Yuzuncu Yil University
Journal of the Institute of Natural & Applied Sciences
https://dergipark.org.tr/en/pub/yyufbed

Research Article

Crop Classification with Attention Based BI-LSTM and Temporal Convolution Neural Network Combination for Remote Sensing Breizhcrop Time Series Data

Amer Najmualdeen Ali Akbar BANDAR*, Ahmet COŞKUNÇAY

Ataturk University, Engineering Faculty, Department of Computer Engineering, 25100, Erzurum, Türkiye
Amer Najmualdeen Ali Akbar BANDAR, ORCID No: 0000-0001-7699-162X,
Ahmet COŞKUNÇAY, ORCID No: 0000-0002-7411-310X
*Corresponding author e-mail: amerngmalden1@gmail.com

Article Info

Received: 01.08.2023
Accepted: 13.02.2024
Online April 2024

DOI:10.53433/yyufbed.1335866

Keywords
Attention mechanism,
Crop classification,
Land use and coverage,
Remote sensing,
Time series

Abstract: In the modern era, remote sensing data has become increasingly useful for determining land use and coverage requirements. Remote sensing data can be used for a variety of purposes, including the classification of crops. It is possible to aggregate remote sensing data for a specific area over time in order to obtain a more complete picture based on the time series of this data. One example of these types of data is the Breizhcrop dataset, which was collected using satellite images acquired by Sentinel 2 over a period of time. This study aims to investigate a neural network based on attention mechanisms using the BI-LSTM layer in conjunction with Temporal-CNN for the classification of crops. The aim of the research is to find a model for corps classification in image-based time series. In line with this goal, in addition to finding features over time, the presented model also needs to produce high-accuracy features at each time step to increase classification. Utilizing the designed neural network, we seek to find local features with the attention mechanism and general features with a second layer. This neural network was validated on the BreizhCrop dataset and we conclude that it performs better than alternative approaches. The proposed method has been compared with Temporal CNN, Star RNN, and Vanilla LSTM networks and it has obtained better results than the mentioned neural networks. Taking advantage of these local and global features that extract with developed model obtained a high accuracy rate of 82%.

Uzaktan Algılama Breizhcrop Zaman Serisi Verileri için Dikkat Tabanlı BI-LSTM ve Zamansal Evrişimli Sinir Ağı Kombinasyonu ile Mahsul Sınıflandırması

Makale Bilgileri

Geliş: 01.08.2023
Kabul: 13.01.2024
Online Nisan 2024

DOI:10.53433/yyufbed.1335866

Anahtar Kelimeler
Arazi kullanımı ve kapsamı,
Dikkat mekanizması,
Mahsul sınıflandırması,
Uzaktan algılama,
Zaman serisi

Öz: Modern çağda, uzaktan algılama verileri, arazi kullanımı ve kaplama gereksinimlerini belirlemede giderek daha fazla kullanışlı hale gelmiştir. Uzaktan algılama verileri, aralarında mahsul sınıflandırması da bulunan çeşitli amaçlar için kullanılabilir. Belirli bir alan için uzaktan algılama verilerini zaman içinde toplamak, bu verilerin zaman serisi temelinde daha kapsamlı bir görüntü elde etmeyi mümkün kılar. Bu tür verilere örnek olarak, bir süre boyunca Sentinel 2 tarafından elde edilen uydu görüntüleri kullanılarak toplanan Breizhcrop veri seti gösterilebilir. Bu çalışma, mahsullerin sınıflandırılması için BI-LSTM katmanı ile Zaman-İlişkili CNN'nin birleşiminde dayanan, dikkat mekanizmaları temelinde bir sinir ağı araştırmayı hedeflemektedir. Araştırmanın amacı, görüntü tabanlı zaman serilerinde mahsul sınıflandırması için bir model bulmaktır. Bu hedef doğrultusunda, zaman içinde özellikler bulmanın yanı sıra, sunulan modelin her zaman adımında yüksek doğrulukta özellikler üretmesi gerekmektedir ki bu da sınıflandırmayı artırsın. Tasarlanan sinir ağı ile yerel özellikleri dikkat mekanizması ile ve genel özellikleri ikinci bir katman ile bulmayı amaçlıyoruz. Bu sinir ağı, BreizhCrop veri seti üzerinde doğrulanmış ve alternatif yaklaşımlara göre daha iyi performans sergilediği sonucuna varılmıştır. Önerilen yöntem, Zaman-İlişkili CNN, Star RNN ve Vanilya LSTM ağları ile karşılaştırılmış ve bahsedilen sinir ağlarından daha iyi sonuçlar elde edilmiştir. Geliştirilen modelle çıkarılan bu yerel ve küresel özelliklerin avantajını kullanarak, %82 gibi yüksek bir doğruluk oranı elde edilmiştir.

1. Introduction

Remote sensing gathers information about the Earth's surface using sensors that do not come in direct contact with the Earth. The process is typically performed using satellites or aircraft equipped with sensors that detect electromagnetic radiation of various wavelengths, including visible light, infrared, and microwave radiation. Agriculturists and farmers are using remote sensing to monitor crop growth and determine changes in vegetation over time using crop classification (Thenkabail et al., 2014).

As a means to monitor crop growth and detect changes in vegetation over time, satellite images have become an important source of information for crop classification. In addition to pixel-based classification (Kussul et al., 2016), object-based classification (Dwivedi et al., 2022), and machine learning-based classification (Huang et al., 2002), various approaches have been proposed for crop classification using satellite images.

Multispectral satellite images are commonly used to derive vegetation indices, such as the Normalized Difference Vegetation Index (NDVI) (Rußwurm & Korner, 2017). Through these indices, it is possible to differentiate among many crops as well as non-crops based on their respective spectral signatures. These spectral features have been used to classify crops using machine learning algorithms, such as Random Forest (Tatsumi et al., 2015) and Support Vector Machines (Huang et al., 2002).

The use of time-series data for crop classification could also include Landsat or Sentinel-2 images, which can be used to analyze vegetation indices, surface temperatures, and other environmental variables. Farmers can make informed decisions regarding crop management and yield optimization by using time-series analysis to understand crop growth and development (Wang et al., 2022).

Breizh is the Breton name for the Brittany region of France, which is located in the northwestern part of the country. The Breizh Crops project was launched in 2014 as a means of supporting sustainable agriculture in the region (BreizhCrops, 2022). Among the variables included are the type of crop, the planting and harvesting dates, irrigation and fertilization schedules, and yield measurements (Rußwurm et al., 2019). A wide range of applications have been developed using this dataset, including crop classification, Land-cover classification (Yan et al., 2022), crop type mapping (Rußwurm et al., 2023) and multimodal deep learning methods for earth observation (Sykas et al., 2021).

