MACHINE LEARNING BASED CROP YIELD PREDICTION USING SPECTRAL IMAGES

Thesis

Submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

ALKHA MOHAN

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE - 575 025

January, 2022

DECLARATION

by the Ph.D. Research Scholar

I hereby declare that the Research Thesis entitled **Machine Learning Based Crop Yield Prediction Using Spectral Images** which is being submitted to the **National Institute of Technology Karnataka, Surathkal** in partial fulfilment of the requirements for the award of the Degree of **Doctor of Philosophy** in Department of Computer Science and Engineering is a bonafide report of the research work carried out by me. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.

Alkha Mohan, 177072 177CO002

Department of Computer Science and Engineering

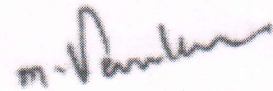Place: NITK, Surathkal.

Date: 10/01/2022

CERTIFICATE

This is to certify that the Research Thesis entitled **Machine Learning Based Crop Yield Prediction Using Spectral Images** submitted by **Alkha Mohan** (Register Number: 177072 177CO002) as the record of the research work carried out by her, is accepted as the Research Thesis submission in partial fulfilment of the requirements for the award of degree of **Doctor of Philosophy**.
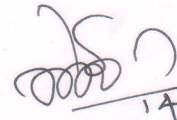
10-01-2022

Dr. M Venkatesan

Research Guide

(Signature with Date and Seal)

14.01.22

Chairman - DRPC

(Signature with Date and Seal)

Chairman
DUGC / DPGC / DRPC
Dept. of Computer Engg.
NITK - Surathkal
Srinivasnagar - 575 025

ACKNOWLEDGEMENTS

Expressing my gratitude is the finishing touch on my dissertation. Research transformed me as a better individual, not only in the technical domain but also on a personal level. Such improvements would not have been possible without the support of the people who have helped me.

First, I would like to express my sincere gratitude to my research supervisor Dr. M. Venkatesan for the continuous support, guidance and motivation throughout my Ph.D. study and related research. His vast knowledge, encouragement and patience definitely helped me to reach to this height. I thank him for identifying my capabilities and giving this opportunity to purse research under him. The amount of consideration that he given during the stages of research paper and thesis writing are impeccable. In fact I could not have imagined having a better advisor and mentor for my Ph.D study.

Now I would like to thank my Research Progress Committee (RPAC) members Prof. P Santhi Thilagam and Dr. P Jidesh for their insightful assessment and suggestions to further improve my work. The points raised by them were extremely valuable and helped me to look at my work in a different angle. I wish to show my appreciation to the Head of the Department Dr. Shashidhar G Koolagudi and other faculty members for their continuing support. I thank all the teaching and non-teaching staff of Computer Science Department, NITK for their help during my research period.

I will always cherish the kind of affection, care and support received from my friends Swathi, Nitesh Naik, Victor Daniel, Aadhirashree, Minimol, Uma Priya and Rashmi. Now I thank my fellow lab mates and group members for the stimulating discussions, knowledge transfer and team work. I wish to extend my special thanks to all the teachers who taught me in schools and colleges. Without them I wouldn't be in this position to write this. In particular, I am grateful to Dr. Binu VP from CE

karunagappally for enlightening me the first glance of research.

Last but not the least, I would like to thank my father Mr. S Mohan, mother N Ambika and brother Alvith for their love and guidance throughout my life. My family had to go through a lot of sacrifices for my studies and achievements and words cannot express how hard those were and how much I am grateful for those sacrifices. I would like to express my gratitude towards my in-laws too for their support and motivation during my studies and in my life. Most importantly, I wish to thank my loving and supportive husband Jayakrishnan who provide unending support and inspiration. Finally I praise the almighty for giving me the vision and strength to continue with confidence. Thank you God.

Alkha Mohan

ABSTRACT

The socio-economic stability of a country heavily dependent on its agricultural outputs. Therefore, each country needs to monitor and maintain agricultural outcomes at an adequate level. The early prediction of crop yield helps the farmers adopt necessary changes in cultivation on a season and ensure food security. The crop yield depends on several parameters, such as vegetation parameters, climatic parameters, soil condition, etc. Spatial and temporal analysis of cropland is necessary for the accurate prediction of yield. The data for such analysis were collected with the help of regular field surveys. Such surveys required more human resources and lack accuracy due to the interpolation method adopted to map the readings to a larger geographical area. The advancement in satellite imaging techniques helps gather temporal data of broad geographical regions with less workforce.

Usage of multispectral sensors in remote sensing helped in accurate discrimination of land objects and vegetations. The higher number of contiguous bands in hyperspectral images(HSI) improve the reconstruction of spectral signature and thereby increase the discrimination power. However, the higher dimensionality nature of HSI increases the computational complexity and leads to the Hughes phenomenon. The evolution of deep learning techniques made a significant impact on HSI classification. Several HSI processing applications rely on various Convolutional Neural Network (CNN) models. Therefore most of the CNN models perform dimensionality reduction (DR) as a preprocessing step. Another challenge in HSI classification is the consideration of both spatial and spectral features for obtaining accurate results. A few 3-D-CNN models are designed to overcome this challenge, but it takes more execution time than other methods. This research work proposes a multiscale spatio-spectral feature-based hybrid CNN model for hyperspectral image classification. Hybrid DR used for optimal band extraction, which performs linear Gaussian Random Projection (GRP) and nonlinear Kernel Principal Component Analysis (KPCA). A novel crop yield prediction model for the Paddy from Moderate Resolution Imaging Spectroradiometer (MODIS)

data and climatic parameters is introduced in this research work. Various vegetation indices (VI) are collected from MODIS data for the crop's entire life cycle. The proposed Temporal Convolutional network (TCN) with a specially designed dilated convolution module predicts the rice crop yield from vegetation indices and climatic parameters. The causal property of TCN and dilated convolution contribute to the multivariate time-based analysis of the crop and results in better performance.

CONTENTS

5 TCN Based Crop Yield Prediction   **81**

6 Conclusions and future scope   **101**

Bibliography   **104**

Publications   **120**

LIST OF FIGURES

LIST OF TABLES

ix

LIST OF ABBREVIATIONS

Abbreviations	Expansion
AA	Average Accuracy
AUC	Area Under Curve
AVHRR	Advanced Very High-Resolution Radiometer
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DNN	Deep Neural Networks
DR	Dimensionality Reduction
EVI	Enhanced Vegetation Index
GPP	Gross Primary Production
GRP	Gaussian Random Projection
HSI	Hyperspectral Images
ICA	Independent Component Analysis
IG	Information Gain
IP	Indian Pines Dataset
ISRO	Indian Space Research Organization
K	Kappa Statistic
KPCA	Kernel PCA
LAI	Leaf Area Index
LDA	Linear Discriminant Analysis
LLE	Locally Linear Embedding
LSTM	Long-Short Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MLR	Multiple Linear Regression
MNF	Minimum Noise Fraction
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
OA	Overall Accuracy
PA	Precision Agriculture
PAR	Photosynthetically Active Radiation
PCA	Principal Component Analysis
PU	Pavia University Dataset
ReLU	Rectified Linear Unit

Abbreviations	Expansion
RG	Relative Greenness
RMSE	Root Mean Square Error
RP	Random Projection
SA	Salinas Dataset
SDE	Sparse Discriminant Embedding
SGDA	Sparse Graph-Based Discriminant Analysis
SRP	Sparse Random Projection
SSNG	Spatial Spectral Neighborhood Graph
SSRN	Spectral-Spatial Residual Network
SSUN	Spectral-Spatial Unified Network
SVM	Support Vector Machine
TCN	Temporal Convolutional Network
TM	Thematic Mapper
VI	Vegetation Indices

CHAPTER 1

INTRODUCTION

Precision Agriculture (PA) was introduced to improve crops' production and ensure food security worldwide. According to the National research council (Council 1997), PA is defined as a decision making and operational process that manages crop production by processing and analyzing spatial and temporal multisource data of higher resolution with the help of modern information technology. Accurate yield prediction is a significant research field in precision agriculture. One of the challenges in crop yield prediction is finding the relationship of crop yield with meteorological and vegetation data. Climate, soil, fertilizer, crop quality, and the plant's health are some of the factors affecting the yield (Xu et al. 2019). Spatial and temporal data analysis is required to monitor these parameters throughout the crop cycle. However, the usage of remote sensing addressed the challenges of data collection and analysis in yield prediction.

