Remote Sensing Applications: Society and Environment 35 (2024) 101225

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

A comprehensive deep learning approach for harvest ready sugarcane pixel classification in Punjab, Pakistan using Sentinel-2 multispectral imagery

Sidra Muqaddas [a], Waqar S. Qureshi [b,*], Hamid Jabbar [a], Arslan Munir [c], Azeem Haider [d]

[a] Department of Mechatronics Engineering, National University of Sciences and Technology, H-12, Islamabad, Pakistan
[b] School of Computer Science, University of Galway, Galway H91 F8DY, Ireland
[c] Department of Computer Science, Kansas State University, Manhattan, KS, USA
[d] Al-Moiz Industries Gulberg Lahore, Pakistan

ARTICLE INFO

Keywords:
Convolution neural network (CNN)
Sentinel-2
Spectral unmixing
Long short-term memory (LSTM)
Normalized difference vegetation index (NDVI)
Deep learning
Sugarcane

ABSTRACT

Sugarcane is an important crop for the production of sugar and ethanol, and its area has increased significantly in recent decades in tropical and subtropical regions. Pakistan is among the top ten producers of sugarcane in the world. Up-to-date and accurate sugarcane maps are critical for monitoring sugarcane acreage, and production and assessing its social, economic, and environmental impacts. A huge amount of work has been published regarding crop monitoring and mapping using remote sensing techniques. This study proposes a deep learning-based framework for pixel-based classification of the sugarcane crop among other popular crops grown in Pakistan (e.g., rice, wheat, and corn) using Sentinel-2 multispectral imagery. The frame work includes selection of Sentinel products (Level-2A), preprocessing, spectral indices extraction, spectral feature compilation, labeling through spectral unmixing and harvest time sugarcane classification. The selection of Sentinel products for each crop field is based on the NDVI values. Different spectral (NDVI, NDWI, DVI, SAVI) and biophysical indices (LAI, FVC) are extracted from these sentinel products. Every pixel is compiled as a 2D feature map containing the time-series (ten-time stamps) evolution of each pixel across twelve spectral bands and six indices. The time-series multispectral feature maps are subjected to bilinear sampling to prepare them for input into different deep learning models. Moreover, the labeling of each pixel is done using linear spectral unmixing to assure the abundance of that relevant crop in each pixel. The data set contains samples from different districts of Pakistan and two combinations of a dataset are formed to check the robustness of the developed methodology i.e., training and testing from the same district and from separate districts. For the first combination of datasets, the F1 score for most of the classification models (Convnet, VGG16, ResNet-50, Inception v3 and LSTM) tested is high nearly about 0.99, and for the second set, LSTM outperformed other models with the F1 score of 0.9. The classified pixels are seamlessly integrated into the classification maps of the respective fields. The harvest time sugarcane classification yields encouraging results when compared to the ConvNext and shows high potential to classify sugarcane among other crops using few numbers of products and is capable of classifying sugarcane from other districts as well.

* Corresponding author.
E-mail address: waqarshahid.qureshi@universityofgalway.ie (W.S. Qureshi).

https://doi.org/10.1016/j.rsase.2024.101225
Received 22 August 2023; Received in revised form 16 December 2023; Accepted 7 May 2024
Available online 22 May 2024

1. Introduction

The global population is growing rapidly, necessitating the adoption of various techniques and methods in the agriculture sector to meet increasing food demands. Crop monitoring and management are crucial for enhancing crop productivity and quality. Agriculture holds significant importance in Pakistan's economy, contributing 24% to the GDP (*Agriculture Statistics | Pakistan Bureau of Statistics*, n. d.) and employing nearly 30% of the country's labor force. However, its contribution to Pakistan's GDP has been declining due to factors such as low productivity, climate change, inadequate infrastructure, pests, and diseases. These factors collectively result in low yields, primarily caused by the lack of adoption of modern monitoring and production techniques practiced in developed countries, and reliance on traditional methods. Pakistan's agriculture sector lacks a centralized database system for the agriculture data and limited amount of accurate and real time data posses challenges to the farmers and policy makers to make informed decisions. Examining historical data in Pakistan highlights that the absence of a policy framework plays a crucial role in the decline of the agricultural system (Rasheed et al., 2021).

Through satellite imagery and data analytics, remote sensing can contribute to the creation of a centralized database system, offering accurate and up-to-date information to farmers and policymakers. In addressing the challenges faced by Pakistan's agriculture sector, integrating advanced technologies such as remote sensing can play a pivotal role. The remote sensing application in agriculture allows for monitoring of crops, assessment of crop health and help in estimating yield.

Sugarcane is one of the top five cash crops in Pakistan, contributing 0.7% to the country's GDP. In the 2020–2021 period, sugarcane was cultivated over an area of 1.2 million hectares, yielding 80 million tons. Apart from sugar and jaggery, sugarcane has other byproducts such as ethanol, alcohol, bagasse, and press mud, which are utilized as fuel, pharmaceutical ingredients, electricity generation, and soil fertility enhancers, respectively. Accurate and timely maps depicting mature sugarcane crops offer valuable information and insights to the sugar industry. These maps can assist in planning harvests, managing the supply chain, estimating yields, allocating resources, and formulating policies. In Pakistan, field agents are tasked with obtaining information on sugarcane availability using mobile apps with offline functionality and GPS tagging. This process is both laborious and time-consuming, typically commencing approximately two months before the harvesting period.

Precision agriculture plays a vital role in identifying crop types, monitoring their growth, estimating yields, and determining irrigation and soil requirements. Traditional methods like ground surveys and aerial imagery capture for crop monitoring are exhausting, time-intensive, and expensive. However, the introduction of remote sensing technology in agriculture since the launch of Landsat-1 in 1972 has made it cost-effective and efficient to monitor and identify crop types and their stages. Crop-type mapping provides valuable insights into crop yields and their economic contributions to the agricultural sector. The European Space Agency launched Sentinel-2A in June 2015 and Sentinel-2B in March 2017, with the aim of improving spectral, spatial, and temporal resolution compared to previous satellites. These satellites provide multi-spectral imagery covering vast land areas, islands, and coastal waters. Coupled with machine learning and deep learning techniques, they enable efficient land cover monitoring without extensive on-ground visits.

Soon after the launch of Sentinel-2, the data was freely available to the research community for establishing effective crop management plans that can help the agriculture sector. The first study involving Sentinel-2 data was the pixel-based classification of crops and object-based classification of tree species using single-date data using a random forest algorithm (Immitzer et al., 2016). Initially, traditional machine learning algorithms were involved to classify the crop types which involved lots of hand-crafted feature extraction. The crop classification algorithms have been applied to single-date data and also to time-series data. Nearly eighty vegetation indices were calculated for a single-date image from spectral bands of Sentinel-2 product to classify crop types by stacking support vector machine (SVM) and random forest (RF), which contributed to higher accuracy than individual performances(Sonobe et al., 2018). SVM demonstrated a promising capability in accurately classifying both single-date and multi-temporal images (Maponya et al., 2020). Dealing with big data imposes a barrier, leading to dimensionality reduction by creating composite images. This sacrifices temporal resolution, but data cube architectures help store, access, and model big data, minimizing losses. A study generated multidimensional data cubes from various medium-resolution satellite data, including S2/MSI. The data cube featuring the original temporal resolution of MSI proved to be more aligned with crop dynamics, making it the optimal choice for obtaining detailed crop mapping within a single growing season (Chaves and Sanches, 2023).

Cloud coverage poses a challenge in satellite imagery as it often leads to missing data. To address this issue, a dynamic time-warping algorithm was applied to uneven time series data of Sentinel-2 to perform object and pixel-based classification (Belgiu and Csillik, 2018). Similarly, the time-weighted dynamic time-warping method was used to identify sugarcane crops in China for the year 2016–2020. The methodology involved the comparison of phenological stages of the sugarcane crop's NDVI time series to the unidentified crop pixels (Moharana, 2021).

The fusion of data from different satellites was utilized to develop an effective classifier for crop mapping. Sentinel-1(SAR) data along with Sentinel-2 (multispectral data) data was used to evaluate the performance of 22 algorithms (Chakhar et al., 2021). Similarly, the Landsat 8, and Sentinel-2A data was used to classify the crops in the southeast region of Spain (Chakhar et al., 2020; Piedelobo et al., 2019). Advancements in data integration methods, including time series of Landsat 8 and Sentinel-1, offer improved mapping opportunities. The combined data approach proved more accurate than using single-sensor inputs (Kordi and Yousefi, 2022).

Besides these conventional machine learning algorithms, deep learning techniques such as, artificial neural networks (ANNs), convolution neural networks (CNNs) and recurrent neural networks (RNNs) have been employed to predict crop classes. These approaches require fewer pre-processing steps and have shown enhanced performance in crop mapping. A densely-connected neural network based algorthim SatRed employed pixel-level classification and demonstrated a high overall accuracy, surpassing seven traditional Machine Learning methods, including Random Forest. The model excelled in predicting various land cover types, with

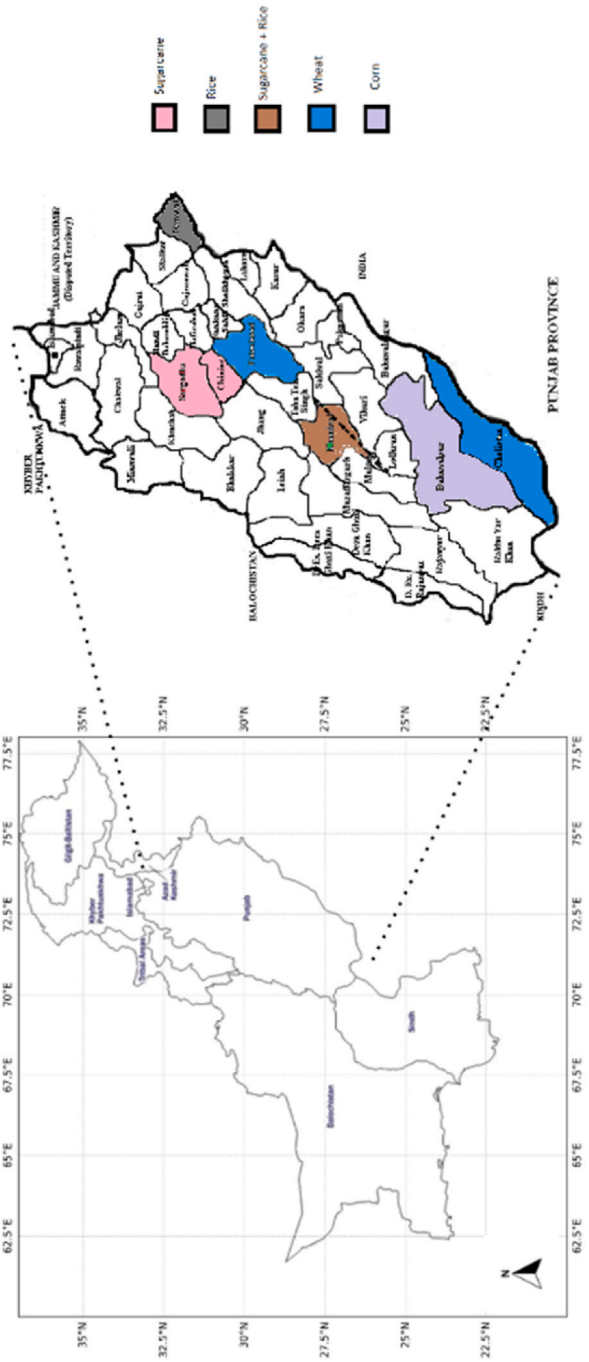

Fig. 1. Study area.