In most of the research that has been done on image time series data, mechanisms and models have been used that pay attention to the characteristics of a region over time, and feature extraction in neural network models has been considered over time. This is despite the fact that in each specific time period, some features are exclusively related to that time and can improve the learning of the model, which has received less attention. In this research, an attempt is made to cover this weakness. In this study, a hybrid neural network based on Bidirectional LSTM (BI-LSTM) attention mechanisms and a temporal convolutional layer is proposed for the classification of remote sensing time series data. Using this method, the main objective is to discover the important features of time series data by examining both directions of input, which is the result of using the attention mechanism in conjunction with the bidirectional layer. Meanwhile, temporal convolution layer neural networks have been utilized to examine all the features and extract information from the entire time series data. For this research, the BraizhCrop dataset (BreizhCrops, 2022) was used to test the methodology. As the title of this data set implies, the main feature of this data set is its time series. The main problem with this data set is that it contains an imbalance in the categories of products that make up the data set. Among the most basic problems with the data set is balancing it, which is one of the most basic problems with using this data set. In the next step, due to the time series of the data, an attention-based method has been introduced, that makes it possible to extract the salient features of each data series in order to categorize each data series using these salient features. We will discuss two concepts in the next sections of the paper, data set balancing and attention-based techniques.

The primary contributions of this paper can be summarized as follows:

1. Combining attention mechanisms with bidirectional methods using Long Short-Term Memory (LSTM) models and creating an attention layer.
2. A hybrid neural network of the attention layer created with a temporal convolutional neural network is presented to increase the accuracy of the classification of remote sensing time series data.
3. Balancing and increasing the classification efficiency of the examined data set to increase the accuracy efficiency of the presented model.

2. Related Works

Time series analysis can assist in identifying changes in land cover over time, such as deforestation, urban expansion, or agricultural land conversion. To understand the drivers of land use change and its environmental impacts, this information is crucial (Yan et al., 2019). Analyzing time series data is useful in evaluating the effectiveness of land management practices, such as conservation programs or land restoration efforts (MohanRajan et al., 2020). Identifying areas in which management practices have a positive impact and areas that require improvement can be achieved by monitoring changes in land cover over time.

The use of time series analysis can serve as a means of supporting land use planning in that it provides information on trends in land use and land cover that can assist in making decisions regarding land use. In light of this information, it is possible to determine which areas should be encouraged or restricted for development based on environmental, social, or economic factors. Overall, time series land cover classification is important for understanding and managing land use change, assessing the effectiveness of land management practices, supporting land use planning decisions, and studying the impacts of climate change on ecosystems (Toh et al., 2018).

Satellite image time series classification is a core issue, which is closely related to many land applications, such as mapping and detecting changes in land cover (Yuan et al., 2019) and identifying vegetation species (Immitzer et al., 2016). According to Devadas et al. (2012), they proposed a method for analyzing multi-temporal remote sensing images to identify changes in crop types during summer and winter. To determine the accuracy of SVM models when used for classification, vegetation index time series are most important. Researchers are already using Deep Learning (DL) to solve a variety of problems in the field of remote sensing. To overcome the problem of unknown classes in the open world affecting hyperspectral imagery classification results, Liu et al. (2020) developed a multitask (classification and reconstruction) DL.

A novel time series classification framework based on multi-task learning with graph convolutional networks by Baroud et al. (2021) proposed a new framework for time series land use and land cover classification. As compared to traditional methods, this framework was tested on several remote sensing datasets and achieved improved accuracy. An approach for spatiotemporal classification of remote sensing imagery using recurrent neural networks is proposed in Jia et al. (2020) study. The results of this study demonstrate that the proposed method can effectively capture long-range dependencies and improve the accuracy of land cover classification. This model obtains an overall accuracy (OA) of 90.46% and a mean intersection-over-union (mIoU) of 0.8073 for Vaihingen. The mentioned model focus on the enhancing the image and not cover the local and global features over the time. A study by Lee et al. (2018) used time series of Landsat spectral indices to classify land cover. In this study, spectral indices of time series are used to classify land cover. They use pixel information for each time images and sampling pixels of images to detect land usage, this method works with classic machine learning methods as Support vector machine (SVM), Random Forest and Decision Tree (DT) methods. The main problem in this work is the lack of feature extraction from the regions and not attention to the Region of Interest (ROI) in Images. This work shows good result based on Kappa Coefficient criterion that improve the traditional approaches 77 percent. According to in their publication, a deep learning-based approach was proposed for crop type classification by using Sentinel-2 time series data. An Ethiopian dataset was used to test the method, and the accuracy of the method was higher than that of traditional methods. ensembles of multiple images from different sensors (Planet-

NICFI/Sentinel-2/Landsat-8) with different spatial resolutions. It is demonstrated in these studies that time series approaches are useful for the classification of land uses and land covers and that deep learning and other advanced techniques can be effective in improving the accuracy of the classifications. The mentioned method shows 90 % accuracy with F1-Score.

The improved random forest algorithm is used to integrate Landsat and Moderate Resolution Imaging Spectroradiometer (MODIS) data for mapping land use and land cover. The improved random forest algorithm and Landsat and MODIS data are used in this study to propose a spatiotemporal fusion approach to land use and land cover mapping in 2019 (Li et al., 2021). Tests were conducted on a dataset from China and the method was found to be highly accurate.

An optimal deep learning model for pixel-based Land Cover and Land Use (LC&CC) is presented in (Mazzia et al., 2019) based on multi-temporal Sentinel-2 imagery of Italy's central north. To compare the results obtained, they tested support vector machines (SVMs), random forests (RFs), kernel SVMs, and gradient boosting machines (XGBoosts). The proposed method performs better than the methods previously mentioned. The overall accuracy achieved by Pixel R-CNN method that proposed by researchers of this study was 96.5%, which showed considerable improvements in comparison with existing mainstream methods.

In most of the presented methods, it is clear that the methods based on deep learning have better performance. The reason for this performance is the extraction of features relative to the regions of interest and they can obtain better features from the images of the region under study (Jia et al., 2020). Among the neural networks, the methods that are more capable of extracting features over time and by extracting features over time focusing on specific areas of an image will be more successful. In this regard, models like LSTM will have better classification ability (Rußwurm et al., 2023). The combination of convolution-based methods for time series data is another group of researches that have shown better performance in this field (Navnath et al., 2022).

Navnath et al. (2022) implement several models, including bidirectional gated recurrent units (GRUs), temporal convolutional neural networks (TCNNs), GRUs combined with TCNNs, TCNNs combined with GRUs, and GRUs combined with TCNNs. The purpose of this paper is to propose an integrated architecture for classifying land cover on Reunion Island based on univariate, multivariate, and pixel coordinates. When combining univariate and multivariate with recurrent neural networks, the performance of the class labels improved with higher F1 scores for each class label. The macro-average F1-score for the mentioned method is 83%, and the weighted average F1 score is 91%

There are two basic challenges associated with remote sensing time series data analysis, as demonstrated in recent studies. The first challenge consists of the extraction of important features using attention, which has improved classification in recent studies, and the second challenge is a lack of balance between classes in the collected data. Due to the nature of the data of agricultural products and the low cultivation of some products, there is an imbalanced data class for training the model in the learning model, which makes it perform poorly. In the following sections, we will explore the methods of resolving imbalances and the mechanism of attention.

