

REVIEW

Agricultural bioinformatics and machine learning techniques in Areca nut disease and crop improvement – a review

G. Archana^{1*} and N. Hemalatha²¹Department of Bioinformatics, St Aloysius Institute of Management and IT, Mangalore, Karnataka²St Aloysius Institute of Management and IT, Mangalore, Karnataka***Correspondence:**G. Archana,
221803archana@staloyusius.ac.in**Received:** 05 April 2024; **Accepted:** 08 August 2024; **Published:** 24 October 2024

Bioinformatics is an interdisciplinary scientific field comprising biology, mathematics, and computer science. It is the application of information technology to the management of biological data that aids in deciphering plant genomes. Biological research, which previously began in laboratories, fields, and botanical clinics, now begins at the computational level, using computers (*in silico*) for data analysis, experimental design, and hypothesis development. Agriculture is the backbone of the nation, so agricultural bioinformatics is one of the fastest-growing scientific fields that use computational approaches to study biology and life sciences. India ranks first in the world for areca nut production and many farmers depend on areca nut cultivation for their livelihood. Areca nut yields are affected by many diseases caused by heavy rainfall and high relative humidity. Early prediction of crop diseases based on weather data helps farmers take preventive measures. Many machine learning methods are used to detect a disease from image data. Expanding knowledge about the molecules and mechanisms associated with specific phenotypic traits and specific responses to biotic or abiotic stresses will be complemented by the predictive power of bioinformatics to influence agricultural practices and improve diagnostic, monitoring, and advocating innovative methods in traceability, enhancing human value and supporting sustainability at low cost. This review briefs about the field of agricultural bioinformatics and the application of machine-learning techniques in the overall crop improvement of areca nut including disease prediction

Keywords: bioinformatics, agriculture, agricultural bioinformatics, areca nut, machine learning

1. Introduction

Bioinformatics is one of the fastest emerging fields of science, being a combination of biology, computer science, and mathematics and a useful area for solving complex biological problems. It can manage to store data in massive amounts, with the help of sophisticated software tools.

The agricultural sector is the most important area of interest where bioinformatics and machine learning techniques are yet to get implemented as in other fields. One can witness many advances in the sector in recent decades such as the production of high-yielding seed varieties, improvements in irrigation, and the introduction of highly efficient fertilizers. But as in all other sectors,

even farmers face several challenges, such as crop disease management, poor infrastructure, and low yields due to climatic instabilities (1).

India depends on its food crops for survival and also on commercial crops for the economy. One such profitable crop is the areca nut grown in most parts of south India and Malenadu. As a seasonal crop, it can produce a high yield on proper cultivation techniques. But like other plants, areca nut also is vulnerable to different diseases such as fruit rot, foot rot, bud rot, yellow leaf disease, leaf dot disease, etc.

Agricultural research has made great strides in recent years with breakthroughs in genomics, precision agriculture and permaculture practices. Integrating these decisions into agriculture promises long-term solutions.

The main objective of the paper is to analyze the field of bioinformatics in agriculture and to analyze the areca nut diseases and types. This article is dedicated to in-depth research on the integration of knowledge and agriculture, focusing on the modern analysis of areca nut diseases. Our mission includes research on the latest developments in the use of bioinformatics in agriculture. Also, the review focuses on the methodologies and technical models created through machine learning and bioinformatics which can either predict the occurrence of diseases to reduce major crop loss or help maintain the cultivation methods according to the environmental conditions for better farming and crop yield.

2. Literature review

2.1. Bioinformatics and agriculture

Genomic analysis of plants and animals produces a huge wealth of data. The tools in bioinformatics can be used to search for the genes within those genomes that are useful for agricultural communities and to decode their functions. This genetic knowledge can be used for the production of stronger and more drought- and disease-resistant crop varieties (2).

The tools can be used to detect metallic contents from metagenomic sequencing of contaminated soil and thereby plays a role in improving soil nutrition (3).

In-silico genomic investigations could be used to find disease-resistant genes and enzymes, as well as their promoters and transcription factors, to improve a plant's defense mechanism (4, 5).

Plants have their defense such as various pre-formed anti-microbial compounds and some other compounds which get activated by pathogen recognition so that they could defend themselves from pathogens and ultimately survive (6).

Thus, agricultural bioinformatics emerges as a very diverse field of study managing a large amount of data in terms of gene sequencing, gene marker, gene mining, etc. With high-end computations, dry lab facilities, and high-speed internet for transferring bulk genomic data, the field is in the phase of expansion. Thus, bioinformatics has high potential to transform the agricultural sector by empowering the farmer through Information and Communication Technologies (2).

