in Table 1.

Table 1: Description of wind turbine blade images

| Type of Image | Number of Images |
|---------------|------------------|
| Reference | 16 |
| Damaged | 30 |
| Edge-Damaged | 58 |
| Erosion | 65 |
| Rough | 130 |
| Total | 299 |



Figure 2: Surface damages on Wind Turbine Blades.

2.2. Solar Panels Dataset

For the defects detection in solar panels, we primarily considered Solar Panel Soiling Image Dataset created by Deep Solar Eye [22]. In general, this dataset has 45,469 images in total which are captured by an RGB camera every 5 seconds for a month with a resolution of 192×192 . These images are captured with various fabricated circumstances, such as sand, dust, soil, and white powder, to demonstrate the soiling impact on the solar panels. Further, to examine the various other defects on the solar panels, we have downloaded images of solar panels which are spoiled by the bird drops or nests, covered by the snow, have

cracks, are affected by the shadow of trees, plants, or buildings and affected by hardened cement as shown in Figure 3. These images are resized to the resolution 72×72 for further image processing. The number of images considered in each training class for labeling is shown in Table 2.

Table 2: Description of Solar panel images

| Type of Image | Number of Images |
|---------------|------------------|
| Clean | 267 |
| Dust | 1204 |
| Cement | 760 |
| Bird Poop | 165 |
| Cracks | 73 |
| Snow | 605 |
| Soil | 980 |
| Shadow | 56 |
| Total | 4110 |

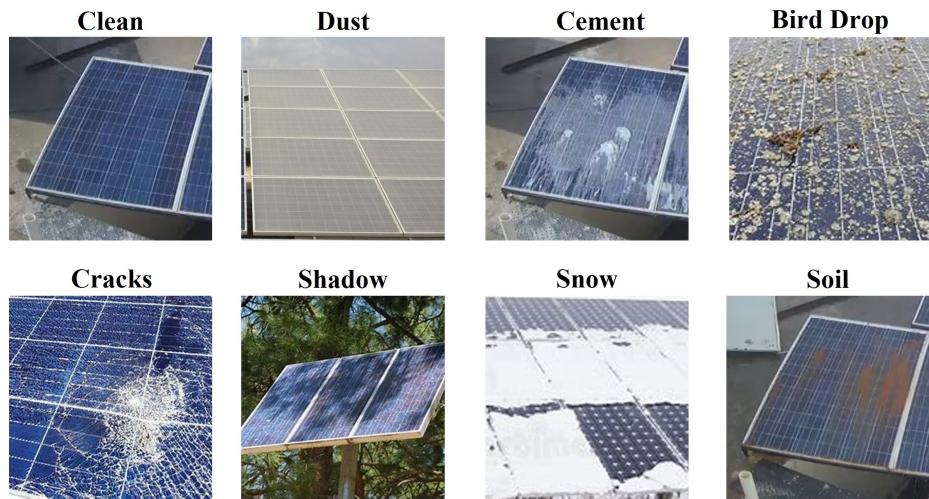


Figure 3: Surface damages on Solar panels.

3. Methods and Materials

In 2017, a transformer was proposed for natural language processing (NLP) on machine translation tasks with great performance [31]. Further, it has been implemented in image processing by Google's research team in 2021 [39], by utilizing the Transformer encoder architecture for the image processing task and named as Vision Transformer (ViT). Transformer in NLP initiate the learning process by measuring the relationship between the 1D input token pairs, while in computer vision, we reshape the images into a sequence of flattened 2D image patches, which are used as the input token for further learning by giving attention in the network. Then these patches are mapped to a constant latent vector with a trainable linear projection, and position embeddings are added to the patch to retain positional information. The implementation process of image classification using ViT is shown in Figure 4.

