CHAPTER 6: INVESTIGATION OF OPTIMAL VEGETATION INDICES FOR RETRIEVAL OF LEAF CHLOROPHYLL AND LEAF AREA INDEX USING ENHANCED LEARNING ALGORITHMS

6.1 INTRODUCTION

Real time monitoring of vegetation parameters such as LAI, LCC, Canopy Water Content (CWC) etc can help in better crop health monitoring and optimized yield production. The chlorophyll content is a salient biochemical property of any plant which determines its health. The chlorophyll content changes at various stages of development and is affected when a crop is exposed to different stresses, which helps farmers and agronomists take safety measures to better the crop yield at crucial stages. The health of plantation is determined by the change in its pigment level [40]. Thus, assessing its amount is necessary for inquiring plant stress and nutritional state for obtaining crop productivity [59].

Biophysical parameters such as LAI, CWC are canopy properties that gives an understanding of the interaction of radiation with vegetation elements. LAI is applicable in evaluating the land and global ecosystems [61], which is used to measure the photosynthetic capacity of vegetation with leaf area as a function. It plays a vital role in estimating foliage cover, monitoring and forecasting crop growth, yield, and biomass production [62, 100]. The knowledge about the spatial distribution of LAI and LCC can give farmers and agronomists pertinent information to use special fertilizers in the field for better yield and crop health [101, 102].

The traditional methods of LAI and LCC retrieval are expensive, destructive, and time-consuming compared to, assessing these parameters non-destructively using remotely sensed vegetation indices (VIs). The VIs exploits the reflectance/absorption values specifically in visible and NIR region of electromagnetic spectrum which provides information about any change in pigments level over its development period. The VIs include ratio, linear combination, orthogonal sets of n equations calculated using data from *n* bands, hybrid vegetation indices, and hyperspectral indices, etc. The advantage of using VIs is to enhance the spectra information by detecting any variation due to physiological and morphological characteristics [46]. Researchers have performed a significant number of studies with the aim of plant species discrimination, leaf vigour evaluation, plant stress assessment, and LCC estimation by developing statistical relations between in-situ data and reflectance measured in the field or using any remote sensors [103-105].

Examining vegetation's spectra using hyperspectral remote sensing (e.g., spectroradiometer) gives promising results to study vegetations biophysical and biochemical properties due to its non-destructive and scalable monitoring approach [106]. Calculating spectral VIs has always been a simple yet efficient method to determine a qualitative and quantitative measure of vegetation growth parameters and health assessment, to name a few applications. Combining this method with statistical enhanced learning algorithms, accuracy can be improved [27, 107].

Couple of studies have been done on physical-based models involving the complex mathematics for vegetation biophysical and biochemical parameters retrieval utilizing vegetation spectral indices [17, 18]. In order to overcome the mathematical complexity of the physical-based model's, nowadays, the robust enhanced learning algorithms such as support vector regression (SVR), partial least square regression (PLSR), random forest regression (RFR), Hybrid neural fussy inference system (HyFIS) are widely used for classification and retrieval purposes in agriculture [7, 24].

SVR is a technique that changes non-linearity regression into a linear problem by the use of various kernel functions, *e.g.,* linear, polynomial, and radial [108]. SVR has several advantages over traditional linear and non-linear regression algorithms, including good generalization ability. However, Random Forest regression is an efficient technique to run over a large number of datasets. Hence, it can handle significant input variables. The technique has proved its efficiency in various agriculture applications [109]. It was also found that RF is less sensitive to noise, and it does not have an overfitting problem. [110].

In contrast, PLSR is a potent technique to reduce multi-collinear variables to a few non-correlated factors. It can be used to predict the vegetation parameters. Although previously researchers have shown that employing all the spectra wavelength for PLSR, it was found that only a few features are sufficient to retrieve meaningful information [102, 111]. Therefore, main vegetation indices can be used to build efficient models. HyFIS is a superior neuro-fuzzy modelling technique that combines the learning power of neural networks with the fuzzy system.

In the presented study, SVR, RFR, PLSR, and HyFIS based vegetation biophysical (LAI) and biochemical (LCC) parameters retrieval for Lentil, Maize, and Mustard crops under different fertilizers and irrigation treatments were investigated. Spectroradiometer FieldSpec4 was used to record the spectra of different crops, with wavelength range of 350nm to 2500nm. To determine the optimal VIs, a ranking-based approach based on the variable in projection (VIP) and R^2 values was applied. Three optimal VIs were given as predictors to the mentioned learning algorithms. The performance of the developed models was evaluated based on mean absolute error (MAE) and coefficient of determination (R^2).

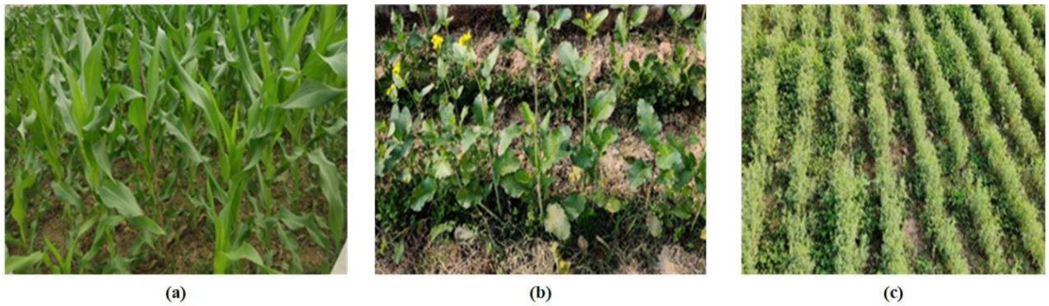

Figure 6.1. **(a)** Maize, **(b)** Mustard, and **(c)** Lentils crops

6.2. METHODOLOGY

6.2.1 Study area and in-situ data collection

For the present study, part of agricultural farm of BHU was chosen as discussed in Chapter 2 (Section 2.3.2). Figure 2.9 (b) is the site in which crop beds (CB) of area 4m×4m was specially prepared to investigate the robustness of hyperspectral indices for LCC and LAI retrieval using different enhanced learning algorithms. The flowchart of the entire methodology is shown in Figure 6.2.

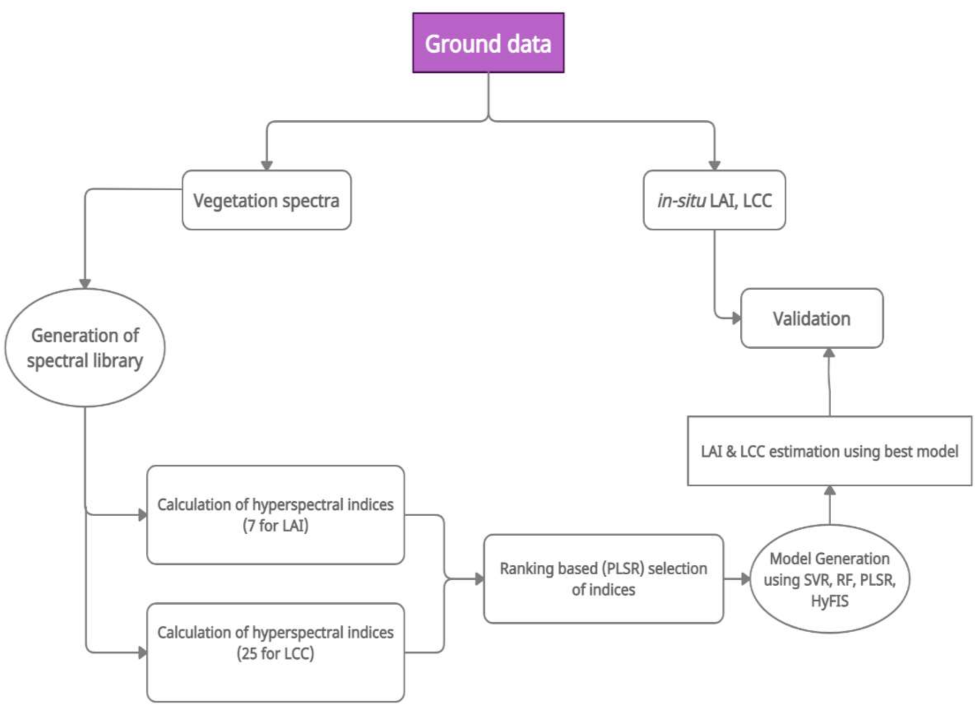

Figure 6.2. Flowchart of the methodology

Three crops, namely Maize, Mustard, and lentils, were selected for in-situ observation of LCC and LAI, with different fertilizer and irrigation treatments. The crop specification such as sowing date and in-situ data acquisition for various phenological stages is tabulated within Table 6.1.

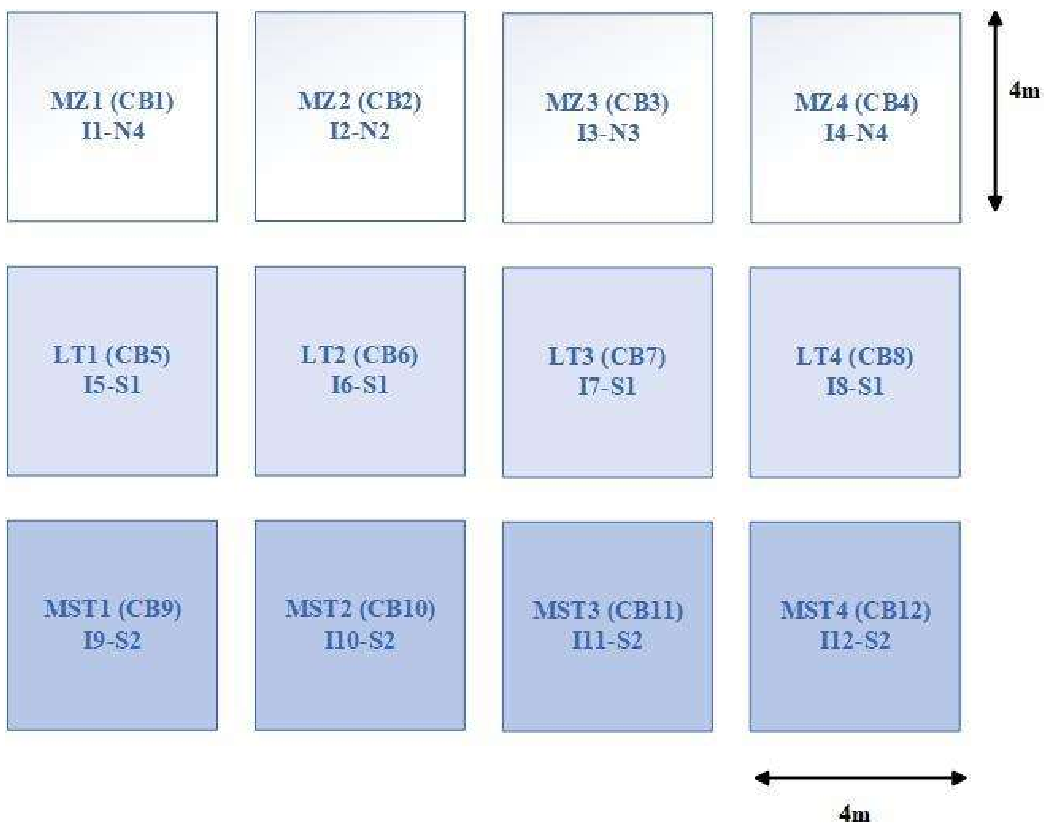

Figure 6.3. Schematic of the crop beds, Maize (MZ), Lentils (LT), and Mustard (MST) under different treatments

Table 6.1. Sowing dates and stages of crops**.**

Crops	**Sowing date**	**Different stages for each crop throughout the data acquisition**		
Mustard	05-12-2019	Inflorescence emergence	Flowering	Ripening

