DC-SAM: DILATED CONVOLUTION AND SPECTRAL ATTENTION MODULE FOR WHEAT SALT STRESS CLASSIFICATION AND INTERPRETATION

Wijayanti Nurul Khotimah
Department of Informatics
Institut Teknologi Sepuluh Nopember (ITS)
Kampus ITS, Sukolilo, Surabaya, 60111, Indonesia
e-mail: wijayanti@if.its.ac.id

ABSTRACT

*Salt stress can impact wheat production significantly and is difficult to be managed when the condition is critical. Hence, detecting such stress whet it is at an early stage is important. This paper proposed a deep learning method called Dilated Convolution and Spectral Attention Module (DC-SAM), which exploits the difference in spectral responses of healthy and stressed wheat. The proposed DC-SAM method consists of two key modules: **(i)** a dilated convolution module to capture spectral features with large receptive field; **(ii)** a spectral attention module to adaptively fuse the spectral features based on their interrelationship. As the dilated convolution module has long receptive fields, it can capture short- and long dependency patterns that exist in hyperspectral data. Our experimental results with four datasets show that DC-SAM outperforms existing state-of-the-art methods. Also, the output of the proposed attention module reveals the most discriminative spectral bands for a given wheat stress classification task.*

Keywords*: dilated convolution, explainable AI (XAI), hyperspectral information, spectral attention, wheat salt stress.*

DC-SAM: MODUL KONVOLUSI DENGAN DILASI DAN SPECTRAL ATTENTION UNTUK KLASIFIKASI DAN INTERPRETASI GANGGUAN PADA GANDUM YANG DISEBABKAN KARENA KELEBIHAN GARAM

Wijayanti Nurul Khotimah
Departmen Informatika
Institut Teknologi Sepuluh Nopember (ITS)
Kampus ITS, Sukolilo, Surabaya, 60111, Indonesia
e-mail: wijayanti@if.its.ac.id

ABSTRAK

*Gangguan pada tanaman gandum karena kebanyakan kandungan garam dapat secara signifikan mengurangi hasil panen. Jika gangguan tersebut sudah parah, akan sangat sulit untuk diselesaikan. Oleh karena itu, pendeteksian keberadaan gangguan tersebut secara dini sangat penting. Paper ini mengusulkan sebuah deep learning model yang disebut dengan Dilated Convolution and Spectral Attention Module (DC-SAM), yang mengekploitasi perbedaan informasi spectral yang berasal dari tanaman gandum yang sehat dan tanaman gandum yang mengalami gangguan. Metode DC-SAM yang diusulkan memiliki dua modul utama: **(i)** modul konvolusi dengan dilasi untuk menangkap fitur spektral dengan bidang reseptif yang besar; **(ii)** modul spectral attention untuk memadukan fitur spektral secara adaptif berdasarkan keterkaitannya. Karena modul konvolusi dilatasi memiliki bidang reseptif yang panjang, modul ini dapat menangkap pola ketergantungan pendek dan panjang yang ada dalam data hiperspektral. Hasil eksperimen kami dengan empat dataset menunjukkan bahwa DC-SAM mengungguli metode-metode sebelumnya. Selain itu, keluaran spectral attention modul yang diusulkan mampu mengungkapkan pita spektral paling diskriminatif untuk tugas klasifikasi gangguan pada gandum.*

Kata Kunci*: konvolusi dengan dilasi, penjelasan pada AI, informasi hiperspectral, attensi pada informasi spectral, gangguang kelebihan garam pada gandum.*

I. INTRODUCTION

SALT stress in wheat crops affects their growth and productivity [1] [2]. Traditionally, it is identified by observing visual symptoms of the crop [3]. However, by the time visual symptoms appear, the crop condition is critical hence it is too late to cure the stress. In other side, in the early stress, plants give responses through tissue degradation and changes in chlorophyll content as well as cellular metabolism [4]. These

http://dx.doi.org/10.12962/j24068535.v21i2.a1219

changes can be captured by hyperspectral sensors and are reflected into spectral information.

Several techniques have been used to extract features from hyperspectral data. These methods include Recurrent Neural Networks (RNNs) [5] [6], Long Short-Term Memory (LSTM) [7] [8] [9], and convolutional neural network (CNN) [10]. However, features from the spectral information that consists of hundreds of narrow bands, where adjacent bands tend to be highly correlated, cannot be captured effectively by those methods because they cannot capture long-range pattern dependencies.

To overcome the problem, we propose a novel deep learning network, dubbed DC-SAM (Dilated Convolution and Spectral Attention Module), that is constructed by using stacked dilated convolutional layers. In the first layer, the dilated convolution layer has a dilation rate of 1, which corresponds to a standard convolution, allowing it to extract the local features. The receptive field for the convolution layer gradually becomes larger by increasing the dilation rate, making it able to extract various levels of global features. The receptive field of our network can cover the data from the first band up to the last band, despite its shallow nature.

Although deep learning models can help solve problems related to feature extraction, their complexity makes them exhibit low explain ability [11]. As a result, it is very difficult to identify the factors determining the model's predictions, making it difficult to trust the model before deploying it ‘in the wild’ [12]. Furthermore, if we know which features affect the model's prediction, then we can determine which spectral bands weight most on the decision made for a given hyperspectral classification problem. The important bands could also be useful in developing a custom multispectral camera [13] for the targeted application.

There have been previous studies done to determine which bands are important for detecting stress in crops. In the study in [13], several machine learning methods were investigated (i.e., ReliefF, SVM-RFE, and Random Forest) to identify important bands for detecting salt stress in wheat crops. The results of these three different methods are mostly similar. However, their detection performance needs to be improved. In contrast, the studies in [14] [15] examined the bands which respond to salinity stress (i.e., NaCl), in soybeans plants using several methods, including student t-criterion and discriminant analysis. They compute the statistical significance between the average values of each spectral index between plants treated with NaCl and without. Only 40 samples were used in their study, the wavelength range was divided into five ranges, and the significance was calculated based on the range area. In addition, this method only focused on finding the significant wavelength, not on detecting stress.

In the case of classification, several studies have been conducted to improve the interpretability of complex models, including deep learning networks. Ribeiro et al. [12] introduced the Local Interpretable Model-agnostic Explanation (LIME) framework. LIME attempts to understand the model by tweaking the input features and observing their effect on changes in the model's predictions. Another approach, SHAP (Shapley Additive exPlanations) [16] was presented to explain model predictions by calculating the contribution of each feature to the model output. Both methods treat the model as a black box, so they do not know exactly how the model works.

