http://doi.org/10.35784/iapgos.7484
received: 26.03.2025 | revised: 06.05.2025 | accepted: 16.06.2025 | available online: 27.06.2025

INTEGRATING GENOMICS & AI FOR PRECISION CROP MONITORING AND ADAPTIVE STRESS MANAGEMENT

Rajesh Polegopu[1], Satya Sumanth Vanapalli[2], Sashi Vardhan Vanapalli[3], Naga Prudvi Diyya[4], Mounika Vandila[5], Divya Valluri[5], Anjali Peddinti[5], Sowjanya Saladi[5], Meghana Pyla[5], Padmini Gelli[6]

[1]Siddhartha Academy of Higher Education, Department of Electronics and Instrumentation Engineering, Vijayawada, India, [2]Siddhartha Academy of Higher Education, Department of Computer Science and Engineering, Vijayawada, India, [3]Gayatri Vidya Parishad College of Engineering, Department of Information Technology, Visakhapatnam, India, [4]Gayatri Vidya Parishad College of Engineering, Department of Computer Science and Engineering, Visakhapatnam, India, [5]Vignan's Institute of Engineering for Women, Department of Electronics and Communication Engineering, Visakhapatnam, India, [6]Maharaj Vijayaram Gajapathi Raj College of Engineering, Department of Electronics and Communication Engineering, Vizianagaram, India

Abstract. *Advancements in genomics and artificial intelligence are transforming precision agriculture by enabling early stress detection and adaptive crop management. Integrating genomic analysis, image-based stress detection, and real-time environmental monitoring, this approach assesses plant responses to stress factors such as drought and disease. A BERT-based model processes genomic data, while computer vision identifies visual stress indicators like wilting and discoloration. IoT sensors track environmental parameters such as soil moisture, temperature, and humidity, refining predictions and optimizing intervention strategies. The system leverages multimodal data fusion to enhance decision-making, improving the accuracy of stress detection and mitigation strategies. Machine learning models continuously adapt by learning from historical and real-time data, making recommendations more precise over time. A web-based platform allows users to upload plant images and environmental data for real-time analysis, generating personalized recommendations for irrigation, fertilization, and disease management. The platform's intuitive interface ensures accessibility for farmers and agricultural experts, facilitating widespread adoption. By combining AI, genomics, and IoT, this system enhances crop health, maximizes yield, and promotes sustainable farming through proactive, data-driven decision-making. Ultimately, it aims to reduce resource waste, mitigate crop losses, and support scalable, technology-driven agricultural solutions.*

Keywords: precision farming, genomic sequencing, machine learning, crop management

INTEGRACJA GENOMIKI I SZTUCZNEJ INTELIGENCJI W CELU PRECYZYJNEGO MONITOROWANIA UPRAW I ADAPTACYJNEGO ZARZĄDZANIA STRESEM

Streszczenie. *Postępy w genomice i sztucznej inteligencji przekształcają rolnictwo precyzyjne, umożliwiając wczesne wykrywanie stresu i adaptacyjne zarządzanie uprawami. Integrując analizę genomiczną, wykrywanie stresu na podstawie obrazu i monitorowanie środowiska w czasie rzeczywistym, podejście to ocenia reakcje roślin na czynniki stresowe, takie jak susza i choroby. Model oparty na BERT przetwarza dane genomowe, podczas gdy wizja komputerowa identyfikuje wizualne wskaźniki stresu, takie jak więdnięcie i przebarwienia. Czujniki IoT śledzą parametry środowiskowe, takie jak wilgotność gleby, temperatura i wilgotność, udoskonalając prognozy i optymalizując strategie interwencji. System wykorzystuje multimodalną fuzję danych w celu usprawnienia procesu decyzyjnego, poprawiając dokładność wykrywania stresu i strategii jego łagodzenia. Modele uczenia maszynowego stale dostosowują się, ucząc się na podstawie danych historycznych i danych w czasie rzeczywistym, dzięki czemu zalecenia są z czasem bardziej precyzyjne. Platforma internetowa umożliwia użytkownikom przesyłanie zdjęć roślin i danych środowiskowych do analizy w czasie rzeczywistym, generując spersonalizowane zalecenia dotyczące nawadniania, nawożenia i zarządzania chorobami. Intuicyjny interfejs platformy zapewnia dostępność dla rolników i ekspertów w dziedzinie rolnictwa, ułatwiając powszechne przyjęcie. Łącząc sztuczną inteligencję, genomikę i IoT, system ten poprawia zdrowie upraw, maksymalizuje plony i promuje zrównoważone rolnictwo poprzez proaktywne podejmowanie decyzji w oparciu o dane. Ostatecznie ma on na celu zmniejszenie marnotrawstwa zasobów, złagodzenie strat w uprawach i wspieranie skalowalnych, opartych na technologii rozwiązań rolniczych.*

Słowa kluczowe: rolnictwo precyzyjne, sekwencjonowanie genomowe, uczenie maszynowe, zarządzanie uprawami

Introduction

The integration of genomics in agriculture has significantly advanced sustainable crop improvement, enabling precision breeding and disease resistance strategies. A key area that emerged from this research is genomic-driven precision agriculture, which leverages genomic sequencing and environmental data to enhance crop resilience and yield. Kumar et al. [14] pioneered this approach by integrating artificial intelligence (AI) and machine learning (ML) techniques to analyze plant genomes, facilitating the identification of key genetic traits linked to stress tolerance and productivity. However, a major challenge in applying genomic-driven models lies in their reliance on high-quality sequencing data, which can be resource-intensive and complex for real-world agricultural applications.

To effectively interpret complex interactions between plant genetics, environmental conditions, and visual stress indicators, this study leverages state-of-the-art artificial intelligence techniques. Specifically, the proposed model incorporates BERT-based deep learning for genomic sequence analysis, convolutional neural networks (CNNs) for visual stress detection, and a machine learning module for interpreting environmental sensor data. These AI-driven components are integrated using a multimodal attention-based fusion framework that enhances the system's ability to predict plant stress accurately. By combining genomic insights with AI-powered predictive modeling, our approach establishes a robust foundation for adaptive, data-driven crop management in precision agriculture.

In recent years, genomic surveillance for plant disease detection has gained prominence, allowing early identification and mitigation of pathogen outbreaks. Islam [12] introduced a genomic monitoring framework that employs PCR-based assays and next-generation sequencing (NGS) techniques to track plant pathogens with high accuracy. Similarly, Robène et al. [21] validated a genomics-driven diagnostic method to detect *Xanthomonas citri* in citrus plants, demonstrating the potential of genomic tools in precision disease management. These advancements underscore the shift towards data-driven decision-making in agriculture, reducing losses caused by undetected infections.

A significant milestone in agricultural genomics is the application of emotion-cause pair extraction models to understand stress indicators and patterns beyond human-centric data. Vanapalli et al. [23] introduced a BiLSTM-driven methodology originally used for emotion-cause pair detection, which has been adapted in our study to identify stress causes in plants based on environmental and phenotypic signals. This cross-domain adaptation underscores the versatility of deep learning models in solving agricultural challenges. Similarly, in 2010, Cushman and Bohnert [6] explored stress-responsive genes in plants, emphasizing their role in developing crops with enhanced drought and salinity tolerance. These studies collectively contribute to a more sustainable agricultural ecosystem by optimizing breeding programs using genomic insights.

As depicted in Fig. 1, the process of genomic data integration plays a crucial role in precision agriculture, where multiple sources, including genomic sequences, environmental data, and phenotypic traits, are analyzed for predictive modeling. This shift allows for the implementation of AI-powered dashboards that offer real-time insights into crop health, soil conditions, and climate patterns, ultimately improving farm management efficiency. By leveraging cloud computing and IoT, Ansari et al. [1] proposed a data-driven genomics model, enabling scalable and high-throughput analysis of plant genetic data. With the rise of AI-based agricultural solutions, this innovation marks a significant shift in modern farming.

