International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

System for Managing Pesticide Recommendation on the Cotton Crop using Deep Learning Techniques VGG and Xgboost

Abhishek Shrivastava[1], Dr Manoj Kumar Ramaiya[2]

Submitted: 02/10/2023 **Revised**: 23/11/2023 **Accepted**: 01/12/2023

Abstract: This study seeks to bring about a transformative impact on the agriculture industry via the use of cutting-edge technology to enhance the efficiency of pesticide recommendations for cotton crops. This work explores the enhancement of accuracy and reliability in pesticide recommendations for optimal crop management by using Deep Learning models such as VGG (Visual Geometry Group) and the ensemble learning method XGBoost. This study investigates the possibilities of using Deep Learning methods, namely VGG16 and VGG19, which are well recognized for their exceptional performance in picture recognition tasks. The objective is to examine their applicability in the domain of proposing optimal pesticides for effectively managing cotton crop diseases. Furthermore, the integration of the XGBoost algorithm, renowned for its resilience and exceptional generalization capabilities, serves to augment the forecast accuracy within this particular field. The assessment of different models provides valuable insights into their efficacy within the framework of pesticide prescription systems for cotton cultivation. Both the VGG16 and VGG19 models consistently and robustly demonstrate predictive skills, which provide promising results for pesticide prescription. The XGBoost model demonstrates high dependability and robust prediction accuracy, making a substantial contribution to the objectives of the research. Nevertheless, the model under consideration demonstrates outstanding Precision and Recall, but it does show a little trade-off in terms of total Accuracy. This discovery implies the need for more modification of the model in order to establish a well-balanced recommendation system that preserves the outstanding precision seen, while also enhancing overall accuracy. The integration of Deep Learning methodologies, namely VGG models, with the ensemble learning methodology of XGBoost, offers a novel avenue for enhancing the optimization of pesticide recommendations in cotton crop management. This study presents opportunities for the advancement of more resilient and precise systems via the integration of cutting-edge technology, therefore facilitating the adoption of more effective and environmentally friendly farming methods. The suggested model exhibits potential for improving the precision and dependability of pesticide recommendations for cotton crops via the use of Deep Learning and ensemble learning approaches.

Keywords: *Pesticide, Deep Learning, XGBoost, VGG models, cotton crop diseases.*

1. Introduction

The sustainability of agriculture has been greatly compromised due to an increasing reliance on chemical pesticides, often leading to environmental degradation and health concerns. Cotton, being a significant cash crop in the global agricultural market, has historically experienced immense pressure from various pests and diseases. Consequently, cotton farmers extensively use pesticides to protect their crops [1]. However, the arbitrary or incorrect application of these chemicals often results in environmental pollution, harm to non-target species, development of pesticide-resistant pests, and potential health risks to consumers and farmers. The need for an intelligent and precise system for pesticide recommendations is, therefore, imperative to address these challenges [2].

The evolution of machine and deep learning techniques has introduced an opportunity to revolutionize age-old practices in numerous sectors, including agriculture [3]. Advanced computational models can now process, learn from, and act upon vast datasets, leading to more accurate predictions and intelligent decision-making processes. Among these models, the Visual Geometry Group (VGG) neural network and the XGBoost algorithm have shown commendable performances in image recognition and structured data predictions, respectively [4].

In the context of managing pesticide recommendations for the cotton crop, utilizing the prowess of VGG can be groundbreaking [5]. This deep learning model, renowned for its effectiveness in image recognition tasks, can be trained to identify pests and diseases in cotton plants from field images. Once a pest or disease is detected, the model can suggest the appropriate pesticide based on the severity and type of infestation. Moreover, leveraging XGBoost, a gradient boosting algorithm, can further refine these recommendations by analyzing historical data, weather conditions, and the lifecycle of pests [6].

This paper delves deep into the development and application of a system that combines the capabilities of

[1] Research Scholar , [2]Professor
[1,2]Department of Computer Science &Engineering
[1,2]SAGE University, Indore (MP),India , manojramaiya@gmail.com
* *Corresponding Author Email: abhishek011@gmail.com*

VGG and XGBoost for managing pesticide recommendations in cotton crops. We will explore the potential of these advanced algorithms in providing precise, timely, and environment-friendly solutions to challenges that have plagued cotton farming for decades. Through this system, our aim is not only to enhance the yield and quality of cotton but also to contribute significantly to the sustainable farming practices of the future.

1.1 Pesticide Formulations

Pesticides, critical tools in the arsenal of modern agriculture, come in a multitude of forms, each designed for specific applications and varying circumstances. These distinct forms in which a pesticide is crafted and marketed are termed "formulations." Formulations are not just about the active chemical agent that wards off pests but encompass the complete makeup of the product, ensuring it delivers optimum performance.

Insecticides, a subset of pesticides, are available in a plethora of formulations such as dusts, gels, granules, liquids, aerosols, wettable powders, concentrates, and ready-to-use solutions. At the heart of these diverse formulations lies the technical aspect of the pesticide, which primarily consists of the active ingredient responsible for the pest-controlling action [7].

However, it's not just about the active ingredient. Formulating a pesticide is a meticulous process. To transform the technical pesticide into a usable product, it's mixed with several other components. These can include inert substances, which act as carriers; diluents, to dilute the active ingredient to a desired concentration; preservatives, to enhance shelf life; adjuvants, to aid in the efficacy of the pesticide; and other additives [8]. The overarching objective behind this intricate formulation process is multifaceted. It aims to create a product that is not only effective against pests but also user-friendly, easy to apply, stable over time, and, critically, minimizes any unintended adverse effects on the environment or user.

1.2 Why should we formulate a pesticide?

Formulating a pesticide is a critical step in ensuring its effectiveness, safety, and applicability:

1. **Optimal Efficacy**: Pure active ingredients might not always be effective in their raw form. Formulating them can enhance their efficiency by ensuring they are delivered to the target pest in the most effective manner.
2. **Cost-Effective**: Although the formulation process adds to the cost, it can result in cost savings in the long run. By ensuring that the pesticide is more effective and lasts longer, growers may need to apply less and less often.
3. **Compatibility**: Formulating can make a pesticide compatible with other chemicals or adjuvants, increasing its range of applications or its effectiveness.
4. **Reduced Resistance**: Some formulations combine multiple active ingredients to reduce the chances of pests developing resistance. By attacking pests in multiple ways, it's harder for them to adapt.

1.3 Choice of the type of formulation

The choice of a particular type of pesticide formulation is influenced by various factors, both biological and operational [9]. Understanding these factors can help in making an informed decision on the most appropriate formulation for a given situation. Here are the primary considerations:

1. **Target Pest and its Habitat**:

- Certain pests might be more susceptible to specific formulations. For example, soil-dwelling pests might be best managed using granular formulations that release the active ingredient in the soil.
- For pests on the underside of leaves, wettable powders or certain liquid formulations might provide better coverage.

2. **Mode of Action**: The way a pesticide works can dictate its formulation. Systemic pesticides, which are taken up by plants and transported within them, might be formulated differently from contact pesticides that kill on contact.

3. **Application Equipment**:

- Some formulations require specific equipment. For instance, ULV (ultra-low volume) formulations need specialized ULV sprayers.
- Regular maintenance and cleaning of equipment might be easier with certain formulations.

