

Automated Health Assessment of Betel (piper betle L.) Leaves using Correlation-based Image Analysis and Disease Severity Estimation

Sandip Bordoloi¹

¹Department of Electrical Engineering, Girijananda Chowdhury University, Azara, Guwahati, INDIA

¹sandip_ee@gcuniversity.ac.in

Abstract: Early detection of foliar diseases is essential for maintaining the quality and commercial value of betel (*Piper betle* L.) leaves. Conventional visual inspection is subjective, time-consuming, and dependent on expert knowledge. This study presents an automated image-based approach for the classification of betel leaf health and estimation of disease severity using digital image processing techniques implemented in the R programming environment. The proposed method employs the *Pliman* package for image preprocessing, resizing, blue-channel index extraction, and disease quantification. A healthy reference leaf image was used to compute the Pearson correlation coefficient between the blue-channel intensity distributions of test and reference images. Based on predefined similarity thresholds, the leaves were categorized as Healthy, Diseased, Torn/Other, or Unclassified. Disease severity was further estimated using healthy and symptomatic reference images. Experimental evaluation was carried out on eleven betel leaf images comprising healthy and diseased samples. The healthy reference image exhibited a similarity score of 1.0, whereas diseased leaves showed similarity values between approximately 0.70 and 0.82. The estimated symptomatic disease severity ranged from approximately 66% to 97%, demonstrating the capability of the proposed approach to distinguish healthy and infected leaves effectively. The developed method provides a rapid, non-destructive, and computationally efficient tool for preliminary disease assessment and has potential applications in precision agriculture, smart farming, and automated crop monitoring system.

Keywords: Betel leaf, Disease severity estimation, Correlation analysis, Plant phenotyping, RGB image analysis, Precision agriculture.

(Article history: Received: : Received 24th July 2026 and accepted 6th August 2026)

I. INTRODUCTION

Betel (*Piper betle* L.) is an economically and culturally important medicinal plant cultivated extensively throughout South and Southeast Asia. The leaves are widely utilized for medicinal preparations, traditional chewing practices, and pharmaceutical applications due to their antimicrobial, antioxidant, and anti-inflammatory properties [1]. Maintaining healthy foliage is therefore essential for ensuring both crop productivity and leaf quality.

Betel cultivation is frequently affected by fungal, bacterial, and physiological disorders that reduce yield and market value [2]. Early identification of disease symptoms enables timely intervention and minimizes economic losses. Traditionally, disease diagnosis relies on visual examination by experienced farmers or agricultural experts. However, manual inspection is subjective, labour-intensive, and often inconsistent under varying environmental conditions [3].

Recent developments in digital image processing have enabled rapid and non-destructive assessment of plant health. RGB imaging combined with computer vision techniques provides valuable information regarding colour variation, texture, and structural characteristics associated with disease symptoms [4]. Digital phenotyping methods have therefore

gained considerable attention for automated crop monitoring because they require relatively inexpensive imaging equipment and minimal computational resources compared with hyperspectral imaging systems [5].

Several image-based approaches have been reported for plant disease detection using colour indices, texture descriptors, segmentation algorithms, and deep learning models [4]. Although deep learning techniques generally achieve high classification accuracy, they often require large annotated datasets and significant computational resources [6]. In contrast, correlation-based image analysis offers a simpler alternative for small datasets while maintaining satisfactory discrimination between healthy and diseased samples [7].

In this work, an automated image analysis framework is developed for assessing the health condition of betel leaves using RGB images. The proposed approach utilizes the *Pliman* package [8] in R to perform image preprocessing, blue-channel feature extraction, correlation-based similarity measurement, and disease severity estimation. The methodology classifies leaves into Healthy, Diseased, Torn/Other, and Unclassified categories while simultaneously estimating disease severity using healthy and symptomatic reference images [8]. The developed system

aims to provide a rapid, low-cost, and non-destructive solution suitable for precision agriculture and routine field monitoring.

The major contributions of this work are as follows:

1. Development of an automated image-processing framework for betel leaf health assessment.
2. Correlation-based classification using blue-channel similarity with a healthy reference image.
3. Quantitative estimation of disease severity from RGB images.
4. Generation of graphical visualizations for disease severity distribution and similarity analysis suitable for agricultural decision support.