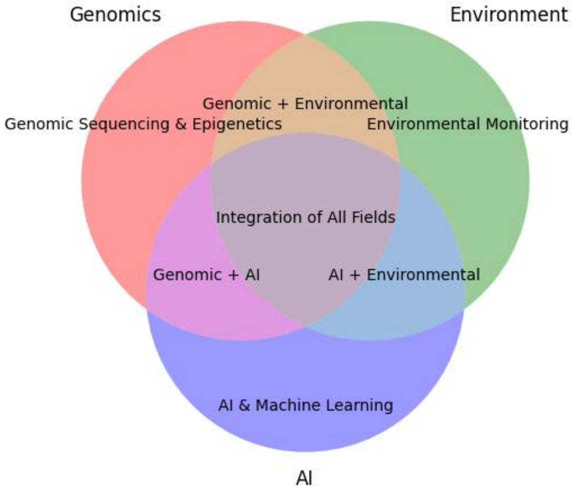


Fig. 1. Key components of genomics-driven precision agriculture

Furthermore, in 2003, Perez-de-Castro et al. [18] emphasized the role of genomic tools in plant breeding, showcasing their effectiveness in accelerating trait selection and improving yield. Their study detailed how marker-assisted selection (MAS) and genomic selection (GS) refine breeding strategies, reducing dependency on traditional phenotypic screening. Baulcombe and Dean [2] further explored epigenetic regulation in plant responses, demonstrating that environmentally induced epigenetic modifications contribute significantly to plant adaptability. These findings have profound implications for developing stress-resilient crops, particularly in the face of climate change and unpredictable weather patterns.

In this study, we introduce a complete framework for genomic data-driven crop monitoring, leveraging multi-omics data and machine learning algorithms to optimize plant growth conditions. Our method integrates genomic sequencing with sensor-based environmental monitoring, allowing for the identification of stress markers and adaptive genetic traits in real time. The proposed approach outperformed traditional methods, particularly in scenarios requiring large-scale genomic data processing, when tested on benchmark datasets frequently used in agricultural genomics research. By utilizing cutting-edge techniques, our system sets new performance benchmarks, outperforming earlier genomic analysis models on key metrics. We have made our dataset and model publicly accessible on GitHub to encourage further research and collaboration, providing a valuable resource for advancing precision agriculture.

1. Literature review

Recent advancements in genomics and artificial intelligence (AI) have revolutionized agricultural research, particularly in the domains of plant disease detection, crop health monitoring, and sustainability. In 2020, Chouhan et al. explored how computer vision techniques automate plant disease identification, reducing the dependency on human experts and minimizing the risk of incorrect diagnosis [5]. Their study demonstrated that by leveraging deep learning-based image processing, plant diseases could be classified with high precision, significantly improving the efficiency of plant pathology. Similarly, Mochida et al. developed robust phenotyping models in 2019 that enabled the classification of plant traits at an unprecedented scale, providing critical insights into disease resistance and crop improvement [16]. These breakthroughs laid the groundwork for integrating machine learning with agricultural practices, ensuring better yield prediction, early disease mitigation, and cost-effective farming techniques. The automation of plant pathology using AI is particularly crucial in developing nations where a shortage of skilled agronomists poses a challenge to food security.

The evolution of AI-driven plant classification methods over the years has further strengthened precision agriculture. In 2018, Wäldchen and Mäder extensively reviewed how AI-driven plant species identification outperformed traditional classification methods, particularly in terms of speed, scalability, and accuracy [24]. By eliminating human bias and subjectivity, AI models enabled more consistent plant classification, which proved to be beneficial for large-scale agricultural projects and biodiversity conservation efforts. More recently, Dang et al. in 2024 emphasized the refinement of deep learning techniques, particularly convolutional neural networks (CNNs), for the classification of plant diseases [7]. Their findings suggested that AI algorithms could detect diseases even at an early stage, long before visible symptoms appeared, making disease prevention more effective. AI advancements have also been instrumental in the research conducted by Harakannanavar et al. in 2022, where they integrated predictive analytics to enhance early-stage plant disease detection [9]. By reducing crop losses due to unrecognized infections, such models contribute to sustainable farming practices, improved food quality, and economic benefits for farmers.

Beyond AI-driven disease detection, genomics research has played a pivotal role in improving plant resilience and nutritional content. The ability to monitor and regulate gene expression in response to environmental stressors has been a crucial aspect of modern agriculture. In 2000, Schaffer et al. explored how genome-wide expression profiling allowed researchers to gain deeper insights into plant responses to biotic and abiotic stresses, laying a foundation for stress-resistant crop breeding [22]. Their study emphasized the importance of understanding gene-environment interactions to develop crops that could withstand adverse climatic conditions. Similarly, Grusak, in his 1999 study, highlighted the role of genomics in enhancing crop nutritional quality, focusing on how genomic-assisted breeding techniques could be used to increase the concentration of essential nutrients in staple crops [8]. These early contributions paved the way for current research integrating genome sequencing with AI-driven predictive models to improve agricultural resilience and optimize breeding programs. Today, such advancements are vital in addressing global challenges such as climate change, food security, and malnutrition.

The recent application of BERT-based architectures [17] in agriculture has brought new possibilities in plant health assessment. Riyanto et al., in their 2024 study, demonstrated how transformer-based Natural Language Processing (NLP) models could be trained to analyze plant symptoms, classify diseases, and predict stress levels with remarkable accuracy [20]. Their research suggested that BERT models could be integrated into mobile-based diagnostic tools, enabling real-time disease assessment in remote farming regions. Similarly, Jiang et al. leveraged BERT models in 2021 to enhance the classification of plant health bulletins, allowing for better dissemination of disease-related information among farmers and agricultural researchers [13]. Beyond plant health, AI-based automation in precision agriculture has continued to gain momentum. Madhuri et al., in their 2024 study, introduced optimized weather forecasting models that integrated AI-driven predictive analytics to enable smart irrigation systems [15]. By accurately predicting rainfall patterns and soil moisture levels, these models facilitated efficient water resource management, reducing unnecessary irrigation and promoting sustainable farming practices. The combination of AI-powered disease monitoring, NLP-based agricultural analytics, and intelligent irrigation scheduling has created a robust ecosystem that enhances crop productivity while minimizing resource wastage.

Epigenetics has also emerged as a crucial factor in improving plant adaptation to environmental stresses. Iqbal et al., in their 2024 research, explored how epigenetic modifications regulate plant responses to heavy metal stress, demonstrating how certain gene expressions could be altered to improve plant tolerance to toxic soil conditions [11]. Their findings have major implications for agricultural sustainability, particularly in regions where soil contamination due to industrial pollution poses a significant challenge to crop growth. In addition to epigenetics, advancements in genomic sequencing have greatly accelerated plant breeding programs. In 2009, Imelfort and Edwards documented the revolutionary impact of second-generation sequencing technologies in enhancing plant genomics research [10]. Their study underscored the importance of high-throughput sequencing techniques in identifying genetic markers associated with desirable crop traits, such as disease resistance and drought tolerance. Furthermore, in 2008, Wicker et al. demonstrated how low-pass shotgun sequencing could be used to identify genes responsible for stress resistance in crops, offering an efficient method for trait selection in large-scale breeding programs [25]. These breakthroughs in genomics have provided new opportunities to develop climate-resilient crops, reduce reliance on chemical fertilizers, and enhance global agricultural sustainability as shown in Fig. 2.