When light hits on the earth's surface, a part of the light may get scattered, reflected, emitted, absorbed, or part of absorbed light emitted from the surface later. The reflection of energy from any material has a unique footprint. It is known as the material's spectral signature, which can identify different materials on the earth's surface. This spectral signature helps in yield prediction from remote sensing data (Fonseca et al. 2009). Spectral images having more bands can be effectively utilized for crop monitoring. Multispectral images are mostly captured in visible and infrared frequencies. These frequency bands are used to calculate vegetation indices such as normalized difference vegetation index (NDVI), leaf area index (LAI), photosynthetically active ra-

diation (PAR), etc. The contiguous spectrum in hyperspectral images (HSI) provides more discriminative power in crop classification. However, analysis of hyperspectral images needs to overcome two significant challenges; computational complexity due to its large volume and curse of dimensionality due to lack of input samples. Basic dimensionality reduction (DR) techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are used for band extraction in HSI. Hyperspectral images are nonlinear, and the application of these linear reduction techniques fails in their processing(Huang et al. 2015). Consideration of both spectral and spatial features in HSI classification also improves accuracy in crop discrimination. 3D-Convolutional Neural Network(CNN) designs help retain both features, while the problem is its computational complexity and proper window size selection. Continuous monitoring of cropland and finding the correlation between each week's vegetation and climatic data is necessary for accurate crop prediction. Existing prediction models fail to carry out this multivariate time-based data analysis.

This thesis introduced and developed novel techniques to overcome the challenges in HSI processing and yield prediction. The work proposes a nonlinear DR technique for optimal band extraction, a multiscale hybrid CNN for crop classification, and a temporal convolutional network (TCN) with a modified dilated convolution module for accurate yield prediction.

1.1 HYPERSPECTRAL IMAGE BAND EXTRACTION

Multispectral images are having around 3-10 spectral components captured beyond the visible spectrum. As the number of bands in multispectral images increases, the bandwidth decreases, and reconstruction of the complete spectral signature becomes easy. Thus researchers started the use of hyperspectral images for crop monitoring(Thenkabail P.S., Smith R.B. 2000), change detection(Liu et al. 2017)(Bruzzone et al. 2016), weather prediction, etc. due to its spectral-spatial features and a large number of bands. The visual representation of HSI and Multispectral Images are given in figure 1.1. Hyperspectral images are contiguous band images that look like a 3D image cube. Two dimensions in this cube are spatial coordinates, and the third dimension is

spectral bandwidth. Thus a single hyperspectral image is the collection of hundreds of images captured in different spectral bandwidth. The contiguous spectral property of these images helps to construct the spectral signature and discriminate objects efficiently. Even though many bands improve the material discrimination, it leads to the curse of dimensionality due to the less number of labeled samples and redundancy in bands(Theodoridis and Koutroumbas 2006). Therefore dimensionality reduction in hyperspectral images is essential to improve the classification accuracy and reduce the computational complexity. There are two DR methods for hyperspectral images named band selection and band extraction, similar to feature selection and feature extraction.
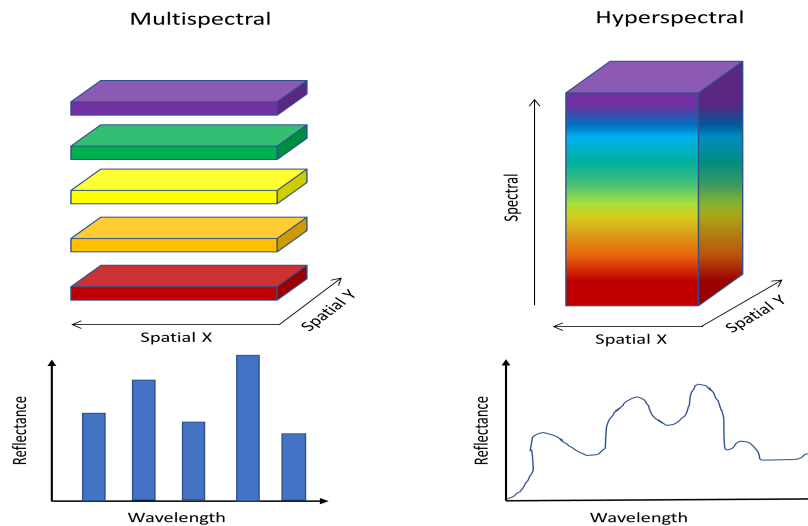
Figure 1.1: Visualization of Multispectral and Hyperspectral Images

The spectral difference between nearby bands in hyperspectral images is significantly less. Therefore some bands are precisely similar to their neighboring bands. Band selection techniques find redundancy in bands and select unique bands from the input image (Datta et al. 2014). These bands have the same input image property, whereas band extraction converts the input image into a lower-dimensional space to avoid higher dimensionality issues. Band extraction methods are more renowned than band selection techniques. Band extraction methods are of two types, namely linear and nonlinear DR (manifold learning) techniques(Datta et al. 2014). Linear dimensionality reduction techniques produce a linear low dimensional mapping of high dimensional input data. Commonly used linear band extraction methods are principal component

analysis (PCA)(Prasad and Bruce 2008)(Martínez and Kak 2001), linear discriminant analysis (LDA)(Martínez and Kak 2001), minimum noise fraction (MNF)(Luo et al. 2016), etc. Most of the linear DR techniques preserve up to second-order statistics of input data without considering the higher-order statistics. Natural images are nonlinear, and linear reduction techniques fail in their processing. Later, researchers started using nonlinear band extraction techniques such as KPCA(Datta et al. 2017)(Bachmann et al. 2005), nonlinear independent component analysis (ICA)(Wang and Chang 2006), LLE(Roweis and Saul 2000)(Bachmann et al. 2005), Isomap(Bachmann et al. 2005), etc. in hyperspectral images. These methods serve the nonlinearity of natural images and preserve the higher-order statistics. However, the major problems of these nonlinear techniques are their computational complexity.

It is essential to model a computationally feasible band extraction technique for HSI. Extracted bands are fed into a classifier to evaluate performance: the classification accuracy, the area under curve (AUC), and time complexity were compared.

1.2 HYPERSPECTRAL IMAGE CLASSIFICATION

The feature extraction in HSI classification was performed either with the help of self-learning classifiers or manually. The classification accuracy of traditional classifiers such as SVM(Melgani and Bruzzone 2004), Random Forest(Ham et al. 2005) is significantly less because it considers only spectral features for classification. Considering both spatial and spectral features in the modified version of SVM named SVM-CK(Camps-Valls et al. 2006)(Fauvel et al. 2012), Markov random field (Tarabalka et al. 2010), etc. improves the classification accuracy. However, most of the methods mentioned above extracted features manually, and it needs an expert's help. The progress in deep learning techniques moves the hyperspectral image classification into a new phase. The training phase of the deep learning method extract features automatically and uses them in the classification step.