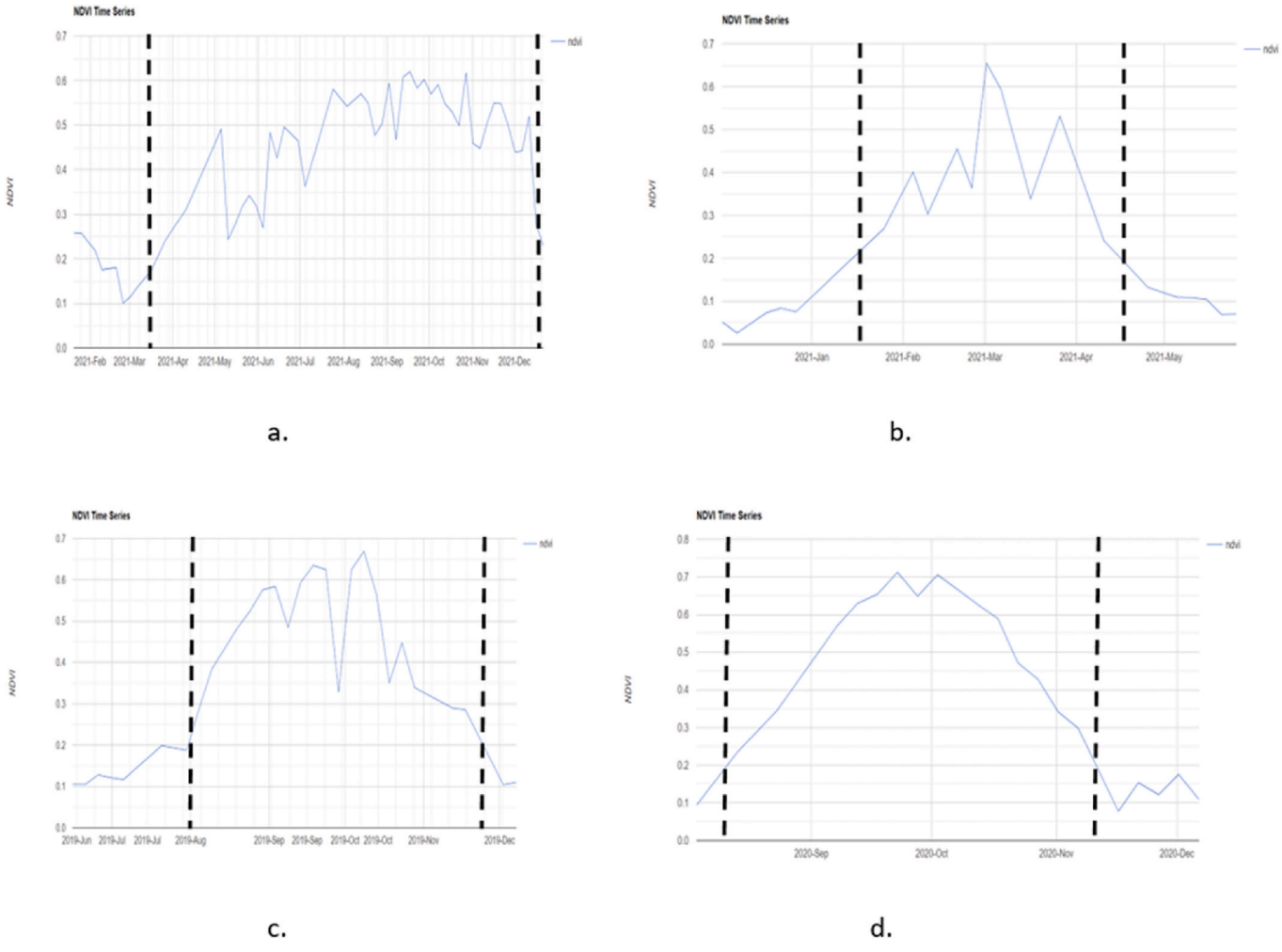

Fig. 2. Time window for Sentinel product selection based on NDVI values for a) sugarcane, b) wheat, c) rice, and d) corn.

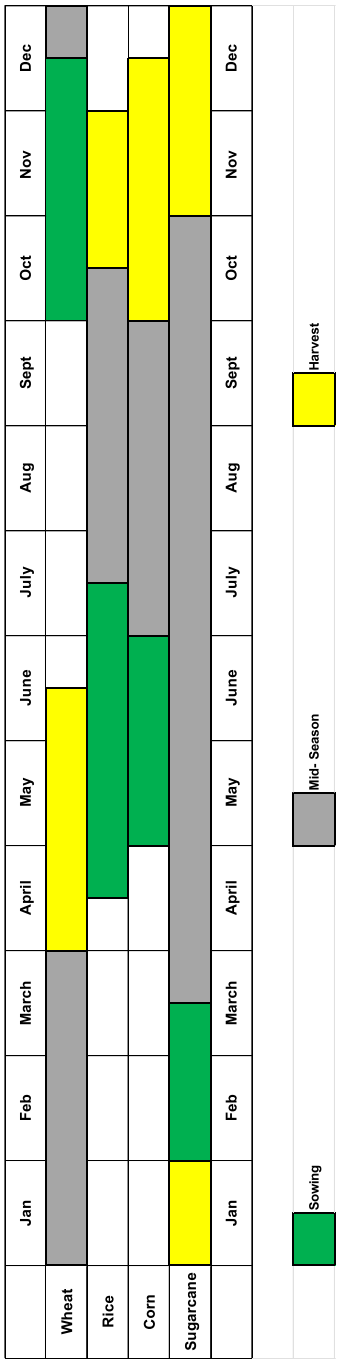

Fig. 3. Crop calendar of investigated crops.

particular success in classifying challenging categories like Fruit crops and Horticulture (Trujillo-Jiménez et al., 2022).

RNNs were used to extract the correlation from the time series data and CNNs are used to extract features based on which classification is done (Mazzia et al., 2019). Various crops were classified by the fusion of these two networks i.e., recurrent-convolutional neural networks by leveraging both temporal and spatial information (Kussul et al., 2020; Li et al., 2023). Most recent studies have shown that the self-attention layer as employed in transformers yield promising performance (Garnot et al., 2022; Ofori-Ampofo et al., 2021).

Similarly, another approach used CNN along with the Geo-convolutional block attention module (Geo-CBAM), where geographical information was integrated with an attention module which focused on informative features and ignored unnecessary information (Wang et al., 2021). CNN with attention layers would be very complex and require complex images whereas our proposed methodology involves separate compilation of each pixel resulting in much simpler images utilizing only ten sentinel-2 multispectral time-series images throughout the growing season of sugarcane.

To classify various crops in Spain, an approach was proposed that encapsulated the time-series information of crop life in the form of an individual image of each pixel (Siesto et al., 2021). Similarly, multi-spectral and time-series data was used to classify rice varieties in Multan Pakistan (Rauf et al., 2022). Both approaches conducted training and testing on the same study area and utilized more than 15 time-series sentinel-2 multispectral images. However, in the former case, the testing year differed, while in the latter case training and testing were conducted for a single growing season using simple convnet(Liu et al., 2022), while other architectures were not explored.

To alleviate the above-mentioned complexities and shortcomings, in this paper, we propose a framework capable of classifying crops from the same or different study areas and different growing seasons, aiming to examine its robustness. Specifically, we focus on pixel-based classification of ready for harvest sugarcane using a reduced number of Sentinel-2 time-series images. Our main contributions in this paper are as follows.

- ➢ Our approach integrates multispectral patterns and the time-series evolution of each pixel, organizing them into a 2-D image. To capture temporal dependencies, we leverage the features of RNNs and long short-term memory (LSTM) networks, allowing each pixel to maintain a record of its previous associated state.
- ➢ In addition to spectral bands, we include six biophysical and spectral indices, stacked together to enhance the efficiency of the developed model. The resulting images are simplified yet contain sufficient information for easy classification using a simple multilayer perceptron (MLP) or CNN in the case of the same district.
- ➢ We have studied sugarcane crops in different growing seasons and in multiple districts (geographical locations) of Punjab, Pakistan. For different districts, we explore various options, including state-of-the-art CNN architectures and LSTM networks. Despite the variations in the dataset, such as location and growing season, the classification results are found to be satisfactory.

This study contributes to the identification of sugarcane around the harvesting period by incorporating ten samples that cover most of its phenological stages, thus allowing to identification of sugarcane fields as ready for harvest. Sugarcane fields that are not in maturity phase might not be classified as ready for harvest fields. Consequently, this method enables the estimation of mature sugarcane availability in a given area using a reduced number of Sentinel-2 multispectral images.

The paper is structured into different sections to explain the proposed framework. In Section 2, the study area and the basic methods used to develop the framework is discussed. Section 3 describes the workflow in order to get the mature sugarcane identified. Section 4 provides a summary of the results obtained from the classification models used in the study. Lastly, Section 5 presents the conclusion and application of proposed framework.