To focus on the informative features, several deep learning networks have leveraged attention modules [17]. For example, the Convolutional Block Attention Module (CBAM) model was used to detect objects in RGB images that have spatial and channel information [17]. CBAM shows that adding an attention module enhances in turn network interpretability. As in CBAM, our network also integrates an attention module to increase network explain ability. Since our data has only spectral information, we design a spectral attention module to exploit the spectral relationship. We plug the spectral attention module at the end of the dilated convolution layer. After that, we visualize the attention module weights to identify which wavelengths or bands are important to our network's decision. We also explain the important band for our model's decision with LIME and SHAP.

In summary, this paper contributions are three-fold: (1) We leverage dilated convolutional layers to capture the features from the spectral information with a long sequence. In contrast to standard convolution, which only captures the local features, our causal dilated convolutional layers can capture both local and global features. (2) We achieve state-of-the-art performance: our experiments show, for example, that our proposed method improves the accuracy to classify wheat salt stress by more than 5% on the CS dataset. The findings demonstrate that the proposed DC-SAM network can detect the stress in wheat even before the visual symptoms arise. (3) We design a spectral attention module in our network to improve the interpretability of our network and identify which wavelengths or bands are essential for our network's decision. The rest of structure of this paper is organized as follows. Section II provides an overview of the related works, including dilated convolution, attention module, LIME, and SHAP. Section III explains the proposed DC-SAM method. Experimental results, performance evaluation, and model explanation are discussed in Section IV. The research findings are concluded in Section V.

II. RELATED WORKS

A. *Attention Mechanism*

An attention mechanism makes a network pay more attention on informative features of an input [18] [19]. Self-attention is an attention mechanism where an input in the input sequence interacts with other inputs in the sequence and learn which inputs the module should pay more attention to. Self-attention is popular in many fields, such as natural language processing and computer vision [20] [21] [22] [23].

In computer vision, a spatial attention module was used to make the network attends most on informative spatial area to make decision [19]. Another work proposed a Convolutional Block Attention Module (CBAM), which uses both spatial attention and channel attention [17]. They argued that spatial attention focuses on 'where' the informative part is, while the channel attention focuses on 'what' is essential given an input image.

In HSI, attention mechanisms have been used by several studies [24] [25] [26] to improve the network performance. However, they could not figure out the importance of a specific band. A spectral attention module is applied in this paper to hyperspectral data that has only spectral information. The spectral attention module is described in Section III.B.

B. LIME and SHAP

It is important to understand the difference between features and interpretable representation before explaining how LIME and SHAP work. LIME uses an interpretable representation that can be interpreted by humans in place of the actual features (original representation) used by a classifier [12]. Text classification, for example, uses features that are complex, such as word embeddings. However, an interpretable representation can be found using a binary vector representing each word's contribution to the class decision (e.g., 1 indicates 'has contribution', 0 indicates 'does not have contribution').

Suppose we want to explain an instance *x*, where $\mathrm{x} \in \mathrm{R}^{\mathrm{d}}$ is the original representation of the instance, and *d* is the number of features. The interpretable representation of the instance is then generated, $\mathrm{x}' \in \{0,1\}^{\mathrm{d}'}$. Given x', by drawing nonzero elements of x' uniformly at random, *N* sample instances (perturbed instances) around x' are produced. The perturbed instances are symbolized by *Z*, and each instance is denoted by z' where $\mathrm{z}' \in \{0,1\}^{\mathrm{d}'}$. Once z' is known, its original representation (*z*) can be recovered.

The LIME process begins with computing the distance between an instance *x* and a perturbed instance *z*. Distances, $\mathrm{D(x,z)}$, can be computed using a variety of functions, for instance, cosine distance for text or L2 distance for images or hyperspectral data. Equation (1) is then used to compute the similarity between *x* and *z*, ($\pi_{\mathrm{x}}(\mathrm{z})$), where σ is the standard deviation.

The original classifier *f* that is usually used for classification is complex, e.g., deep networks. Because the original classifier is difficult to explain, LIME uses simpler explainable models (surrogate models) [27]. A surrogate model is defined as $\mathrm{g} \in \mathrm{G}$, where *G* is a class of potentially interpretable models such as decision tree and linear regression. Consider *f(z)* to be the label of a perturbed sample *z* with the original classifier, and *g(z')* to be the label of an interpretable version of *z* with the explanation model; the measure of how unfaithful *g* is in representing *f* is defined by the expression $\mathrm{L(f,g,\pi_x)}$ that is computed by Equation (2).

The best representation of instance *x* using explanation models in *G* is $\xi(\mathrm{x})$ (Equation (3)), where $\Omega(\mathrm{g})$ represents the complexity of *g*. For example, in a decision tree, $\Omega(\mathrm{g})$ is its depth, and in a linear regression, $\Omega(\mathrm{g})$ is its non-zero weight [12].

$$\pi_x(z) = \exp\left(-\frac{D(x,z)^2}{\sigma^2}\right) \tag{1}$$

$$L(f,g,\pi_x) = \sum_{z,z' \epsilon Z} \pi_x(z)(f(z) - g(z'))^2 \tag{2}$$

$$\xi(x) = \underset{g \epsilon G}{\operatorname{argmin}} L(f,g,\pi_x) + \Omega(x) \tag{3}$$

Like LIME, SHAP uses an interpretable model to explain the prediction of the original model. SHAP utilizes the idea of the Shapley values to model the importance of a feature [16]. The Shapley values use all possible combinations of inputs to measure all possible predictions of an instance. Therefore, its exact computation is challenging, but it does guarantee the accuracy and consistency of the importance of the features. The LIME[1] and SHAP[2] modules are available online.

[1] https://github.com/marcotcr/lime

[2] https://github.com/slundberg/shap

http://dx.doi.org/10.12962/j24068535.v21i2.a1219

III. PROPOSED METHODOLOGY

A flowchart of our proposed work, DC-SAM, is shown in **Error! Reference source not found.**. The input of the training, testing, and explaining phases are a spectral vector, sized B × 1, where B is the number of bands. We consider these inputs as a sequence of spectral channels.

In the training phase, the training inputs are exploited to train the DC-SAM network. First, each input is convolved by using 1D convolution with 24 output channels and kernels of size 3, producing an intermediate feature with size S × 1 × 24, i.e., $\mathrm{F}^{S\times1\times24}$, where S is the spectral feature size. The kernels example representation used in the first convolution layer is shown in **Error! Reference source not found.**.

The intermediate feature is used as an input to the spectral dilated convolution module (illustrated in **Error! Reference source not found.**), which produces a refined feature. Then, a spectral attention module (see **Error! Reference source not found.**) is applied to the refined feature to produce a spectral attention map. A pixel-wise multiplication is then operated between the refined feature and the spectral attention map. The result is then processed by a pooling operation and a classifier consisting of a dense layer with ReLU and a softmax layer. The classifier output is a label prediction, which is then compared with the true label to produce training loss. The training loss is further utilized to update the DC-SAM training parameters. These processes are repeated with a certain iteration size to build a trained DC-SAM model.

