Explainable AI Based Wheat Crop Monitoring Using Vegetation Indices, Machine Learning and LSTM-Based Temporal Forecasting

Vijaya Ahire[1], Dr. Vijaya Musande[2]

[1,2]Computer Science & Engineering, JNEC, MGM University, Chhatrapati Sambhajinagar, Maharashtra, India.
Email ID: vahire@mgmu.ac.in[1], vmusande@mgmu.ac.in[2]

Abstract
Accurate crop health monitoring is critical to increase the productivity in agriculture and for precision farming. This research introduces a machine learning and deep learning-based model explainable system for monitoring wheat crops. Twenty-five Sentinel-2A and Sentinel-2B images were acquired over three wheat fields, 8-acre, 2.5-acre and 2-acre in extent. Vegetation indices were extracted after processing using AOI clipping and cloud masking, that is NDVI, NDRE, GNDVI, EVI, MCARI and CIgreen. In the case where no labels exist, unsupervised K-Means clustering methods were used to create the crop health classes. The classes were then generated and classified using Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and XGBoost Models. Fold cross validation method was used and the results of five-fold cross validation showed that SVM gave the best classification accuracy of 92.0%, and Random Forest and XGBoost gave 90.67%. SHAP analysis determined that the most influential features are the NDVI, NDRE, MCARI, and Chlorophyll. Moreover, the 100 epochs-trained LSTM model yielded an MAE of 0.1797 in its role of NDVI prediction. It was seen that the proposed combination of clustering, explainable machine learning, and temporal deep learning shows its effectiveness to assess health and predict performance of wheat crop.

Keywords: Explainable AI, SHAP, LSTM, ML

1. Introduction

Wheat is one of the most significant cereal crops all over the world and one of the most critical food species in terms of world food security [1]. Monitoring of the crop health within a suitable period of time is of critical importance to ensure crop productivity, minimise crop losses and precision agriculture practices [2]. Traditional field monitoring approaches can be time-consuming, intensive and too impractical to apply in vast farming areas. Vegetation indices give an efficient alternative that can monitor constantly the crop conditions from satellite imagery [3].

The most common vegetation indices used to evaluate plant photosynthetic activity, biomass, and the concentration of chlorophylls or crop vigor are Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference Red Edge Index (NDRE), Enhanced Vegetation Index (EVI), Modified Chlorophyll Absorption Ratio Index (MCARI), CIgreen and Chlorophyll. Interpreting multiple vegetation indices simultaneously and retrieve meaningful information about crop health is, however, challenging [4].

With the development of machine learning and deep learning in recent years, multi-temporal remote sensing data are able to be analyzed automatically. However, most of these studies use health labels predetermined in the laboratory, which can cause bias and extensive field observations [5]. However, to overcome this constraint, the present study introduces an unsupervised-supervised hybrid model which brings K-Means clustering together with machine learning classification, SHAP based explainability and Long Short-Term Memory (LSTM) forecasting for monitoring wheat crops in a field with varying sizes. The proposed framework auto-classifies the health of multiple crops using vegetation indices, compares the performance of various machine learning classifiers, extracts the key spectral indices with the help of explainable AI and predicts the future vegetation dynamics using temporal deep learning models.

2. Literature Review

For precision agriculture, use of remote sensing based estimation of wheat biophysical parameters has become a matter of much interest. In a study done in southern Iran, interested vegetation indices (VI) derived from Sentinel-2 data were combined with machine learning algorithms, such as SVM, ANN, and DNN, to estimate leaf area index (LAI), leaf dry weight (LDW), specific leaf area (SLA), and leaf specific weight (SLW) in wheat under various management practices. Results showed that the DNN model achieved better performance compared to the other models by estimating LAI with R = 0.80, RMSE = 1.19 and MAE = 0.98 when tested in a validation dataset. NDVI had the highest correlation with wheat leaves-based parameter, and proved the efficiency of deep learning approach with Sentinel-2 image in monitoring plant physiology parameters [6].

In order to enhance large-scale forecasting of crop yield, researchers developed multi-branch deep learning framework called DeepAgroNet which used met, sat, and soil data to predict winter wheat crop yield for the region of southern Pakistan. The CNN, RNN, and ANN models were used for the framework and were trained on yield data for the period 2017 to 2022. Of all the models considered, CNN's model showed the best prediction ability, an R2 of 0.77, and a

forecast accuracy of 98% one month prior to harvest. By combining multi-source data using the deep learning framework, the researchers showed that the accuracy of the yield prediction can be boosted and it could help sustainable agriculture management practice [7].
In another study, they tried to prediction the yield of winter wheat based on machine learning and UAV multispectral imagery. Six models were developed by using 16 vegetation indices extracted at different phases (heading, Flowering and Filling phases). The results showed that performance of the RF was best in comparison to traditional machine learning algorithms, and prediction accuracy and stability of the 1D-CNN model was highest in all input variable combinations. Most of the prediction errors of the CNN model were less than ±0.1 $t{\cdot}ha^{-1}$, revealing its ability to well describe the nonlinearity between the vegetation indices and the level of crop yields [8].