II. THEORETICAL OVERVIEW

Research on automated betel leaf health assessment primarily utilizes deep learning and machine learning to classify diseases [9]. Current methodologies emphasize high-accuracy image processing, while limitations include model overfitting, computational intensity, and the need for diverse, large-scale datasets [10].

Automated health assessment of *Piper betle* L. leaves relies on diverse image processing and classification architectures, ranging from traditional machine learning to advanced deep neural networks [11]. Researchers have frequently employed preprocessing techniques to enhance image quality and manage noise, such as Contrast Limited Adaptive Histogram Equalization (CLAHE), Gaussian Mixture Models (GMM), Adaptive Differential Filtering (ADF), and Principal Component Analysis (PCA) for dimensionality reduction [12, 13].

Feature extraction and classification strategies include:

- Deep Learning models: Architectures such as EfficientNet, VGG16, ResNet50, InceptionV3, and specialized models like E-SwishNet and BetelProNet are commonly deployed [11-15].
- Lightweight architectures: Models using depth-wise separable convolutions, such as BLCNN and MobileNetV3-based frameworks, are utilized to balance high classification accuracy with real-time computational efficiency [14],[17].
- Machine Learning techniques: Traditional approaches, including Extreme Learning Machines (ELM), Support Vector Machines (SVM), Logistic Regression, and Self-Organizing Maps (SOM) using RGB and HSV color space features, GhostNetV2-Based Lightweight Deep learning framework, have been implemented to categorize health status [13],[18],[19], [20].

Recent studies demonstrate significant progress in the precision of automated disease identification, with several

models achieving high accuracy rates across various disease categories.

TABLE 1 RESEARCH FINDINGS FOR BETAL LEAVES DISEASE CLASSIFICATION

MODEL	REPORTED ACCURACY	KEY FINDINGS
EfficientNet B5	98.84%	High accuracy achieved via transfer learning [15]
E-SwishNet	99.34%	Superior precision and recall performance [12]
BetelProNet	98.00%	Efficient inference speed and F1-score [14]
BLCNN	96.02%	Effective balance of speed and accuracy [17]
SOM Method	97.20%	Successful classification using RGB/HSV features [19]

These findings suggest that deep learning, particularly when utilizing transfer learning or custom-built CNNs, provides a robust mechanism for early detection, which is essential for minimizing crop losses and ensuring the quality of medicinal herbal supplies [11,12,15,17].

Despite high reported accuracy, current research faces several technical and practical constraints. A primary challenge is the occurrence of overfitting or underfitting in specific architectures, such as MobileNetV2 and EfficientNetV2B3, which can cause total failure in disease classification [11]. Furthermore, computational intensity remains a concern; while some models provide superior performance, they often require significant training time or resources compared to lightweight alternatives [15,17].

Additional limitations include:

- Data variability: Models must remain adaptable across diverse data sources and varying symptoms that correlate with plant age [14],[20].
- Diagnostic scope: While many studies achieve high precision, they are often limited to a specific subset of diseases, necessitating broader frameworks to handle the wide range of fungal, bacterial, and viral threats facing betel cultivation [11-14].
- Implementation Issues: Transitioning from controlled image datasets to field-based, real-time monitoring requires systems that can maintain high performance under varying environmental conditions [12,17,18].

III. MATERIAL AND METHOD

A. Image Dataset

Digital images of healthy and diseased betel (*Piper betle* L.) leaves were collected under controlled lighting conditions using a high-resolution Mobile Camera (50 MP). The dataset consisted of eleven leaf images representing varying degrees of disease severity. One healthy leaf image served as the reference sample, while one symptomatic leaf image was used for disease severity estimation. The remaining images were analysed individually by the proposed algorithm.

Raw sample images from the Image dataset are shown in Figure 1.

B. Image Pre-processing

The collected image dataset is then pre-processed. Raw field images typically suffer from cluttered backgrounds, varying illumination, and noise. Preprocessing techniques utilize noise filtering, contrast enhancement, and background segmentation to isolate the leaf region. The background of the images was converted to blue color in order to implement the *Pliman* tool [8] in the image dataset. Each RGB image was resized to a resolution of 300×300 pixels to ensure uniform dimensionality before feature extraction. Standardizing image dimensions minimized computational complexity and enabled direct pixel-wise comparison between the reference and test images.