Maize	01-12-2019	Emerging	Kernel developing	Cob development
Lentils	06-12-2019	Nodding	Flowering	Seeding

The ground data acquisition started on January 14th, 2020, and concluded on March 1st, 2020. The sampling started when all of the crops were in the early vegetative stage. Data acquisition was planned to coincide with different phenological stages of the crops to witness any change in vegetation properties. A few locations were pinned for the repetitive collection of data points for every camping visitation in each plot. The schematics of lentils, maize and mustard crop beds under different fertilizer and irrigation treatments is shown in Figure 6.3. The recommended dose was given for each plot with some changes in both kinds of treatments. Different combination of irrigation and fertilizer treatments were given to CB to determine if the learning algorithms can determine how the choice of optimum absorption bands change for different treatments. For Maize, it was given as 150kg of Nitrogen (N), 60kg of phosphate (P), and 80kg of potassium (K), i.e. NPK (150, 60, 80) per hectare. Irrigation treatments for Maize crop beds (CB1 to 4) were given as, saturation (up to field capacity) (I1), alternate wetting and drying (I2), 5 cm of irrigation at 7 days interval (I3), and 2.5 cm of irrigation at 7 days interval (I3); N1 is recommended dose of fertilizers (NPK) and Municipality solid waste (MSW) along with Zinc solubilizers (NPK + MSW + Zn solubilizers), N2 is NPK+MSW, N3 is NPK + MSW (PSB) where PSB is phosphate solubilizing bacteria, and N4 is NPK. In case of Lentil crop beds, fertilizing treatments for each crop beds are NPK (20, 40, 20) per hectare (S1). However, the irrigation treatments are given differently in each crop beds, one time in CB1 (I5), two times irrigation was given to CB2 (I6), one time in CB3 (4.5 cm) (I7), and three times was given to the CB2 (9cm) (I8). The fertilizer treatment for Mustard crop beds are the NPK (40, 20, 20) per hectare, (S2). In case of Mustard irrigation treatments given are different as in the case of Lentils crop beds. Irrigation of Mustard CBs are given as one time in CB1 (2.5 cm) (I9) and CB3 (4.5 cm) (I10), three times was given to the CB2 (9cm) (I11), and two times in CB4 (6.0 cm) (I12).

The in-situ observation of LAI and LCC was systematically taken for three different crop phenological stages from each crop bed from LAI-2200C plant canopy analyzer and SPAD, respectively. The in-situ LAI (i.e., 4-observations) and LCC (i.e., 20-SPAD observations) values were later averaged to one value corresponding to each plot for each data acquisition.

6.2.2 Leaf spectral data

The leaf level spectral radiance was generated at a direction normal to the adaxial leaf surface using Fieldspec 4, under a suitable weather condition (January to March 2020) during the winter season. The readings were obtained from 11:00 am to 1:00 pm IST during bright and clear weather. The spectra were calibrated adequately with a white reference of 99%. The contact probe attached to the spectroradiometer has a diameter of 25 millimetres and an instantaneous field of view (IFOV) of 10 mm and has its halogen light source. The distance between the contact probe and crop canopy was kept below 5 cm in the measurement experiment.

The spectral measurements at leaf level were collected from Maize (n=20, five leaves from each maize crop bed), lentils (n=40, ten from each crop bed), and Mustard (n=20, five from each bed) for every data acquisition. Further, the leaf level spectra were averaged to one value for each crop bed, this way, one spectral library was generated for each crop bed.

6.2.3 Selection of vegetation indices

The calculated VIs were selected based on their characteristics/sensitivity to study LCC and LAI. Different VIs were computed using the leaf spectra, for instance, ratio indices, normalized difference ratio indices, modified ratio/normalized difference ratio indices, hybrid VIs, and hyperspectral VIs. Ratio indices and normalized difference ratio indices are based on reflectance in the Red band (maximum absorption due to chlorophyll pigment) and NIR (maximum reflection due to leaf cellular structure) part of the spectra and known as best classical VIs. The hybrid VIs were usually used to study greenness measurement over the entire phenology of vegetation. However, the hyperspectral VIs were used to study the characteristics of vegetation's red edge reflection pattern of vegetation. The observed blue shift and red shift at the blue and red edge inflection point are associated with various phenological stages. The computed VIs based on the leaf scale related to LCC and canopy scale related to LAI are shown in Tables 6.2 and 6.3. Furthermore, a comprehensive comparative analysis was carried out to prove the potential of listed vegetation indices for predicting various vegetation parameters, namely, LCC and LAI using different enhanced learning techniques, namely, SVR, PLSR, RFR, and HyFIS.

Table 6.2. VIs at canopy related to LAI

Index	Formulation	References
Greenness Index (*GI*)	$\frac{R_{554}}{R_{677}}$	Smith et al. (1995) [112]
Modified Chlorophyll Absorption Ratio Index (*MCARI*)	$\left((R_{700} - R_{670}) - 0.2(R_{700} - R_{550})\right)\frac{R_{700}}{R_{670}}$	Daughtry et al. (2000) [75]
MSR2	$\frac{R_{750}}{R_{705}} - \frac{1}{\sqrt{\frac{R_{750}}{R_{705}} + 1}}$	Chen (1996) [113]

Normalized Difference Vegetation Index (*NDVI*)	$\frac{R_{800} - R_{680}}{R_{800} + R_{680}}$	Tucker (1979) [114]
Renormalised Difference Vegetation Index (*RDVI*)	$\frac{R_{800} - R_{680}}{\sqrt{R_{800} + R_{680}}}$	Roujean and Breon (1995) [115]
Simple Ratio (*SR*)	$\frac{R_{800}}{R_{680}}$	Jordan (1969) [116]
Transformed Vegetation Index (*TVI*)	$0.5(120(R_{750} - R_{550}) - 200(R_{670} - R_{550}))$	Broge and Leblanc (2000) [74]

Table 6.3. VIs at canopy /leaf scale related to chlorophyll content

Index	**Formulation**	**References**
Chlorophyll Absorption Ratio Index (*CARI*)	$R_{700}\sqrt{\frac{a*670 + R_{670} + b}{R_{670}}}\sqrt{(a^2 + 1)}$ $a = \frac{R_{700} - R_{550}}{150}; b = R_{550} - a*550$	Kim et al. (1994) [117]
Datt1	$\frac{R_{850} - R_{710}}{R_{850} + R_{680}}$	Datt (1999) [96]
Datt2	$\frac{R_{850}}{R_{710}}$	
Greenness Index (*GI*)	$\frac{R_{554}}{R_{677}}$	Smith et al. (1995) [112]
Gitelson1	$\frac{1}{R_{677}}$	Gitelson et al.(1999) [97]
Gitelson2	$\frac{(R_{750} - R_{800})}{(R_{695} + R_{740})} - 1$	Gitelson et al.(2003) [98]

Modified Chlorophyll Absorption Ratio Index (*MCARI*) *MCARI/OSAVI*	$((R_{700} - R_{670}) - 0.2 * (R_{700} - R_{550}))\frac{R_{700}}{R_{670}}$	Daughtry et al. (2000) [75]
MCARI2	$((R_{750} - R_{705}) - 0.2 * (R_{750} - R_{550}))\frac{R_{750}}{R_{705}}$	Wu et al. (2008) [118]
Modified NDVI (*mNDVI*)	$\frac{R_{800} - R_{680}}{R_{800} + R_{680} - 2R_{445}}$	Sims and Gamon (2002) [72]
Maccioni	$\frac{R_{780} - R_{710}}{R_{780} + R_{680}}$	Maccioni et al. (2001) [119]
Normalised Difference Vegetation Index (*NDVI*)	$\frac{R_{800} - R_{670}}{R_{800} + R_{670}}$	Tucker (1979) [114]
NDVI3	$\frac{R_{682} - R_{553}}{R_{682} + R_{553}}$	Gandia et al. (2004) [120]
Optimised Soil-Adjusted Vegetation Index (*OSAVI*)	$(1 + 0.16)\frac{R_{800} - R_{670}}{R_{800} + R_{670} + 0.16}$	Rondeaux et al. (1996) [121]
OSAVI2	$(1 + 0.16)\frac{R_{750} - R_{705}}{R_{750} + R_{705} + 0.16}$	Wu et al. (2008) [118]
Simple Ratio (*SR*) *SR1*	$\frac{R_{750}}{R_{700}}$	Gitelson and

SR2	$\frac{R_{752}}{R_{690}}$	Merzlyak (1997) [99]
SR3	$\frac{R_{750}}{R_{554}}$	McMurtey et al.(1994) [122]
SR4	$\frac{R_{700}}{R_{670}}$	Zarco-Tejada and Miller (1999) [44]
SR6	$\frac{R_{750}}{R_{710}}$	
Structure Insensitive Pigment Index (*SIPI*)	$\frac{R_{800} - R_{445}}{R_{800} + R_{445}}$	Peñuelas et al. (1995) [123]
Transformed Chlorophyll Absorption reflectance Index (*TCARI*) *TCARI/OSAVI*	$3 * \left((R_{700} - R_{670}) - 0.2 * (R_{700} - R_{550})\right) * \frac{R_{700}}{R_{670}}$	Haboudane et al. (2002) [76]
TCARI2 *TCARI2/OSAVI2*	$((3 * ((R_{750} - R_{705}) - 0.2 * (R_{750} - R_{550}) * \frac{R_{750}}{R_{705}}$	Wu et al. (2008) [118]

6.3. STATISTICAL ANALYSIS

6.3.1 Support Vector Regression (SVR)

The SVR was developed in the late 90s, which uses a statistical learning approach for non-linear mapping of given problems. It does not require any assumption on the response variable's distribution. Our focus is on obtaining a continued-valued function in the SVR that maps observable with pair of input parameters into high (n-) dimensional feature space by a non-linear mapping. Further, a linear model is generated into this feature space by ε- insensitive loss This linear function f(x,ω) can be represented as

$$f(x, \omega) = \sum_{i=0}^{n} \omega_i \varphi_i + b \tag{6.1}$$

where, $\varphi_i(s)$, i=1,2,3,….,n, represents high dimensional feature space, mapped non-linearly, and b stands for bias term. Further, maximization $\frac{1}{\omega}$ or minimization $\|\omega\|^2$ is described by introducing a non-negative slack variable ξ_j, ξ_j^*, j=1,2,…,m to measure any deviation of samples outside ε-insensitive area. The minimization problem is formulated as

$$\min \frac{1}{2}\|\omega\|^2 + C\sum_{j=0}^{n}\left(\xi_j + \xi_j^*\right) \tag{6.2}$$

$$\text{Subjected to} \begin{cases} y_i - f(s_i, \omega) \le \varepsilon + \xi_j^* \\ f(s_i, \omega) - y_i \le \varepsilon + \xi_j \\ \xi_j, \xi_j^* \ge 0, j = 1,2, \dots, m \end{cases}$$

C is a regularization constant. The Lagrange multiplier is used to get a new function unhindered by constraints to find the extremum of a function with associated constraints. After introducing the multiplier and minimizing the function, the equation becomes,

$$f(s) = \sum_{i=0}^{n}(\alpha_i - \alpha_i^*)\, K\left(s_i, s_j\right) \tag{6.3}$$

where α_i and α_i^* are Lagrange's multipliers, and $K\left(s_i, s_j\right) = \varphi(s_i).\varphi\left(s_j\right)$ is the kernel functions, i.e., radial, linear and polynomial. $\varphi(s_i)$ and $\varphi\left(s_j\right)$ represent different spaces.