3. Study Area

Based on this, three wheat-growing areas were selected to obtain information on the diversity of crop size and condition in wheat fields. The largest of the three regions is spread in Harsul region located at 19.929420° N, 75.353233° E, under Chattrapati Sambhajinagar area of Maharashtra and is about 8 acres in area. This site has typical semi-arid climate where wheat cultivation is affected by rainfall and irrigation. Moreover, two smaller areas (2.5 acres and 2 acres as shown in figure 1) were selected for evaluation and these areas are located in Jatwada located at 19.9555° N, 75.2849° E, in Maharashtra. Variation in these in terms of fields sizes, farming practices and crop variability.
In the three regions the cropland was monitored using satellite imagery on a repeated basis to gain insight into the crop condition patterns. The choice of multiple regions with varying spatial coverage allows for the evaluation of multiple patterns of crop health and helps in strengthening the analysis. This also helps in evaluating how crop health monitoring using machine learning algorithms with vegetation indexes are affected by the variability in field size and homogeneity.

Figure 1: Study Area

4. Methodology

Methodology section in details explain proposed framework as shown in figure 2. This framework consist of data collection, followed with required preprocessing, spectral feature extraction through vegetation indices, unsupervised clustering followed with ML classification techniques implementation. Novel approach in our framework starts after it. We have used 5-fold cross validation technique, then we perform comparison among it. SHAP Explainability Analysis is done. For temporal forecasting LSTM is used.

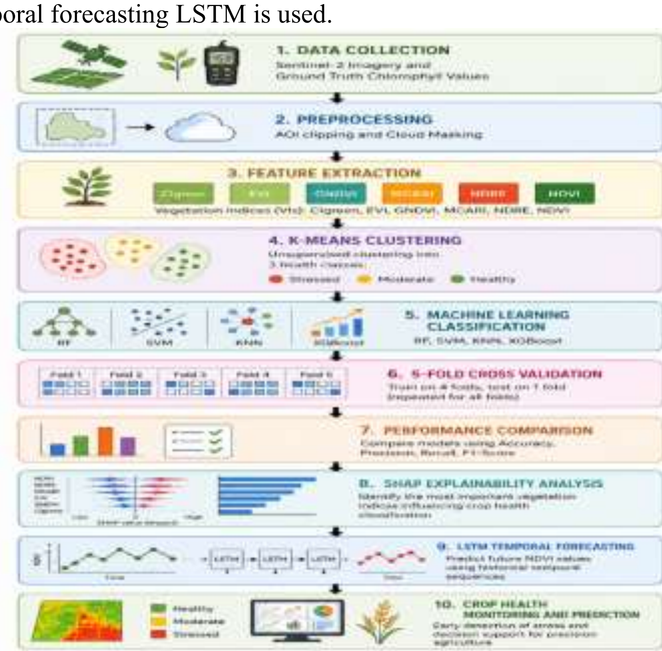

Figure 2: Proposed Methodology

4.1. Data Collection

The multi-temporal Sentinel-2 satellite data was acquired to track the growth and overall health of the crops throughout the 2025-2026 wheat growing season. The study was conducted from November 2025 to February 2026 including the important vegetative and reproductive growth phases of the wheat crop. We have received a total of 25 Sentinel-2A and Sentinel-2B images from the Copernicus Open Access Hub. High spatial resolution multispectral data from Sentinel-2 imagery is suitable for vegetation monitoring and crop health assessment.

Three cultivation areas having three different sizes, 8 acre, 2.5 acre and 2 acre wheat fields were selected for the study. The chosen fields allowed to assess the crop health monitoring strategies at different farm levels. Vegetation analysis and spectral response validation in crop growth conditions were achieved through simultaneous ground truth chlorophyll readings when satellite acquisitions were taken.

Sentinel-2 carries spectral bands in the visible and NIR and also red-edge bands (bands in between the visible and NIR) that are specifically useful for the production of vegetation indices related to the concentration of chlorophyll, biomass accumulation, canopy vigor and photosynthetic activity. These spectral characteristics were used as the foundation in subsequent extraction of the vegetation indices and analysis of the health of the crops [9].

4.2. Preprocessing

To reduce the amount of irrelevant areas for agricultural analysis and to improve the quality of the images, pre-processing was performed. There was a preprocessing workflow involving Area of Interest (AOI) clipping and cloud masking.

4.2.1. AOI Clipping

The region of congealed images acquired from Sentinel-2 is bigger than the geographical area where the wheat is cultivated. Hence, to remove the unnecessary portion of each satellite image, AOI clipping was used to retain only the study areas. Three AOIs were drawn and clipped, each one pertaining to a certain wheat field:

- AOI 1: 8-acre wheat field
- AOI 2: 2.5-acre wheat field
- AOI 3: 2-acre wheat field

To decrease the computational complexity, clipping was performed and only the crop pixels in the study area were used for further processing [10].