During testing, the trained DC-SAM model is used to classify the test inputs, and the output is a prediction label. To obtain the performance measurements, the prediction and the true labels are computed. We can then draw a heat-map of the spectral attention module output by using the test input and the trained DC-SAM model to identify the bands that the model focuses on to produce an explanation. Examples of model explanations produced by LIME and SHAP based on a test input and the trained DC-SAM model are shown in Figure 7 and Figure 8.

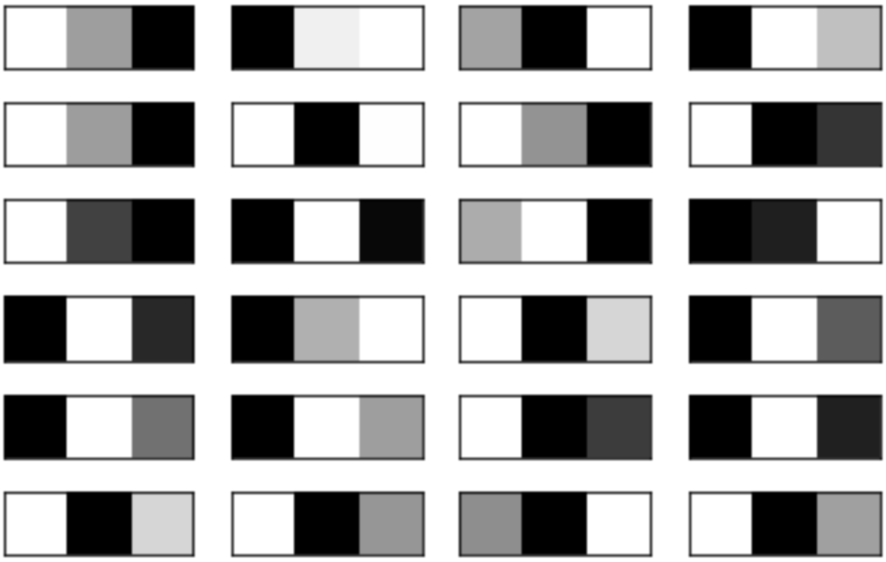


Figure 2 The sample of kernels for processing spectral input.

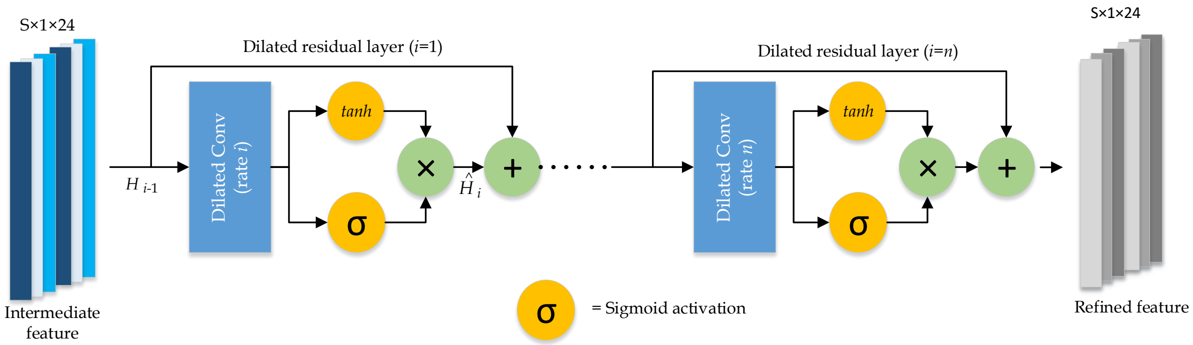
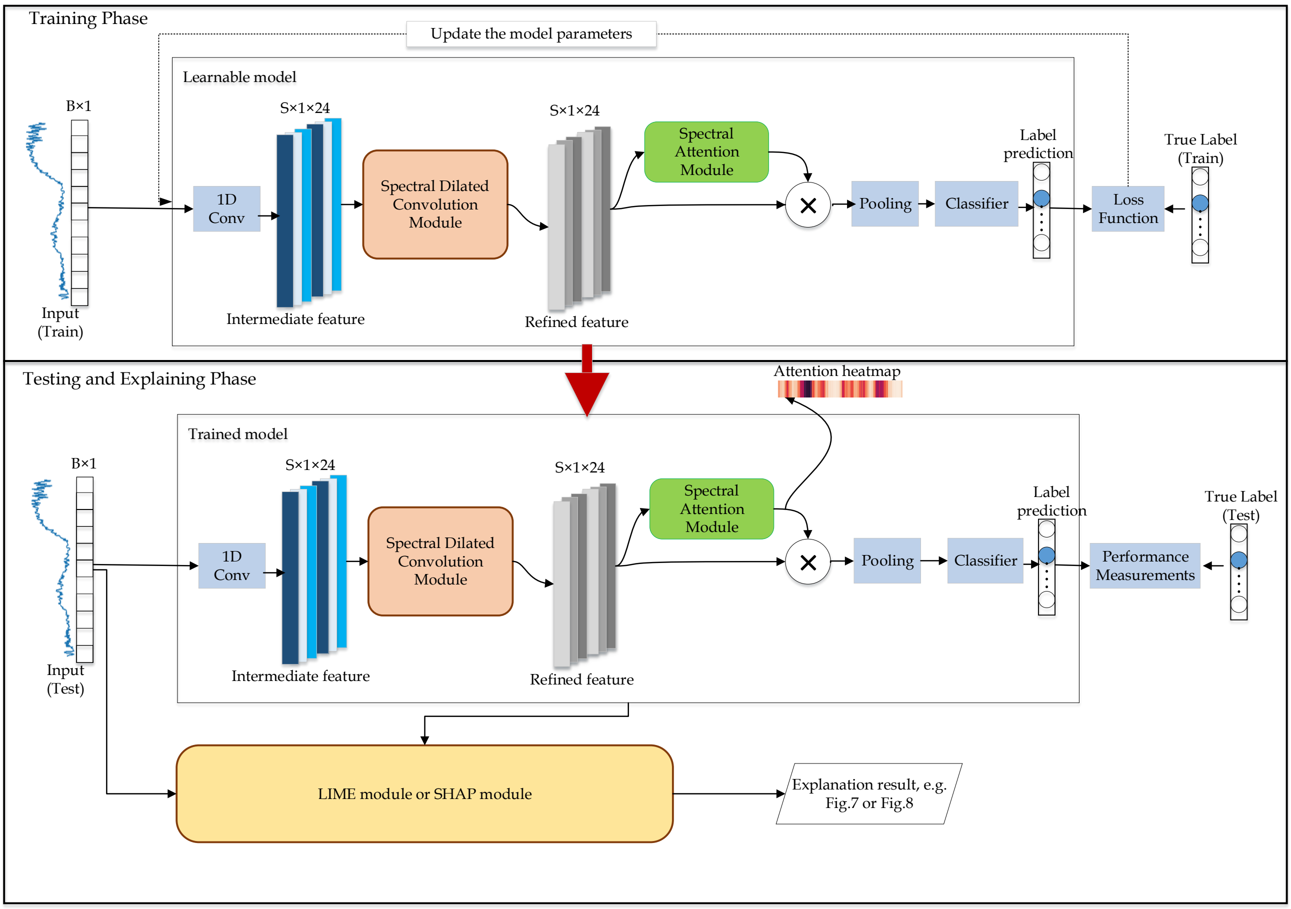