6.3.2 Random Forest Regression (RFR)

The RFR is a deep machine learning technique that uses a non-parametric approach to forecast data samples. It generates several regression trees using samples from the training data. These trees are the sets of conditions given from the roots of the tree to leaves. The RFR model blends the results of regression trees and estimates a better result. Each tree was trained using a separate bootstrap sample drawn at random from the data set. In general, two-thirds of the dataset is used to train the model, rest of the one-third dataset is used to create out-of-the-bag error estimates, which is further used for model validation [124]. Mathematically, RF is presented as

$$\text{RF prediction} = \frac{1}{\mathrm{k}} \sum_{\mathrm{k}=1}^{\mathrm{k}} \mathrm{k}_{\mathrm{th}} \text{ tree response} \tag{6.4}$$

where k is the index run over individual trees, and k_{th} is response function.

6.3.3 Partial Least Square Regression (PLSR)

The PLSR is an iterative statistical technique developed in 1966 by Wold et al., 2001 [93], which uses a linear multivariate model to associate data matrices and, as a result, reduces collinearity and noise in a dataset. This two-step method reduces predictors of a dataset to a smaller number of uncorrelated components before performing least squares regression on them instead of relying on the original data.

PLSR generalizes and combines the features of principal component regression and multilinear regression. The tool is quite useful when there are many variables, and there is much collinearity between them. This regression builds a linear model as follows,

Its algorithm is explained in Chapter 5, Section 5.2.4 (d).

6.3.4 Hybrid neural fuzzy inference system (HyFIS):

HyFIS is an adaptive neuro-fuzzy system; it combines the learning power of a neural network with the fuzzy logic system. The learning criteria of HyFIS is a two-fold method. In the first part, rule finding phase is achieved using the knowledge acquisition module. The second part is the parameter learning phase which is further used to tune membership function to achieve the desired level of performance [125]. The knowledge acquisition module uses Wang and Mendel's (WM) technique. WM technique divides the input and output data into a proper number of Fuzzy regions and assigns a fuzzy membership function to each of them. Structure and parameter learning is a supervised learning method that employs gradient descent-based learning algorithms. This function creates a model that includes a rule database and membership function parameters. The Mamdani model is used in the HyFIS rules for the antecedent and consequent sections. Furthermore, HyFIS uses a Gaussian membership function. So, the mean and variance of the Gaussian function are two parameters that are optimized.

6.3.5 Performance indices

The performance of the developed enhanced learning regression models has been evaluated in terms of mean absolute error (MAE) and coefficient of determination (R^2) as shown in Equations (6.5) and (6.6).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_{obs} - f_{est}| \tag{6.5}$$

$$R^2 = \left[\frac{\frac{1}{n}\sum_{i=1}^{n}(f_{obs}.f_{est})}{\sigma_{obs}.\sigma_{est}}\right]^2 \tag{6.6}$$

where (f_{obs}, σ_{obs}) and (f_{est}, σ_{est}) are the in-situ data and estimated data with its corresponding standard deviation.

6.4. RESULTS AND DISCUSSION

The significant effect of different irrigation and treatments lead to the variation in biochemical properties of the plant. Figure 6.3 represents the variation in the value of in-situ LCC at different phenology of Maize, Mustard, and Lentils as per their treatment and irrigation. No significant change was observed in the crops' physiological status due to the different treatments used for Mustard, Maize, and lentils crops, i.e. temporal variation was showing very similar pattern for LAI with the treatments given to crops. Therefore, LAI values were computed by averaging the value of in-situ LAI of each crop bed. Figure 6.5 shows the temporal variation of LAI at different stages of Maize, Mustard, and lentils. In Figures 6.4 and 6.5, both LAI & LCC of the Lentils crop showed a similar temporal trend. They increased throughout the nodding stage and then started decreasing after flowering. In addition, for Mustard crop beds, both parameters were found decreasing since the data acquisition started from emergence, followed by flowering and ripening. Furthermore, in Maize crop beds, both parameters increased in the same pattern with age.

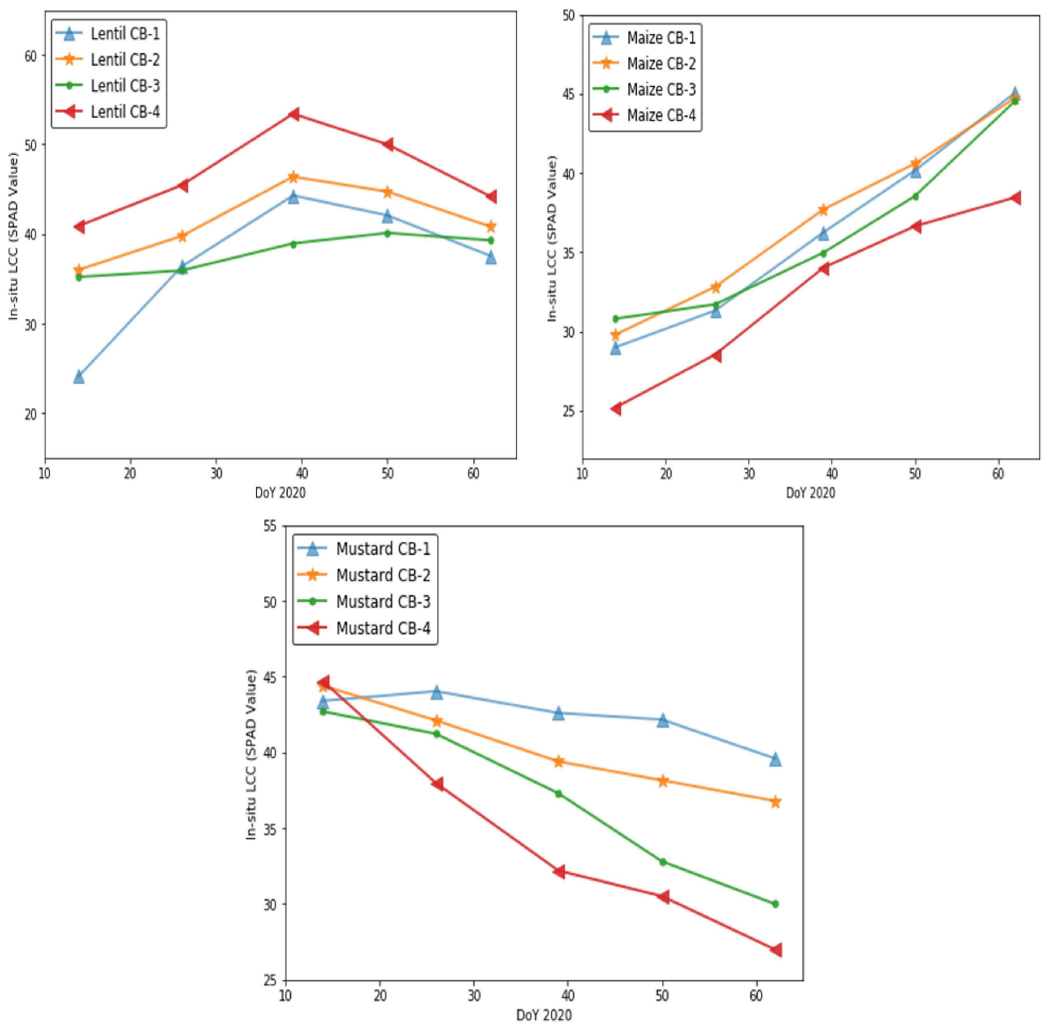

Figure 6.4. Temporal variation of in-situ LCC for various crop beds.

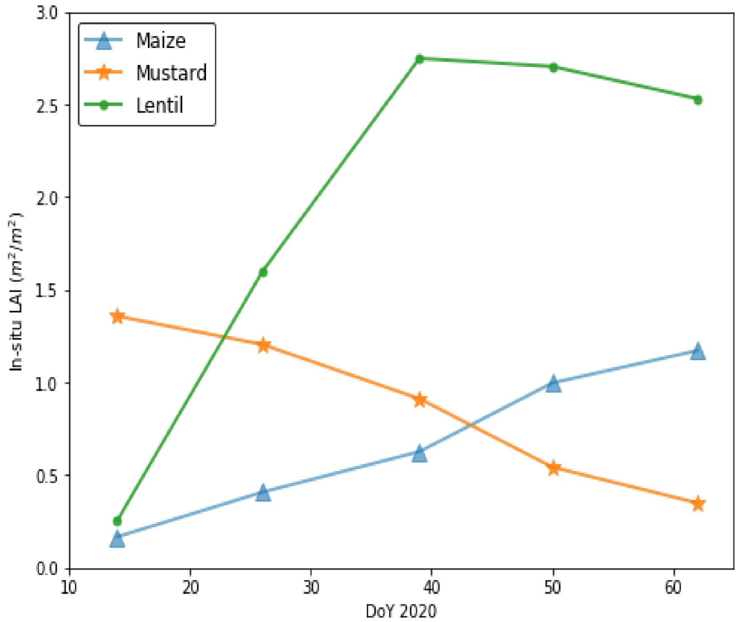

Figure 6.5. Temporal variation of averaged in-situ LAI for various crop beds.

6.4.1 Partial least square (PLS) and coefficient of determination (R^2) analysis-based selection of optimal VIs

In order to evaluate the potential of each computed VIs to predict LAI and LCC, the ranking-based comprehensive comparative analysis was carried out utilizing the PLS regression optimization technique. The PLSR extracted sensitive information from the available sets of VIs and scored them in descending order based on their variable significance in projection (VIP) score. The VIP scores were calculated for regression analysis for selected 25 and 7 hyperspectral VIs to predict LCC and LAI, respectively, for each crop bed, based on the combination of previous studies that shows the sensitivity of LAI and LCC w.r.t. the VIs. The hyperspectral VIs with a high VIP value suggests that it is crucial in vegetation parameters retrieval. It was also found that almost all the top three high VIP scoring indices were the same as that of the top three highly R^2 valued indices for each experiment.

For LAI, MSR2 was found highly sensitive to all species of crops with high VIP=1.12; R^2=0.93 in case of Lentils, VIP=1.19; R^2=0.78 for Maize, and VIP=1.16; R^2=0.99 for mustard crops, as shown in Figure 6.6. It indicates that MSR2 is sensitive to identify the temporal phenology of the various crops. After MSR2, the VIs, namely SR, TVI, and NDVI, were found strongly correlated with LAI of Maize, Mustard, and Lentils crops, respectively, according to their VIP values. The red-edge area is distinguished by a significant increase in the reflectance of green vegetation between the red spectral region's local minimum reflectance band and the NIR spectral region's maximum reflectance band [21, 126]. This area of the electromagnetic spectrum provides crucial information on fresh biomass amount and LAI than other parts of the spectrum. For different treatments CB in same species, it was found that top three VIs were almost same with varying ranks but they all came within top three or four, so for one species all the data were averaged to be treated as one. In conclusion, the VIs using the red-edge reflectance band were more responsive to LAI than the VIs made of blue, green, red reflectance bands.