2.1. Imbalanced dataset

Handling imbalanced datasets is a significant challenge in machine learning algorithms. Various methods can be employed to address this challenge, including under-sampling, which reduces instances in the majority class, and oversampling, which increases instances in the minority class. Synthetic sampling techniques can create new instances in the minority class, but careful selection of methods and parameters is necessary for large datasets. Cost-sensitive learning algorithms modify the learning process to prioritize the minority class. Ensemble methods combine multiple classifiers to improve performance, and anomaly detection identifies instances significantly different from the majority class. These approaches offer strategies for effectively handling imbalanced datasets.

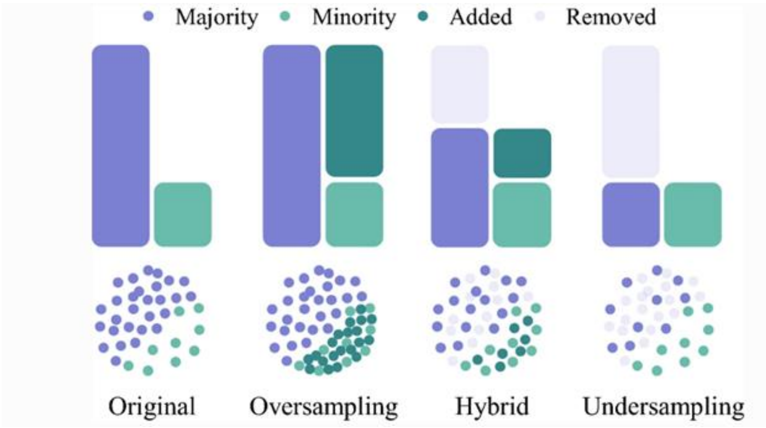


Figure 1. Different methods of handling imbalanced datasets (Werner de Vargas et al., 2023).

It is also important to carefully evaluate the performance of the learning algorithm on a validation set in addition to these approaches. An analysis of performance indicators such as precision, recall, F1-score, and area under the receiver operating characteristic curves (ROCs) can provide insight into the algorithm's performance when dealing with imbalanced datasets. The business context and the costs associated with false positives and false negatives should also be considered when selecting an approach to handle imbalanced datasets.

2.2. Attention Mechanism

Deep learning models based on attention have gained a lot of popularity in recent years. In tasks such as machine translation, as well as tasks where input and output sequences vary in length and certain parts of the input sequence are more important in generating particular parts of the output sequence, these methods are particularly useful (Mei et al., 2019). Typical neural networks treat all inputs equally, but attention-based neural networks selectively focus on certain inputs while ignoring others (Zhu et al., 2020). The input and output sequences are aligned by a soft or "learned" alignment. An attention-based neural network generates output elements based on a separate "attention mechanism" that learns to weigh the importance of the different parts of the input sequence. Attention mechanisms work by applying weights to input sequences to compute the weighted sum of the inputs, which is then used to determine the output (Cheng et al., 2020). Attention mechanisms can be categorized in several ways, but "soft" or "content-based" attention is one of the most common. According to this method, attention weights are calculated based on the similarity measure between the current decoder state and each input element. An input element with a higher similarity is given a higher weight. It has been found that attention-based neural networks are effective for a wide range of time series tasks (Li et al., 2019).

Time series tasks, such as forecasting, anomaly detection, and classification, have been shown to benefit from attention mechanisms. Attention mechanisms can be used for time series data by treating a sequence of time steps as the input sequence and focusing on the most relevant time steps at each step of the decoding process. A similar approach is used in natural language processing where the model attends to different parts of the input sequence throughout the decoding process. In time series forecasting, for example, the model can be trained to predict a series' value at a future time step. The attention mechanism can be used to identify the most relevant past time steps to consider during the prediction process at each step of the decoding process. The self-attention mechanism can also be used to capture the temporal dependencies within time series data. Long-term dependencies can be captured using self-attention mechanisms, which allow the model to attend to different parts of the time series at different levels of abstraction.

3. Methodology

As shown in Figure 2, a series of data preparation processes were carried out in order to prepare the input data into a form that would fit the neural network. In order to ensure a higher degree of accuracy for the neural network, a series of data preparation processes have first been carried out. A number of steps are taken in order to normalize the existing data. To improve the performance and convergence of neural networks, data normalization is commonly used as a preprocessing technique. An input data set is normalized by scaling it to have a mean of zero and a standard deviation of one, or by scaling it to a specific range, such as [0, 1], [-1, 1]. The effects of data normalization in neural networks include:

- Enhanced convergence: Normalization can make neural networks more stable and faster to converge. Input data that is not normalized can lead to unbalanced network weights, causing slow convergence or divergence. By ensuring that inputs are on the same scale, normalization prevents this.
- Overfitting can also be reduced through normalization. Overfitting happens when a neural network learns to fit data too closely, leading to poor performance on new, unseen data. As a result of normalization, outliers are less impactful and the model is less sensitive to input variations.
- Enhanced generalization: Normalization can also benefit neural networks. Normalization can help the network learn more robust and generalizable features that apply to a wider range of inputs by reducing the impact of outliers and ensuring the inputs are on a similar scale.

A balance between the classes in the dataset will be applied in the next step of the process. Balancing is important for preventing the bias of the network towards one particular side while also ensuring that the network has the same efficiency to be able to detect all the classes equally.

It is now time to prepare the data for classification, which is the next stage of data preparation, and, because the data of this study are time series, these data must be categorized into time series categories to enter into our network. For this reason, we need to reformat the data to make it suitable for this goal. Following the preparation of the data, it is separated into two different categories - training and testing - where the training data is used as input to the neural network, and the test data is used to determine if the neural network model that has been extracted is effective. Among the neural networks that are used, there is a section devoted to attention, and there is also a section devoted to convolutional neural networks. There are two types of convolutional neural networks: temporal convolutional neural networks, which extract features from the entire time series data set, and attentional convolutional neural networks, which extract the most significant portions of a sequential data set. After concatenating these two sets of features, the fully connected layer of the model will then use the fully connected network to map these two sets of features to labels in the final model. Finally, once the trained neural network has been completely trained, we will use a variety of criteria for evaluation. A few of these criteria include precision, recall, F1 score, and kappa score, which we will use for the evaluation of the neural network.