2.2. Areca nut cultivation and its importance in India

India ranks first in the world with a 50% of areca nut production. Many farmers depend on it for their living. The Southeast countries such as India, China, and Malaysia are the main producers of areca nut. In India, it is widely grown in states such as Karnataka, Kerala, Goa, Meghalaya, Assam, and West Bengal. It is available in dried, cured, and fresh

forms. A land area of 518,000 hectares yields up to 853,000 metric tons of areca nut in India annually (7).

The areca nut is consumed for mastication with betel leaves and used in religious rites (8). Nowadays the areca nut is used not only as a commercial crop but also as a side ingredient for many innovative food items such as tea, coffee, sweets, etc. Its extract is also used as a coloring agent for cloth. Moreover, research is going on since science has got some shreds of evidence regarding the anti-cancer properties of areca nut.

2.3. Areca nut diseases

All plants are vulnerable to diseases that may be caused due to insect pests or parasites. The vulnerability of the areca nut plant toward disease is a serious issue among farmers. Diseases such as fruit rot, foot rot, yellow leaf disease, bud rot, etc. are some commonly encountered diseases.

A brief account of diseases acquired by areca nut is as follows:

Fruit rot and bud rot disease: It is a disease caused due to a fungus named *Phytophthora* spp. The disease is characterized by rotting and heavy shedding of immature nuts and buds. It develops during the monsoon season and can occur due to a lack of fertilizers (7) (Patil Balanagouda et al.).

Yellow leaf disease: A disease that causes yellow coloration on leaves. It occurs due to the loss of chlorophyll caused by pathogens. As a result, the leaves become smaller, yellow, stiff, wither, and fall off (9) (Ni, X.; Quisenberry et al.).

The diseases can also be developed through climatic conditions and improper cultivation habits.

There can be a development of fruit rot disease due to heavy rain as well as a lack of chemical fertilizers.

Drastic temperature increases can lead to abnormalities in areca nut plants.

Excessive use of herbicides and fertilizers can damage the soil contents and areca roots. (Diaz-Perez et al.).

Inefficient irrigation methods such as drought damage caused by non-irrigation or waterlogging caused by excessive irrigation can cause yellow leaf disease. (Hohender., Xiao et al.).

2.4. Machine learning in agriculture

The increase in population led to an increase in the demand for digital agriculture. Agricultural tasks were found to be improved through advances in machine learning.

The forecasting of crop disease was able to be carried out by means of models using algorithms of machine learning such as Support Vector Machine (SVM), Decision tree algorithm, artificial neural networks, etc.

The steps involved in model creation:

The steps involve the preparation and validation of datasets related to areca nut disease and weather.

The data can be collected from sources such as

- (1) Weather data from weather stations
- (2) Disease data from farmers
- (3) Data from the literature

- The collected data are integrated using a rule-based method.
- The dataset is subjected to pre-processing.
- Later, the dataset is validated using machine learning techniques.

Rajashree Krishna et al. worked on an Areca nut disease (fruit rot) dataset along with weather data of a region in Dakshina Kannada, Udupi, and conducted machine learning studies on it. The prepared dataset from different sources such as weather stations, farmers' interviews, and previous research was subjected to pre-processing by using daily-based weather data as time series data. Data validation was carried out by considering weather as an independent variable and disease score as the dependent variable. They used Multiple Regression, Support Vector Regression, DTR, MLP, RFR, etc., for disease incidence prediction, before and after the feature selection. They concluded that RFR is a more accurate and efficient method for dataset validation and also that the fewer the features, the lower the chances for errors (9).

Vinod et al. worked on building a model for the prediction of areca nut disease based on weather parameters such as temperature, humidity, rainfall, and wind flow and soil moisture. They used the Support Vector Machine Regression (SVMR) algorithm and Random Forest Classifier (RFC) algorithm to perform prediction. A sensor namely DHT-11 was used for collecting environmental data. They collected the required data from the region of Vittal, Dakshina Kannada, for one year for fruit rot disease prediction. The sensors of the fields were used for the collection of data such as temperature, humidity, and rainfall. The raw data were processed and converted into a tabular format for producing an input dataset for the model. The data were classified based on the disease conditions to extract the actual conditions. A regression algorithm was used for comparing the data values for finding similarities in the pattern of conditions. A value greater than 8 on a scale of 10 indicated the possibilities of disease occurrence. The efficient evaluation of the model was done through manual confirmation (10).

Advantages of the Random Forest algorithm (RFA)

- Random forests appear to provide high accuracy in classification and propagation. It creates multiple decision trees and combines their predictions, reducing the risk of overfitting and increasing overall model accuracy.

- It is a versatile algorithm because of its numerical and categorical features and therefore can be applied to a variety of data types.
- The holistic nature of random forests helps reduce overfitting, a common problem in decision trees. The effects of noise and exercise can be reduced by averaging or polling the estimates of many trees (11).