4.2.2. Cloud Masking

Vegetation indices are very sensitive to changes in spectral reflectances caused by contamination with clouds. For this reason, cloud masking was implemented in all of the Sentinel-2 images before processing. This was achieved by using the information provided in the Sentinel-2 quality assessment files where cloud-covered and shadowed pixels were identified and removed. This process made sure that only observations free of clouds were used to compute vegetation index and for temporal analysis [11].

4.3. Feature Extraction

The spectral bands in the visible, near-infrared (NIR) and red-edge are provided in high resolution by Sentinel-2 and are well suited for applications in crop monitoring. This study included six vegetation indices which were calculated for each of the dates in the data set: CIgreen, EVI, GNDVI, MCARI, NDRE, and NDVI [4]. These indices relate to various physiological and biochemical properties of wheat crop and give complementary assessment of wheat health. Table 1 shows mathematical model of each vegetation indices [12]. The obtained vegetation indices were used as input features to implement the K-Means clustering, machine learning classification, SHAP-based feature importance analysis and LSTM forecasting.

Table 1: Vegetation Indices

Vegetation Index	Formula	Sentinel-2 Bands Used	Purpose in Crop Monitoring
NDVI (Normalized Difference Vegetation Index)	(NIR − Red) / (NIR + Red)	B8, B4	Measures vegetation vigor, biomass, and overall crop health. Widely used for monitoring crop growth and stress conditions.
GNDVI (Green Normalized Difference Vegetation Index)	(NIR − Green) / (NIR + Green)	B8, B3	Sensitive to chlorophyll concentration and nitrogen status. Useful for assessing crop productivity and nutrient availability.
NDRE (Normalized Difference Red Edge Index)	(NIR − Red Edge) / (NIR + Red Edge)	B8, B5	Detects subtle variations in chlorophyll content and plant stress, particularly during advanced growth stages.
EVI (Enhanced Vegetation Index)	2.5 × ((NIR − Red) / (NIR + 6 × Red − 7.5 × Blue + 1))	B8, B4, B2	Improves sensitivity in dense vegetation areas and reduces atmospheric and soil background effects.
MCARI (Modified Chlorophyll Absorption Ratio Index)	[(RE − Red) − 0.2 × (RE − Green)] × (RE / Red)	B5, B4, B3	Estimates chlorophyll absorption and photosynthetic activity. Useful for detecting nutrient deficiency and stress conditions.
CIgreen (Green Chlorophyll Index)	(NIR / Green) − 1	B8, B3	Measures leaf chlorophyll content and nitrogen concentration. Effective for monitoring crop nutritional status.

4.4. K-Means Clustering

Using an unsupervised learning method called K-Means clustering, wheat crop health was automatically classified using the six vegetation indices that were extracted from the data (CIgreen, EVI, GNDVI, MCARI, NDRE, and NDVI). Ground truth health labels were not available and so K-Means was used to try to uncover natural patterns and similarities in the multi-temporal spectral data. The vegetation index features were standardized, that is, equal contribution of each variable to cluster creation, that is, prior to clustering. A fixed number of K = 3 was selected (clusters of Healthy, Moderate and Stressed crop health). Observations were iteratively subdivided according to the nearest centroid of the clusters (minimizing WCSS). The mean of the NDVI, NDRE and chlorophyll corresponding to the generated clusters were used for analyzing the generated clusters, and the clusters with higher vegetation index (NDVI and NDRE) were colored as healthy and clusters with lower values as stressed. The resulting cluster labels were used as pseudo-ground truth classes for subsequent supervised machine learning based classification processes for automated assessment of crop health without having to label it extensively in the field [13].

4.5. Machine Learning Classification

Four machine learning algorithms were used for analysis: Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and XGBoost.

4.5.1. Random Forest (RF): RF is an important an ensemble learning model which constructs multiple decision trees and aggregates their outputs together [14]. The prediction is given by following mathematical model represented in equation 1:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (1)$$

where $h_t(x)$is the prediction of the t^{th}tree.

4.5.2. Support Vector Machine (SVM): SVM is supposed to target on to finding the optimal hyperplane that maximizes the margin between classes [15], mathematical equation for it is represented in equation 2:

$$f(x) = w^T x + b \quad (2)$$

where $w$is the weight vector and $b$is the bias.