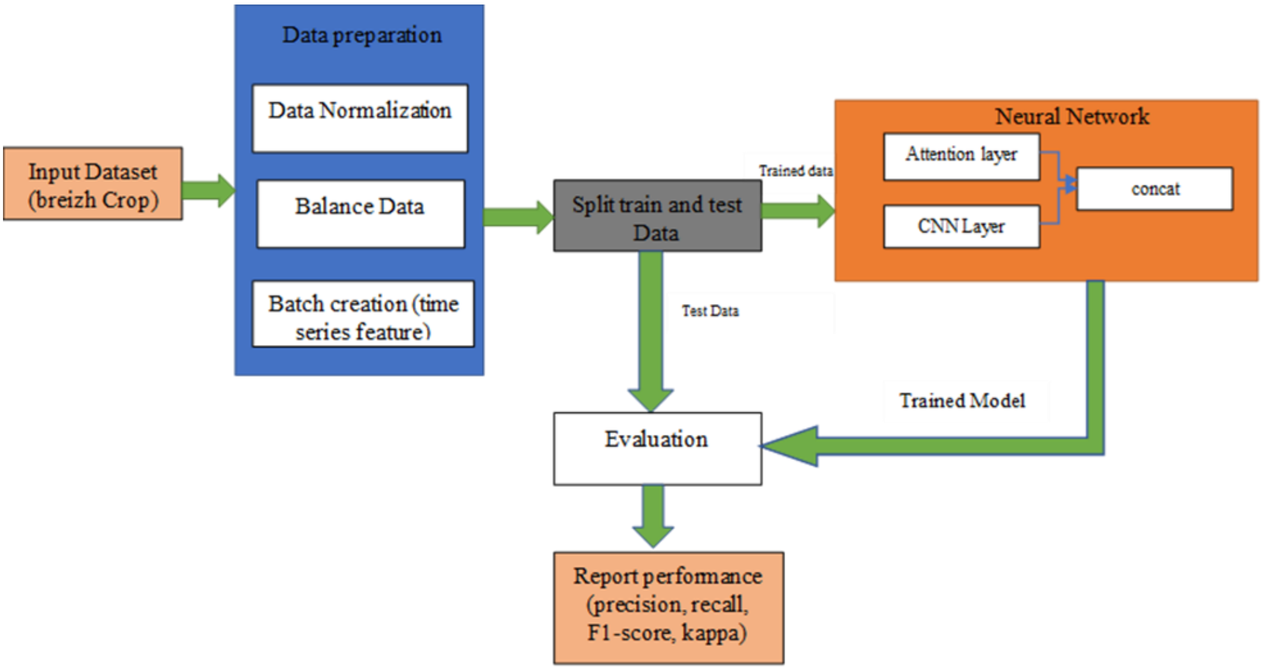


Figure 2. The steps of model training and evaluation methodology.

3.1. Solution for imbalanced class labels problem

In most agricultural areas, agriculture is dominated by a few crops such as corn, meadows, and wheat, which are extensively cultivated. However, other types of vegetation should also be classified at a reasonable level of accuracy by the local authorities (Rußwurm & Korner, 2017). Due to the limited amount of agricultural land available for some crops, the dataset is unbalanced. Moreover, some products are cultivated on large areas of agricultural land because they are highly demanded. An Under-sampling approach is used in this study to reduce categories with a large number of classes. Under-sampling methods involve reducing the number of instances in the majority class (i.e., the class with more instances) so that it is closer in size to the minority class (i.e., the class with fewer instances) (Junsomboon & Phienthrakul, 2017). Under-sampling method mechanism is shown in Fig. 3. To balance the dataset, classes with a small number of deleted data are considered. Because the data is a time series, the production of artificial data or oversampling is not suitable for this type of data. The techniques available for this type of data have not worked well and cause the efficiency of the neural network to decrease. For this reason, in this research, to avoid reducing the accuracy of the neural network, the classes with a small number are not considered and will be removed from the dataset.

Figure 3. Under-sampling from major class in dataset.

3.2. Neural network architecture

A parallel neural network architecture is proposed with two different architectures. The first architecture involves a BI-LSTM layer that implements an attention mechanism. As a result of this structure, we can extract the effective features from time series data and use a dropout layer followed by another LSTM layer to sort and flatten the sequence of these features. The extracted features are entered sequentially and one-dimensionally into the concatenate layer by this layer. Also, the Temporal-CNN layer is applied in parallel on the same batch of input data. As a result of designing this network, we are able to extract the general characteristics of the time series data, thereby providing a more general insight into the order of the time series data in parallel with the characteristics of the attention layer. A flattened and one-dimensional output from this layer is then passed on to the concatenate layer. Each time series data set is mapped to the Softmax activation function based on features extracted from two independent neural networks. Figure 6 shows the general structure of this architecture.

An advantage of the BI-LSTM architecture over a simple LSTM architecture is that the effective features of the time series are obtained from both sides of the time series input, rather than from only one side. Thus, the extracted features are revised in a two-way manner. By using the Temporal-CNN neural network, the more general characteristics of the time series are transferred to the next layer with a more general perspective. It is therefore possible to extract both local and general features from time series data simultaneously. As a result of the features obtained, mapping to the labels becomes possible from both a general and a local perspective. As a general summary, the superiority of the proposed method can be summarized as follows: (a) navigation of effective local features from beginning to end and vice versa (b) extraction of general features (c) the simultaneous use of two categories of features to identify the label.

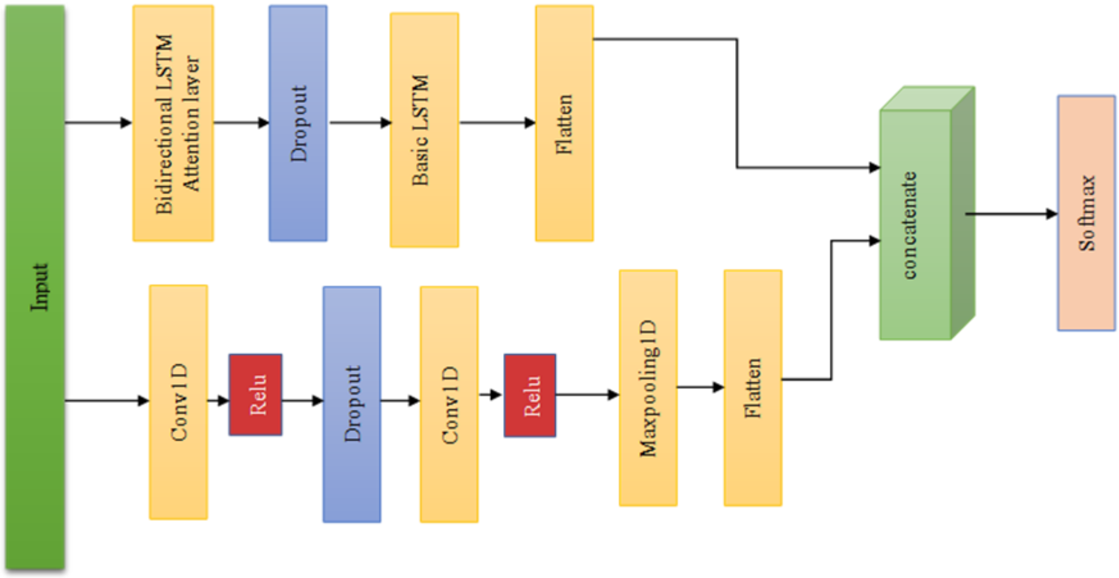


Figure 4. Proposed Bi-LSTM attentional temporal CNN architecture.