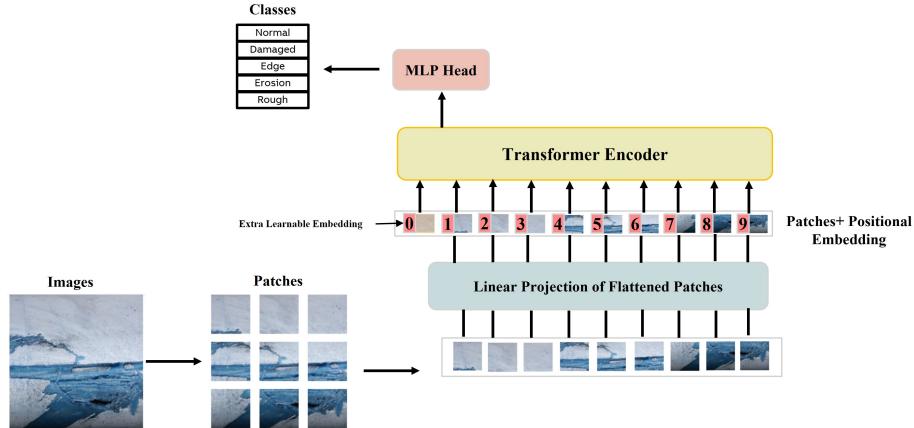


Figure 4: ViT framework with wind turbine blade image as input.

3.1. Architecture of Transformer Encoder

The transformer encoder architecture block comprises of stacked multi-head attention layers, a feed-forward neural network, a shortcut connection, and a normalization layer for both the encoder and decoder as shown in Figure 5.

3.1.1. Encoder Stack

The encoder comprises of six identical layers; each layer performs a multi-head mechanism followed by a positional fully connected feed-forward layer then employed by a residual connection throughout the layer and finally normalization using $\text{LayerNorm}(X + \text{Attention}(X))$, where X is the input of self-attention layer.

3.1.2. Decoder Stack

The decoder also contain six identical layers, each layer also performs a similar multi-head mechanism followed by a positional fully connected feed-forward layer, but in addition, it incorporates multi-head attention over the output of the encoder stack. They also employed by a residual connection throughout the layer and finally followed by normalization.

3.1.3. Self-Attention

In this layer, the input vector is transformed into three vectors: Query Vector (Q), Key Vector (K), and Value Vector (V) of dimension $D_Q = D_K = D_V = D_{\text{Model}} = 512$. The derived vectors are stacked into their respective matrices Q_m, K_m and V_m of size $D_{\text{Model}} \times (N + 1)$ where N is the number of image patches [40]. With the help of these matrices, we could compute the attention function as follows:

- Compute scores between Query and Key Matrices to determine the degree of attention, $S_m = D_m \cdot K_m^T$.
- Normalize the scores for stabilizing the gradient for improvising the training performance, $S_n = S_m / \sqrt{D_{\text{Model}}}$.
- Transform the normalized scores into probabilities with Softmax function, $P_S = \text{Softmax}(S_n)$.
- Finally, obtain the weighted value matrix, $W_m = V_m \cdot P_s$.

The complete process is explained as combined Eq.(1) and in Figure 6(left) as:

$$SelfAttention(Q_m, K_m, V_m) = \frac{Softmax(Q_m \cdot K_m^T)}{\sqrt{D_{Model}}} \cdot V_m \quad (1)$$

