A novel strategy for classifying spectral-spatial shallow and deep hyperspectral image features using 1D-EWT and 3D-CNN

Sandeep Kumar Ladi1,2 · G K Panda3 · Ratnakar Dash4 · Pradeep Kumar Ladi5

Received: 3 August 2022 / Accepted: 6 October 2022 / Published online: 20 October 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Hyperspectral images (HIs) are used in diverse disciplines, such as resource handling, land cover analysis, food science, anomaly detection, and precision agriculture. Researchers have been working on a number of visual processing and machine intelligence algorithms to handle this type of data as efficiently as feasible. Deep learning approaches have advanced significantly in the field of machine vision, which is also having a big impact on the analysis of hyperspectral data. To increase its discriminative potential for HI classification, this work suggests a powerful 3D-CNNs (Convolutional Neural Networks) architecture, where the shallow features extracted using 1D-EWT (Empirical Wavelet Transform) are served as input, and the ultimate output of the CNN are projected class-related outcomes. The framework is known as PEC, where P stands for PCA, E for 1D-EWT, and C for 3D-CNN model. Prior to features extraction, the HI undergoes spectral dimension reduction via Principal Component Anaalysis (PCA). To forecast segmentation for a volumetric area of a 3D HI sample, 3D CNNs use 3D convolutional kernels. The usage of more parameters by these CNNs means that the ability to employ interslice context that can boost performance. The CNN model’s parameters are optimised using a limited training set. The newly proposed PEC framework achieves a considerable overall accuracy of 99.58% and 99.94% (percentage increase of 0.08% and 0.54%) and Kappa score of 99.51% and 99.92 (percentage increase of 1.5% and 0.8%) compared to similar approaches for two real-world data sets IP and PU respectively with 30% training samples.

Keywords Hyperspectral images (HIs) · Principal Component Analysis (PCA) · 1D-EWT (Empirical Wavelet Transform) · PEC framework · 3D-CNN (Convolutional Neural Network)

Introduction
In practice, hyperspectral imaging (HI) refers to the process of capturing and analyzing an image at sizable wavelengths, allowing the image to be split into hundreds of colors (limited to four colours for multispectral imaging such as red, green, blue and near infrared). The techniques of spectroscopy (to detect materials based on how light interacts with a subject), enable the HI to gather sufficient bands of data to accurately represent each pixel of an image.

A drone equipped with a hyperspectral camera may identify plant diseases, weeds, soil erosion issues, water control issues, and predict crop yields in agriculture applications. HI is a non-destructive, non-contact technique that can be utilised without harming the object being examined. While multispectral data sets typically consist of roughly 5-10 bands with very high bandwidths, data sets of hyperspectral origin typically contain over hundred spectral bands and relatively low (5-10 nm) bandwidths. Most frequently employed spectral ranges include visible and infrared(near and short wave).

Hyperspectral sensors are passive components that gather data in the form of a collection of images that correspond to various electromagnetic spectrum bands. A hyperspectral data cube is created by combining these images; this cube can then be processed and examined to analyze the spectral data for a variety of uses.

Tan (2017) et al., discuss HI based advancements and identified HI tools for social, environmental and military
applications for past three decades. The vast majority of the Earth can be imaged using modern sensor technology, which has exceptional spatial, spectral, and temporal resolution. These characteristics enable the effective application of hyperspectral imaging in a plethora of remote sensing applications that necessitates the determination of the physical characteristics of various complex surfaces as well as the identification of visually comparable materials with precise spectral signatures. Spectral resolution and spatial resolution are two characteristics of hyperspectral images (HIs).

HI has two spatial dimensions ($H_x$ and $H_y$) measuring the geometric connection of the image pixels to one another and one spectral dimension ($S_z$), that reveals the changes in image pixel values with relation to wavelength (Khan et al. 2018).

One of the objectives of HI classification is to identify the class label for every pixel within the image and categorized as the key issues for the bulk of these applications. The fundamental issue in HI-based classification is coping with few training samples, high dimensional data, and expensive human labeling. (Paolletti et al. 2019).

To retrieve relevant spectral properties, band selection and dimension reduction are two most preferable approaches being suggested in Chang et al. (1999); Bai et al. (2015) and Harshanyi and Chang (1994); Tan et al. (2013) respectively. It is challenging to appropriately categorise the objects using only spectral data because of their different spatial distribution and spectral variation (Rajan et al. 2008). Spatial elements are becoming increasingly important in HI classification with the advancement of computer based intelligent image processing and vision (Fauvel et al. 2008). A plethora of spatial feature extraction strategies have been devised in recent years, including the wavelet transform (Zhu and Yang 1998), the gabor filter (Li and Du 2014) and the grey-level co-occurrence matrix (Pesaresi et al. 2008).

Applications of wavelet transform has been evolving since three decades and employed in many fields. Authors in Pittner and Kamarthi (1999); Mallat (1999) discuss such applications for analyzing signals and images, data compression, pattern recognition and human vision. For the purpose of HI classification, many other feature extraction methodologies have also been devised, all of which emphasize on the localization characteristics of the wavelet transform (Hsu and Ths 2003; Hsu et al. 2002). The regional spectral disparity of HI curve in diverse spectral channels at every wavelength band can be automatically evaluated and delivers instrumental details for HI classification in wavelet-based feature extraction techniques. Hsu and Ths (2003) provides additional information on wavelet-based feature extraction techniques. The aforementioned spatial feature extraction methods using the wavelet transform, Gabor filter and Grey-level co-occurrence matrix are together referred as the “handcrafted” or “Shallow” features (Xu et al. 2018). These features are distinct from those that are obtained from various layers of a deep learning model.