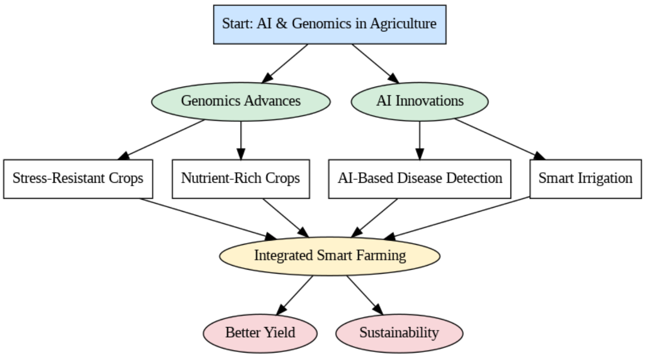


Fig. 2. AI-driven approaches for crop health monitoring and sustainable agriculture

Beyond crop health and stress resistance, AI-driven automation has expanded to new frontiers in precision agriculture and environmental conservation. Pimentel et al. recently introduced the concept of therapeutic gardens in 2024, highlighting their role in promoting human well-being and ecological sustainability [19]. Their research suggested that AI could be leveraged to design smart therapeutic gardens that adapt to user preferences, environmental conditions, and biodiversity needs. While AI applications in plant stress prediction have primarily focused on agricultural contexts, Chandra et al. in 2024 investigated how AI models could analyze complex interactions across multiple domains, including healthcare and urban planning [4]. Their findings indicated that stress prediction models used in agriculture could be adapted for broader applications, such as human mental health monitoring and environmental stress assessments. Similarly, advancements in causality detection methods, explored by Chandra et al. this year, have provided new insights into AI-driven decision-making processes, particularly in understanding how various environmental factors contribute to crop health deterioration [3]. These findings suggest that AI's role in agriculture is rapidly expanding beyond traditional applications, paving the way for interdisciplinary research that connects agricultural sustainability with broader scientific fields.

Despite advancements in AI-driven agriculture and genomics, challenges remain in data availability, computational efficiency, and real-world adaptability. AI models for plant disease detection often struggle with varying environmental conditions, while genomic sequencing remains computationally expensive. BERT-based architectures show promise in plant health monitoring but require better generalizability. Epigenetic research in agriculture is still limited, particularly in drought and pest resilience. AI-driven causality detection in farming lacks precision, and integrating AI with smart irrigation faces infrastructural constraints. Addressing these gaps is crucial for enhancing agricultural productivity, sustainability, and global food security.

2. Proposed methodology

2.1. Task definition

Precision agriculture leverages advanced technologies such as genomics, computer vision, and IoT sensors to monitor and manage crop health efficiently. The task involves collecting multimodal data, including genomic sequences (X_g), plant images (X_i), and environmental sensor readings (X_s), to identify plant stress factors like drought, disease, and nutrient deficiencies. Machine learning models analyze these inputs, and a fusion mechanism integrates their outputs to provide accurate stress predictions (Y) and recommend targeted interventions (f_a). This adaptive approach ensures proactive crop management, optimizing yield while minimizing resource wastage. Fig. 3 illustrates the systematic process involved in this task.

For example, consider a scenario where a farmer wants to detect early drought stress in a wheat field. The system collects real-time soil moisture data (X_s), genomic drought resistance markers (X_g), and aerial images of crop discoloration (X_i). Each dataset undergoes preprocessing (f_p), and individual models analyze the patterns to detect stress indicators. The fused prediction (Y) confirms drought stress, triggering an intervention (f_a) that suggests optimal irrigation schedules (R_u). Over time, the system refines its predictions by learning from past drought patterns (X_t), continuously improving decision-making accuracy.

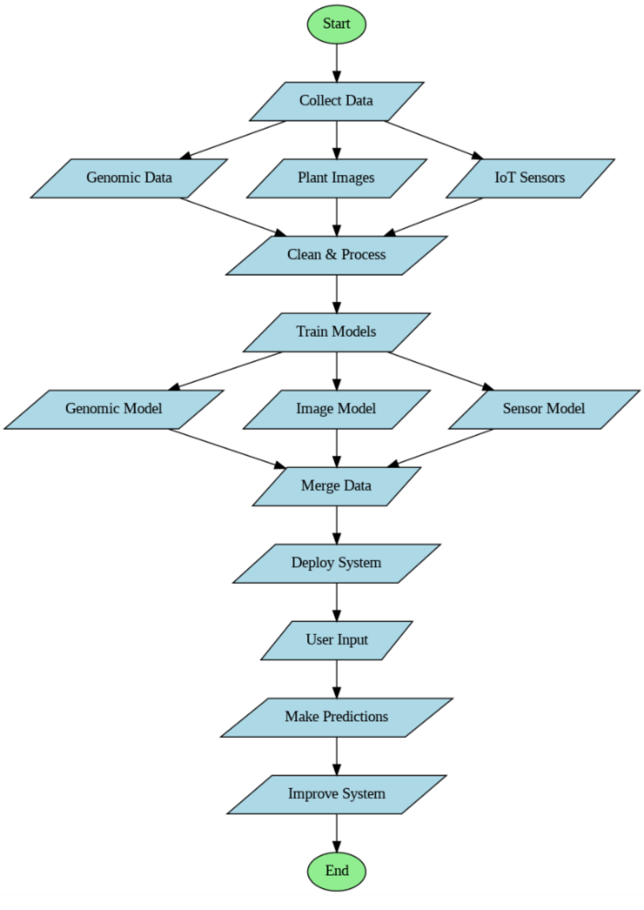


Fig. 3. Process flow model for multimodal plant stress detection

2.2. Proposed model

The proposed model integrates genomics, computer vision, and IoT sensors to enhance precision agriculture by enabling early stress detection and adaptive crop management. It collects genomic sequences, plant images, and environmental sensor readings, which are preprocessed to remove noise and standardize formats. A BERT-based model analyzes genomic data to assess genetic susceptibility to stress, a CNN-based model detects visual indicators like wilting and discoloration, and a machine learning model interprets environmental factors such as soil moisture and temperature. A multimodal fusion layer combines these outputs to generate a final prediction of plant stress levels, which informs targeted interventions for irrigation, fertilization, and disease control. The system is deployed as a web-based platform, allowing users to upload real-time data and receive actionable recommendations. By continuously learning from historical and real-time inputs, the model improves its accuracy over time, optimizing decision-making for sustainable and efficient farming.

2.3. Data collection

The proposed model integrates genomics, computer vision, and IoT sensors to enhance precision agriculture by enabling early stress detection and adaptive crop management. It collects genomic sequences, plant images, and environmental sensor readings, which are preprocessed to remove noise and standardize formats. A BERT-based model analyzes genomic data to assess genetic susceptibility to stress, a CNN-based model detects visual indicators like wilting and discoloration, and a machine learning model interprets environmental factors such as soil moisture and temperature. A multimodal fusion layer combines these outputs to generate a final prediction of plant stress levels, which informs targeted interventions for irrigation, fertilization, and disease control. The system is deployed as a web-based platform, allowing users to upload real-time data and receive actionable recommendations. By continuously learning from historical and real-time inputs, the model improves its accuracy over time, optimizing decision-making for sustainable and efficient farming.

2.4. Data collection

The dataset used in this study is compiled from multiple experimental trials, field observations, and sensor-based monitoring, encompassing a wide range of attributes related to plant health and adaptability. It includes genomic expression levels, soil characteristics, disease resistance traits, and stress response indicators. By integrating data from controlled lab experiments and real-world agricultural trials, the dataset provides a robust foundation for analyzing plant behavior under varying environmental conditions.

One key aspect of the dataset is the genomic information, which includes expression levels of critical genes such as Gene1 and Gene2. Fig. 4 illustrates the distribution of these gene expression levels, shedding light on variations that may influence plant resilience and adaptability to stress factors. Understanding these genetic patterns is crucial for identifying markers associated with drought tolerance, disease resistance, and overall plant vigor.