C. Software Environment

The complete image analysis was implemented in the R programming language using the following libraries:

- *pliman* [8] for digital plant phenotyping
- *ggplot2* for visualization
- *dplyr* for data manipulation
- *gridExtra* for figure generation

The *Pliman* (Plant Image Analysis) package in R is an open-source tool designed to automate and streamline the analysis of plant images. It provides functions for batch image manipulation, measuring leaf area, assessing disease severity, counting objects, and analyzing field phenotyping imagery [8].

The workflow was executed on a personal computer using the Rstudio development environment. The workflow is shown in Figure 2.

D. Blue-Channel Feature Extraction

The blue-channel intensity component was extracted from every resized RGB image using the image indexing function provided by the *pliman* package. Let $I_B(x, y)$ represent the blue-channel intensity at pixel location (x, y) .



Figure 1: Raw Sample Images from Image Dataset

E. Correlation-Based Similarity Analysis

The similarity between the healthy reference image and each test image was determined using the Pearson correlation coefficient,

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where

- x_i denotes the blue-channel intensity of the healthy reference image,
- y_i denotes the blue-channel intensity of the test image.

Higher correlation values indicate greater similarity to the healthy leaf.

F. Leaf Classification

Leaf health status was assigned according to the similarity coefficient obtained from correlation analysis.

- Healthy: Similarity > 0.90
- Diseased: $0.60 < \text{Similarity} \leq 0.90$
- Torn/Other: Similarity ≤ 0.60
- Unclassified: Similarity unavailable

This threshold-based decision strategy provides an efficient means of distinguishing healthy and infected leaves.

G. Disease Severity Estimation

Disease severity was estimated using the *pliman* package [8] by comparing each image with healthy and symptomatic reference leaves.

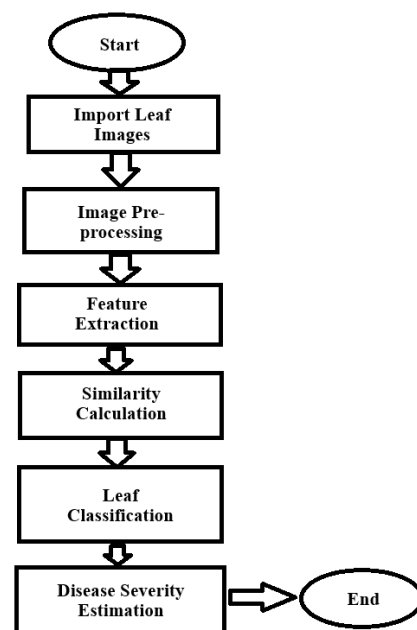


Figure 2: Workflow

The disease severity was calculated as

$$\text{Disease Severity (\%)} = \frac{\text{Symtomatic Area}}{\text{Total Leaf Area}} \times 100 \quad (2)$$

The algorithm reports both healthy and symptomatic percentages, allowing quantitative assessment of disease progression.

H. Data Visualization

Three graphical representations were generated to analyse the results:

- Disease severity across all betel leaf samples.
- Relationship between similarity score and disease severity.
- Disease severity distribution for different health categories.

The plots facilitate interpretation of the classification performance and disease progression

IV. RESULTS AND DISCUSSION

The proposed image-processing framework was evaluated using eleven RGB images of betel (Piper betle L.) leaves consisting of healthy and diseased samples. The analysis involved blue-channel feature extraction, correlation-based similarity measurement, health status classification, and disease severity estimation. The developed workflow automatically generated quantitative measurements together with graphical visualizations for interpretation.

A. Leaf Classification Based on Similarity Analysis

The Pearson correlation coefficient between the blue-channel intensity distributions of the healthy reference leaf and each test image was used as the primary feature for health classification. The healthy reference image achieved a similarity score of 1.000, indicating complete agreement with itself. Diseased leaves exhibited similarity values ranging from approximately 0.696 to 0.816, whereas several images were assigned to the Unclassified category because similarity values could not be computed successfully during feature extraction.

The selected threshold values effectively separated healthy and diseased samples. The healthy leaf was correctly identified, while leaves exhibiting visible disease symptoms were classified as diseased. A sample classification is shown in Figure 3

This demonstrates that the blue-channel intensity distribution is sufficiently sensitive to detect changes in leaf pigmentation caused by disease progression. Unlike conventional visual inspection, the proposed automated

method provides objective and repeatable measurements independent of operator experience.