Further, for LCC, the variation in the R^2 and VIP scores of VIs as a result of substantial fertilizer's effect on the biochemical property of selected crop species were also studied by utilizing the PLSR. Figures 6.7, 6.8, and 6.9 show indices values and their corresponding VIP scores and R^2 values w.r.t. chlorophyll content for each crop bed. The VIs such as MCARI, TCARI2/OSAVI2, TCARI2, and SR1 showed high sensitivity with VIP values 1.01, 2.07, 1.03, and 1.81 for the Lentils CB-1, 2, 3, and 4, respectively. In addition, for Maize CB1, 2, 3, and 4, the VIs, namely MCARI2, SR1, SR2, and SIPI, were found highly correlated with VIP values 1.51, 1.05, 1.68, and 1.30, respectively. Moreover, The VIP score of SR4, SR4, MCARI, and MCARI/OSAVI was found highly sensitive to the LCC value of Lentils CB-1, 2, 3, and 4, respectively. In conclusion, the VIs indices made of combining green (550nm), red (670, 690, 700nm), and red-edge (705,750nm) bands were found highly responsive to the temporal LCC values of lentils and maize crop beds. In contrast, the VIs made of green, red, and NIR (800nm) bands were most sensitive to mustard crop beds' temporal LCC values. The findings can be justified by the fact that in the red-edge region, the inflection point position depends on the chlorophyll concentration, which occurs mostly in between 690nm-740nm. In the IR region, the reflectance is much higher due to intracellular structures of the leaves because inner structures act as a diffused reflector in the near IR region. A comparison of 4-crop beds of Lentils, Maize, and Mustard revealed high sensitivity (i.e., in terms of high VIP values) to optimum sets of three VIs for Lentils CB-2 under treatment I6-S1, Maize CB-2 under I4-N4 and Mustard CB-4 under I10-S2.

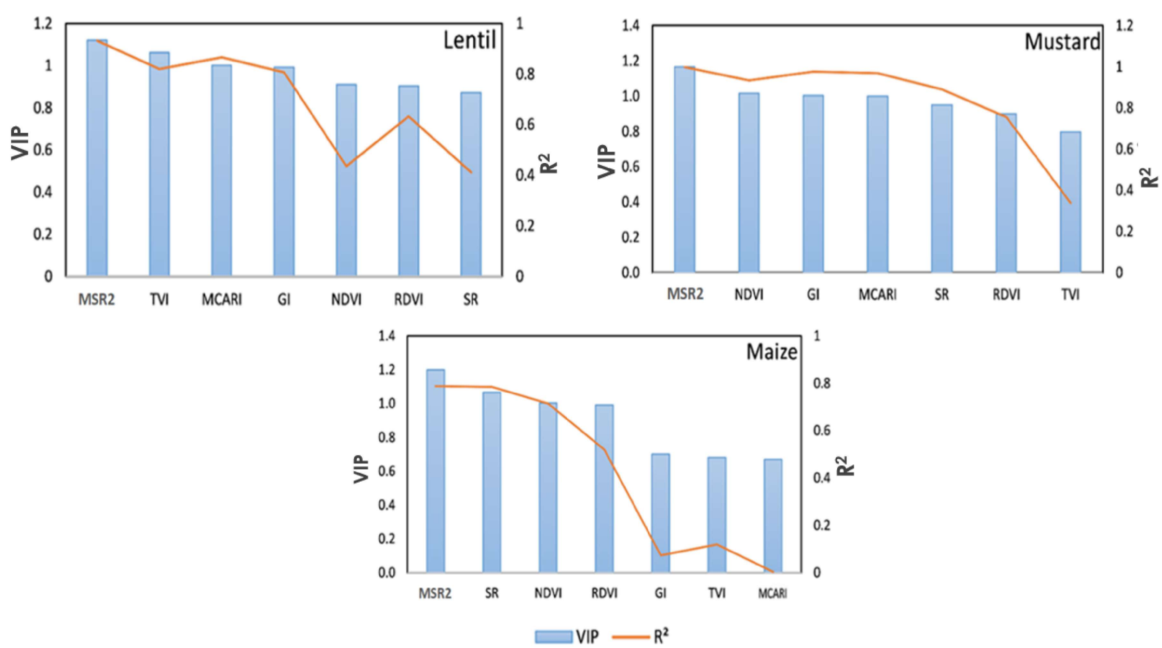

Figure 6.6. VIP of canopy scale hyperspectral VIs related to LAI and their R^2 variation utilizing PLSR for Lentils, Maize, and Mustard crop.

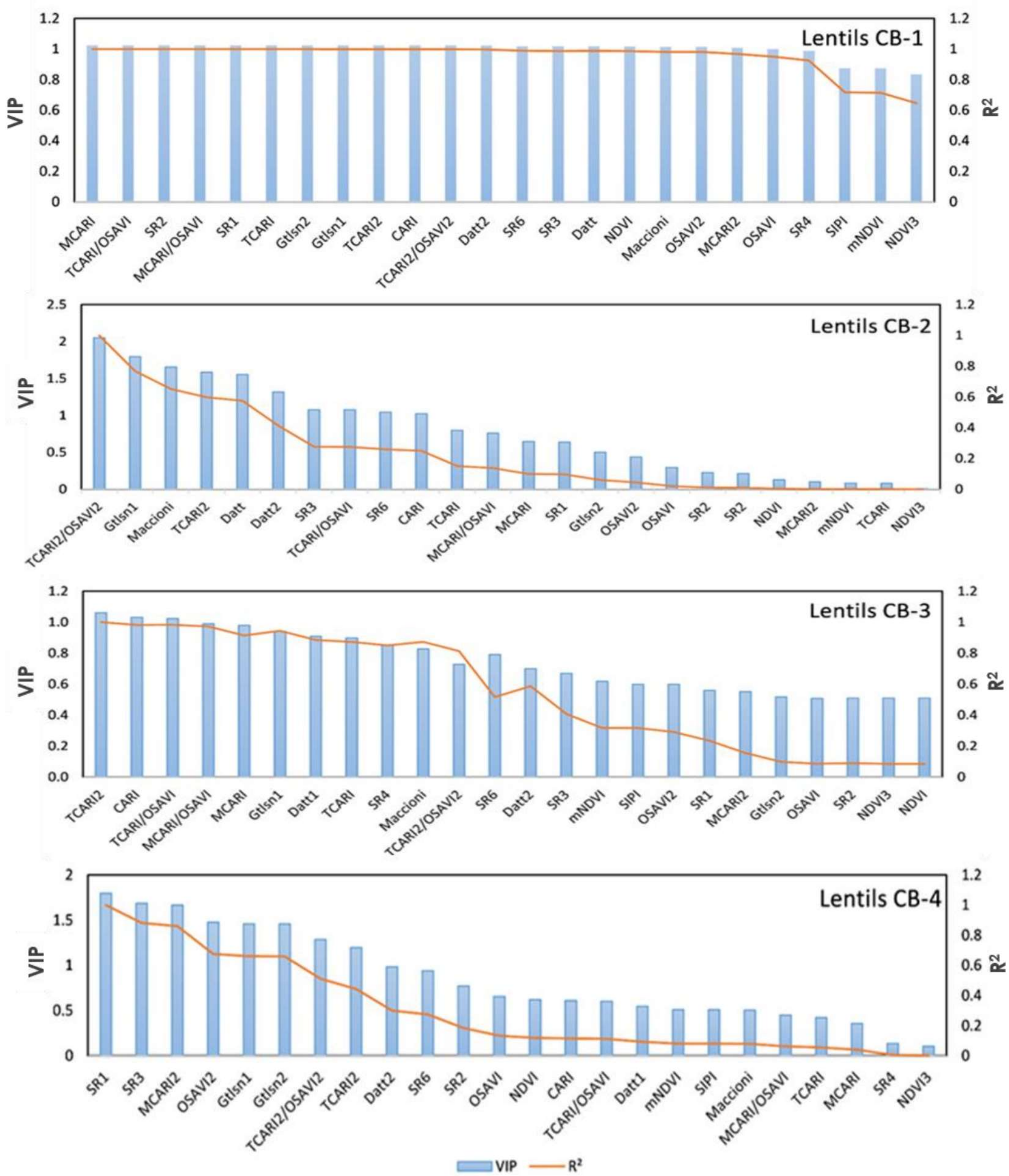

Figure 6.7. VIP of canopy/leaf scale hyperspectral VIs related to LCC and their R^2 variation utilizing PLSR for Lentils CB-1, CB-2, CB-3, and CB-4.

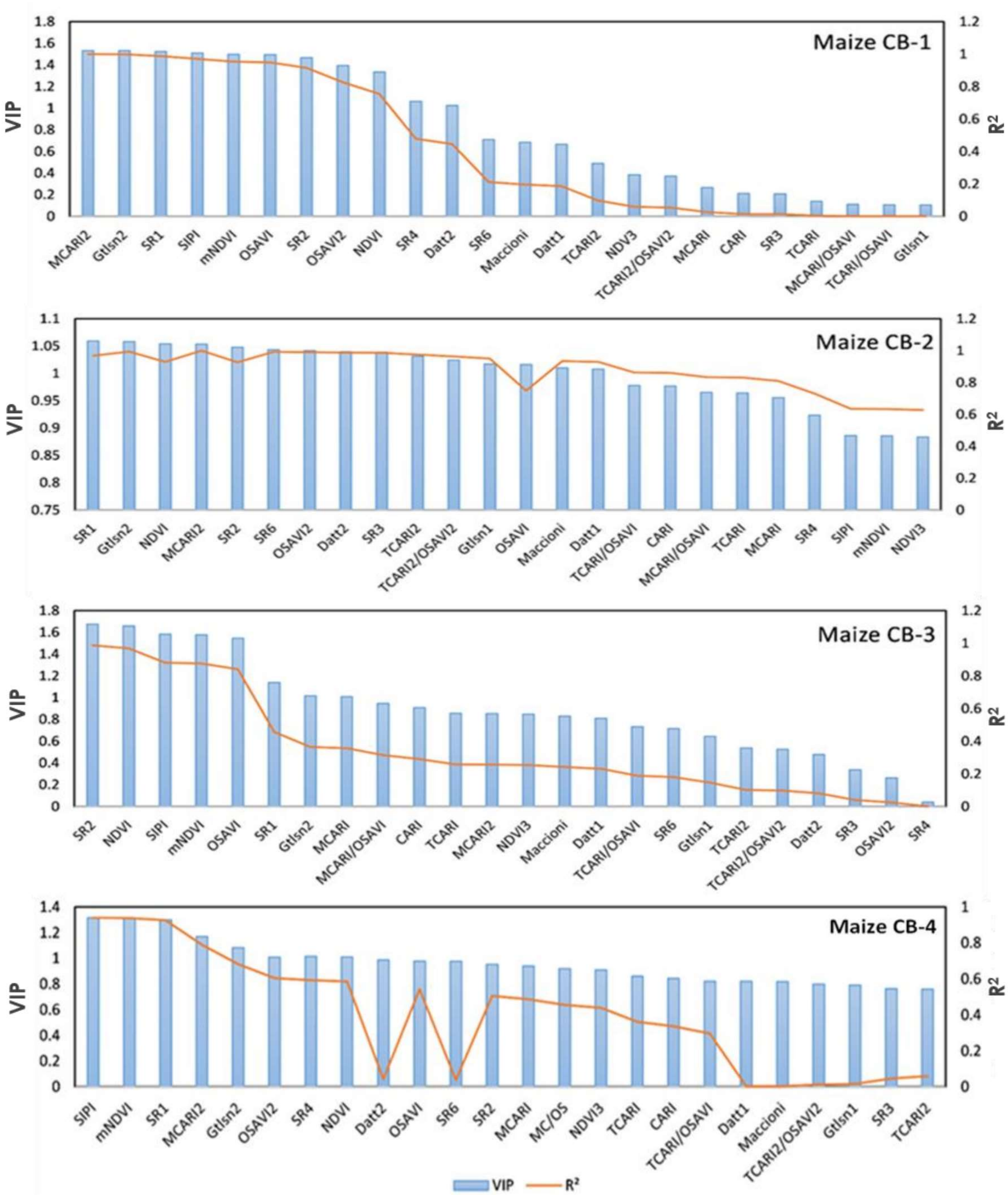

Figure 6.8. VIP of canopy/leaf scale hyperspectral VIs related to LCC and their R^2 variation utilizing PLSR for Maize CB-1, CB-2, CB-3, and CB-4.

