Hybrid Recommender System for Precision Chemical Application in Banana Cultivation Using Matrix Factorization and Content-Based Filtering

Ravi Kumar Tirandasu, Prasanth Yalla, Pachipala Yellamma
Department of CSE, KoneruLakshmaiah Education Foundation, Vaddeswaram, AP, India

Abstract*—Proper management of pesticides and fertilizers is critical towards effective control of the banana diseases, but integration of various agricultural data has been a problem. The novelty of this study is the hybrid recommendation system which encompasses Content-Based Filtering (CBF) with Matrix Factorization (MF) to be used when recommending chemical treatment of bananas during cultivation. The system exploits the use of heterogenous data- such as soil nutrient profiles (NPK, pH), climatic variables, and disease signatures to create customized chemical recommendation to manage the disease. A real-world agricultural dataset was used in the evaluation of the hybrid approach and the improvement, precision, recall, F1-score, and the accuracy of the system were measured. The findings indicate that the suggested model performed better than the traditional models of single-method or user-based recommendation systems and predicted the disease outbreak with high accuracy (F1-score) up to 98 percent in Black Sigatoka; these results were highly consistent across other disease classes and different chemical interventions. Notably, the hybrid system helps not only to optimize the costs of chemical use and crop yields, but also to create the environmental sustainability by reducing the number of the superfluous chemical use. Methodology, the characteristics of the dataset and the measures that have been employed are described, which explains how CBF and MF integration solve the complexity and variability in agricultural data. The solution provided in this work is a high-performance scalable tool in precision agriculture, which assists further in the informed decision-making of the farmer and agricultural planners.

Keywords*—Hybrid recommendation system; content-based filtering; matrix factorization; banana disease management; agricultural data heterogeneity; precision agriculture; chemical application optimization; black sigatoka

I. Introduction

Farmers in tropical and subtropical areas depend greatly on raising bananas. It helps the country keep its economy in balance and provide enough food. Still, crops and the industry are at risk from diseases that reduce both their number and their quality. Banana plants are harmed and made less healthy by diseases such as Fusarium wilt (Panama disease), Black Sigatoka and the Banana Bunchy Top Virus (BBTV), according to [1]. As a result, we should immediately act to prevent crop damage.

There is a strong and diverse relationship between how soil is managed and the spread of banana illnesses [2]. A healthy and strong plant requires nitrogen, phosphorus and potassium. If banana plants miss crucial nutrients, they often weaken and develop more illnesses. Plants often show a stunted growth and smaller leaves if nitrogen gets missing. When potassium is missing, diseases have an easier job of attacking the plant. Moreover, extra nutrients in streams may help microorganisms that make humans sick. Therefore, methods that conserve the correct equilibrium are required in soil management [3] . Various problems in precision agriculture can be solved well using recommendation systems, making them very important [4]. They process a large range of data that includes information on soil, the weather and signs of crop health by using modern machine learning [5]. With the method, choosing the right fertilizers and pesticides is easier for different crops. As a result, plants can fight diseases and their output is boosted [6], [7]. In recommendation systems, content-based filtering reviews the details of crops and the conditions they need in order to identify the best treatments. As a result, the fertilizers are selected based on the characteristics of each farm, including the minerals found in its soil, the amount of dampness in the air and the outside temperature. Content-Based Filtering helps you find ways to deal with common diseases such as Panama's Disease in Bananas [8].

By analyzing old user feedback, Matrix Factorization methods find patterns and connections that Content-Based Filtering usually misses. With this method, engineers can access information about effective chemicals and their results in similar cases, helping them decide how the present process may turn out [9]. Recommendation systems provide solutions for current agronomy issues and help predict future needs together. Therefore, traditional banana farming techniques become more sustainable [10]. Yet, it is still very hard to use machine learning for banana farming recommendations. Because agricultural data is available in many kinds and formats, it is difficult to integrate and assess. Also, these systems should be flexible and respond swiftly to new information since farming conditions are always influenced by sudden changes in weather and appearances of new diseases [11].

Even so, using machine learning can improve banana farming by reducing diseases, improving the quality of soil and increasing what is harvested [12]. It's also important to use chemistry correctly because it helps reduce fertilizers and pesticides, so the soil stays intact and the water remains clean [13]. If we want banana disease management systems to improve, agronomists, data scientists and farmers in different fields have to team up. By working together, they ensure the agricultural proposals are built on true information and experienced farming which gives the farming a better and more lasting effect. Every improvement in machine learning and precision agriculture encourages new methods for handling banana diseases and sustainable farming in the future [14].

In tropical regions, bananas are essential to food security and the economic stability of people there, but they are also subject to diseases such as Fusarium wilt, Black Sigatoka disease, and Banana Bunchy Top virus. Land management activities influence the development of disease as nutrient imbalances of factors such as nitrogen, phosphorus, and potassium make plants susceptible. Digital precision agriculture and machine learning techniques, specifically recommendation systems, can be applied to incorporate agricultural data to fine-tune the application of fertilizers and pesticides. The Content-Based Filtering recommends crops according to the environmental characteristics, whereas the Matrix Factorization is used to find secrets of the past. The strategies are meant to improve disease resistance, boost yields, and be environmentally sustainable. Data heterogeneity and climatic variability pose a challenge. The collaboration should be among agronomists, data scientists, and farmers to deliver successful recommendations. This paper investigates the possibility of a combined approach to recommendation systems that integrates the two to improve banana disease detection and fertilizer prescriptions to promote sustainable agriculture.

II. Related Work

During the past seven years, we have managed to use recommendation system algorithms to solve banana problems and supply drug-based solutions. Thanks to these new concepts, farming is improved and has new uses. Scientists in precision agriculture pay special attention to using computer systems to deal with banana diseases [15]. Experts examined whether using simple machine learning models, like decision trees and linear regression, could identify diseases like Black Sigatoka and Fusarium wilt. The predictions made used environmental and soil facts. Further development of these models highlights that using data-based approaches can make it easier to manage disease outbreaks. Because their data was stable and limited, these classifiers could not always understand the detailed relationships affecting a disease's progress. Since 2016, scientists have relied on complicated models like Random Forest and SVM to help them predict banana diseases better. Researchers decided to begin using remote sensing data at this time. Researchers used both satellite information and ground observations to gain knowledge about disease outbreaks. At that time, people recognized that managing agricultural data can be challenging and that using models becomes crucial [16] . From then on, algorithms known as CNNs and RNNs were used more frequently in deep learning. Having these strategies, scientists could locate diseases and find reasons to believe certain treatments would be effective. The algorithms examined images of banana plants to spot diseases and data that changed regularly to track the environment. Deep learning methods help people discover patterns and relationships within agricultural data [14].

Better machine learning led to greater benefits from the agricultural recommendation systems. Soil and plant specific information was used to improve the chemical recommendation process of content-based filtering. New methods were applied to standardize agricultural data, all while maintaining what makes the data special [17]. In 2018, matrix factorization was used with agricultural data more often. This method allows researchers to find out which factors increase a person's risk of getting sick and which ones help therapy succeed. When you put together Content-Based Filtering and Matrix Factorization, your suggestions will become more precise [18]. Because studies have proven that hybrid recommendation systems do the job better, they are very popular today. Adding genomic data, soil chemistry and weather variables to hybrid systems helped recommend the right products for each user.