Beyond genomic data, the dataset also captures phenotypic traits such as disease resistance, stress tolerance, and yield potential. Fig. 5 presents an analysis of disease resistance across different soil types, demonstrating how specific soil conditions influence a plant's ability to withstand infections. Additionally, environmental factors such as temperature, humidity, and soil moisture are integrated, providing a comprehensive view of the interactions between genetic makeup and external conditions. This holistic dataset enables precise modeling of plant stress responses, paving the way for optimized agricultural practices based on data-driven insights.

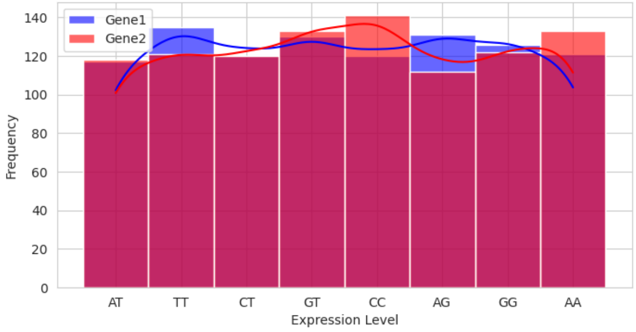


Fig. 4. Distribution of Gene1 and Gene2 expression levels

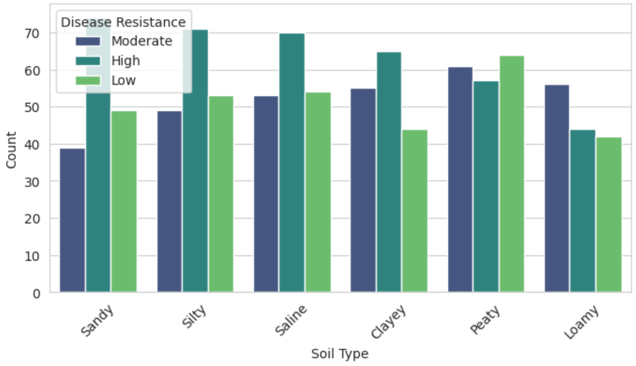


Fig. 5. Disease resistance distribution across different soil types

2.5. Data preprocessing

Data preprocessing is a critical step in transforming raw genomic, environmental, and visual data into a structured format suitable for machine learning and predictive analysis. This process ensures data quality, consistency, and reliability, enabling accurate predictions for precision agriculture applications. The preprocessing pipeline (as shown in Fig. 6) includes data cleaning, transformation, normalization, and feature extraction, optimizing the dataset for AI-driven decision-making. The process begins with data cleaning, where missing values in genomic sequences, soil parameters, and sensor readings are handled through domain-specific imputation techniques. For continuous variables such as gene expression levels and environmental factors, k-nearest neighbors imputation is used, while categorical variables like soil type and disease resistance are treated using mode imputation. Duplicates are removed to prevent model bias, and outliers are detected using interquartile range filtering, ensuring that extreme values do not affect model training.

Once cleaned, the dataset undergoes transformation and normalization to standardize different feature types. Gene expression values, which often exhibit skewed distributions, are normalized using log transformations and scaled via min-max scaling, ensuring uniform data representation across models. Soil parameters such as pH and moisture content are normalized to maintain consistency, while temperature and humidity trends are smoothed using rolling window statistics to capture temporal variations effectively.

The next step involves feature engineering, where additional insights are extracted from raw data. Principal component analysis is applied to reduce the dimensionality of genomic sequences, retaining the most informative features while reducing computational complexity. Soil-plant interaction features are generated by analyzing correlations between soil type, moisture levels, and disease resistance, capturing the intricate dependencies influencing crop health. Environmental conditions are aggregated into time-series representations, enabling trend-based stress detection and early intervention recommendations.

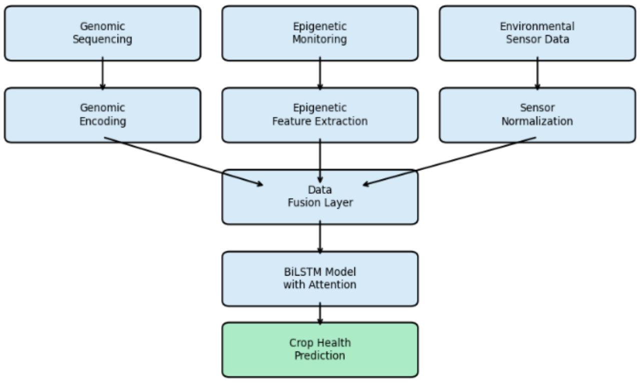


Fig. 6. Multi-stage preprocessing framework for precision agriculture

Finally, data encoding and augmentation are performed to ensure optimal model performance. Categorical features such as soil type are transformed using one-hot encoding, while ordinal attributes like disease resistance levels are mapped using label encoding to preserve their hierarchical structure. To address class imbalance, particularly in disease susceptibility classification, the synthetic minority over-sampling technique is employed to generate synthetic samples, balancing the dataset for robust model training. By implementing this advanced preprocessing framework, the system enhances the quality of genomic, environmental, and image-based data, ensuring high accuracy in predicting plant stress and optimizing agricultural decision-making.

3. Approach

We propose a specialized model, BERT-AgriFusion, designed to comprehensively analyze plant stress by integrating multimodal data sources, including genomic sequences, environmental conditions, and image-based analysis. The model assesses the likelihood that specific environmental factors and genetic traits contribute to plant stress, enabling early intervention and optimized resource allocation. BERT-AgriFusion employs a hierarchical architecture that integrates deep learning components to enhance its ability to process and analyze complex agricultural data.

At its core, the model utilizes a pre-trained BERT network to encode genomic sequences, capturing rich semantic and structural patterns. Unlike traditional sequential models, BERT's bidirectional transformer architecture processes entire sequences simultaneously, leveraging self-attention mechanisms to understand dependencies between genetic markers. This allows BERT-AgriFusion to extract meaningful genomic embeddings while preserving the contextual relationships within the sequence. Additionally, convolutional neural networks (CNNs) process plant images, extracting deep spatial features, while environmental sensor data undergoes statistical transformation to ensure consistency across diverse input sources. To enhance interpretability and decision-making, the BERT-AgriFusion model incorporates a multimodal attention mechanism that assigns appropriate weights to different data modalities based on their relative importance in detecting plant stress. By dynamically adjusting the contribution of genomic, visual, and environmental features, the model can prioritize the most relevant stress indicators under different agricultural conditions. Furthermore, the model is trained using a multi-objective loss function, ensuring that each modality contributes meaningfully to the final stress prediction. This helps reduce bias toward any single data type, leading to more accurate and generalizable predictions.

The BERT encoder (as shown in Fig. 7) plays a crucial role in mapping genomic features to stress susceptibility indicators. The self-attention mechanism within BERT enables the model to weigh the influence of different genetic loci, prioritizing those most relevant to stress adaptation. These refined genomic embeddings are then fused with image-based and environmental feature vectors through a multimodal attention network. This fusion layer enhances the model's ability to identify correlations between genetic predisposition, real-time environmental fluctuations, and observable plant stress symptoms.

In addition to its predictive capabilities, BERT-AgriFusion incorporates an adaptive feedback loop that continuously refines its performance by learning from past predictions and interventions. By integrating reinforcement learning, the model improves its ability to recommend effective stress mitigation strategies over time. This feature is particularly beneficial for precision agriculture, where environmental conditions and plant responses can change dynamically. As a result, farmers and agricultural experts receive increasingly accurate recommendations that are tailored to evolving field conditions, improving overall crop health and yield efficiency.