4. Experiment

In this section, three temporal models, CNN, vanilla LSM, and Star Recurrent Neural Network (RNN), are applied to validate the effectiveness of the neural network outlined in the previous section. These methods have been used in a variety of studies to classify crops based on remote sensing data. For the evaluation of the proposed methodology, the Breizhcrop dataset was used. Consequently, different classification metrics will be used to compare the methods. Throughout all runs, classification results have been reported on Cohen's kappa score(κ), f1 score, precision and recall. A brief description of each model is provided below.

*1) Temporal CNN (TCN) (*Lea et al., 2016*):* The Temporal Convolutional Neural Network (TCN) is a type of neural network that models sequential data, such as time series and natural language processing. To process input sequences, TCNs employ a convolutional neural network (CNN) architecture. As opposed to other recurrent neural networks (RNNs), such as LSTMs or GRUs, TCNs are better at capturing long-term dependencies in sequential data. Compared to RNNs, TCNs are faster and more memory-efficient because they can process input sequences in parallel.

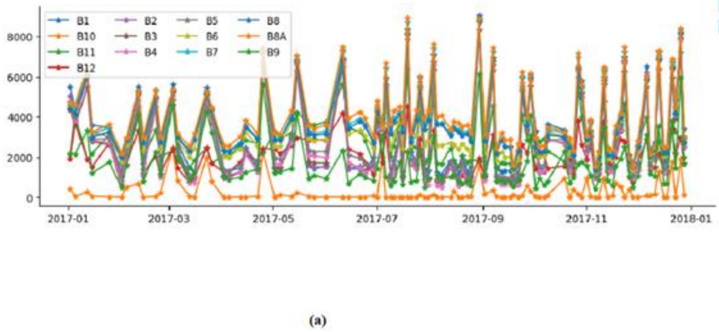
*2) Star RNN (*Tran et al., 2023*):* Several models for sequence data are based on recurrent neural networks (RNNs). Similar to feedforward networks, recurrent neural networks have also become common in which multiple recurrent layers are stacked to obtain higher-level abstractions of the data. It is important to note, however, that this is only effective for a few layers. There is a risk of model performance degrading when more than a few recurrent units (e.g., LSTM cells) are combined, as vanishing or exploding gradients usually occur during training. Multilayer RNNs are examined and the magnitude of gradients as they propagate is examined. In addition to improving performance, STAR cells enable significantly deeper recurrent architectures. RNNs that use Start cells are referred to as Star RNNs.

*3) Vanilla LSTM (*Ngoc Hai et al., 2020*):* A vanilla LSTM is the simplest version of the LSTM architecture, consisting of three gates: input, forget, and output. The gates control the information flowing through the network's internal memory, the cell state. Input gates determine how much new information is added to the cell state, forget gates determine how much information is discarded, and output gates determine how much information is output. Vanilla LSTM has the advantage of handling input sequences of varying lengths as well as capturing long-term dependencies in sequential data. Applications of Vanilla LSTM include speech recognition, machine translation, and video analysis. The vanilla LSTM is powerful to model sequential data, but it can be difficult to train and needs a lot of attention when tuning the hyperparameters. In addition to LSTMs, gated recurrent units (GRU) and peephole LSTMs have additional gates and modifications to the formula for updating the cell state.

LSTMs are recurrent neural networks designed to capture long-term dependencies in sequential data and address the vanishing gradient problem.
All tests designed on a system with a core I7 8550 CPU with a 1.8 GHz processing rate and 12 GB of RAM and Debian Linux operating system have been performed. The following subsections will examine the data set and then analyze the classification results using a variety of performance metrics. All implementations and tests are available in the GitHub repository*.

4.1. Dataset

The proposed method has been evaluated using the Breizhcrop dataset. There are four regions in this set (Frh01, Frh02, Frh03, and Frh04). A series of farms have been tracked through remote sensing data in this dataset. We have selected 50 data series with 13 features each, arranged in sequence order. There are 12 thousand different data items in 13 classes in this data set. Figure 5a shows an example of data related to the data set, which is based on a field where corn was planted in 2017 and can be viewed on a map (Figure 5b)



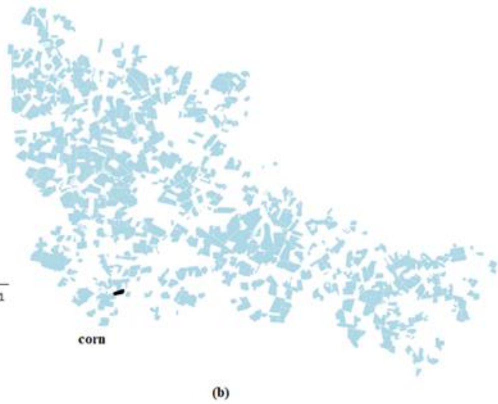


Figure 5. (a) Data of emitted bands in remote sensing images over time (b) The field sample from the dataset where corn is planted.

We have removed six classes with a small number of data and no specific product categories so that the data could be evaluated and balanced. Under-sampling is used for classes with major sizes to sample the remaining 7 classes. Table 1 shows balanced data for evaluation. A total of 37,348 numbers of data will be used for training and evaluating the presented model.

Table 1. Dataset information used for evaluation

Label	Crop Type	Train Set	Test Set	Total
1	Barley	4964	1014	5978
2	Wheat	4560	1005	5565
3	Rapeseed	2656	580	3236
4	Corn	5126	1207	6333
5	Misc.	3864	1069	4933
6	Permanent meadow	4014	1103	5117
7	Temporary meadow	4983	1203	6186
Total		**30167**	**7181**	**37348**

* https://github.com/amerNN/Timeseries

4.2. Evaluation metrics

To evaluate the performance of classification models, precision, recall, and F1 scores are commonly used. In addition to revealing the model's capability to classify instances correctly, these metrics also provide insights into the accuracy of the classification. In addition, in cases where the distribution of classes in the dataset appears to be imbalanced, the Kappa score is used as another metric to quantify the level of agreement between predicted and true labels (Lever, 2016).

- Precision is a measure of the proportion of positive instances predicted correctly out of all that are predicted as positive. As a result, it measures the accuracy of the model in identifying true positives while minimizing false positives. Calculation of precision consists of:

$$Precision = \frac{TP}{(TP + FP)} \tag{1}$$

- The recall, also known as sensitivity or true positive rate, is the percentage of predicted positive instances that have been verified as true. A model is evaluated based on its ability to identify all positive instances without missing any. Recall is calculated as:

$$Recall = \frac{TP}{(TP + FN)} \tag{2}$$

- F1 Score is a combined metric that takes both recall and precision into account. In order to calculate the harmonic, mean of precision and recall, we find out the balance of the model's performance. The F1 score is calculated as:

$$F_1\ Score = \frac{2 * Recall * Precision}{Precision + Recall} \tag{3}$$

F1 scores range from 0 to 1, with 1 representing the best possible performance.