Figure 1 The architecture of the spectral dilated convolution module, which consists of n dilated residual layers. Each dilated residual layer contains a dilated convolutional layer and two activation functions, 'tanh' and '\sigma' (sigmoid)



A. Spectral Dilated Convolution

Error! Reference source not found. illustrates the architecture of the spectral dilated convolution module. The input of this module is an intermediate feature with size S × 1 × 24, and the output is a refined feature with the same size. Our spectral dilated convolution module consists of *n* dilated residual layers. Each dilated residual layer contains one dilated convolutional layer and two activation functions, '*tanh*' and 'σ' (sigmoid).

As in WaveNet [28], our dilated convolutional layers use a dilation factor of 2^{i-1}, where *i* is the dilated residual layer number. The dilation factor that increases exponentially with depth results in the exponential growth of the receptive field, and thus stacking the dilated residual layers will increase the model capacity.

In contrast to WaveNet, which uses causal dilated convolution, we use acausal dilated convolution. WaveNet uses causal dilated convolution since it assumes that an input at a timestep *t* is only conditioned by the inputs at all previous timesteps. However, in our case, we consider that the information at a particular band is correlated to the adjacent bands (the previous and the next bands), since [9] pointed out that their network that utilized both previous and latter information to explore the spectral information performs better than the network that only utilized previous information. As a result, we used acausal dilated convolution. We used *tanh* and σ for activation function, like in the gated PixelCNN [28] [29]. These works have shown that *tanh* and σ improve the network's performance [28] [29].

B. Spectral Attention Module

The output of the spectral dilated convolution module, $F \in R^{S\times1\times M}$, consists of *M* channels (in our architecture *M=24*) and *S* spectral features. The spectral attention module then computes the global average pooling along the channel axis to generate an efficient feature descriptor producing $F_{avg} \in R^{S\times1}$. We further implement a convolutional layer, which can extract the inter-spectral relationships between features, to generate a spectral attention map, $M_{sp} \in R^{S\times1}$. The spectral attention map encodes which feature to emphasize or suppress. Equation (4) shows the formula to compute the spectral attention map, where $f^3(.)$ represents a 1D convolution with filter

Figure 3 The overall architecture of the proposed network, which consists of two main modules, i.e., a spectral dilated convolutional module to capture the long-range dependencies between spectral bands and a spectral attention module to give weight to each spec

size 3. The final feature map, $F^{f} \in R^{S\times1\times M}$ is computed by using Equation (5).

$$M_{sp}(F) = \sigma\left(f^{3}(F_{avg})\right) \tag{4}$$

$$F^{f} = F \odot M_{sp}(F) \tag{5}$$

IV. Experiments and Analysis

A. Experimental Setting

Datasets: We evaluated DC-SAM on four datasets for wheat salt stress classification: Chinese Spring (CS), Aegilops columnaris (co(CS)), Ae.speltoides auchery (sp(CS)), and Kharchia dataset [30]. These datasets can be accessed freely[3]. The dataset also has a file that contains the details of the corresponding wavelengths associated with each band [30]. We then used the wavelength information as the name of features instead of the band's number in the model explanation (see Section IV E).

Evaluation Protocols and Performance Measurements: For the experiments, alike [13], we used 70% data as training samples and 30% data as testing samples. In each experiment, we applied 10-fold cross-validation. As preprocessing, we utilized a standardization technique to rescale data to have a mean of 0 and a standard deviation of 1. For training, we used Adam optimizer with a learning rate of 0.0003, the batch size was 256, and the number

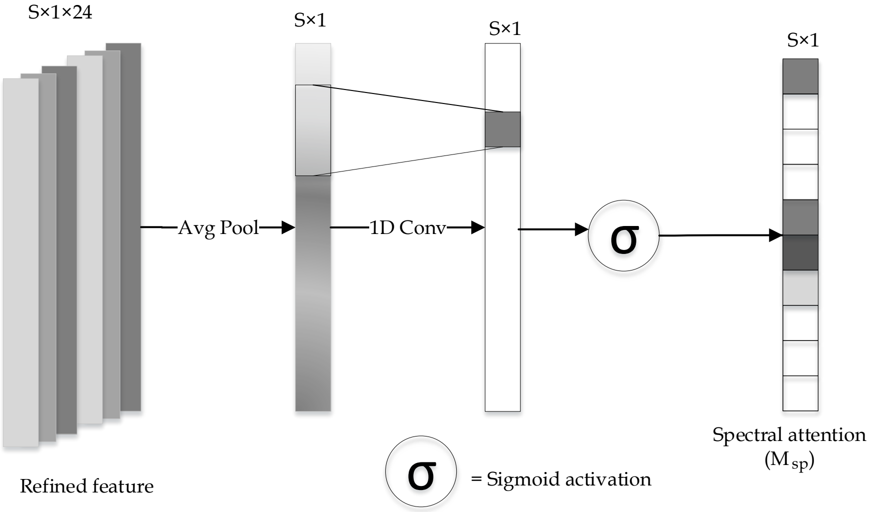


Figure 5 Spectral Attention Module

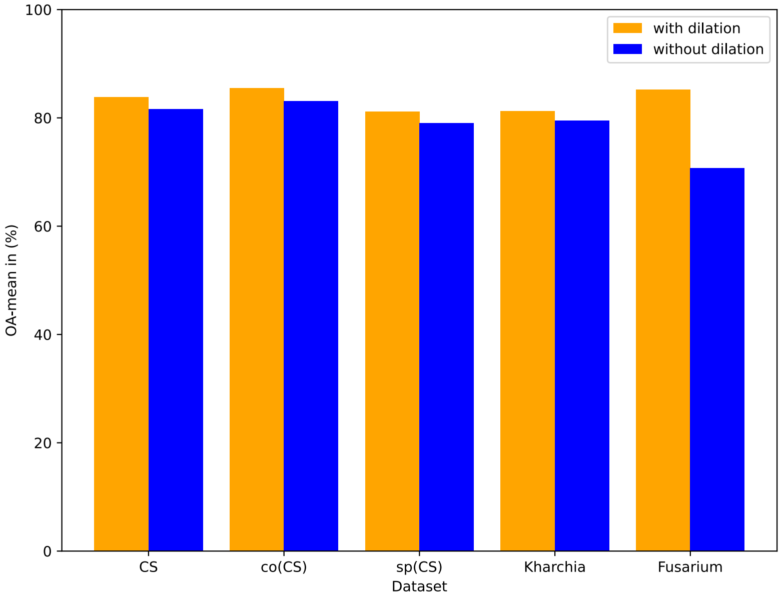


Figure 4 Performance comparison of our proposed method for different numbers of dilated convolutional layers. OA-mean is the average of OA from 10-fold experiments.