4.5.3. K-Nearest Neighbors (KNN): KNN model is aims to classifies a data point based on the majority class of its nearest neighbors [16], following is equation 3 represents it:

$$\hat{y} = \text{mode}(y_1, y_2, \ldots, y_k) \quad (3)$$

4.5.5. XGBoost: XGBoost is a gradient boosting technique that builds models sequentially to minimize a loss function [17]. Equation 4 represents its mathematical function:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \quad (4)$$

4.6. Five-Fold Cross Validation

The evaluation of machine learning models was performed using a five-fold cross-verification approach to attain the robustness and unbiased evaluation. The entire data set was then divided into five mutually exclusive data sets with the same number of observations in each. For each iteration four subsets were used for training model and a remaining subset was used for validating model. This was repeated with five different sets, with each set getting used exactly once as the validation set. The mean accuracy and standard deviation from the accuracy of all the folds were used to quantify the classification performance. To avoid overfitting and to obtain a good estimate of generalisation capability of evaluated classifiers across varying data distributions, the cross-validation procedure was used [18].

4.7. Performance Comparison

Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Extreme Gradient Boosting (XGBoost) classifiers' performance was comparatively evaluated against the crop health classes generated using K-Means clustering. The extracted vegetation indices were used in training the models and the evaluation was performed by the five-fold cross validation. The evaluation was primarily carried out based on the classification accuracy and standard deviation. Results demonstrated that comparative analysis of classifiers helped find the most suitable machine learning classifier to classify the wheat crop health and found the stability and robustness of the model in a validation fold of different classifiers.

4.8. SHAP Explainability Analysis

SHapley Additive exPlanations (SHAP) analysis was conducted to add interpretability to the models and gain an understanding of the impact of the vegetation indices in determining the health of the crop. SHAP (Shapeley Additive Permutation) explains is a game-theoretic explainable artificial intelligence method which provides quantitative feature importance measures for the predictions of model [19]. It calculates, for each feature, the Shapley value that is calculated to assess the contribution of the feature within all combinations of features. For this study, the XGBoost classifier was used and SHAP analysis was applied to determine the relative importance of the vegetation indices used:

NDVI, NDRE, MCARI, EVI, GNDVI, and CIgreen. The obtained feature importance rankings gave insights into spectral variables most responsible for distinguishing healthy, moderate and stressed wheat crop conditions thus enhancing the transparency and interpretability of the classification framework.

4.9. LSTM-Based Temporal Forecasting

A Long Short-Term Memory (LSTM) network was used for predicting future NDVI values to take care of temporal dependences in changes of crop growth. LSTM is a special type of recurrent neural network that is able to include episodic data and long-term temporal relationships, thanks to a memory cell and gates [20]. Multi-temporal Sentinel-2 data were used to extract the NDVI historical observations as input sequences used to train the model. The network was composed of LSTM layers, and the outputs were condensed with a dense layer and dropout regularization for continuous prediction of NDVI. The Adam optimization algorithm and the Mean Squared Error (MSE) loss function were used for fitting the model and the Mean Absolute Error (MAE) was used to check the forecasting performance. Use of the LSTM model allowed predicting future vegetation status and thus proactively monitoring the crop and early assessing the potential crop stress during the wheat growing season.

5. Result and Discussion

The complete result and discussion section consist of K-means clustering, Machine Learning Models performnace, SHAP-Based Explainability Analysis and LSTM-Based Temporal Forecasting.

5.1. K-Means Clustering

The extracted vegetation indices (CIgreen, EVI, GNDVI, MCARI, NDRE, and NDVI) were segmented using K-Means clustering as the number of different health conditions in the wheat crop was unknown and ground-truth labels were not available. Using spectral similarity, the algorithm formed three clusters (K = 3) which correspond to different crop health conditions. The three indices, NDVI, NDRE and Chlorophyll, sensitive to vegetation vigor, chlorophyll concentration and crop physiological state were used to analyze cluster characterization by calculating their mean values. The table 2 represents clustering technique. Also figure 3 shows K-Means crop health clusters,

Table 2: Mean Spectral Characteristics of K-Means Clusters

Health Category / Class	NDVI	NDRE	Chlorophyll
Stressed - Cluster 0	0.3718	0.2198	0.4611
Healthy- Cluster 1	0.8238	0.6052	0.5593
Moderate- Cluster 2	0.6653	0.4523	0.5735