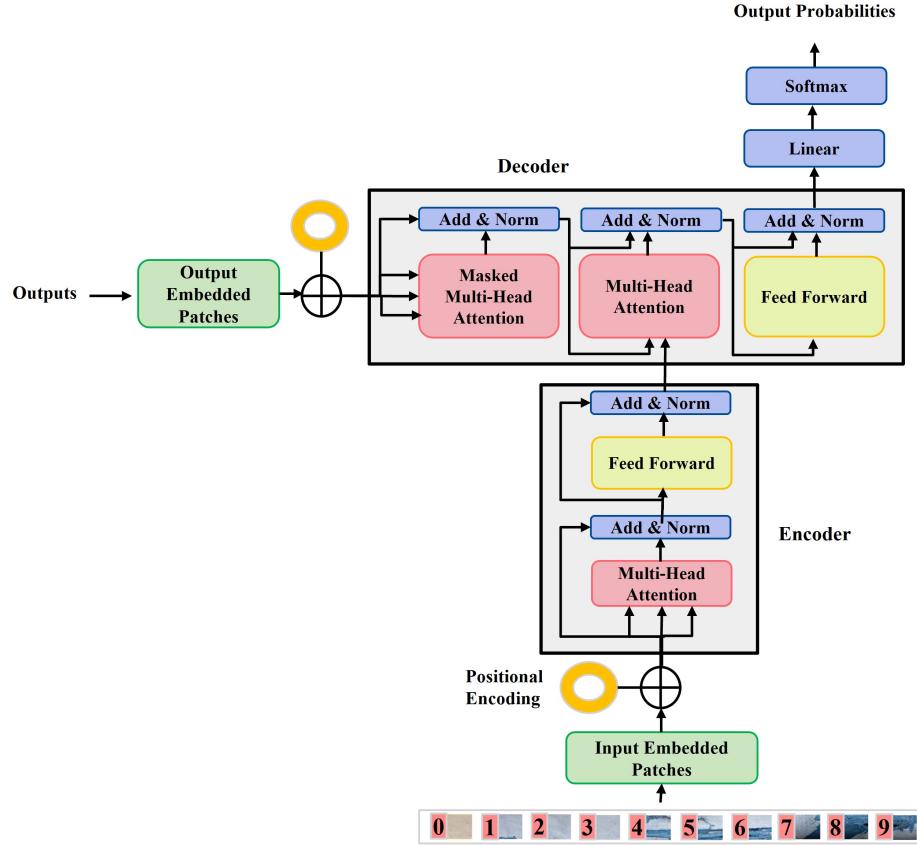


Figure 5: Architectural Diagram of Transformer Encoder- Decoder.

3.2. Multi-Head Attention

Using the single attention function $x - times$, D_{Model} - dimensional queries, keys, and values are linearly projected with different learnable linear projections to their respective dimensions. Further, these projected variants perform the attention function in parallel and result in V -dimensional output values as shown

in Figure 6(right). Multi-head attention facilitates the model to acquire complete information from the represented layers at various positions and illustrated as:

$$\text{Multi}_{\text{Head}(Q_m, K_m, V_m)} = \text{Concat}(\text{head}_1, \dots, \text{head}_x)W^0 \quad (2)$$

where, $\text{head}_k = \text{Attention}(Q_m W_k^{Q_m}, K_m W_k^{K_m}, V_m W_k^{V_m})$
and $W_k^{Q_m} \in R^{D_{Model} \times D_Q}$; $W_k^{K_m} \in R^{D_{Model} \times D_K}$; $W_k^{V_m} \in R^{D_{Model} \times D_V}$.

In this work, we used eight parallel attention layers ($x = 8$),

$$\therefore D_Q = D_K = D_V = \frac{D_{Model}}{x} = \frac{512}{8} = 64.$$

3.3. Feed-Forward Networks

Each Multi-head attention layer is connected to the feed-forward network as shown in Figure 5. It is composed of two linear transformations having a RELU activation function between them.

$$F_{FN}(i) = \max(0, iW_1, b_1) \cdot W_2 + b_2 \quad (3)$$

3.4. Positional Embedding

Input patch images are embedded with positional encodings at the bottom of the encoder and decoder stacks with the dimension D_{Model} . These positional encodings could be implemented by various approaches [41], but here we considered sine and cosine functions at different frequencies as:

$$PE(pos, 2k) = \sin(pos/10000^{2k/D_{model}}) \quad (4)$$

$$PE(pos, 2k + 1) = \cos(pos/10000^{2k/D_{model}}) \quad (5)$$

where, pos is the position and k is the dimension.

Thus, the final architecture of ViT is shown in Figure 4, which includes all the layers as discussed above, where linear image patches are passed through a dense layer to achieve encoded vectors by integrating them with positional embedding. Then positional encoded patches are passed through Transformer

encoder layers to get the contextual vector. Then at the final stage, this vector is passed through a multi-layer head to get the final image classification.

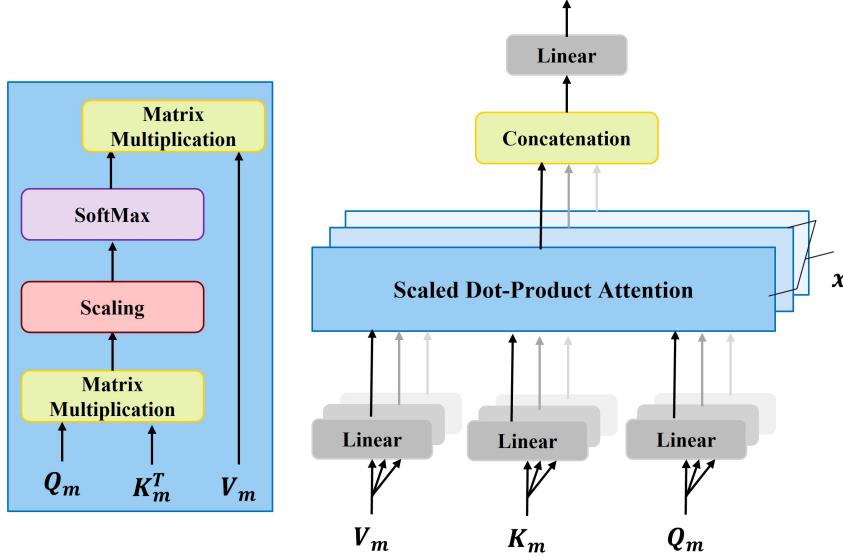


Figure 6: (left) Scaled Dot-Product Attention, (right) Multi-Head Attention consists of several attention layers running in parallel.