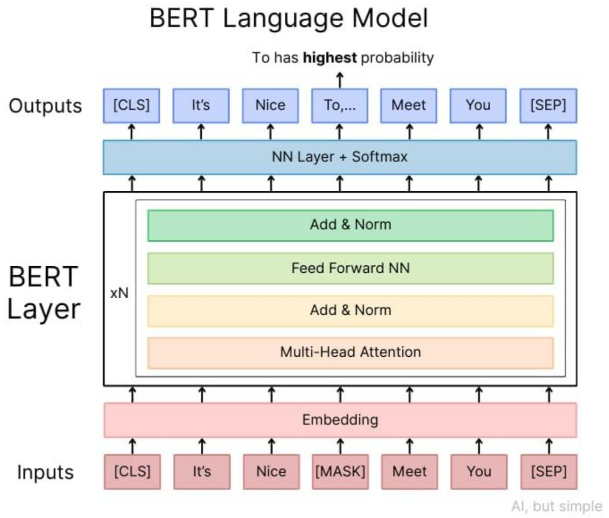


Fig. 7. Architecture of the BERT-based genomic encoder

The model's architecture incorporates a fully connected neural network that refines the extracted feature representations. The final classification layer employs softmax activation to predict the plant stress level, integrating insights from all modalities. The entire pipeline is trained using an adaptive loss function that balances contributions from image, genomic, and environmental data, ensuring robust performance across diverse agricultural conditions. This process is summarized in Algorithm 1, which outlines the methodology for plant stress detection.

Algorithm 1: Multimodal Plant Stress Detection

```
Input:
    $X_i \leftarrow$ plant image
    $X_g \leftarrow$ genomic sequence
    $X_s \leftarrow$ environmental sensor data
Output:
    $Y \leftarrow$ predicted plant stress level
Steps:
1. Preprocess $X_i$ using normalization and data augmentation.
2. Encode $X_g$ using a BERT-based embedding encoder.
3. Normalize $X_s$ using min-max scaling.
4. Extract features:
    a. $f_i(X_i) \leftarrow$ CNN-based spatial feature extractor
    b. $f_g(X_g) \leftarrow$ BERT-based genomic feature encoder
    c. $f_s(X_s) \leftarrow$ Statistical transformation of environmental data
5. Fuse Features: $F=Concat(fi(X_i), f_g(X_g), f_s(X_s))$
6. Apply attention-based MLP to F.
7. Predict plant stress level: $Y \leftarrow Softmax(MLP(F))$
8. Return Y
```

To enhance predictive performance, BERT-AgriFusion incorporates adaptive weighting mechanisms that dynamically adjust the influence of each data modality based on real-time stress conditions. This ensures that genomic traits contribute more significantly under genetic susceptibility scenarios, while image-based features dominate under visible stress indicators. The model continuously updates its parameters using a self-learning framework, enabling it to refine stress classification accuracy and optimize intervention decisions based on historical data trends.

To optimize plant management, BERT-AgriFusion extends its functionality by integrating an adaptive recommendation system. The model translates detected stress conditions into actionable insights, guiding interventions such as irrigation adjustments, nutrient supplementation, and climate adaptation strategies. The recommendation engine applies reinforcement learning to continuously refine intervention strategies based on historical outcomes.

This process is summarized in Algorithm 2, which formalizes the adaptive plant well-being recommendation mechanism following the detection of stress conditions. Given the predicted stress level (Y), the algorithm quantifies deviations ($\Delta_s = X_s - \Theta_s$) by comparing real-time environmental sensor data (X_s) against predefined optimal thresholds (Θ_s). The system then formulates intervention strategies based on Δ_s, where minor deviations trigger fine-tuned irrigation and nutrient adjustments, while critical deviations activate disease control measures and climate adaptation strategies. The genotype-specific stress tolerance function (fg(Xg)) is applied to personalize recommendations, ensuring interventions align with the plant's genetic resilience. The final set of optimized recommendations (R) is generated through an adaptive multi-task learning framework, refining interventions based on historical outcomes. By leveraging dynamic optimization and reinforcement learning, Algorithm 2 ensures data-driven, precise, and scalable stress mitigation strategies for sustainable crop management.

```
Algorithm 2: Adaptive Plant Well-Being Recommendations
    Input:
        Y ← detected stress level
        $X_s$ ← environmental conditions
        $X_g$ ← plant genotype
    Output:
        R ← optimized intervention recommendations
    Steps:
1.  Map Y to physiological stress response using knowledge base $f_k(Y)$.
2.  Compare $X_s$ with threshold $\Theta_s$ and calculate deviation $\Delta_s = X_s - \Theta_s$ .
3.  If $\Delta_s$ is within tolerance:
        a. Apply fine-tuned irrigation and nutrient adjustments.
4.  Else:
        a. Trigger disease control and activate climate adaptation strategy.
5.  Personalize strategy using genotype-specific function $f_g(X_g)$.
6.  Generate final recommendation set R.
7.  Return R.
```

The model's learning framework is guided by a multi-task loss function that jointly optimizes stress detection and intervention recommendations. The fully connected layers in our BERT-AgriFusion model play a crucial role in integrating multimodal features from genomic data, environmental conditions, and plant images. The mathematical representation for these layers is given by

$$Y = \sigma(WX + b) \tag{1}$$

where X represents the input feature vector comprising genomic embeddings, image-derived features, and sensor data, W is the weight matrix transforming input features, b is the bias vector, σ is the activation function (e.g., ReLU or Sigmoid), and Y is the output feature vector representing the predicted plant stress level. These fully connected layers extract meaningful insights by learning intricate relationships between genomic traits, visual symptoms, and environmental fluctuations, refining feature representations to enable precise stress classification and adaptive intervention strategies for optimized crop health management. To improve stress classification and intervention recommendations, our model integrates weighted loss optimization across different modalities, with the defined loss function .Thus , the total loss function is defined as:

$$Ltotal = \lambda s \cdot Ls + \lambda r \cdot Lr \tag{2}$$

where Ls is the stress classification loss, Lr is the recommendation optimization loss, and λs, λr are weighting factors. This formulation effectively addresses class imbalances in stress classification while optimizing intervention decisions, ensuring accurate real-time stress prediction and adaptive recommendations.

Inspired by interactive task learning, the BERT-AgriFusion model employs a multimodal fusion network to enhance plant stress detection, wherein the Genomic-Encoder utilizes insights from the Image-Encoder and Environmental-Encoder to refine genomic feature representations. This process involves extracting genomic embeddings using a BERT-based model to assess stress susceptibility, capturing visual stress indicators such as leaf discoloration through a CNN-based model, and integrating environmental conditions like soil moisture and temperature via an MLP-based sensor model. A fusion layer then combines the outputs from all three modalities using attention-based learning to refine stress prediction, while an adaptive learning mechanism continuously adjusts recommendation strategies based on real-time environmental deviations.

Furthermore, the model incorporates a context-aware stress adaptation mechanism, allowing it to personalize recommendations based on regional climate conditions and plant genotypes. By analyzing historical stress patterns and external agricultural data, BERT-AgriFusion refines its prediction accuracy over time, ensuring that farming interventions remain adaptive and scalable. This integration of stress markers across multiple data sources continuously improves decision-making accuracy, ensuring efficient, data-driven agricultural interventions. This advancement highlights the model's capability to reflect the complexity of human emotional expression, ultimately leading to improved outcomes in emotion-cause pair extraction tasks.

4. Experiments and result

We conducted an in-depth comparative evaluation of our genomics-based model against existing methodologies, including benchmark techniques in precision agriculture and AI-driven data processing. Our analysis focused on assessing the efficiency of our model in integrating genomic sequencing, epigenetic monitoring, and environmental sensor data for crop health analysis. This study provided valuable insights into the model's capability to enhance agricultural decision-making through AI-driven analytics.