[3] https://conservancy.umn.edu/handle/11299/195720

of iterations was 200. For evaluation, we computed the F1 measure of control (C0) and salt (C1) classes, Overal Accuracy (OA), and Average Accuracy (AA) to evaluate the proposed method's performance.

B. Ablation Analysis: Impact of dilation on performance

Error! Reference source not found. shows the impact of dilation on performance. Using the depth of 7, we evaluated the model in two scenarios: with dilation and without dilation. Their architectures were the same, but for the model without dilation, a constant dilation rate of 1 was used instead of 2^{i-1} where *i* is the depth of the layer. The figure shows that the model with dilation achieved better performance than the respective model without dilation. The dilated convolution improves OA by more than 2% on CS, co(CS), and sp(CS) datasets. On Kharchia dataset, the OA of with dilation model is around 1% higher compared with a standard convolution.

The results indicate that the dilated convolution is more suitable for hyperspectral data than the standard convolution. The dilated convolutions enable the network to have a larger receptive field than the standard convolution, and therefore it can capture global features and longer dependencies between bands.

C. Ablation Analysis: Impact of spectral attention module on performance

The proposed architecture uses a convolutional-based spectral attention module to weigh the high-level spectral features. In this experiment, we investigate the impact of the attention module on the classification performance. **Error! Reference source not found.** shows that the model's performance with attention is better than its performance without attention only in some cases. The spectral attention module does not adversely impact the classification performance. Most importantly, the attention module still meets its main purpose, which is to weigh the spectral features. As a result, the network can reveal the relative importance of bands (see Section IV E).

Table 1 Performance comparison with and without spectral attention module

Dataset	Performance	With spectral attention	Without spectral attentions
CS	OA	83.85+-0.60	83.72+-0.49
	AA	83.71+-0.60	83.58+-0.48
	F1C0	82.31+-0.81	82.37+-0.58
	F1C1	85.14+-0.50	84.87+-0.49
co(CS)	OA	85.46+-0.97	85.70+-1.02
	AA	84.19+-1.10	84.64+-1.29
	F1C0	88.61+-0.78	88.87+-0.88
	F1C1	79.88+-1.31	79.97+-1.43
sp(CS)	OA	81.15+-0.49	81.18+-0.74
	AA	81.16+-0.45	81.20+-0.76
	F1C0	81.99+-0.57	82.07+-0.84
	F1C1	80.21+-0.66	80.19+-0.76
Kharchia	OA	81.21+-0.48	81.13+-0.52
	AA	80.47+-0.60	80.34+-0.60
	F1C0	74.59+-0.86	74.55+-0.40
	F1C1	85.08+-0.46	85.00+-0.47

D. Comparison with existing methods

This experiment compared our proposed architecture with existing state-of-the-art architectures for the early detection of plant stress. We focused on models that use only spectral information because we focus to learn vegetation interaction with spectral reflectance. For CS, co(CS), sp(CS), and Kharchia datasets, we compared our model with a model that treats spectral information as a vector and uses the standard 1D convolution to extract the features [10]. In contrast to the standard 1D convolution used in Section IVB}, the 1D convolution in [10] used a very long filter size (number of bands/4). We also compared the proposed method with the spectral-residual network (sRN) [31], LSTM [32] and spectralFormer [33]. The result is shown in Table 2, which proves that the proposed method produces the highest performance. Moreover, our proposed method outperforms SFS Forward by a large margin of 6% in terms of F1-mean on the CS dataset.

DC-SAM is superior to the other existing methods because it is able to capture both local and global features and it has longer receptive field making its capacity higher. Additionally, because our method has fewer parameters, it is suitable for problems with a small number of training samples, as opposed to spectralFormer, a type of Transformer that needs a lot of training samples. Considering that the datasets for salt stress detection are taken before visual symptoms appear, the good results indicate the potential for detecting stress early with hyperspectral information is promising. An early crop management solution can then be implemented to minimize crop production loss.

However, if the train and test samples are from different environments or growth stages, the performance may be compromised. As an example, the model is trained with data from hydroponically grown crops and then tested against data from real-life fields, or the model is trained with data from early-stage crops and tested against late-stage crops. The reason is that some growth factors (e.g., soil type, lighting, irrigation) and the growth stage influence crop spectral behavior. As a result, domain shifts may occur between the train and test samples. Future studies are required to solve the domain shift problem.

Table 2 Performance comparison between our proposed method and other methods.

