International Journal of Applied Research in Bioinformatics
Volume 13 • Issue 1 • January-December 2024

Big Data Visualization for Black Sigatoka Disease of Bananas and Pathogen–Host Interactions (PHI) of Other Plants

Richard S. Segall
https://orcid.org/0000-0001-7627-2609
Arkansas State University, USA

Soichiro Takahashi
https://orcid.org/0009-0002-7309-2797
Arkanas State University, USA

Prasanna Rajbhandari
https://orcid.org/0009-0009-3394-2137
Arkansas State University, USA

ABSTRACT

Black sigatoka is a leaf spot disease affecting banana plants that has caused significant yield reductions of up to 50% (Arman et al., 2023). This research presents data visualizations of 8,761 data points related to black sigatoka in banana plants, encompassing attributes such as time, canopy temperature, and relative humidity. The paper also reviews related work, including the application of data mining to plant studies, the use of deep learning neural networks for plant disease data, and the use of machine learning for predicting crop yields and detecting disease. Additionally, it presents big data visualizations for 20,952 values obtained from the web-accessible Pathogen–Host Interactions database (PHI-base), covering various plant disease categories and providing an epidemiological analysis of prevalent causative agents in specific plant diseases.

KEYWORDS

Big Data, Black Sigatoka (Banana), Canopy Temperature (CT) Data Visualization, Pathogen–Host Interaction (PHI), Plant Pathogens, Relative Humidity (RH)

INTRODUCTION

Many researchers have investigated applications of data mining, neural networks for deep leaning, and machine learning (ML) for detection of plant pathogens. However, this paper presents the applications of techniques for big data visualization to selected the significant leaf spot disease of black sigatoka for banana and also interactions between pathogens and host plants. Some related work is discussed below and, in the tables, presented in the next section.

Big Data for Plant Omics

Ohyanagi et al. (2023) edited a book on plant omics focusing on advances in big data biology. The book discusses deep learning approaches in plant omics research, including deep learning on plant images and genetic sequences for classifications and regressions, object detection, and image

DOI: 10.4018/IJARB.361940

segmentation. It also provides an online resource guide for plant omics databases. Zheng and Terauchi (2023) explored plant–pathogen interactions from the new perspective of "omics."

Machine Learning (ML) for Plant Disease Detection

Ghosh et al. (2021) provided a survey on plant disease detection using ML approaches. Mehta et al. (2023) researched artificial intelligence (AI)-based plant disease detection and classification using pretrained models. Tripathi et al. (2021) presented a survey of deep learning methods for plant disease detection. Verma et al. (2019) created a deep learning-based mobile application for plant disease diagnosis with a case study on tomato plants. Additionally, Deka (2022) presented a study on plant DNA barcoding.

Bowers et al. (2022) conducted an in-depth study on surveillance for the early detection of high-consequence pests and pathogens, including plant pathogens, featuring a case study on soybean rust. Yakimovich (2021) used ML and AI for the prediction of host–pathogen interactions (PHI). Sperschneider (2020) used ML in plant–pathogen interactions for empowering biological predictions.

Deep Learning for Plant Disease Detection

Sangeeta et al. (2023) investigated plant leaf disease detection using deep learning. Kour et al. (2023) studied the implementation of ML techniques such as artificial neural networks, linear discriminant analysis, and logistic regression for wheat yield prediction. Kaundal et al. (2022) created a comprehensive deep learning platform named deepHPI for accurate prediction and visualization of host–pathogen protein–protein interactions.

Previous Related Work by Authors

Previous work on data mining related to the environmental stress tolerances of plants has been presented by Segall et al. (2004, 2008), Segall and Nonis (2004), and Segall (2005a, 2005b). In addition, Segall (2006) and Segall and Zhang (2006) have presented studies on microarray databases in bioinformatics. The authors Segall and Rajbhandari (2024) discussed the applications of image processing for plant diseases of four different plant categories.

BACKGROUND AND RELATED WORK

What Is Data Visualization?

Data visualization involves presenting sets of primarily quantitative raw data in a schematic form. It is an essential strategy for many business organizations. Common tools for data visualization include charts (e.g., bar, line, and pie), plots (e.g., scatter, bubble, and box), and maps (e.g., dot distribution maps), some of which are used in the research presented in this paper. Tableau and Power BI by Microsoft are two of the most popular data visualization tools.

Related Work

A comparative study of performances of plant disease detection and classification techniques for big data was presented by Demilie (2024).

Hasan et al. (2022) provided a comprehensive review of plant leaf disease detection using image processing, discussing the algorithms for image processing and AI. Other investigators who used image processing for detection of plant diseases include Bharate and Shirdhonker (2017).

Nagila and Mishra (2023) and Kumar (2022) discussed the effectiveness of ML and image processing in detecting leaf disease. Sahu et al. (2021) discussed challenges and issues in plant disease detection using deep learning. Aldakheel et al. (2024) studied the detection and identification of plant leaf disease using an algorithm named YOLOv4 (meaning, "you only look once") that compared its performance with established target identification methods.

Related Content

Pharmaco-EcoMicrobiology and Its Potential Role in Medical and Environmental Sciences
Mohd. Shahid, Hridesh Mishra, Hemant Kumar Mishra, Trivendra Tripathi, Haris M. Khan, Farrukh Sobiaand Anuradha Singh (2012). *Pharmacoinformatics and Drug Discovery Technologies: Theories and Applications (pp. 291-302).*
www.irma-international.org/chapter/pharmaco-ecomicrobiology-its-potential-role/64079

Efficient Computational Construction of Weighted Protein-Protein Interaction Networks Using Adaptive Filtering Techniques Combined with Natural Selection-Based Heuristic Algorithms
Christos M. Dimitrakopoulos, Konstantinos A. Theofilatos, Efstratios F. Georgopoulos, Spiridon Likothanassis, Athanasios Tsakalidisand Seferina P. Mavroudi (2012). *International Journal of Systems Biology and Biomedical Technologies (pp. 20-34).*
www.irma-international.org/article/efficient-computational-construction-weighted-protein/67104

Introduction to Classification Error Estimation
(2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations (pp. 334-340).*
www.irma-international.org/chapter/introduction-classification-error-estimation/53912

Application of Genetic Algorithm in Denoising MRI Images Clouded with Rician Noise
Debajyoti Misra, Ankur Gangulyand Dewaki Nandan Tibarewala (2016). *Biomedical Image Analysis and Mining Techniques for Improved Health Outcomes (pp. 14-38).*
www.irma-international.org/chapter/application-of-genetic-algorithm-in-denoising-mri-images-clouded-with-rician-noise/140483

The Benefits of Big Data Analytics in the Healthcare Sector: What Are They and Who Benefits?

Andrea Darrel, Margee Hume, Timothy Hardieand Jeffery Soar (2015). *Big Data Analytics in Bioinformatics and Healthcare (pp. 406-439).*

www.irma-international.org/chapter/the-benefits-of-big-data-analytics-in-the-healthcare-sector/121469