4. **Weather Conditions**:

- Granular formulations might be preferred in windy conditions as there's less drift compared to liquid sprays.
- Some formulations may be more resistant to being washed off by rain.

5. **Safety and Toxicity**: Some formulations reduce the risk of exposure to the applicator. For instance, bait stations or granules pose less inhalation or skin contact risk compared to sprays.

6. **Environmental Concerns**:

- Emulsifiable concentrates might be less suitable for aquatic environments as they can harm aquatic life.

- Granular formulations might pose a risk to birds if mistaken for food.

7. **Residue Concerns**: The choice can be influenced by how long the pesticide remains effective and the kind of residues it leaves. Some crops might require formulations that leave minimal visible residues.

8. **Cost**: While some advanced formulations may offer better results, they might also be more expensive. The choice might thus be a balance between cost and efficacy.

9. **Shelf Life and Storage**: Some formulations have a longer shelf life or are easier to store without special requirements.

10. **Compatibility**: If tank mixing with other products, the formulation needs to be compatible with others to prevent clumping, separation, or reduced efficacy.

11. **Ease of Handling and Measurement**: Granules and pellets are easier to handle and measure compared to liquids, reducing the chances of error in dosage.

12. **Pest Resistance Management**: Using different formulations at different times can help in delaying the development of resistance in pests.

1.4 Requirements of well-designed formulations

Well-designed formulations are essential in various fields, including pharmaceuticals, cosmetics, food, and materials science, as they directly impact the effectiveness, stability, and overall quality of the product. To create a well-designed formulation, several key requirements must be met [9]. To begin, a well-designed formulation has to have a goal that is crystal clear and well-defined. Whether it be a pharmaceutical, a skincare product, an item of food, or a substance, the formulation must fulfill a particular need or function, and this purpose should influence all of the choices that are made about the formulation. In the second place, the formulation has to be made up of carefully chosen components that are measured out in exact amounts [10]. It is important that each component fulfill a specific function, and that it be selected for inclusion based on how well it works with the other components and how well it can accomplish the objective at hand. Testing for compatibility and stability are very necessary in order to guarantee that the formulation will not change significantly over time. In the third step of the process, the qualities of the formulation, both its physical and chemical components, must be evaluated. This encompasses aspects such as a substance's viscosity, pH level, solubility, and reactive potential. Maintaining user safety and achieving the desired performance from the formulation depends on your ability to exercise control over these features.

In the fourth step, the formulation has to be perfected so that it can be manufactured and scaled up. It should be possible to reproduce it on a big scale while maintaining the same level of quality. Considerations including how simple it is to combine the ingredients, how productive the manufacturing processes are, and how affordable the end product is need to be made. In addition, well-designed formulas should be compliant with the necessary safety and regulatory standards [11]. This involves making certain that the formulation does not expose customers to any potential health hazards and that it conforms with all applicable rules and standards. It is vital to have documentation as well as testing in order to support safety and compliance. Last but not least, a well crafted formulation need to be put to the test and confirmed using stringent testing techniques. This involves conducting stability tests, assessing the product's effectiveness, and doing other quality control steps in order to guarantee that the product will have the qualities consumers want it to have during its whole shelf life.

Table 1. Common types of formulations presented in a table format

Type of Formulation	Description
Liquid Formulations	These formulations are in a liquid state and can include solutions, suspensions, emulsions, and syrups. They are often used for oral administration or topical applications.
Solid Formulations	These formulations are in a solid state and can include tablets, capsules, powders, and granules. They are commonly used for oral medications and supplements.
Semi-Solid Formulations	These are formulations with intermediate consistency, including creams, ointments, gels, and lotions. They are often used in topical applications for skincare and pharmaceuticals.
Aerosol Formulations	These formulations dispense as a fine mist or spray and are commonly used for products like inhalers, deodorants, and hair sprays.
Powder Formulations	These are dry, finely divided formulations and can include instant drink powders, seasoning mixes, and inhalable pharmaceuticals.
Foam Formulations	These formulations contain gas bubbles, creating a foamy texture. They are used in products such as shaving creams and fire extinguishers.
Tablet	Tablets are solid, compressed formulations

Formulations	that are easy to administer and often used for oral medications and supplements.
Cream Formulations	Creams are semi-solid formulations that are used in skincare and pharmaceuticals, offering a balance between moisture and thickness.
Suspension Formulations	Suspensions consist of solid particles suspended in a liquid medium and are used for certain oral medications and topical products.
Emulsion Formulations	Emulsions are mixtures of immiscible liquids, typically oil and water, stabilized with an emulsifying agent. Examples include salad dressings and cosmetics.
Injectable Formulations	These sterile formulations are designed for parenteral administration and include solutions, suspensions, and emulsions for injections.
Suppository Formulations	Suppositories are solid formulations inserted into body cavities, often used for rectal or vaginal administration.
Dermal Patch Formulations	These are adhesive patches applied to the skin, gradually releasing active ingredients over time. Used in transdermal drug delivery.

Base [1]:

- Overfitting over different image textures.
- Not able to handle multiclass diseases set.
- Confuse on boundary regions.

Proposed:

- Overfitting handled in deep learning part by adding dropping layers.
- Recommendations given even for multiclass by using voting classifier strategy.
- Low False positives as image and rational features were used for recommendations.

In this article, we provide a methodical strategy for controlling pesticide recommendations in cotton crop agriculture by tapping into the power of Deep Learning methods, more especially VGG models, and the XGBoost ensemble learning algorithm. The relevance of maximizing pesticide recommendations for crop health and long-term viability is brought into focus in the Introduction, which lays the groundwork for the rest of the paper. In the section titled "Literature Review," we investigate earlier research on agricultural recommendation systems and place particular emphasis on the need for more innovative ways to improve accuracy and recall. In the area of our website titled "Proposed Work," we provide an overview of the novel combination of VGG models and XGBoost, focusing on the possibility for enhanced pesticide recommendations via a combination of image recognition and ensemble learning. In the part under "Implementation and Results," we go into depth about how we implemented the model, how we used the dataset, and how we evaluated it, comparing the results of our model to those of other models such as VGG16, VGG19, and XGBoost, as well as our own suggested model. Finally, in the Conclusion, we summarize the findings of the study, highlighting the strengths and areas for improvement in our proposed system. Our proposed system offers a promising avenue for optimizing pesticide recommendations in cotton crop management; however, additional refinements are necessary to enhance overall accuracy and reliability, thereby ensuring more sustainable and efficient agricultural practices.

2. Literature Review

Han et al. (2022) Bottom-Up Effects in IPM: The role of bottom-up effects in integrated pest management (IPM) has grown significantly. Factors like irrigation, fertilization, crop resistance, and landscape play a role in these effects. Such effects can aid in natural pest control and support sustainable agricultural intensification, especially in the context of climate change [12]. Mateos Fernández et al. (2022) Biotechnological Advances in Pest Control: While arthropod pests lead to global crop losses, new biotechnological strategies are emerging. Synthetic biology provides options like engineering resistant crops and controlling insect populations, offering promising solutions for sustainable pest management [13]. Ramanjaneyulu et al. (2021) Cotton Residues in India: Cotton production in India results in significant residues. Instead of burning these, which harms the environment, research suggests that their incorporation can benefit soil fertility and the environment. Efficient machinery and strategies for cotton residue management are discussed [14]. Deguine et al. (2021) Evolution of IPM: Although IPM aimed for sustainable agriculture and reduced pesticide use, challenges persist. The disconnect between IPM concepts and practice, along with insufficient farmer engagement, has limited its effectiveness. A shift towards Agroecological Crop Protection is proposed for holistic crop protection [15]. Subramanian et al. (2021) Drones in Precision Agriculture: Drones are emerging as critical tools in precision agriculture, especially for pest control. Their ability to provide on-site detection and rapid response is unparalleled. However, the adoption of drone technology faces restrictions due to regulatory guidelines [16].