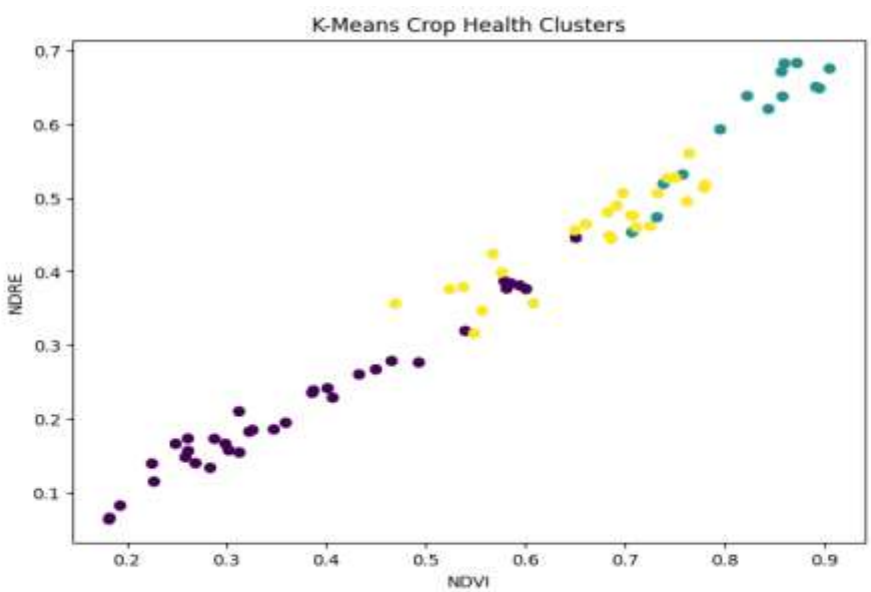

Figure 3: K-Means Crop Health Clusters

From the centroid analysis of the cluster, the Cluster 0 showed the lowest NDVI (0.3718), NDRE (0.2198), and Chlorophyll (0.4611), which showed poor vegetation vigor and less content of chlorophyll, this means that the crop might be stressed. This group was, therefore, classified as Stressed class. Cluster 1 had the highest NDVI (0.8238) and NDRE (0.6052) values indicating high canopy cover, high photosynthetic activity and good growth of crops; so, it is considered as Healthy class. Cluster 2 had moderate NDVI (0.6653), NDRE (0.4523) and highest average chlorophyll value (0.5735) indicating moderate crop vigor and transitional growth characteristics. Thus, this cluster was designated to the Moderate health class.

5.2. Evaluation of machine learning models

Four machine learning classifiers—Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Extreme Gradient Boosting (XGBoost)—were trained and tested on the health classes of crops K-Means. To get reliability results and to reduce overfitting, the five fold cross validation technique was used. The mean classification accuracy and standard deviation were used to evaluate.

Table 3: Performance Comparison of Machine Learning Models

Model	Mean Accuracy (%)	Standard Deviation
SVM	92%	0.1067
Random Forest	90.67%	0.08
XGBoost	90.67%	0.1162
KNN	89.33%	0.1373

From the results shown in Table 3, the mean classification accuracy of the Support Vector Machine was the highest (92.0%) which has the better space-separating ability in the multidimensional feature space of the vegetation-indices. The good accuracy achieved by SVM indicates that spectral representation of healthy, moderate and stressed wheat crops are highly class separable in kernel based decision boundary.

The mean accuracy of the Random Forest and XGBoost classifiers were 90.67% which was the same. The lowest standard deviation (0.0800) occurred for random forest, which shows its robustness and stability across the validation folds. By contrast, XGBoost yielded a slightly larger standard deviation of 0.1162, which means that it is a bit more sensitive to the fluctuations of training data.

KNN presented the lowest robustness and higher reliance on the neighborhood structure of the feature space that showed the highest standard deviation (0.1373) while the mean was 89.33%. KNN achieved satisfactory performance in the classification, but the deviation resulted by the folds is higher than the other classifiers.

Overall, the classification results show that vegetation indices extracted from them contain sufficient discrimination information for the health assessment of various types of crops, and SVM can be the most precise classification model when compared to other machine learning models.

5.3. XGBoost Performance Exclusively

A train-test evaluation was carried out with the created crop health classes to further assess the ability of the XGBoost classifier. The overall accuracy of the classifier was 93.33%, which is a good performance rate.

Table 4: XGBoost Performance

Class	Precision	Recall	F1-Score
Cluster 0	1	1	1
Cluster 1	1	0.67	0.8
Cluster 2	0.83	1	0.91
Overall Accuracy	**0.93**	**0.93**	**0.93**

As represented in table 4,one result was perfect precision and recall for Cluster 0, which proved an accurate recognition of one crop health, using the XGBoost classifier. Cluster 2 yielded an F1 value of 0.91, thus showing high classification ability. Of note, however, is the comparatively lower recall for Cluster 1 (moderate health conditions), indicating minor overlap with adjacent health conditions. However, when considering all the classes, the weighted F1 score of 0.93 proves that the use of vegetation indices for the classification of wheat crops is effective.

5.4. SHAP-Based Explainability Analysis

SHAP (SHapley Additive exPlanations) analysis is used with XGBoost classifier to enhance the model interpretability. SHAP measures the importance of each vegetation index towards the classification of the crop health and score Shapley values using cooperative game theory.

