Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Guavafusionformer: A Dual-Branch Soil-Climate Fusion Network for Intelligent Guava Health Classification

ARCHANA DIRGULE

archana.dirgule@dypatil.edu

D.Y. Patil Deemed to be University, Navi Mumbai, India.

Amarsinh Vidhate

D.Y. Patil Deemed to be University, Navi Mumbai, India.

Archana Pakhare

Tukaram Chavan

Research Article

Keywords: Guava Health Classification, Soil-Climate Fusion, Dual-Branch Network, Adaptive Feature Fusion, Deep Learning, Precision Agriculture, Intelligent Agriculture

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11317368/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

GUAVAFUSIONFORMER: A DUAL-BRANCH SOIL-CLIMATE FUSION NETWORK FOR INTELLIGENT GUAVA HEALTH CLASSIFICATION

Archana Dirgule[1*], Amarsinh Vidhate[2], Archana Pakhare[3*], Tukaram Chavan[4]

[1*]Department of Computer Science and Engineering, Ramrao Adik Institute of Technology, D.Y. Patil Deemed to be University, Navi Mumbai, India.

[2]Department of Computer Science and Engineering, Ramrao Adik Institute of Technology, D.Y. Patil Deemed to be University, Navi Mumbai, India.

[3*]Department of Artificial Intelligence and Analytics School of Computing, MIT–Art,Design & Technology University, Pune, India.

[4]Department of Computer Science and Engineering, Shree Siddheshwar Women's College of Engineering, Solapur, DBATU, Lonere, India.

[1*]Corresponding Author Email:

[1] archana.dirgule@dypatil.edu ,[3] archana.pakhare@mituniversity.edu.in

Abstract

Guava is a fruit economically relevant in both tropical and subtropical climates and accurate assessment of health is critical to ensure sustainable production and yield. However, traditional methods might be inadequate in capturing complex relationships existing in diverse soil-climatic variables. This paper presents GuavaFusionFormer, a two-branch soil-climate fusion model for intelligent classification of guava health. The model consists of nitrogen, phosphorus, potassium, pH, soil moisture, temperature, humidity, and rainfall fed into the Dual-Branch Contextual Attention Encoder (DCAE), Adaptive Feature Enhancement (AFE), Multi-scale Dependency Learning (MDL), Adaptive Soil-Climate Fusion (ASCF), and Gated Residual Network (GRN). Employing the Guava Soil-Climate Disease Dataset collected from Maharashtra, India, the framework makes classification of eight guava health and disease states. The GuavaFusionFormer model obtains 0.9754% accuracy, 0.9757% macro precision, 0.9754% macro recall, and 0.9755% macro F1-score, and its specificity and FPR are 0.9965% and 0.35%, respectively. In addition to evaluating its classification capability via comparative experiments on traditional and Transformer-based deep learning approaches, SHAP and t-SNE analysis provides an interpretation of feature contribution and representation separability for guava health.

Keywords: Guava Health Classification; Soil-Climate Fusion; Dual-Branch Network; Adaptive Feature Fusion; Deep Learning; Precision Agriculture; Intelligent Agriculture.

1. Introduction

Fruit crops are an important component of agriculture and provide a significant source of food, especially in tropical and subtropical countries, where they are essential for food and nutritional security and agricultural sustainability [1]. Guava is one of these crops due to its economic importance, nutritional value, adaptability and productivity under different agro-climatic conditions [2]. However, guava plants require proper soil conditions and a suitable climate for healthy growth and fruit development [3]. The properties of the soil, like nitrogen (N), phosphorous (P), potassium (K), soil pH, and moisture, affect nutrient availability and plant growth, while climatic factors, such as temperature, relative humidity and rainfall, affect physiological development, disease progression and environmental stress [4, 5]. Hence, a soil-climatic analysis is crucial to obtain a holistic knowledge of guava health and to enable prompt agricultural management decisions.

Most of the conventional methods of guava health assessment rely on periodic field inspection, manual observation, and independent analysis of environmental parameters [6]. These methods may not fully reflect the combined effects of soil and climatic factors on the health of plants [7]. The dependency structure and statistical properties of soil variables and climatic variables can differ, as well as the interaction between them based on the environmental condition and the health state of guava plants [8]. It is also hard to derive higher-order relationships between below-ground soil parameters and above-ground climatic parameters from the analysis of these diverse parameters individually. Moreover, traditional feature integration techniques may

just merge the heterogeneous information and do not identify the relative importance of each modality under various health conditions. These constraints inspire the development of intelligent data-driven methods that can learn and fuse complementary soil and climate data for robust guava health assessment.

The recent developments in the field of Artificial Intelligence (AI) and deep learning present new opportunities for automated monitoring and decision-support systems in agriculture [9]. Machine-learning techniques have proven to be effective in detecting non-linear relationships between environmental variables and have been used in more tasks relating to agricultural classification and prediction [10]. Yet, traditional machine-learning methods might not be able to automatically learn higher-order dependencies in heterogeneous environmental observations [11]. Deep-learning models offer automatic representation learning and can capture informative patterns without having to use hand-crafted features entirely [12]. Recent years have seen the use of more recently invented architectures that exhibit great potential for structured tabular data, such as attention-based and Transformer architectures, which can learn contextual relations among individual attributes [13, 14]. However, the process of applying a single learning pathway directly to the heterogeneous soil and climate data might not fully retain the modality-specific properties. Moreover, feature-learning and/or fusion techniques are insufficient to handle adaptive feature enhancement, multi-scale dependency modelling and context-dependent cross-modal interaction, important for discriminating a wide range of guava health conditions.

To solve these problems, this research work introduces a novel dual-branch soil-climate fusion network called GuavaFusionFormer for intelligent health monitoring of guavas. The proposed framework is intended to learn informative representations from soil and climatic observations of different modalities while retaining the salient features of each modality. It adopts independent Soil and Climate branches, and each branch carries out the Dual-Branch Contextual Attention Encoder (DCAE) to conduct contextual representation learning. The learned representations are then refined by Adaptive Feature Enhancement (AFE) and used to process representations for learning multi-scale dependencies (MDL) at different representation-levels. The resulting modality-specific representations are then combined via Adaptive Soil–Climate Fusion (ASCF) that creates bidirectional cross-modal relationships and dynamically controls the importance of soil and climate data. Lastly, a Gated Residual Network (GRN) is used for post-fusion representation refinement prior to guava health classification into eight classes.

The current research study distinguishes itself from other research work performed earlier in guavas by employing heterogeneous soil and climate data to classify guava's healthy state and diseased condition. In contrast to using image-based visual data, the proposed framework processes structured environmental data from two different sources, soil and climate. The GuavaFusionFormer framework is devised to extract modality-dependent and conditional cross-modal interactions between them prior to performing fusion of the same, thus addressing the issue of heterogeneity of soil and climate data.

The major findings of this study are presented below:

- GuavaFusionFormer is designed as a dual-branch soil–climate deep-learning framework to learn comprehensive representations from heterogeneous agricultural environmental observations, whilst also retaining modality-specific information.
- A Dual-Branch Contextual Attention Encoder (DCAE) is proposed to learn relationships between different feature-specific tokens for each modality and generate context-aware latent representation of soil and climate attributes.
- A module based on Adaptive Feature Enhancement (AFE) has been embedded that selectively enhances informative representation components and suppresses less informative or redundant responses by learnable feature gating.
- The soil and climate branches use Multi-scale Dependency Learning (MDL) independently before cross-modal interaction, to learn the complementary dependencies between these modalities at various hierarchies.
- An Adaptive Soil–Climate Fusion (ASCF) module is designed to build cross-modal relationships in both directions by employing interaction with attention; instead of just simple feature concatenation, the relative contribution of soil and climate representations is dynamically adjusted.
- To ensure reliable health classification, a Gated Residual Network (GRN) is added after fusion for nonlinear representation refinement and selectively preserving discriminative components.
- The proposed framework performs eight-class guava health classification, comprising Healthy, Anthracnose, Canker, Rust, Leaf Spot, Fruit Scab, Stylar End Rot, and Nutrient Stress, thereby providing a comprehensive environmental-data-driven approach to guava health monitoring.

This study aims to design an AI-based intelligent health assessment system for guava plants using structured soil and climate observations. The proposed GuavaFusionFormer can learn the environmental representation of each modality, model the intra-modal dependency for each modality, build adaptive associations between soil and climatic information, and finally enable the creation of a fused representation for finer discrimination of soil health. The framework is tested with the corresponding guava

soil and climate observations and evaluated using accuracy, precision, recall and F1-score to measure its effectiveness in differentiating the eight health conditions.

Section 2 discusses the related works in guava health assessment and AI-based agricultural monitoring. The proposed framework GuavaFusionFormer and the methodological components are described in Section 3. The experimental results of the study, along with the discussion, are included in Section 4. Finally, Section 5 sums up the study and provides directions for future research and directions for limitations.

2. Related Works

Based on the most recent research about guava health and disease classification, it can be broadly classified into the following types: Conventional machine learning-based approaches, Deep convolutional feature learning, transfer learning-based approaches, and advanced feature representation strategies. The above-mentioned studies indicate the growing potential of artificial intelligence in the field of automated assessment of guava diseases and fruit quality. Most of the current methods, however, are based on visual data, and they have been mainly developed to learn discriminative patterns using single modalities.

2.1. Traditional Machine Learning and Feature-based Classification

Early and hybrid methods have been investigated to combine deep feature extraction with traditional machine learning classifiers for guava disease recognition. Kılcı and Koklu [18] extracted the features by using Inceptionv3 and then employed artificial neural networks (ANN), support vector machines (SVM), k-nearest neighbors (kNN), random forest and decision tree classifiers. The findings showed that SVM and ANN have better classification accuracy with a lower number of misclassifications between anthracnose, fruit fly and healthy guava categories. The sequential architecture is not necessarily beneficial to the simultaneous learning of hierarchical and contextual feature interactions, although this approach proved successful for combining deep feature extraction with conventional classifiers.

2.2. A Guava Disease Classification system based on deep CNN

Subsequently, deep convolutional neural networks (CNNs) have emerged as important tools for automatically obtaining visual representations related to disease. Doutoum et al. [19] studied 1834 guava leaf images from five disease classes on four different pretrained CNN architectures, and EfficientNet-B3 performed best. Likewise, Pathmanaban et al. [20] proposed a CNN model for the quality assessment of guava fruits with digital and thermal imaging under various maturity and damage levels. The studies showed that CNNs can indeed learn discriminative visual features without needing to manually engineer them. However, traditional CNN-based models tend to focus on extracting spatial features from a single visual input and may lack the ability to simultaneously learn complementary fine-grained and high-level contextual features.