Figure 3: Comparison of a diseased leaf with a healthy (reference) leaf

B. Disease Severity Estimation

The disease severity was showed considerable variation among the analysed samples. The symptomatic area ranged from approximately 66.7% for the healthy reference image to values exceeding 96% for severely infected leaves. Several diseased samples exhibited severity values above 90%, indicating extensive symptomatic regions, while moderately infected leaves showed disease severity between 74% and 81%.

The severity measurements correspond well with the visual appearance of the leaf images. Leaves having greater discoloration and damaged tissue produced larger symptomatic percentages, whereas comparatively healthier leaves exhibited lower disease severity. These findings demonstrate that the implemented digital phenotyping approach is capable of quantifying disease progression rather than merely identifying the presence of disease.

C. Relationship Between Similarity and Disease Severity

The relationship between similarity score and disease severity is illustrated in Figure 2. An overall inverse relationship can be observed between the two parameters. As disease severity increases, the similarity with the healthy reference leaf generally decreases.

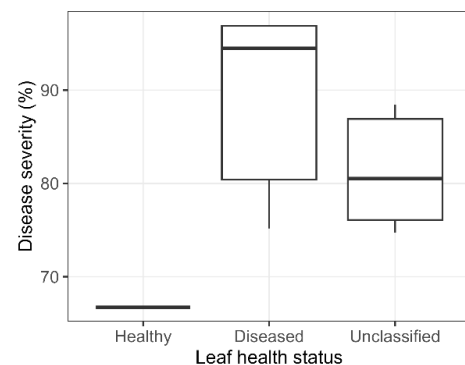


Figure 4: Box plot of the disease severity of the leaf samples.

TABLE 2 LEAF SEVERITY CLASSIFICATION

FILENAME	STATUS	SIMILARITY	SEVERITY_HEALTHY	SEVERITY_SYMPTOMATIC	SEVERITY
Leaf 1	Healthy	1	33.287789	66.71221	66.71221
Leaf 10	Unclassified	NA	13.08557	86.91443	86.91443
Leaf 11	Unclassified	NA	23.920259	76.07974	76.07974
Leaf 2	Diseased	0.6961703	3.027547	96.97245	96.97245
Leaf 3	Diseased	0.7278758	5.51738	94.48262	94.48262
Leaf 4	Diseased	0.8002329	19.590905	80.4091	80.4091
Leaf 5	Diseased	0.7589317	3.105279	96.89472	96.89472
Leaf 6	Diseased	0.8163685	24.821797	75.1782	75.1782
Leaf 7	Unclassified	NA	19.483841	80.51616	80.51616
Leaf 8	Unclassified	NA	25.286822	74.71318	74.71318
Leaf 9	Unclassified	NA	11.557857	88.44214	88.44214

Leaves exhibiting similarity values close to 0.80 showed comparatively lower disease severity than those with similarity values around 0.70. This behaviour is expected because disease symptoms alter the colour distribution and pixel intensity of the leaf surface, thereby reducing the correlation with the healthy reference image.

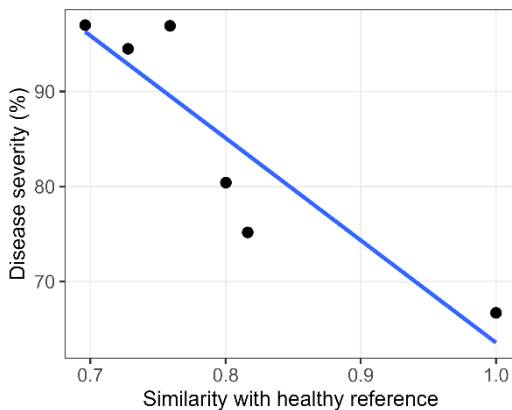


Figure 5: Disease severity vs. Leaf similarity graph

Although only a limited number of samples possessed valid similarity values, the observed trend indicates that correlation-based similarity can serve as a useful indicator of disease progression. Increasing the dataset size would further improve the statistical reliability of this relationship and facilitate predictive modelling of disease severity.

D. Disease Severity Distribution Among Health Categories

Figure 5 presents the distribution of disease severity for different health status categories. Diseased leaves exhibit consistently higher symptomatic percentages compared with the healthy sample. The boxplot further illustrates the variability of disease severity within the diseased class, reflecting differences in infection intensity among individual leaves.