3.5. Matrices for Model Analyses

To assess the model, we have computed the following scores:

- Accuracy- It is widely used to analyze the model effectiveness, which compares the total number of accurate predictions with respect to the total number of guesses.
- Recall- It measures the success of prediction under misbalancing. Mathematically, the ratio between truly classified positive cases to the sum of true positive (TP) and false negative (FN) is defined as:

$$Recall_{Score} = \frac{TP}{(TP + FN)} \quad (6)$$

- Precision- It measures the model's ability which would not label the positive sample as negative. Mathematically, it is expressed as the ratio of

true positive to the total predicted positive (including true positive (TP) and false positive (FP)) defined as:

$$Precision_{Score} = \frac{TP}{(TP + FP)} \quad (7)$$

- F1- It measures the precision and recall harmonic mean [42] as:

$$F1_{Score} = \frac{2TP}{(2TP + FN + FP)} \quad (8)$$

- Cohen's Kappa- It is a score that measures inter-annotator agreement, which tells how effective your classifier model is performing in comparison to the classifier which randomly performs the classification. Mathematically, it is expressed as:

$$Cohen_Kappa_{Score} = \frac{P_o - P_e}{1 - P_e} \quad (9)$$

where, P_o is the observed agreement and P_e is the expected agreement.

- Matthew's Correlation Coefficient- It is the most effective and truthful score to evaluate any classifier model. Mathematically, it is expressed as:

$$Matthew_Corr_{Score} = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (10)$$

4. Results and Discussion

All the analyses were performed on NVIDIA RTX 3060ti 8 GB and CPU Intel core I5- 12600K, RAM of 32 GB DDR4 on 64-bit Windows-10 operating system. For training, we used 100 epochs in each run and used 80% of the dataset for training and 20% for testing. Table 3 lists the hyperparameters that were applied during the training phase.

Table 3: Hyperparameters of the ViT Model

| Hyperparameters | Values |
|--------------------|--|
| Batch Size | 32 |
| Number of Epochs | 100 |
| Optimizer | AdamW |
| Learning Rate | 0.001 |
| Weight Decay | 0.0001 |
| Transformer Layers | 8 |
| Heads | 4 |
| Project Dimension | 64 |
| Image Size | Wind Turbine Blades- 256, Solar Panels- 72 |
| Patch Size | Wind Turbine Blades- 16, Solar Panels- 8 |

The solar panels are resized to 72×72 , and wind turbine blades are resized to 256×256 , further flattened into 2D image patches of size 16×16 and 8×8 respectively, as shown in Figure 7. Furthermore, these training images are passed through a dense layer to get the positional embeddings and then finally fed to the transformer encoder. All the mentioned scores in section 3.5 are computed to analyze the ViT model are tabulated in Table 4, and accuracy and cross-entropy curves with respect to epochs for training and validation datasets are shown in Figure 8 for solar panels and wind turbine blades.

Table 4 shows that the model accuracy is 97.33% for wind turbine blade images and 98.66% for solar panel images, which are considered good for any classifier model; other scores are also above 96% for both types of images, which proves that the ViT model is superior and effective. However, Figure 9 reports each class evaluation metric on the testing set (recall, precision, and f1-score) for both types of images.