- In statistics, the Kappa score, also called Cohen's kappa coefficient, is used to measure whether predicted and true labels agree, taking into account the possibility of chance agreement. Typically, it is used when datasets are unbalanced, and accuracy alone may be misleading. Values closer to 1 indicate a greater degree of agreement than chance, values closer to 0 indicate agreement by chance, and negative values indicate worse agreement than chance. It is computed based on the observed agreement (the number of events predicted correctly) and the expected agreement (the number of events predicted by chance). The formula for the Kappa score is:

$$Kappa = \frac{Observed\ Agreement - Expected\ Agreement}{1 - Expected\ Agreement} \tag{4}$$

In this calculation, we assume that the predicted labels and the actual labels are randomly distributed.

Different evaluation metrics provide insight into the model's performance, and the choice depends on the specific requirements of the problem. Precision might be more important in scenarios of false positives, while recall might be more important in false negatives. It is useful when assessing agreement beyond chance to use the F1 score, and especially when assessing imbalanced datasets to use the Kappa score.

4.3. Classification performance

In this section, using 20% of the data in the dataset, the validation of our method has been done, and on the other hand, the proposed neural network architecture has been compared with other existing methods. The results of this experiment are shown in Table 2. According to this test, the proposed neural network has performed better in four classification evaluation metrics. As it can be seen, the LSTM network alone does not work properly and when it is used for attention, it has greatly increased the accuracy of the network. On the other hand, the Temporal-CNN network has performed better on its own, but with the addition of the proposed attention layer, the efficiency of this network has also increased. The parameters and base layers of each model shown in Table 2 and Table 3. Performance metrics of classification methods.

Table 2. Parameters for different methods

Model	Parameter and layers	Description
Temporal CNN	Kernel Size	The kernel size refers to the width × height of the filter mask.
	Conv1D Layer	This layer creates a convolution kernel that is convolved with the layer input over a single spatial (or temporal) dimension to produce a tensor of outputs.
Star RNN	Learning Rate	learning rate is a tuning parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum of a loss function.
	Star Layer	stackable recurrent (STAR) cell allows for substantially deeper recurrent architectures.
Vanilla LSTM	Number of sequence	Number of memory Cell for sequence of data that input to model.
	LSTM layer	An LSTM layer is an RNN layer that learns long-term dependencies between time steps in time-series and sequence data.
Bi-LSTM Attentional temporal CNN	Concatator	returns a single tensor that is the concatenation of all inputs features.
	BI-LSTM layer	BI-LSTM layer is an RNN layer that learns bidirectional long-term dependencies between time steps of time-series or sequence data.

Table 3. Classification evaluation reports for different methods

Metric	Temporal CNN	Star RNN	Vanilla LSTM	Bi-LSTM Attentional temporal CNN
Precision	0.8	0.71	0.73	0.82
Recall	0.8	0.72	0.71	0.82
F1-Score	0.8	0.71	0.79	0.82
Kappa	0.73	0.69	0.73	0.76

According to Figure 6, the confusion matrix is specified for all four neural networks tested. This matrix shows the correct detection percentages for each class in its main diameter. It has been shown that the proposed method has the greatest accuracy for all classes (Figure 6d), followed by the LSM and Temporal-CNN methods (Figure 6a, Figure 6b) that perform better for the classes, respectively, and the Star-RNN method (Figure 6c) later has good accuracy for all classes, but its performance is less than the proposed method. The main reason for the accuracy obtained for the proposed architecture is the simultaneous use of local and global features. In the TemporalCNN model, although it focuses more on global features and trains the model based on the features that exist in a specific area over time. On the other hand, the vanilla LSTM model also does not have the attention mechanism, and although it is trained over time, it does not consider the features related to each image independently. This is despite the fact that the Star RNN method, despite focusing on local and global features over time, does not consider those features together due to the architecture that constitutes them, and uses each of them independently in training. Meanwhile, in addition to considering two types of global and local features, the proposed method can correctly use the integration of these features, which is in the form of an attention layer and B-LSTM together and in the integration. It has increased the efficiency of the proposed model.

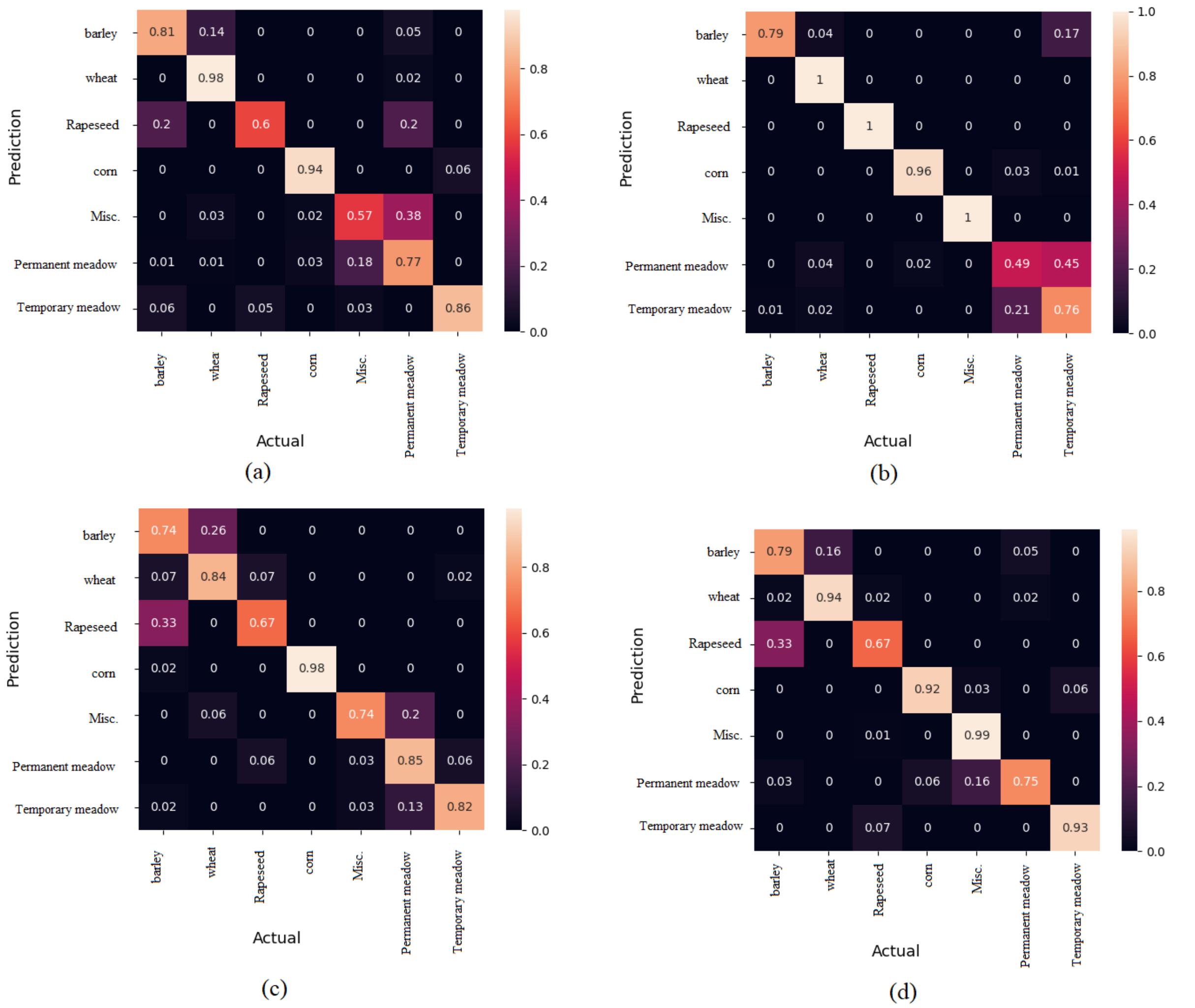


Figure 6. Confusion matrix for (a) Temporal CNN (b) Vanilla LSTM (c) Star RNN (d) BI-LSTM attentional temporal CNN.