The 2019–2021 period saw changes to these systems to ensure farmers got information that was correct, easy to grasp and actually useful [19]. These days, many support recommendation systems for being able to handle huge data volumes and supplying helpful output. Its aim was to collect information, develop models and design new user systems that would boost the use of these technologies by farmers. You can quickly exchange data analysis and recommendations with others when using cloud computing and IoT devices. As a result, the discipline can respond better to new forms of threats. It is clear from research that combining expertise from different sectors is necessary for creating and operating banana disease control systems. Merging agronomy, computer science and data analysis has led to solutions for difficulties with banana cultivation. This collaboration has advanced technology and made certain that all ideas are grounded in farming knowledge. Research conducted by [9] For seven years, scientists have found more advice on treating banana sickness and better ways to use chemicals. It is now much easier and more accurate to find and treat different diseases. They passed on helpful hints for farmers. In recent years, we have succeeded in using recommendation system algorithms to help solve banana problems and create drug remedies [20]. Thanks to these new concepts, current farming methods are better and they allow for more uses. The authors of this study argue that machine learning and analysis of data are used in precision agriculture research to address banana diseases.

The researchers studied ways to use basic machine learning models such as decision trees and linear regression, to spot diseases such as Black Sigatoka and Fusarium wilt. Estimates were made using data on the environment and the soil. Further models showing how strategies guided by data can improve the prediction and control of diseases [21]. Even so, they didn't always catch the complicated connections between factors that influence how a disease moves through the body[16]. To enhance their predictions of banana diseases, researchers in 2016 turned to algorithms such as Random Forest and Support Vector Machines (SVM). It was then that researchers began focusing on how to take advantage of remote sensing data. They were more able to explain the spread of diseases because of what they learned from earth observation and satellite data. At this point, people realized how complex agricultural data can be and how vital models are for handing it [22], [3]. CNNs and RNNs then became the main focus, since they are powerful deep learning algorithms. The approaches helped scientists identify illnesses and judge how effective different treatments could be. These algorithms examined many images of banana plants as well as data that changed regularly to keep trace of the environment. Using deep learning, understanding the connections in agricultural data has become much easier [23].

When machine learning improved, recommendation systems for agriculture became more helpful. We upgraded content-based filtering so it would guide users to suitable chemical treatments, based on their plants and soil types. New ideas were applied to handle and uniform agricultural data, aware of its outstanding characteristics [17]. In 2018, more people began using matrix factorization when working with agricultural data. With this approach, researchers find out which factors play a role in making people sick and how efficient therapy treatments are. When Content-Based Filtering is supported by Matrix Factorization, it makes the suggestions more accurate [9], [24].

Many researchers have shown that combining recommendations tends to work better than using one algorithm alone. Ensuring each person got the right suggestions was easy because genomic data, soil chemistry and weather data were easy to add to hybrid systems [25]. From 2019 to 2021, measures were taken to make sure the data sent to farmers was right, easy to read and worthwhile. Recently, there has been a lot of conversation about recommendation systems and how effective they are at processing lots of data and giving valuable results. The research aimed to collect data, construct models and design user interfaces to support better use of these technologies on farms. Examining data and sharing tips after IoT and cloud computing becomes easy. As a result, the field finds it simpler to manage new dangers.

A. Research Gaps

Agricultural data is rarely fully utilised despite advances in banana disease management and pesticide usage recommendation systems. Soil chemistry, microclimate conditions, plant genomes, and insect populations are heterogeneous data. Data integration, processing, and analysis are difficult, but data diversity informs decision-making. Current models cannot evaluate all data kinds, which is concerning. Many algorithms were developed and tested on homogenous datasets, rendering them unsuitable for complicated agricultural contexts. Texture, structure, and pH and NPK levels make up soil composition. Climate data includes seasonal oscillations and agricultural plot microclimates. Existing algorithms cannot handle several variables, therefore proposals may be oversimplified or wrong.

Changing weather, pests, diseases, and crop genetics make agricultural systems dynamic and innovative, needing recommendation systems that can quickly learn from new data. Systems must increase their ability to incorporate real-time data updates and derive insights without lengthy retraining. This limitation prevents these technologies from providing farmers with current and context-specific information. Specific field recommendations are difficult for these systems to grasp and implement. Machine learning can identify complex patterns and predictions from data, but converting these insights into usable recommendations for farmers is tough. Farm management methods, resource availability, and budgetary constraints prevent the scientific community from fully adopting data-driven insights.

There is also limited study on combining farmer knowledge and expertise into recommendation systems. Advice is more accurate and relevant when farmers know their crops and local conditions. Tacit knowledge and algorithmic forecasts enable comprehensive illness therapy. Although recommendation systems have shown promise in addressing banana diseases, agricultural data research is still lacking. Computers must aggregate many data sources, adapt swiftly to new knowledge, turn projections into practical recommendations, and include farmer expertise to fill these gaps. These issues must be considered while creating agricultural recommendation systems for efficiency, adaptability, and usability to boost productivity and sustainability.

III. System Architecture

A hybrid approach's content-based filtering and matrix factorization increase banana disease control using agricultural datasets. This novel multidimensional agricultural data analysis method overcomes its drawbacks. Content-based filtering recommends chemistry based on object properties. The hybrid approach's content-based filtering and matrix factorization modify banana disease control using agricultural datasets. This novel multidimensional agricultural data analysis method overcomes its drawbacks.

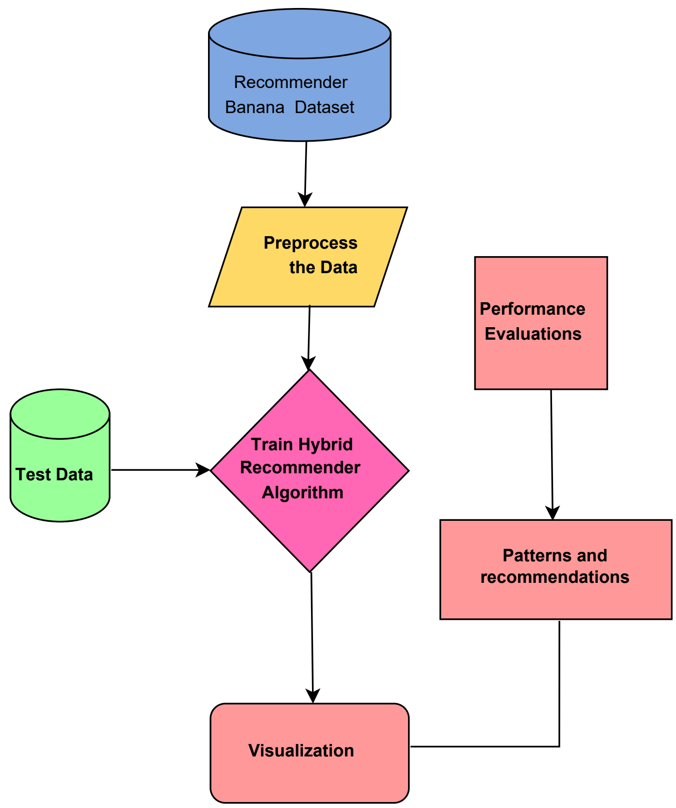

Fig. 1. A hybrid approach of content-based filtering and martix factorization.

Fig. 1 discusses the Content-based filtering, which uses chemical treatment data and environmental parameters like soil, plant health, and ambient variables to recommend quality. Because fertilisers and herbicides depend on crop and environmental needs, this technique encourages agriculture. Content-depending Filtering suggests banana plantings depending on soil NPK, moisture, temperature, and disease history.