2.3. Transfer Learning and Lightweight Deep Learning Models

Transfer learning has also been widely used to help with the recognition of guava diseases and decrease the need to train deep networks from scratch. In guava disease detection, a CNN (ResNet-101) was used by Ahmed et al. [21] used augmentation to create a balanced dataset of 4,632 images in three classes: Anthracnose, Fruit Fly, and Healthy. To make the model interpretable, Grad-CAM was also applied to visualize the regions of the image that contribute to the disease in the image. Similarly, Mustak un-Nobi et al. [22] introduced GLD-Det, which is a lightweight transfer-learning model that uses a modified MobileNet model with extra pooling, batch normalization, dropout, and ReLU layers. The model was trained to be good on two benchmark datasets, and the predictions were visualised using Grad-CAM. Although these benefits exist, image-centric transfer-learning approaches are still largely the norm, and the characteristics of the disease may not be well distinguished. In addition, light architectures might hinder deeper hierarchical and contextual representation learning.

2.4. Multimodal and Context-Aware Agricultural Learning

Although image-based disease recognition is the major focus of agricultural research, more recently, heterogeneous environmental information, such as soil and climatic data, has been considered for plant monitoring and agricultural prediction. The soil composition (nutrients, pH, moisture) gives information about the environment in which the plant is growing, while the temperature, humidity and rain reflect the surrounding climatic conditions. Combining these complementary sources can provide a more comprehensive representation of plant health than relying exclusively on visual observations. But there are significant differences between the statistical properties and information structure of heterogeneous agricultural modalities. Concatenation of features or processing of individual features may then be unable to maintain all the information associated with the modality of the data or to capture the context in which soil and climate conditions are related.

2.5. Research Gap and Motivation

While previous research has demonstrated promising results in guava disease and fruit-health classification, several methodological and data-representation gaps remain. First, most existing guava studies are image-centric, whereas comparatively limited attention has been given to the joint modelling of structured soil and climate information for guava health assessment. Second, conventional or single-pathway feature-learning approaches may be inadequate for capturing the hierarchical and multi-scale dependencies present in heterogeneous agricultural attributes. Third, limited adaptive interaction between heterogeneous environmental representations may restrict the ability of existing models to learn context-dependent relationships between different agricultural conditions. In addition, fixed or undifferentiated feature integration may retain redundant or less discriminative information and may not sufficiently exploit the complementary characteristics of different modalities. Furthermore, the use of soil and climate information for eight-class guava health classification represents a comparatively underexplored multimodal tabular setting in the existing guava literature. These limitations motivate the development of a framework that can preserve modality-specific characteristics, learn multi-scale intra-modal dependencies, establish adaptive cross-modal interactions, and progressively refine the integrated representation for robust eight-class guava health classification.

3. Materials and Methods

This study utilizes heterogeneous agricultural data comprising soil and climate observations associated with eight guava health conditions, namely Healthy, Anthracnose, Canker, Rust, Leaf Spot, Fruit Scab, Stylar End Rot, and Nutrient Stress. The soil modality contains nitrogen (N), phosphorus (P), potassium (K), soil pH, and soil moisture, whereas the climate modality includes temperature, relative humidity, and rainfall. The two modalities are independently preprocessed through missing-value imputation, outlier detection, and normalization, followed by Synthetic Minority Over-sampling Technique (SMOTE) applied exclusively to the training partition to mitigate class imbalance. A new GuavaFusionFormer, a dual-branch soil–climate fusion network, is proposed to learn complementary environmental representations for intelligent guava health classification. The proposed framework independently processes the soil and climate modalities through a Dual-Branch Contextual Attention Encoder (DCAE) to learn context-aware feature representations, followed by Adaptive Feature Enhancement (AFE) to selectively emphasize informative components. Subsequently, Multi-scale Dependency Learning (MDL) captures fine-grained and broader hierarchical dependencies within each modality while preserving modality-specific characteristics. The resulting representations are then integrated through Adaptive Soil–Climate Fusion (ASCF), which employs bidirectional cross-attention and adaptive fusion mechanisms to establish contextual interactions between soil and climate information. Finally, a Gated Residual Network (GRN) performs post-fusion nonlinear refinement before an eight-class fully connected classification head and Softmax activation generate the final guava health prediction. Through this dual-branch and progressively refined learning strategy, GuavaFusionFormer aims to preserve distinctive soil and climate information while effectively learning their complementary relationships for robust and fine-grained multi-class guava health classification. Figure 1 depicts the overall architecture of the proposed GuavaFusionFormer framework.

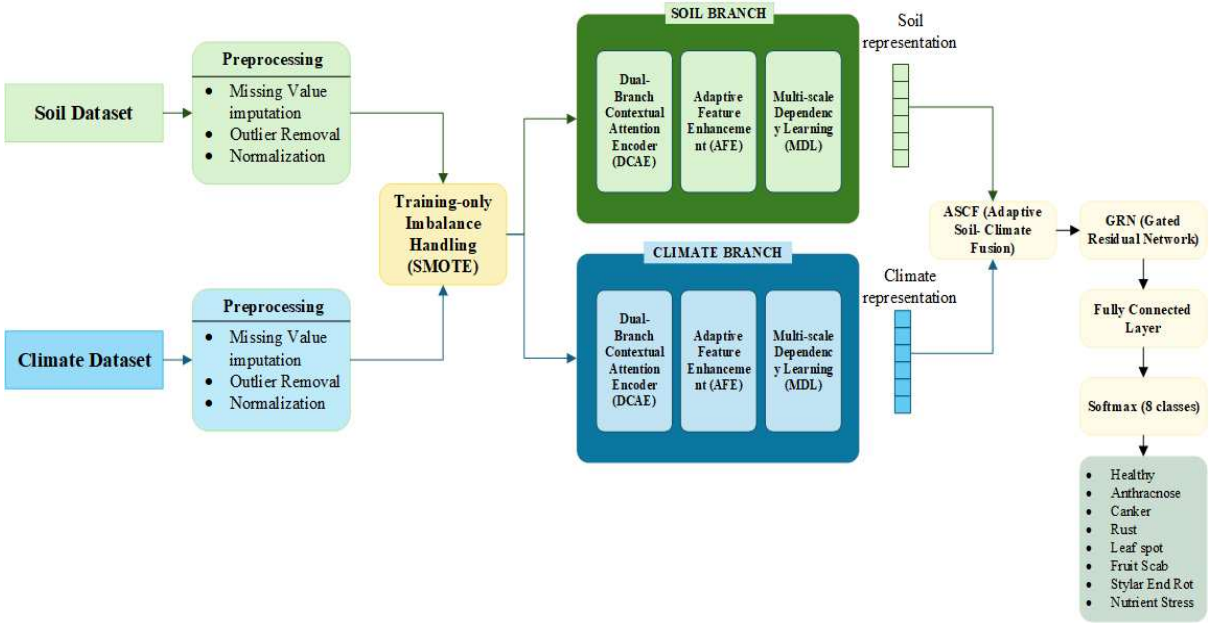


Figure 1. *Overall architecture of the proposed methodology*

3.1. Materials

The data set that has been used for this research work is called “Guava Soil-Climate Disease Data Set” [23]. It consists of 12,000 structured observations that are related to health conditions of guavas under different soil and climatic conditions. The data set was created with the purpose of reflecting the growing conditions of guavas in Maharashtra, India in the rainy season. For systematic representation, ten representative environments of orchards are being considered and 1,200 observations are generated for each of the orchards resulting in the number of 12,000 observations. These observations reflect the health condition of around 120 guava plants in each of the orchard environments and multiple observations are used for variation in soil and climatic conditions during the mentioned rainy season period from June to September. There are five soil-related attributes, which are nitrogen (N), phosphorus (P), potassium (K), soil pH, and soil moisture, and three climate-related attributes, which are temperature, relative humidity, and rainfall. The ninth attribute represents the guava health or disease category. Eight target classes are Healthy, Anthracnose, Canker, Rust, Leaf Spot, Fruit Scab, Stylar End Rot, and Nutrient Stress. The soil features were defined using agronomic measures of nutrient and soil quality, while the climate features were defined using environmental variations of temperature, relative humidity, and rainfall. Given that the dataset is a structured computationally generated dataset, it means that the individual observations cannot be considered as individually measured readings made using named physical equipment. In addition, the orchard environments can be considered as representative growing condition profiles rather than individually identified commercial orchards. This makes it possible to define heterogeneous soil-climate conditions using the same feature definition framework that enables multimodal learning. Disease or Health class of each individual observation has been determined by the eight-class classification system defined in the dataset. The obtained dataset enables unified mapping from the environmental features to guava health classes and can thus be used to test supervised multi-class classifiers. This dataset is represented in two complimentary modalities. Specifically, soil is composed of N, P, K, soil pH, and soil moisture, while climate is composed of temperature, relative humidity, and rainfall. Importantly, these modalities are kept as separate inputs in the early stage of processing to maintain modality-specific statistics. The obtained feature representations are then fed to the proposed GuavaFusionFormer network to adaptively fuse the inputs to learn complementary relationships between soil and climate features for guava health classification. The source dataset consists of 12,000 observations. To address class imbalance problem, only the training set is augmented using SMOTE. Importantly, validation and testing sets remain untouched. Thus, the source dataset observations and the synthetic observations obtained during the process of model training are two separate things. The full set of dataset characteristics is presented in Table I.

Table 1. ***Characteristics of the Guava Soil-Climate Disease Dataset***

Category	Description
Dataset name	Guava Soil-Climate Disease Dataset
Data source	Kaggle
Study region	Maharashtra, India
Agricultural environment	Tropical and subtropical
Number of observations	12,000
Number of attributes	9
Soil attributes	N, P, K, pH, Moisture
Climate attributes	Temperature, Humidity, Rainfall
Target attribute	Disease
Number of classes	8
Health/disease classes	Healthy, Anthracnose, Canker, Rust, Leaf Spot, Fruit Scab, Stylar End Rot, Nutrient Stress
Learning task	Multi-class classification
Data modalities	Soil and Climate

3.2. Preprocessing Techniques

In order to guarantee consistency of the data and facilitate the model learning process, a system of systematic data preprocessing was completed before data were used in the proposed GuavaFusionFormer framework. Soil and climate attributes were separated as two different modalities in the environment because they are different in terms of statistical properties and are heterogeneous with respect to soil and climate. This is a modality-specific processing which prevents the two sources from interacting in advance as they should, and allows the two branches to retain their own environmental properties.