In the Figure 7, the predicted products for each field are specified according to its label. 7 different labels have been used to identify each farm product. In Table 4, we have compared the related works that have recently worked and reported on the time series data of images with the proposed method in terms of accuracy and efficiency. The proposed method has performed almost 0.17 better than other existing methods. The Triplet-LSTM method, which uses three consecutive LSTM layers to extract features, or the transformers method, which works with the Encode and Decode mechanism, is another approach that has shown better performance. On the other hand, the Vanilla LSTM method, which was introduced earlier, has been used for this dataset, and the results show that the proposed method performed better than other methods. All the mentioned methods focus on feature extraction and the presented method has performed better in terms of feature extraction approach at any time and also over time.

Figure 7. The classification result after prediction for sample region (a) BI-LSTM attentional temporal CNN (b) Star RNN (c) Vanilla LSTM (d) Temporal CNN.

Table 4. Classification evaluation reports for different methods

Author	Type of Corps	Method	Precision
Vaswani et al. (2017)	Wheat, Barley, Corn, Fodder, Fallow, Misc, Orchards, Cereals, Perm. Meadows, Protein crops, Rapeseed, Temp. Meadows, Vegetables	Transforms	0.69
Bozo et al. (2020)	Barley, Wheat, Rapeseed, Corn, Misc., Permanent meadow, Temporary meadow	TripletLSTM	0.642
Rußwurm et al. (2023)	Wheat, Barley, Corn, Fodder, Fallow, Misc, Orchards, Cereals, Perm. Meadows, Protein crops, Rapeseed, Temp. Meadows, Vegetables.	Vanilla-LSTM	0.63
Proposed Method	Barley, Wheat, Rapeseed, Corn, Misc., Permanent meadow, Temporary meadow	Bi-LSTM Attentional temporal CNN	0.82

4.4. Computational complexity

The training durations of the various tested approaches are presented in Table 5, while the testing durations are displayed in real time (measured in second) for each approach. The training time for all models is almost the same. Only the Star-RNN network had more training time, and the number of neural network parameters compared shows that the proposal has better training speed and accuracy than other methods, even though it has fewer parameters. There were only two parameters of the Star-RNN neural network that performed worse than the other methods: the time and the parameters.

Table 5. Training real time duration and neural network parameters

Method	Training Duration (second)	Number of Network Parameters
Temporal CNN	46	321,601
Star RNN	130	567,103
Vanilla LSTM	39	125,833
Bi-LSTM Attentional temporal CNN	49	255,965

5. Conclusions

In the field of machine learning, it is challenging to classify crop based on time series data derived from remote sensing data. There has been a great deal of effort put into improving neural networks in this regard. The purpose of this research was to train the network on the effective features over time by using an attention mechanism on the LSTM layer. Meanwhile, the features of time series data were extracted using a temporal CNN neural network, which was combined with the previous approach to increase the efficiency of the network. Attention layer construction is made more accurate by using a bidirectional layer, which scrolls the data from the first to the last and from the last to the first, and this increases the accuracy of the neural network. Based on the results of the experiment conducted on the Breizhcrop dataset, the neural network was found to be more efficient based on the metrics of precision, recall, F1 score, and Kappa. Alternatively, the training speed and number of parameters of the network indicate that these two parameters are optimal when compared with the accuracy obtained from validation. One of the biggest limitations of this model is the requirement that each time series have a certain amount of data. A new model can be developed that eliminates this limitation. A second limitation of the proposed method is related to the use of other features, such as indices such as NDVI, which can improve the model's accuracy by adding new indices based on different spectral data. A limitation of this method can be attributed to the fact that it requires data with the same time series. However, there may not be enough data available for some areas. Additionally, another important limitation of this model is that it requires several long-term time series to achieve high accuracy, but there are models that can achieve high accuracy with fewer series. The use of different attention mechanisms could have a greater impact than the type of data in future research. Time series data can also be processed using different deep neural network models. Additionally, it is important to propose the effective dataset balancing methods for this type of data.

References

Baroud, S., Chokri, S., Belhaous, S., & Mestari, M. (2021). A brief review of graph convolutional neural network based learning for classifying remote sensing images. *Procedia Computer Science*, 191, 349-354. doi:10.1016/j.procs.2021.07.047

Bozo, M., Aptoula, E., & Cataltepe, Z. (2020). A discriminative long short term memory network with metric learning applied to multispectral time series classification. *Journal of Imaging*, 6(7), 68. doi:10.3390/jimaging6070068

BreizhCrops. (2022). BreizhCrops - Smart Agriculture. https://www.breizhcrops.fr/en/ Access date: 01.01.2024.

Cheng, D., Xiang, S., Shang, C., Zhang, Y., Yang, F., & Zhang, L. (2020, April). *Spatio-temporal attention-based neural network for credit card fraud detection.* Proceedings of the AAAI Conference on Artificial Intelligence, 34(01), 362-369. doi:10.1609/aaai.v34i01.5371

Devadas, R., Denham, R. J., & Pringle, M. (2012). Support vector machine classification of object-based data for crop mapping, using multi-temporal Landsat imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *39*, 185-190. doi:10.5194/isprsarchives-XXXIX-B7-185-2012

Dwivedi, A. K., Singh, A. K., & Singh, D. (2022, July). *An object based image analysis of multispectral satellite and drone images for precision agriculture monitoring*. IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia. doi:10.1109/IGARSS46834.2022.9884674

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), 725-749. doi:10.1080/01431160110040323

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing*, 8(3), 166. doi:10.3390/rs8030166

Jia, D., Cheng, C., Shen, S., & Ning, L. (2022). Multitask deep learning framework for spatiotemporal fusion of NDVI. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-13. doi:10.1109/TGRS.2021.3140144

Junsomboon, N., & Phienthrakul, T. (2017, February). *Combining over-sampling and under-sampling techniques for imbalance dataset.* Proceedings of the 9th International Conference on Machine Learning and Computing, Singapore. doi:10.1145/3055635.3056643

Kussul, N., Lemoine, G., Gallego, F. J., Skakun, S. V., Lavreniuk, M., & Shelestov, A. Y. (2016). Parcel-based crop classification in Ukraine using Landsat-8 data and Sentinel-1A data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6), 2500-2508. doi:10.1109/JSTARS.2016.2560141

Lea, C., Vidal, R., Reiter, A., & Hager, G. D. (2016, October). *Temporal convolutional networks: A unified approach to action segmentation.* Computer Vision–ECCV 2016 Workshops, Amsterdam, Netherlands. Publishing. doi:10.1007/978-3-319-49409-8_7

Lee, J. K., Acharya, T. D., & Lee, D. H. (2018). Exploring land cover classification accuracy of landsat 8 image using spectral index layer stacking in hilly region of South Korea. *Sensors & Materials*, 30(12), 2927-2941. doi:10.18494/SAM.2018.1934

Lever, J. (2016). Classification evaluation: It is important to understand both what a classification metric expresses and what it hides. *Nature Methods*, 13(8), 603-605.