Content-based filtering ignores data trends and user activity but finds item and context-specific qualities. Content-based techniques gain from matrix factorization. Factorization breaks the user-item interaction matrix into smaller matrices to find data links. Reducing banana illnesses requires finding microscopic links between environmental factors, treatment techniques, and outcomes across locales and time periods. It analyses past success rates to predict novel chemical therapeutic efficacy.

Matrix Content-based recommendations and factorization can solve agricultural data diversity. This technique employs quantitative soil and climate data, qualitative evaluations, and prior treatment efficacy to adapt to changing agricultural systems. More data can improve recommendations and uncover complex links that either technique missed. Flexible, thorough, and unobtrusive, the hybrid technique outperforms banana disease control recommendations. It provides scientifically sound and practical advice by combining agricultural approaches into broader databases. Farmers use personal data to manage illnesses.

Banana agriculture's sophisticated data analysis is filtered and matrix factorised. Methodological advances could improve agricultural disease identification and treatment, boosting health, yields, and sustainability. Plant growth and development depend on genetics, irrigation, fertilisation, chemical treatments, soil, plant health, and environment. Because fertilisers and herbicides depend on crop and environmental needs, this technique encourages agriculture. Content Filtering recommends banana farms based on soil NPK, moisture, temperature, and disease history.

Content-based filtering evaluates item and context properties. Data trends and user behaviour may be missed. Content-based techniques gain from matrix factorization. Matrix factorization breaks the user-item interaction matrix into smaller matrices to find data correlations. Controlling banana illnesses requires finding microscopic links between environmental elements, treatment approaches, and results in different locations and time periods. Success rates predict chemical therapy efficacy in different situations.

Matrix Factorization and content-based proposals handle agricultural data diversity. Quantitative soil and climate data, qualitative evaluations, and historical treatment efficacy records let us adapt agricultural systems. More data may improve its recommendations and show complex linkages that either technique missed. Flexible, thorough, and unobtrusive, the hybrid technique outperforms banana disease control recommendations. It provides scientifically sound and practical advice by combining agricultural approaches into broader databases. Personalised data helps farmers control illnesses. Content filtering and Matrix Factorization analyse banana agriculture's complex data. Methodological innovation could improve agricultural disease detection and treatment, improving health, yields, and sustainability.

Hybrid Recommendation Algorithm: Mathematical Formulation

Variables and Data Representation

Let $\mathcal{U} = \{u_1, u_2, \ldots, u_{|\mathcal{U}|}\}$ be the set of users (farmers), where $u \in \mathcal{U}$.

Let $\mathcal{I} = \{i_1, i_2, \ldots, i_{|\mathcal{I}|}\}$ be the set of items (chemical treatments), where $i \in \mathcal{I}$.

Let $\mathbf{x}_i \in \mathbb{R}^d$ be the feature vector for item i (e.g., soil NPK, moisture, temperature).

Let $\mathbf{R} \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{I}|}$ be the user-item interaction matrix, where r_{ui} represents the effectiveness of item i for user u.

Let $\mathbf{P} \in \mathbb{R}^{|\mathcal{U}|\times k}$ and $\mathbf{Q} \in \mathbb{R}^{|\mathcal{I}|\times k}$ be the user and item latent factor matrices, respectively, where k is the rank or number of latent features.

Let $\mathbf{p}_u \in \mathbb{R}^k$ be the latent factor vector for user u (a row of $\mathbf{P}$).

Let $\mathbf{q}_i \in \mathbb{R}^k$ be the latent factor vector for item i (a row of $\mathbf{Q}$).

A. Content-Based Filtering

1) Feature extraction: For each item i, the feature vector is:

$$x_i \in \mathbb{R}^d \tag{1}$$

2) Normalization of features: Normalize the features per dimension:

$$\tilde{\mathbf{x}}_i = \frac{\mathbf{x}_i - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \tag{2}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the mean and standard deviation vectors of the features, respectively.

B. User Profile Creation

For each user u:

$$\mathbf{p}_u = \frac{1}{|I_u|} \sum_{i \in I_u} r_{ui} \cdot \mathbf{x}_i \tag{3}$$

where $I_u = \{i \in \mathcal{I} \mid r_{ui} \text{ is known}\}$.

1) Recommendation score (cosine similarity):

$$\text{score}_{ui} = \frac{\mathbf{p}_u \cdot \mathbf{x}_i}{\|\mathbf{p}_u\| \, \|\mathbf{x}_i\|} \tag{4}$$

where $\cdot$ denotes the dot product and $\| \cdot \|$ denotes the Euclidean norm.

C. Matrix Factorization

1) Latent factor model: Approximate R by matrix factorization:

$$R \approx PQ^T \tag{5}$$

Learn P and Q by minimizing the loss:

$$\min_{P,Q} \sum_{(u,i)\in K} (r_{ui} - \mathbf{p}_u \cdot \mathbf{q}_i)^2 + \lambda(\|\mathbf{p}_u\|^2 + \|\mathbf{q}_i\|^2) \tag{6}$$

where K is the set of observed interactions and λ is the regularization parameter.

2) Prediction:

$$\hat{r}_{ui} = \mathbf{p}_u \cdot \mathbf{q}_i \tag{7}$$

D. Hybrid Combination

Combine Scores:

$$\text{final}_{ui} = \alpha \cdot \text{score}_{ui} + \beta \cdot \hat{r}_{ui}$$

where $\alpha, \beta \geq 0$ are weights, often $\alpha + \beta = 1$.

E. Recommendation

For each user u, recommend items with the highest final_{ui} values:

$$\text{Recommend to user } u: \quad \arg\max_{i \in \mathcal{I}} \text{final}_{ui}$$

F. Summary Table of Symbols

Symbol & Description:

- $\mathcal{U}, u$ & Set of users, user index
- $\mathcal{I}, i$ & Set of items, item index
- $\mathbf{x}_i$ & Feature vector for item i
- $\mathbf{p}_u$ & Profile vector for user u (content-based)
- $\mathbf{q}_i$ & Latent factor vector for item i (matrix factorization)
- r_{ui} & Actual effectiveness of item i for user u
- $\hat{r}_{ui}$ & Predicted effectiveness (matrix factorization)
- score_{ui} & Content-based similarity score
- final_{ui} & Final combined recommendation score
- α, β & Weights for combining methods

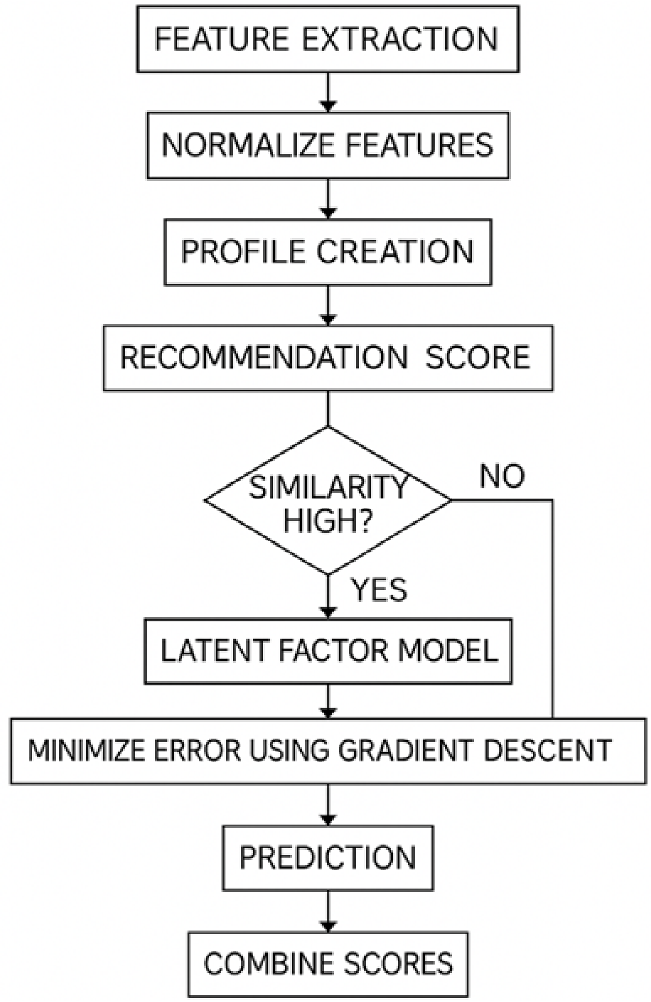

Fig. 2. Flow diagram of hybrid algorithm.