Initially, the loss of observations was detected separately in the soil and climate modalities. All the input variables were numerical, so the missing values were filled in with a training-set-fitted imputation approach. When median imputation is employed, the imputed feature value is defined as

$$x_{i,j}^{imp} = \begin{cases} x_{i,j}, & \text{if } x_{i,j} \text{ is observed} \\ median\left(X_j^{train}\right), & \text{if } x_{i,j} \text{ is missing} \end{cases} \quad (1)$$

where $x_{i,j}$denotes the value of feature $j$for observation i, and X_j^{train}represents the corresponding feature values obtained from the training partition. The pre-processing step is not influenced by the information in the validation and testing partitions as the imputation statistic is only fitted to the training partition. After filling in the missing values, abnormal values were detected separately for the soil and climate modalities. The Interquartile Range (IQR) criterion was used to determine extreme observations. For feature j, the IQR is calculated as

$$IQR_j = Q_{3,j} - Q_{1,j} \quad (2)$$

where $Q_{1,j}$and $Q_{3,j}$denote the first and third quartiles, respectively. The lower and upper limits for detecting potential outliers are defined as

$$L_j = Q_{1,j} - 1.5\left(IQR_j\right) \quad (3)$$

$$U_j = Q_{3,j} + 1.5\left(IQR_j\right) \quad (4)$$

Accordingly, an observation is considered an outlier when

$$x_{i,j} < L_j \text{ or } x_{i,j} > U_j \quad (5)$$

The outlier boundaries are computed separately for each modality (soil, climate) to avoid extreme values in one modality affecting the statistical properties of the other modality. All the numerical variables were afterwards standardized following the elimination of outliers, so that all environmental variables were measured on a comparable scale. Min–max normalization is expressed as

$$x_{i,j}^{norm} = \frac{x_{i,j} - x_j^{min}}{x_j^{max} - x_j^{min}} \quad (6)$$

where $x_j^{min}$and x_j^{max}denote the minimum and maximum values of the feature j, respectively, calculated from the training data. The transformation brings the values of the soil and climatic features into the range [0,1] and thus minimizes the impact of the variations between the numerical scales of soil and climate parameters.

3.2.1 Training-Only Class Balancing Using SMOTE

The SMOTE procedure is used for solving the problem of class imbalance just in the training subset. Initially, the original dataset is split into three subsets: 80% for training, 10% for validation and 10% for independent testing. This means that among 12,000 source observations 9,600 observations will be used for training, 1,200 will be used for validation and 1,200 will be used for independent testing. It is important to note that the goal of SMOTE is not the balancing of the independent test subset but the creation of a more balanced dataset in terms of minority and majority class samples for optimizing the model. The validation and test subsets remain intact and are used for model selection and evaluation, accordingly. This means that no new synthetic observations created by SMOTE procedure will be added to these subsets.

Observations might differ among all eight health/disease categories, which could cause the class size variance during learning. In such cases, the soil and climate modalities are preprocessed independently and their training features are merged into one training subset. Then SMOTE is applied to this training subset. For a selected minority-class observation x_i, a synthetic observation is generated using one of its nearest minority-class neighbors x_{nn}as

$$x_{\text{new}} = x_i + \lambda(x_{nn} - x_i) \quad (7)$$

where $x_i$represents an existing minority-class training sample formed from the corresponding soil and climate attributes, x_{nn}denotes a selected nearest minority-class neighbor from the same minority class, and $\lambda$is a random value uniformly distributed between 0 and 1. The synthetic data point is thus on the line segment joining the minority data point and its chosen neighbour in the feature space. With SMOTE applied to the merged training representation, the number of minority-class training samples is augmented while maintaining the relationship between the soil and climate characteristics. This balanced training representation is divided into two separate soil and climate representations and fed to the Soil Branch and Climate Branch for learning modality-specific features. It must be stressed that SMOTE is only performed on the training set, without changing the original distribution of the validation and testing sets.

Following data preprocessing and SMOTE operation only in the training set, the soil and climate features are separately provided to the Dual-Branch Contextual Attention Encoder (DCAE).

3.3. Dual-Branch Contextual Attention Encoder

After the modality-specific pre-processing and training-only SMOTE, balanced soil and climate observations are separately sent to the Dual-Branch Contextual Attention Encoder (DCAE) as illustrated in Figure 2. The goal of the DCAE is to convert the numerical agricultural attributes into learnable feature representations and build relationships between the attributes in each modality in the respective context. Soil and climate conditions are all different environmental factors, so the two arms are independent of each other when performing the encoding operation. This enables the model to map out soil environment and climate environment dependencies without mixing them up prematurely.

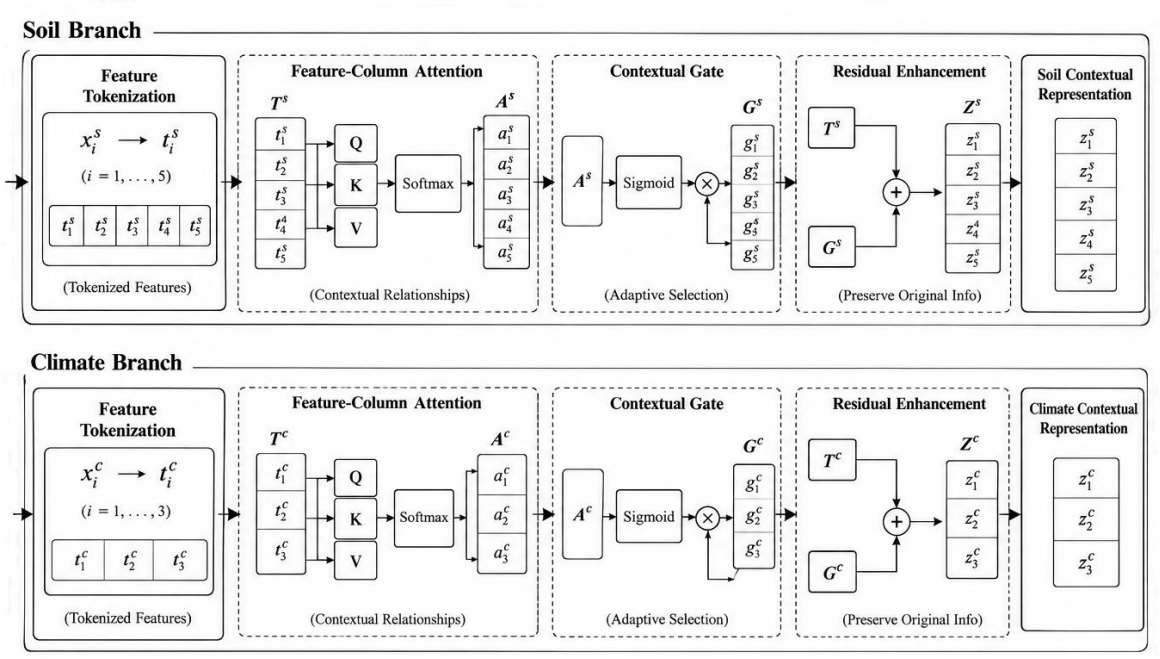


Figure 2. *Dual-Branch Contextual Attention Encoder (DCAE) module. The module contains modality-specific feature tokenization, feature-column attention, contextual gating, and residual enhancement units to learn intra-modal relationships, selectively emphasize contextual information, and preserve informative representations from the soil and climate features.*

Primarily, the numerical attributes are transformed into latent feature tokens through a feature-tokenization operation. For a modality m, where $m \in \{s, c\}$represents soil and climate, respectively, the token representation is obtained as

$$z_j^m = W_j^m x_j^m + b_j^m \tag{8}$$

where x_j^mdenotes the j-th numerical attribute, $W_j^m$and $b_j^m$are learnable projection parameters, and z_j^mrepresents the corresponding latent feature token. The resulting tokens act as a shared representation in which the attention mechanism can work with diverse numerical attributes. Then the tokenized features are fed to feature-column attention to find context-aware relations between the attributes:

$$A^m = Softmax\left(\frac{Q^m (K^m)^T}{\sqrt{d_k}}\right) V^m \tag{9}$$

where Q^m, K^m, and V^mrepresent the query, key, and value projections of the modality-specific feature tokens, respectively, and $d_k$denotes the key dimension. This operation allows the selective attention to other attributes in the same modality. The soil branch can learn contextual relationships between N, P, K, pH, and moisture, for instance, and the climate branch can learn relationships between temperature, humidity, and rainfall. This contextual attention output is then passed through a contextual gate implemented as a sigmoid, and then enhanced with the original representation:

$$Z_{DCAE}^m = \sigma\left(W_g^m A^m + b_g^m\right) \odot A^m + Z^m \tag{10}$$

where Z^mdenotes the original token representation and $\odot$ represents element-wise multiplication. The amount of contextual information that is propagated to the next stage is controlled by the gate, and the residual connection retains useful information from the original representation of the token. This makes that DCAE results in a contextual representation that weaves together the original feature information with the learned intra-modal relationships.

3.4. Adaptive Feature Enhancement

The contextual representations produced by DCAE are then shown to the Adaptive Feature Enhancement (AFE) module (as depicted in Figure 3). AFE's main goal is to identify informative contextual responses for guava classification into health classes and automatically control the subsequent contribution. This is especially significant for different types of soil and climate, because not all soil and climate properties have the same effect under various guava health conditions. AFE assigns different weights to each of its learned responses.