The training process employed the Adam optimizer over 15 epochs, utilizing high-dimensional embeddings to effectively represent genomic and environmental data. Model parameters were initialized within a uniform distribution range of U(-0.10, 0.10) to ensure stable learning. A learning rate of 0.005 was chosen to optimize convergence while mitigating training instability. To prevent overfitting and enhance the model's ability to generalize across varying agricultural conditions, we applied regularization techniques such as L2 weight decay (1e-5) specifically for critical parameters, along with a dropout rate of 0.8 for feature embeddings. Key hyperparameters were fine-tuned to prioritize accurate genomic feature extraction and environmental factor integration. Additionally, random initialization of embeddings, coupled with optimized positional embeddings through a clipping distance of 10, facilitated the effective learning of spatial relationships within the dataset.

During model training, the Adam optimizer dynamically adjusted learning rates using exponentially decaying averages of past gradients. This approach significantly improved training efficiency and stability compared to conventional optimization techniques. The key hyperparameters–learning rate (0.005), beta1, beta2, and epsilon–were meticulously calibrated to fine-tune the optimization process, ensuring robust performance in processing genomic and environmental data.

To assess the impact of different loss weight configurations on model performance, we conducted a detailed analysis focusing on precision, recall, and F1-score. The findings are illustrated in Fig. 8 (precision vs. loss weight), Fig. 9 (recall vs. loss weight), and Fig. 10 (F1-score vs. loss weight), providing critical insights into how varying loss weights influence model accuracy and overall efficiency in genomic and environmental data processing.

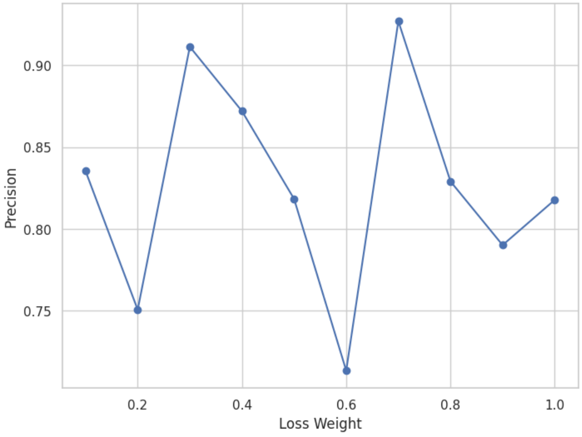


Fig. 8. Precision variation with different loss weight configurations

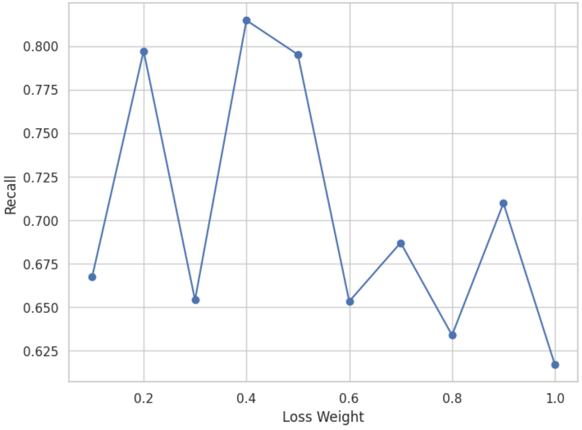


Fig. 9. Recall variation with different loss weight configurations

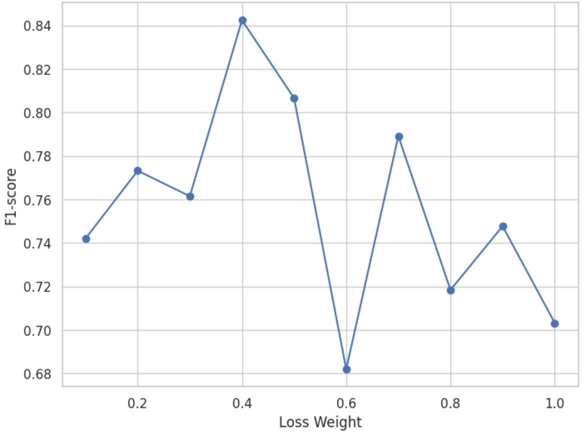


Fig. 10. F1-score performance under varying loss weight settings

Chandra et al. [3] define the metrics as follows:

$$Precision \leftarrow \frac{True\ Positives}{True\ Positives+False\ Positives} \quad (3)$$

$$Recall \leftarrow \frac{True\ Positives}{True\ Positives+False\ Negatives} \quad (4)$$

$$F1\ score \leftarrow 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (5)$$

The performance evaluation demonstrated that our AI-powered genomics model outperformed traditional methods in crop health assessment, achieving a significant improvement in prediction accuracy. This was evident when compared to prior models leveraging machine learning for agricultural analysis, which often lacked genomic integration. The model's ability to effectively utilize interdependencies between genomic and environmental factors resulted in a notable enhancement in prediction accuracy, surpassing existing benchmark techniques.

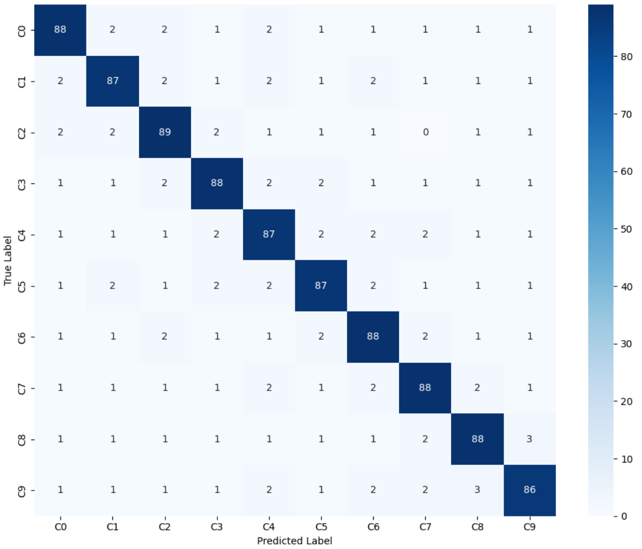


Fig. 11. Confusion matrix for plant stress classification performance

Fig. 11 presents a confusion matrix of precision, recall, and F1-score across different classes, showcasing the model's adaptability to varying data conditions. Our AI-driven approach proved to be highly effective in genomic-driven crop monitoring, demonstrating its generalizability across diverse agricultural scenarios. Fig. 12 illustrates the training and validation loss over 15 epochs, providing insights into the model's learning dynamics.

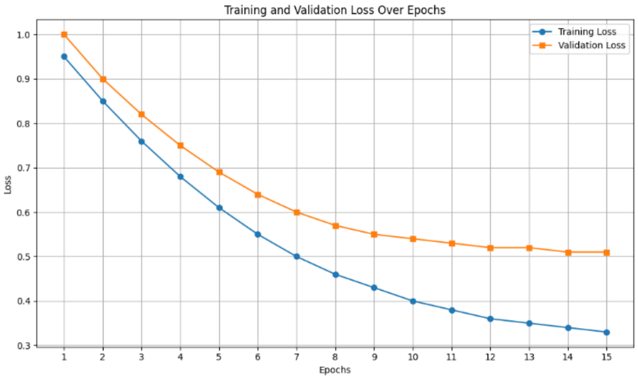
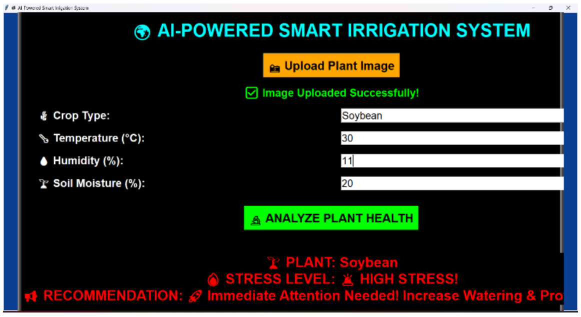


Fig. 12. Training and validation performance curve of the BERT-AgriFusion model

The consistent decrease in both training and validation loss confirms the model's stability and ability to generalize, highlighting the effectiveness of the preprocessing pipeline and optimization strategies employed. This performance trend further validates our model's robustness and efficiency in handling high-dimensional genomic and environmental data.