Due to the significant contributions and achievements of deep learning techniques in the realms of Visual Information processing, a practical alternative method to automatically extract valuable characteristics from HI has emerged (Krizhevsky et al. 2012; Ji et al. 2013; Sun et al. 2014; Sandeep Kumar et al. 2022).

Authors in Fukushima and Miyake (1982) introduced convolutional neural networks (CNN), and since then a widespread use in a variety of fields have been observed. Some includes, activity recognition (Papakostas et al. 2016; Yudistira and Kurita 2017), sentence classification (Kim 2014), text recognition (Zhou et al. 2017), face recognition (Ranjan et al. 2018), object detection and localization (Milyaev and Laptev 2017; Zhou et al. 2017) and image characterization (Druzhkov and Kustikova 2016).

Hyperspectral data exhibits the characteristics of non-stationary and non-linearity. As the Empirical Mode of Decomposition (EMD) suggested by Huang et al. in Huang et al. (1998), offers solutions for non-linear and non-stationary signals, the adaptive techniques of Empirical Wavelet Transforms (EWT) and EMD have immense potential for HI applications. EMDs also implemented for dimensionality reduction (Gormus et al. 2012), separation of harmonics from damaged bearings (Wang et al. 2014; Immovilli et al. 2010), and basic levels of HI classifications (Demir and Erturk 2008; Demir and Erturk 2010; Erturk et al. 2012; Jonnadula et al. 2020).

Numerous studies also evaluate the Ensemble Empirical Mode Decomposition (EEMD), Empirical Mode Decomposition (EMD), and EWT from the application perspective of bearing defect diagnostics. Studies in Prabhakar and Gee-tha (2017) observe that EWT outperforms in comparison to EEMD and EMD, where the time complexity in the former model is significantly low.

In this study, we use EWT-based shallow feature extraction method to handle the frequency band features of hyperspectral images. Further, to extract and classify the deep features, EWT-based features are further fed into a deep learning architecture. This work improves the study of HIs in adapting (Jayapriya et al. 2020) which was meant for classification of EWT-based HIs using deep learning. The proposed framework of classification outperforms to six previous models of analysis.

The paper’s remaining content is laid out as follows. Section 2 illustrates the suggested framework comprising of established methods, proposed mathematical models and details on the classification process modeled in the study of analysis. The HI dataset(s) used, experimental conditions, and the framework’s findings are all detailed in Section 3. Experimental analysis of the framework as outlined in section-3 exhibits effectively even with the limited training samples and learns to acquire discriminative hierarchical spectral-spatial features. Section 3.4 illustrates the
computational complexities of various algorithms used in the suggested framework and the previously implemented method (Jayapriya et al. 2020). The block diagram of the proposed PEC framework is shown in Fig. 4. Finally, Section 4 summarises the conclusions and discussions.

Proposed Framework

Due to the high dimensionality of the hyperspectral image data and the scarcity of training examples, HI applications struggle with class imbalance. To deal with such issues we propose a novel framework. The framework adapts three distinct techniques, as discussed above. In the first stage of the framework, the spectral dimension is lowered via principal component analysis (PCA). We elaborated the adapted dimensionality reduction process in Section 2.1. The data is then transformed using 1D-EWT to extract shallow features. Shallow feature extraction with EWT is the core methodology adapted in the framework and discussed in details in Section 2.2. Input frequencies suitable for deep hierarchical features extraction and classification, finally used in 3D-CNNs. 3D-CNN model-based deep feature extraction and classification methodologies are outlined in Section 2.3. The detailed description of the proposed framework is explained as below:

Dimensionality reduction

Let $\text{Img} \in \mathbb{R}^{M \times N \times d}$ represent the spectral-spatial HI cube, where $\text{Img}$ represents the initial input, $M, N$ and $d$ indicate the corresponding rows, columns and channels. Each HI pixel in the $\text{Img}$ cube has $d$ spectral values. The groundtruth for the $\text{Img}$ takes one of the values from the set of $C$ classes, $Y = y_1, y_2, ..., y_C$. The HI data, $\text{Img}$ is subjected to the conventional PCA along spectral bands as a means to reduce the spectral redundancy. The PCA retains the spatial dimension while decreasing the amount of spectral bands to $m$ where ($d \gg m$) (refer block diagram in Fig. 1). As the spatial content is crucial to identify any item and to preserve it, the spectral bands are just decreased. $I_p \in \mathbb{R}^{M \times N \times m}$ is the updated input after PCA, where $M, N$ and $m$ represent to width, height and quantity of spectral bands respectively. These values indicate PCA reduced data cube. Spectral Dimension reduction improves classification performance while reducing overfitting and computational expenses. We use $I_p$ to represent this dimensionally shrunk image. Further details on PCA and the underlying mathematics can be referred in Jonathon (2014).

Feature extraction with 1D-EWT

In the signal processing domain, the Empirical Wavelet Transform (EWT) proposed by Gilles (2013) has been widely adapted for numerous applications. During 2014, Gilles et al. (2014). further expanded the model to get utilise the EWT model exclusively for visual processing applications. The seminal work presented by Prabhakar and Geetha (2017); Jayapriya et al. (2020) adapts the EWT for classification of Hyperspectral data.