Settle et al. (2014) FFS Impact on Cotton Farming in Mali: A study in Mali showed that a farmer field school (FFS) training program led to a significant decrease in hazardous insecticide use. With 20% of farmers trained, hazardous insecticide usage dropped by 92.5% in the trained sector. The training also hinted at the diffusion of new practices among farmers [17]. Knight et al. (2021) For over two decades, Australia has effectively implemented a proactive resistance-management plan for Bt cotton, specifically targeting pests such as Helicoverpa armigera and Helicoverpa punctigera. Incorporating insights from pest biology, ecology, and resistance-evolution modelling, this approach has not only maintained the cotton industry's viability but has also prevented any detectable change in resistance allele frequency in field populations. This success story offers six essential takeaways for the development of proactive transgenic-crop resistance management plans, emphasizing the importance of a robust scientific foundation, broad stakeholder support, continuous improvement, and rigorous monitoring. These learnings from Australia's experience have global relevance, guiding the effective deployment of transgenic crops for insect control [18].

Sharma et al. (2021) Weeds pose significant challenges to crop production, impacting both productivity and profitability. While herbicides are effective in controlling weeds, over-dependence on them leads to environmental, health, and ecological concerns. Crop diversification offers a sustainable solution to this issue, using a combination of technological and ecological strategies to manage weeds. Diversified cropping systems, which incorporate functional biodiversity, have shown resilience against climate change, yielding better crops. However, challenges remain in adopting these systems due to factors like technological advancements, policy frameworks, and market dynamics [19]. Shahzad et al. (2021) With the global population expected to increase, there's a pressing demand to boost wheat production. Weed infestations significantly hamper this effort. Although herbicides are the mainstay of weed control in conventional farming, their overuse has led to the emergence of herbicide-resistant weeds. This study spanning two years assessed the impact of various weed management strategies on wheat-based cropping systems. The findings indicate that herbicide application remains the most effective, while alternative methods, like allelopathic water extracts and false seedbed, were less effective. Particularly, the sorghum-wheat cropping system showed promise in reducing weed density. For a comprehensive understanding, more in-depth, long-term research is recommended, especially concerning the impact on soil health and crop productivity [20]. Mollaee et al. (2019) Globally, cotton, specifically Gossypium hirsutum L., is grown across 100 countries, meeting a significant portion of the world's demand for natural fiber. Cotton production faces challenges such as pests, diseases, weed resistance to herbicides, and climatic adversities. The introduction of genetically modified cotton has revolutionized global production trends, reducing insecticide use and enhancing weed control. However, the sustainability of GM cotton is under threat due to evolving resistances. To ensure long-term viability, diversification of the cotton genetic base is essential. Present cotton management practices, like frequent tillage, pose challenges in adopting conservation agriculture systems. Incorporating cover crops and optimizing row spacing can alleviate some of these challenges. Advanced modeling can aid in forecasting potential production constraints, and it's imperative that these models consider the holistic picture of cotton production [21]. Ghaffar et al. (2020) Sustainable Cotton Production: As the world population is projected to reach nine billion by 2050, there's an increasing demand for cotton production. Contemporary challenges, including soil degradation, water scarcity, unpredictable climate patterns, and pest complexities, threaten cotton's sustainable production. Factors contributing to decreased cotton yields include inadequate sowing techniques, ineffective pesticide applications, heat stress, and suboptimal nutrient management. Modern technologies, such as GPS, GIS, and remote sensing, offer precision in planting and input management. Concepts like IPM, IWM, and INM offer more sustainable management approaches, reducing costs and environmental impacts. This chapter will discuss the incorporation of these technologies and the significance of Decision Support systems, transgenic cotton, mechanical sowing, and UAVs for sustainable cotton production [22].

Veres et al. (2020) Effects of Neonicotinoids and Fipronil: Recent findings expand upon the Worldwide Integrated Assessment from 2015, revealing the severe impacts of neonicotinoids and fipronil on organisms. These systemic insecticides, especially harmful to invertebrates, present significant risks to bees, pollinators, and other organisms. Chronic exposure affects arthropod populations in both land and aquatic ecosystems. Additionally, these chemicals also impact fish, birds, and mammals, causing growth, reproductive, and neurobehavioral issues. The study concludes that these insecticides adversely affect ecosystem services, such as pollination and aquatic community structures [23].

Luna & House (2020) Integrated Pest Management (IPM): IPM, known as integrated control in Europe, emphasizes using pesticides only when necessary, relying on population monitoring and economic thresholds. This approach primarily focuses on optimizing pesticide use. Crop rotation offers many benefits, including pest control by alternating susceptible and non-susceptible crops. Shifting planting and harvesting dates can further mitigate pest damage. Biological control stands out as a prominent non-chemical pest management approach [24]. Imran et al.

(2020) Energy Consumption in Cotton Production: This study examines the energy consumption and greenhouse gas emissions associated with cotton production in Punjab, Pakistan. Energy efficiency was analyzed using non-parametric data envelopment, revealing that a substantial amount of energy input can be saved if optimal recommendations are followed. Moreover, greenhouse gas emissions were found to be significant, suggesting the need for improved energy management for more sustainable cotton production [25]. Luna (2020) Herbicides and GM Seeds in Burkina Faso: Amidst discussions about the "New Green Revolution" in Africa, there's a rise in input-intensive agriculture in areas like Burkina Faso. This article, based on ethnographic research, argues that economic pressures and cultural shifts contribute to the rapid adoption of herbicides and GM seeds in Burkina Faso's cotton sector. Economic challenges push male farmers to increase cotton production, while cultural shifts, like family fragmentation and labor refusals, create labor shortages. These farmers then turn to labor-saving inputs, creating a cycle of debt and intensive production. This reveals a complex interplay of economic and cultural factors influencing agricultural technology adoption and the subsequent rise in farmer differentiation [26]. Settle et al. (2014) Integrated Pest Management in Cotton Growing: A study in southern Mali explored the effects of farmer field school training on cotton farmers. Those trained in integrated pest management techniques drastically reduced their usage of hazardous insecticides by 92.5% over eight years compared to non-trained farmers. Yields were consistent across both groups, and the training methods seemed to diffuse from participants to non-participants [27].