Method	$\mathbf{F1_{C0}}$	$\mathbf{F1_{C1}}$	$\mathbf{F1-mean}$	**OA**	**AA**
CS					
1D CNN [10]	79.50±0.86	83.02±0.74	81.26±0.77	81.43±0.77	81.27±0.78
Group LSTM [32]	76.94±0.81	80.86±0.61	78.90±0.66	79.09±0.65	78.91±0.65
sRN [31]	79.66±0.60	81.87±0.95	80.77±0.68	80.84±0.70	80.81±0.62
spectralFormer [33]	77.18±1.70	80.17±1.24	78.68±1.25	78.81±1.20	78.77±1.12
SFS_forward [13]	78.87	76.55	77.71	-	-
Proposed Method	**82.31±0.81**	**85.14±0.50**	**83.72±0.63**	**83.85±0.60**	**83.71±0.60**
co(CS)					
1D CNN [10]	86.08±0.66	74.60±1.00	80.34±0.80	82.01±0.77	80.58±0.90
Group LSTM [32]	84.22±0.44	70.94±0.81	77.58±0.56	79.55±0.51	77.89±0.59
sRN [31]	84.67±0.85	70.52±2.05	77.59±1.23	79.85±0.99	78.54±1.13
spectralFormer [33]	86.23±0.79	74.50±2.68	80.36±1.70	82.13±1.28	80.80±1.16
Proposed Method	**88.61±0.78**	**79.88±1.31**	**84.25±1.03**	**85.46±0.97**	**84.19±1.10**
sp(CS)					
1D CNN [10]	78.92±0.61	76.16±0.71	77.54±0.63	77.62±0.62	77.65±0.63
Group LSTM [32]	76.33±0.93	73.25±0.93	74.79±0.78	74.89±0.78	74.92±0.83
sRN [31]	77.99±0.82	74.83±1.25	76.41±0.96	76.52±0.94	76.58±0.93
spectralFormer [33]	77.52±1.35	75.22±1.21	76.37±1.10	76.44±1.12	76.49±1.19
Proposed Method	**81.99±0.57**	**80.21±0.66**	**81.10±0.50**	**81.15±0.49**	**81.16±0.45**
Kharchia					
1D CNN [10]	70.49±0.74	82.55±0.51	76.52±0.61	78.07±0.60	76.99±0.67
Group LSTM [32]	66.84±0.98	79.86±0.72	73.35±0.81	74.94±0.80	73.56±0.86
sRN [31]	69.38±1.11	82.58±0.54	75.98±0.67	77.80±0.58	76.91±0.71
spectralFormer [33]	67.26±1.79	81.04±1.08	74.15±1.31	76.00±1.23	74.83±1.36
Proposed Method	**74.59±0.86**	**85.08±0.46**	**79.83±0.53**	**81.21±0.48**	**80.47±0.60**

E. Model Explanation

We have shown that our proposed model achieves good performance on all of the datasets. As deep learning models have low explain ability, we used a spectral attention module to form an output heatmap displaying which features influence most the model's decision. For validation, we used LIME and SHAP to explain the model's decision [16] [27].

For example, given a spectral sequence as input (Figure 6), Figure 7 shows its LIME explanation. The feature value shown in Figure 7 is the feature value of the input after the standardization process. From the bar chart, we can see that if w_579.35 ≥ 0.63, then the wavelength will contribute to class ‘1’. The feature value of wl_579.35 is 2.18. Hence, the wavelength contributes to class ‘1’ with a weight of 0.16. Here, value 0.63 is the threshold value of wl_579.35. The further the feature value from the threshold, the higher the weight. The figure shows that the prediction output of a given LIME input is ‘1’, shown in orange colour, with a prediction probability of 1.00. The wavelengths that support the decision to class ‘1’ are also highlighted in orange. We can see from the figure that wavelengths 579.35, 599.88, 784.63, and 597.83 contribute to class ‘1’ with a weight of 0.16, 0.13, 0.1 and 0.1, respectively. The rest wavelengths with orange colour contribute to class ‘1’ with a small weight (less than 0.09). Hence, we only consider wavelengths that contribute on supporting the decision are 579.35, 599.88, 784.63, and 597.83.

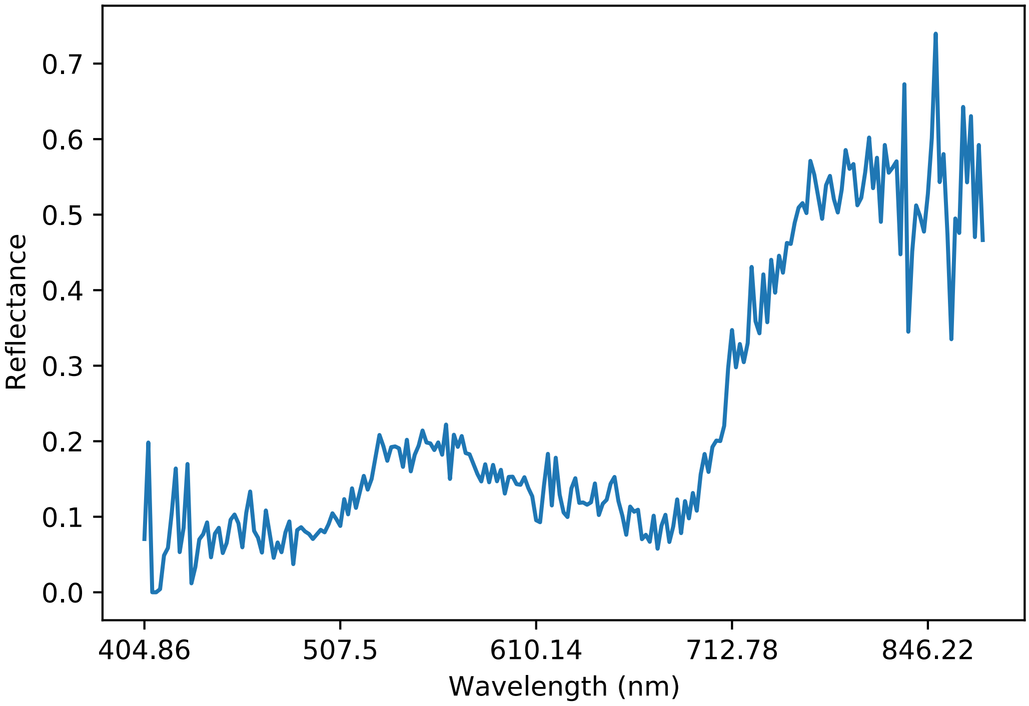


Figure 6 An example of spectral input, which has spectral information from wavelength 404.86 to 874.96. The label of this input is 1 (i.e., salt treatment).

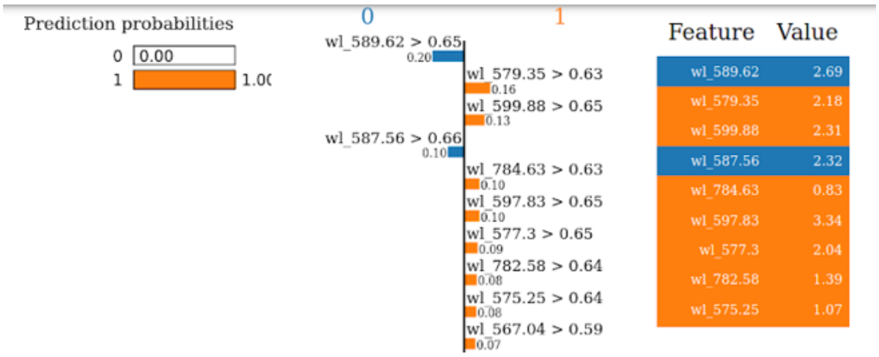


Figure 7 Our model predicts that the spectral input belongs to class `1', and LIME highlights the wavelength in the spectral input that led to the prediction. The bar chart represents the most relevant wavelengths. The colour indicates which class the wavelength.