Building a wavelet filter bank based on Fourier supports that are discovered in the processed signal spectrum is the fundamental concept of EWT. The empirical wavelets are dilated variations of a single parent wavelet in the temporal domain, just like the traditional wavelets. Empirical detection of the corresponding dilation components instead than adhering to a predetermined scheme is the novel aspect of the empirical wavelet. The crux of EWT is adaptive Fourier spectrum segmentation of the original signal, followed by filtering. A set of wavelets is created using the 1D-EWT based on arbitrary divisions of the 1D Fourier domain. There are $N$ sequential intervals (different size bandwidths) between 0 and $\pi$, $\omega_0 = 0$ and $\omega_N = \pi$. The remaining $(N-1)$ limits are found by looking for the local maximum spectral values. Set $M$ as the overall number of maxima; if $M$ is greater than or equal to $N$, the first $N$ maxima are kept; if $M$ is less than $N$, all maxima are kept and $N$ is revised. The intermediate frequency is therefore determined as $\omega_n (n = 1,2,3,....,N-1)$ between two local maxima (Lopez-Gutierrez 2022). The detailed illustration of the transformation process can be found in Lopez-Gutierrez (2022); Mohapatra et al. (2022).

The aforementioned components (detail and approximation) is utilised to recover the original signal because they correspond to the various frequency components/modes of EWT. Each mode of the EWT is retrieved and used as a feature for additional
investigations. The modes are arranged by frequency from low to high. In the proposed framework, 1-D EWT transform, with \( N_e = 2 \) is applied on each band of \( I_p \) giving rise to \( I_{p, \text{ewt}} \) where \( N_e \) represents the number of modes. After being formed, modes are layered, therefore if \( N_e \) is the number of modes, stacks containing modes, \((2^{N_e-1})\) combinations will exist. As per the observations by Prabhakar and Geetha (2017), these stacks find suitability to get used as features for HI classification. It is next to select the first \( K \) spectral bands of \( I_{p, \text{ewt}} \). In the experimental analysis we select and focus \( K \) as 30 for IP and 15 for rest datasets and as per the conclusions derived from Roy et al. (2020) for selecting number of Principal Components for Hyperspectral classification. As a means to use image classification methods, the selected bands of the transformed Image, \( I_{p, \text{ewt}} \) is split into overlapping tiny 3D chunks. The label of centered pixel derives the truth labels of the image. Further, in using \( I_{p, \text{ewt}} \) 3-D other adjacent patches \( P \in R(S \times S \times K) \) were derived. They encompass the \( S \times S \) window or spatial extent as well as all \( K \) channels.

**Deep feature extraction and classification with 3D-CNN**

The use of convolutional neural networks (CNNs) has demonstrated promising results for a number of computer vision and image classification tasks. The CNN offers the advantage of being able to extract spatial features from the data with its kernel, which is not possible with other networks. With CNNs, edges, colour distributions, etc. can be detected in an image, making them particularly robust for recognizing and classifying images.

CNNs can be categorized based on the dimension of convolutional kernel that is utilized. In 1D-CNN, kernel slides along one dimension. To predict a single slice HI map, 2D-CNN use two dimension convolutional kernels. As a result of the fact that 2D convolutional kernels take one slice as input, they are inherently unable to leverage context from adjacent slices. Since they can leverage context across the height and width of a slice, 2D CNNs cannot leverage context from adjacent slices. A 3D-CNN addresses these issues by predicting segmentation for a volumetric patch in a scan using 3D convolutional kernels. Moreover, this is due to the volumetric nature of HI data and the presence of spectral dimension, 2D CNN is found to be restrictive to successfully extract key features from spectral channels. As per the studies of Ji et al. (2013); Fırat et al. (2022), 3D-CNN is capable enough to deal with both spectral and spatial characteristics as compared to 2D-CNN which simply takes spatial information only. Because these CNNs use a greater number of parameters, using interslice context can improve performance, but at the cost of computational complexity. With these features, it motivated to get use the 3D-CNN. At the same time the dimension reduction methodology adopted in our framework tackles to the challenges of computational complexity being inherited in 3D-CNN classification model.

We refer the schematic representation of 3D-CNN model in Fig. 2 being adapted for human action recognition applications (Ji et al. 2013). The model uses a three dimensional filter which convolves with three dimensional volumetric data.

The schematic representation of 3D-CNN classification model adapted in our framework is displayed in Fig. 3. The suggested deep learning model of PEC framework has two 3D-CNNs, two dropout layers, a flatten layer and three dense layers. Here, the 3D convolution generates a feature map based on spatio-temporal information of the input volumetric sample. High-level features in the data are represented in the convolutional layer output. Fully-connected layer learns the non-linear combinations using these features. Dropouts are the regularization technique being used to prevent overfitting in this model. A specific proportion of the network’s neurons are switched at random when dropouts are introduced. The neurons’ input and output connections are also disabled. This improves the model’s feature learning capability.

**Proposed PEC Framework for HI Classification**

The road map of the proposed PEC framework is shown in Fig. 4. The framework adapts three distinct techniques, as
discussed above. In the first stage of the framework, the spectral dimension is lowered via principal component analysis (PCA). We elaborated the adapted dimensionality reduction process in Section 2.1. The data is then transformed using 1D-EWT to extract shallow features. Shallow feature extraction with EWT is the core methodology adapted in the framework and discussed in detailed in Section 2.2. Input frequencies suitable for deep hierarchical features extraction and classification finally used in 3D-CNNs. 3D-CNN model-based deep feature extraction and classification methodologies are outlined in Section 2.3.

**PCA Modeling**

Let $M, N, d$ be the rows, columns and channels or spectral values of image cube, $R^{M \times N \times d}$.