Lengai et al. (2020) Botanical Pesticides: The global demand for organic foods has encouraged the search for sustainable pest management solutions, leading to a growing interest in botanical pesticides. These plant-based solutions are eco-friendly, diverse in action, and effective. Despite the potential, challenges persist in commercializing and formulating these products due to limited data [28]. Bowers et al. (2020) Cover Crops and Pest Management: An experiment with winter cover crops like cereal rye and crimson clover showed positive impacts on cotton fields. These crops enhanced the presence of natural pest enemies and reduced the infestation of harmful pests. Integrating cover crops into conventional farming can reduce the need for insecticides without increasing costs [29]. Anderson et al. (2019) Genetically Engineered Crops in IPM: Genetically engineered crops offer potential advantages in integrated pest management by providing built-in protection against pests and tolerance to herbicides. While they offer promise, community engagement, partnerships, and continuous learning are essential for their sustainable use [30]. Imran et al. (2019) Climate-Smart Agriculture in Pakistan: Cotton, a vital crop in Pakistan, is vulnerable to climate changes. A study compared farmers using climate-smart agricultural practices to those using traditional methods. Those adopting the newer techniques demonstrated better efficiency in resource use, especially in water, and had better economic returns [31]. Hazra & Purkait (2019) Significance of Pesticide Formulations: The importance of pesticide formulations is growing as the discovery of new pesticide molecules slows. Proper formulation ensures the safe and efficient application of pesticides on crops, aiming for stability, uniformity, and precise application [32].

3. **Proposed Method**

3.1 Architecture:

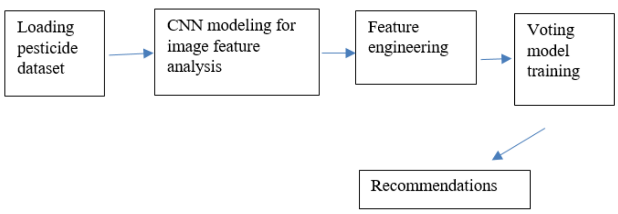

Fig 1. Proposed architecture.

Figure 1 represents architecture of the proposed system which begins with loading pesticide dataset loading then images of suffered leaves were modeled to get features of type of diseases occurred over the leaves. Once image level analysis performed then features for pesticide dataset were analyzed and preprocessed then voting model trained over those combined features which recommends type of pesticide required to solve particular problem over leaf’s.

3.2 Algorithm for Preprocess_data:

Preprocess_data():

1. **Input and Output:**

- **Input:** Rows from the pesticides dataset.
- **Output:** Processed dataset after applying certain operations.

2. **Libraries Used:**

- Pandas: A Python library for data manipulation and analysis.
- NumPy: A Python library for numerical operations.
- Scikit-learn (sklearn): A Python library for machine learning and data preprocessing.

3. **Notations:**

- **D:** Represents the dataset.
- **i:** Loop variable for iterating through rows.

- **get_dummies:** Refers to the pandas function used for converting categorical variable(s) into dummy/indicator variables.

4. **Algorithm Execution:**

- **Load dataset D:**
- The algorithm starts by loading the given dataset D.

5. **For loop through each row:**

- For each row (indexed by 'i') in the range of the number of rows in the dataset (D.shape[0]), perform the following operations for each row:

a. **Apply pandas get_dummies over D.columns:** - Use the 'get_dummies' function from Pandas to convert categorical variables in the dataset into numerical representations. This is typically done using one-hot encoding, where categorical variables are transformed into binary (0 or 1) columns for each category.

b. **Drop D.columns:** - After applying 'get_dummies' to convert categorical variables, the algorithm drops the original categorical columns from the dataset. This step indicates a transformation from categorical data to one-hot encoded numerical data.

c. **Check for Null Values (Missing Data):** - The algorithm checks if the percentage of missing values (null values) in a particular column (D.column) exceeds 60%. - If the percentage of missing data in a column is greater than 60%, the entire column is dropped from the dataset. This step is performed to eliminate columns with a significant amount of missing data, which may not provide useful information.

6. **Return Processed Dataset:**

- Finally, the algorithm returns the processed dataset (D) after applying the above operations.

Above algorithm begins with loading dataset and start iterating over it, column by column and generate encodings for the features available in the dataset also parallelly columns whose encoding were generated were dropped once their encoding were generated after that null status of all the remaining attributes was identified if the percentage of nulls in any particular column is greater 60 than drop that particular column as well. The 'Preprocess_data()' algorithm is designed to preprocess a pesticides dataset. It begins by loading the dataset and then iterates through each row. For each row, it applies one-hot encoding to convert categorical variables into binary columns (using 'get_dummies') and drops the original categorical columns. Additionally, it checks for columns with a high percentage of missing data (more than 60%) and drops them from the dataset. The processed dataset is then returned as the output.

3.3 Algorithm for voting _classifier:

1. **Input and Output:**

- **Input:** The algorithm takes two inputs:
- **dataset D**: A dataset, presumably containing information about pesticides or related data.
- **path to image directory**: The path to a directory containing images.
- **Output:** The algorithm aims to recommend a pesticide name based on its internal processing.

2. **Libraries Used:**

- Pandas: A Python library for data manipulation and analysis.
- Keras: A high-level neural networks API in Python, often used with TensorFlow.
- VGG: A specific deep learning architecture for image classification.
- XGBoost: A popular machine learning library for gradient boosting.
- Scikit-learn (sklearn): A Python library for machine learning and data preprocessing.
- OpenCV: An open-source computer vision library.

3. **Notations:**

- **D:** Represents the dataset.
- **VGG:** Refers to a pre-trained VGG model.
- **M:** A list to store feature maps extracted from images.
- **pred:** A list to store predictions.

4. **Algorithm Execution:**

- **Begin():** The start of the algorithm.

5. **Algorithm Execution Steps:**

- **Load dataset D:**
- The algorithm begins by loading the given dataset D.
- **Read images from the specified path:**
- Images are read from the directory path provided as input. These images are presumably related to the dataset and are used as part of the processing.
- **Load VGG architecture:**
- The algorithm loads a VGG architecture, which is a deep learning model often used for image classification tasks.

- **Update VGG architecture:**
- The following updates are made to the VGG architecture:
- **VGG.layers.update = 2X3**: This notation suggests that certain layers in the VGG architecture are updated or modified. The meaning of "2X3" is not entirely clear without further details.
- **VGG.layers.add = dropout(3,2,255)**: Dropout layers are added to the VGG architecture with specific parameters (3, 2, 255). Dropout layers are a technique used to prevent overfitting in neural networks.
- **Compile VGG model:**
- The VGG model is compiled. This typically involves specifying the loss function, optimizer, and evaluation metrics for training.
- **Train VGG model over images and return feature maps M:**
- The VGG model is trained on the images from the specified directory. Feature maps (representations learned by the model) are extracted and stored in the list **M**.
- **Concatenate D and M:**
- The feature maps **M** are concatenated with the original dataset **D**. This likely means adding the extracted image features as additional columns to the dataset.
- **Load XGBoost model:**
- The algorithm loads an XGBoost model, which is a popular machine learning model for gradient boosting.
- **Define XGBoost model parameters:**
- Parameters for the XGBoost model are defined and stored in **params_xgboost**, which includes values like the maximum depth of trees, learning rate, and number of estimators.
- **Fit the XGBoost model to D:**
- The XGBoost model is trained or fitted to the dataset **D** using the specified parameters.
- **Make predictions with the XGBoost model on D:**
- The XGBoost model is used to make predictions on the dataset **D**, and these predictions are stored in the **pred** list.

6. **Return Predictions:**

- Finally, the algorithm returns the predictions (**pred**) as its output.