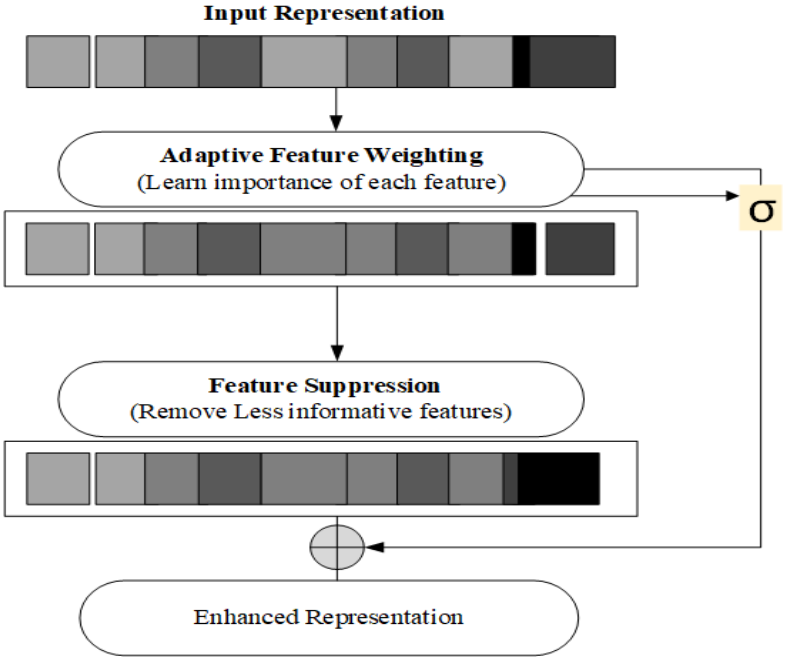


Figure 3. *Adaptive Feature Enhancement (AFE) module. The module employs adaptive feature weighting and feature suppression mechanisms to identify informative responses and reduce less relevant information from the modality-specific soil and climate representations.*

The adaptive feature weights are first generated from the DCAE representation as

$$\alpha^m = \sigma(W_a^m Z_{DCAE}^m + b_a^m) \quad (11)$$

where W_a^m and b_a^m are learnable parameters and α^m represents the importance coefficients assigned to the corresponding feature responses. Higher scores signify greater contributions and lower scores imply less influence from less informative responses. These learned weights are used to multiply the contextual representation, and after the soft thresholding process, this improved representation is fed into the Multi-scale Dependency Learning (MDL) module. The output is then sent back to the RNN, after multiplying with the learned weights and performing the soft-thresholding operation as shown in the diagram:

$$\tilde{Z}^m = S_\tau(\alpha^m \odot Z_{DCAE}^m) \quad (12)$$

where $S_\tau(\cdot)$ denotes the soft-thresholding function and τ represents the threshold parameter. This soft-thresholding operation gradually truncates low responses of features instead of hardcoding a binary threshold. This enables AFE to silence less interesting replies while preserving relevant feature information. Finally, residual enhancement is used for maintaining the contextual information acquired by DCAE:

$$Z_{AFE}^m = \tilde{Z}^m + Z_{DCAE}^m \quad (13)$$

Features presented in the resulting representation are thus enhanced adaptively while presenting less of the weak response. This allows the following module in the MDL to act on a more informative and discriminative representation.

3.5. Multi-scale Dependency Learning

The soil and climate representations are then passed to the Multi-scale Dependency Learning (MDL) module. The dependency learning is performed at the local, medium, and global scale in MDL as illustrated in the proposed architecture. Multiple dependency scales is a motivation because there might be different context levels of relationships between agricultural variables. Fine grain relationships can offer feature interactions at a more local level, broader at a more comprehensive level for environmental dependencies:

$$D_l^m = Conv_l(Z_{AFE}^m),\ D_m^m = Conv_m(Z_{AFE}^m),\ D_g^m = Conv_g(Z_{AFE}^m) \quad (14)$$

where $\mathrm{Conv}_l(\cdot)$, $\mathrm{Conv}_m(\cdot)$, and $\mathrm{Conv}_g(\cdot)$denote the dependency-learning operations associated with the local, medium, and global scales, respectively. The multi-scale responses are subsequently aggregated:

$$D^m = \phi\left(D_l^m \parallel D_m^m \parallel D_g^m\right) \quad (15)$$

where $\parallel$denotes concatenation and $\phi(\cdot)$represents the aggregation/projection operation. This allows information learned at different scales to be jointly represented rather than relying on a single dependency level. The final modality-specific dependency representations are therefore expressed as

$$D_s = MDL(Z_{AFE}^s),\ D_c = MDL(Z_{AFE}^c) \quad (16)$$

where $D_s$and $D_c$denote the resulting soil and climate dependency representations. Through this multi-scale learning process, the model captures complementary fine-grained and broader dependencies within each modality before allowing the two modalities to interact.

3.6. Adaptive Soil–Climate Fusion

The representations of soil and climate dependency that are created are then provided to the Adaptive Soil–Climate Fusion (ASCF) module. The two learned modalities are explicitly integrated through ASCF to establish meaningful cross-modal interactions. Instead of directly concatenating the soil and climate representations, ASCF uses cross-attention to determine relevant information between the modalities. Bidirectional cross-attention is used to obtain the soil-conditioned climate representation and the climate-conditioned soil representation. Bidirectional cross-attention is used to achieve soil-conditioned climate representation and climate-conditioned soil representation. This two-way operation also allows the soil representation to selectively include relevant climate information, and vice-versa allows the climate representation to include relevant soil information.

The cross-modal information is then integrated with the global contextual information gained from both modalities:

$$G = f_{global}(D_s \parallel D_c) \quad (17)$$

where $f_{\mathrm{global}}(\cdot)$represents the global context integration operation. The global context gives a general overview of the state of the environment and is designed to highlight the pairwise cross-modal interactions. Finally, adaptive importance weighting is used to control the contribution of the cross-modal and global representations:

$$Z_F = \beta \odot C + (1-\beta) \odot G\ ,\ \beta = \sigma\left(W_f[C \parallel G] + b_f\right) \quad (18)$$

where $C = C_{s\leftarrow c} \parallel C_{c\leftarrow s}$denotes the combined cross-modal representation and βrepresents the learned importance coefficient. This adaptive mechanism allows the network to dynamically regulate the contribution of complementary information rather than relying on fixed fusion weights. The output $Z_F$therefore represents the unified soil–climate representation used for subsequent refinement.

3.7. Gated Residual Network

ASCF then fuses these representations and sends the fused representation to the Gated Residual Network (GRN). The main intention of GRN is to further improve the fused representation and build a more discriminative representation for classification. ASCF is used to correlate soil and climate data; GRN is the transformation and storage of the fused data before prediction. The fused representation is first subject to a non-linear transformation:

$$H = f(W_h Z_F + b_h) \quad (19)$$

where $W_h$and $b_h$are learnable parameters and $f(\cdot)$denotes the nonlinear transformation. This operation allows the network to learn the higher level, nonlinear relationships in the fused soil–climate representation. A feature-wise gate is then created to control the modified representation:

$$g_r = \sigma(W_r H + b_r) \quad (20)$$

where $g_r$represents the learned gating coefficient. The gate regulates the amount of each transformed response and the information that is useful is retained while the less useful responses are lessened. This gated representation is then fused with the original fused representation by means of the residual connection:

$$Z_R = (g_r \odot H) + Z_F \quad (21)$$

where $Z_R$represents the final refined representation. The residual pathway helps preserve useful information from ASCF, while the gated nonlinear pathway introduces additional discriminative refinement. Thus, GRN provides the final representation used

by the classification head. The refined representation $Z_R$is finally passed to the classification head, which converts the learned representation into predictions for the eight guava health conditions. The proposed architecture contains a fully connected layer and a Softmax layer for the classification head. The fully connected layer converts the refined representation into eight class-specific scores:

$$o = W_c Z_R + b_c \tag{22}$$

where $W_c$and $b_c$denote the learnable parameters of the fully connected layer and $o$represents the resulting eight-dimensional score vector. The output scores are converted into normalized class probabilities using Softmax:

$$P(y = k \mid X) = \frac{exp(o_k)}{\sum_{j=1}^{8} exp(o_j)} \tag{23}$$

where $o_k$represents the score corresponding to the k-th guava health class and $P(y = k \mid X)$denotes its predicted probability. The final predicted class is selected according to the highest-class probability:

$$\hat{y} = arg \max_{k \in \{1,\ldots,8\}} P(y = k \mid X). \tag{24}$$

The eight categories that are predicted include Healthy, Anthracnose, Canker, Rust, Leaf Spot, Fruit Scab, Stylar End Rot, Nutrient Stress.

Algorithm 1: Proposed GuavaFusionFormer Framework

Input: Soil feature matrix X_s, climate feature matrix X_c, ground-truth labels Y, and learnable parameters Θ.

Output: Predicted guava health class $\hat{Y}$.

Procedure:

1. Input the soil feature matrix $X_s$and climate feature matrix X_c.
2. Perform modality-specific preprocessing on $X_s$and $X_c$by applying missing-value imputation, abnormal-observation detection, and normalization using training-set statistics.
3. Apply SMOTE exclusively to the training partitions of the soil and climate modalities, while keeping the validation and testing distributions unchanged.
4. For each modality $m \in \{\text{soil}, \text{climate}\}$:

   a. Transform each numerical attribute into a learnable feature token:

   $$Z^m = \text{FeatureTokenization}(X^m)$$

   b. Generate query, key, and value representations:

   $$Q^m, K^m, V^m = \text{LinearProjection}(Z^m)$$

   c. Compute intra-modal contextual attention:

   $$A^m = \text{Softmax}\left(\frac{Q^m (K^m)^T}{\sqrt{d_k}}\right) V^m$$

   d. Generate the contextual gate:

   $$G^m = \text{Sigmoid}\left(W_g^m A^m + b_g^m\right)$$

   e. Perform gated residual enhancement:

   $$H^m = Z^m + G^m \odot A^m$$

   f. Apply Adaptive Feature Enhancement:

   $$E^m = \text{AFE}(H^m)$$

   g. Apply Multi-scale Dependency Learning:

   $$M^m = \text{MDL}(E^m)$$

5. End the modality-specific processing and retain the learned soil and climate representations.
6. Establish adaptive cross-modal interaction using Adaptive Soil–Climate Fusion:

$$F = \mathrm{ASCF}(M^{soil}, M^{climate})$$

7. Refine the fused representation using the Gated Residual Network:

$$R = \mathrm{GRN}(F)$$

8. Generate eight class-specific scores:

$$S = W_c R + b_c$$

9. Convert the class scores into probabilities:

$$P = \mathrm{Softmax}(S)$$

10. Determine the predicted guava health class:

$$\hat{Y} = \mathrm{argmax}(P)$$

11. If training mode is enabled:

Compute the multi-class cross-entropy loss:

$$L = \mathrm{CrossEntropy}(P, Y)$$

12. Update the learnable parameters Θusing backpropagation and the selected optimization strategy.
13. Return the predicted guava health class $\hat{Y}$.