Let $A \in R^{d}$ and $L_{N}$ be all N-dimensional linear subspace. Thus, $N^{th}$ principal subspace ($L_{N}$) can be defined as:

$$l_{N} = \arg\min_{b \in L_{N}} E(\min_{b \in L_{N}} \| A - B \|^2)$$  

where $\hat{A} = A - \mu$ and $\mu = E(A)$

$$X_{\text{Reduced-Dimension}}T_{N}(A) = \mu + \Pi_{l_{N}}A$$  

where $\Pi_{l_{N}}A$ is the projection of A onto $L_{N}$

Let $\sum = E(A - \mu)(A - \mu)^{T}$ be the covariance matrix where $\mu = E(x)$

Let $\lambda_{1} \geq \lambda_{2} \geq \lambda_{3} \geq \ldots \geq \lambda_{d}$ be the ordered eigen values of $\sum$ and

Let $V = \{v_1, v_2, \ldots, v_d\}$ be the corresponding eigenvector set and D be diagonal matrix with $D_{jj} = \lambda_{j}$, then the spectral decomposition of $\sum$ can be defined as:

$$I_{p} = VDV^{T} = \sum_{n=1}^{d}\lambda_{n}v_{n}v_{n}^{T}$$

where $\sum_{n=1}^{d}I_{p} \ll d$ and $Image_{i} \in R^{M \times N \times d}$ w.r.t set of groundtruth HIs

**1D-EWT Modeling**

Let $U_{s}(n, t)$ be the detail coefficient where:

$$U_{s}(n, t) = \langle v, \psi_{n} \rangle = \int v(T)\psi_{n}(T-t)dT = (\hat{v}(\omega)\psi_{n}(\omega))^{-1}$$  

(4)

Let $U_{s}(0, t)$ be the approximation coefficient where:

$$U_{s}(0, t) = \langle v, \phi_{1} \rangle = \int v(T)\phi_{1}(T-t)dT = (\hat{v}(\omega)\phi_{1}(\omega))^{-1}$$  

(5)

Hence, $I_{p,\text{ewt}} = U_{s}(0, t)$, where $N_{c} = 2$, on $I_{p}$ and $|\text{bands}| = 2^{N_{c}-1}$

**3D-CNN Modeling**

The activation value $d_{i,j,k}^{L}$ at output position $(i,j,k)$ of the $L^{th}$ filter when convolving with the input HI, $X$ is defined as:

$$d_{i,j,k}^{L} = F_{RLA}\left( \sum_{m=i}^{i+(F-1)} \sum_{c} \sum_{d=1}^{d} \sum_{c=1}^{c} w_{c,m,n,d}^{L}X_{c,m,n,d} + b_{k} \right)$$

(6)

where RLA is the Rectified Linear Activation Function and $F = \begin{cases} p & \text{if } p \geq 0 \\ 0 & \text{if } p < 0 \end{cases}$

$$Map_{\text{predicted}} = d_{i,j,k}^{L}$$  

(7)

Figure 4 is the empirical representation of our proposed framework.

Algorithm-1 represents the step-wise operational approach of the proposed PEC framework for hyperspectral image classification which in turn adapts above mathematical modeling of PCA, 1D-EWT and 3D-CNN.

**Experimental Setup and Outcomes**

To experiment the undertaken study and to validate the proposed PEC framework, an extensive testing is done using Google Colab Pro online platform with Keras Deep learning framework and allied Python libraries on three HI datasets,
Indian Pines (IP), Salinas Scene (SA) and University of Pavia (PU) datasets.

The performance analysis on the classification is evaluated with three accuracy metrics, overall accuracy (OA), average accuracy (AA), and kappa coefficient (k). We refer the evaluating parameters, as follows.

\[
OA = \frac{\text{Sum of correctly classified values}}{\text{Total number of values}}
\]

\[
AA = \frac{\text{Sum of accuracy for each class predicted}}{\text{Number of class}}
\]

In other words, the average accuracy can be termed as the average of each accuracy per individual class.

\[
k = \frac{N\sum_{i=1}^{n}m_{i,i} - \sum_{i=1}^{n}(G_{i} \times C_{i})}{N^{2} - \sum_{i=1}^{n}(G_{i} \times C_{i})}
\]

where, \(i\) denotes the class number, \(N\) represents total classified values compared to truth values, \(m_{i,i}\) are values correctly classified and located on diagonals of error matrix. \(C_{i}, G_{i}\) are the total number of predicted values and total true values of the class \(i\), respectively.

**Dataset description**

The first dataset in our experimental analysis i.e., Indian Pines (IP) dataset recorded in western portion of Indiana during 1995, between 0.4 and 2.5 \(\mu m\) in wavelength. It has 200 spectral bands (excluding 20 water absorbing spectral bands) and a \(145 \times 145\) spatial dimension. There are sixteen classes in total. The groundtruth of IP is shown in Fig. 5(a).

The second dataset of our experiment comprises with 610 \(\times\) 340 pixel University of Pavia (PU). There are nine classes in total. The image contains 103 spectral bands (wave length range 430-860nm) and spatial resolution of 1.3m. The groundtruth of PU is shown in Fig. 5(b).

The third dataset, The Salinas dataset (SA) represents to the Salinas Valley in California. The spatial size \(512 \times 217\) contains 204 spectral bands. This is a high-resolution dataset with sixteen distinct classes and the capturing distance was around 3.7 metres. The groundtruth of SA is shown in Fig. 5(c).

**Experimental settings**

In the experimental settings, we used PCA algorithm of Scikit-learn library Pedregosa et al. (2011) to reserve first \(m\) PCs of input image \(Img\), to produce the dimensionally reduced image, \(I_p\). For the above three datasets IP, PU and SA, the values of \(m\) are set to 30, 15, and 15 respectively in order to minimise computing costs. The whiten parameter of the PCA is set to true. Whiten parameter aids in raising forecasting accuracy. For feature extraction in 1D-EWT, the EWT1D() method of the Python package ewtpy (Carvalho et al. 2020) is applied to individual spectral bands of the \(I_p\).