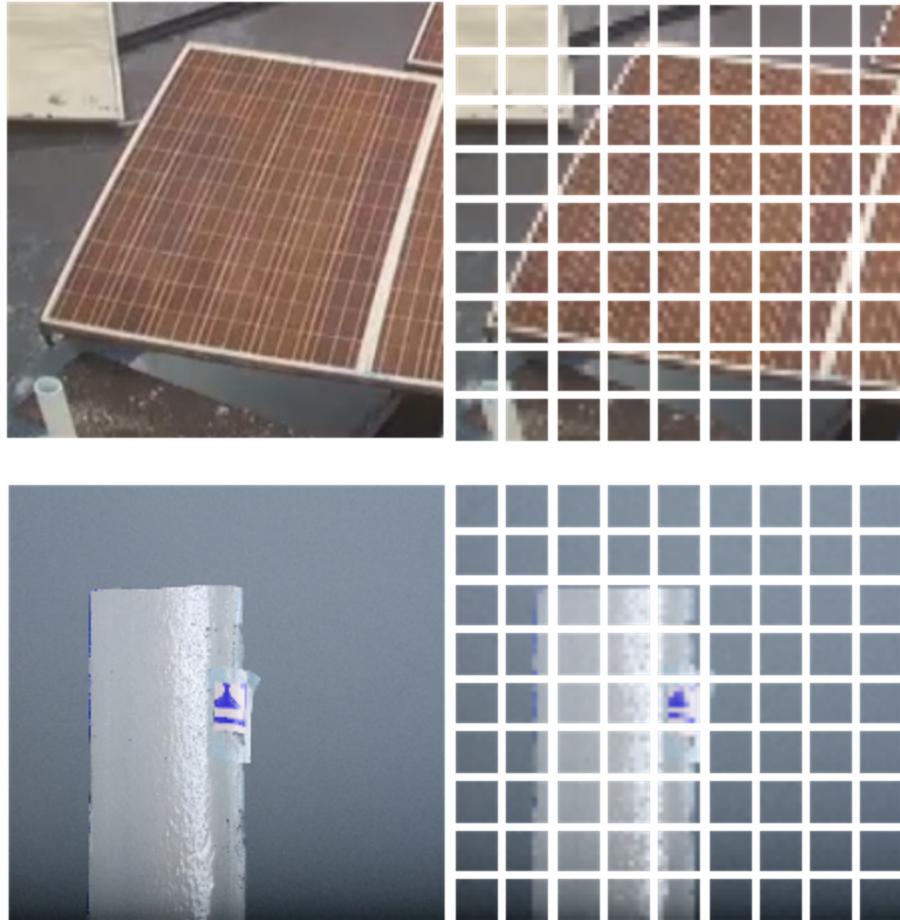


Figure 7: (top) Solar Panel Image of size 72×72 divided into 81 patches of size 8×8 , (bottom) Wind Turbine Blade Image of size 256×256 divided into 256 patches of size 16×16 .

Table 4: Computed scores for both the datasets

| Scores | Solar Panels Images | Wind Turbine Blade Images |
|-----------------------|---------------------|---------------------------|
| Accuracy | 0.9866 | 0.9733 |
| Recall | 0.9900 | 0.9700 |
| Precision | 0.9900 | 0.9800 |
| F1 | 0.9900 | 0.9700 |
| Cohen's Kappa | 0.9828 | 0.9627 |
| Matthew's Correlation | 0.9828 | 0.9635 |

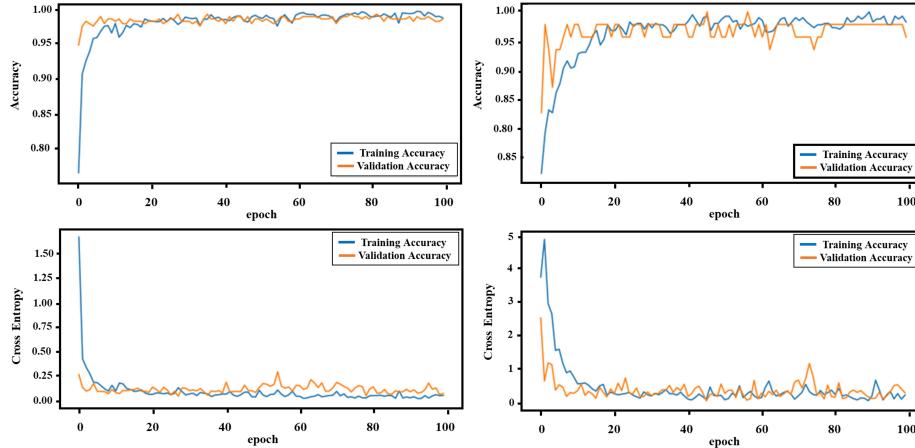


Figure 8: Accuracy and loss curve with respect to epochs of training and validation for (left) Solar Panels images, and (right) Wind Turbine blades images.