2. Study area and materials

Our study area spans seven districts of Punjab, Pakistan (29.3544° N, 71.6911° E to 33.5651° N, 73.0169° E) as illustrated in Fig. 1. The temperature of the region ranges from 0 to 26 °C in winters and hovers between 28 and 48 °C in summers and has an annual precipitation of 532.384 mm. Sugarcane field data is taken from the districts of Faisalabad, Sargodha, and Khanewal. Field data for other crops are taken (rice from Narowal and Khanewal), wheat from Faisalabad and Cholistan, and corn from Bahawalpur. Sentinel-2 Level-2A products with a cloud coverage of less than 8% are downloaded from the open-access hub (Copernicus Data Space Ecosystem, n.d.). Each product covers an area of 100 × 100 km^2 and two types of products are available Level-1C and Level-2A. Sentinel-2, Level-1C product has thirteen spectral bands, four bands (red (B2), green (B3), blue (B4) and near-infrared (B8)) having 10 m spatial resolution, six bands (red edge (B5), near-infrared NIR (B6, B7, B8A) and short-wave infrared SWIR (B11 and B12)) having 20m, and three bands (coastal aerosol (B1), water vapor (B9) and SWIR-cirrus (B10)) having 60m spatial resolution. Sentinel-2, Level-2A product has 12 bands as band 10 is used in atmospheric correction to obtain the bottom of atmosphere reflectance.

Ground truth was obtained from the Agriculture Robot Lab at NCRA (*NCRA*| National Centre of Robotics and Automation, n.d.), that contained georeferenced field data and sowing date, from which polygons were drawn, and NDVI time-series graphs were obtained from the date of sowing to the estimated harvesting period as shown in Fig. 2. From the NDVI crop signature, a time window was obtained for each field having a threshold NDVI value of 0.2, ten Sentinel-2 Level-2A products were downloaded between those time windows.

Sugarcane in Pakistan is around a 10–12-month crop so 10 Sentinel-2 tiles with a time gap of almost one month are downloaded. Since wheat is a 5–6-month crop, rice is 4-5-month and corn is 3-4-month crop so the time duration gap varies accordingly. Fig. 3 provides a visual representation of the planting and harvesting periods for these four crops. The total field area is around 270 acres, with a sugarcane field area covering 130 acres, and wheat, rice, and corn fields have areas of 45, 55, and 42 acres respectively. The field

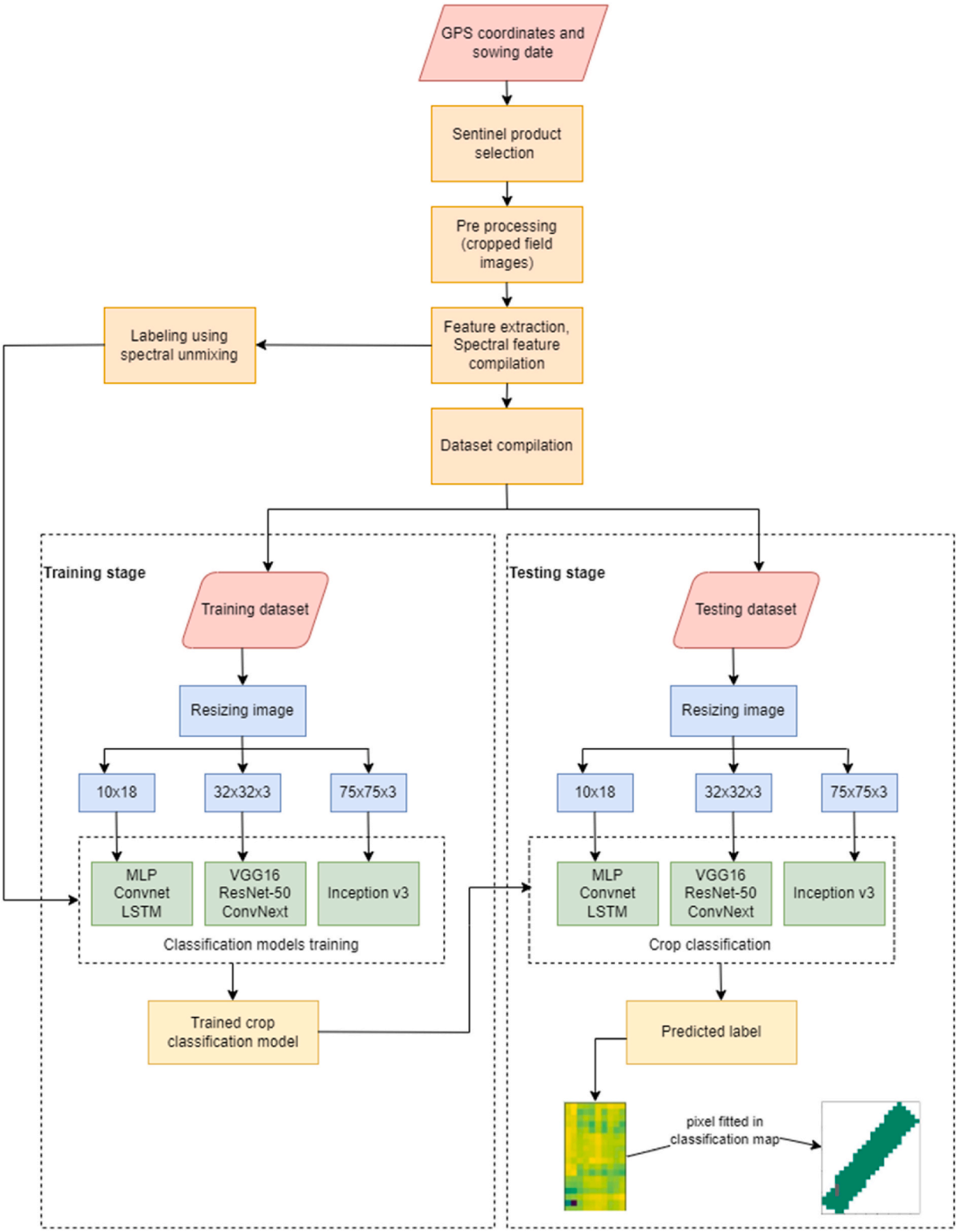

Fig. 4. Block diagram of the proposed framework for crop classification.

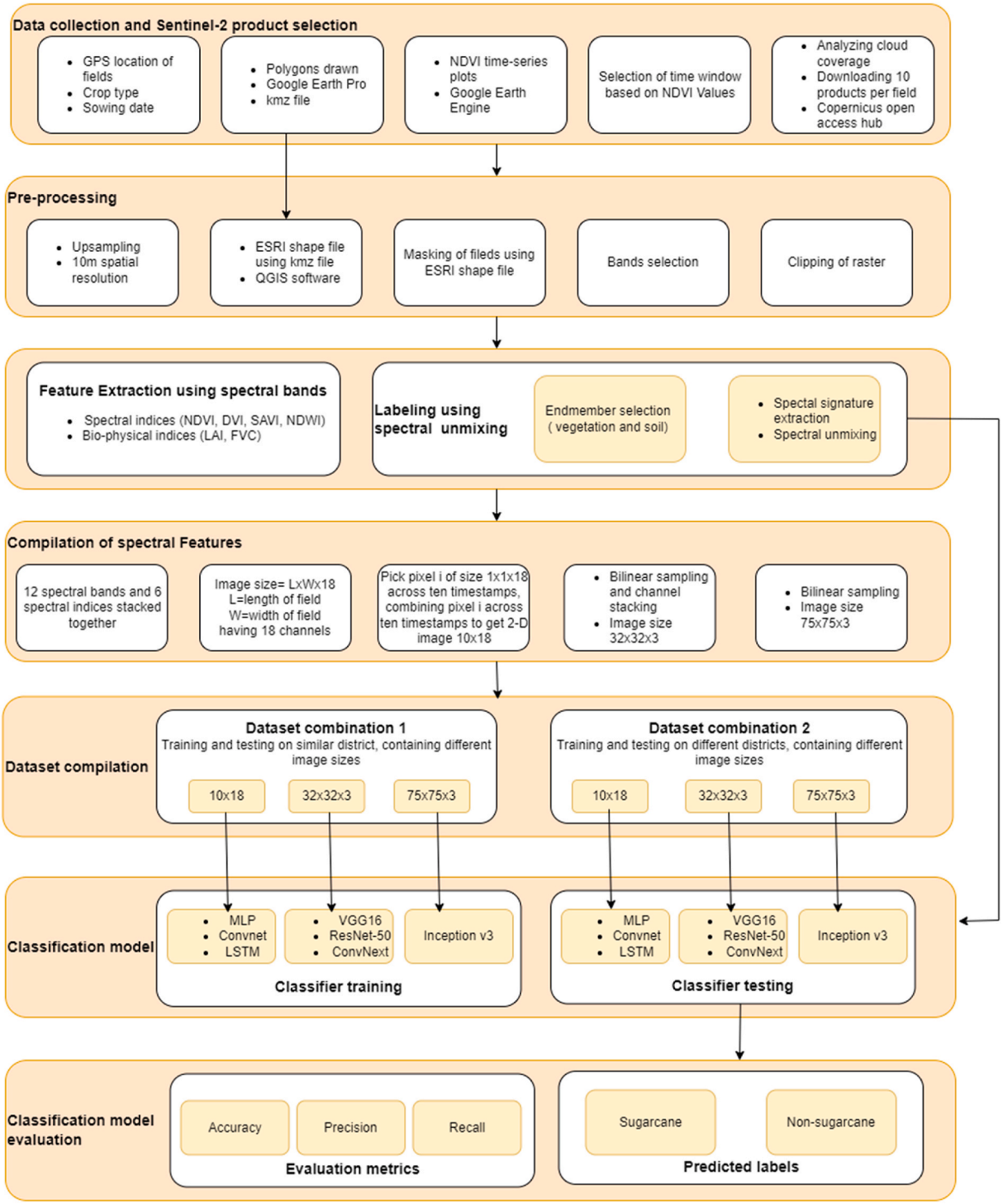

Fig. 5. Detailed workflow of the proposed framework.

data is scattered across the province of Punjab and is not confined to a specific location as opposed to other studies where a specific area is defined and all the crops in the vicinity are mapped using the same Sentinel-2 tile.

2.1. Spectral indices

Spectral indices are the mathematical ratio of different combinations of spectral bands in multispectral imagery and are applied to

each pixel. These spectral indices are used to extract features to map and classify crop types. The **normalized difference vegetation index** (NDVI) measures the density and health of vegetation by calculating the difference between the NIR and red channel of multispectral imagery (Tucker, 1979). **Normalized difference water index** (NDWI) reflects the water content present in vegetation by measuring the difference between the green and NIR channels and it helps to indicate the drought condition in crops (Gao, 1995). **Differenced vegetation index (DVI)** is the simplest of all indices and is sensitive to vegetation changes with respect to background i.e., soil (Richardson and Wiegand, 1977). **Soil-adjusted vegetation index (SAVI)** minimizes the influence of soil brightness by applying a soil brightness correction factor (Huete, 1988). The **leaf area index (LAI)** is a biophysical index that is the ratio of the leaf area to the ground area, it enumerates the number of leaves in the canopy (Williams, 1946). **Fractional vegetation cover (FVC)** tells the percentage of soil covered by green vegetation.