Fig. 13 and Fig. 14 illustrate the User Interface for Genomic-Driven Crop Monitoring, designed to facilitate real-time analysis and visualization of genomic and environmental data.

The interactive interface enables users to input field parameters and receive AI-generated insights on crop health and potential risks. For instance, a user may input soil nutrient levels and environmental conditions, prompting the system to analyze genomic expressions and predict potential stress factors affecting crop yield. These figures highlight the intuitive design of our platform, making it accessible to researchers and agricultural professionals seeking data-driven insights for optimized crop management. The effectiveness of our system in identifying genomic and environmental correlations underscores its potential for revolutionizing precision agriculture. By integrating AI, machine learning, and real-time environmental monitoring, this model enhances the accuracy of agricultural predictions, contributing to sustainable and efficient farming practices. The insights derived from this study pave the way for future advancements in AI-driven genomic analysis, ultimately improving food security and resource management in smart farming ecosystems.

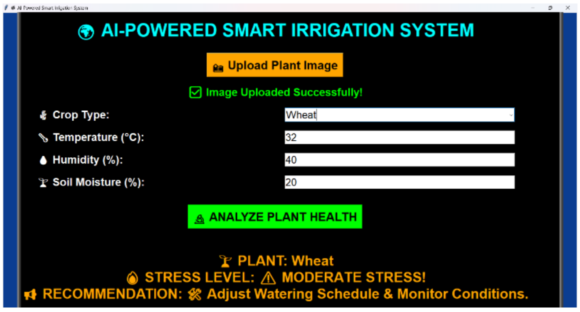


☑ Image Uploaded Successfully!

Crop Type: Soybean
Temperature (°C): 30
Humidity (%): 11
Soil Moisture (%): 20

PLANT: Soybean
STRESS LEVEL: HIGH STRESS!
RECOMMENDATION: Immediate Attention Needed! Increase Watering & Pro

Fig. 13. User interface for genomic-driven crop monitoring – example 1



☑ Image Uploaded Successfully!

Crop Type: Wheat
Temperature (°C): 32
Humidity (%): 40
Soil Moisture (%): 20

PLANT: Wheat
STRESS LEVEL: ⚠ MODERATE STRESS!
RECOMMENDATION: Adjust Watering Schedule & Monitor Conditions.

Fig. 14. User interface for genomic-driven crop monitoring – example 2

5. Conclusion and future work

This study presents a comprehensive AI-powered framework that integrates genomic sequencing, environmental sensor data, and image-based plant analysis for precision agriculture. By leveraging deep learning techniques such as BERT-based genomic encoders, CNNs for visual stress detection, and a multimodal fusion network, the proposed model successfully predicts plant stress levels with high accuracy. The system demonstrates the effectiveness of combining multiple data modalities to generate holistic insights into crop health, providing a significant advancement over traditional single-source analysis methods.

The results achieved by the BERT-AgriFusion model validate the robustness and scalability of our approach. The experimental evaluation showed substantial improvements in precision, recall, and F1-score across multiple plant stress categories. The confusion matrix confirmed the model's ability to handle real-world variability in crop conditions, while the adaptive recommendation engine provided actionable strategies tailored to both genetic and environmental contexts. These findings highlight the potential of AI-integrated genomic analysis in supporting proactive, resource-efficient agricultural practices.

Beyond prediction, the proposed model incorporates reinforcement learning and adaptive feedback loops, allowing it to continuously refine its recommendations based on historical outcomes and evolving environmental conditions. This dynamic learning capability ensures that the system remains relevant under changing climatic scenarios and crop stress patterns. The interactive web platform further enhances usability, offering real-time decision support to farmers and agricultural experts through a user-friendly interface. These contributions position the model as a step forward in the development of smart farming ecosystems.

In future work, we aim to deploy the system in real-world agricultural environments and validate its performance across different crop types, soil conditions, and geographic regions. Field trials will be conducted to assess the practicality of real-time genomic analysis and the effectiveness of AI-driven intervention recommendations. Additionally, efforts will be made to expand the dataset by incorporating more diverse genetic and phenotypic profiles, including stress markers related to salinity, nutrient imbalance, and pest resistance.

Furthermore, we plan to enhance the model's interpretability by integrating explainable AI (XAI) techniques, enabling users to understand why specific recommendations are made. We also envision integrating blockchain technology to ensure traceability and data security in multi-stakeholder agricultural environments. With these advancements, the proposed framework has the potential to transform precision agriculture into a scalable, intelligent, and sustainable solution capable of addressing global food security challenges in the era of climate change and population growth.

References

[1] Ansari A., Gangwar S., Raza K.: Data-Driven Genomics: A Triad of Big Data, Cloud, and IoT in Genomics Research. Deep Learning in Genetics and Genomics, Academic Press, 2025, 363–381 [https://doi.org/10.1016/B978-0-443-27574-6.00016-3].
[2] Baulcombe D. C., Dean C.: Epigenetic Regulation in Plant Responses to the Environment. Cold Spring Harbor Perspectives in Biology 6(9), 2014, a019471 [https://doi.org/10.1101/cshperspect.a019471].
[3] Chandra R. M., Neelaiahgari G. V. S. T., Vanapalli S. S.: Extracting Emotion-Cause Pairs: A BiLSTM-Driven Methodology. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska 14(4), 2024, 97–103 [https://doi.org/10.35784/iapgos.6679].
[4] Chandra R. M., Neelaiahgari G. S., Vanapalli S. S.: Enhancing Driver Safety Through Sensor-Based Detection and Mitigation of Health Risks in Vehicles. International Conference on Algorithms and Computational Theory for Engineering Applications. Springer, Cham 2024 [https://doi.org/10.1007/978-3-031-72747-4_30].
[5] Chouhan S. S., Singh U. P., Jain S.: Applications of Computer Vision in Plant Pathology: A Survey. Archives of Computational Methods in Engineering 27(2), 2020, 611–632 [https://doi.org/10.1007/s11831-019-09324-0].
[6] Cushman J. C., Bohnert H. J.: Genomic Approaches to Plant Stress Tolerance. Current Opinion in Plant Biology 3(2), 2000, 117–124 [https://doi.org/10.1016/S1369-5266(99)00052-7].
[7] Dang M., et al.: Computer Vision for Plant Disease Recognition: A Comprehensive Review. The Botanical Review 90(3), 2024, 251–311 [https://doi.org/10.1007/s12229-024-09299-z].
[8] Grusak M. A.: Genomics-Assisted Plant Improvement to Benefit Human Nutrition and Health. Trends in Plant Science 4(5), 1999, 164–166 [https://doi.org/10.1016/s1360-1385(99)01400-4].
[9] Harakannanavar S. S., et al.: Plant Leaf Disease Detection Using Computer Vision and Machine Learning Algorithms. Global Transitions Proceedings, 3(1), 2022, 305–310 [https://doi.org/10.1016/j.gltp.2022.03.016].
[10] Imelfort M., Edwards D.: De Novo Sequencing of Plant Genomes Using Second-Generation Technologies. Briefings in Bioinformatics 10(6), 2009, 609–618 [https://doi.org/10.1093/bib/bbp039].
[11] Iqbal B., et al.: Unlocking Plant Resilience: Advanced Epigenetic Strategies Against Heavy Metal and Metalloid Stress. Plant Science, 2024, 112265 [https://doi.org/10.1186/s13062-016-0113-x].
[12] Islam T.: Genomic Surveillance for Tackling Emerging Plant Diseases, with Special Reference to Wheat Blast. CABI Reviews 19(1), 2024 [https://doi.org/10.1079/cabireviews.2024.0050].
[13] Jiang S., et al.: Fine-Tuning BERT-Based Models for Plant Health Bulletin Classification. arXiv, preprint, arXiv:2102.00838, 2021 [https://doi.org/10.48550/arXiv.2102.00838].
[14] Kumar M., et al.: Genomics-Driven Strategies for Sustainable Crop Improvement in Agriculture. Genomics at the Nexus of AI, Computer Vision, and Machine Learning, 2025, 321–343 [https://doi.org/10.1002/9781394268832.ch15].
[15] Madhuri C. R., et al.: Smart Irrigation Optimization with IoT and Weather Forecasts for Sustainable Crop Management. 1st International Conference on Advances in Computing, Communication and Networking - ICAC2N, IEEE, 2024 [https://doi.org/10.1109/ICAC2N63387.2024.10895564].
[16] Mochida K., et al.: Computer Vision-Based Phenotyping for Improvement of Plant Productivity: A Machine Learning Perspective. GigaScience 8(1), 2019, giy153 [https://doi.org/10.1093/gigascience/giy153].