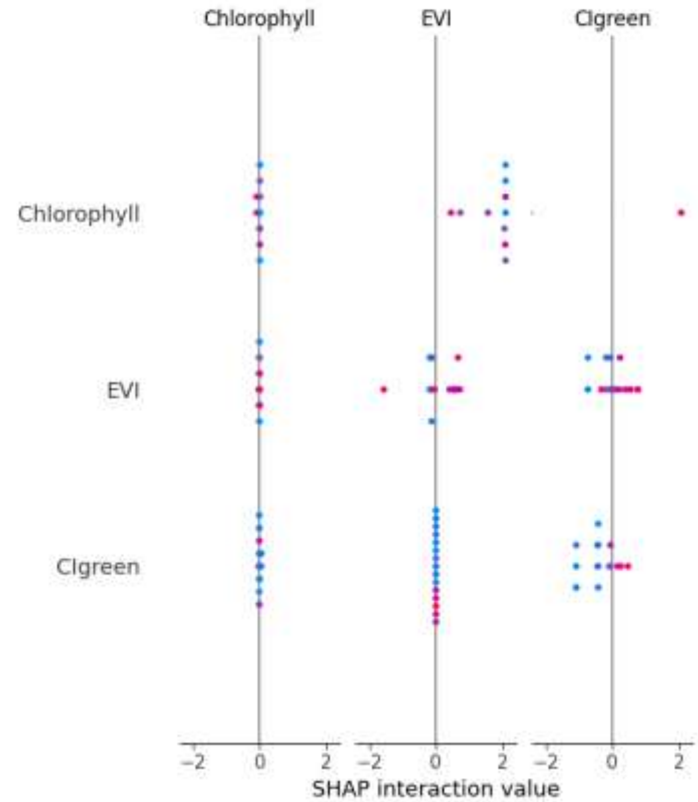

Figure 4: SHAP Interaction Values

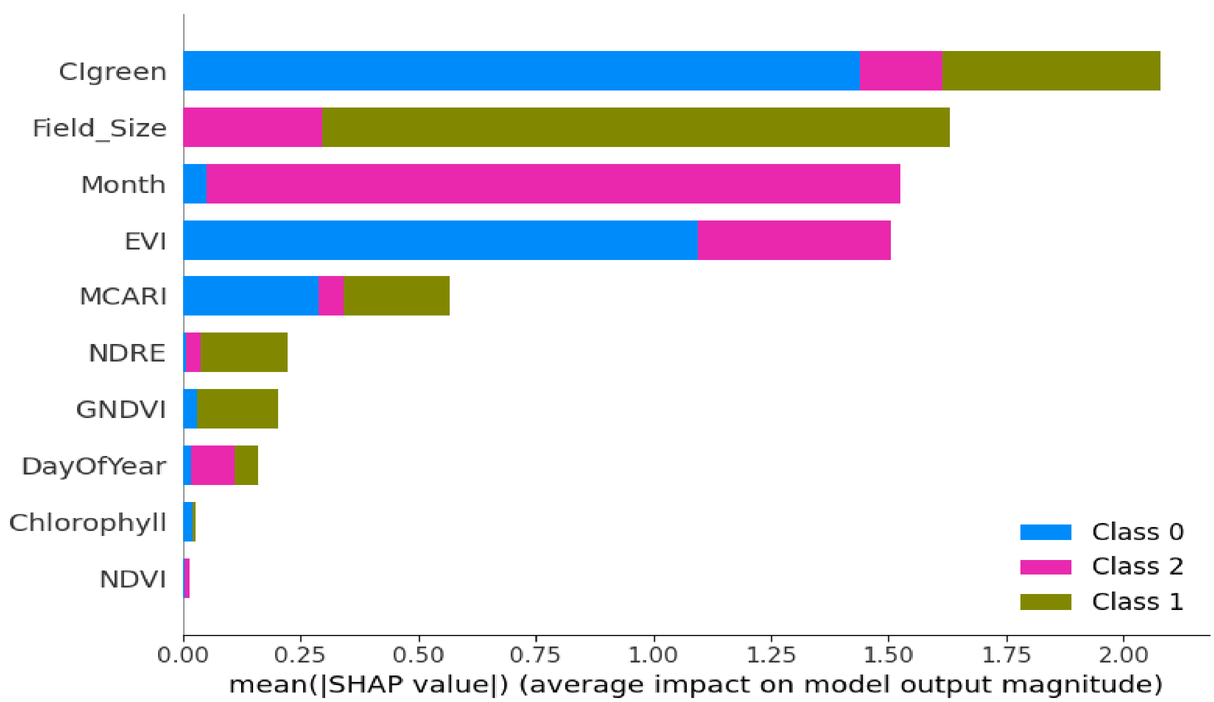

Figure 5: SHAP Summary Plot

Figure 4 indicates SHAP Interaction Values while figure 5, SHAP summary plots indicated that features such as NDVI, NDRE, MCARI, Chlorophyll, and GNDVI were some of the most influential features that affected the classification performance. NDVI and NDRE had the greatest influence on model predictions because of its high sensitivity to the vigour of the canopy, the amount of chlorophyll and the photosynthetic activity. Likewise, MCARI and Chlorophyll responded well to changes in crop nutrition status and physiological stress. The explainability analysis validates that the biologically relevant vegetation indices related to wheat growth and health are the key variables being used by the machine learning model. This improves the transparency of the proposed framework and gives confidence that the crop health predictions that can be generated are accurate.