The preprocessing technique and classification model has a significant role in classification accuracy and computational complexity. Most of the HSI classification models use PCA as a preprocessing technique, hence nonlinear features of the image are lost.

Another major problem in HSI classification is considering both spectral and spatial features. The window size selection for 3D-CNN is challenging. If the window size is large, it considers more local similarity, but the presence of noise may lead to misclassification, and small window size-based feature extraction lacks correlated features.

A nonlinear dimensionality reduction is a mandatory preprocessing for HSI classification(as explained in 1.1) to reduce the computational complexity in classification. Introduction of Multiscale spatial and spectral feature extraction with different window sizes can reduce the misclassification and retain correlated features. The performance of HSI classification models is compared using the evaluation parameters such as Overall Accuracy(OA), Average Accuracy(AA), and Kappa Statistic(K).

1.3 CROP YIELD PREDICTION

Machine learning (ML) algorithms have gained much attention in recent years and are used for crop yield prediction. ML techniques analyze climatic and vegetative data to decide the type of crop to cultivate, time of cultivation, and management of the crop. Machine learning-based prediction models trained using prior data aid the prediction of yield for current conditions. Although there are many machine learning-based prediction models, the most commonly used techniques in yield predictions are decision trees (Shekoofa et al. 2014), random forest (Jeong et al. 2016), regression models (Johnson et al. 2016) and neural networks (Bose et al. 2016).

Manual collection of data through field surveys is tedious and often leads to inaccurate prediction of yield. However, remote sensing data usage helps to extract the features for a wide area with less overhead. Vegetation indices and climatic data have an equal and vital role in yield prediction. The use of remote sensing data in most research works helped to identify the vegetation parameters accurately. Still, proper vegetation index selection and identification of their dependency in yield remain tricky. Recent researches in yield prediction focused mainly on deep learning techniques, especially CNN and Long-Short Term Memory (LSTM) based prediction models, due to their less prediction error. However, current CNN-based prediction designs lack accurate multivariate time-based analysis. The LSTM models overcome these limitations and work

for time-based analysis. However, to analyze the data for a considerable period, the LSTM model has to increase its receptive field; this leads to higher time complexity.

The Development of a multivariate time-based prediction model that accounts for climatic and vegetation parameters is mandatory for accurate crop yield forecasting. The prediction results can be compared using the coefficient of determination(R^2), mean absolute error (MAE), and root mean square error (RMSE). A better prediction model must exhibit higher R^2 and lesser MAE and RMSE.

1.4 MOTIVATION

Precession farming analyses the environmental and climatic parameters to apply agricultural inputs to improve crop productivity and ensure food security. The early prediction of crop yield plays a crucial role in improving productivity in a season. Such prediction techniques can guide farmers to apply appropriate fertilizers, pesticides, and required irrigation to overcome crop damage and yield drop. The advancement made in remote sensing has its impact in aiding precision agriculture. Although such developments are made in science, precision agriculture is still in its infant stage and needs studies and research to improve the techniques further. This is one of the motivations to carry out an extensive study on precision agriculture with remote sensing data.

The deep study on literature survey enlightens that the impact of hyperspectral remote sensing for vegetation analysis is tremendous. The presence of multiple spectral bands in HSI correctly identifies the crop area, which is vital for accurate vegetation classification and yield prediction. However, the curse of dimensionality in HSI is a backdrop in processing them, and this motivates to introduce a feature preserving high efficient dimensionality reduction technique for HSI.

Accurate HSI classification requires the consideration of both spatial and spectral features. The introduction of deep learning models eases such complex feature extraction. However, the existing deep learning classification models fail in multi-scale spatio-spectral feature extraction and result in high computational complexity. These shortcomings motivate the design of a novel multi-scale hybrid CNN that works even for lesser training samples with minimal run time.

6

In 2018 Indian Space Research Organization (ISRO) launched a new satellite named HysIS equipped with high-resolution hyperspectral sensors aimed for earth surface study. Although the availability of data from HysIS is strictly for internal research and defense purposes, once the data is in the public research domain, the above DR and classification techniques make a huge impact.

Crop yield is heavily dependent on both climatic and vegetation parameters. Continuous monitoring of such parameters from the sowing to the harvesting stage is mandatory for accurate yield forecast. The existing prediction models fail in such multi-variate time-based analysis, and this motivates to find an efficient prediction model that performs a multi-variate time-based analysis.

1.5 APPLICATIONS

The applications of spectral remote sensing are vast and applicable to multiple areas of science. Spectral images are widely used in atmospheric studies to understand the presence of water vapor and aerosols and analyze the properties of the cloud. The vegetation parameters such as chlorophyll content, leaf water index, leaf area index, cellulose, and pigment content can be estimated with spectral images, and it helps in ecological studies.

Another vital contribution of spectral imaging is in geology; identifying various minerals and soil types can be carried out with the aid of hyperspectral images mainly (Wan et al. 2021). Coastal water areas can be studied with hyperspectral data to find the presence of chlorophyll, phytoplankton, dissolved organic materials, suspended sediments(Abou El-Magd and El-Zeiny 2014). The contribution of spectral remote sensing is not only limited to the above scientific areas but also in the artic, sea-ice fraction identification, glacier analysis, and melting predictions, forest fire, and biomass burning, etc.

The commercial applications of spectral imaging include mineral exploration, forest cover analysis, and agriculture. The techniques put forward by this study helps to tackle the DR and classification challenges in hyperspectral image analysis and contribute to all application areas discussed above.

Precision agriculture (PA) helps design a decision support system for the farmers to schedule agricultural inputs based on climatic and vegetation factors. Precision agriculture helps in attaining higher crop productivity and thus by achieving food security worldwide. PA helps in sustainable agricultural development, reduced chemical usage in the field, production of quality crops, reduced environmental pollution, requirement-based plotwise farming, and efficient resource usage. Early crop yield prediction is one of the paths to achieve precision farming. The proposed prediction model relates vegetation and climatic parameters to perform a multi-variate time-based analysis to forecast early yield.

1.6 CHALLENGES

Remote sensing-based crop yield prediction techniques are relatively new. Many research works are going on in this domain to overcome remote sensing-based crop yield prediction challenges. This section briefs some of the challenges in the domain.

1.6.1 Availability of spectral images

Hyperspectral images' availability is significantly less compared to other remote sensing images. Most of the hyperspectral missions are launched for a short period, and it covers a particular area of the earth for a specific application. A complete hyperspectral mission that covers the entire earth is still not existing. Therefore, the usage of large training samples for deep learning-based classification is a challenge in hyperspectral image processing.

1.6.2 Lack of spectral library

Every object on the earth's surface has a unique footprint for its spectral reflectance, called a spectral signature. Early classification techniques used this spectral signature profile to discriminate the objects in a spectral image. However, the spectral library contains only a limited spectral signature, and thus it is a challenge for correct classification of objects using the spectral signature. The spectral signature of different objects may look similar due to atmospheric noise in image capture, leading to misclassification.

1.6.3 Curse of dimensionality

Hyperspectral images having more than a hundred contiguous spectral bands. Some of these spectral bands are redundant, and some may not contain relevant information. This redundancy and sparsity in data lead to overfitting the classification model. Even though the multi-band information is present in hyperspectral images, labeled samples' availability is significantly less. This becomes a challenge in hyperspectral image processing.