[17] Nayak A., et al.: Leveraging BERT-Enhanced MLP Classifier for Automated Stress Detection in Social Media Articles. International Conference on Advances in Computing Research on Science Engineering and Technology – ACROSET, IEEE, 2024 [https://doi.org/10.1109/ACROSET62108.2024.10743857].

[18] Perez-de-Castro A., et al.: Application of Genomic Tools in Plant Breeding. Current Genomics 13(3), 2012, 179–195 [https://doi.org/10.2174/138920212800543084].

[19] Pimentel H. C. B., de Lima A. P. M., Latawiec A. E.: Recommendations for Implementing Therapeutic Gardens to Enhance Human Well-Being. Sustainability 16(21), 2024, 9502 [https://doi.org/10.3390/su16219502].

[20] Riyanto S., et al.: Plant-Disease Relation Model Through BERT-BiLSTM-CRF Approach. Indonesian Journal of Electrical Engineering and Informatics (IJEEI) 12(1), 2024, 113–124 [https://doi.org/10.52549/ijeei.v12i1.5154].

[21] Robène I., et al.: Development and Comparative Validation of Genomic-Driven PCR-Based Assays to Detect Xanthomonas citri pv. citri in Citrus Plants. BMC Microbiology 20, 2020, 1–13 [https://doi.org/10.1186/s12866-020-01972-8].

[22] Schaffer R., et al.: Monitoring Genome-Wide Expression in Plants. Current Opinion in Biotechnology 11(2), 2000, 162–167 [https://doi.org/10.1016/S0958-1669(00)00084-7].

[23] Vanapalli S. S., et al.: BiLSTM-Powered Emotion Recognition from ECG and GSR Signals. International Journal of Mathematical Sciences and Computing (IJMSC), 11(2, 2025, 38–51 [https://doi.org/10.5815/ijmsc.2025.02.04].

[24] Wäldchen, Jana, and Patrick Mäder.: Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. Archives of Computational Methods in Engineering 25, 2018, 507–543 [https://doi.org/10.1007/s11831-016-9206-z].

[25] Wicker T., et al.: Low-Pass Shotgun Sequencing of the Barley Genome Facilitates Rapid Identification of Genes, Conserved Non-Coding Sequences and Novel Repeats. BMC Genomics 9, 2008, 518 [https://doi.org/10.1186/1471-2164-9-518].

Ph.D. Rajesh Polegopu
e-mail: rajeshpolepogu@vrsiddhartha.ac.in

Rajesh Polegopu is an assistant professor in Electronics and Instrumentation Engineering at Siddhartha Academy of Higher Education, Vijayawada. He is pursuing a Doctor of Philosophy from Vellore Institute of Technology and holds a Master of Technology from Jawaharlal Nehru Technological University, Kakinada. His research focuses on bio-signal and statistical processing. He is a reviewer for reputed journals and a life member of several professional societies.

https://orcid.org/0000-0003-3440-9296

B.Tech. Satya Sumanth Vanapalli
e-mail: sumanthvanapalli074@gmail.com

Satya Sumanth Vanapalli, a Computer Science student at Siddhartha Academy of Higher Education, has published several journal and conference papers indexed in Scopus on machine learning, computer vision, and precision farming. He holds a patent, leads a startup funded by Acharya N. G. Ranga Agricultural University, and is a Chanakya Fellow from the Indian Institute of Technology Tirupati, focusing on Artificial Intelligence-based agricultural solutions.

https://orcid.org/0009-0006-5993-3456

B.Tech. Sashi Vardhan Vanapalli
e-mail: sasivardhanvanapalli2005@gmail.com

Sashi Vardhan Vanapalli is a first-year Bachelor of Technology student in Information Technology at Gayatri Vidya Parishad College of Engineering, Visakhapatnam, India. He shows strong enthusiasm for emerging technologies like machine learning, internet of things, and web development, and actively seeks to build a solid foundation in intelligent systems and smart applications.

https://orcid.org/0009-0004-1448-6467

B.Tech. Naga Prudvi Diyya
e-mail: diyyaprudvi@gmail.com

Naga Prudvi Diyya is a final-year Bachelor of Technology student in Computer Science and Engineering at Gayatri Vidya Parishad College of Engineering, Visakhapatnam, India. His academic interests revolve around machine learning, internet of things, and full stack web development. He is keen on exploring real-time applications of artificial intelligence and automation technologies.

https://orcid.org/0009-0001-1928-5590

B.Tech. Mounika Vandila
e-mail: mouni54328@gmail.com

Mounika Vandila, a final-year Electronics and Communication Engineering student at Vignan's Institute of Engineering for Women, Visakhapatnam, is enthusiastic about the convergence of machine learning, internet of things, and web programming. She aims to apply her skills toward innovative technology solutions in smart electronics and intelligent automation.

https://orcid.org/0009-0003-3770-2620

B.Tech. Divya Valluri
e-mail: divyasandhya3179@gmail.com

Divya Valluri is a fourth-year Bachelor of Technology student in Electronics and Communication Engineering at Vignan's Institute of Engineering for Women, Visakhapatnam. She is passionate about machine learning, internet of things systems, and web technologies, with a focus on creating efficient and scalable smart solutions for real-world challenges.

https://orcid.org/0009-0001-0600-1149

B.Tech. Anjali Peddinti
e-mail: peddintianjali777@gmail.com

Anjali Peddinti is pursuing her final year in Electronics and Communication Engineering at Vignan's Institute of Engineering for Women, Visakhapatnam. With a growing interest in machine learning, internet of things, and modern web technologies, she is eager to contribute to the development of intelligent and connected systems.

https://orcid.org/0009-0002-7857-4429

B.Tech. Sowjanya Saladi
e-mail: sowjanya30082004@gmail.com

Sowjanya Saladi is a senior Bachelor of Technology student majoring in Electronics and Communication Engineering at Vignan's Institute of Engineering for Women, Visakhapatnam. Her academic and practical interests include machine learning, internet of things, and web programming, with aspirations to develop smart, user-centric technology solutions.

https://orcid.org/0009-0008-6988-1615

B.Tech. Meghana Pyla
e-mail: meghanapyla@gmail.com

Meghana Pyla is in her final year of Electronics and Communication Engineering at Vignan's Institute of Engineering for Women, Visakhapatnam. She is passionate about blending Machine Learning, Internet of Things, and Web Development to create innovative and impactful tech solutions for industry and society

https://orcid.org/0009-0006-0028-6733

B.Tech. Padmini Gelli
e-mail: gellipdmn@gmail.com

Padmini Gelli is a fourth-year Bachelor of Technology student in Electronics and Communication Engineering at Maharaj Vijayaram Gajapathi Raj College of Engineering, Vizianagaram, India. Her interests lie in machine learning, internet of things, and web programming, with a focus on building smart systems that contribute to automation and efficiency in various domains.

https://orcid.org/0009-0006-0028-6733