End Algorithm

4. Results and Discussions

A framework called GuavaFusionFormer was proposed and implemented for intelligent multi-class guava health classification using the Guava Soil–Climate Disease Dataset. The experiments concentrated on the classification ability, stability of training, learned modality-specific representations, adaptive soil–climate fusion and effectiveness of the suggested feature-learning modules. Besides traditional classification metrics, the behavior of the dual-branch contextual encoding, adaptive feature enhancement, multi-scale dependency learning, and adaptive soil–climate fusion mechanisms was explored to gain insight on the contribution of the proposed architecture.

4.1. Experimental Setup

The proposed GuavaFusionFormer framework is applied to Kaggle Notebook environment with Python 3.11 and PyTorch. Model training is done on a single NVIDIA Tesla T4 GPU with 16 GB GDDR6 memory and 4 vCPUs, 29 GB system RAM and CUDA 12.8. The entire Guava Soil–Climate Disease Dataset is first obtained by dealing with missing values, abnormal observations separately for soil and climate modality and normalizing the numerical attributes based on statistics obtained from the training data. Soil/climate information is considered as two modalities and pre-processing operations are carried out separately on the two data branches. The processed data is then split into 80% train, 10% valid and 10% test subsets with the distribution of the eight guava health classes remaining the same. SMOTE technique only works on the training partition, while the validation and testing partition are left as they are.

The feature-tokenisation, feature-column attention, contextual gating, residual enhancement, AFE, MDL, ASCF, GRN and fully connected classification layers are trained on the training set. In the training process, the model will learn the nonlinear relationships among soil attributes with each other and also learn the nonlinear relationship of climate attributes with each other without any external influence, and in the ASCF module, it will establish the cross-modal relationship. The validation set is employed to keep an eye on the model convergence and to choose the best model configuration during the training process, and the testing set is kept completely independent of model training and is only used for model performance evaluation at the end. The final model is tested on the unseen testing samples, and the following performance measures are used: Accuracy, Precision, Recall, F1-Score, Specificity, Cohen's Kappa and Confusion Matrix.

4.2. Hyperparameter Configuration

Table 2 shows the main hyperparameters involved in training GuavaFusionFormer. The model uses the Adam optimizer which has a controlled learning rate and weight decay so that the optimization process is stable and to prevent overfitting. For this classification, the objective is guava health classification in eight classes, hence the objective function used here is cross-entropy. The mini-batch training strategy is used to make full use of the GPU memory.

Table 2. *Hyperparameter Configuration of GuavaFusionFormer*

Hyperparameter	Value
Initial Learning Rate	0.001
Maximum Learning Rate	0.001
Weight Decay	0.0001
Batch Size	128
Number of Epochs	50
Loss Function	Cross-Entropy Loss
Train/Test Split	80% / 20%
Dropout	0.15
Classifier Activation	GELU

The learning rate is initialized at 0.001 and maximum learning rate is 0.001, weight decay is 0.0001, used to support stable optimization to prevent overfitting. The training samples are processed efficiently via mini-batch size of 128 and the model is trained for 50 epochs. The objective function used for the task of 8-class guava health classification is Cross-Entropy Loss. The data are split into training (80%) and testing (20%). We apply dropout with 0.15 to enhance the generalization of the model and employ GELU as the classifier activation function for feature transformation, thus adding nonlinearity.

4.3. Dataset Distribution and Exploratory Analysis

The distribution of the five soil (N, P, K, pH and Moisture) and three climate (Temperature, Humidity and Rainfall) attributes over the eight guava health and disease classes is shown in figure 4. Density plots illustrate the distribution of the numerical characteristics across the disease conditions and rainfall boxplots are the summary plots of the median, spread and potential outliers across classes. These distributions are useful for getting a first impression of the variability and discriminative properties of the soil-climate data.

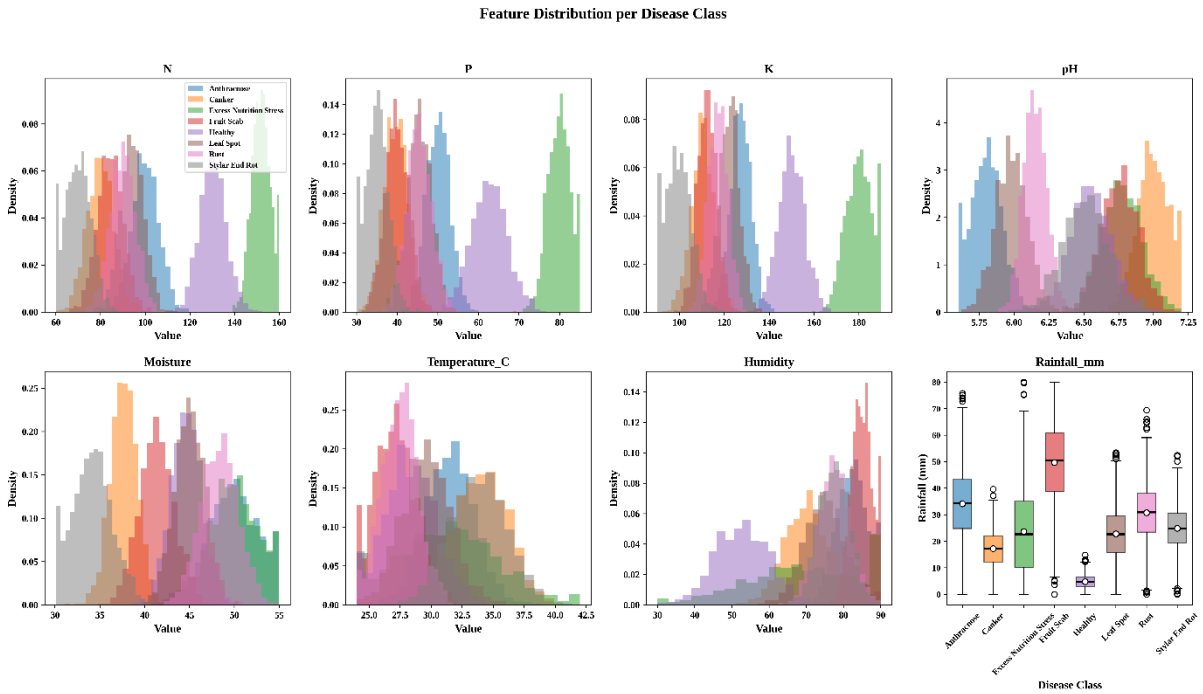


Figure 4. Feature Distribution Across Guava Disease Classes

The feature distributions show some significant class-wise shifts for N, P, K, pH and Moisture, which suggest that the soil conditions are different for various health classes. The other two variables, Temperature and Humidity, also have significant

distributional overlap, with some classes having clearly defined concentration ranges. Rainfall shows variation in median values and dispersion between classes, and a few values seem to be potential outliers. The results of the observed variation and partial overlap within the classes suggest that a learning model is needed which can describe the complex nonlinear relationships between soil and climate attributes.

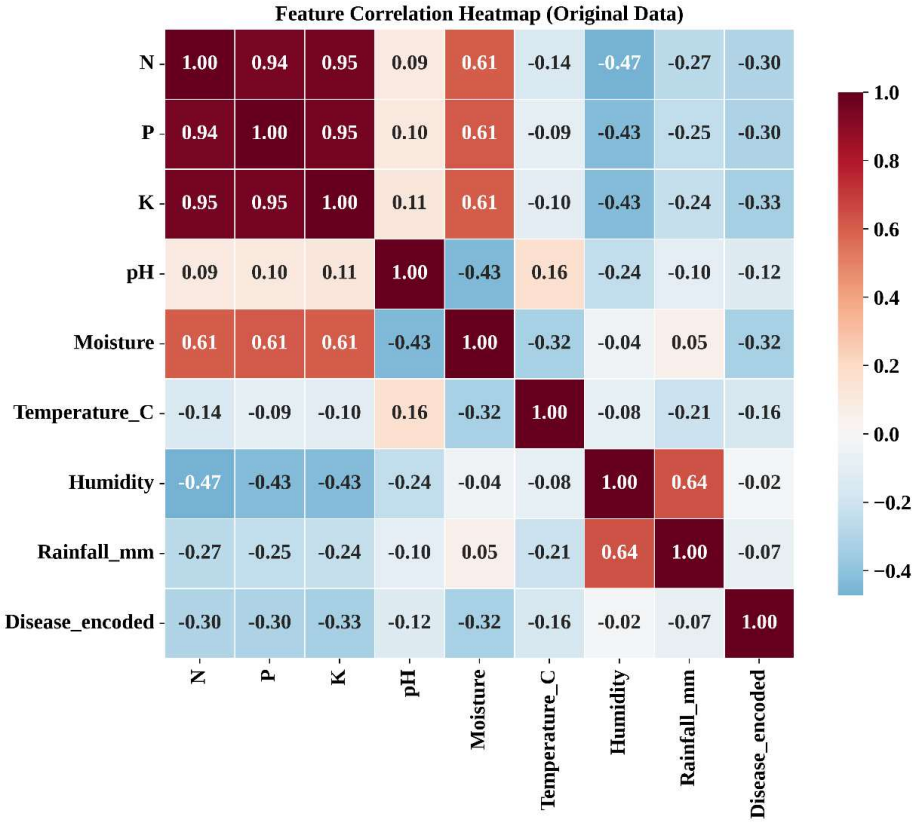
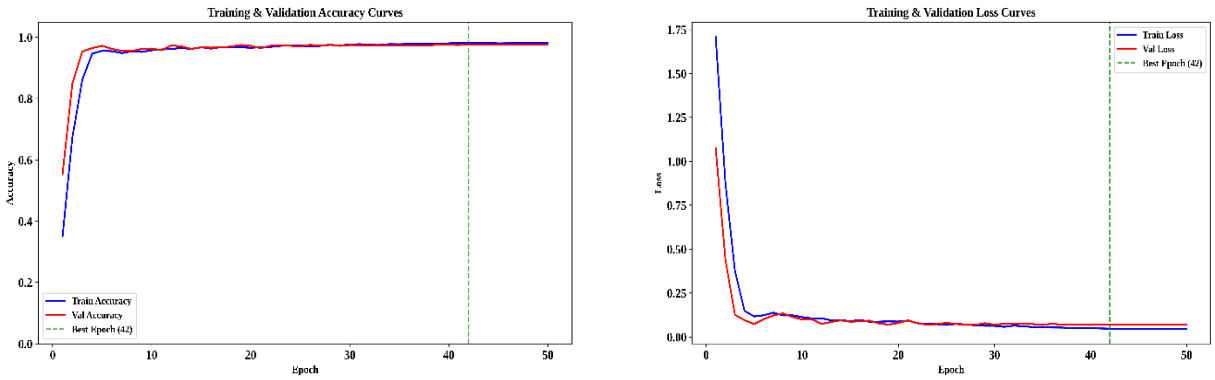


Figure 5. *Feature Correlation Heatmap of the Original Guava Soil–Climate Dataset*

Figure 5 shows the correlation heatmap for the five soil attributes, three climate attributes and the encoded disease attribute. The analysis allows to estimate the level of relation between the environmental variables, and to detect possible relations within and between soil and climate modalities. The relationships are helpful for the design of the modality-specific and cross-modal feature-learning stages of GuavaFusionFormer. The correlation between N, P, and K is very high (0.94–0.95), and moderate (0.61) in each soil moisture. Humidity and rainfall are most positively correlated of all the climate variables at 0.64. The disease-encoded variable is negatively correlated with N (−0.30), P (−0.30), K (−0.33), and moisture (−0.32) and relatively weak in the negative direction with temperature (−0.16), humidity (−0.02), and rainfall (−0.07). The different strengths of correlation and cross-modal relationships suggest a model that can grasp both the modality-specific relationships and adapt to interactions between soil and climate properties.