1.6.4 Lack of high-resolution images

In current scenarios, hyperspectral image sensors can provide only low-resolution samples. Therefore, a single pixel in a low-resolution hyperspectral image may cover a large area (approximately up to 250m) in the field. Thus, classifying this single pixel to a particular class can be error-prone because that area may cover multiple objects of interest (mixed pixel).

1.6.5 Proper selection and identification of vegetation parameters

The crop yield is heavily dependent on the vegetation parameters of the land area; thus, accurate identification of these parameters is critical. The vegetation parameters that influence crop yield vary for each variety of crops and its cultivation method. Hence an expert identification is a must for deciding such vegetation parameters for yield prediction.

1.6.6 Accurate collection of climatic data

Climatic data are usually measured for a large area such as a district or block. The accurate estimation of these parameters on the study area requires placement of sensors and field survey that contributes to accurate yield prediction.

1.7 HIGHLIGHTS OF THE PRESENT RESEARCH WORK

- An in-depth literature survey on challenges in hyperspectral image analysis (dimensionality reduction, classification) and various crop yield prediction models using remote sensing data.

- Introducing a nonlinear dimensionality reduction technique for HSI having a lesser computational overhead.

- Proposing a multiscale hybrid CNN model with minimal learnable parameters for HSI classification.

- Presenting a modified TCN based prediction model for crop yield using vegetation and climatic parameters.

1.8 BRIEF OVERVIEW OF THESIS CONTRIBUTIONS

The significant contributions of this thesis include the challenges in crop yield prediction from remote sensing data. This work introduces novel techniques that address the dimensionality reduction and classification challenges in HSI. A multivariate time-based crop yield prediction model for the paddy field is also presented. The brief details of the work are given below.

1.8.1 Nonlinear dimensionality reduction

Dimensionality reduction is a crucial stage in HSI processing, and linear techniques fail in preserving higher-order statistics of data. The available nonlinear techniques are time-consuming for large datasets. The hybrid dimensionality reduction technique using random projection (RP) and kernel PCA retains the data's higher-order statistics. Experiments were done on three hyperspectral datasets using Gaussian Random Projection (GRP) and Sparse Random projection (SRP) based hybrid models. The results demonstrate that the hybrid method reduces nonlinear DR techniques' computational complexity with better classification accuracy.

1.8.2 Spatio-spectral feature-based HSI classification

The integrated usage of spatial and spectral features increases the classification performance of HSI. The novel hybrid CNN model in the research work considers both spatial and spectral features for HSI classification. The dependency on spatial features is found out using multi-scale receptive fields. The hybridization of 3D-CNN and 2D-CNN reduces the number of learning parameters. Multi-scale Hybrid CNN is compared against

the various state-of-the-art techniques on three widely used datasets. It is found that the proposed technique is computationally less complicated and produces a better result for classification.

1.8.3 Multivariate time-based yield prediction

A multi-temporal data analysis is critical for accurate yield forecasting. Existing machine learning and deep learning prediction models lack such multi-temporal analysis. The proposed radical TCN based yield prediction model for Paddy gave a promising yield prediction result with limited running time. The dilated convolution in TCNN helped to find the long-term historic dependencies of various indices in yield. The model compared against other machine learning and deep learning techniques and outperformed all of them.

1.9 ORGANIZATION OF THE THESIS

The thesis advances in 6 chapters. An outline of each chapter is given below.

- **Chapter 1 : The Introduction** section covers the need for precision agriculture from remote sensing data. The difficulties in processing remote sensing images such as multispectral and hyperspectral are discussed. Motivation, applications, and challenges in remote sensing data-based yield prediction are also addressed. The chapter ends with a brief overview of research contributions and a thesis outline.

- **Chapter 2 : Literature Review** section mainly consists of a detailed review of the dimensionality reduction in HSI, HSI classification techniques, and crop yield prediction models. The section discusses the identified research gaps and the scope of the research work. The datasets used in the proposed research work are familiarized in this section.

- **Chapter 3 : Band Extraction for Hyperspectral Images** includes the proposed hyperspectral band extraction algorithm and their design details. The section contains the details of the experiments and comparisons carried out to evaluate the model. The analysis of the result is discussed in this section with appropriate

conclusions.

- **Chapter 4 : Hybrid CNN Based HSI Classification** covers novel multi-scale spatio-spectral deep learning model for HSI classification. The presented model's performance is compared against the available state-of-the-art techniques and analyzed the results to reach conclusions.

- **Chapter 5 : TCN Based Crop Yield Prediction** discuss the model and design of a multivariate time-based prediction model for rice yield prediction. This chapter also provides the need for dilated convolution in yield prediction. The performance of the model is presented with appropriate analysis and discussions.

- **Chapter 6 : Conclusions and Future Scope** chapter summarize the contributions and findings of this research work. This chapter also provides insights into the future scope and directions for crop yield prediction from remote sensing data.

CHAPTER 2

LITERATURE REVIEW

Chapter 1 introduced the need for precision agriculture and presented a brief overview of hyperspectral band extraction, classification, and crop yield prediction. Motivation for the research work, along with significant applications and challenges of machine learning-based crop yield prediction, are listed in chapter 1. This chapter aims to give a deep perception of recent methodologies proposed and developed in the literature of hyperspectral dimensionality reduction, hyperspectral image classification, and machine learning-based crop yield prediction. The most commonly used hyperspectral image datasets are familiarizing in this chapter. A set of research gaps evolved from a thorough literature review is listed at the end of the chapter, along with the problem definition and objectives.

2.1 DIMENSIONALITY REDUCTION IN HYPERSPECTRAL IMAGES: A RE-VIEW

Measuring crops' biophysical and biochemical attributes for productivity calculation, stress monitoring, etc., through the land survey is a tedious task. However, remote sensing-based measurement is less time-consuming and more accurate. Burgan and Hartford (1993) presented a report to monitor vegetation greenness using Advanced Very High-Resolution Radiometer (AVHRR) five-band satellite data. They used NDVI and Relative Greenness (RG) to analyze the greenness of the crop. In 1994, Thenkabail et al. used Landsat-5 Thematic Mapper(TM) data for corn and soybean yield measure, LAI, wet mass, and dry mass. TM data produce about 35% of the variability in yield

(THENKABAIL et al. 1994). Corn and soybean yield in the south of Iowa City is measured using NDVI by Zhang et al. (1999). They captured the image using ADAR 5500 4-band digital camera and concluded that as the spatial resolution of images reaches 9 meters, it gives a more accurate prediction. Thenkabail et al. (2000) found a relationship between crop characteristics and vegetation indices using a 512 band spectroradiometer. These narrowband spectrum-based models overshot the accuracy of all previous wideband image-based analyses. Later, researchers started using hyperspectral images for various agricultural applications such as crop condition monitoring, weed discrimination, disease detection, yield estimation, etc. The higher number of bands of HSI helps in discriminating crops using clearly defined spectral signature. However, as the number of bands increases, the number of training data required to maintain minimum functionality and statistical confidence for classification purposes must be high. Since hyperspectral images' availability is significantly less, the only solution for the classification problem is dimensionality reduction.