Since a higher number decomposition level results to a high computations, the number of modes \(N_e\) is limited to 2. The parameters are set to 10 for lengthFilter and 5 for sigmaFilter. The \((2^{N_e} - 1)\) coefficient feature matrices produced by the 1-D EWT procedure each have the same spatial dimensions as the \(I_p\). We choose the first \(K\) high frequency components of \(I_{p,ewt}\)’s spectral bands. The selected bands of the transformed image, \(I_{p,ewt}\), are extracted into small, overlapping 3-D patches whose truth labels are determined by the label of the centre pixel in order to employ image classification techniques. Our 3-D neighbouring patches \(P \in R^{S \times S \times K}\) from \(I_{p,ewt}\) have been...
generated, that span over the whole of $K$ spectral bands as well as the $S \times S$ spatial breadth.

As discussed, the proposed Deep learning model consists of two 3D-Convolutional layers. Four filters of dimension $3 \times 3 \times 7$ are present in the first 3D convolution layer. Eight filters of dimension $3 \times 3 \times 3$ are present in the second 3D convolution layer. Both the 3D convolution layers employ ReLU activation. The first 3D convolution layer receives the input samples and extracts the features maps using different numbers of kernels. These features are fed to second 3D convolution layers. Feature maps from the previous layer are re-passed to a dropout layer at a 40% dropout rate and flattened to feed the first fully connected layer (Dense). Output features are once again fed into second fully connected layer (Dense) comprising 128 neurons with ReLU activation followed by second dropout layer with a rate of 40%. Finally, the model obtains the expected output from the final fully connected layer using the softmax activation function. Random batches with a size of 1024 are employed for training. Adam with 0.01 learning rate and 1e-4 decay is used as a batch optimizer while the Categorical cross-entropy loss function is employed as the loss function. 100 epochs have been used for training.

(a) Groundtruth of IP  
(b) Groundtruth of PU  
(c) Groundtruth of SA

Fig. 5 Groundtruth of IP, PU, and SA dataset
**Classification outcomes and analysis**

In Table 1 we present the experimental outcomes of PEC framework with thirty percent and ten percent training samples in respect to the undertaken datasets (IP, PU and SA).

Figure 6 portrays the PEC framework based predicted images for undertaken datasets with thirty and ten percent training along with their groundtruth(s).

In Table 2 we compare the findings of Jayapriya et al. (2020) with those of the proposed PEC framework.

---

**Table 1 Classification outcomes of PEC framework**

| Dataset | Accuracy | Training (30%) | Training (10%) |
|---------|----------|----------------|----------------|
| IP      | OA(%)    | 99.58          | 96.35          |
|         | AA(%)    | 99.68          | 95.74          |
|         | K × 100  | 99.51          | 95.83          |
|         | OA(%)    | 99.94          | 99.93          |
| PU      | AA(%)    | 99.89          | 98.54          |
|         | K × 100  | 99.92          | 99.19          |
|         | OA(%)    | 99.99          | 99.96          |
| SA      | AA(%)    | 99.99          | 99.97          |
|         | K × 100  | 99.99          | 99.96          |

**Table 2 Comparison of the PEC framework**

| Dataset | Accuracy Metric | EWT-CNN-RF | EWT-CNN-ELM | EWT-CNN-MSVM | PEC |
|---------|-----------------|------------|-------------|--------------|-----|
| IP      | OA              | 95.3       | 100         | 99.5         | 99.58 |
|         | K × 100         | 94.9       | 98.2        | 98           | 99.51 |
| PU      | OA              | 95.3       | 100         | 99.4         | 99.94 |
|         | K × 100         | 94.9       | 98.5        | 98.4         | 99.92 |

---

The method suggested by Jayapriya et al. (2020) uses EWT for shallow feature extraction, 2D CNN for deep feature extraction, and associated three classification algorithms, Random forest, Multi-SVM, and ELM. In contrast, the proposed PEC framework uses 3D Convolutional layers along with 1D-EWT and PCA for the extraction and classification of deep spectral and spatial features. Therefore the PEC framework is not too expensive computationally. In Table 3, the classification effectiveness of the suggested approach is compared to that of several techniques, such as SSRN Zhong et al. (2018), 2D-CNN Makantasis et al. (2015), 3D-CNN Ben Hamida et al. (2018), and Multi-scale 3D-CNN He et al. (2017).

We cite some of the technical validity and development of the approach in reference to the developments of allied approaches. The values for the competing approaches in the

---

**Fig. 6** Predicted maps of PEC classification framework: IP, PU & SA

(a) IP: with 10% training
(b) PU with 10% training
(c) SA with 30% training
(d) SA with 10% training
Earth Science Informatics (2022) 15:2289–2301

Table 3 Classification outcomes of three Datasets with 30% training samples

| Accuracy Metric | 2D-CNN | 3DCNN | M3D-CNN | SSRN | PEC |
|-----------------|--------|-------|---------|------|-----|
| IP Dataset      |        |       |         |      |     |
| OA(%)           | 89.48  | 91.10 | 95.32   | 99.19| 99.58|
| AA(%)           | 86.14  | 91.58 | 96.41   | 98.93| 99.68|
| $\kappa \times 100$ | 87.96 | 89.98 | 94.7    | 99.07| 99.51|
| PU Dataset      |        |       |         |      |     |
| OA(%)           | 97.86  | 96.53 | 95.75   | 99.90| 99.94|
| AA(%)           | 96.55  | 97.57 | 95.08   | 99.91| 99.86|
| $\kappa \times 100$ | 97.16 | 95.51 | 94.50   | 99.87| 99.91|
| SA Dataset      |        |       |         |      |     |
| OA(%)           | 97.38  | 93.96 | 94.79   | 99.98| 99.99|
| AA(%)           | 98.84  | 97.01 | 96.25   | 99.97| 99.99|
| $\kappa \times 100$ | 97.08 | 93.32 | 94.20   | 99.97| 99.99|

The aforementioned tables are obtained in line with the data of Roy et al. (2020). It may be noted here that, the SSRN is a supervised deep learning framework (Zhong et al. 2018), that decreases the issue of diminishing accuracy shown in earlier deep learning model. Particularly, the back propagation of gradients is facilitated by the residual blocks, which link up every other 3-D convolutional layer via identity mapping. In order to extract more discriminative features, the SSRN treats spectral characteristics and spatial information in two independent blocks. To further regularise the learning process and enhance the classification accuracy of trained models, SSRN model mandates batch normalisation on each convolutional layer.