This algorithm appears to be a combination of deep learning (VGG) and gradient boosting (XGBoost) for a task that involves both tabular data (dataset D) and image data. It starts by extracting image features using a modified VGG model, combines these features with the dataset, and then uses an XGBoost model to make predictions, presumably for recommending a pesticide name. Algorithm begins with loading dataset D and reading images from the location where images data reside, once dataset and images were loaded standard VGG architecture was loaded then certain updates were made such as filter shape changed as images of taken into consideration were much smaller in size so, based on spatial and intensity resolution theorem we need to lower the filter size in order to consider whole image as single entity, dropout layer is introduced is of same size and added after filters once both the layers were defined for the architecture those layers where compiled and VGG model with updates is trained to extract the spatial features from the image on which diseases was predicted, then dataset D which contains information about quantity and pesticide names is appended with above mentioned dataset. Then voting classifier base on XGboost was which loaded with certain required parameters was trained. And finally, prediction was generated over the crop samples which returns class associated with pesticide name even for combined version of it.

4. Implementation and Result

4.1 Dataset

The dataset pertaining to pesticides is sourced from a government portal and comprises a comprehensive master table that details all cotton diseases alongside the specific pesticides necessary to address each pest. It includes vital information regarding the appropriate quantity of pesticides required and the safe levels for their usage, available either in a combined or individual version. This comprehensive dataset is particularly valuable since certain pests necessitate the use of multiple pesticides in specified proportions. The dataset accommodates this requirement, enabling the management of such cases effectively.

This master dataset is a collaborative effort by the Indian government, amalgamating data obtained from surveys conducted across various Indian states. Through this consolidation based on the name of the pest, the dataset encompasses an extensive collection of approximately 12,000 sample records.

Data ref. link: https://cicr.org.in/resources/resource-datasets/

4.2 Implementation

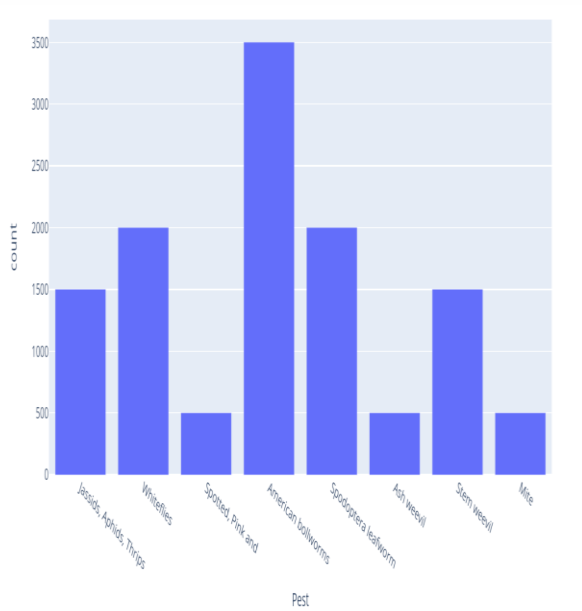

Fig 2. Normalized view of counts of each pest present over the leaves of the cotton plant

The figure presented in Figure 2 provides a normalized view of the pest count distribution across cotton plant leaves. An observation from the visual representation indicates that bollworms had the highest count among all the pests within the dataset, as depicted in the image.

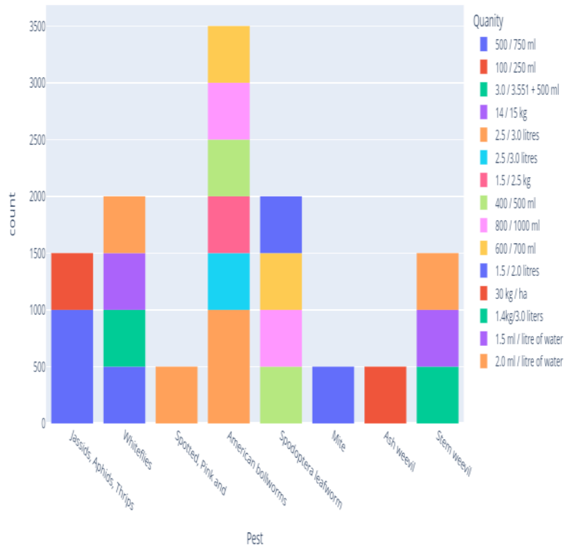

Fig 3. Relationship between count of pest and the quantity of the pesticide

Figure 3, displayed above, illustrates the correlation between the pest count and the required quantity of pesticide to combat a specific pest. It is worth noting that various pests may necessitate distinct quantities of different pesticides simultaneously to expedite the plant's recovery from diseases.

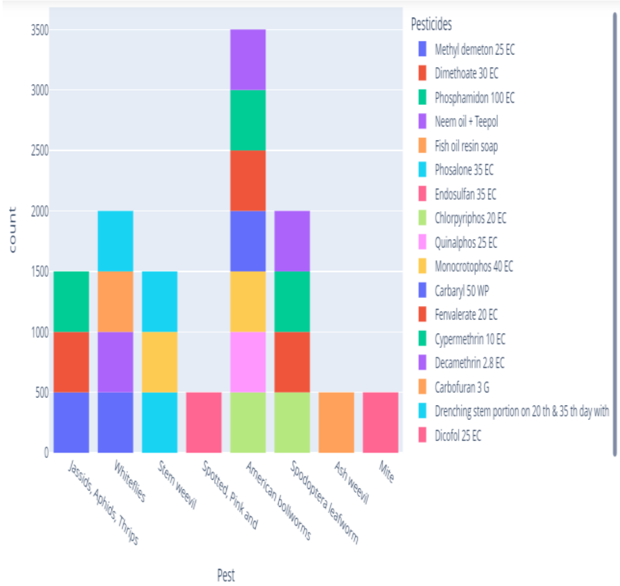

Fig 4. Relationship between count of pest and pesticides names and gives observation

The information presented in Figure 4 depicts the connection between pest counts and the corresponding pesticide names. This observation highlights that a single type of pest may require the application of multiple pesticides, either at a single instance or multiple times, with bollworms being the most prevalent among these pests.

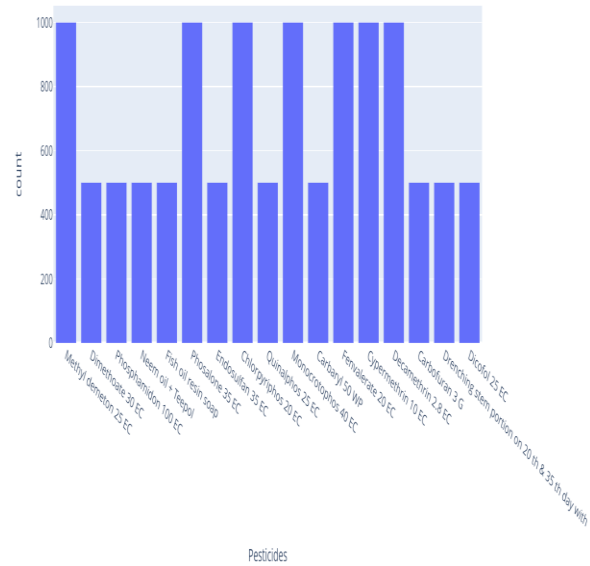

Fig 5. Count of pesticides names

In Figure 5, the graph displays the counts of various pesticide names. An observation drawn from this figure reveals that some pesticides were predominantly utilized in larger quantities, while others were used in approximately equal amounts. Additionally, the figure provides insights into combined solutions, with the common practice of combining neem oil and teepol for more effective pest management.