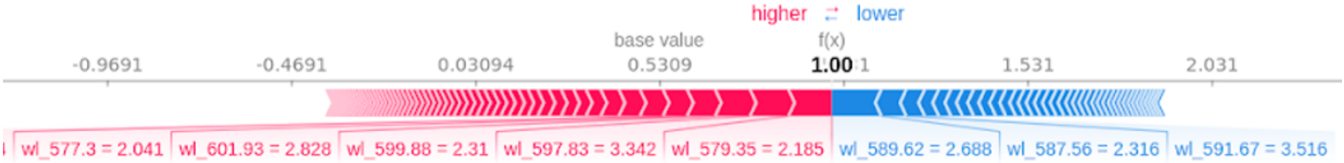


Figure 8 Our model predicts that the spectral input belongs to class ‘1’. The SHAP values encode the wavelength's support to the model prediction. Wavelength 579.35,597.83, 599.88, and 601.93 support most to the prediction of class ‘1’.

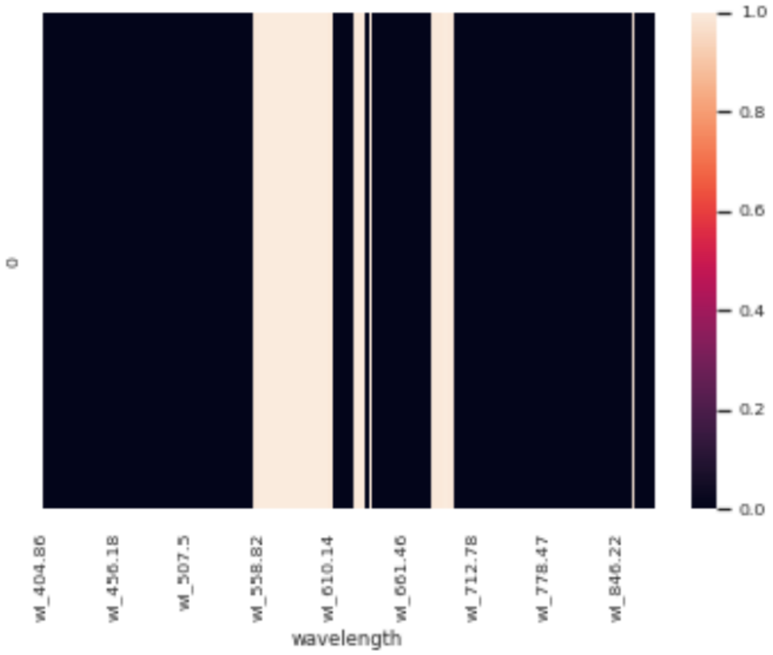
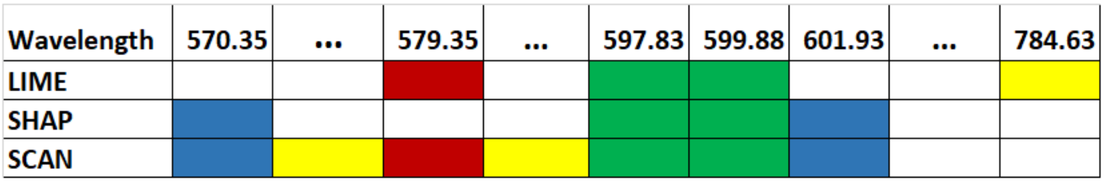


Figure 9 The attention wavelengths range produce by spectral attention module in the DC-SAM, where we selected the most important bands by setting a high threshold, 0.97, to the attention values.

Figure 8 shows the feature importance by SHAP technique from the same spectral input, with the features that improve predictions are represented in pink, while features that degrade the prediction are shown in blue. Furthermore, the visual size shows the magnitude of the feature's effect. Figure 8 shows that the most significant wavelengths that support prediction are 570.35, 597.83, 599.88, and 601.93.

We also generated the spectral attention map to see which bands the model gave more attention. Figure 9 displays an example of spectral attention map visualization. Specifically, we selected the most important bands by setting a high threshold, 0.97, to the attention values. The figure shows that DC-SAM pays more attention to the wavelength range of 555-610, around 630-640, and 680-700.

A cursory look at Figure 7, Figure 8, and Figure 9 may indicate that the results of a few specific feature bands look different. In order to have a deeper insight, we represent the specific feature bands produced by each method e.g., in Figure 10, where the important wavelengths considered by each technique are presented in colours. The wavelengths where LIME, SHAP, and DC-SAM intersect colour with green, the wavelengths where SHAP and DC-SAM intersect colour with blue, the wavelengths where LIME and DC-SAM intersect colour with red, and the wavelengths that do not intersect colour with yellow. The figure shows that all important wavelengths considered by SHAP intersect with DC-SAM. For LIME, three out of four wavelengths that are considered as important intersect with DC-SAM. DC-SAM explanation result is still consistent with LIME and SHAP because most of important wavelengths produced by LIME and SHAP are in the range 555-610 (see Figure 10). But none of the specific bands produced by LIME and SHAP lay in the range of 630-640, and 680-700. We then draw explanation from more samples to clearly understand the explanation.



Wavelength	570.35	...	579.35	...	597.83	599.88	601.93	...	784.63
LIME			🟥		🟩	🟩			🟨
SHAP	🟦				🟩	🟩	🟦		
SCAN	🟦	🟨	🟥	🟨	🟩	🟩	🟦		

Figure 10 The wavelengths that are considered as important by LIME, SHAP, DC-SAM. Green represents wavelengths where LIME, SHAP, and DC-SAM intersect. Blue = wavelengths where SHAP and DC-SAM intersect; Red= the wavelengths where LIME and DC-SAM intersect, and Yellow = the wavelength where none intersects.

V. Conclusion

The paper focuses on the spectral information and uses the spectral response of the crops to detect stress before visible symptoms appear. It proposes a deep learning architecture with dilated convolutional layers to extract the spectral features for salt stress classification. The main idea is using dilated 1D convolution on the spectral data to capture the short- and long-dependencies between bands. Our experiments on CS, co(CS), sp(CS), and Kharchia datasets show that the dilated convolution produces higher performance than the standard convolution.

The paper also proposes a spectral attention module to explain DC-SAM's prediction by showing which bands the model gives more attention to. From experiments, we demonstrate that 1) Some explanations by LIME, SHAP exactly lay in the range where DC-SAM pays attention to. 2) Several more samples produce a similar explanation, where the bands considered as giving high support to the prediction by LIME, SHAP, and DC-SAM have intersections. 3) Some samples with low confidentiality in their classification result have different explanations with DC-SAM, LIME, and SHAP. The possible reason is DC-SAM, LIME, and SHAP face difficulties explaining the prediction result when the data is also difficult to classify.