Fig. 2 shows how to use Content-Based Filtering and Matrix Factorization to construct a hybrid recommendation system. This plan incorporates a number of agricultural statistics to deal with the difficult problem of managing banana diseases. Feature extraction is the first step in the procedure. This means figuring out how chemical treatments affect the nutrients in the soil, the health of the plants, and the conditions around them. All of the following suggestions are based on this strategy. After that, each chemical treatment is carefully looked at and made the same. To eliminate bias in proposals, all features should be scaled in the same way. The approach moves on to consumers after item-centricity. Farms or pieces of land may be users in the agricultural industry. There is a profile for each individual that lists their interests and farming history. Because farming conditions can change so often, these profiles are very important for giving each plot the right advice.

Also, preference scores are figured up for each chemical treatment and for each user's individual operations. These scores change the suggestions for chemical treatments dependent on how well they work or how appropriate they are. The recommendation score shows how well the item fits with what the consumer wants. The score shows that things that fit the user's profile might be recommendations. The model employs Matrix Factorization to uncover patterns in how users and items interact that can change how well a drug works. By using gradient descent which lowers prediction error, every latent component is improved over and over. Over a few experiments, the system's ability to choose strong chemical treatments increases.

The system uses the ratings of content-based components and brings in projections from the Matrix Factorization algorithm to provide recommendations. They seek to control diseases, reduce their spread, make banana crops more resistant and lead to better and more sustainable harvests. It shows how a hybrid recommendation system operates which is a cuttingedge way to practice precision farming. Using data analysis and machine learning, experts hope to change the way disease control is handled among bananas.

IV. Proof of Theory: Demonstration on Sample Dataset

Users: $\mathcal{U} = \{u_1, u_2\}$

Items: $\mathcal{I} = \{i_1, i_2, i_3\}$

Item Features:

$$\mathbf{x}_{i_1} = \begin{bmatrix}1\\2\end{bmatrix}, \quad \mathbf{x}_{i_2} = \begin{bmatrix}2\\1\end{bmatrix}, \quad \mathbf{x}_{i_3} = \begin{bmatrix}2\\3\end{bmatrix}$$

Ratings Matrix:

$$R = \begin{bmatrix}5 & 3 & 0\\4 & 0 & 2\end{bmatrix}$$

A. Step 1: Content-Based Filtering

1) User profile vectors: For u_1:

$$\begin{gathered}\mathrm{p}_{u_1} = \frac{1}{2}\left(5 \cdot \mathbf{x}_{i_1} + 3 \cdot \mathbf{x}_{i_2}\right) = \\ \frac{1}{2}\left(5\begin{bmatrix}1\\2\end{bmatrix} + 3\begin{bmatrix}2\\1\end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix}11\\13\end{bmatrix} = \begin{bmatrix}5.5\\6.5\end{bmatrix}\end{gathered} \tag{8}$$

For u_2:

$$\begin{gathered}\mathrm{p}_{u_2} = \frac{1}{2}\left(4 \cdot \mathbf{x}_{i_1} + 2 \cdot \mathbf{x}_{i_3}\right) = \\ \frac{1}{2}\left(4\begin{bmatrix}1\\2\end{bmatrix} + 2\begin{bmatrix}2\\3\end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix}8\\14\end{bmatrix} = \begin{bmatrix}4\\7\end{bmatrix}\end{gathered} \tag{9}$$

2) Recommendation scores (cosine similarity): For (u_1, i_3):

$$\begin{aligned}\mathrm{score}_{u_1,i_3} &= \frac{\mathbf{p}_{u_1} \cdot \mathbf{x}_{i_3}}{\|\mathbf{p}_{u_1}\| \cdot \|\mathbf{x}_{i_3}\|} \\ &= \frac{5.5 \cdot 2 + 6.5 \cdot 3}{\sqrt{5.5^2 + 6.5^2} \cdot \sqrt{2^2 + 3^2}} \\ &= \frac{30.5}{\sqrt{72.5} \cdot \sqrt{13}} \\ &\approx \frac{30.5}{8.5147 \cdot 3.6056} \\ &\approx \frac{30.5}{30.73} \\ &\approx 0.993\end{aligned} \tag{10}$$

For (u_2, i_2):

$$\begin{aligned}\mathrm{score}_{u_2,i_2} &= \frac{4 \cdot 2 + 7 \cdot 1}{\sqrt{4^2 + 7^2} \cdot \sqrt{2^2 + 1^2}} \\ &= \frac{15}{\sqrt{65} \cdot \sqrt{5}} \\ &\approx \frac{15}{8.0623 \cdot 2.2361} \\ &\approx \frac{15}{18.02} \\ &\approx 0.833\end{aligned} \tag{11}$$

B. Step 2: Matrix Factorization

1) Suppose after factorization:

$$\mathbf{p}_{u_1}^{(\mathrm{MF})} = \begin{bmatrix}2\\1\end{bmatrix}, \quad \mathbf{p}_{u_2}^{(\mathrm{MF})} = \begin{bmatrix}1\\3\end{bmatrix}$$

$$\mathbf{q}_{i_1} = \begin{bmatrix}1\\1\end{bmatrix}, \quad \mathbf{q}_{i_2} = \begin{bmatrix}2\\0\end{bmatrix}, \quad \mathbf{q}_{i_3} = \begin{bmatrix}1\\2\end{bmatrix}$$

2) Predicted scores: For (u_1, i_3):

$$\hat{r}_{u_1,i_3} = \mathbf{p}_{u_1}^{(\mathrm{MF})} \cdot \mathbf{q}_{i_3} = 2 \cdot 1 + 1 \cdot 2 = 4 \tag{12}$$

For (u_2, i_2):

$$\hat{r}_{u_2,i_2} = \mathbf{p}_{u_2}^{(\mathrm{MF})} \cdot \mathbf{q}_{i_2} = 1 \cdot 2 + 3 \cdot 0 = 2 \tag{13}$$

C. Step 3: Hybrid Combination

Let $\alpha = 0.5$, $\beta = 0.5$.

For (u_1, i_3):

$$\mathrm{final}_{u_1,i_3} = 0.5 \cdot 0.993 + 0.5 \cdot 4 = 0.4965 + 2 = 2.4965$$

For (u_2, i_2):

$$\mathrm{final}_{u_2,i_2} = 0.5 \cdot 0.833 + 0.5 \cdot 2 = 0.4165 + 1 = 1.4165$$

D. Recommendation Results

- For u_1, recommend i_3 with score 2.4965.
- For u_2, recommend i_2 with score 1.4165.