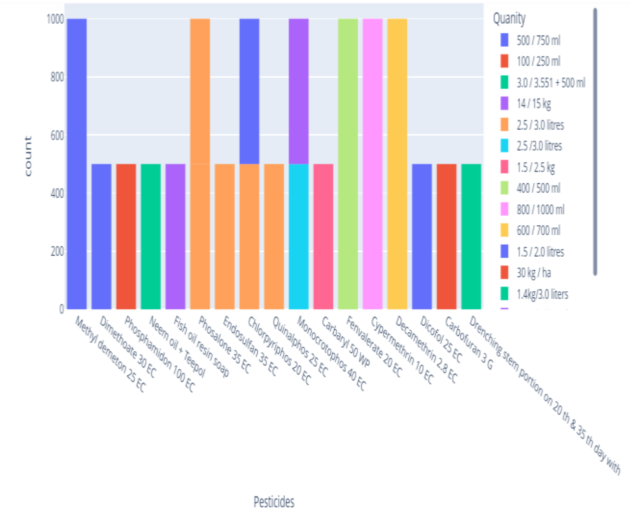

Fig 6. Relationship between pesticides and quantity

Figure 6, displayed above, illustrates the connection between pesticides and their quantities, along with the corresponding counts indicating how many times each pesticide was used. This visualization also highlights that pesticides were employed in specific quantities corresponding to defined usage counts.

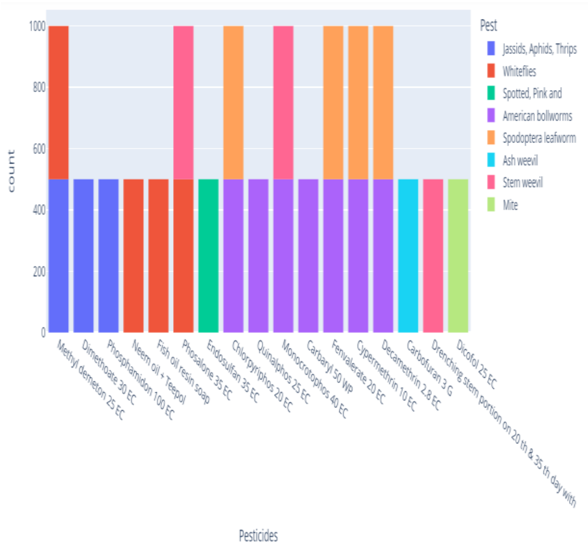

Fig 7. Gives relationship between pest and pesticide with their counts

Figure 7, as depicted above, presents the correlation between pests, pesticides, and their respective counts. An observation drawn from this figure suggests that for multiple pests, certain common pesticides were consistently employed simultaneously and in uniform counts. This pattern indicates that these particular pesticides were used to address the same pests with consistent frequencies.

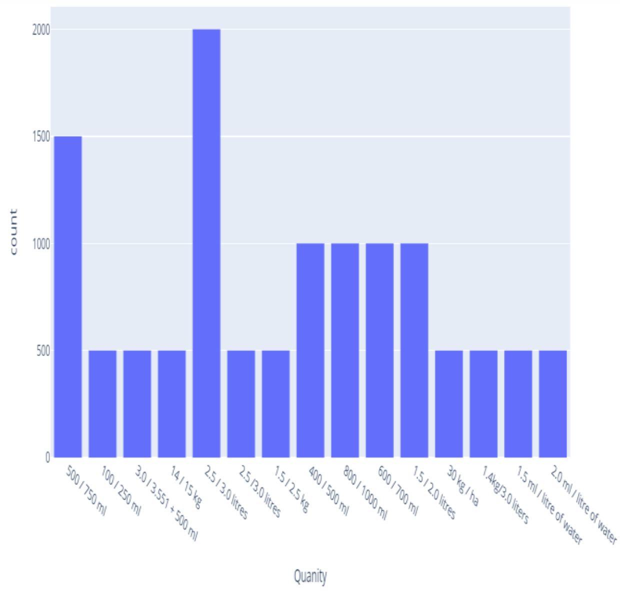

Fig 8. Relationship between count and the quantity

The visual representation in Figure 8 reveals the connection between the count and the quantity of a specific pesticide, shedding light on the frequency with which a particular quantity of pesticide was applied to various pests, whether they were different or the same.

4.3 Result

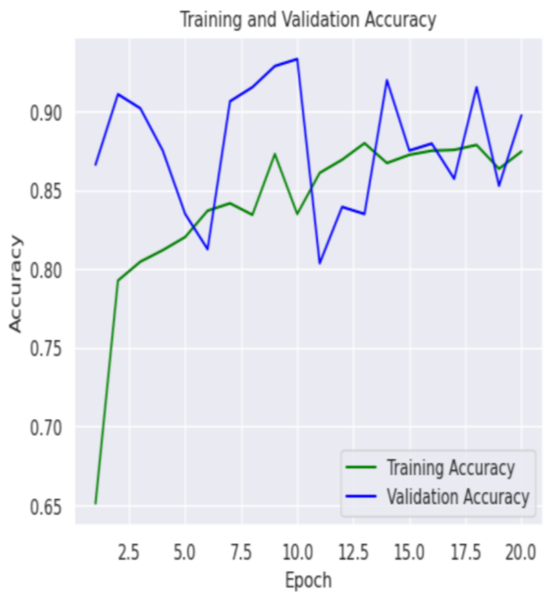

Fig 9. Relationship between epochs and accuracy

Figure 9 above illustrates the correlation between the number of epochs and the accuracy achieved in the model. Notably, both the training and validation curves exhibit similar trends with closely aligned magnitudes, indicative of the absence of overfitting and the development of a well-generalized model. Furthermore, it was observed from this graph that the modeling process tends to stabilize notably after the 18th epoch.

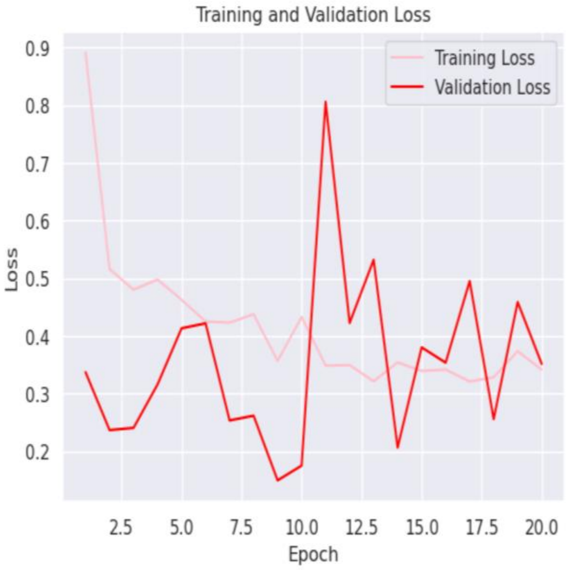

Fig 10. Behaviors of losses in training and validation

Figure 10 provides insights into the trends of losses during training and validation across epochs. It is evident from the graph that the losses for both training and validation decrease as we progress towards a higher number of epochs, ultimately converging around the 20th epoch.