Similarly, as per Makantasis et al. (2015), 2D-CNN based classification technique that auto-hierarchically produces high-level features from HI(s). This approach uses multi-layer perceptron to carry out the classification task and convolutional neural network to encode the spectral and spatial information of the pixels. In Ben Hamida et al. (2018), authors investigated how well the DL-architecture perform in classifying the RS hyperspectral dataset, and developed a novel 3D Deep Learning strategy which caters to process both spectral and spatial information simultaneously. Some of suggested models achieve better classification rate than state-of-the-art methods with reduced processing costs. A Multiscale 3D deep convolutional neural network (M3D-DCNN) suggested by He et al. (2017) not only improves the classification but also promises to attain large-scale datasets by jointly learning both 2D Multi-scale spatial feature and 1D spectral feature from HI data in an end-to-end strategy. The model produced cutting-edge results on the standard datasets despite not using any hand-crafted features or pre- or post-processing like PCA, sparse coding, etc.

Table 4 Effect of number of frequency bands (K) on the Accuracy metrics

| Accuracy Metric | K = 20 | K = 30 | K = 40 | K = 50 | K = 60 |
|-----------------|--------|--------|--------|--------|--------|
| For 30% Training Samples:IP Dataset |        |        |        |        |        |
| OA(%)           | 99.65  | 99.69  | 99.57  | 99.54  | 99.73  |
| AA(%)           | 99.73  | 99.82  | 99.68  | 98.81  | 99.45  |
| $\kappa \times 100$ | 99.61 | 99.64  | 99.51  | 99.48  | 99.70  |
| For 10% Training Samples:IP Dataset |        |        |        |        |        |
| OA(%)           | 95.82  | 95.12  | 95.23  | 95.78  | 96.5   |
| AA(%)           | 93.83  | 93.91  | 93.58  | 93.68  | 95.64  |
| $\kappa \times 100$ | 95.23 | 94.42  | 94.54  | 95.17  | 96.00  |

The influential factor in accepting 30% training set for the framework is analysed in respect to three undertaken datasets. For simplicity, the effect of number of frequency bands (K) on accuracy metrics for one dataset (say, IP) is shown in the Table 4. Figure 7(a), (b) and (c) illustrate how accuracy metrics vary with different training samples for various frequency bands K.

Complexity Analysis

Table 5 is the self-explanatory representation on computational complexities of the associated methods used by competing models and suggested framework. $K_s$, $C$, and $L$ stand for kernel size and channels, respectively, in 3D CNN. $N$ stands for the input channel, $W$ for the image’s width, and $H$ for its height. $K_s$, $W$ and $H$ are recognised as width, height and input image in the 2D-CNN. Number of training samples are represented as $n$ in SVM and RF, whereas $p$ represents to number of available features. $n_{trees}$ indicate the number of trees. By exploitation random hidden weights, the ELM has a time complexity of $L$ and $N$, where $L$ is the number of hidden layer neurons and $N$ is the number of training data. In the next entry of Table 5, $N$ represents signal samples. $N_{mra}$ denotes the multiresolution analysis (MRA) or signal decomposition. The last row of the table represents computational complexity of PCA algorithm applied on an input matrix with $N \times D$ dimension.

The proposed PEC model using 1D-EWT for feature extraction has a computational complexity advantage of factor N over the 2D-EWT employed in Jayapriya et al. (2020). The classification approach used in Jayapriya et al. (2020) includes 2D CNNs and additional machine learning algorithms, including MSVMs, ELMs, and Random Forest, but the suggested PEC model uses only 3D CNNs for classification. As multiple classifiers have been suggested by Jayapriya et al. (2020) whereas the proposed PEC technique uses 3D CNNs, so exact comparison is not possible.
Conclusions and discussions

Experimental findings of the proposed PEC framework are contrasted with those of other cutting-edge work carried out by Jayapriya et al. (2020). The suggested PEC framework surpasses by producing better performance oriented classification accuracy like Overall Accuracy (OA) Alberg et al. (2004), Average Accuracy (AA) Fung and LeDrew (1988), and Kappa coefficient (k) Fung and LeDrew (1988). There is immense potential for enhancing classification efficiency by implementing the discussed approaches to enhance spatio-temporal feature learning.
hyperparameter tailoring. Besides the suggested classification model, customised deep learning models can also be included.

Author contributions Sandeep Kumar Ladi: Conceptualization, Writing - Original Draft, Software, Validation, Formal analysis, Investigation . G K Panda: Project administration, Supervision, Validation, Writing - Review & Editing. Ratnakar Dash: Conceptualization, Methodology, Supervision, Writing - Review & Editing.