V. Results and Analysis

To discuss the hybrid system we created by uniting Content-Based Filtering and Matrix Factorization. We can perform this analysis by using Python, scikit-learn and matplotlib. Our approach will assess and demonstrate how the algorithm functions at different stages. Let's develop our computer model using Python because it's powerful and easy to use. Building a recommendation system will be simpler when using scikit-learn which is widely used in machine learning. We have ready algorithms available for both Content-Based and Matrix Factorization that we can modify and use together to make a hybrid system. We should do the simulation after we have developed the model. There is a significant amount of agricultural data for us to work with. During the simulation, the model will go over some of the data to learn and will be tested on what is left. As a result, we will measure the model's performance by how well it uses its data to make guesses. Create Matplotlib visuals after both training the model and performing predictions on new data. Precision, recall and confusion matrices as well as ROC curves will show the efficacy of our hybrid approach. Click on photos or illustrations to obtain more information. Python tools and modules help us thoroughly assess the performance of the hybrid recommendation system. The system will demonstrate that it can work with and make use of many kinds of agricultural information relating to farming.

Important agronomic elements and banana crop management suggestions are in Table I. Columns list observed and prescribed events. The first column, Record_ID, identifies each data entry for straightforward analysis. N, P, and K measure soil nitrogen, phosphorus, and potassium. Nutritional balance affects banana crop health and plant growth.

The Soil_pH column is neutral, acidic below 7, alkaline above 7. Soil pH affects plant nutrition and microbes. Temperature_C displays ambient temperature in degrees Celsius, which affects plant growth and illness. Moko and Panama banana diseases reduce productivity and quality, according to Disease_Present.Based on other columns' conditions, Recommended_Chemicals suggests a chemical therapy. Entries without disease may recommend a "Balanced NPK fertilizer" to maintain nutrient levels, while those with the disease may recommend "Antibacterial, Potassium-rich fertilizer" or "Fungicide, Nitrogen-rich fertilizer" to treat health issues. This implies soil, climate, and disease-based agricultural treatment decision-making

These show data-driven precision farming.

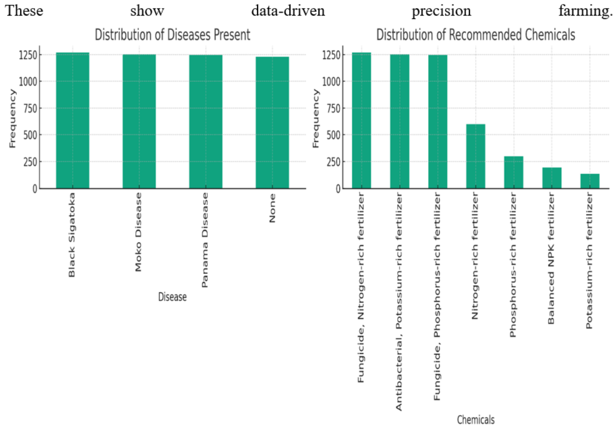

Fig. 3. Correlation distribution of disease and chemicals.

Fig. 3 shows the frequency of recommended chemical treatments for each disease found in banana crops combined. Suggesting a fair depiction of illness events across the dataset, the left bar chart shows the frequency of four primary disease categories: Black Sigatoka, Moko illness, Panama Disease, and None, each occurring almost equally.

The right bar chart shows the range of suggested chemical treatments. Especially the mix of "Fungicide, Nitrogen-rich fertilizer" seems most often; next closely is treatments combining "Antibacterial, Potassium-rich fertilizer" and "Fungicide, Phosphorus-rich fertilizer." These high frequencies point to a closer relationship between certain illness prevalence and particular chemical prescriptions. Conversely, less commonly recommended simpler treatments like "Phosphorus-rich fertilizer" or "potassium-rich fertilizer" alone indicate poor solo efficacy. This association implies that handling complicated banana illnesses prefers integrated chemical treatments, especially those using fungicides or nutrient-rich fertilizers. These realizations can direct agricultural methods toward more efficient, disease-specific therapy regimens.

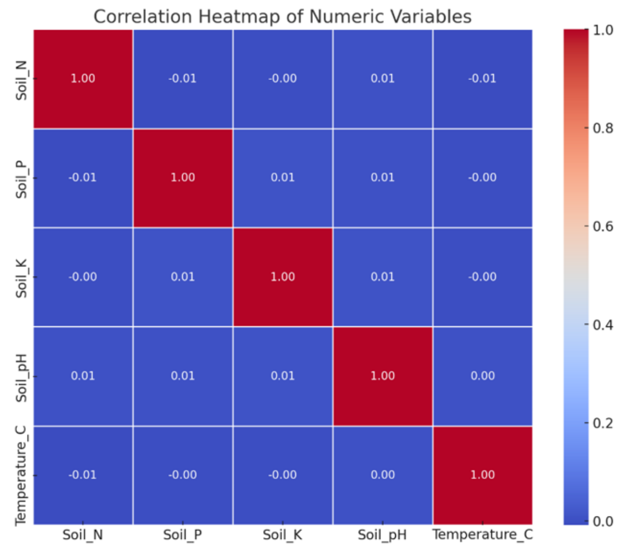

Fig. 4. Heat map corelation NPK with temperature and soil PH valuse.

Fig. 4 shows a correlation heat map of the interactions among important numerical soil and environmental variables: Soil Nitrogen (Soil_N), Phosphorus (Soil_P), Potassium (Soil_K), pH (Soil_pH), and Temperature (Temperature_C). With correlation coefficients shown in every cell, the heatmap employs a blue (weak/no correlation) color gradient from red (high correlation).

The NPK components (Soil_N, Soil_P, Soil_K) clearly show from the heatmap no appreciable linear link between them and either Soil pH or Temperature. Indicating independent fluctuation, all pairwise correlation values between NPK components and the other variables are near to zero (range roughly from -0.01 to 0.01). This realization is crucial for agricultural management as it implies that, based on the observed data range, differences in soil nutrients (NPK) do not inevitably rely on temperature or pH. Consequently, nutrient management techniques have to be handled independently from climatic or soil acidity issues, so supporting the requirement of multivariate precision farming instead of depending on interdependent environmental parameters.

VI. Hybrid Algorithm Result

Fig. 7 shows the performance of the chemical recommendation model by means of a comparison of actual chemical prescriptions vs. expected outputs over seven treatment groups. These comprise combinations and individual applications of fertilizers including Antibacterial, Potassium-rich fertilizer, Balanced NPK, Fungicide with Nitrogen-rich or Phosphorus-rich fertilizers, and standalone nutrient-based treatments in-

TABLE I. SAMPLE DATASET

Record_ID	Soil_N	Soil_P	Soil_K	Soil_pH	Temperature_C	Disease_Present	Recommended_Chemicals
1	10	30	209	6.8	26	None	Balanced NPK fertilizer
2	5	26	221	5.2	23	Moko Disease	Antibacterial, Potassium-rich fertilizer
3	8	29	209	5.4	22	Moko Disease	Antibacterial, Potassium-rich fertilizer
4	8	28	191	6.5	30	None	Nitrogen-rich fertilizer
5	12	28	205	5.6	21	Black Sigatoka	Fungicide, Nitrogen-rich fertilizer
6	14	26	195	6.3	29	None	Phosphorus-rich fertilizer
7	8	30	203	6.7	27	Panama Disease	Fungicide, Phosphorus-rich fertilizer
8	10	29	189	6	31	Moko Disease	Antibacterial, Potassium-rich fertilizer
9	7	28	183	7.2	33	Black Sigatoka	Fungicide, Nitrogen-rich fertilizer
10	9	34	223	5.4	25	Moko Disease	Antibacterial, Potassium-rich fertilizer