2.2. Classification models

2.2.1. VGG16

VGG16 is a CNN architecture consisting of 21 layers, including 13 convolutional layers, 5 max pooling layers, and 3 dense layers. Its unique feature is the emphasis on 16 wt layers, which are the learnable parameters. VGG16 maintains a consistent arrangement of convolution and max pool layers, utilizing 3×3 filters with a stride of 1 for convolutions and 2×2 filters with a stride of 2 for max pooling. The convolution layers are configured with 64, 128, 256, and 512 filters respectively. Three fully connected layers follow the convolutional stack, with the first two having 4096 channels each and the third performing classification with 1000 channels representing classes (Deng et al., 2009). The architecture concludes with a soft-max layer (Krizhevsky et al., 2012).

2.2.2. ResNet-50

ResNet-50 is a deep CNN that consists of 50 layers. It introduces the concept of residual connections, which help address the problem of vanishing gradients in deep networks. It introduces a bottleneck design to optimize training speed and reduce the number of parameters(He et al., 2016).

2.2.3. Inception v3

Inception v3 was proposed by (Szegedy et al., 2016). It employs inception modules, which consist of parallel convolutional layers with diverse filter sizes, enabling the network to extract features from various spatial scales. Additionally, the architecture incorporates techniques such as factorized convolutions and auxiliary classifiers to enhance training and overall performance.

2.2.4. Long short-term memory

LSTM is a type of RNN that can learn long-term dependencies in sequence data (Hochreiter and Schmidhuber, 1997). LSTM can store selective information with the help of gate functionality, which enables saving short-term or long-term information, and helps to avoid the vanishing gradient problem in traditional RNNs. They are used in a wide variety of applications, that is, language modeling, sequence classification, image captioning, etc.

2.2.5. ConvNext

With the advancement in deep learning technologies, RNN and LSTM are being replaced mostly by transformers (Vaswani et al., 2017) in the NLP domain. CNNs ruled the computer vision domain for over a decade. Despite the difference in the domain of vision and language, the network architecture from two different domains converged into new architecture vision transformers (Dosovitskiy et al., 2020). When compared with convnets in image classification they proved to have promising results as compared to CNN but don't perform well in image segmentation. To overcome this problem Swin transformer was proposed by (Liu et al., 2021) which reintroduced the concept of sliding windows to transformers and performed well in problems that are beyond classification, describing the essence of convnet has not vanished. So ConvNext is proposed in which ResNet50 was used as a baseline architecture and progressed with the hierarchical construction similar to the Swin transformer (Liu et al., 2022).

3. Methodology

3.1. Sentinel-2 product selection

Our proposed crop classification framework is presented in Fig. 4 and a detailed workflow is shown in Fig. 5, which involves selecting satellite images on those dates where our crop field NDVI value is greater than 0.2 and cloud coverage is less than 8 percent. In Pakistan, during the months of July and August cloud coverage is very large usually around 70–100%. During this time the products are analyzed thoroughly if the area of interest, that is, the concerned field is cloud-free then those Sentinel-2 products are used in dataset compilation.

3.2. Preprocessing

All the preprocessing and feature extraction are done using the Sentinel application platform (SNAP) (Brockmann Consult, 2014). All the bands of Sentinel products having a resolution lower than 10m are upsampled to 10m resolution using the bilinear up-sampling method. Reprojection is not done as the product is already in UTM/WGS84 coordinate reference system. Polygons of each field are drawn in Google Earth Pro (*Google Earth Pro*, 2022) and then converted to ESRI shapefile in QGIS software which is further used to clip the area of interest from the Sentine-2, Level 2A products as shown in Fig. 6. After preprocessing and polygon clipping fields are obtained that contain 12 spectral bands i.e., red, blue, green, NIR, SWIR, etc.

3.3. Feature extraction

Using spectral bands several spectral indices, that is, DVI, SAVI, NDVI, NDWI, LAI, and FVC are obtained in SNAP. All the values of these indices are normalized between the range 0–1 as it can be visualized in Fig. 7.

3.4. Spectral feature compilation

Spectral bands of the cropped fields and the corresponding six indices are stacked together. Each pixel is compiled separately where each column represents spectral bands and indices and each row corresponds to the evolution of that pixel throughout the crop life. Since 10 Sentinel products on the basis of NDVI value, covering whole phenological stages of each crop are selected and used in image compilation along with 18 spectral bands and indices, arranging them in 2D matrix results in an image size of 18×10. For transfer learning images are upsampled using bilinear interpolation to a size of $32 \times 32 \times 3$. For CNN architectures like VGG16 and ResNet-50, the smallest image size in Keras framework should be $32 \times 32 \times 3$ and for Inception v3 it should be $75 \times 75 \times 3$. A comprehensive graphical representation of spectral features compilation is shown in Fig. 8.

3.5. Labeling

In multispectral images, those pixels whose spectral values are a combination of two or more materials are referred to as mixed pixels whereas pixels that correspond to a single object reflectance are pure pixels. Spectral signatures of green vegetation and soil are extracted in SNAP and then an abundance map (proportion of each endmember present in each pixel) of each field at the peak of the crop is calculated using these two endmembers, that is, green vegetation and soil as shown in Fig. 9. Despite careful field geo-referencing that considers crop field boundaries, there is a risk of misidentifying weed or tree areas as green vegetation due to the relatively large pixel size of 10×10 m. Abundance values, in the green vegetation abundance map, greater than 0.5 correspond to the crop pixel and less than 0.5 are treated as soil pixels.

3.6. Training and testing datasets

Our dataset has a lot of variation in terms of location, sowing year, and crop stage during Sentinel image selection. The dataset is balanced having almost the same number of synthetic images for the sugarcane and non-sugarcane classes that include wheat, rice, and corn. The first set of experiment datasets from all the locations is compiled together and then divided into training and validation datasets using a ratio of 65:20 and 15% of data is allocated for testing. For the second set of experiments sugarcane data from Khanewal and Sargodha districts are used for training purposes and tested on the data from Chiniot district. Similarly, for other classes, the data for training purposes is taken from different districts and tested in different districts.

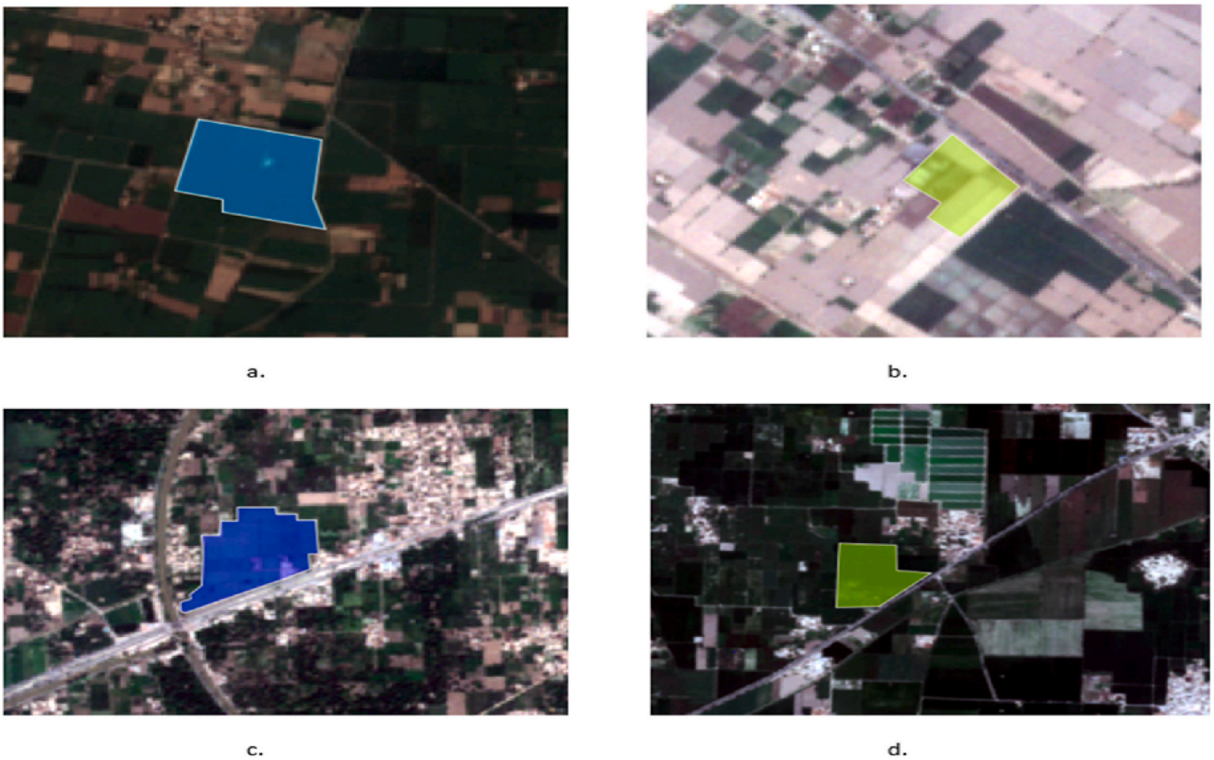

Fig. 6. Polygons drawn in Google Earth Pro and imported in SNAP for clipping of sentinel product to get the fields for a) sugarcane, b) wheat, c) rice, and d) corn.