Declarations

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

Ethics approval The authors ensure that accepted principles of ethical and professional conduct have been followed.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Alberg AJ, Park JW, Hager BW et al (2004) The use of “overall accuracy” to evaluate the validity of screening or diagnostic tests. J Gen Intern Med 19(5p1):460–465

Bai J, Xiang S, Shi L et al (2015) Semisupervised pair-wise band selection for hyperspectral images. IEEE J Select Topic Appl Earth Observ Remote Sens 8(6):2798–2813

Ben Hamida A, Benoit A, Lambert P et al (2018) 3-d deep learning approach for remote sensing image classification. IEEE Trans Geosci Remote Sens 56(8):4420–4434. https://doi.org/10.1109/TGRS.2018.2818945

Carvalho VR, Moraes MF, Braga AP et al (2020) Evaluating five different adaptive decomposition methods for eeg signal seizure detection and classification. Biomed Signal Proc Control 62(102):073. https://doi.org/10.1016/j.bspc.2020.102073

Chang CI, Du Q, Sun TL et al (1999) A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. IEEE Trans Geosci Remote Sens 37(6):2631–2641

Demir B, Ertürk S (2010) Empirical mode decomposition of hyperspectral images for support vector machine classification. IEEE Trans Geosci Remote Sens 48(11):4071–4084

Demir B, Erturk S (2008) Empirical mode decomposition pre-process for higher accuracy hyperspectral image classification. In: IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium, pp II–939–II–941. https://doi.org/10.1109/IGARSS.2008.4779150

Druzhkov P, Kustikova V (2016) A survey of deep learning methods and software tools for image classification and object detection. Pattern Recognit Image Anal 26(1):9–15

Elgamal T, Hefeeda M (2015) Analysis of pca algorithms in distributed environments. arxiv: https://arxiv.org/abs/1503.05214

Ertürk A, Gülli MK, Ertürk S (2012) Hyperspectral image classification using empirical mode decomposition with spectral gradient enhancement. IEEE Trans Geosci Remote Sens 51(5):2787–2798

Fauvel M, Benediktsson JA, Chanussot J et al (2008) Spectral and spatial classification of hyperspectral data using svms and morphological profiles. IEEE Trans Geosci Remote Sens 46(11):3804–3814

Fırat H, Asker ME, Hanbay D (2022) Classification of hyperspectral remote sensing images using different dimension reduction methods with 3d/2d cnn. Remote Sens Appl Soc Environ 25(100):694. https://doi.org/10.1016/j.rsase.2022.100694

Fukushima K, Miyake S (1982) Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets. Springer, Berlin, p 267–285

Fung T, LeDrew E (1988) For change detection using various accuracy. Photogramm Eng Remote Sens 54(10):1449–1454

Gilles J (2013) Empirical wavelet transform. IEEE Trans Signal Process 61(16):3999–4010

Gilles J, Tran G, Osher S (2014) 2d empirical transforms. wavelets, ridgelets, and curvelets revisited. SIAM J Imag Sci 7(1):157–186

Gormus ET, Canagarajah N, Achim A (2012) Dimensionality reduction of hyperspectral images using empirical mode decompositions and wavelets. IEEE J Select Topic Appl Earth Observ Remote Sens 5(6):1821–1830

Harsanyi JC, Chang CI (1994) Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. IEEE Trans Geosci Remote Sens 32(4):779–785

He M, Li B, Chen H (2017) Multi-scale 3d deep convolutional neural network for hyperspectral image classification. In: 2017 IEEE International Conference on Image Processing (ICIP), pp 3904–3908. https://doi.org/10.1109/ICIP.2017.8297014

Springer
Hsu PH, Tseng YH, Gong P (2006) Spectral feature extraction of hyperspectral images using wavelet transform. Journal of Aerial Surveying and Telemetry 11(1):93–109. https://doi.org/10.6574/JPRS.2006.11(1.8)

Pai-Hui Hsu, Yi-Hsing Tseng & Peng Gong (2002) Dimension Reduction of Hyperspectral Images for Classification Applications, Geographic Information Sciences, 8:1, 1-8. https://doi.org/10.1080/1082400209480567

Huang NE, Shen Z, Long SR et al (1998) The empirical mode decomposition and hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc London Ser A Math Phys Eng Sci 454(1971):903–995

Immovilli F, Bellini A, Rubini R et al (2010) Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison. IEEE Trans Indus Appl 46(4):1350–1359

Iosifidis A, Tefas A, Pitas I (2015) On the kernel extreme learning machine classifier. Pattern Recogn Lett 54:11–17. https://doi.org/10.1016/j.patrec.2014.12.003

Jayapriya K, Jacob IJ, Darney PE (2020) Hyperspectral image classification using multi-task feature leverage with multi-variant deep learning. Earth Science Informatics 13:1093–1102

Ji S, Xu W, Yang M et al (2013) 3d convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell 35(1):221–231

Jonathon S (2014) A tutorial on principal component analysis. arxiv. http://arxiv.org/abs/1404.1100

Jonnadula H, Kumar LS, Panda GK et al (2020) Hyperspectral image classification bi-dimensional empirical mode decomposition and deep residual networks. In: 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), pp 1–6. https://doi.org/10.1109/AISP48273.2020.9073241

Khan MJ, Khan HS, Yousaf A et al (2018) Modern trends in hyperspectral image analysis: A review. IEEE Access 6:14,118–14,129. https://doi.org/10.1109/ACCESS.2018.2812999

Kim Y (2014) Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. https://doi.org/10.3115/v1/D14-1181

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Pereira F, Burges C, Bottou L, et al (eds) Advances in Neural Information Processing Systems, vol 25. Cur-ran Associates, Inc., https://proceedings.neurips.cc/paper/2012/file/c399862d3b9db6b76c8436e924a68c45b-Paper.pdf

Li W, Du Q (2014) Gabor-filtering-based nearest regularized subspace for hyperspectral image classification. IEEE J Sel Topic Appl Earth Observ Remote Sens 7(4):1012–1022