The unclassified images were primarily associated with unsuccessful similarity computation rather than incorrect disease detection. This suggests that future improvements should focus on enhancing image preprocessing and feature extraction to reduce missing similarity values. Incorporating additional colour indices or texture descriptors could further improve classification robustness under varying lighting conditions.

V. CONCLUSION

The proposed workflow successfully integrates digital image preprocessing, colour-based feature extraction, correlation analysis, disease classification, and severity estimation into a single automated framework. The method requires only RGB images and does not depend on expensive imaging equipment or computationally intensive deep learning models.

The principal advantages of the proposed approach include:

- Fully automated image analysis using open-source R software.
- Objective disease severity estimation from RGB images.
- Simple correlation-based classification requiring minimal computational resources.
- High suitability for rapid screening of betel leaves in laboratory or field environments.

However, several limitations were observed. The dataset consisted of few images, limiting the statistical generalization of the results. Additionally, some images produced missing similarity values, indicating the need

For improved preprocessing or alternative feature extraction methods. Future work will focus on expanding the image database, incorporating multiple colour indices

and texture features, and integrating machine learning algorithms for enhanced classification accuracy.

REFERENCES

- [1] P. Biswas et al., "Betelvine (Piper betleL.): A comprehensive insight into its ethnopharmacology, phytochemistry, and pharmacological, biomedical and therapeutic attributes," *Journal of Cellular and Molecular Medicine*, vol. 26, no. 11, pp. 3083–3119, May 2022, doi: 10.1111/jcmm.17323. Available: <https://doi.org/10.1111/jcmm.17323>
- [2] A. Kumar et al., "Challenges and economic feasibility of betelvine cultivation in Uttar Pradesh," *International Journal of Economic Plants*, vol. 10, no. May, 2, pp. 160–169, May 2023, doi: 10.23910/2/2023.0513c. Available: <https://doi.org/10.23910/2/2023.0513c>
- [3] S. Ghimire, R. Lamsal, A. Sankuratri, P. Lakkarsu, and S. Vutla, "Artificial Intelligence-Driven Plant Disease Detection and Diagnosis: A comprehensive review of deep learning approaches, multimodal sensing technologies, and future perspectives in precision agriculture," *The Plant Pathology Journal*, vol. 42, no. 2, pp. 121–137, Mar. 2026, doi: 10.5423/ppj.rw.01.2026.0004. Available: <https://doi.org/10.5423/ppj.rw.01.2026.0004>
- [4] M. Shafay et al., "Recent advances in plant disease detection: challenges and opportunities," *Plant Methods*, vol. 21, no. 1, p. 140, Oct. 2025, doi: 10.1186/s13007-025-01450-0. Available: <https://doi.org/10.1186/s13007-025-01450-0>
- [5] J. L. Araus, J. Jauregui-Besó, J. Segarra, and F. Junliang, "Affordable crop phenotyping matters," *Crop and Environment*, p. 100145, May 2026, doi: 10.1016/j.crope.2026.100145. Available: <https://doi.org/10.1016/j.crope.2026.100145>
- [6] J. Li, P. Jiang, Q. An, G.-G. Wang, and H.-F. Kong, "Medical image identification methods: A review," *Computers in Biology and Medicine*, vol. 169, p. 107777, Dec. 2023, doi: 10.1016/j.compbiomed.2023.107777. Available: <https://doi.org/10.1016/j.compbiomed.2023.107777>
- [7] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5, p. 91, Apr. 2023, doi: 10.3390/computers12050091. Available: <https://doi.org/10.3390/computers12050091>
- [8] A. Upadhyay et al., "Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture," *Artificial Intelligence Review*, vol. 58, no. 3, Jan. 2025, doi: 10.1007/s10462-024-11100-x. Available: <https://doi.org/10.1007/s10462-024-11100-x>