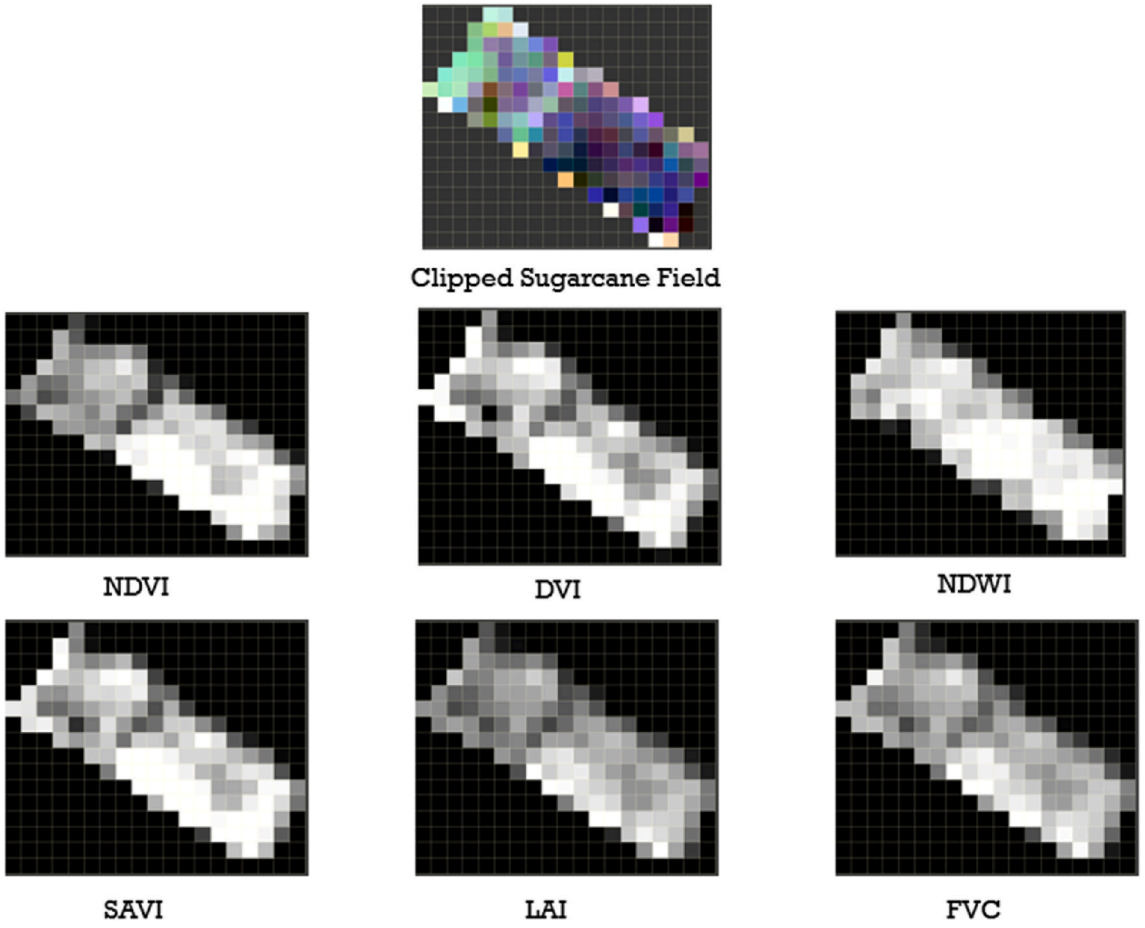

Fig. 7. Spectral and bio-physical Indices.

3.7. Classifier training

After the compilation and division of the dataset in training and validation units, it was fed through various already developed CNN architectures, simple sequential convnet, multi-layer perceptron, and LSTM. The CNN architectures used in this study are VGG16, ResNet-50, Inception V3, and ConvNext. To transfer the image classification knowledge previously attained by these models, these pre-trained models are used to classify sugarcane among other crops. In first case, we are using the pre-trained portion of the feature extractor by freezing those layers and training only the classification layers. In the second case, we are training all the layers of these CNN models on our dataset. We have trained and optimized all the classification models using Keras framework on Google Colab.

4. Results and discussion

4.1. Evaluation metrics

For the performance evaluation of classification models used in our framework, accuracy, precision, recall and F1 score are used and are computed in Eq. 1, Eq. 2, Eq. 3 and Eq.4. Accuracy is the most frequently employed evaluation metric in classification, representing the ratio of correctly predicted labels to the total number of labels. Precision is a measure that assesses the correctness of predictions by calculating the proportion of true positives (TP) to the total number of positive predictions (TP + FP). A higher precision value indicates more accurate predictions with a lower rate of false positives. It quantifies the model's effectiveness in correctly classifying positive instances.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

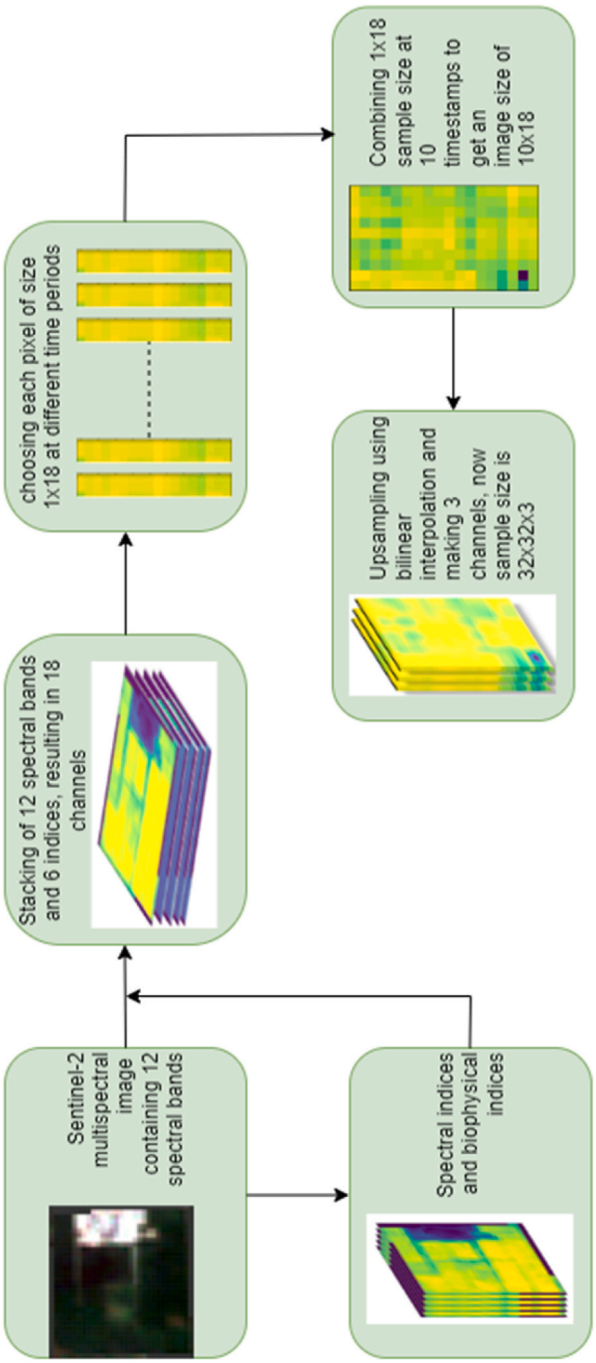

Fig. 8. Image compilation, from a clipped field containing 12 spectral bands to $32 \times 32 \times 3$ sized image.

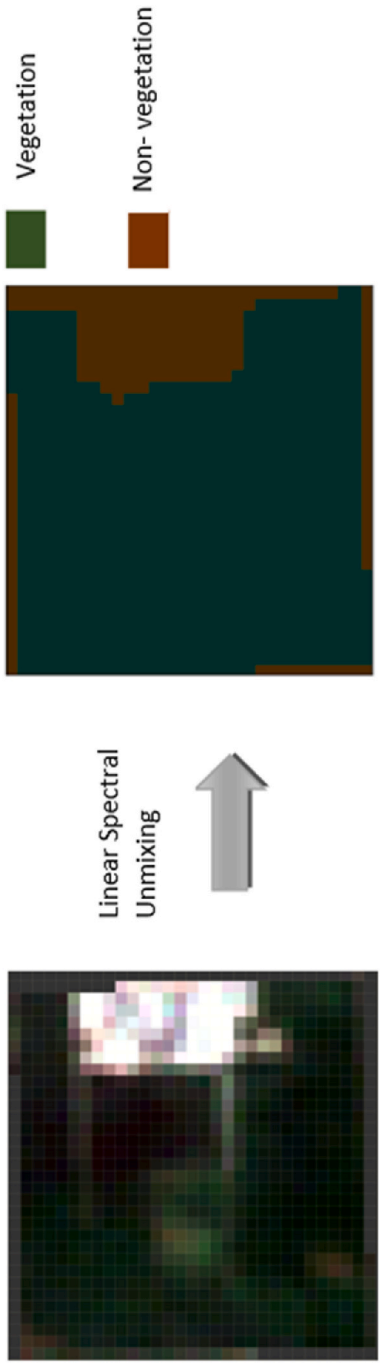

Fig. 9. Labeling of sentinel images using spectral unmixing.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Recall measures the ratio of true positives (TP) to the total number of actual positive cases (TP + false negatives (FN)). A higher recall value signifies that the model correctly identifies a larger proportion of positive cases as positive. It evaluates the model's capability to capture the majority of positive instances. The F1 score is a metric used to evaluate the performance of a classification model. It combines precision and recall into a single value, providing a balanced measure of the model's accuracy. It finds particular application in scenarios where there exists a disparity in the quantity of positive and negative instances within the dataset.

4.2. Experimental results and discussion

For the first set of dataset combinations, datasets from all the districts are collected together and then divided into training and validation datasets and tested on the fields from the same district. Formerly we have employed 1D data to multilayer perceptron to perform classification. Each column of the 2D time-series multispectral feature map is trained as a single sample of size 1×18, representing 12 spectral bands and 6 indices of each pixel. Then the image of size 10×18 is flattened and subsequently fed to a multilayer perceptron with an input sample size of 1×180. In the same way, 2D time-series multispectral feature maps are classified using a simple CNN and some state-of-the-art architectures of CNN. Evaluation metrics and input sample size are shown in Table I and confusion matrix for all the models are shown in Fig. 10.

Our test results for all three CNN architectures (VGG16, ResNet-50 and Inception v3) are similar. Our sugarcane pixels are classified among all other crop pixels i.e., wheat, rice, and, corn, with the F1 score of 0.99 for above mentioned three CNN architectures and precision close to 1 representing sugarcane is accurately predicted with nearly zero false positive predictions as presented in Table I'. Moreover, the results are almost the same with the transfer learning and parameter training. For comparison purposes, the dataset is also classified using LSTM, having ten-time stamps and a number of features equal to eighteen. Classification map of sugarcane and non-sugarcane test field for dataset combination 1 using ResNet-50 is shown in Fig. 11. In this experiment, 928 pixels of sugarcane field are labeled as vegetation and soil pixels, soil pixels are discarded and remaining pixels are compiled as 2D spectral feature maps and tested on ResNet-50 trained model. Similarly, a total of 568 pixels of non-sugarcane test field is compiled as 2D spectral features, and tested on same trained model.