A considerable amount of researches were carried out to extract optimal bands from hyperspectral images. Dimensionality reduction (DR) is mainly classified into two categories, namely linear and nonlinear dimensionality reduction. In linear reduction, the higher dimensional data is converted to a lower dimension using some linear transformations. Principal Component Analysis (PCA) is the most common linear dimensionality reduction method (Khodr and Younes 2011). PCA considers only up to second-order statistics of data i.e,mean and variance. Hyperspectral images are nonlinear, and it follows higher-order statistics, and thus PCA fails for HSI. Maximum noise fraction(MNF) (Green et al. 1988) is another DR technique used for HSI follows only second order statistics. Overcome the drawbacks of PCA and MNF, Wang and Chang (2006) introduces independent component analysis (ICA) based dimensionality reduction technique for hyperspectral images. They introduced three algorithms called ICA-DR1, ICA-DR2, and ICA-DR3 using random initial projection vectors. In ICA-DR1, the number of retained independent components (ICs) are estimated using the virtual dimensionality (VD) concept(Chein-I Chang and Qian Du 2004). They found that an excellent upper approximation of the number of ICs for better classification is twice

the number of VD. The ICA-DR2 method iteratively runs the FastICA technique on the image to find the most common ICs until ICs remain unchanged. The generation of random order ICs in DR1 and DR2 overcome by using a specific manner initial projection vector selection. These techniques resolve the challenges in ICA-based dimensionality reduction but reduce the classification accuracy.

Based on the knowledge of nonlinearity in hyperspectral images, many nonlinear techniques such as Laplacian Eigenmap (Belkin and Niyogi 2003), Locally Linear Embedding (LLE) (Roweis and Saul 2000), ISOMAP (Bachmann et al. 2005) became popular. Mohan et al. (2007) introduced a modified nonlinear dimensionality reduction technique named spatially coherent locally linear embedding. Each band image in HSI are looks similar to a 2D image, and hence the probability of a pixel and its neighbor belongs to the same class is high. Instead of Euclidean distance in classical LLE, Mohan et al. (2007) used spatially coherence distance to find the relationship between a pixel and its immediate neighbors. This method effectively reduces the size of data up to 75% and improves the classification accuracy by 15%. Here the selection of immediate neighbor depends on the size of the image patch ($n = 3$ in this work), and the value of n depends on classification accuracy.

Fractal dimension-based dimensionality reduction of hyperspectral images put forwarded by Ghosh and Somvanshi (2008). They developed a single fractal dimension image from multiple bands of the hyperspectral image. The entropy of the fractal dimension image is higher than any other existing methods, and it improves classification accuracy. However, the fractal dimension depends on the noise factor, and it may lead to misclassification. Junying and Ning (2008) applied the wavelet transform-based noise removal method in the input image before calculating the fractal dimension. The proper choice of noise removal filter became a problem for them. Phillips et al. (2008) used MNF with adaptive filters to maximize the SNR value and applied discriminant analysis to find the optimal bands. This filtered data improved the classification accuracy from 88% to 99%. However, classification accuracy depends on the proper selection of the number of bins and filter size.

Considering the merits of various DR techniques, researchers introduced some hy-

brid dimensionality reduction methods. Koonsanit et al. (2012) used the integration of Information Gain (IG) bands and PCA bands for dimensionality reduction. PCA maximizes the covariance between bands to find optimal bands whereas IG based method assume that bands with a high value of information gain are the optimal bands. However, the redundancy in IG bands is notorious. Hence authors find optimal bands from input data separately and the intersection of two sets of results considered as final optimal bands. This method gave better results for small multimission satellite images.Fuzzy rough set model is another nonlinear dimensionality reduction method. It preserves the meaning of data, and processing needs no additional information. Lodha and Kamlapur (2014) used fuzzy rough set method for hyperspectral dimensionality reduction and its computation time was high compared to other techniques.

Ly et al. (2014) employed a supervised Sparse Graph-based Discriminant Analysis (SGDA) for dimensionality reduction in hyperspectral images. A modified version of SGDA named as block SGDA (BSGDA), which improved the classification accuracy through constrained sparse representation within samples in a cluster. The construction of ℓ_1-graph for SGDA and BSGA is a significant computational burden. Huang et al. (2015) inherited the merits of sparsity in natural images and introduced a nonlinear Sparse Discriminant Embedding (SDE) technique for hyperspectral dimensionality reduction. SDE preserves the sparse reconstructive relation through ℓ_1-graph and enhances the intermanifold separability of data. SDE is a supervised technique, and this method is useful only if prior knowledge about the classes are available. Based on homogeneity and mutual information, Nhaila et al. (2015) found the optimal bands in hyperspectral images. Dimensionality Reduction via Regression(DRR), a new technique that preserves the volume of useful information proposed by Laparra et al. (2015) for HSI band extraction. The classfication accuracy of HSI using DRR varies based on selection of regression function.

Bourennane and Fossati (2015) proposed a multilinear method to perform both denoising and dimensionality reduction on hyperspectral images with the help of wavelet transforms and Wiener filter. There is no guaranty that hyperspectral images are always distorted by white noise only. The multidimensional wavelet transform is not

efficient for non-white noises. Hence researchers proposed a new method, which initially whitened the noise in the image and then applies wavelet and Wiener filter on this whitened image. The proposed method takes more computational time due to the use of brute force method for optimal parameter selection. Wang and Wang (2015) proposed Global Mixture Coordination Factor Analysis(GMCFA) to retain maximum spectral information in HSI. This method passes through three steps; first, converting the high dimensional data to low dimensional manifold using mixture factor analysis. In the second phase, the optimal parameters in low dimensional manifold calculated using Expectation- Maximization (EM) method. The manifold is aligned in global parametrization field by global coordinate factor analysis in the third phase and hence achieve dimensionality reduction. Estimation of Expectation- maximization parameter is the main bottleneck of this technique.

Graph-based discriminant analysis became very popular recently(Chen et al. 2016; Feng et al. 2017; Ly et al. 2014). Affinity matrix calculation using Euclidean distance is not a good measure to project high dimensional data to low dimensional data. Feng et al. (2017) proposed a spectral similarity based measurement for high dimensional to low dimensional projection. The spectral similarity between two pixels calculated from chosen bands leads to better discrimination. Classification accuracy of this method depends on the sparseness controlling parameters η and γ. Semi-supervised double sparse graph based dimensionality reduction technique consider both positive and negative structure relationship of data. The k-nearest neighbor method is used to map unlabeled data to pseudo labeled maps. Sparsity reduces the error in between pseudo label and real label up to an extend (Chen et al. 2016).

Murinto and PA (2017) introduced another hybrid model using SVM and discriminant ICA. Discriminant ICA uses the Fisher criterion and sum of marginal negative entropy independent components for reduction. Hybridization with SVM achieves structural risk minimization and independence maximization, however, this method is slow in multi-class classification. Kernel PCA is another nonlinear dimensionality reduction technique, depend on the number of patterns selected for kernel matrix calculation. The proper selection of initial patterns using density-based spatial clustering reduces the

computational cost of DR and improves the classification accuracy. Therefore, Datta et al. (2017) introduced a new hybrid model using kernel PCA and DBSCAN. Finding more suitable kernel is still existing as a problem in this research work. Supervised dimensionality reduction techniques lead to overfitting when the number of labeled data is less. Wu and Prasad (2018) proposed Semi-Supervised Local Fisher Discriminant Analysis (SSLFDA) to discriminate both labeled and unlabeled data using pseudo labeling concept. They used the Dirichlet Process Mixture Model (DPMM) based clustering for false labeling instead of k-nn in Chen et al. (2016). It preserves more discriminative information for pseudo labels than soft labels but fails in active learning.