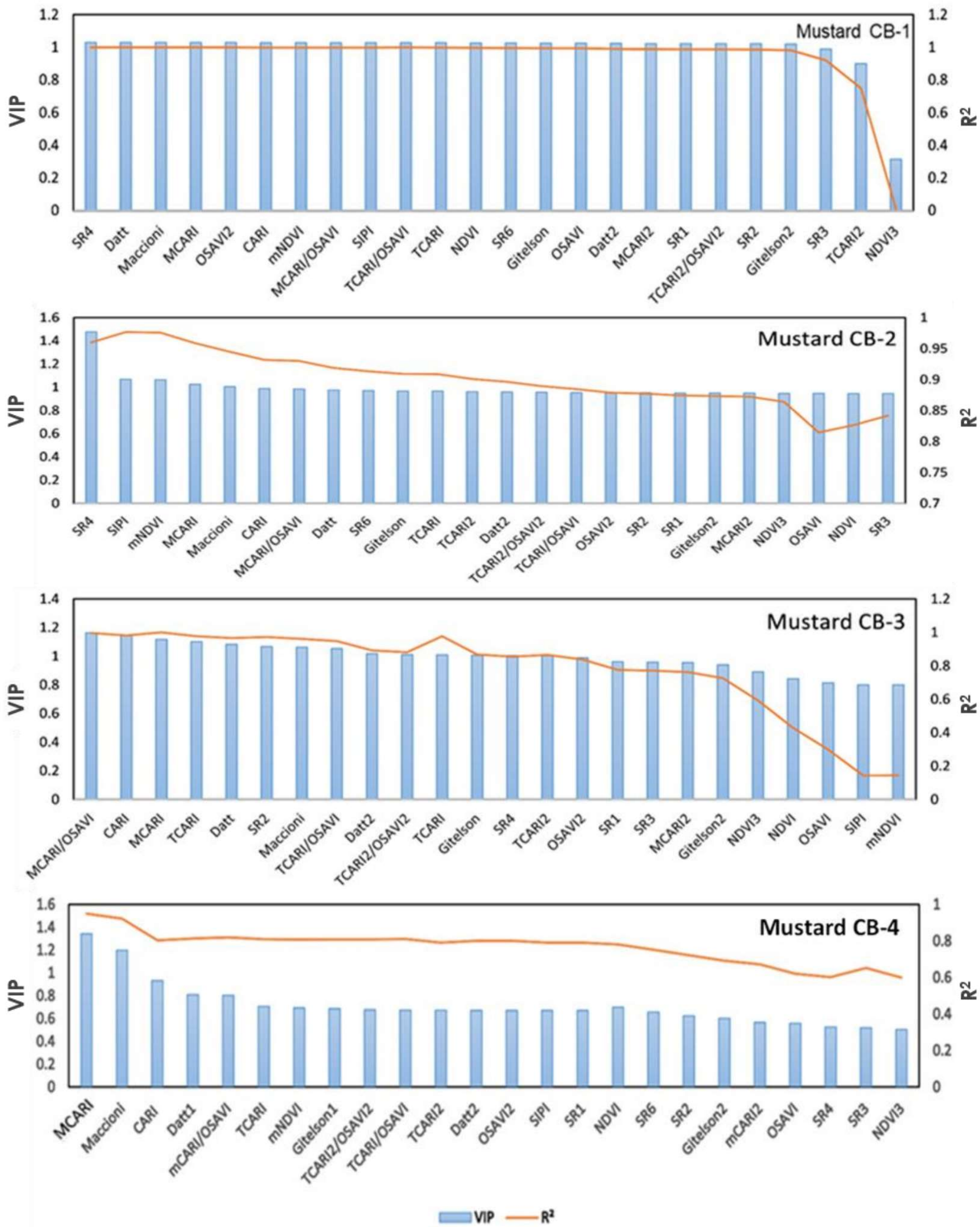

Figure 6.9. VIP of canopy/leaf scale hyperspectral VIs related to LCC and their R^2 variation utilizing PLSR for Mustard CB-1, CB-2, CB-3, and CB-4.

6.4.2 Predictors, response variable, and calibration/validation data sets for model development

The optimum sets of three VIs having high VIP and R^2 value as predictors and in-situ LAI and LCC as response variables were selected to build enhanced learning algorithms. The linear interpolation was carried out to generate temporal datasets of predictors and response variables at different phenology of each crop species from 14 DoY 2020 to 52 DoY 2020 daily. Further, a total of 48 measurements of LCC, LAI, and optimal VIs were divided into two parts for calibration and validation of enhanced learning algorithms. For forward modelling, 70 % of datasets were selected to calibrate the algorithms. Whereas, remaining 30% of datasets were used for validation of algorithms in inverse modelling. Thus, the input variables dimension given to the regression methods is a matrix of 48×3. The calibration data sets were determined regularly to cover various phenology stages of crops under different irrigation and fertilizer treatment. The SVR was executed using three kernel functions, i.e., radial (SVR _Rad), linear (SVR _Li), and polynomial (SVR _Poly). The forward modelling of the SVR using radial, linear, and polynomial kernels mainly focuses on the optimizing model parameters (C, ε, d, and γ). C and ε are hyper-parameters of the SVR, which regularizes the regression network. The notation d and γ represent the kernel parameters. These parameters were optimized to minimize the n-fold cross-validation error using calibration datasets. In addition, the RFR focuses on regularizing stopping parameters and determining an optimum number of trees to produce a stable averaged square error for estimation. Furthermore, the PLSR optimizes the predictive ability by including the optimum number of components required in the model. An n-fold cross-validation strategy was used to evaluate PLSR performance. The HyFIS parameters and training structure uses a gradient descent-based algorithm. This algorithm focuses on building a model with its rule database and parameters membership function.

6.4.3 Enhanced learning algorithms Inversion for LAI and LCC retrieval

6.4.3.1 LAI retrieval of Lentil, Maize, Mustard crop using SVR, RF, PLSR, and HyFIS models

The optimization of SVR model parameters in forward modelling uses 8-fold cross-validation method for Lentils, Maize, and Mustard LAI retrieval is shown in Table 6.7. However, for RFR, Figure 6.10 shows the average squared error against the number of trees. The stopping value ranges from 5 child nodes (minimum) to 100 child nodes (maximum) and is used to stop the number of trees from growing in 10 cycles for mean error calculation. In order to calculate the average squared error, a total of 100 trees were chosen. After 60 trees, a consistent average squared error was observed for all crops, and this tree number was deemed significant for the lowest error. In the case of PLSR, 7-fold cross-validation was used to find a significant number of components. The regression is further carried out by calculating the regression coefficients. HyFIS model generally uses the Gaussian membership function. Therefore, only two parameters (i.e., mean and variance) of Gaussian function were optimized in the presented investigation.

The performance of three optimal VIs combinations (i.e., SR, MSR2, and NDVI for Maize; GI, MSR2, and NDVI for Mustard; MCARI, MSR2, and TVI for lentils) using the enhanced learning algorithms was evaluated based on MAE and R^2 value. The whisker plot between in-situ and retrieved LAI of Lentils, Maize, and Mustard crops using SVR-Li, SVR-Poly, SVR-Rad, RFR, PLSR, and HyFIS models are shown in Figure 6.11. In addition, the numerical values of R^2 between in-situ and estimated LAI for each model was also added within the whisker plot. The high LAI retrieval accuracy of Lentils and maize crops was found using the SVR using the radial kernel model followed by SVR using the linear kernel, HyFIS, PLSR, and RFR. However, for Mustard crop, the high LAI retrieval accuracy was found using the SVR-Rad model followed by SVR-Li, PLSR HyFIS, and RFR. On comparing results of SVR-Rad, the LAI retrieval showed promising results with MAE= 0.017; R^2 =0.99 for Mustard; MAE= 0.020; R^2= 0.99 for Maize, MAE=0.036; R^2=0.99 for lentils.

6.4.3.2 LCC retrieval of Lentils, Maize and Mustard crop beds using SVR, RF, PLSR, and HyFIS models

The optimized SVR model parameters with radial, linear and polynomial kernel using 8-fold cross-validation method for LCC retrieval of Lentils, Maize, and Mustard crop beds are shown in Tables 6.4, 6.5, and 6.6, respectively. Figures 6.12, 6.13, and 6.14 represent average square error against the number of trees for lentils, Maize, and Mustard crop beds under different fertilizers and irrigation treatments. The predictive ability of PLSR and HyFIS were also investigated for LCC retrieval by optimizing their respective model parameters. A comprehensive evaluation technique was required to assess the retrieval accuracy of LCC using the SVR-Li, SVR-Poly, SVR-Rad, RFR, PLSR, and HyFIS models for each crop bed of Maize, Mustard, and Lentils under varying irrigation and treatment conditions. The top three optimal VIs selected for four crop beds of Lentils, Maize, and Mustard under different fertilizers and irrigation treatment were found different, as shown in Figures 6.7, 6.8, and 6.9. Therefore, the comparative performance of three optimal VIs set w.r.t. their crop beds within the enhanced learning algorithms is illustrated in the form of a whisker plots, as shown in Figures 6.15, 6.16, and 6.17. Based on the computed performance indices (i.e., MAE and R^2), the retrieval of LCC using SVR-Rad, SVR-Li, PLSR, and HyFIS were found quite well, whereas RFR and SVR-Poly models performed poorly. On comparing results based on the regression models' performance indices, the high LCC retrieval accuracies with MAE=0.193 and R^2=0.99 were found for Maize CB-3 using the SVR-Rad model followed by SVR-Li PLSR, HyFIS, SVR-poly, and RFR. However, for Lentils CB-2 and mustard CB-2, High retrieval accuracy with MAE=0.10 and R^2=0.99 and MAE=0.067 and R^2=0.9**,** respectively, were found using the SVR-Rad model. Interestingly, out of four crop beds of lentil, Maize, and Mustard crop, the crop bed having high VIP and R^2 values of three optimum selection of VIs was found best for vegetation biochemical parameter retrieval. In conclusion, the SVR-Rad model outperformed the SVR-Li, PLSR, HyFIS, SVR-Poly, and RFR models in terms of robustness and extremely stable outliers for vegetation biophysical and biochemical parameters retrieval using hyperspectral data.