For the second combination of datasets, the results are not as impressive as those of the first combination. This dataset combination has a high variance and noise that leads to low comparable accuracies as compared to the first combination. In this case, LSTM outperforms the convnets and MLP by allowing the F1 score of 0.90, thus learning the time sequence dependencies in the data. VGG16 and ResNet-50 performance are nearly equal, that is, 0.87 and 0.86 respectively as they managed to extract an adequate amount of the relevant features from different crops despite variations in location, sowing year, and crop stage at the sentinel product selection. In the case of Inception v3 using Keras framework, the input sample size must be at least $75 \times 75 \times 3$. When a 10×18 image size is upsampled using the interpolation method a lot of noise is added into the data and because of this the model fails to fit the test data accurately. Low accuracy as compared to the first case is due to the high variance in the dataset. Fitting the model to different training sets leads to slight changes in model parameters and variance can be reduced by adding data, similar to the test data, to the training dataset. Evaluation metrices of all the classification models for dataset combination 2 are shown in Table II.

LSTM trained model is tested on sugarcane fields from Chiniot and rice field from Khanewal. Classification map of sugarcane and non-sugarcane test field for dataset combination 2 using ResNet-50 is shown in Fig. 12.

The performance of ConvNext was subpar in both training and testing on the same dataset, whereas other architectures excelled. For ConvNext, the accuracy for both dataset combinations are approximately 51%, indicating that the ConvNext model fails to capture the features present in the training data. However, after applying data augmentation techniques, the training accuracy has improved to 91% and the test accuracy to 78.20%. This suggests that the model is capable of performing well when the dataset size is increased. Fig. 13 illustrates a graphical representation of the training and test accuracy for the dataset combination 2 and it is presented in tabular form in Table III.

From the results, it can be seen that to make our classifier more robust for sugarcane and other crops, data from all over Pakistan should be taken and included in our training data. The same crop characteristics may vary from area to area as soil, variety of the crop, and crop duration are not similar. Moreover, the number of products should also be increased to make our model learn every single

Table 1
Evaluation metrics along with input sample size for dataset combination 1.

Input sample size	Architecture	Test accuracy	Precision	Recall	F1 Score
1×18	MLP	0.77	0.97	0.56	0.71
1×180	MLP	0.94	1.00	0.89	0.94
$10 \times 18 \times 1$	Simple convnet	0.99	0.99	0.99	0.99
10×18	LSTM	0.92	1.00	0.85	0.92
10×18	LSTM + MLP	0.98	1.00	0.97	0.98
$32 \times 32 \times 3$	VGG16	0.99	0.99	0.99	0.99
$32 \times 732 \times 3$	Resnet50	0.99	1.00	0.98	0.99
$75 \times 75 \times 3$	Inception V3	0.99	1.00	0.99	0.99

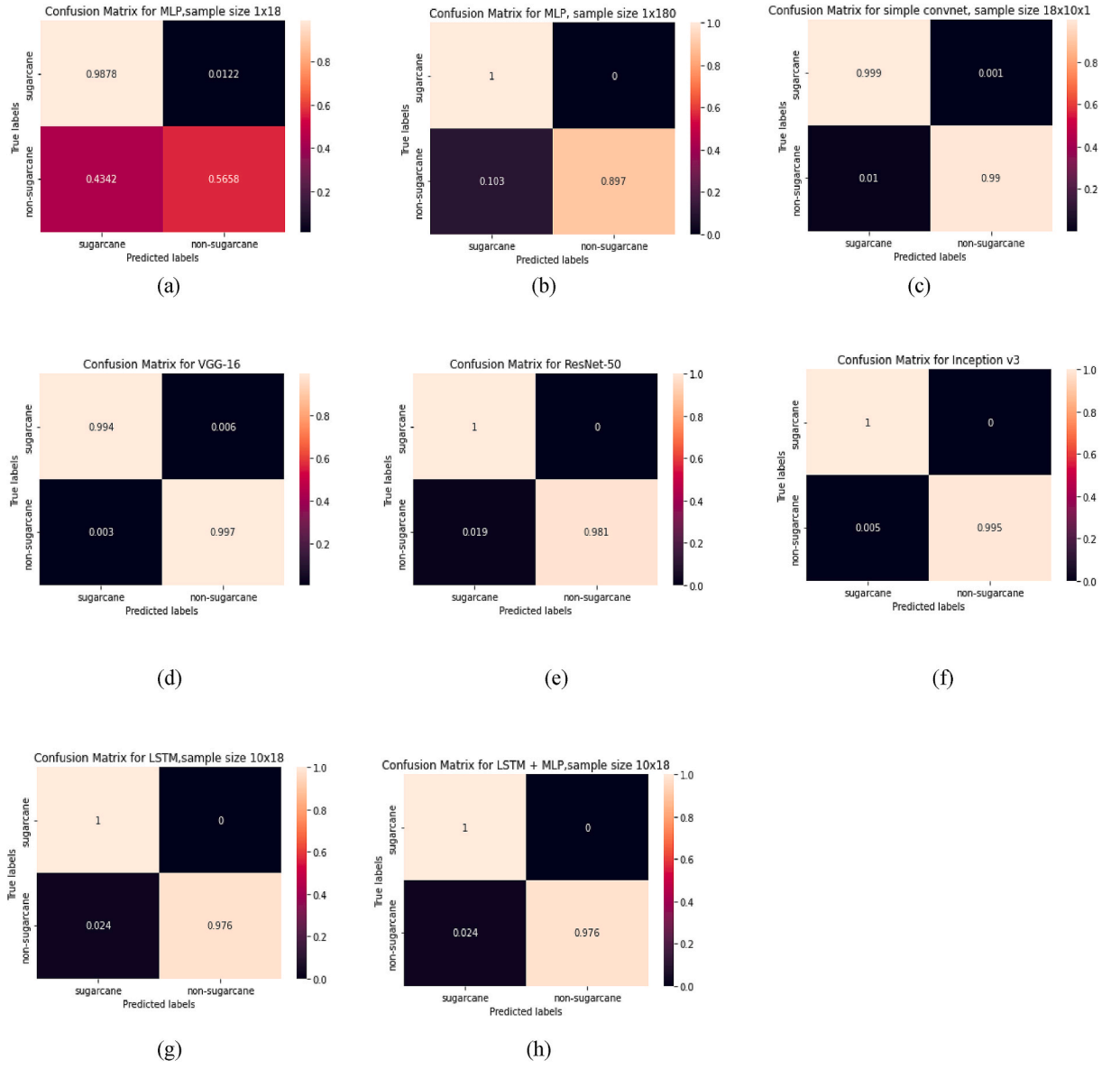

Fig. 10. Confusion matrix for the architectures used for classification of dataset combination 1. (a) MLP with input sample size 1×18, (b) MLP with input sample size 1×180, (c) simple convnet with input sample size $10 \times 18 \times 1$, (d) VGG16 with input sample size $32 \times 32 \times 3$, (e) ResNet-50 with input sample size $32 \times 32 \times 3$, (f) Inception v3 with input sample size $75 \times 75 \times 3$, (g) LSTM with input sample size 10×18, (h) LSTM + MLP with input sample size 10×18.

attribute of crop phenology and acquire a better understanding of crop characteristics. The performance of classification models for both the dataset combination is shown in Fig. 14.

Our developed methodology shows promising results interms of classification accuracy as compared to the previous studies as our data has a lot of variability and we have transformed 13 channel multispectral images to 2D spectral maps which are capable of classifying with higher accuracy even with simple ANNs. We conducted an intense study to incorporate data from various districts and different variety of crops and tried different classification algorithms to get the most optimum classification maps which resulted in higher classification results.

Our developed methodology demonstrates promising results in terms of classification accuracy compared to previous studies. Given the sentinel-2 images with high spatial resolution and multispectral capabilities, we transformed 13-channel multispectral images into 2D spectral maps, enabling more accurate classification even with simple artificial neural networks (ANNs). Through an extensive study incorporating data from various districts and diverse crop varieties, we experimented with different classification algorithms to obtain optimal classification maps, leading to improved classification results.

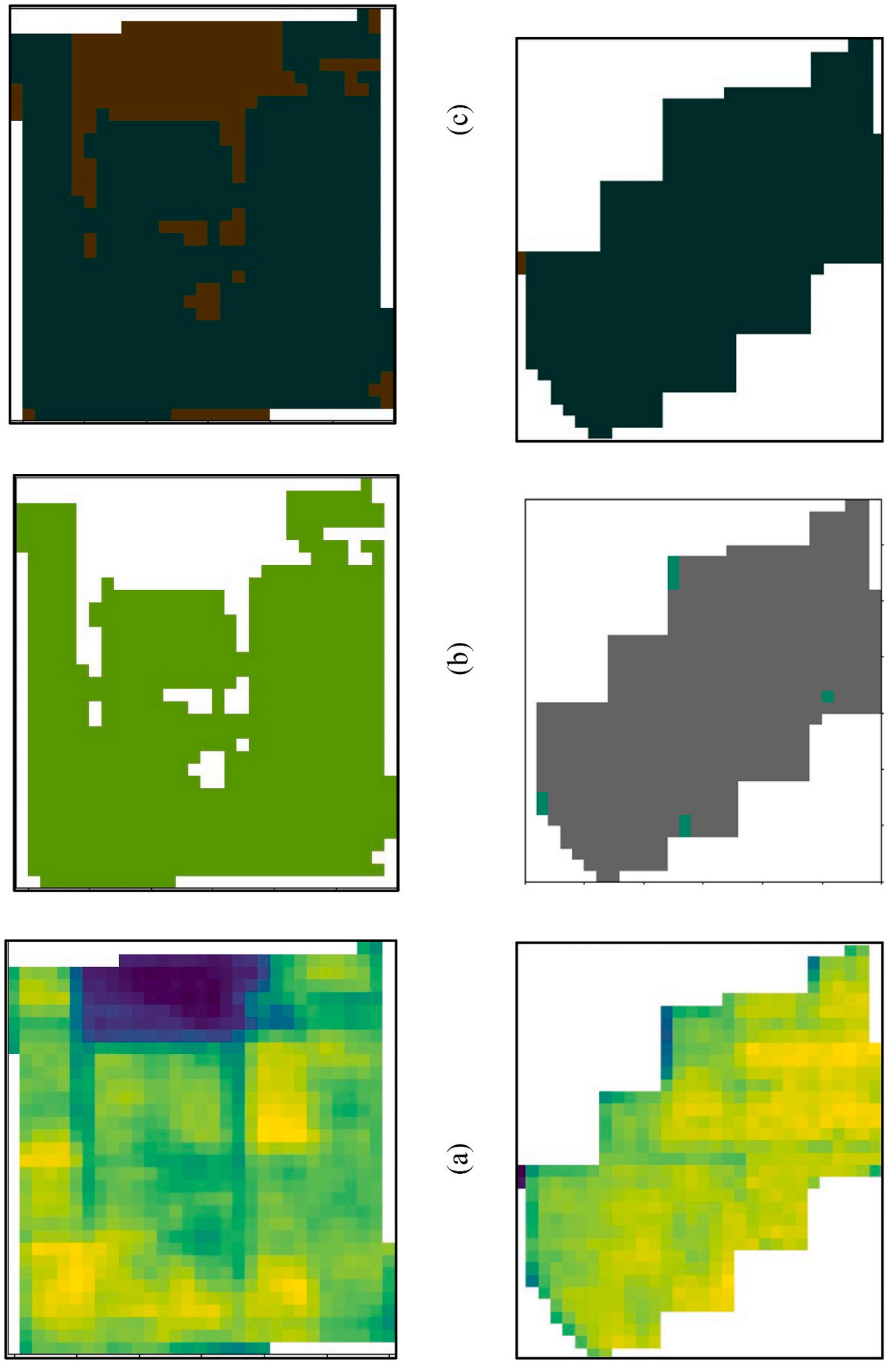

Fig. 11. Dataset combination 1: a) Cropped multispectral image of sugarcane field, (b) Labeling using spectral unmixing, (c) Classification map using ResNet-50, where green pixels represent sugarcane and grey represents non-sugarcane, d) Cropped multispectral image of non-sugarcane field, e) Labeling using spectral unmixing, f) Classification map, where green pixels represent sugarcane and grey represents non-sugarcane