A survey on various discriminant analysis-based dimensionality reduction on HSI is carried out by Li et al. (2018). Fast and robust principal component analysis on Laplacian graph (FRPCALG) is a band selection technique proposed by Sun and Du (2018). The small projected dimension using sparse random projection in FRPCALG reduces the method's computational complexity without compromising the classification accuracy. Li et al. (2019) introduced another graph-based dimensionality reduction for hyperspectral images by considering both spatial and spectral features. Table 2.1 summarizes various research works done for dimensionality reduction in hyperspectral images in recent decades.

2.2 HYPERSPECTRAL IMAGE CLASSIFICATION: A REVIEW

Dimensionality reduction is the preprocessing step for hyperspectral image analysis. DR technique in HSI reduces the computational complexity of further processing. Each hyperspectral image covers a wide geographical area. Crop area is an essential measure in crop yield calculation. Therefore a major step in crop yield prediction is the segmentation of different crop areas in HSI using classification techniques. Classification of hyperspectral images can be accomplished using supervised, semi-supervised or unsupervised learning techniques. The availability of hyperspectral images and its ground truth is comparatively less, hence supervised classification technique are tidious. This section discuss about various hyperspectral image classification techniques proposed in this decades.

Table 2.1: Summary of Research Works Done in Dimensionality Reduction of Hyperspectral Images

Research work	Approach used	Band Selection	Band Extraction
Wang and Chang (2006)	Independent Component Analysis		✓
Mohan et al. (2007)	Spatially Coherent Locally Linear Embedding		✓
Ghosh and Somvanshi (2008)	Fractal based Dimensionality Reduction		✓
Junying and Ning (2008)	Wavelet based noise removal+Fractal dimension		✓
Phillips et al. (2008)	MNF with adaptive filter+discriminant analysis	✓	
Koonsanit et al. (2012)	Information gain band+PCA	✓	
Lodha and Kamlapur (2014)	Fuzzy roughset model	✓	
Ly et al. (2014)	Sparse Graph based Discriminant Analysis		✓
Huang et al. (2015)	Sparse Discriminant Embedding		✓
Nhaila et al. (2015)	Homogeneity+mutual information	✓	
Laparra et al. (2015)	Dimensionality reduction via regression		✓
Bourennane and Fossati (2015)	Wavelet transform+Wiener filter		✓
Wang and Wang (2015)	Global mixture coordination factor analysis		✓
Chen et al. (2016)	Semi-supervised double sparse graph		✓
Feng et al. (2017)	Spectral similarity based discriminant analyis		✓
Murinto and PA (2017)	SVM+ Discriminant ICA		✓
Datta et al. (2017)	Kernel PCA+DBSCAN		✓
Wu and Prasad (2018)	Semi-supervised Local Fisher Discriminant Analysis		✓
Sun and Du (2018)	Fast and robust principal component analysis on Laplacian graph (FRPCALG)		✓
Li et al. (2019)	Spectral spatial neighborhood graph (SSNG)		✓

Gomez-Chova et al. (2003) proposed a combination of the supervised and unsupervised technique for HSI classification. In the unsupervised technique, the image size is reduced by uniform spatial decimation. Then applied C-means clustering algorithm on the data to find initial clusters. These clusters are used as initial seeds for Expectation Maximization (EM) algorithm and iteratively repeat this EM algorithm for density estimation. They used classification Gaussian Maximum Likelihood method for classification. Once the classification over, the training spectral signature data of different crops used for labeling. Therefore, this phase is a supervised method. The proposed semi-supervised technique gave an excellent result in robustness and accuracy. However, this method is more suitable for classes with the Gaussian mixture distribution.

Hyperspectral classification using predefined spectral signature is another method. Rao (2008) classifies three varieties of rice, chilies, sugar cane and cotton plants in Andra Pradesh, India using spectral signature. It gave accuracy up to 86-88%. Due to noise addition, the spectral signature of different varieties of rice and sugar cane crop was the same and it reduces classification accuracy. Hadoux et al. (2012) applied Partial Least Square LDA (PLS-LDA) on spectral information for weed and wheat classification. They used a reference surface set and spectral preprocessing to avoid the variable lighting condition. This model provides high robustness. However, it causes a classification error due to the variability of images taken at different time. The classification accuracy of hyperspectral images depends on image resolution. Alganci et al. (2013) investigated the difference between pixel-based and object-based classification techniques for crop identification. They used Maximum Likelihood, Spectral Angle Mapper and SVM for pixel-based classification. The object-based classification with high resolution gives better accuracy than others. Spectral angle mapper technique compares the spectral similarity between every two bands in an image by measuring the angle between the vectors. Since it is a pixel-based approach, it takes more time for processing and affected by noise (Boitt 2014). Supervised waveform classification is another method suggested by Moharana and Dutta (2014). Clustering technique used in this approach reduces the noise and critical bands in HSI found using band-band correlation. Proper selection of the similarity metric in clustering technique improves the accuracy.

20

Sateesh and Sridhar (2014) used Niche Hierarchical Artificial Immune System (NHAIS) based clustering technique for hyperspectral image classification. First, they performed PCA on the input image for dimensionality reduction. Then each pixel and its weighted neighbors are used for clustering iteratively. These weighted neighbor methods reduce the edge degradation in images. However, NHAIS is not useful for randomly distributed pixels and isolated pixels in DR images. Kernel-based SVM technique, an object-based image analysis paradigm for crop mapping from hyperspectral images put forwarded by Liu and Bo (2015). They used object mean, texture and geometrical features for classification. This method provides overall accuracy of 90.3%. A hybrid model of segmentation and classification approach offered by Zhang et al. (2016). They used feature band set selection and object-oriented approach for crop discrimination. It outperformed MNF based optimization, but the quality of segmentation controls the classification accuracy.

2014 onwards deep learning techniques introduced for hyperspectral image analysis. Chen et al. (2014) proposed the concept of deep learning for hyperspectral image classification for the first time. Their model is a hybridization of PCA, deep learning and logistic regression. PCA helps to find optimal bands in hyperspectral images. They extract the spatial features and spectral features separately and classify the image using separate features. They introduced a joint spectral-spatial classification by selecting dominating spatial features of each pixel and by merging spectral characteristics to them. The training time for this method is comparatively high. However, this approach opens a new window for researchers to work on spatial-spectral classification. Contextual deep learning is another spatial- spectral feature based classification proposed by Ma et al. (2015). This classification method is entirely a supervised technique, which passes through three steps namely, spatial information extraction, spectral feature mining, and feature smoothness. Since lots of unlabeled data available for hyperspectral classification, this deep network is not adequate for HSIs.

A five-layered deep Convolutional Neural Network(CNN) was implemented for HSI classification in spectral domain by Hu et al. (2015). This network classifies the image purely based on supervised spectral signature features. Classification accuracy depends

on the number of training samples, and thus it fails to classify categories with a small number of samples. Chen et al. (2016) developed multilayered CNN for hyperspectral image classification. They extracted both spectral and spatial features for crop discrimination. However, the limited number of input samples leads to overfitting. Virtual samples and post-classification processing may improve accuracy and robustness. Zhao and Du (2016) suggested a combination of CNN and Balanced Local Discriminant Embedding(BLDE) technique. BLDE is used for spectral feature extraction, and CNN is for spatial features. They used PCA for dimensionality reduction and used these optimal bands for spatial feature extraction in CNN. These two features were stacked and fed into multiclass classifier for classification. This approach improves the classification accuracy and fails in optimal observation scale selection.