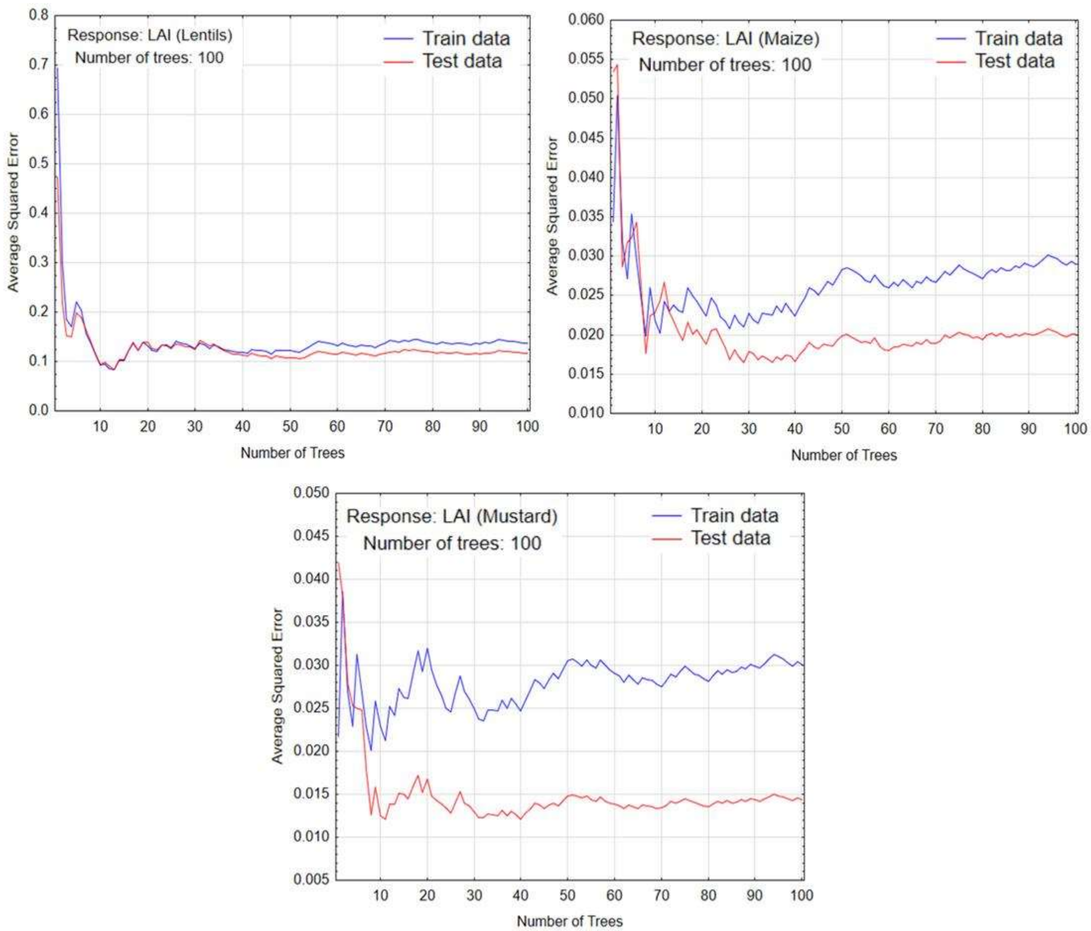

Figure 6.10. RFR optimization by minimizing average squared error using forward tree selection technique for Lentil, Maize, and Mustards LAI retrieval.

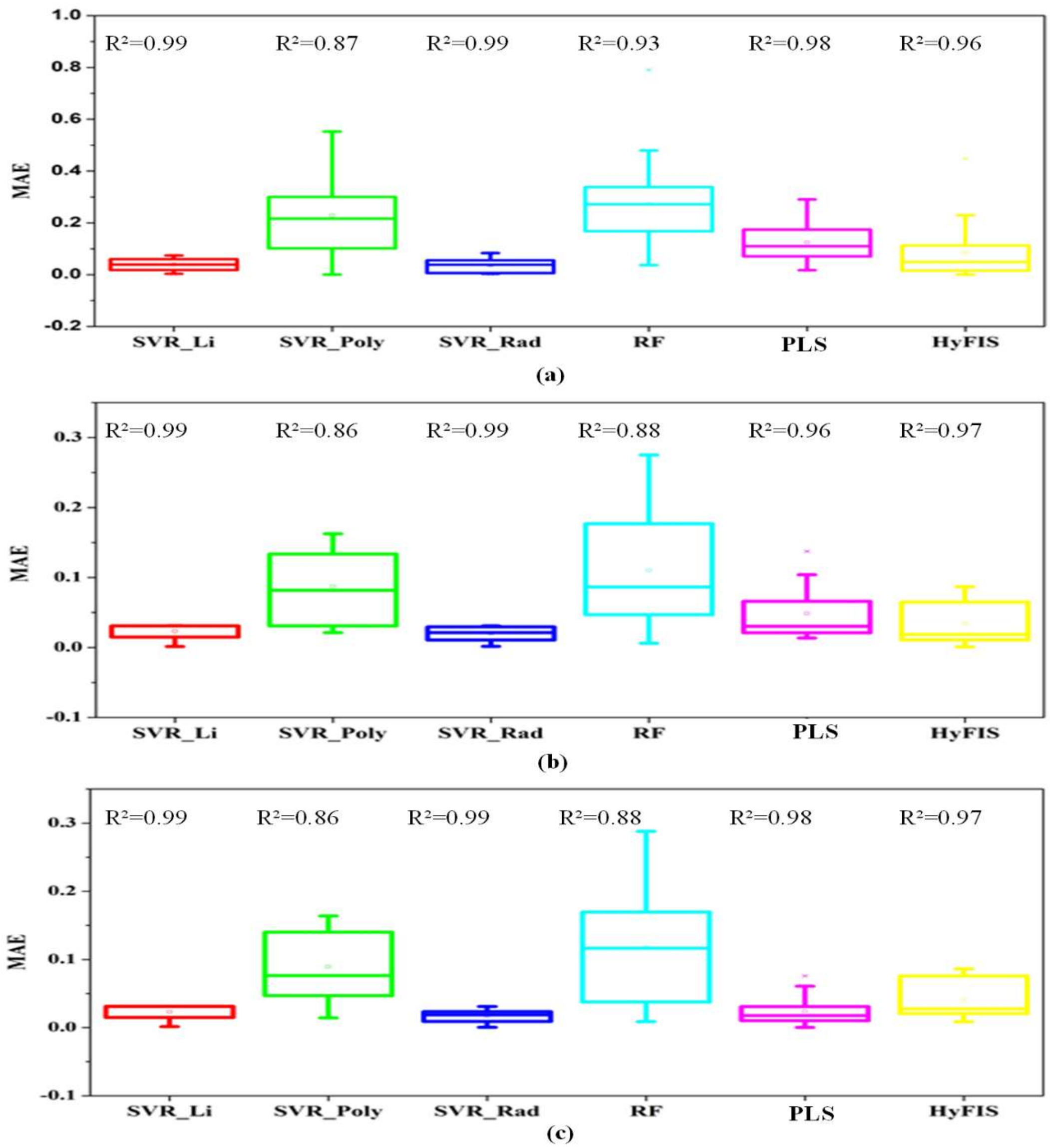

Figure 6.11. Performance of all enhanced learning algorithms for LAI retrieval using three optimum Vis for (a) Lentils, (b) Maize, and (c) Mustard crop.

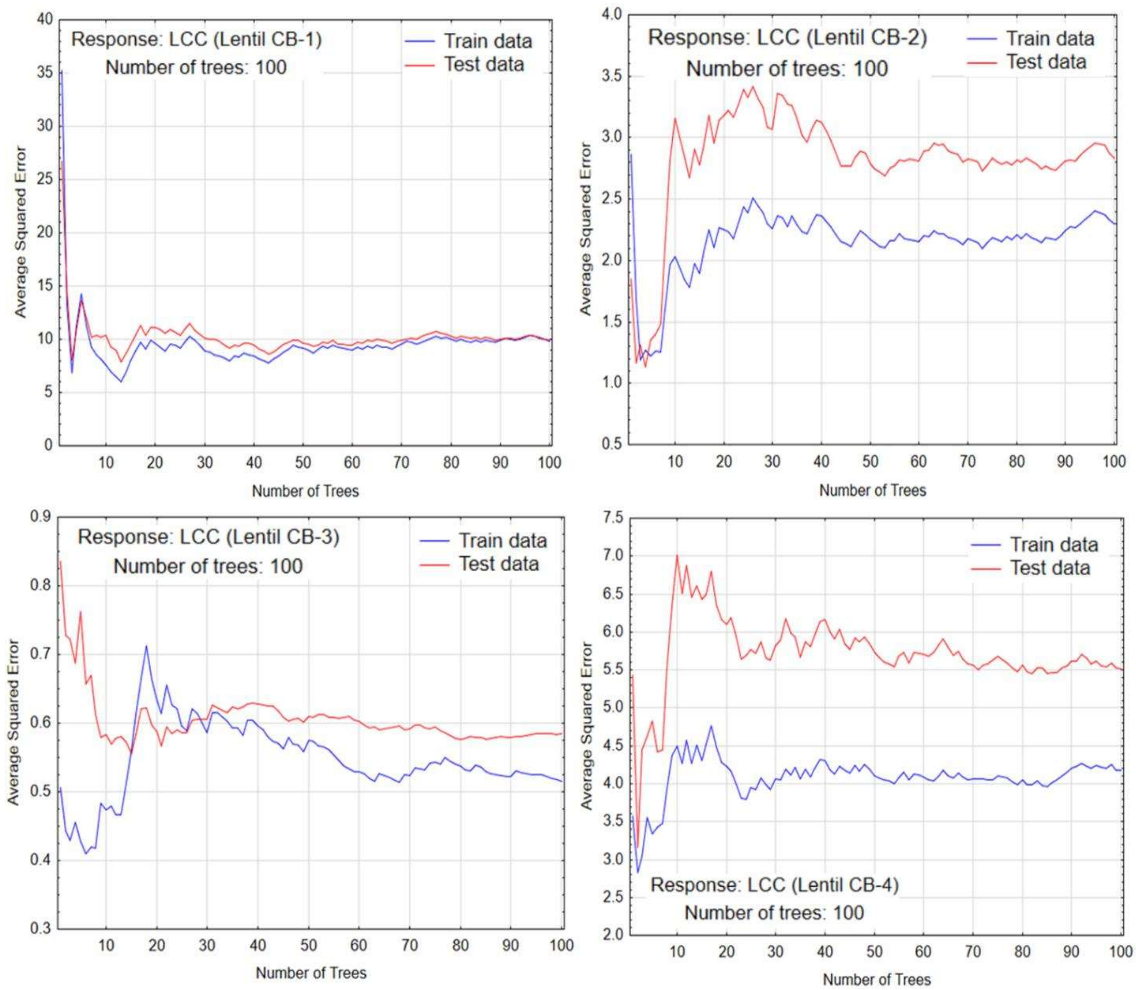

Figure 6.12. RFR optimization by minimizing average squared error using forward tree selection technique for Lentils CBs LCC retrieval.

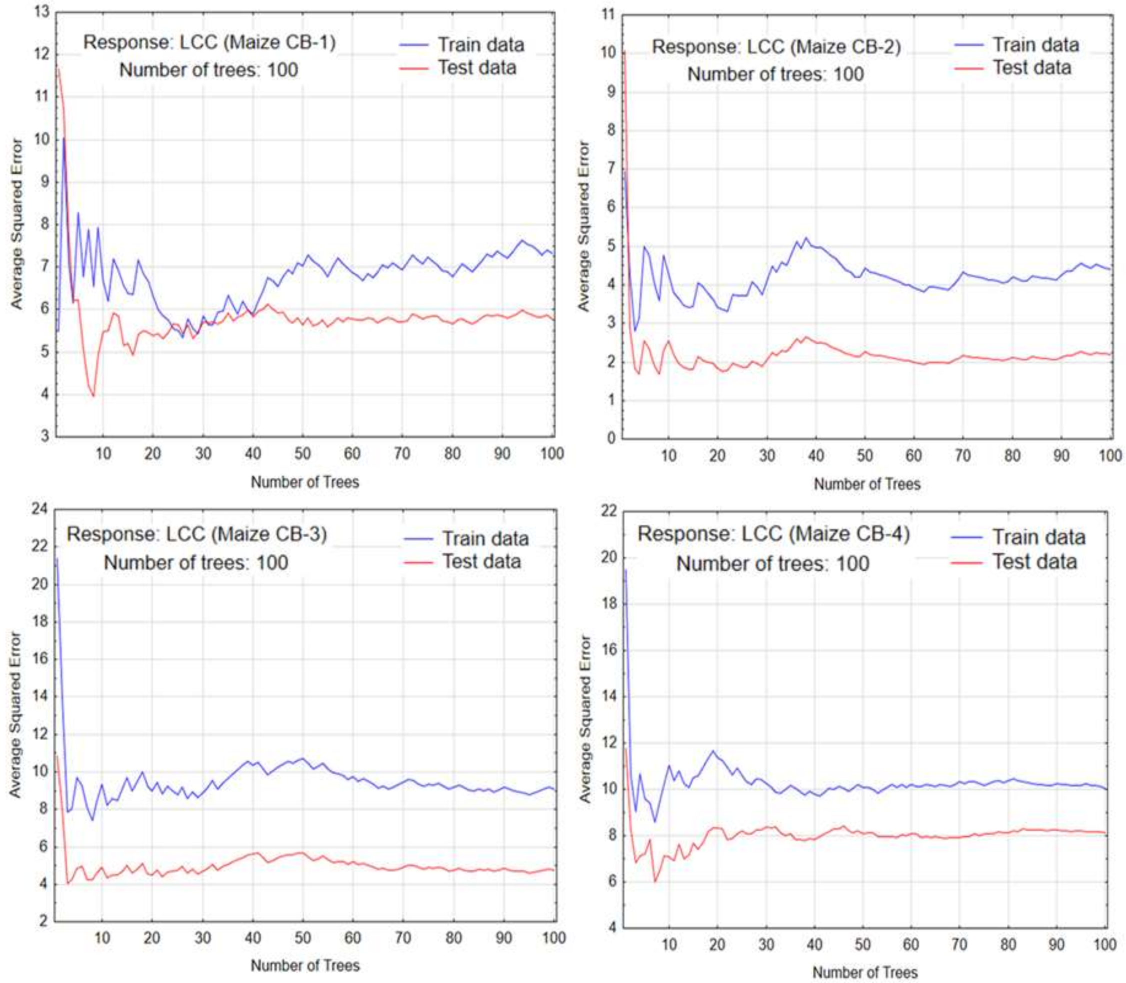

Figure 6.13. RFR optimization by minimizing average squared error using forward tree selection technique for Maize CBs LCC retrieval.