Table 2
Evaluation metrics along with input sample size for dataset combination 2.

Input sample size	Architecture	Accuracy	Precision	Recall	F1 Score
10×18	MLP	0.66	0.67	0.64	0.65
10×18	Simple convnet	0.72	0.71	0.73	0.72
10×18	LSTM	0.90	0.87	0.94	0.90
$32 \times 32 \times 3$	VGG16	0.88	0.85	0.91	0.87
$32 \times 32 \times 3$	ResNet-50	0.85	0.98	0.78	0.86
$32 \times 32 \times 3$	ConvNext	0.78	0.75	0.85	0.79
$75 \times 75 \times 3$	Inception v3	0.68	0.70	0.66	0.68

5. Conclusions

Accurate information about the readiness of sugarcane fields for harvesting ensures that the crops are harvested at the optimal time. This leads to improved yield and quality of sugarcane, as harvesting at the right maturity level maximizes sugar content and reduces losses due to overripening or delays. It also provides valuable insights for long-term planning and market forecasting.

In this paper, we have proposed a deep learning-based framework for identifying ready to harvest sugarcane fields among other popular crops grown in Pakistan. Proposed framework includes Sentinel-2 Level 2A product selection based on NDVI time-series plots followed by preprocessing and spectral feature extraction in SNAP. The temporal and multispectral features are arranged in 2D images referred to as 2D time-series multispectral feature maps in this study. These feature maps are classified using multiple classification models and the developed framework shows positive potential to classify sugarcane feature maps from other major crop feature maps when it comes to the same district. Despite the challenges posed by our dataset's variability in location and dates, and the limited number of products used, our methodology excelled in feature extraction and pixel-based classification, yielding impressive results. Through experimentation, it has been determined that a quantity of ten time-series samples is sufficient for extracting the essential features throughout the entire crop life cycle for each crop. Furthermore, additional tests are conducted using five and seven time-series samples to accurately map crops, but the obtained results are not satisfactory; thus, making ten time-series samples a right choice to obtain sufficient multitemporal features.

The classification task focused on distinguishing sugarcane from crops like wheat, rice, and corn, which have shorter growth periods of 4–6 months, unlike sugarcane that takes 10–12 months to reach maturity in Pakistan. To ensure fairness, we have collected an equal number of Sentinel-2 tiles for all crops resulting in a large time gap in the case of sugarcane. This methodology involves feature extraction and compilation of each pixel as an image and works well even with MLP and simple convnet, allowing the F1 score of 0.94 and 0.99, respectively, when trained and tested from the same district. Results indicate that deep CNN architectures and LSTM perform well in recognizing sugarcane from other districts whereas MLP and simple Convnet perform poorly. The sugarcane classification task achieves impressive results using LSTM, VGG16, and ResNet-50 models, with the F1 scores of 0.90, 0.87, and 0.86, respectively.

To improve the performance of the proposed framework with the capability of identifying mature sugarcane fields across Pakistan, data from most of the sugarcane growing districts can be added in the training dataset. Moreover, the choice of ten samples should cover all growing phases of sugarcane, that is, germination, tillering, grand growth and ripening and maturation phase. Furthermore, it's important to note that our current study is confined to Punjab, Pakistan, and has yet to undergo comprehensive testing utilizing data from other sugarcane-growing regions, both within and outside Pakistan. Future research endeavors will address this limitation by expanding the scope to include lower Punjab and Sind, where sugarcane cultivation differs in terms of land characteristics and the varieties cultivated by farmers. This broader examination will provide a more robust understanding of the spectral features' collective behavior, shedding light on specific features that significantly contribute to enhanced classification results.

Ethical statement

The research conducted complies with all relevant national and international ethical guidelines and regulations. Where applicable, informed consent was obtained from all participants involved in the study. Any data or information that could identify individual participants has been anonymized or omitted. The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request, and they comply with all relevant data protection and privacy regulations.

The authors acknowledge the use of OpenAI's ChatGPT model to assist in improving the readability and language of the manuscript. This assistance was limited to language enhancement and did not influence the scientific content, data interpretation, or conclusions of the research.

CRediT authorship contribution statement

Sidra Muqaddas: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – original draft. **Waqar S. Qureshi:** Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Hamid Jabbar:** Supervision, Writing – review & editing. **Arslan Munir:** Supervision, Visualization, Writing – review & editing. **Azeem Haider:** Data curation, Project administration, Resources.

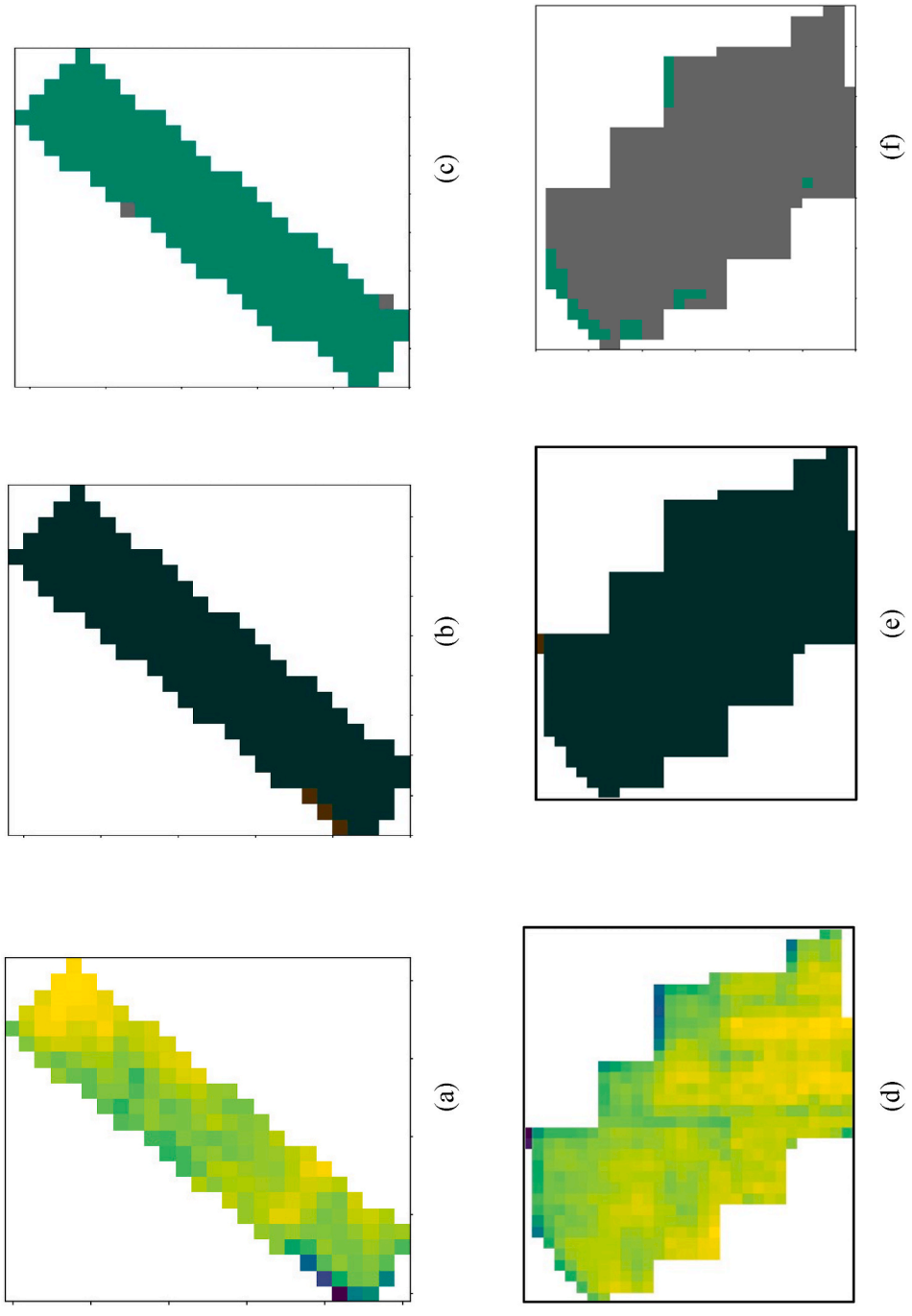

Fig. 12. Dataset combination 2: a) Cropped multi-spectral image of sugarcane field, (b) Labeling using spectral unmixing, (c) Classification map using LSTM, where green pixels represent sugarcane and grey represents non-sugarcane, d) Cropped multispectral image of non-sugarcane field, e) Labeling using spectral unmixing, f) Classification map, where green pixels represent sugarcane and grey represents non-sugarcane

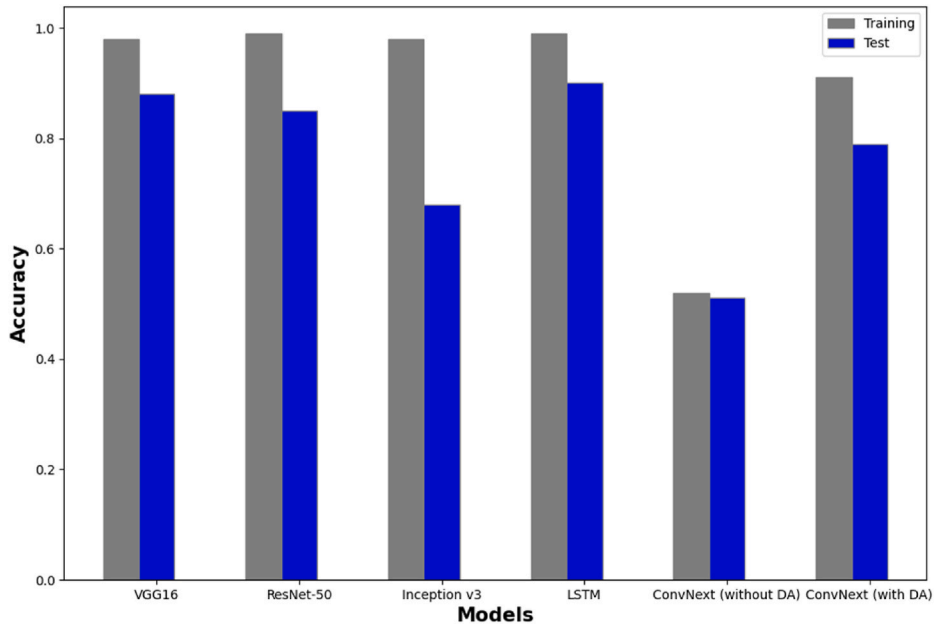

Fig. 13. Comparison of train and test accuracy for different classification models for dataset combination 2, (DA: data augmentation).