Lopez-Gutierrez R, Rangel-Magdaleno J, Morales-Perez CJ et al (2022) Induction machine bearing fault detection using empirical wavelet transform. Shock Vibrat 2022:1–12. https://doi.org/10.1155/2022/6187912

Maji P, Mullins R (2018) On the reduction of computational complexity of deep convolutional neural networks. Entropy (Basel) 20(4):305

Makantasis K, Karantzalos K, Doulamis A et al (2015) Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp 4959–4962. https://doi.org/10.1109/IGARSS.2015.7326945

Mallat S (1999) A wavelet tour of signal processing. Elsevier, Amsterdam https://doi.org/10.1016/B978-0-12-374370-1.00001-8

Milyaev S, Laptev I (2017) Towards reliable object detection in noisy images. Pattern Recognit Image Anal 27(4):713–722

Mohapatra S, Kumar Pati G, Mishra M et al (2022) Gastrointestinal abnormality detection and classification using empirical wavelet transform and deep convolutional neural network from endoscopic images. Ain Shams Eng J:101942. https://doi.org/10.1016/j.asej.2022.101942

Paolletti M, Haut J, Plaza J et al (2019) Deep learning classifiers for hyperspectral imaging: A review. ISPRS J Photogram Remote Sens 158:279–317. https://doi.org/10.1016/j.isprsjprs.2019.09.006

Papakostas M, Giannakopoulos T, Makedon F et al (2016) Short-term recognition of human activities using convolutional neural networks. In: 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp 302–307. https://doi.org/10.1109/SITIS.2016.56

Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: Machine learning in python. J Mach Learn Res 12:2825–2830

Pesaresi M, Gerhardinger A, Kayikciire F (2008) A robust built-up area presence index by anisotropic rotation-invariant textural measure. IEEE J Select Topic Appl Earth Observ Remote Sens 1(3):180–192

Pittner S, Kamarthi S (1999) Feature extraction from wavelet coefficients for pattern recognition tasks. IEEE Trans Pattern Anal Mach Intell 21(1):83–88. https://doi.org/10.1109/34.745739

Prabhakar TN, Geetha P (2017) Two-dimensional empirical wavelet transform based supervised hyperspectral image classification. ISPRS J Photogramm Remote Sens 133:37–45. https://doi.org/10.1016/j.isprsjprs.2017.09.003

Rajan S, Ghosh J, Crawford MM (2008) An active learning approach to hyperspectral data classification. IEEE Trans Geosci Remote Sens 46(4):1231–1242

Ranjan R, Sankaranarayanan S, Bansal A et al (2018) Deep learning for understanding faces: Machines may be just as good, or better, than humans. IEEE Signal Proc Mag 35(1):66–83

Roy SK, Krishna G, Dubey SR et al (2020) Hybridsing: Exploring 3-d-2-d cnn feature hierarchy for hyperspectral image classification. IEEE Geosci Remote Sens Lett 17(2):277–281. https://doi.org/10.1109/LGRS.2019.2918719

Sandeep Kumar L, Panda G, Dash R et al (2022) A novel grey wolf optimisation based cnn classifier for hyperspectral image classification. Multimed Tools Appl 81:1–24. https://doi.org/10.1007/s11042-022-12628-2

Sun Y, Wang X, Tang X (2014) Deep learning face representation from predicting 10,000 classes. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1891–1898

Tan SY (2017) Developments in Hyperspectral Sensing, Springer International Publishing, Cham, pp 1137–1157. https://doi.org/10.1007/978-3-319-23386-4_101

Tan K, Li E, Du Q et al (2013) Hyperspectral image classification using band selection and morphological profiles. IEEE J Select Topic Appl Earth Observ Remote Sens 7(1):40–48

Wang T, Liang M, Li Ji et al (2014) Rolling element bearing fault diagnosis via fault characteristic order (fco) analysis. Mech Syst Signal Proc 45(1):139–153

www.thekerneltrip.com (2018) Computational complexity of machine learning algorithms. https://www.thekerneltrip.com/machine/learning/computational-complexity-learning-algorithms/. Online Accessed 9 Sep 2022

Xu Y, Du B, Zhang F et al (2018) Hyperspectral image classification via a random patches network. ISPRS J Photogram Remote Sens 142:344–357

Yudistira N (2017) Kurita T (2017) Gated spatio and temporal convolutional neural network for activity recognition: towards gated multimodal deep learning. EURASIP J Image Vid Proc 1:1–12
Zhong Z, Li J, Luo Z et al (2018) Spectral-spatial residual network for hyperspectral image classification: A 3-d deep learning framework. IEEE Trans Geosci Remote Sens 56(2):847–858. https://doi.org/10.1109/TGRS.2017.2755542

Zhou X, Gong W, Fu W et al (2017) Application of deep learning in object detection. In: 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), pp 631–634 https://doi.org/10.1109/ICIS.2017.7960069

Zhu C, Yang X (1998) Study of remote sensing image texture analysis and classification using wavelet. Int J Remote Sens 19(16):3197–3203

Authors and Affiliations

Sandeep Kumar Ladi1,2 · G K Panda3 · Ratnakar Dash4 · Pradeep Kumar Ladi5

1 Department of ECE, Biju Patnaik University of Technology, Odisha 769015 Rourkela, India
2 Department of EECE, GITAM Deemed to be University, 530045, A.P., Visakhapatnam, India
3 Department of CSE, MITS, 765017, Odisha, Rayagada, India
4 Department of CSE, NIT, 769008, Odisha, Rourkela, India
5 Department of CSE, School of Engineering and Technology, GIET University, 765022 Gunupur, India

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.