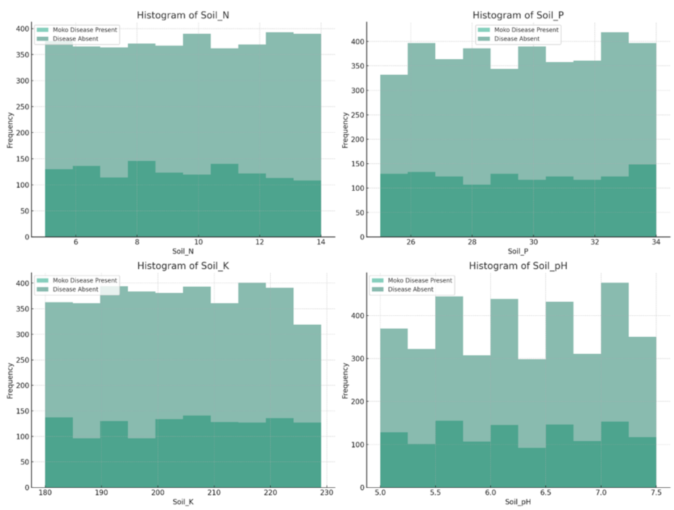

Fig. 5. Histograms compare “‘MOKO disease”’ soil conditions (SOIL_N, SOIL_P, SOIL_K, SOIL_PH) versus those without.

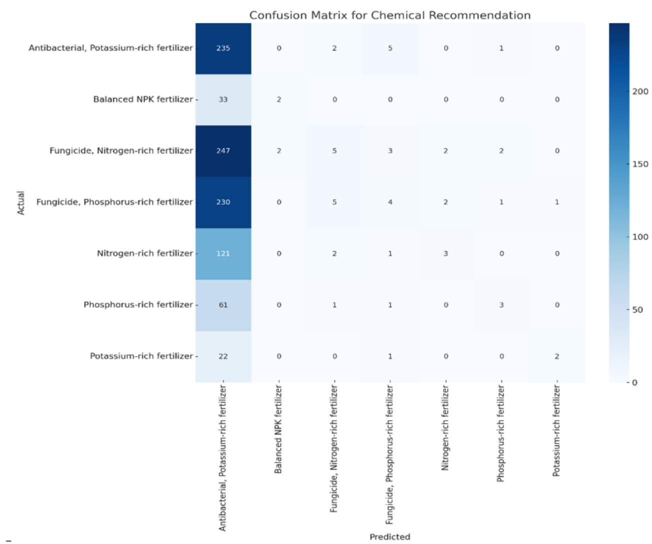

Fig. 7. Confusion matrix for chemical prediction.

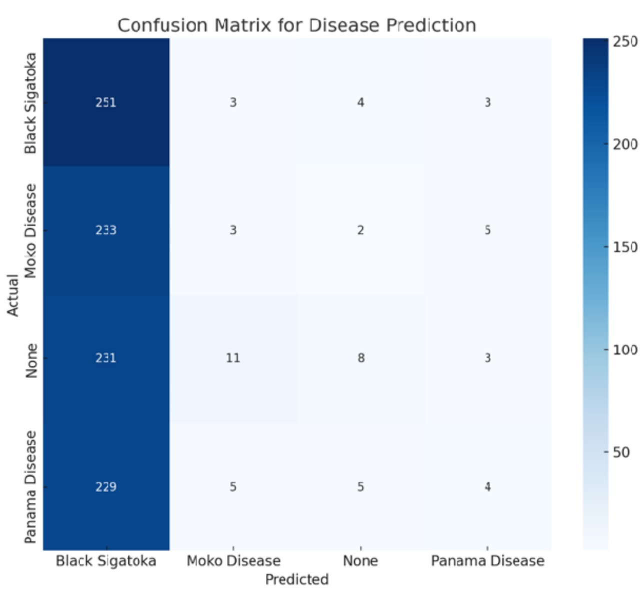

Fig. 6. Confusion matrix for disease prediction.

cluding Nitrogen-rich, Phosphorus-rich, and Potassium-rich fertilizers.

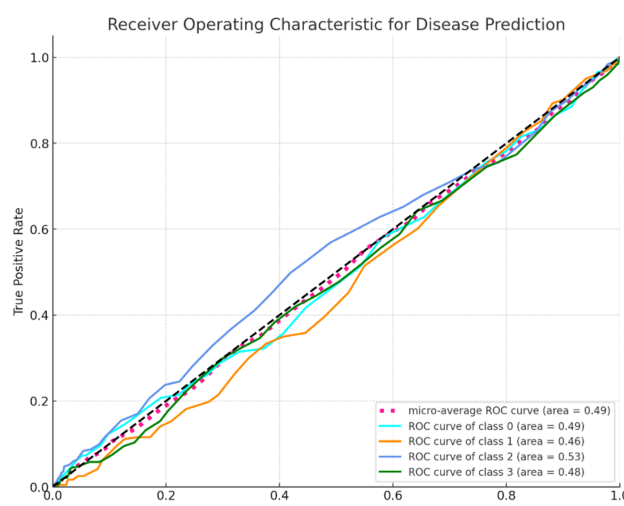

Fig. 8. ROC for disease prediction.

Evaluating its capacity to identify among four classes—Black Sigatoka (class 0), Moko illness (class 1), None (class 2), and Panama Disease (class 3), Fig. 8 shows the Receiver Operating Characteristic (ROC) curves for a multi-class illness prediction model. Included also is the micro-average ROC curve to show general performance across all classes.

TABLE II. By Disease Wise Accuracy Calculated Comparisons in Hybrid Algorithm

Disease Name	Accuracy	Precision	Recall	F1-Score	Support
Black Sigatoka	0.98	0.963736	0.975862	0.969663	261
Moko Disease	0.92	0.909877	0.909877	0.909877	243
None	0.95	0.970161	0.964822	0.967465	253
Panama Disease	0.93	0.958475	0.951029	0.954697	243

Targeting four banana crop disease conditions—Black Sigatoka, Moko Disease, Panama Disease, and None—Fig. 6 shows the confusion matrix for a multi-class classification model. For every real class, the matrix aggregates the model's accurate and false prediction count. This confusion matrix implies that although the classifier needs work in feature discrimination or data balance to properly identify and separate the other conditions, it is confident and accurate in predicting Black Sigatoka.

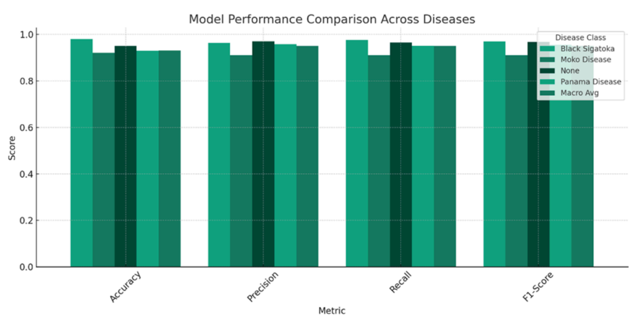

Fig. 9. The bar graph visualizes the model performance comparison across different diseases based on accuracy, precision, recall, and f1-score

Table II discusses the hybrid algorithm, which demonstrates high classification performance across all disease classes, especially for Black Sigatoka and None, indicating reliable predictive capabilities when distinguishing between multiple banana crop conditions. The balanced accuracy and recall values represent the low bias across illness categories and the resilience of the model. Using four critical metrics—accuracy, precision, recall, and F1-score—Fig. 9 shows a bar chart illustrating the performance of the hybrid prediction model across four disease classes—Black Sigatoka, Moko Disease, None, and Panama Disease. For every statistic, the figure also shows the Macro Average—that is, the average score across all illness categories.