Deep Belief Network(DBN) based classification is another deep learning approach used for hyperspectral image analysis. Spatial updated DBN is a HSI classification technique, which using spatial information within spectrally similar contiguous pixel (Mughees et al. 2017). In this approach, hyperspectral image segmented into spatially similar groups with similar spectral features and object-level features were computed for classification. A systematic approach of hyper-segmentation depends on the classification accuracy of this method. Li et al. (2018) proposed a novel hyperspectral classification using optimal DBN and Texture Feature Enhancement(TFE). First, they grouped the bands into different sets based on band group method and performed texture feature enhancement using guided filters. These texture enhanced hyperspectral images were used as an input for classification by DBN.

Most of the CNN based approaches consider both spectral and spatial features, while still they fails in dimensionality optimization(Pan et al. 2018; Paoletti et al. 2018; Yu et al. 2017). Multi-grained Network(Mug Net) combine different grain's spatial and spectral information for classification. This network used the semi-supervised technique for unlabeled samples (Pan et al. 2018). Li et al. (2017b) used 3D CNN for HSI classification by considering both spectral and spatial features together. They considered the hyperspectral image as a cube and performed the convolution using 3D kernels. However, this method can't classify unlabeled samples. Ji et al. (2018) proposed

another 3D CNN concept by considering both spatial and temporal features for crop classification in hyperspectral images. It improves the efficiency of manual learning. Pixel pair feature based hyperspectral image classification is a novel method suggested by Li et al. (2017). In this method, a pixel and its neighbors are trained using CNN, and test pixels are classified using a voting strategy. This method's classification accuracy is better than conventional CNN, but compromises the computational complexity.

Hyperspectral Reconstruction CNN (HRCNN) based classification (Li et al. 2017a) has four steps namely normalization, band selection, reconstruction, and classification. This method used the first PCA image as a reference image and calculate the widely used six gray level co-occurrence feature for all the bands. They compare the co-occurrence features of each band with a reference image feature and select the bands with a minimum difference as optimal ones. HRCNN using four convolutional layers and each is responsible for patch extraction and representation, feature enhancement, non-linear mapping and reconstruction orderly. These reconstructed images are classified using Extreme Learning Machine(ELM). This technique is faster than all the state of art methods and provides better accuracy with less amount of training samples.

Band-Adaptive Spectral-Spatial Network(BASS Net) (Santara et al. 2017) is an end to end deep learning network developed to overcome the challenges in hyperspectral image classification. This architecture extract band specific spectral-spatial features for classification. The entire network design divided into three blocks. The aim of block1 is spectral feature selection and band partitioning. Here the 3D volume input image is partitioned into nonoverlapping bands of specific size using their spectral features. Block2 has a specific number of parallel networks. Each band fed into each fully connected network of block2 for spectral-spatial learning. Block3 concatenate the output of block2 and perform classification. This network has better classification performance and faster convergence than other deep learning networks.

Jiao et al. (2017) presented a novel spatial-spectral based hyperspectral image classification technique using the deep fully convolutional network. First, they used VGG-verydeep16 based network to extract in-depth multiscale spatial information. Then these spatial features were fused with spectral features by weighted fusion method. This

fusion features fed into classifier for pixel-wise classification. This method performs effectively in highly nonlinear and diverse images. Chen et al. (2018) proposed another spatial-spectral based hyperspectral image classification. They compute the spatial similarity in each band using a spatial window and then find the spatial and spectral features in each image blocks. These features are merged using convolutional layers and a fully connected layer is used for classification. Moreover, this network design combined the network with SVM for better classification and introduced a new adaptive spatial window size selection algorithm too. The window size is strict into two discrete values is a constraint of this work.

Spectral-spatial Residual Network (SSRN) (Zhong et al. 2018), Spectral-spatial Unified Network (SSUN) (Xu et al. 2018), Unsupervised spatial-spectral feature learning using 3D-convolutional autoencoder (Mei et al. 2019), spectral- spatial LSTM Zhou et al. (2019), spatial-spectral ConvLSTM (Hu et al. 2019), capsule network based HSI classification (Paoletti et al. 2019a), Deep pyramidal network based HSI classification Paoletti et al. (2019b) and Hybrid SN (Roy et al. 2020) are some of the latest works carried out in HSI classification by considering both spatial and spectral features together. These works are outperformed the existing classification model that consider any one of the feature alone. Table 2.2 summarizes various research works done for classification of hyperspectral images in recent decades.

2.3 CROP YIELD PREDICTION MODELS: A REVIEW

Machine learning (ML) algorithms have gained much attention in recent years and are used for crop yield prediction. ML techniques analyze climatic and vegetative data to decide the type of crop to cultivate, time of cultivation, and management of the crop. Machine learning-based prediction models trained using prior data aid the prediction of yield for current conditions. Although there are many machine learning-based prediction models, the most commonly used techniques in yield predictions are decision trees (Shekoofa et al. 2014), random forest (Jeong et al. 2016), regression models (Johnson et al. 2016) and neural networks (Bose et al. 2016).

Deep learning, the subset of machine learning, created a tremendous impact on

Table 2.2: Summary of Research Works Done in Hyperspectral Image Classification

Research work	Approach used
(Gomez-Chova et al. 2003)	Semi-supervised classification(EM+Gaussian Mixture model)
(Rao 2008)	Hyperspectral reflectance data
(Hadoux et al. 2012)	Partial Least Square LDA on spectral information
(Alganci et al. 2013)	Object based classification
(Boitt 2014)	Spectral Angle Mapper
(Moharana and Dutta 2014)	Supervised waveform classification
(Sateesh and Sridhar 2014)	NHAIS
(Liu and Bo 2015)	Kernel based SVM
(Zhang et al. 2016)	Feature Band set+Object oriented approach
(Chen et al. 2014)	Deep Neural Network
(Ma et al. 2015)	Contextual deep learning
(Hu et al. 2015)	Convolutional Neural Network(CNN)
(Chen et al. 2016)	CNN
(Zhao and Du 2016)	Balanced local discriminant embedding (BLDE)+ CNN
(Mughees et al. 2017)	Deep Belief Network(DBN)
(Li et al. 2018)	DBN+Texture feature enhancement
(Pan et al. 2018)	MugNet
(Li et al. 2017a)	Deep CNN
(Li et al. 2017)	Pixel pair feature+CNN
(Li et al. 2017b)	3D CNN
(Ji et al. 2018)	3D CNN
(Santara et al. 2017)	Bass net
(Jiao et al. 2017)	Deep CNN
(Chen et al. 2018)	Spectral-Spatial CNN
(Zhong et al. 2018)	Spectral-Spatial Residual Network (SSRN)
(Xu et al. 2018)	Spectral-Spatial Unified network (SSUN)
(Mei et al. 2019)	3D-Convolutional Autoencoder
(Zhou et al. 2019)	Spectral-Spatial LSTM
(Hu et al. 2019)	Spatial-Spectral ConvLSTM
(Paoletti et al. 2019a)	Capsule Network
(Paoletti et al. 2019b)	Deep Pyramid Network
(Roy et al. 2020)	Hybrid SN

many research fields. Deep neural networks(DNN) are similar to artificial neural networks used in machine learning. However, they have a large number of intermediate layers to analyze the correlation between features. Since yield prediction is a tedious task and yield depends on many parameters linearly and nonlinearly, researchers started using deep learning techniques. The majority of deep learning models used for crop yield prediction can be categorized into three. They are DNN based, convolutional neural network (CNN) based, and long-short term memory (LSTM) based (van Klompenburg et al. 2020). CNN is a sort of deep neural network having convolutional layers and pooling layers. CNN is a self-learned network mainly used in image data. The pooling layer in CNN helps to downsample the data, thus generalize the feature map and avoid overfitting (Brownlee 2019). LSTM is a recurrent neural network (RNN) structure extensively used for temporal analysis of input data. Since the yield heavily depends on temporal climatic and vegetative data, LSTM is a suitable yield prediction method. It fundamentally explores the relationship between past long-period data and present data.