Li, Y., Zhu, Z., Kong, D., Han, H., & Zhao, Y. (2019). EA-LSTM: Evolutionary attention-based LSTM for time series prediction. *Knowledge-Based Systems*, 181, 104785. doi:10.1016/j.knosys.2019.05.028

Li, R., Xu, M., Chen, Z., Gao, B., Cai, J., Shen, F., ..., & Chen, D. (2021). Phenology-based classification of crop species and rotation types using fused MODIS and Landsat data: The comparison of a random-forest-based model and a decision-rule-based model. *Soil and Tillage Research*, 206, 104838. doi:10.1016/j.still.2020.104838

Liu, C., Zeng, D., Wu, H., Wang, Y., Jia, S., & Xin, L. (2020). Urban land cover classification of high-resolution aerial imagery using a relation-enhanced multiscale convolutional network. *Remote Sensing*, 12(2), 311. doi:10.3390/rs12020311

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). *Applied Sciences*, 10(1), 238. doi :10.3390/app10010238

Mei, X., Pan, E., Ma, Y., Dai, X., Huang, J., Fan, F., ..., & Ma, J. (2019). Spectral-spatial attention networks for hyperspectral image classification. *Remote Sensing*, 11(8), 963. doi:10.3390/rs11080963

MohanRajan, S. N., Loganathan, A., & Manoharan, P. (2020). Survey on Land Use/Land Cover (LU/LC) change analysis in remote sensing and GIS environment: Techniques and Challenges. *Environmental Science and Pollution Research*, 27(24), 29900-29926. doi:10.1007/s11356-020-09091-7

Navnath, N. N., Chandrasekaran, K., Stateczny, A., Sundaram, V. M., & Panneer, P. (2022). Spatiotemporal assessment of satellite image time series for land cover classification using deep learning techniques: a case study of Reunion Island, France. *Remote Sensing*, 14(20), 5232. doi:10.3390/rs14205232

Ngoc Hai, P., Manh Tien, N., Trung Hieu, H., Quoc Chung, P., Thanh Son, N., Ngoc Ha, P., & Tung Son, N. (2020, October). *An empirical research on the effectiveness of different LSTM architectures on Vietnamese stock market.* Proceedings of the 2020 1st International Conference on Control, Robotics and Intelligent System, Xiamen, China. doi:10.1145/3437802.3437827

Rußwurm, M., & Korner, M. (2017, July). *Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA. doi:10.1109/CVPRW.2017.193

Rußwurm, M., Pelletier, C., Zollner, M., Lefèvre, S., & Körner, M. (2019). Breizhcrops: A time series dataset for crop type mapping. *arXiv preprint arXiv:1905.11893*. doi:10.48550/arXiv.1905.11893

Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., & Tavenard, R. (2023). End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 445-456. doi:10.1016/j.isprsjprs.2022.12.016

Sykas, D., Papoutsis, I., & Zografakis, D. (2021, July). *Sen4AgriNet: A harmonized multi-country, multi-temporal benchmark dataset for agricultural earth observation machine learning applications.* IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium. doi:10.1109/IGARSS47720.2021.9553603

Tatsumi, K., Yamashiki, Y., Torres, M. A. C., & Taipe, C. L. R. (2015). Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture*, 115, 171-179. doi:10.1016/j.compag.2015.05.001

Thenkabail, P. S., Gumma, M. K., Teluguntla, P., & Irshad, A. M. (2014). Hyperspectral remote sensing of vegetation and agricultural crops. *Photogrammetric Engineering & Remote Sensing (TSI)*, 80(8), 695-723.

Toh, F. A., Angwafo, T. E., Ndam, L. M., & Antoine, M. Z. (2018). The socio-economic impact of land use and land cover change on the inhabitants of Mount Bambouto Caldera of the Western Highlands of Cameroon. *Advances in Remote Sensing*, 7(1), 25-45. doi:10.4236/ars.2018.71003

Tran, H. D., Choi, S. W., Yang, X., Yamaguchi, T., Hoxha, B., & Prokhorov, D. (2023, May). *Verification of recurrent neural networks with star reachability*. 26th ACM International Conference on Hybrid Systems: Computation and Control, San Antonio, TX, USA. doi:10.1145/3575870.3587128

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L.,Gomez, ..., & Polosukhin, I. (2017) Attention is all you need. *arXiv preprint arXiv:1706.03762*. doi:10.48550/arXiv.1706.03762

Wang, X., Zhang, J., Xun, L., Wang, J., Wu, Z., Henchiri, M., ..., & Yu, X. (2022). Evaluating the effectiveness of machine learning and deep learning models combined time-series satellite data for multiple crop types classification over a large-scale region. *Remote Sensing*, 14(10), 2341. doi:10.3390/rs14102341

Werner de Vargas, V., Schneider Aranda, J. A., dos Santos Costa, R., da Silva Pereira, P. R., & Victória Barbosa, J. L. (2023). Imbalanced data preprocessing techniques for machine learning: a systematic mapping study. *Knowledge and Information Systems*, 65(1), 31-57. doi:10.1007/s10115-022-01772-8

Yan, J., Wang, L., Song, W., Chen, Y., Chen, X., & Deng, Z. (2019). A time-series classification approach based on change detection for rapid land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158, 249-262. doi:10.1016/j.isprsjprs.2019.10.003

Yan, J., Liu, J., Wang, L., Liang, D., Cao, Q., Zhang, W., & Peng, J. (2022). Land-cover classification with time-series remote sensing images by complete extraction of multiscale timing dependence. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 1953-1967. doi:10.1109/JSTARS.2022.3150430

Yuan, Y., Lin, L., Chen, J., Sahli, H., Chen, Y., Wang, C., & Wu, B. (2019). A new framework for modelling and monitoring the conversion of cultivated land to built-up land based on a hierarchical hidden semi-Markov model using satellite image time series. *Remote Sensing*, 11(2), 210. doi:10.3390/rs11020210

Zhu, T., Luo, W., & Yu, F. (2020). Convolution-and attention-based neural network for automated sleep stage classification. *International Journal of Environmental Research and Public Health*, 17(11), 4152. doi:10.3390/ijerph17114152