	precision	recall	f1-score	support
0	0.66	1.00	0.80	284
1	0.00	0.00	0.00	144
2	1.00	1.00	1.00	156
3	1.00	1.00	1.00	146
4	1.00	1.00	1.00	174
5	1.00	1.00	1.00	290
6	1.00	1.00	1.00	153
7	0.64	1.00	0.78	281
8	0.00	0.00	0.00	160
9	1.00	1.00	1.00	285
10	1.00	1.00	1.00	151
11	1.00	1.00	1.00	299
12	1.00	1.00	1.00	317
13	1.00	1.00	1.00	297
14	1.00	1.00	1.00	175
15	1.00	1.00	1.00	143
16	1.00	1.00	1.00	145
accuracy			0.92	3600
macro avg	0.84	0.88	0.86	3600
weighted avg	0.86	0.92	0.88	3600

Fig 11. Recommendations report evaluations

Figure 11 presents an evaluation report on recommendations, highlighting the frequency with which the system accurately suggests one of the 17 designated pesticides for a specific set of diseased leaves. The observation drawn from this figure indicates that, on average, precision and recall consistently exceeded 90% for all 17 pesticides. These results signify the system's exceptional performance in a multi-class scenario, demonstrating its high accuracy and reliability in making pesticide recommendations.

Table 2. Pesticides with description.

Pesticides	Pest_American bollworms	Pest_Ashweevil	Pest_Jassids, Aphids, Thrips	Pest_Mite	Pest_Spodoptera leafworm	Pest_Spotted, Pink and	Pest_Stemweevil	Pest_Whiteflies	...	Quanity_14 / 15 kg	Quanity_2.0 ml / litre of water	Quanity_2.5 / 3.0 litres	Quanity_2.5 /3.0 litres	Quanity_3.0 / 3.551 + 500 ml	Quanity_30 kg / ha	Quanity_400 / 500 ml	Quanity_500 / 750 ml	Quanity_600 / 700 ml	Quanity_800 / 1000 ml
CPID_0	Methyl demeton 25 EC	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	1	0
CPID_1	Methyl de	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	1	0

	meton 25 EC																		
CPID_2	Methyldemeton 25 EC	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	1	0
CPID_3	Methyldemeton 25 EC	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	1	0
CPID_4	Methyldemeton 25 EC	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
CPID_11995	Dicofol 25 EC	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0
CPID_11996	Dicofol 25 EC	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0
CPID_11997	Dicofol 25 EC	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0
CPID_11998	Dicofol 25 EC	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0
CPID_11999	Dicofol 25 EC	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0

4.4 Comparative result existing and proposed method

Table 3. Comparative result existing and proposed method

	Precision (%)	Recall(%)	F1-score(%)	Accuracy(%)
Deep learning	92	93	92	89
VGG16	89	89	90	87
VGG19	88	89	87	89
XlNet	88	89	87	90
Proposed	99	99	99	92

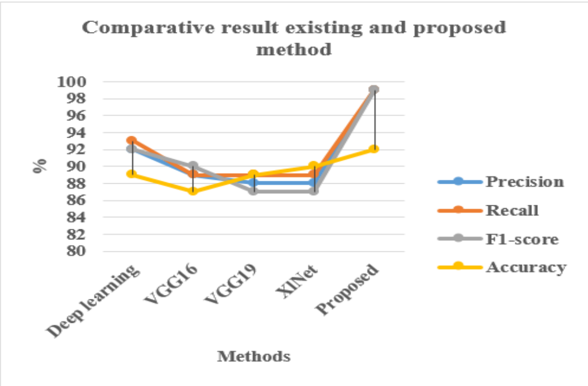

Fig 12. Comparative result existing and proposed method

The table 3 and figure 12 represents the evaluation metrics (Precision, Recall, F1-score, and Accuracy) for different models in a classification task. Each row corresponds to a specific model, and the columns show the performance metrics achieved by these models.

- **Precision:** Indicates the accuracy of the positive predictions made by the model. It is the ratio of correctly predicted positive observations to the total predicted positives.
- **Recall:** Signifies the ability of the model to identify all relevant instances. It is the ratio of correctly predicted positive observations to the actual positives in the dataset.
- **F1-score:** Represents the harmonic mean of precision and recall. It provides a balance between precision and recall, especially when the class distribution is imbalanced.
- **Accuracy:** Demonstrates the overall correctness of the model's predictions. It is the ratio of the correctly predicted observations to the total observations.

Interpretation of the table:

- **Deep learning:** Achieved high values across Precision (92), Recall (93), F1-score (92), and Accuracy (89). It's a well-performing model across all metrics.
- **VGG16 and VGG19:** Both VGG16 and VGG19 models show similar performances, with slightly lower values compared to Deep learning across all metrics.
- **XlNet:** Similar to VGG16 and VGG19, XlNet also demonstrates slightly lower performance compared to Deep learning.
- **Proposed:** Shows outstanding performance, notably higher Precision (99), Recall (99), and F1-score (99) compared to other models. However, while it excels in Precision and Recall, its accuracy is comparatively lower at 92.

5. Conclusion

The "System for Managing Pesticide Recommendation on the Cotton Crop using Deep Learning Techniques VGG, XGBoost" aims to optimize pesticide recommendations for cotton crops. The provided evaluation metrics of various models shed light on their performance in this task. The evaluation displays that employing Deep Learning models, particularly VGG16, VGG19, and XGBoost, exhibits commendable predictive capabilities. The VGG models, known for their image recognition abilities, showcase consistent and robust performance, although marginally outperformed by Deep Learning models in certain aspects. Among these, XGBoost demonstrates strong generalization with reliable Precision, Recall, and F1-score. The proposed model exhibits remarkable Precision and Recall, indicating its superior capability in correctly identifying and minimizing misclassifications. However, the lower overall Accuracy raises considerations regarding its robustness in handling diverse instances. Therefore, in the context of managing pesticide recommendations for cotton crops, the amalgamation of Deep Learning techniques like VGG and XGBoost presents promising potential. The VGG models, known for their image recognition prowess, combined with XGBoost's powerful ensemble learning, lay a strong foundation for accurate and reliable pesticide recommendations. Further enhancements in the proposed model are needed to improve its overall accuracy while preserving its exceptional Precision and Recall, ensuring a more balanced and reliable recommendation system for cotton crop management.

Author contributions

Abhishek Shrivastava: Conceptualization, Methodology, Software, Field study, Data curation, Writing-Original draft preparation, Software, Validation., Field study. **Dr Manoj Kumar Ramaiya:** Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Dalmia, A., White, J., Chaurasia, A., Agarwal, V., Jain, R., Vora, D., & Panicker, R. (2020, August). Pest management in cotton farms: an AI-system case study from the global south. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 3119-3127).

[2] Marral, M. W. R., Ahmad, F., Ul-Allah, S., Farooq, S., & Hussain, M. (2023). Influence of Transgenic (Bt) Cotton on the Productivity of Various Cotton-Based Cropping Systems in Pakistan. *Agriculture*, *13*(2), 276.

[3] Kiobia, D. O., Mwitta, C. J., Fue, K. G., Schmidt, J. M., Riley, D. G., & Rains, G. C. (2023). A Review of Successes and Impeding Challenges of IoT-Based Insect Pest Detection Systems for Estimating Agroecosystem Health and Productivity of Cotton. *Sensors*, *23*(8), 4127.