Table 3
Comparison of training and test accuracy for dataset combination 2.

Input sample size	Architecture	Training accuracy %	Test accuracy %
$32 \times 32 \times 3$	VGG16	98.40	82.25
$32 \times 32 \times 3$	ResNet-50	99.01	82.24
$75 \times 75 \times 3$	Inception v3	98.11	64.23
$32 \times 32 \times 3$	LSTM	99.60	90.00
$32 \times 32 \times 3$	ConvNext (without data augmentation)	51.91	51.32
$32 \times 32 \times 3$	ConvNext (with data augmentation)	91.41	78.20

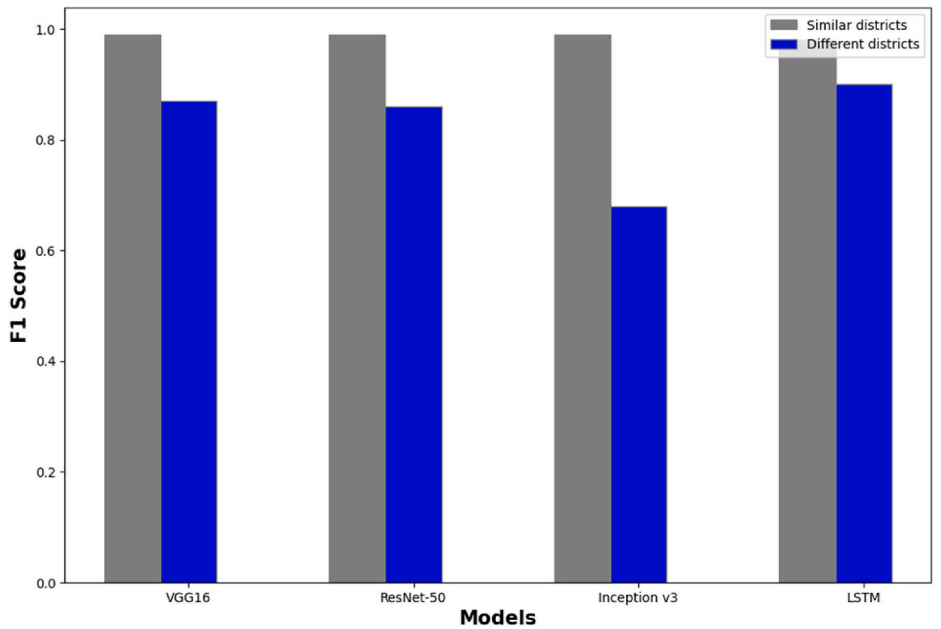

Fig. 14. Performance of classification models on both dataset combinations, that is, (similar districts and different districts).

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgement

This research is supported by the Higher Education Commission of Pakistan and the National Centre of Robotics and Automation, Pakistan under Grant Number DF 1009-0031. We would like to express our gratitude to all those who contributed to this research. Special thanks to our colleagues and collaborators for their invaluable insights and support.We also acknowledge the use of OpenAI's ChatGPT model, which assisted in enhancing the readability and language of the manuscript. This support was limited to language refinement and did not influence the scientific content or conclusions of the study.

References

Agriculture Statistics, Pakistan Bureau of Statistics. (n.d.). Retrieved June 2, 2023, from https://www.pbs.gov.pk/content/agriculture-statistics.
Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Rem. Sens. Environ. 204, 509–523.
Brockmann Consult, S.S., 2014. *SNAP* (9.0.0).
Chakhar, A., Hernández-López, D., Ballesteros, R., Moreno, M.A., 2021. Improving the accuracy of multiple algorithms for crop classification by integrating sentinel-1 observations with sentinel-2 data. Rem. Sens. 13 (2), 243.
Chakhar, A., Ortega-Terol, D., Hernández-López, D., Ballesteros, R., Ortega, J.F., Moreno, M.A., 2020. Assessing the accuracy of multiple classification algorithms for crop classification using Landsat-8 and Sentinel-2 data. Rem. Sens. 12 (11), 1735.
Chaves, M.E.D., Sanches, I.D., 2023. Improving crop mapping in Brazil's Cerrado from a data cubes-derived Sentinel-2 temporal analysis. Remote Sens. Appl.: Society and Environment 32. https://doi.org/10.1016/j.rsase.2023.101014.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., others, 2020. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv Preprint ArXiv:2010.11929*.
Gao, B.-C., 1995. Normalized difference water index for remote sensing of vegetation liquid water from space. Imaging Spectrometry 2480, 225–236.
Garnot, V.S.F., Landrieu, L., Chehata, N., 2022. Multi-modal temporal attention models for crop mapping from satellite time series. ISPRS J. Photogrammetry Remote Sens. 187, 294–305.
Google Earth Pro, 2022. 7.3.6.9345 (Google LLC).
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 770–778.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Rem. Sens. Environ. 25 (3), 295–309.
Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. Rem. Sens. 8 (3), 166.
Kordi, F., Yousefi, H., 2022. Crop classification based on phenology information by using time series of optical and synthetic-aperture radar images. Remote Sens. Appl.: Society and Environment 27. https://doi.org/10.1016/j.rsase.2022.100812.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25.
Kussul, N., Lavreniuk, M., Shumilo, L., 2020. Deep recurrent neural network for crop classification task based on sentinel-1 and sentinel-2 imagery. IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium 6914–6917.
Li, Q., Tian, J., Tian, Q., 2023. Deep learning application for crop classification via multi-temporal remote sensing images. Agriculture 13 (4), 906.
Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision 10012–10022.
Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A convnet for the 2020s. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11976–11986.
Maponya, M.G., Van Niekerk, A., Mashimbye, Z.E., 2020. Pre-harvest classification of crop types using a Sentinel-2 time-series and machine learning. Comput. Electron. Agric. 169, 105164.
Mazzia, V., Khaliq, A., Chiaberge, M., 2019. Improvement in land cover and crop classification based on temporal features learning from Sentinel-2 data using recurrent-convolutional neural network (R-CNN). Appl. Sci. 10 (1), 238.
Moharana, S., Kambhammettu, B.V.N.P., Chintala, S., Sandhya Rani, A., Avtar, R., 2021. Spatial distribution of inter-and intra-crop variability using time-weighted dynamic time warping analysis from Sentinel-1 datasets. Remote Sens. Appl. 24, 100630.
NCRA | National Centre of Robotics and Automation. (n.d.). Retrieved June 2, 2023, from https://ncra.org.pk/.
Ofori-Ampofo, S., Pelletier, C., Lang, S., 2021. Crop type mapping from optical and radar time series using attention-based deep learning. Rem. Sens. 13 (22), 4668.
Copernicus Data Space Ecosystem | Europe's eyes on Earth. (n.d.). Retrieved May 30, 2024, from: https://dataspace.copernicus.eu/.
Piedelobo, L., Hernández-López, D., Ballesteros, R., Chakhar, A., Del Pozo, S., González-Aguilera, D., Moreno, M.A., 2019. Scalable pixel-based crop classification combining Sentinel-2 and Landsat-8 data time series: case study of the Duero river basin. Agric. Syst. 171, 36–50.
Rasheed, N., Khan, S.A., Hassan, A., Safdar, S., 2021. A decision support framework for national crop production planning. IEEE Access 9, 133402–133415. https://doi.org/10.1109/ACCESS.2021.3115801.
Rauf, U., Qureshi, W.S., Jabbar, H., Zeb, A., Mirza, A., Alanazi, E., Khan, U.S., Rashid, N., 2022. A new method for pixel classification for rice variety identification using spectral and time series data from Sentinel-2 satellite imagery. Comput. Electron. Agric. 193, 106731.
Richardson, A.J., Wiegand, C.L., 1977. Distinguishing vegetation from soil background information. Photogramm. Eng. Rem. Sens. 43 (12), 1541–1552.
Siesto, G., Fernández-Sellers, M., Lozano-Tello, A., 2021. Crop classification of satellite imagery using synthetic multitemporal and multispectral images in convolutional neural networks. Rem. Sens. 13 (17), 3378.
Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., Mochizuki, K., 2018. Crop classification from Sentinel-2-derived vegetation indices using ensemble learning. J. Appl. Remote Sens. 12 (2), 26019.
Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.

Trujillo-Jiménez, M.A., Liberoff, A.L., Pessacg, N., Pacheco, C., Díaz, L., Flaherty, S., 2022. SatRed: new classification land use/land cover model based on multi-spectral satellite images and neural networks applied to a semiarid valley of Patagonia. Remote Sens. Appl.: Society and Environment 26. https://doi.org/10.1016/j.rsase.2022.100703.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Rem. Sens. Environ. 8 (2), 127–150.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Wang, Y., Zhang, Z., Feng, L., Ma, Y., Du, Q., 2021. A new attention-based CNN approach for crop mapping using time series Sentinel-2 images. Comput. Electron. Agric. 184, 106090.

Williams, R.F., 1946. The physiology of plant growth with special reference to the concept of net assimilation rate. Ann. Bot. 10 (37), 41–72.