5.5. LSTM-Based Temporal Forecasting

A Long Short-Term Memory (LSTM) network was created to add temporal learning capabilities to the crop monitoring system and predict future NDVI values. Historical NDVI sequences were collected from the multi-temporal Sentinel-2 observations and used for the model training. The model was composed of two LSTM layers with 64 and 32 neurons, respectively and a dropout-regularized dense output layer. Table 5 shows LSTM performance.

Table 5: LSTM Forecasting Performance

Metric	Value
Training Epochs	100
Evaluation Metric	MAE
Mean Absolute Error (MAE)	0.1797

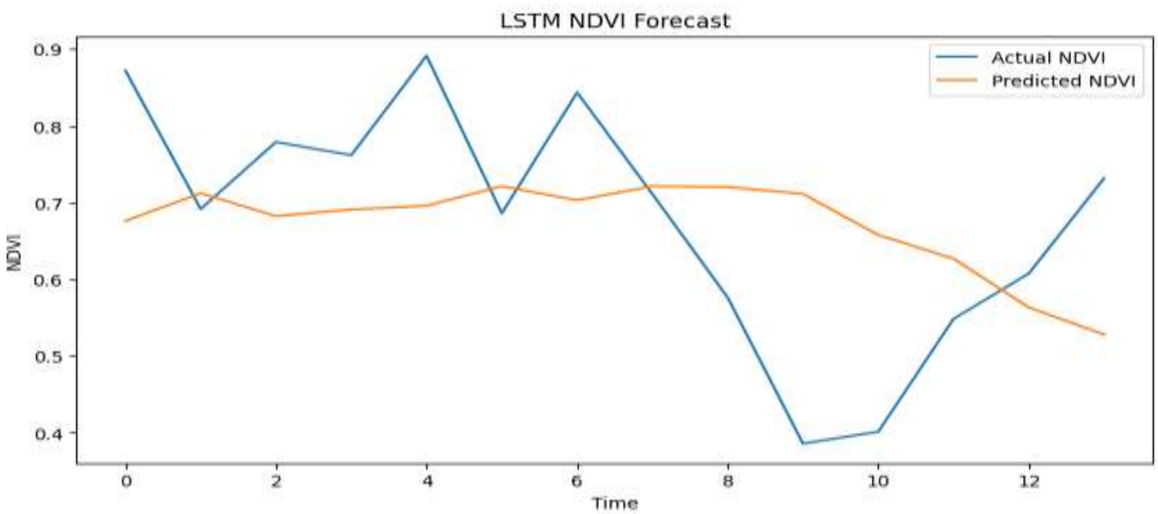

Figure 6: LSTM Forecasting Performance

The model was trained with a maximum of 100 epochs, using Mean Squared Error (MSE) as the loss function and the Adam optimizer. To make sure that no overfitting occurs, early stopping was implemented to restore the optimal model weights. The network gradually acquired temporal dependency that exists within the NDVI time series during training to achieve converging validation performance. The LSTM model yielded the Mean Absolute Error (MAE) of 0.1797 which shows the acceptable forecasting performance for short term NDVI prediction. Figure 6 indicates LSTM forecasting performance. Given the limited temporal observations over the three wheat fields, the achieved forecasting accuracy shows that the network was able to capture trends in vegetation and crop growth dynamics. The model was able to learn seasonality of crop vigor and produce predictions that matched the NDVI observations quite well. Forecasting results show the potential of deep learning approaches for early assessment of crop condition and temporal monitoring of crops. The proposed framework can aid proactive manageability of agriculture and assist in taking the decisions in time during critical growth stages.

6. Conclusion

This study created an integrated framework for wheat crop monitoring by integrating vegetation indices derived from the Sentinel-2 images, k-means based clustering, machine learning based classification, explainability using SHAP approach and LSTM based temporal forecasting. Using the spectral characteristic the unsupervised clustering approach had successfully classified wheat crop conditions into healthy, moderate and stressed classes. The models compare each other and the results show that the machine learning model of Support Vector Machine (SVM) has the highest classification accuracy rate (92.0%) and the classification stability rates of Random Forest and XGBoost are similar. These results were confirmed using a SHAP analysis which revealed that indices derived around the red-edge (NDVI, NDRE, MCARI) and chlorophyll (MCARI) had the greatest influence on crop health assessment. Furthermore, LSTM model trained for 100 epochs had an MAE of 0.1797, showing its ability to depict the temporal crop growth pattern and predicting crop conditions in the future. The results obtained show that the proposed explainable and predictive framework is effective to support the crop health monitoring and decision making in wheat field for various size wheat fields.