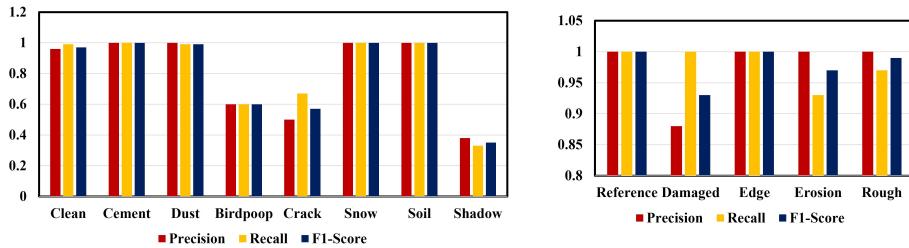


Figure 9: Reports Precision, Recall, F1-Scores for each class labels for (left) Solar Panels images, and (right) Wind Turbine blades images.

These scores are computed based on the confusion matrix, and it helps to estimate the error distribution across the classes. Figure 10 illustrates the confusion matrices for both types of datasets. The rough class is highly sensitive from the wind turbine blade image dataset, and mostly the misclassification is observed for reference images. On the other hand, the dust class is highly sensitive for the solar panel image dataset, and misclassification was mainly observed for shadow images.

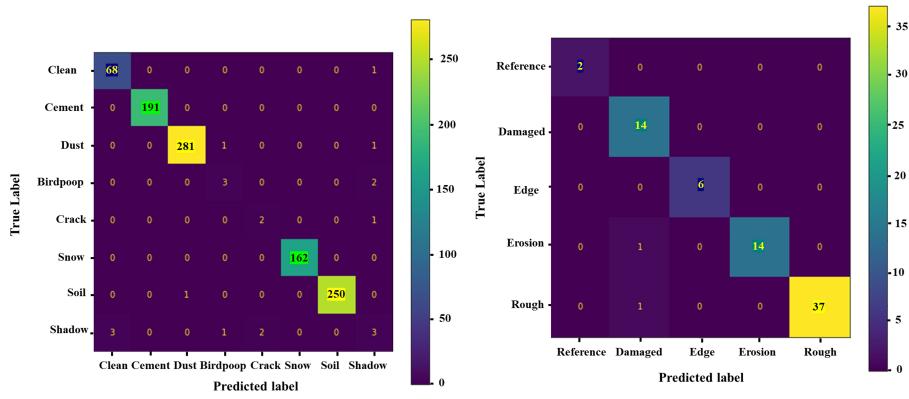


Figure 10: Confusion matrix showing all the labels for (left) Solar Panels images and (right) Wind Turbine blades images.

5. Conclusion

To ensure high generation and low maintenance cost for renewable energy assets, regular monitoring and image detection based on the drill of drone-inspected images is important. In this paper, we explored a dataset of wind turbine blades and solar panels to implement Vision Transformer (ViT), a multi-class image classification using a deep learning model in computer vision. The results showed that the ViT model has effectively learnt to classify the damages in renewable energy assets with more than 97% of accuracy. Thus, this smart inspection in electrical distribution systems installed for renewable energy sources would help reduce the maintenance cost, generate more power and enhance the life of renewable energy assets. Mainly, this work projected the strength of Vi-

sion Transformer in the field of active electrical distribution systems to detect the damages and take the corrective measures to curtail them. Thus, this model would come out as an early smart system to detect the structural damage on the surface of renewable energy assets of the large-scale power utility to supervise proper safety and maintenance of the plant.

Acknowledgement

Divyanshi Dwivedi and K. Victor Sam Moses Babu would like to thank ABB Ability Innovation Centre, Hyderabad for the financial support in research. The author Mayukha Pal would like to thank the ABB Ability Innovation Center, Hyderabad, for their support in this work.

CRediT authorship contribution statement

Divyanshi Dwivedi: Methodology, Software, Data curation, Writing- Original draft. **K. Victor Sam Moses Babu:** Methodology, Data curation, Writing- Original draft. **Pradeep Kumar Yemula:** Supervision, Writing- Reviewing and Editing. **Pratyush Chakraborty:** Supervision, Writing- Reviewing and Editing. **Mayukha Pal:** Conceptualization, Methodology, Project administration, Validation, Supervision, Writing- Reviewing and Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] I. E. Agency, World Energy Outlook 2022, <https://iea.blob.core.windows.net/assets/830fe099-5530-48f2-a7c1-11f35d510983/WorldEnergyOutlook2022.pdf> (2022).