- [9] T. Olivoto, "Lights, camera, pliman! An R package for plant image analysis," *Methods in Ecology and Evolution*, vol. 13, no. 4, pp. 789–798, Jan. 2022, doi: 10.1111/2041-210x.13803. Available: <https://doi.org/10.1111/2041-210x.13803>
- [10] S. M. Hassan et al., "A survey on different plant diseases detection using machine learning techniques," *Electronics*, vol. 11, no. 17, p. 2641, Aug. 2022, doi: 10.3390/electronics11172641. Available: <https://doi.org/10.3390/electronics11172641>
- [11] M. Gayakwad et al., "Applying the transfer learning models on the dataset on the effect of diseases on Nagvel-betel (Piper betle) leaves," *Data in Brief*, vol. 62, p. 111987, Aug. 2025, doi: 10.1016/j.dib.2025.111987. Available: <https://doi.org/10.1016/j.dib.2025.111987>
- [12] P. B and K. K, "An Integrated Framework of ADF and PCA for Betel Leaf Disease Classification Using E-Swish Net. 2025, pp. 1–6, doi: 10.1109/ic3it66137.2025.11341484. Available: <https://doi.org/10.1109/ic3it66137.2025.11341484>
- [13] M. A. Malek, S. Basak, and S. S. Reya, "An Approach to Identify Diseases in Betel Leaf Using Deep Learning Techniques. 2022, pp. 1–5, doi: 10.1109/sti56238.2022.10103348. Available: <https://doi.org/10.1109/sti56238.2022.10103348>
- [14] B. Bhaskar, S. Kusuma, K. A. Prasad, A. Naresh, and C. V. Subbaiah, "BetelProNet Framework: efficient Deep Learning model for Betel Leaf Disease Detection," *International Research Journal of Multidisciplinary Scope*, vol. 06, no. 04, pp. 856–870, Jan. 2025, doi: 10.47857/irjms.2025.v06i04.05947. Available: <https://doi.org/10.47857/irjms.2025.v06i04.05947>
- [15] R. H. Hridoy, Md. T. Habib, Md. S. Rahman, and M. S. Uddin, "Deep Neural Networks-Based Recognition of Betel Plant Diseases by Leaf Image Classification," *Lecture Notes on Data Engineering and Communications Technologies*, pp. 227–241, Jan. 2022, doi: 10.1007/978-981-16-9605-3_16. Available: https://doi.org/10.1007/978-981-16-9605-3_16
- [16] R. T. Wahyuningrum, I. H. Ayani, A. Bauravindah, I. A. Siradjuddin, and I. S. Faradisa, "Classification of betel leaf diseases based on convolutional neural network to increase production herbal spice materials," *Register Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 11, no. 1, pp. 13–28, May 2025, doi: 10.26594/register.v11i1.4653. Available: <https://doi.org/10.26594/register.v11i1.4653>
- [17] R. H. Hridoy, T. Habib, I. Jabiullah, R. Rahman, and F. Ahmed, "Early Recognition of Betel Leaf Disease using Deep Learning with Depth-wise Separable Convolutions," *2021 IEEE Region 10 Symposium (TENSYP)*, pp. 1–7, Aug. 2021, doi: 10.1109/tensymp52854.2021.9551009. Available: <https://doi.org/10.1109/tensymp52854.2021.9551009>
- [18] F. David and M. A. Mukunthan, "Betel Leaf Diseases Classification using Machine Learning Algorithm: A Feasible Approach," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 40, no. 1, pp. 74–86, Feb. 2024, doi: 10.37934/araset.40.1.7486. Available: <https://doi.org/10.37934/araset.40.1.7486>
- [19] D. I. Mulyana, A. Saepudin, and M. B. Yel, "Health detection of betal leaves using Self-Organizing Map and Thresholding Algorithm," *Journal of Applied Engineering and Technological Science (JAETS)*, vol. 4, no. 1, pp. 180–189, Sep. 2022, doi: 10.37385/jaets.v4i1.957. Available: <http://dx.doi.org/10.37385/jaets.v4i1.957>
- [20] "GhostNetV2-Based Lightweight Deep learning framework for automated Betel Leaf disease classification," *IEEE Conference Publication IEEE Xplore*, May 06, 2026. Available: <https://ieeexplore.ieee.org/abstract/document/11549057/>

AUTHOR PROFILE



Sandip Bordoloi is working as an Associate Professor at Girijananda Chowdhury University, Assam India. He received his PhD degree in Electrical and Electronics Engineering from Assam Don Bosco University (India) in 2015. His research interest is in sensor and instrumentation, energy analytics, data analytics and machine learning.