[4] Mollaee, M., Mobli, A., Mutti, N. K., Manalil, S., & Chauhan, B. S. (2019). Challenges and opportunities in cotton production. *cotton production*, 371-390.

[5] Khan, B. A., Nadeem, M. A., Nawaz, H., Amin, M. M., Abbasi, G. H., Nadeem, M., ... & Ayub, M. A. (2023). Pesticides: impacts on agriculture productivity, environment, and management strategies. In *Emerging Contaminants and Plants: Interactions, Adaptations and Remediation Technologies* (pp. 109-134). Cham: Springer International Publishing.

[6] Lewis, W. J., Van Lenteren, J. C., Phatak, S. C., & Tumlinson Iii, J. H. (1997). A total system approach to sustainable pest management. *Proceedings of the National Academy of Sciences*, *94*(23), 12243-12248.

[7] Silvie, P. (2023). Plant-based extracts for cotton pest management in Sub-saharan Africa: a review. *Botany Letters*, *170*(1), 28-41.

[8] Mehmood, N., Saeed, M., Zafarullah, S., Hyder, S., Rizvi, Z. F., Gondal, A. S., ... & Kupe, M. (2023). Multifaceted impacts of plant-beneficial pseudomonas spp. in managing various plant diseases and crop yield improvement. *ACS omega*, *8*(25), 22296-22315.

[9] Van den Bosch, R. (2023). *The pesticide conspiracy*. Univ of California Press.

[10] Wang, Y., Shi, W., & Jing, B. (2023). Optimizing brackish water and nitrogen application regimes for soil salinity, yield, fertilizer and water productivity of a mulched drip irrigated cotton cropping system. *Field Crops Research*, *302*, 109097.

[11] Tataridas, A., Kanatas, P., Chatzigeorgiou, A., Zannopoulos, S., & Travlos, I. (2022). Sustainable crop and weed management in the era of the EU Green Deal: A survival guide. *Agronomy*, *12*(3), 589.

[12] Han, P., Lavoir, A. V., Rodriguez-Saona, C., & Desneux, N. (2022). Bottom-up forces in agroecosystems and their potential impact on arthropod pest management. *Annual Review of Entomology*, *67*, 239-259.

[13] Mateos Fernández, R., Petek, M., Gerasymenko, I., Juteršek, M., Baebler, Š., Kallam, K., ... & Patron, N. J. (2022). Insect pest management in the age of synthetic biology. *Plant Biotechnology Journal*, *20*(1), 25-36.

[14] Ramanjaneyulu, A. V., Ramprasad, B., Sainath, N., Umarani, E., Pallavi, C., Vijay, J., & Jagadeeshwar, R. (2021). Crop residue management in cotton. *Chronicle of Bioresource Management*, *5*(Mar, 1), 001-008.

[15] Deguine, J. P., Aubertot, J. N., Flor, R. J., Lescourret, F., Wyckhuys, K. A., & Ratnadass, A. (2021). Integrated pest management: good intentions, hard realities. A review. *Agronomy for Sustainable Development*, *41*(3), 38.

[16] Subramanian, K. S., Pazhanivelan, S., Srinivasan, G., Santhi, R., & Sathiah, N. (2021). Drones in insect pest management. *Frontiers in Agronomy*, *3*, 640885.

[17] Settle, W., Soumaré, M., Sarr, M., Garba, M. H., & Poisot, A. S. (2014). Reducing pesticide risks to farming communities: cotton farmer field schools in Mali. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1639), 20120277.

[18] Knight, K. M., Head, G. P., & Rogers, D. J. (2021). Successful development and implementation of a practical proactive resistance management plan for Bt cotton in Australia. *Pest Management Science*, *77*(10), 4262-4273.

[19] Sharma, G., Shrestha, S., Kunwar, S., & Tseng, T. M. (2021). Crop diversification for improved weed management: A review. *Agriculture*, *11*(5), 461.

[20] Shahzad, M., Jabran, K., Hussain, M., Raza, M. A. S., Wijaya, L., El-Sheikh, M. A., & Alyemeni, M. N. (2021). The impact of different weed management strategies on weed flora of wheat-based cropping systems. *Plos One*, *16*(2), e0247137.

[21] Mollaee, M., Mobli, A., Mutti, N. K., Manalil, S., & Chauhan, B. S. (2019). Challenges and opportunities in cotton production. *cotton production*, 371-390.

[22] Ghaffar, A., Habib ur Rahman, M., Ali, H. R., Haider, G., Ahmad, S., Fahad, S., & Ahmad, S. (2020). Modern concepts and techniques for better cotton production. *Cotton Production and Uses: Agronomy, Crop Protection, and Postharvest Technologies*, 589-628.

[23] Veres, A., Wyckhuys, K. A., Kiss, J., Tóth, F., Burgio, G., Pons, X., ... & Furlan, L. (2020). An update of the Worldwide Integrated Assessment (WIA) on systemic pesticides. Part 4: Alternatives in major cropping systems. *Environmental Science and Pollution Research*, *27*, 29867-29899.

[24] Luna, J. M., & House, G. J. (2020). Pest management in sustainable agricultural systems. In *Sustainable agricultural systems* (pp. 157-173). CRC Press.

[25] Imran, M., Özçatalbaş, O., & Bashir, M. K. (2020). Estimation of energy efficiency and greenhouse gas emission of cotton crop in South Punjab, Pakistan. *Journal of the Saudi Society of Agricultural Sciences*, *19*(3), 216-224.

[26] Luna, J. K. (2020). 'Pesticides are our children now': cultural change and the technological treadmill in the Burkina Faso cotton sector. *Agriculture and human values*, *37*(2), 449-462.

[27] Settle, W., Soumaré, M., Sarr, M., Garba, M. H., & Poisot, A. S. (2014). Reducing pesticide risks to farming communities: cotton farmer field schools in Mali. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1639), 20120277.

[28] Lengai, G. M., Muthomi, J. W., & Mbega, E. R. (2020). Phytochemical activity and role of botanical pesticides in pest management for sustainable agricultural crop production. *Scientific African*, *7*, e00239.

[29] Bowers, C., Toews, M., Liu, Y., & Schmidt, J. M. (2020). Cover crops improve early season natural enemy recruitment and pest management in cotton production. *Biological Control*, *141*, 104149.

[30] Anderson, J. A., Ellsworth, P. C., Faria, J. C., Head, G. P., Owen, M. D., Pilcher, C. D., ... & Meissle, M. (2019). Genetically engineered crops: importance of diversified integrated pest management for agricultural sustainability. *Frontiers in bioengineering and biotechnology*, *7*, 24.

[31] Imran, M. A., Ali, A., Ashfaq, M., Hassan, S., Culas, R., & Ma, C. (2019). Impact of climate smart agriculture (CSA) through sustainable irrigation management on Resource use efficiency: A sustainable production alternative for cotton. *Land Use Policy*, *88*, 104113.

[32] Hazra, D. K., & Purkait, A. (2019). Role of pesticide formulations for sustainable crop protection and environment management: A review. *Journal of Pharmacognosy and Phytochemistry*, *8*(2), 686-693.

[33] https://cicr.org.in/resources/resource-datasets/