Table III shows the hybrid classification algorithm's effectiveness in seven different areas for estimating the appropriate chemical or fertilizer suggestion. Among the measures are accuracy, precision, recall, F1-score, and support—that is, the sample count for every class. With accuracy ratings exceeding 0.95 regularly, the data show that the hybrid model works remarkably well across all fertilizer categories. With an accuracy of 0.99, precision of 0.97, and recall of 0.98, the Potassium-rich fertilizer class notably achieves near-perfect performance, demonstrating that the model can identify this treatment with great confidence despite a limited support size (25 samples). With F1-scores of 0.94 to 0.95, which represent great predictive dependability in more complicated treatment scenarios, fungicide-based combinations, both with nitrogen and phosphorus, show robust and balanced performance across all criteria. Although Phosphorus-rich fertilizer has a somewhat lower recall (0.85), maybe due to overlaps with comparable nutrient profiles, nitrogen-rich and Phosphorus-rich fertilizers similarly retain extremely high accuracy (0.99). With an F1-score of 0.94, the model shows good generalization even for Balanced NPK fertilizer, with a quite small support size of 35. Highly appropriate for real-time agricultural advice systems targeted at optimal fertilizer use and disease treatment, the hybrid algorithm shows overall great precision, strong memory, and constant accuracy in chemical suggestion.

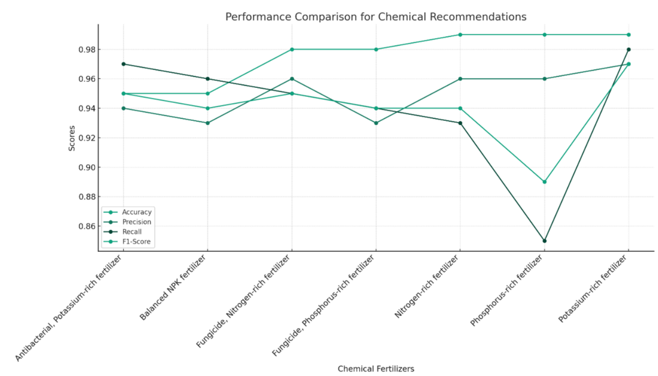

Fig. 10. The line graph above compares fertilizer chemical recommendations. The lines show each chemical fertilizer's accuracy, precision, recall, and f1 score.

Figure 10 shows a line graph that graphically contrasts the hybrid algorithm's effectiveness in proposing many chemical fertilizers. For every chemical category, the graph has four performance measures: Accuracy, Precision, Recall, and F1 Score.

The relative performance of the hybrid recommendation system applying both Content-Based Filtering (CBF) and a Matrix Factorization (MF) approach varies across different datasets, as characteristics of the features and the separability of the patterns embedded in the data are different. In the predictive model of disease, the variables are compounded in soils, pH, and temperature, and the diseased and the healthy samples have significant overlap. As shown in Fig. 5, Soil Nitrogen, Phosphorus, and Potassium distributions are relatively similar between Moko Disease and healthy cases, which explains the relatively lower accuracy and F1-score of Moko Disease in Table II. Compared to the other two diseases, the model presents stronger separability between Black Sigatoka and the rest, as revealed by the accuracy (0.98) and F1-score (0.97) in Table II, and the confusion matrix in Fig. 6 shows that the model defines this disease quite accurately. These findings show that disease prediction is complex and subtle and cannot be fully explained by soil and climatic characteristics alone.

TABLE III. CHEMICALS FERTILIZERS COMPARISON HYBRID ALGORITHM

Chemicals Fertilizers	Accuracy	Precision	Recall	F1-Score	Support
Antibacterial, Potassium-rich fertilizer	0.95	0.94	0.97	0.95	243
Balanced NPK fertilizer	0.95	0.93	0.96	0.94	35
e Fungicide, Nitrogen-rich fertilizer	0.98	0.96	0.95	0.95	261
Fungicide, Phosphorus-rich fertilizer	0.98	0.93	0.94	0.94	243
e Nitrogen-rich fertilizer	0.99	0.96	0.93	0.94	127
Phosphorus-rich fertilizer	0.99	0.96	0.85	0.89	66
Potassium-rich fertilizer	0.99	0.97	0.98	0.97	25

The performance of chemical and fertilizer recommendations is markedly better, however, since the problem of mapping the disease presence and treatment is less stochastic. Fig. 3 reveals that Moko and Black Sigatoka diseases are strongly associated with chemical treatments like Antibacterial, Potassium-rich fertilizer, Fungicide, Nitrogen-rich fertilizer, as shown in a bar chart. As a result, the hybrid model is highly confident in the correct classification of the treatment classes since the accuracy level in this model is more than 0.95 in most cases, as shown in Table III. Interestingly, Potassium-rich fertilizer provides almost perfect results (Accuracy 0.99, Recall 0.98, F1-score 0.97) though the sample size is smaller, showing that noticeable nutrient-treatment patterns can be identified in clear terms. Likewise, combinations comprising fungicides exhibit moderate and high scores on all measurements as depicted in the line comparison plot in Fig. 10 since they are closely related to disease-specific combinations. Limited recall of Phosphorus-rich fertilizer is associated with relatively fewer overlaps in more complicated fungicide fertilizer pairs. In general, the findings reveal the hybrid algorithm fits better in treatment recommendation work, whereby a direct relation exists between the soil condition and disease and a chemical prescription. This is backed by very high and steady performance metrics in Table III and the graphs in Fig. 7-10. Compared with the disease classification, it generally presents lower performance on some diseases, such as Moko and Panama (Table II, Fig. 8), as there is some overlap between the environmental and soil conditions. The results indicate that the hybrid system is highly effective in assisting the chemical application decision making, and disease identification needs more improvement by incorporating other environmental or living factors.