4.4. Training and Validation Performance

The training and validation performance is evaluated with respect to learning behavior, the GuavaFusionFormer model convergence, and the generalization ability of the proposed model is analyzed. The training and validation accuracy and loss curves are checked throughout the epochs to assess the stability of training and validation, convergence behavior, and possible signs of overfitting.



(a) (b)

Figure 6: *Training and validation performance of GuavaFusionFormer: (a)Accuracy and (b) Loss*

To evaluate the learning behavior, shown in Figure 6, the convergence and generalization capability of the proposed GuavaFusionFormer framework and the training and validation performance, are analyzed. Training and validation accuracy and loss curves are analyzed over epochs to check optimization stability, convergence and the possibility of overfitting.

4.5. Confusion Matrix Analysis

The prediction capability of GuavaFusionFormer is assessed using confusion matrix analysis for eight health and disease classes of guavas. It gives detailed information about samples which are correctly classified and misclassified, especially at the health condition with similar soil and climatic properties.

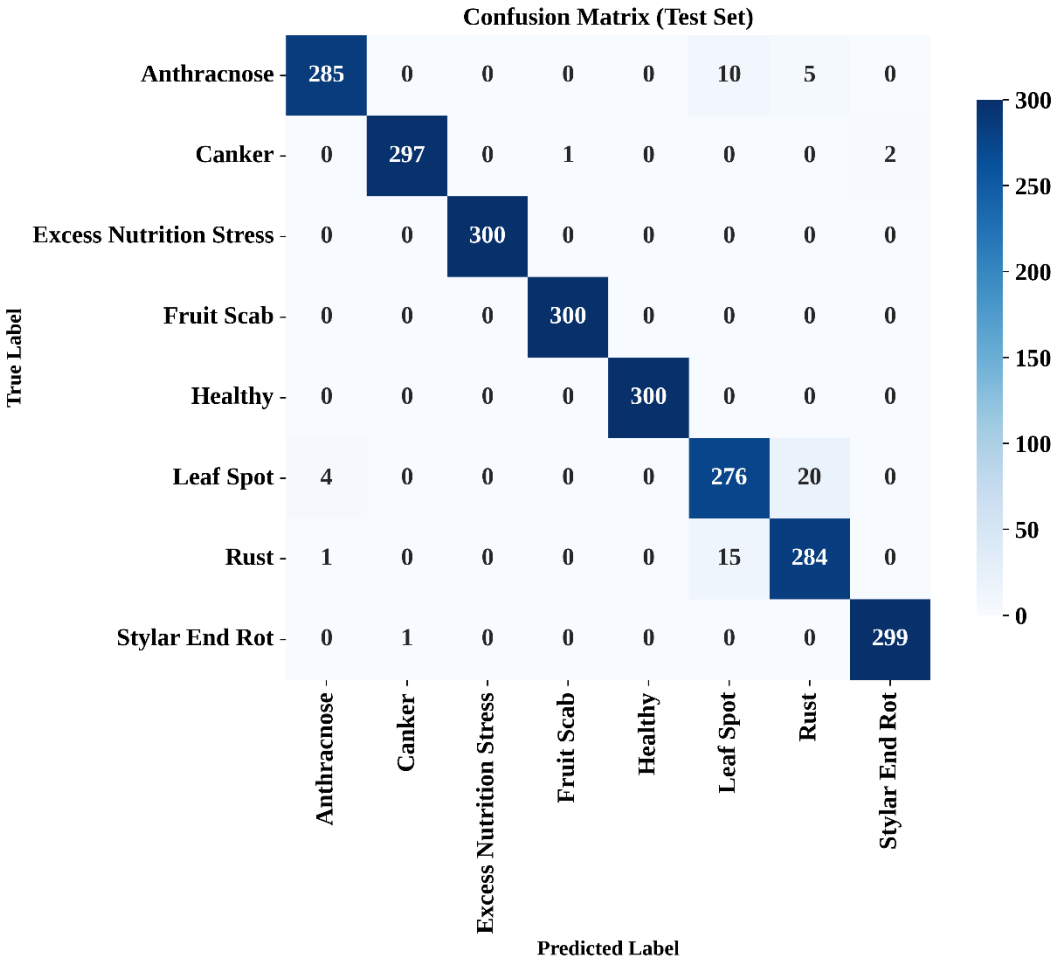


Figure 7. *Confusion Matrix analysis*

The prediction accuracy of the proposed GuavaFusionFormer on each class in the eight-guava health and disease category is illustrated by the Confusion matrix in Figure 7. The confusion matrix highlights that the proposed model exhibits excellent

classification capability among all the eight classes of guava disease conditions. From 2,400 test samples, 2,341 samples are classified correctly, giving rise to an overall accuracy rate of around 97.54%. Nutrition Stress, Fruit Scab, and Healthy are classified perfectly with 300/300 test samples classified correctly for each category. In the case of Canker, 297 samples are classified correctly, out of which only one sample is wrongly classified into Fruit Scab and another two into Stylar End Rot. Stylar End Rot also exhibits good performance with 299 samples being classified correctly and one being classified wrongly as Canker. As far as Anthracnose is concerned, 285 samples have been classified correctly, but 10 samples have been classified wrongly as Leaf Spot and 5 samples as Rust. In the case of Leaf Spot, 276 samples are classified correctly while 4 samples are incorrectly classified as Anthracnose and 20 samples as Rust. Likewise, 284 samples are classified correctly under Rust category, while 1 sample is misclassified as Anthracnose and 15 samples as Leaf Spot.

4.6. Precision, Recall, F1-Score, and Specificity Analysis

For each of the eight classes, the precision, recall, and F1-Score are evaluated to determine the reliability of class-wise predictions and balanced classification performance of GuavaFusionFormer. The class-wise prediction reliability and balanced classification performance of GuavaFusionFormer in the eight-guava health and disease categories is analysed with precision, recall and F1-Score. Table 3 shows that GuavaFusionFormer has achieved class-wise precision, Recall, F1-Score, and ROC-AUC Performance.

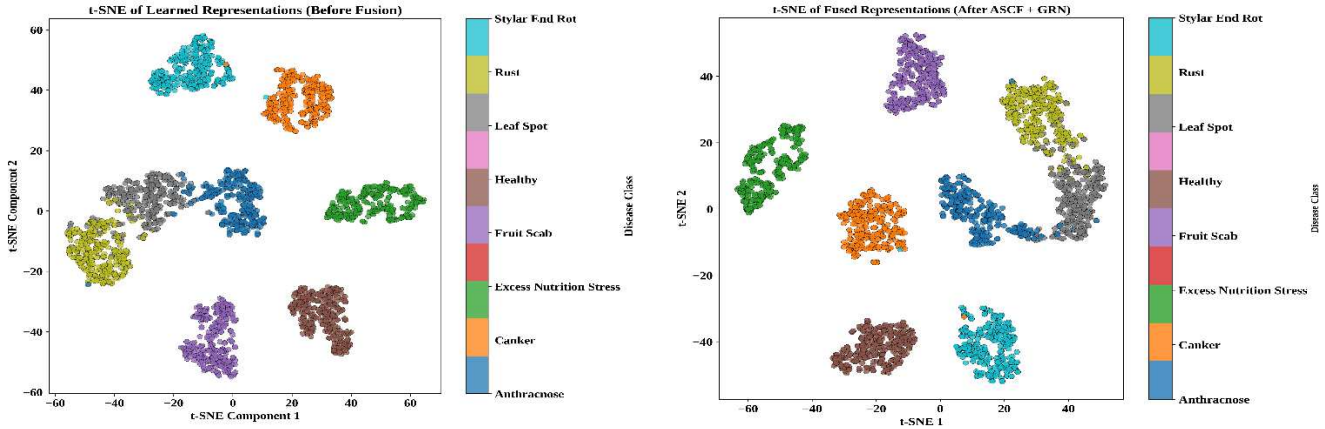
Table 3. *Class-wise Precision, Recall, F1-Score, and ROC-AUC Performance of GuavaFusionFormer*

Class	Precision	Recall	F1-Score	Specificity	NPV	FNR	FPR
Anthracnose	0.982759	0.95	0.966102	0.997619	0.992891	0.05	0.002381
Canker	0.996644	0.99	0.993311	0.999524	0.998573	0.01	0.000476
Nutrition Stress	1	1	1	1	1	0	0
Fruit Scab	0.996678	1	0.998336	0.999524	1	0	0.000476
Healthy	1	1	1	1	1	0	0
Leaf Spot	0.916944	0.92	0.918469	0.988095	0.988566	0.08	0.011905
Rust	0.919094	0.946667	0.932677	0.988095	0.992348	0.053333	0.011905
Stylar End Rot	0.993355	0.996667	0.995008	0.999048	0.999524	0.003333	0.000952

The results of class-wise show outstanding results for all eight classes of health of guava. For Nutrition Stress and Healthy 100% is reached for all parameters, which means perfect classification. For Canker, Fruit Scab, and Stylar End Rot, 0.9933%, 0.9983%, and 0.9950% F1-scores are reached, respectively. For Anthracnose, Leaf Spot, and Rust, 0.9661%, 0.9185%, and 0.9327% are reached, respectively, with some misclassifications. High specificity and low FPR values mean good discriminative properties and reliable prediction of class.