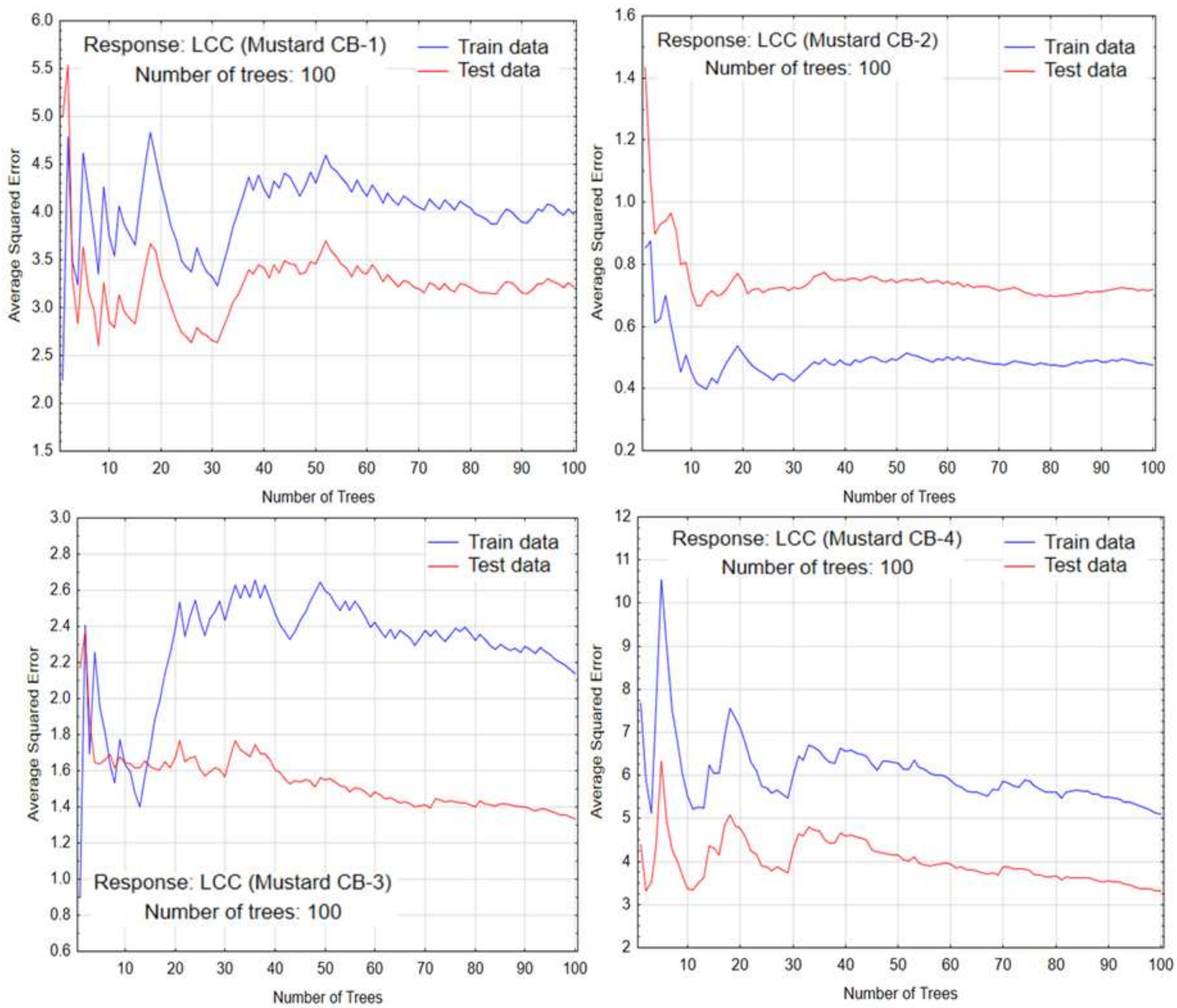

Figure 6.14. RFR optimization by minimizing average squared error using forward tree selection technique for Mustard CBs LCC retrieval.

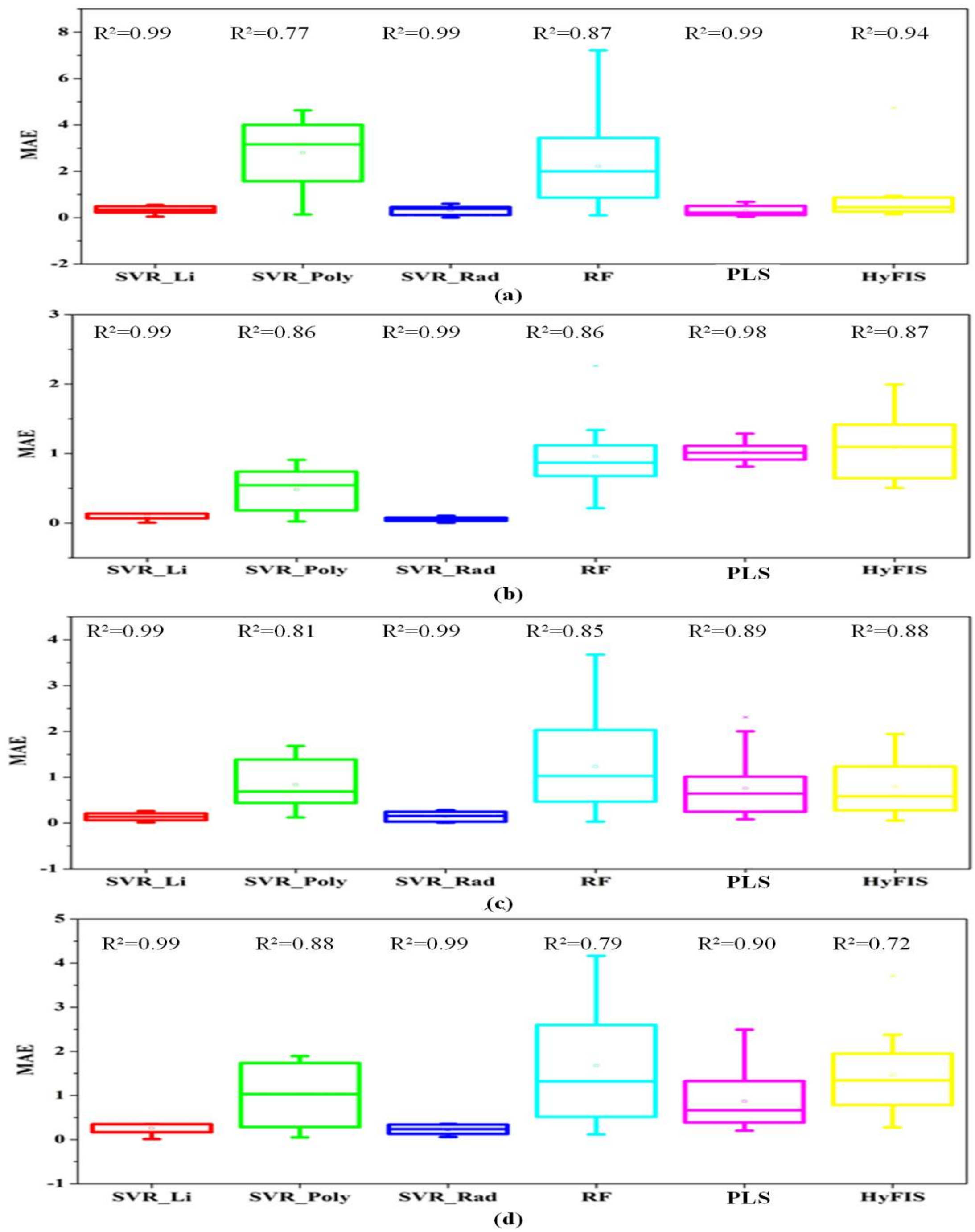

Figure 6.15. Performance of all enhanced learning algorithms for LCC retrival using three optimum VIs for Lentils (a) CB-1; (b) CB-2; (c) CB-3; and (d) CB-4.