- [2] D. M. Reddy, D. Dwivedi, P. K. Yemula, M. Pal, Data-driven approach to form energy resilient smart microgrids with identification of vulnerable nodes in active electrical distribution network, arXiv preprint arXiv:2208.11682 (2022).
- [3] S. E. International. Green power plant of europe [online, cited 20.11.2022].
- [4] E. Observatory-NASA. The largest solar power plant in europe [online, cited 20.11.2022].
- [5] A. Renewable. Adani Solar Park [online, cited 12.11.2022].
- [6] U. Nations. Sustainable Development Goals [online, cited 12.11.2022].
- [7] H. B. Lund, B. V. Mathiesen, Energy system analysis of 100% renewable energy systems-the case of denmark in years 2030 and 2050, *Energy* 34 (2009) 524–531.
- [8] A. B. Asghar, X. Liu, Adaptive neuro-fuzzy algorithm to estimate effective wind speed and optimal rotor speed for variable-speed wind turbine, *Neurocomputing* 272 (2018) 495–504.
- [9] M. M. Rezaei, M. Behzad, H. Moradi, H. Haddadpour, Modal-based damage identification for the nonlinear model of modern wind turbine blade, *Renewable Energy* 94 (2016) 391–409.
- [10] H. Li, W. Zhou, J. Xu, *Structural Health Monitoring of Wind Turbine Blades*, Springer International Publishing, Cham, 2014, pp. 231–265.
- [11] Y. Yajie, C. Hui, Y. Xingyu, W. Tao, S. G. Shuzhi, Defect identification of wind turbine blades based on defect semantic features with transfer feature extractor, *Neurocomputing* 376 (2020) 1–9.
- [12] Y. Zhou, Y. Liu, G. Han, Z. Zhang, Face recognition based on global and local feature fusion, in: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, pp. 2771–2775.

- [13] T. Hong, C. R. Dyer, A. Rosenfeld, Texture primitive extraction using an edge-based approach, 1979.
- [14] K. Emrith, M. J. Chantler, P. R. Green, L. T. Maloney, A. D. F. Clarke, Measuring perceived differences in surface texture due to changes in higher order statistics., *Journal of the Optical Society of America. A, Optics, image science, and vision* 27 5 (2010) 1232–44.
- [15] T. Randen, J. Husoy, Filtering for texture classification: a comparative study, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (4) (1999) 291–310.
- [16] X. Wang, T. X. Han, S. Yan, An hog-lbp human detector with partial occlusion handling, in: 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 32–39.
- [17] M. Pal, Y. Tiwari, T. V. Reddy, P. Sai Ram Aditya, P. K. Panigrahi, An integrative method for covid-19 patients' classification from chest x-ray using deep learning network with image visibility graph as feature extractor, *Preprints from medRxiv and bioRxiv* (2022). doi:<https://doi.org/10.1101/2021.11.17.21266472>.
- [18] M. Pal, An image visibility graph based method for genomic data sequencing and classification: its applications to detect sars-cov-2 mutant virus variants (2021). doi:[10.13140/RG.2.2.14379.16168/1](https://doi.org/10.13140/RG.2.2.14379.16168/1).
- [19] V. Rodriguez-Galiano, M. Chica-Olmo, F. Abarca-Hernandez, P. Atkinson, C. Jeganathan, Random forest classification of mediterranean land cover using multi-seasonal imagery and multi-seasonal texture, *Remote Sensing of Environment* 121 (2012) 93–107.
- [20] D. Dwivedi, P. K. Yemula, M. Pal, A methodology for identifying resiliency in renewable electrical distribution system using complex network, *arXiv preprint arXiv:2208.11682* (2022).