4.7. Learned Feature Representation Analysis

Visualizations of the learned features' representation from the proposed GuavaFusionFormer were done through t-SNE to explore the separation of the eight classes of guava health and diseases based on the learned feature space. As seen in Figure 8, t-SNE visualization has revealed clear clusters in the classes which imply that the learned fusion features have captured discriminative soil-climate features of the guava health classification.



(a) (b)

Figure 8. *t-SNE Visualization of Learned Feature Representations: (a) Before Adaptive Soil-Climate Fusion; (b) After Adaptive Soil-Climate Fusion*

The t-SNE visualizations show the latent feature representations learned by GuavaFusionFormer, both before and after the ASCF and GRN stages. The pre-fusion representations in Figure. 8(a) form relatively distinct clusters of the eight guava health and disease classes, suggesting the capture of meaningful class-related information during the modality-specific feature-learning stages. The fused representations of ASCF and GRN in Figure. 8(b) show clearly separated class clusters with small, compact within-class distributions, and limited overlap between different classes. The proposed cross-modal fusion and post-fusion refinement stages in the proposed framework result in more discriminative representations of multi-class guava health classification as revealed by the enhanced organization of latent space after fusion.

4.8. SHAP-Based Feature Importance Analysis

The SHAP analysis evaluates the contribution of individual soil and climate attributes to the predictions generated by GuavaFusionFormer. Mean absolute SHAP values are used to compare the relative importance of the features across the eight-guava health and disease classes.

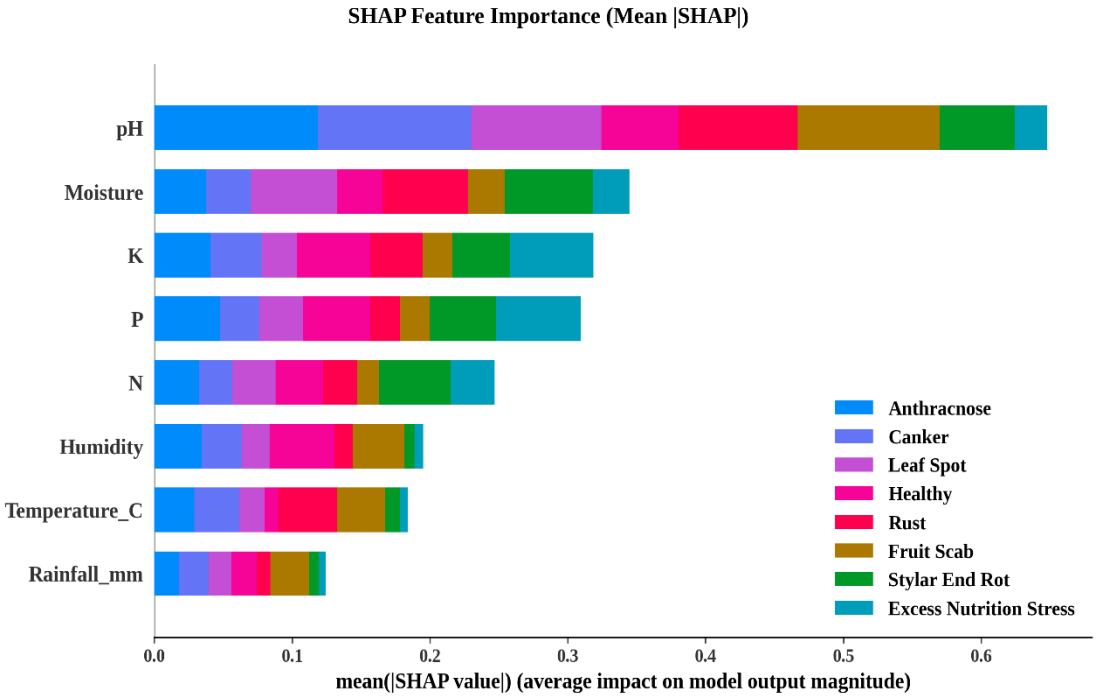
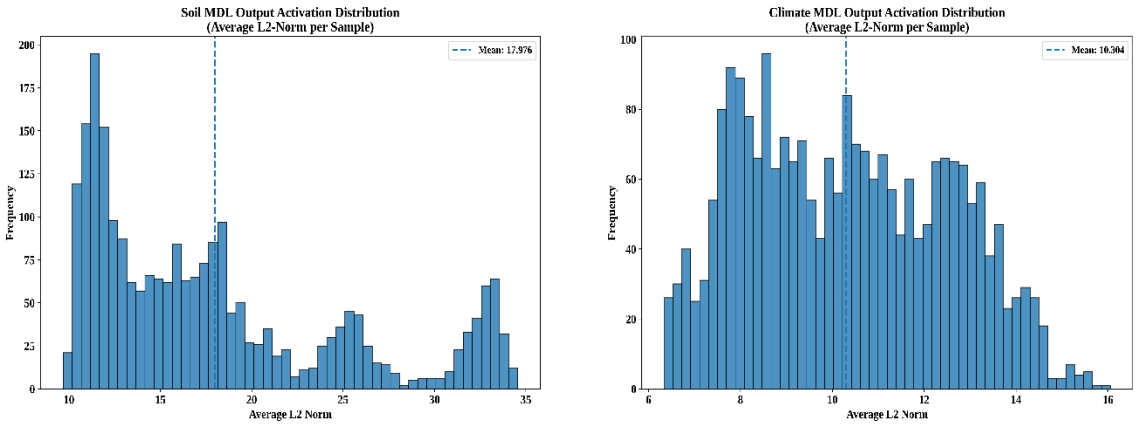


Figure 9. *SHAP-Based Feature Importance of Soil and Climate Attributes for Guava Health Classification*

The overall mean absolute SHAP value for pH is the highest (as shown in Figure 9), meaning that this feature contributes the most to the model output among the features considered. Moisture, K, P and N are next in importance, with significant contributions from soil modality. Humidity and Temperature_C moderately contribute to the climate attributes, while Rainfall_mm has the smallest absolute contribution to the climate SHAP. The class-wise stacked contributions also show that the importance of the features can differ by disease class, which demonstrates that GuavaFusionFormer can make use of the different soil and climate parameters depending on the disease class predicted.

4.9. Soil–Climate Multi-Scale Dependency Activation Analysis

Soil-Climate Multi-Scale Dependency Activation Analysis examines the activation patterns learned by the MDL module from the soil and climate branches. It shows how strongly the model represents dependency information at different scales within each modality.



(a) (b)

Figure 10. *Distribution of Multi-Scale Dependency Learning (MDL) Output Activations: (a) Soil Branch; (b) Climate Branch*

The soil MDL outputs in Figure. 10(a) are broader, with an activation range of about 9 to 35, and with a mean L2-norm of 17.976. The distribution possesses a quite high concentration of samples at the lower magnitude of the activation, with several wider regions of activation. Figure. 10(b) indicates that the climate MDL outputs are located within a smaller range of about 7–16, with a mean L2-norm of 10.304, compared to the other data. There are different activation magnitudes and distributions across the two branches, suggesting that the MDL module is not learning representations of the dependency of soil and climate information, but rather modality-specific representations.

4.10. Overall Classification Performance Analysis

Overall classification performance is used to evaluate the ability of GuavaFusionFormer to predict the overall performance of the fusion model on the eight-guava health and disease categories listed in Table 4. The accuracy, reliability and classification error features of the model are evaluated based on multiple performance measures.

Table 4. *Overall classification Performance of the GuavaFusionFormer*

Metric	**Value**
Accuracy	0.9754
Precision (Macro)	0.9757
Recall (Macro)	0.9754
F1-Score (Macro)	0.9755
Specificity (Macro)	0.9965
NPV (Macro)	0.9965
FNR (Macro)	0.0246
FPR (Macro)	0.0035
Specificity (Micro)	0.9965
NPV (Micro)	0.9965
FNR (Micro)	0.0246
FPR (Micro)	0.0035

The model gives an accuracy of 0.9754%, which is an indication that the classifier performs well in general. The macro precision, recall, and F1 score at 0.9757%, 0.9754%, and 0.9755%, respectively, are an indication of balanced performance for the eight categories. The model scores 0.9965% macro specificity and NPV, which means that there is a very small likelihood of having any false positives and negatives. This is also evidenced by the FNR of 0.0246% and FPR of 0.0035%.

4.11. Ablation Study

The ablation study is performed to analyze the contribution of major components in the proposed GuavaFusionFormer framework as provided in Table 5. The full model is compared with reduced variants, where one module is removed at a time, and the rest of the model architecture and experimental conditions are unchanged. The analysis includes the following metrics

to understand the impact of modality-specific contextual encoding, adaptive feature enhancement, multi-scale dependency learning, adaptive soil–climate fusion, and post-fusion gated refinement: Accuracy, Precision, Recall and F1-Score.

Table 5. *Ablation study of the GuavaFusionFormer framework*

Model Variant	Accuracy	Precision	Recall	F1-Score
without DCAE	0.9423	0.9465	0.9444	0.9488
without AFE	0.9512	0.9585	0.9554	0.9552
without MDL	0.9626	0.9656	0.9658	0.9652
without ASCF	0.9378	0.9415	0.9315	0.9345
without GRN	0.9656	0.9612	0.9611	0.9695
Complete GuavaFusionFormer	0.9754	0.9757	0.9754	0.9755

The experimental results for ablation show that the entire GuavaFusionFormer performs best with 0.9754% accuracy, 0.9757% precision, 0.9754% recall and 0.9755% F1-score. However, the exclusion of single module lowers the performance with the greatest drop in the case of ASCF, followed by DCAE and AFE, showing their importance to multimodal representations and fusion. In general, the experiment results have shown that the combination of DCAE, AFE, MDL, ASCF, and GRN modules affects the performance of the proposed framework.