3. Conclusion

Bioinformatics handles more data and information than many other theoretical disciplines and involves integration of knowledge gained in various fields and subfields such as biology, information science, physics, chemistry, pharmacy, electronics, statistics, agronomy, and medicine (1, 12).

The field is dynamic in nature, therefore it evolves fast and continuously to fit and adapt to the experimental design. In the future, the field of bioinformatics and nanobiotechnology will find technological convergence in agricultural innovation, which highlights the importance of bioinformatics in identifying and finding strategies to combat bacterial and fungal pathogens that attack different crops.

In the realm of agriculture, with a primary focus on the areca nut in this review, the application of machine learning techniques stands out as a beacon for efficient prediction of abnormal conditions in plant crops. Notably, there exists a gap in the current research landscape, with limited studies and models specifically tailored to areca nut cultivation. Given the regional confinement of areca nut cultivation, it might have escaped the broader research spotlight. However, the emergence of new diseases necessitates immediate attention to prevent economic downturns (13, 14).

Looking forward, the prospects for agricultural bioinformatics, coupled with machine learning, are monumental. These technologies not only serve as potent tools for data storage but also hold the key to creating predictive models that can revolutionize agricultural practices. The imperative to address emerging challenges in crop health underscores the significance of ongoing and future research in this field. As we stand on the cusp of innovation, the synergy between bioinformatics, machine learning, and agriculture promises a transformative journey toward sustainable and resilient farming practices.

Since the cultivation is confined to certain regions of the country, it can be assumed to be of not much focus. But while considering new diseases arising to the crop, it becomes the demand of the day to work for resolving the problem, which in turn prevents the economic fall. In future, there is enormous scope for agricultural bioinformatics along with machine learning as it can serve as a potential means to store data, create models, and predict the conditions (15–17).

Author contributions

Both authors contributed equally to this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Cagliarini A, Rush A. *Economic development and agriculture in India. RBA Bulletin*. Sydney, NSW: Reserve Bank of Australia (2011).
2. Singh VK, Singh AK, Chand R, Kushwaha C. Role of bioinformatics in agriculture and sustainable development. *Int J Bioinforma Res*. (2011) 3:221–6.
3. Schloss PD, Handelsman J. Status of the microbial census. *Microbiol Mol Biol Rev*. (2004) 68:686–91.
4. Kummerfeld SK, Teichmann SA. DBD: a transcription factor prediction database. *Nucleic Acids Res*. (2006) 34:D74–D81.
5. Pandey SP, Somssich IE. The Role of WRKY Transcription Factors in Plant Immunity. *Plant Physiol*. (2009) 150:1648–55.
6. Matsumura H, Reich S, Ito A, Saitoh H, Kamoun S, Winter P, et al. Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *PNAS*. (2003) 100:15718–23.
7. Balanagouda P, Vinayaka H, Maheswarappa HP, Narayanaswamy H. Phytophthora diseases of arecanut in India: Prior findings, present status and future prospects. *Indian Phytopathol*. (2021) 74:561–72.
8. Krishna R, Prema KV, Gaonkar R. Areca nut disease dataset creation and validation using machine learning techniques based on weather parameters. *Eng Sci*. (2022) 19:205–14.
9. Lei S, Luo J, Tao X, Qiu Z. Remote sensing detecting of yellow leaf disease of Arecanut based on UAV multisource sensors. *Remote Sens*. (2021) 13:4562.
10. Kanan LV. Arecanut Yield Disease Forecast using IoT and Machine Learning. *Int J Sci Res Eng Technol*. (2021) 2:11–5.
11. Geetha V, Punitha A, Abarna M, Akshaya M, Illakiya S, Janani A. An effective crop prediction using random forest algorithm. *Proceedings of the 2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*. Piscataway, NJ (2020).
12. Kumar D. Development of Agricultural Bioinformatics in India: Issues and Challenges. *Asian Biotechnol Dev Rev*. (2018) 20:3.
13. Orozco A, Morera J, Jiménez S, Boza R. A review of bioinformatics training applied to research in molecular medicine, agriculture and biodiversity in Costa Rica and Central America. *Brief Bioinform*. (2013) 14:661–70.
14. Iquebal MA, Jaiswal S, Mukhopadhyay CS, Sarkar C, Rai A, Kumar D. Applications of bioinformatics in plant and agriculture. *Plant Omics*. (2015) 2015:755–89.
15. Brass A. Bioinformatics education—a UK perspective. *Bioinformatics*. (2000) 16:77–8.
16. Altman RB. A curriculum for bioinformatics: the time is ripe. *Bioinformatics*. (1998) 14:549–50.
17. Counsell D. A review of bioinformatics education in the UK. *Brief Bioinform*. (2003) 4:7–21.