- [21] A. A. Mansur, M. R. Amin, K. K. Islam, Determination of module rearrangement techniques for non-uniformly aged pv arrays with sp, tct, bl and hc configurations for maximum power output, in: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019, pp. 1–5.
- [22] S. Mehta, A. P. Azad, S. A. Chemmengath, V. Raykar, S. Kalyanaraman, Deepsolareye: Power loss prediction and weakly supervised soiling localization via fully convolutional networks for solar panels, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 333–342.
- [23] W. Zhang, S. Liu, O. Gandhi, C. D. Rodríguez-Gallegos, H. Quan, D. Srivivasan, Deep-learning-based probabilistic estimation of solar pv soiling loss, *IEEE Transactions on Sustainable Energy* 12 (4) (2021) 2436–2444.
- [24] H. P.-C. Hwang, C. C.-Y. Ku, J. C.-C. Chan, Detection of malfunctioning photovoltaic modules based on machine learning algorithms, *IEEE Access* 9 (2021) 37210–37219.
- [25] A. Shihavuddin, M. R. A. Rashid, M. H. Maruf, M. A. Hasan, M. A. ul Haq, R. H. Ashique, A. A. Mansur, Image based surface damage detection of renewable energy installations using a unified deep learning approach, *Energy Reports* 7 (2021) 4566–4576.
- [26] K. Zhou, S.-K. Oh, W. Pedrycz, J. Qiu, Data preprocessing strategy in constructing convolutional neural network classifier based on constrained particle swarm optimization with fuzzy penalty function, *Engineering Applications of Artificial Intelligence* 117 (2023) 105580.
- [27] M. Pal, A. K. Bharati, S. S. Belkhode, Data analytics for smart electrical protection systems, WO2021229487A1 (18.11.2021).
- [28] T. Ahila, A. Subhajini, E-gcs: Detection of covid-19 through classifica-

- tion by attention bottleneck residual network, *Engineering Applications of Artificial Intelligence* 116 (2022) 105398.
- [29] Y. Wang, Z. Li, F. Li, P. Yang, J. Yue, Towards fusing fuzzy discriminative projection and representation learning for image classification, *Engineering Applications of Artificial Intelligence* 114 (2022) 105137.
- [30] Z. Li, Z. Zhang, Z. Fan, J. Wen, An interactively constrained discriminative dictionary learning algorithm for image classification, *Engineering Applications of Artificial Intelligence* 72 (2018) 241–252.
- [31] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, ArXiv abs/1706.03762 (2017).
- [32] R. Castro, I. Pineda, W. Lim, M. E. Morocho-Cayamcela, Deep learning approaches based on transformer architectures for image captioning tasks, *IEEE Access* 10 (2022) 33679–33694.
- [33] Y. Yu, Y. Li, J. Wang, H. Guan, F. Li, S. Xiao, E. Tang, X. Ding, C²-capsvit: Cross-context and cross-scale capsule vision transformers for remote sensing image scene classification, *IEEE Geoscience and Remote Sensing Letters* 19 (2022) 1–5.
- [34] Z. Xue, X. Tan, X. Yu, B. Liu, A. Yu, P. Zhang, Deep hierarchical vision transformer for hyperspectral and lidar data classification, *IEEE Transactions on Image Processing* 31 (2022) 3095–3110.
- [35] R. Castro, I. Pineda, W. Lim, M. E. Morocho-Cayamcela, Deep learning approaches based on transformer architectures for image captioning tasks, *IEEE Access* 10 (2022) 33679–33694.
- [36] B. S. Deo, M. Pal, P. K. Panigrahi, A. Pradhan, Supremacy of attention based convolution neural network in classification of oral cancer using histopathological images, *Preprints from medRxiv and bioRxiv* (2022).
doi:<https://doi.org/10.1101/2022.11.13.22282265>.

- [37] B. S. Deo, M. Pal, P. K. Panigrahi, A. Pradhan, An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification, Preprints from medRxiv and bioRxiv (2022).
[doi:<https://doi.org/10.1101/2022.11.13.22282266>](https://doi.org/10.1101/2022.11.13.22282266).
- [38] I. A. Nikolov, M. Nielsen, J. Garnæs, C. B. Madsen, Wind turbine blade surfaces dataset, 2020.
- [39] A. Kolesnikov, A. Dosovitskiy, D. Weissenborn, G. Heigold, J. Uszkoreit, L. Beyer, M. Minderer, M. Dehghani, N. Houlsby, S. Gelly, T. Unterthiner, X. Zhai, An image is worth 16x16 words: Transformers for image recognition at scale, 2021.
- [40] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, D. Tao, A survey on vision transformer, IEEE Transactions on Pattern Analysis and Machine Intelligence (2022) 1–1.
- [41] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, M. De Vos, Seqsleepnet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging, IEEE Transactions on Neural Systems and Rehabilitation Engineering 27 (3) (2019) 400–410.
- [42] H. Huang, H. Xu, X. Wang, W. Silamu, Maximum f1-score discriminative training criterion for automatic mispronunciation detection, IEEE/ACM Transactions on Audio, Speech, and Language Processing 23 (4) (2015) 787–797.