4.12. Comparative Analysis with Existing Deep Learning Methods

A comparative experiment was carried out with other deep learning methods with the same Guava Soil-Climate Disease dataset to further validate the effectiveness of the proposed GuavaFusionFormer framework. The methods selected span conventional deep neural, 1D-CNN, recurrent, attention-based, and tree-based learning architectures: DNN, 1D-CNN, LSTM, GRU, BiLSTM, Transformer, TabTransformer, Random Forest (RF), Gradient Boosting (GB), XGBoost, TabNet, and FT-Transformer. Multiclass classification capability was evaluated on eight guava health and disease categories by comparison using Accuracy, Precision, Recall and F1-Score. Comparative results are shown in Table 6.

Table 6. *Comparative performance of GuavaFusionFormer with existing deep learning methods*

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
DNN	0.9315	0.9321	0.9304	0.9310
1D-CNN	0.9442	0.9450	0.9431	0.9439
LSTM	0.9518	0.9526	0.9507	0.9515
GRU	0.9556	0.9562	0.9548	0.9554
BiLSTM	0.9589	0.9595	0.9581	0.9587
Transformer	0.9647	0.9653	0.9639	0.9646
TabTransformer	0.9698	0.9702	0.9691	0.9696
Random Forest (RF)	0.9588	0.9585	0.9111	0.9124
Gradient Boosting (GB)	0.9444	0.9325	0.9487	0.9152
TabNet	0.9257	0.9214	0.9222	0.9365
FT-Transformer	0.9254	0.9147	0.9124	0.9441
GuavaFusionFormer	0.9754	0.9757	0.9754	0.9755

GuavaFusionFormer had the best results in terms of Accuracy (97.54%), Precision (97.57%), Recall (97.54%) and F1-Score (97.55%). Of the other baseline models, TabTransformer and Transformer performed relatively well, while DNN and 1D-CNN had relatively poor results. The RF, GB, TabNet, and FT-Transformer were implemented independently and performed relatively poorly. Overall, it is evident that GuavaFusionFormer shows consistent classification ability in all four measures of performance.

4.13 Limitations and Future Directions

The proposed study is bound by limitations related to the availability, size, and diversity of the guava soil-climate data set used for evaluation purposes. Limited number of observations for each health and disease class will restrict the extent to which the learning model would be able to capture all the variations in guava cultivation conditions. Moreover, the current data set comprises of five soil features and three climate features while many other environmental features, physiological attributes of plants, and field cultivation attributes are left out. The current evaluation is done using only structured soil-climate observations from the current data set and might not be capturing all variations occurring in different geographic regions, at different seasons, with different cultivation methods, and environmental settings. Further studies could focus on building larger and more varied guava soil-climate data sets covering various geographic regions, different seasons, growth stages, and various cultivation settings. The GuavaFusionFormer approach could be extended by including other environmental features and plant-related attributes along with temporal observations and heterogenous sensor readings.

5. Conclusion

This proposed framework of GuavaFusionFormer provides an intelligent method of guava health condition classification based on heterogeneous soil and climate data and shows its ability to learn different context from environmental data for multi-class health condition classification. This framework handles soil and climate features independently, learns different context representations, identifies different scales dependencies, learns different cross-modality relations, and refines features after fusion for eight-class guava health condition classification.

The performance of the proposed framework is validated through the following results:

- Accurate classification was attained by GuavaFusionFormer with 97.54% accuracy for all eight-guava health and disease conditions.
- The model had an accuracy of 97.57%, 97.54%, and 97.55% in macro-average precision, macro-average recall, and macro-average F1-score, respectively, suggesting consistent classification for all the classes considered.
- Accuracy of 99.65% in macro-average specificity and FPR of 0.35% demonstrate the high capacity of correctly classifying non-target classes at the same time with a very low false-positive rate.
- The incorporation of DCAE, AFE, MDL, ASCF, and GRN enables contextual modality-specific learning, adaptive feature enhancement, multi-scale dependency learning, adaptive soil-climate interaction, and post-fusion representation refinement, respectively.
- Insightful analysis of learned feature separability, relative feature contribution, and modality-specific representations can be obtained using t-SNE, SHAP, and MDL activation analysis.
- Conclusively, GuavaFusionFormer is highly efficient in fusing soil and climatic data of different modalities for guava health classification across eight categories. The dual-branch model retains the specific properties of different modalities as well as the ability to adaptively interact between the soil and climatic data and gradually refine the representation.

Declaration of Competing Interests

The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Dataset Availability:

The dataset used in this study, Guava Soil–Climate Disease Dataset, is publicly available on Kaggle and can be accessed at: https://www.kaggle.com/datasets/archanadirgule45/guva-soil-w432.

Use of Generative Artificial Intelligence

The authors declare that no generative artificial intelligence tools were used in the preparation of this manuscript.

Consent to Publish declaration

Not Applicable

Clinical trial

Not Applicable

Ethics and Consent to Participate declarations

Not Applicable.

References

[1] Alrashdi, I., Sharawi, M., Ali, A. M., Sallam, K. M., Arain, B., & Abdel-Basset, M. (2026). Utilizing deep learning models for early detection and classification of fruit diseases: towards sustainable agriculture and enhanced food quality. Scientific Reports, 16(1), 8167.

[2] Mazur, P., Gozdowski, D., & Wójcik-Gront, E. (2022). Soil electrical conductivity and satellite-derived vegetation indices for evaluation of phosphorus, potassium and magnesium content, pH, and delineation of within-field management zones. Agriculture, 12(6), 883.

[3] Roussin-Léveillée, C., Rossi, C. A., Castroverde, C. D. M., & Moffett, P. (2024). The plant disease triangle facing climate change: a molecular perspective. Trends in Plant Science, 29(8), 895-914.

[4] Ahmed, S., Jhatial, M. J., Bashir, F., Usman, A., Ali, I., & Memon, S. (2025). Multi-Class Guava Classification Using Convolutional Neural Networks for Precision Agriculture. International Journal of Artificial Intelligence & Mathematical Sciences, 4(2), 40-60.

[5] Biswas, J., Islam, M., Bangabashi, M. M., Akter, M., Nishi, T. S., Sheikh, M. K., ... & Anwar, M. M. (2026). GuavaVision AI: An Explainable Deep Learning Framework for Automated Classification, Lesion Localization, and Segmentation of Guava Diseases. bioRxiv, 2026-06.

[6] Cravero, A., Pardo, S., Galeas, P., López Fenner, J., & Caniupán, M. (2022). Data type and data sources for agricultural big data and machine learning. Sustainability, 14(23), 16131.

[7] Zeng, Y., Verhoef, A., Vereecken, H., Ben-Dor, E., Veldkamp, T., Shaw, L., ... & Su, Z. (2025). Monitoring and modeling the soil-plant system toward understanding soil health. *Reviews of Geophysics*, *63*(1).

[8] Han, S., & Wang, H. (2025). Application of deep learning technology in monitoring plant attribute changes. Sustainability, 17(17), 7602.

[9] Roman, A., Rahman, M. M., Haider, S. A., Akram, T., & Naqvi, S. R. (2025). Integrating feature selection and deep learning: a hybrid approach for smart agriculture applications. Algorithms, 18(4), 222.

[10] Das, S., Pattanayak, S., & Behera, P. R. (2022). Application of machine learning: a recent advancement in plant diseases detection. Journal of Plant Protection Research, 122-135.

[11] Afzaal, H., Rude, D., Farooque, A. A., Randhawa, G. S., Schumann, A. W., & Krouglicof, N. (2025). Improved crop row detection by employing attention-based vision transformers and convolutional neural networks with integrated depth modeling for precise spatial accuracy. Smart Agricultural Technology, 11, 100934.

[12] Li, H., Song, Y., Wang, Z., Li, M., & Yang, W. (2024). Development of an online prediction system for soil organic matter and soil moisture content based on multi-modal fusion. Computers and Electronics in Agriculture, 227, 109514.

[13] Dolatabadian, A., Neik, T. X., Danilevicz, M. F., Upadhyaya, S. R., Batley, J., & Edwards, D. (2025). Image-based crop disease detection using machine learning. Plant Pathology, 74(1), 18-38.

[14] Xu, J., Wen, S., Su, Y., Liang, H., Mo, L., Chen, Y., ... & Liu, K. (2026). Ascorbic acid and melatonin synergistically maintain postharvest guava quality by regulating cell wall, antioxidant, and transcriptomic pathways. npj Science of Food.

[15] Ali, A., Das, B., Dhakar, M. K., Naik, S. K., Patel, V. B., Mishra, G. P., ... & Bhatt, B. P. (2024). Enhancing soil health and fruit yield through Tephrosia biomass mulching in rainfed guava (Psidium guajava L.) orchards. Scientific Reports, 14(1), 13913.

[16] Deantonio-Florido, L. Y., Bautista-Montealegre, L. G., Cardona, W. A., Lesmes-Suárez, J. C., Monroy-Cárdenas, D. M., Ospina-Parra, C. E., & Salinas-Velandia, D. A. (2025). Examining Heavy Metals in Guava: Soil Quality and Climatic Effects. Journal of Soil Science and Plant Nutrition, 25(3), 6001-6013.

[17] Preet, M. S., Kumar, R., Valipour, M., Singh, V. P., Singh, A. K., Iqbal, R., ... & Sabagh, A. E. (2022). Soil nutrient status and morphometric responses of guava under drip irrigation and high-tech horticultural techniques for sustainable farming. Hydrology, 9(9), 151.

[18] Kılcı, O., & Koklu, M. (2025). Guava fruit disease classification using deep learning and machine learning models. Research in Agricultural Sciences, 56(3), 217-226.

[19] Doutoum, A., Eryiğit, R., & TUĞRUL, B. (2023). Classification of guava leaf disease using deep learning. World Scientific and Engineering Academy and Society (WSEAS), 20(38).

[20] Pathmanaban, P., Gnanavel, B. K., & Anandan, S. S. (2022). Guava fruit (Psidium guajava) damage and disease detection using deep convolutional neural networks and thermal imaging. The Imaging Science Journal, 70(2), 102-116.

[21] Ahmed, M., Ahmed, F., Naz, N. S., Mazhar, T., Khan, M. A., Khan, M. A., ... & Abbas, M. (2025). Automated Guava Disease Detection Using Transfer Learning With ResNet-101. Food Science & Nutrition, 13(12), e71348.

[22] Mustak Un Nobi, M., Rifat, M., Mridha, M. F., Alfarhood, S., Safran, M., & Che, D. (2023). Gld-det: Guava leaf disease detection in real-time using lightweight deep learning approach based on mobilenet. Agronomy, 13(9), 2240.

[23] https://www.kaggle.com/datasets/archanadirgule45/guva-soil-w432