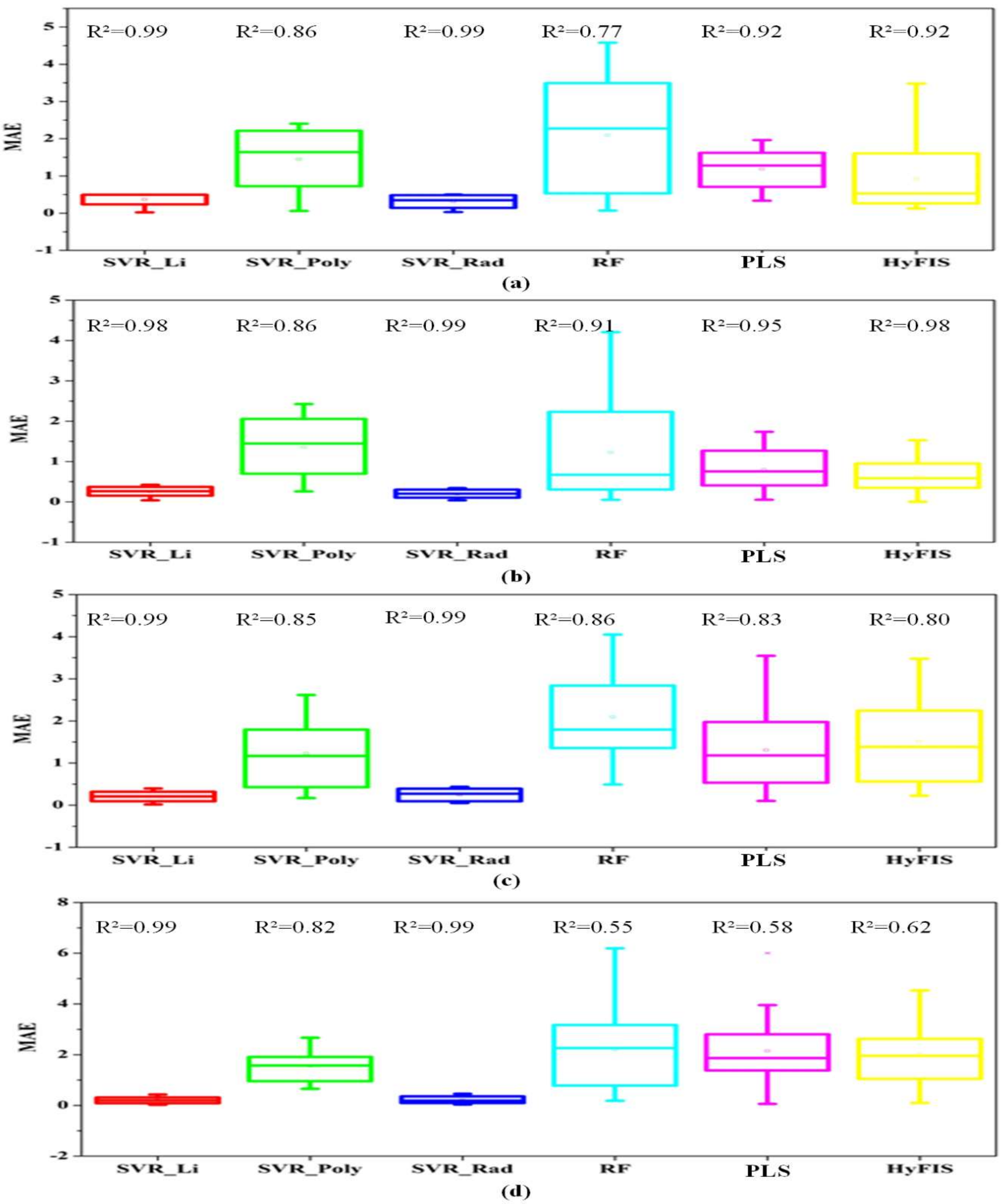

Figure 6.16. Performance of all enhanced learning algorithms for LCC retrival using three optimum VIs for Maizes (a) CB-1; (b) CB-2; (c) CB-3; and (d) CB-4.

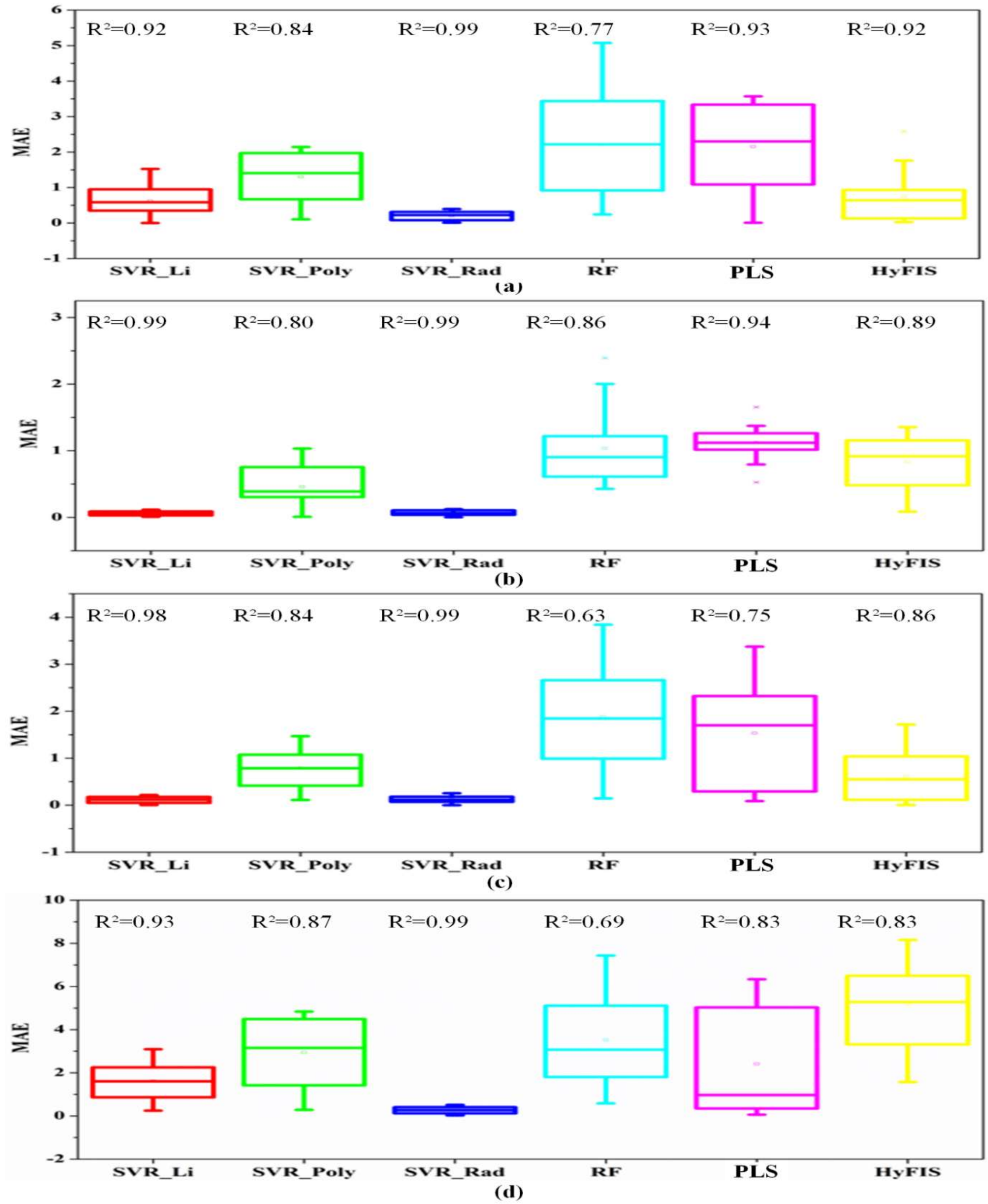

Figure 6.17. Performance of all enhanced learning algorithms for LCC retrival using three optimum Vis for Mustards (a) CB-1; (b) CB-2; (c) CB-3; and (d) CB-4.

Table 6.4. Optimized SVR model parameters for Mustard's LCC estimation.

Mustard	Kernels	Optimized parameters
CB-1	Linear	$\varepsilon = 0.1$, C = 4
	Polynomial	$\varepsilon = 0.3$, C =12, $\gamma = 0.33$, d=3
	Radial	$\varepsilon = 0.1$, C = 4,$\gamma = 0.33$
CB-2	Linear	$\varepsilon = 0.1$, C = 32
	Polynomial	$\varepsilon = 0.3$, C = 16, $\gamma = 0.33$, d=3
	Radial	$\varepsilon = 0.1$, C = 4, $\gamma = 0.33$
CB-3	Linear	$\varepsilon = 0.1$, C = 4
	Polynomial	$\varepsilon = 0.5$, C = 16, $\gamma = 0.33$, d=3
	Radial	$\varepsilon = 0.1$, C = 4, $\gamma = 0.33$
CB-4	Linear	$\varepsilon = 0.9$, C = 32
	Polynomial	$\varepsilon = 0.9$, C = 32, $\gamma = 0.33$, d=3
	Radial	$\varepsilon = 0.1$, C = 4, $\gamma = 0.33$

Table 6.5. Optimized SVR model parameters for Maize's LCC estimation.

Maize	Kernels	Optimized parameters
CB-1	Linear	$\varepsilon = 0.1$, C = 4
	Polynomial	$\varepsilon = 0.4$, C = 8, $\gamma = 0.33$, d=3
	Radial	$\varepsilon = 0.1$, C = 4,$\gamma = 0.33$
CB-2	Linear	$\varepsilon = 0.1$, C = 4
	polynomial	$\varepsilon = 0.3$, C = 32, $\gamma = 0.33$
	Radial	$\varepsilon = 0.1$, C = 4,$\gamma = 0.33$
CB-3	Linear	$\varepsilon = 0.1$, C = 4
	polynomial	$\varepsilon = 0.2$, C = 32, $\gamma = 0.33$, d=3
	Radial	$\varepsilon = 0.1$, C = 4,$\gamma = 0.33$

CB-4	Linear	ε = 0.1, C = 32
	polynomial	ε = 0.3, C = 32, γ = 0.33, d=3
	Radial	ε = 0.1, C = 4,γ = 0.33

Table 6.6. Optimized SVR model parameters for Lentil's LCC estimation.

Lentils	Kernels	Optimized parameters
CB-1	Linear	ε = 0.1, C = 4
	polynomial	ε = 0.7, C = 4, γ = 0.33, d=3
	Radial	ε = 0.1, C = 4,γ = 0.33
CB-2	Linear	ε = 0.1, C = 4
	polynomial	ε = 0.4, C = 32,γ = 0.33, d=3
	Radial	ε = 0.1, C = 4,γ = 0.33
CB-3	Linear	ε = 0.1, C = 4
	polynomial	ε = 0.5, C = 32, γ = 0.33, d=3
	Radial	ε = 0.1, C = 4,γ = 0.33
CB-4	Linear	ε = 0.1, C = 4
	polynomial	ε = 0.5, C = 32, γ = 0.33, d=3
	Radial	ε = 0.1, C = 4,γ = 0.33

Table 6.7. Optimized SVR model parameters for Maize, Mustard and lentils LAI estimation.

Crops	Kernels	Optimized parameters
Maize	Linear	ε = 0.1, C = 4
	polynomial	ε = 0.3, C = 32, γ = 0.33, d=3
	Radial	ε = 0.1, C = 4,γ = 0.33
Mustard	Linear	ε = 0.1, C = 4

	polynomial	ε = 0.3, C = 32,γ = 0.33, d=3
	Radial	ε = 0.1, C = 4,γ = 0.33
Lentils	Linear	ε = 0.1, C = 4
	polynomial	ε = 0.2, C = 16, γ = 0.33, d=3
	Radial	ε = 0.1, C = 4,γ = 0.33

6.5. CONCLUSION

The potential of different enhanced learning algorithms was exploited in this work to estimate LCC and LAI. The 25-hyperspectral VIs at both leaf and canopy level for LCC and 7- hyperspectral VIs for LAI at canopy level were rigorously assessed to find out optimum VIs over Maize, Mustard, and Lentils under different irrigation and fertilizers treatment. The PLSR and R^2 approach were used to extract the sensitive information from the available sets of VIs and scored them in descending order with respect to their VIP value. Furthermore, the robustness of the optimum sets of three VIs having high VIP and R^2 values were evaluated via generated models. Selected optimum vegetation indices based on their sensitivity were found to have excellent crop LAI and LCC estimation capability. Modification of classical VIs using NIR and red-edge (namely, MSR2) was found highly sensitive to crop LAI. Whereas, due to various fertilizers and different irrigation treatments, several VIs such as MCARI, TCARI2/OSAVI2, TCARI2, SR1 for lentils crop beds; MCARI2, SR1, SR2, and SIPI for Maize crop beds; and SR4, MCARI/OSAVI and MCARI for Mustard crop beds were found sensitive with high VIP values to crops LCC. Due to the difference in fertilizer and irrigation treatment to the crop beds of the same species, we get different sensitive VIs for each crop bed for LCC, whereas in case of LAI the top three Vis chosen were almost same for each species. The finding of indices sensitivity can be explained by the fact that each crop's spectral response is altering because of the varying nutrition of vegetation. The analysis showed that the SVR model using radial kernel was best for estimating LCC and LAI, followed by the SVR model using linear kernel function and PLSR. The least performing algorithms are SVR with polynomial kernel function, HyFIS model, and RFR generated model. The performance of each model was evaluated based on MAE and R^2 values. The investigation of robust vegetation indices over crops under different treatments could further collaborate with the future hyperspectral mission, Airborne Visible Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) and Hyperspectral Infrared Imager (HyspIRI) for LAI and LCC parameters retrieval.