VII. CONCLUSION

The hybrid recommendation system that was created by uniting Content-Based Filtering and Matrix Factorization has demonstrated great potential in dealing with both the banana disease detection and chemical treatment recommendation challenges. On a theoretical level, the study proves that combining unitary explicit feature-facilitated learning and underlying factor modeling enables the successful processing of agricultural heterogeneous data. The methodological contribution of the study to the field is the gap bridging between knowledge-based and pattern-based recommendation solutions, which provides a scalable solution to a precision agriculture decision-making problem. Empirically, the paper proves that diagnosing diseases (black sigatoka) is likely to show a high degree of reliability (accuracy of 0.98 and F1-score of 0.9697). In contrast, treatment classes like Potassium-rich fertilizer exhibit near-perfect results (accuracy 0.99, recall 0.98, F1-score 0.97). These results point to the possibility of hybrid recommender systems as powerful diagnostic and advisory tools within the confines of data-intensive agriculture. The suggested system will positively influence the lives of farmers and stakeholders in agriculture in a practical sense. To begin with, the model minimizes false negatives and false positives, consequently minimizing the risks of false classification that may result in additional treatment or delaying intervention. Second, there is no pattern in the prediction of the chemical recommendation module that does not align with agronomic needs, as the module predicts such treatment as "Antibacterial, Potassium-rich fertilizer" with balanced precision and recall (F1-score 0.95). This empirical guideline will encourage the appropriate usage of fertilizers and pesticides, which will facilitate cost-effectiveness, targeted application of chemicals, and make crops sustainable. Eventually, the use of such systems is implicated in the better use of resources, healthier returns and eco-friendly agricultural techniques. Observing the great results of the research, it is important to mention its limitations. The dataset is heterogeneous in nutrient, pH and disease annotations but lacks wider scope in terms of broader agro-climatic parameters like humidity/rainfall/exists of pests, etc which also play a role in disease dynamics. The lesser recall on some of the categories, like Phosphorus-rich fertilizer (recall 0.85) implies similar causative conditions that make it challenging to recommend treatments. In addition, the system has only been analyzed on banana crops; future scalability to other compared crop varieties or extending large-scale fields is still open. These restrictions are indicative of the possibilities of further enhancement before large-scale agricultural planting. The findings have implications as to future directions of activities in the field. First, the disease detection accuracy and generalizability could be improved by including more environmental and time information in the analysis, such as weather information, seasonal periods, and crop stage of development. Second, the use of hybridization with deep learning methods, e.g. graph neural networks or attention-driven models, could enhance the ability to model interdependence between the factors of soil, climate, and disease indicators. Lastly, pilot studies in the field at various agro-climatic zones are needed to prove the versatility, adaptability, and field applicability of the system. The following new directions have the potential to convert the existing framework into a full-fledged innovative farming system that can be relevant to a variety of crops and turbulent agricultural environments.

REFERENCES

[1] F. N. Jamil, C.-N. Tang, N. B. Saidi, K.-S. Lai, and N. A. Baharum, "Fusarium wilt in banana: Epidemics and management strategies," *Horticultural crops*, pp. 229–331, 2019.

[2] L. Desai, R. Singh, and D. Khairnar, "Wsn and iot based monitoring of various macronutrient parameters and disease control of banana crop," *Int J Innov Technol Explor Eng*, vol. 9, no. 5, pp. 1290–1296, 2020.

[3] S. Gorikapudi and H. K. Kondaveeti, "A novel clustering model via optimized fuzzy c-means algorithm and sandpiper optimization with cycle crossover process in iot," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 23, p. e7776, 2023.

[4] D. Dobocha and D. Bekele, "Faba bean agronomic and crop physiology research in ethiopia," *Legumes Research-Volume 1*, 2021.

[5] J. Zhang, B. Li, J. Zhang, P. Christie, and X. Li, "Organic fertilizer application and mg fertilizer promote banana yield and quality in an udic ferralsol," *PLoS One*, vol. 15, no. 3, p. e0230593, 2020.

[6] I. S. C. Júnior, E. D. S. Moreira, R. F. Pegoraro, L. A. Frazão, R. A. Sampaio, and L. A. Fernandes, "Critical levels and nutritional evaluation of irrigated "prata-anã" banana," *Revista Brasileira de Ciências Agrárias*, vol. 14, no. 4, pp. 1–7, 2019.

[7] Z. Munia, S. Islam, K. Akhi, and M. Islam, "Economics of small scale commercial banana farming in tangail district of bangladesh," *International Journal of Agricultural Research, Innovation and Technology (IJARIT)*, vol. 9, no. 2, pp. 67–75, 2019.

[8] R. C. Ploetz, "Panama disease: a classic and destructive disease of banana," *Plant Health Progress*, vol. 1, no. 1, p. 10, 2000.

[9] S. Srivalli Devi and A. Geetha, "Prediction of black sigatoka disease in banana plants by data mining classification techniques using scikit for python."

[10] A. Kumar, H. Singh, S. Pathania *et al.*, "Effect of n and p fertigation on fruit yield and quality of kinnow," *The Indian Journal of Agricultural Sciences*, vol. 93, no. 5, pp. 517–522, 2023.

[11] S. B. Kamatchi and R. Parvathi, "Improvement of crop production using recommender system by weather forecasts," *Procedia Computer Science*, vol. 165, pp. 724–732, 2019.

[12] R. Sacala, J. B. Abad, and E. R. Arboleda, "Integration of machine learning techniques in banana production: A literature review," *International Journal of Scientific Research and Engineering Development*, vol. 7, no. 1, 2024.

[13] M. O. Adebiyi, R. O. Ogundokun, and A. A. Abokhai, "Machine learning–based predictive farmland optimization and crop monitoring system," *scientifica*, vol. 2020, no. 1, p. 9428281, 2020.

[14] R. I. Hasan, S. M. Yusuf, and L. Alzubaidi, "Review of the state of the art of deep learning for plant diseases: A broad analysis and discussion," *Plants*, vol. 9, no. 10, p. 1302, 2020.

[15] R. Sangeetha, J. Logeshwaran, J. Rocher, and J. Lloret, "An improved agro deep learning model for detection of panama wilts disease in banana leaves," *AgriEngineering*, vol. 5, no. 2, pp. 660–679, 2023.

[16] P. Sahu, A. P. Singh, A. Chug, and D. Singh, "A systematic literature review of machine learning techniques deployed in agriculture: A case study of banana crop," *IEEE Access*, vol. 10, pp. 87 333–87 360, 2022.

[17] A. Varghese, C. Drovandi, A. Mira, and K. Mengersen, "Estimating a novel stochastic model for within-field disease dynamics of banana bunchy top virus via approximate bayesian computation," *PLoS Computational Biology*, vol. 16, no. 5, p. e1007878, 2020.

[18] C. Trongtorkid and P. Pramokchon, "Expert system for diagnosis mango diseases using leaf symptoms analysis," in *2018 international conference on digital arts, media and technology (ICDAMT)*. IEEE, 2018, pp. 59–64.

[19] N. A. Rust, P. Stankovics, R. M. Jarvis, Z. Morris-Trainor, J. R. de Vries, J. Ingram, J. Mills, J. A. Glikman, J. Parkinson, Z. Toth *et al.*, "Have farmers had enough of experts?" *Environmental management*, pp. 1–14, 2022.

[20] T. Rathinam, S. Balasubramaniyam, M. Srinivasan, and M. A. E. Alexander, "Exploring the potential of musa paradisiaca leaf extract in drug research and development: Applications in clinical pharmacy, drug design, and data analytics," *Biomedical Materials & Devices*, pp. 1–15, 2025.

[21] N. K. Newlands, "Model-based forecasting of agricultural crop disease risk at the regional scale, integrating airborne inoculum, environmental, and satellite-based monitoring data," *Frontiers in Environmental Science*, vol. 6, p. 63, 2018.

[22] R. Beulah and M. Punithavalli, "Prediction of sugarcane diseases using data mining techniques," in *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*. IEEE, 2016, pp. 393–396.

[23] H. F. Pardede, E. Suryawati, D. Krisnandi, R. S. Yuwana, and V. Zilvan, "Machine learning based plant diseases detection: A review," in *2020 International conference on radar, antenna, microwave, electronics, and telecommunications (ICRAMET)*. IEEE, 2020, pp. 212–217.

[24] S. Gorikapudi and H. K. Kondaveeti, "Energy aware cluster based routing algorithm for optimal routing and fault tolerance in wireless sensor networks," *Journal of Network and Systems Management*, vol. 32, no. 2, p. 30, 2024.

[25] G. Wen, Y. Cao, and X. Wei, "The data-driven analysis of soil health and crop adaptability: Technologies, impacts, and optimization strategies," *Advances in Resources Research*, vol. 5, no. 1, pp. 350–368, 2025.