7. References

1. Barbedo, J. G. A. (2025). A review of artificial intelligence techniques for wheat crop monitoring and management. Agronomy, 15(5), 1157.
2. Jamil, M., Rehman, H., Saqlain Zaheer, M., Tariq, A., Iqbal, R., Hasnain, M. U., ... & Elshikh, M. S. (2023). The use of Multispectral Radio-Meter (MSR5) data for wheat crop genotypes identification using machine learning models. Scientific Reports, 13(1), 19867.
3. Gill, H. S., Bath, B. S., Singh, R., & Riar, A. S. (2024). Wheat crop classification using deep learning. Multimedia Tools and Applications, 83(35), 82641-82657.
4. Vidican, R., Mălinaș, A., Ranta, O., Moldovan, C., Marian, O., Ghețe, A., ... & Cătunescu, G. M. (2023). Using remote sensing vegetation indices for the discrimination and monitoring of agricultural crops: a critical review. Agronomy, 13(12), 3040.
5. Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A. B., Qin, X., ... & Loupian, E. (2023). Challenges and opportunities in remote sensing-based crop monitoring: A review. National Science Review, 10(4), nwac290.
6. Jamali, M., Soufizadeh, S., Yeganeh, B., & Emam, Y. (2023). Wheat leaf traits monitoring based on machine learning algorithms and high-resolution satellite imagery. Ecological Informatics, 74, 101967.
7. Ashfaq, M., Khan, I., Shah, D., Ali, S., & Tahir, M. (2025). Predicting wheat yield using deep learning and multi-source environmental data. Scientific Reports, 15(1), 26446.
8. Li, Z., Chen, Z., Cheng, Q., Fei, S., & Zhou, X. (2023). Deep learning models outperform generalized machine learning models in predicting winter wheat yield based on multispectral data from drones. Drones, 7(8), 505.
9. Nduku, L., Munghemezulu, C., Mashaba-Munghemezulu, Z., Ratshiedana, P. E., Sibanda, S., & Chirima, J. G. (2024). Synergetic use of sentinel-1 and sentinel-2 data for wheat-crop height monitoring using machine learning. AgriEngineering, 6(2), 1093-1116.
10. Kim, J. Y. (2024). Open-source software for satellite-based crop health monitoring. Journal of Biosystems Engineering, 49(4), 419-433.
11. Gao, X., Chi, H., Huang, J., Han, Y., Li, Y., & Ling, F. (2024). Comparison of cloud-mask algorithms and machine-learning methods using Sentinel-2 imagery for mapping paddy rice in Jianghan Plain. Remote Sensing, 16(7), 1305.
12. Skendžić, S., Zovko, M., Lešić, V., Pajač Živković, I., & Lemić, D. (2023). Detection and evaluation of environmental stress in winter wheat using remote and proximal sensing methods and vegetation indices—A review. Diversity, 15(4), 481.
13. Singh, R., Yadav, V. K., Kumar, P., Shekhar, S., Chauhan, V., & Kumar, A. (2025). Inquisition on principal component and K-mean clustering analysis for yield and its contributing traits of bread wheat (Triticum aestivum L.). Electronic Journal of Plant Breeding, 16(2), 234-240.
14. Zhu, G., Zhao, C., Zhou, L., Li, Z., & Zhu, H. (2025). Winter wheat yield prediction at a county scale using time series variation features of remote sensing spectra and machine learning. European Journal of Agronomy, 170, 127751.
15. Samad, N., & Shafique, M. (2024). MACHINE LEARNING AND REMOTE SENSING INTEGRATION FOR EARLY DETECTION OF CROP DISEASES AND PEST OUTBREAKS. Gomal Journal of Agriculture and Biology, 2(02), 88-106.
16. Chaurasia, A., & Kaur, A. A Hybrid Machine Learning Approach for Accurate Crop Yield Forecasting. In Multimodal Artificial Intelligence in Precision Agriculture (pp. 225-234). CRC Press.
17. Sangeetha, T., & Ezhumalai, P. (2025). Optimizing plant health monitoring: improved accuracy and the computational efficiency with stacked machine learning models and feature filtering. Bulletin of Electrical Engineering and Informatics, 14(2), 1428-1436.
18. Jamil, M., Ahsan, Z., Saeed, M. N., Raza, A., Migdady, H., Daoud, M. S., ... & Abualigah, L. (2024). Wheat crop genotype and age prediction using machine learning with multispectral radiometer sensor data. Agronomy Journal, 116(4), 1643-1654.
19. Qushtom, H., Hasasneh, A., & Masri, S. (2025). Enhanced wheat disease detection using deep learning and explainable AI techniques. Computers, Materials, & Continua, 84(1), 1379.
20. Dhal, S. B., Kalafatis, S., Braga-Neto, U., Gadepally, K. C., Landivar-Scott, J. L., Zhao, L., ... & Bhandari, M. (2024). Testing the performance of LSTM and ARIMA models for in-season forecasting of canopy cover (CC) in cotton crops. Remote Sensing, 16(11), 1906.