References

[1] M. Sarwat, A. Ahmad, M. Z. Abdin and M. M. Ibrahim, Stress Signaling in Plants: Genomics and Proteomics Perspective, 2 ed., Springer International Publishing, Cham, 2013.

[2] N. Suzuki, R. M. Rivero, V. Shulaev, E. Blumwald and R. Mittler, "Abiotic and biotic stress combinations," *New Phytologist,* vol. 203, no. 1, pp. 32-43, 7 2014.

[3] Y. Wang, H. Wang and Z. Peng, "Rice diseases detection and classification using attention based neural network and bayesian optimization," *Expert Systems with Applications,* vol. 178, p. 114770, 9 2021.

[4] A.-K. Mahlein, "Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping," *Plant Disease,* vol. 100, no. 2, pp. 241-251, 2016.

[5] A. Graves, A. R. Mohamed and G. Hinton, "Speech recognition with deep recurrent neural networks," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings,* pp. 6645-6649, 10 2013.

[6] V. Chow, "Predicting auction price of vehicle license plate with deep recurrent neural network," *Expert Systems with Applications,* vol. 142, p. 113008, 3 2020.

[7] Z. C. Lipton, J. Berkowitz and C. Elkan, "A Critical Review of Recurrent Neural Networks for Sequence Learning," *arXiv preprint arXiv:1506.00019,* 5 2015.

[8] F. Zhou, R. Hang, Q. Liu and X. Yuan, "Hyperspectral image classification using spectral-spatial LSTMs," *Neurocomputing,* vol. 328, pp. 39-47, 2 2019.

[9] Q. Liu, F. Zhou, R. Hang and X. Yuan, "Bidirectional-Convolutional LSTM Based Spectral-Spatial Feature Learning for Hyperspectral Image Classification," *Remote Sensing,* vol. 9, no. 12, p. 1330, 12 2017.

[10] W. Hu, Y. Huang, L. Wei, F. Zhang and H. Li, "Deep Convolutional Neural Networks for Hyperspectral Image Classification," *Journal of Sensors,* vol. 2015, pp. 1-12, 7 2015.

[11] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion,* vol. 58, pp. 82-115, 6 2020.

[12] M. T. Ribeiro, S. Singh and C. Guestrin, ""Why should i trust you?" Explaining the predictions of any classifier," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* Vols. 13-17-Augu, pp. 1135-1144, 8 2016.

[13] A. Moghimi, C. Yang and P. M. Marchetto, "Ensemble Feature Selection for Plant Phenotyping: A Journey from Hyperspectral to Multispectral Imaging," *IEEE Access,* vol. 6, pp. 56870-56884, 2018.

[14] I. Iliev, D. Krezhova, T. Yanev, E. Kirova and V. Alexieva, "Response of chlorophyll fluorescence to salinity stress on the early growth stage of the soybean plants (Glycine max L.)," *4th International Conference on Recent Advances Space Technologies,* pp. 403-407, 2009.

[15] D. Krezhova, I. Iliev, T. Yanev and E. Kirova, "Assessment of the effect of salinity on the early growth stage of soybean plants (Glycine max L.)," *RAST 2009 - Proceedings of 4th International Conference on Recent Advances Space Technologies,* pp. 397-402, 2009.

[16] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems,* Vols. 2017-Decem, pp. 4766-4775, 5 2017.

[17] S. Woo, J. Park, J.-Y. Lee and I. S. Kweon, "CBAM: Convolutional Block Attention Module," *Proceedings of the European Conference on Computer Vision (ECCV),* pp. 3-19, 2018.

[18] J. Hou, G. Wang, X. Chen, J.-H. Xue, R. Zhu and H. Yang, "Spatial-Temporal Attention Res-TCN for Skeleton-based Dynamic Hand Gesture Recognition," *Proceedings of the European Conference on Computer Vision (ECCV) Workshops,* p. 0, 2018.

[19] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings,* p. 0, 12 2017.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems,* vol. 30, pp. 5999-6009, 6 2017.

[21] J. Cheng, L. Dong and M. Lapata, "Long Short-Term Memory-Networks for Machine Reading," *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings,* pp. 551-561, 1 2016.

[22] Z. Lin, M. Feng, C. N. dos Santos, M. Yu, B. Xiang, B. Zhou and Y. Bengio, "A Structured Self-attentive Sentence Embedding," *arXiv preprint arXiv:1703.03130,* 3 2017.

[23] A. P. Parikh, O. Täckström, D. Das and J. Uszkoreit, "A Decomposable Attention Model for Natural Language Inference," *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings,* pp. 2249-2255, 6 2016.

[24] L. Mou, P. Ghamisi and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, no. 7, pp. 3639-3655, 7 2017.

[25] L. Mou and X. X. Zhu, "Learning to Pay Attention on Spectral Domain: A Spectral Attention Module-Based Convolutional Network for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 58, no. 1, pp. 110-122, 1 2020.

[26] Q. Liu, Z. Li, S. Shuai and Q. Sun, "Spectral group attention networks for hyperspectral image classification with spectral separability analysis," *Infrared Physics and Technology,* vol. 108, p. 103340, 8 2020.

[27] I. Kakogeorgiou and K. Karantzalos, "Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing," *International Journal of Applied Earth Observation and Geoinformation,* vol. 103, p. 102520, 12 2021.

[28] A. Van Den Oord, S. Dieleman, Zen and H, "WaveNet: A generative model for raw audio," *SSW,* no. 2, p. 125, 2016.

[29] A. van den Oord, N. Kalchbrenner, O. Vinyals, L. Espeholt, A. Graves and K. Kavukcuoglu, "Conditional Image Generation with PixelCNN Decoders," *Advances in Neural Information Processing Systems,* pp. 4797-4805, 6 2016.

[30] A. Moghimi, C. Yang, M. E. Miller, S. F. Kianian and P. M. Marchetto, "A Novel Approach to Assess Salt Stress Tolerance in Wheat Using Hyperspectral Imaging," *Frontiers in Plant Science,* vol. 9, p. 1182, 8 2018.

[31] W. N. Khotimah, M. Bennamoun, F. Boussaid, F. Sohel and D. Edwards, "A high-performance spectral-spatial residual network for hyperspectral image classification with small training data," *Remote Sensing,* vol. 12, no. 19, p. 3137, 10 2020.

[32] Y. Xu, L. Zhang, B. Du and F. Zhang, "Spectral-Spatial Unified Networks for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 56, no. 10, pp. 5893--5909, 2018.

[33] D. Hong, Z. Han, J. Yao and L. Gao, "SpectralFormer: Rethinking Hyperspectral Image Classification with Transfprmers," *IEEE Transactions on Geoscience and Remote Sensing,* p. 1, 2021.