In 2019 Khaki and Wang designed a DNN model to predict the yield from genotype and environmental factors provided by Syngenta Crop Challenge. The proposed deep network has 21 hidden layers with 50 neurons in each layer. Since there are numerous environmental factors available in the dataset, the researchers performed a feature selection technique using a guided backpropagation algorithm and maintained the prediction accuracy (Khaki and Wang 2019). Tsouli Fathi et al. (2020) formulated a deep learning model for crop yield prediction in the Medeterenian region agro-chemical and climatic data. They compared the proposed DNN with existing ML techniques and achieved a low prediction error. Saravi et al. put forwarded a DNN model to predict maize yield by considering precipitation and irrigation during the crop's life cycle. This model reduced the root mean square error in prediction. However, neglecting other agro-climatic factors that influence yield (Saravi et al. 2020) is a major setback. Winter wheat yield prediction within the season is another DNN based prediction model proposed by Wang et al. They analyzed crop, soil, climatic data, and VI for prediction. The model compared against AdaBoost, random forest, SVM and found that the sug-

gested model performed better (Wang et al. 2020b). The extensive studies conducted by Bhojani and Bhatt (Bhojani and Bhatt 2020) resulted in the introduction of new activation functions to predict wheat yield using DNN. They evaluated the yield using different climatic parameters and found the best pair of features for each geographical area. The combination of multimodal data such as thermal, RGB, and spectral helps to extract more number features to predict yield accurately(Maimaitijiang et al. 2020).

Manual collection of data through field surveys is tedious and often leads to inaccurate prediction of yield. However, the usage of remote sensing data helps to extract the features for a wide area with less overhead. Convolutional layers in CNN are the best aid to extract features automatically. The CNN in (Rahnemoonfar and Sheppard 2017), is designed to count tomatoes from captured images. They developed a modified Inception-Resnet architecture and trained it using synthetic images. The model displayed the ability to count fruits even from occluded and shaded images accurately. Nevavuori et al. used NDVI and RGB images to predict wheat and barley yield for individual croplands. They designed a CNN based deep learning model for yield prediction without considering the climatic and soil features(Nevavuori et al. 2019). Similarly, Yang et al. (2019) proposed another CNN model to predict rice grain yield. The authors compared the RGB-based CNN model with the conventional regression model and found it futuristic. (Chen et al. 2019), (Terliksiz and Altýlar 2019), (Shidnal et al. 2019), (Yalcin 2019), and (Kang et al. 2020) are other studies related to yield prediction using CNN.

In 2018 Wang et al. predicted soybean yield in Argentina with the help of remote sensing data. They used LSTM cells in the proposed recurrent neural network and applied transfer learning technique to predict soybean yield in Brazil(Wang et al. 2018). Schwalbert et al. (2020) also forecasted the soybean yield in Brazil using the LSTM model in 2020. However, they used both climatic and satellite data to predict yield, and the model outperformed other machine learning and deep learning techniques. The combined use of remote sensing, meteorological, and phenological data for yield prediction in LSTM models reduced the prediction error and helped early prediction. Jiang et al. (2020), and Zhang et al. (2020) designed LSTM models to predict maize yield by

considering the combination of all features. CNN-LSTM hybridization was the next step in crop yield prediction techniques. The end-of-season soybean yield prediction model is designed by Sun et al., and the model uses CNN for spatial data analysis and LSTM for phenological dependency calculation. Therefore the hybrid model is more accurate than existing deep learning techniques(Sun et al. 2019). A similar CNN-LSTM model is proposed in (Wang et al. 2020a) for winter wheat yield prediction. The CNN part in the proposed model is used for static feature analysis and the LSTM for dynamic data analysis.

2.4 RESEARCH GAPS

Some of the important research gaps identified from the above literature review are listed below.

- A new dimensionality reduction technique that reduces computational complexity and improves classification accuracy is necessary for hyperspectral images

- 3D-CNN designs help to retain both spectral and spatial features, whereas the number of learnable parameters in 3D-CNN models are very high

- The availability of hyperspectral images and its groundtruth in very less, and image patches are used for most of the supervised classification techniques.

- Image patch size selection is another challenging issue. If the window size is large, the more local similarity is considered and it lead to misclassification due to noise.

- If the model consider smaller image patches for feature extraction, it fails to identify the correlation between pixels.

- Proper selection of vegetation parameters and climatic factors for yield prediction model is necessary

- Continuous monitoring of crop land is needed for accurate prediction and most of the existing prediction models fails continuous monitoring.

- DNN based models fails in multivariate time based analysis

- Increase in receptive field of LSTM models increase the time complexity of the model.

2.5 PROBLEM STATEMENT

The primary aim of this work is to design a crop yield prediction model from satellite images and climatic data.

Discrimination of different crop area from high dimensional Hyperspectral images and calculation of different vegetation index parameters are necessary for accurate prediction of crop yield. However,the processing of hyperspectral images is computationally complex due to its high dimensionality. Therefore research work aims to design a crop yield prediction model using optimal bands of hyperspectral images, vegetation indices and climatic data.

The objectives of the work are:

1. Extract the optimal bands from hyperspectral images for crop classification using nonlinear dimensionality reduction techniques.

2. Crop area discrimination using deep learning technique by considering spectral and spatial features.

3. Design a machine learning based prediction model using vegetation indices and climatic factors

2.6 RESOURCES USED

This section familiarizes the datasets used in this research work for experimenral analysis.

2.6.1 Hyperspectral Image Datasets

Mainly three hyperspectral images and its groundtruth are used for dimensionality reduction and classification. They are Indian pines(IP), Pavia University(PU), Salinas(SA) and Houston dataset. IP, PU and SA collected from the website `http://`

29
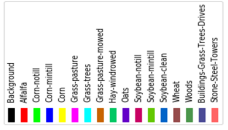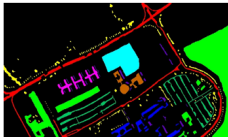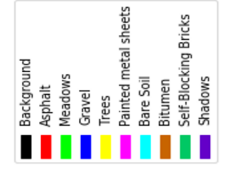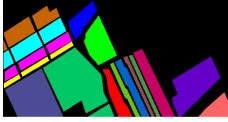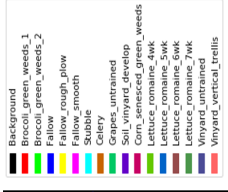
`www.ehu.eus/ccwintco/index.php.`

Indian pine dataset is captured from Indian pine test site using AVIRIS sensors. Each spectral image in Indian pine is of size 145×145. Sensor acquired a total of 224 spectral bands within the wavelength range 0.4-2.5 micrometers. Out of these 224 bands, some of the bands (104–108, 150–163, and 220) are in the region of complete water absorption and these are not helpful for classification process. After removing the twenty water absorption bands, the remaining two hundred bands are used for experiments. IP dataset contain 16 classes, most of the area is covered by vegetation. The classes are Alfalfa, Corn-notill, Corn-mintill, Corn, Grass-pasture, Grass-trees, Grass-pasture-mowed, Hay-windrowed, Oats, Soybean-notill, Soybean-mintill, Soybean-clean, Wheat, Woods, Buildings-Grass-Trees-Drives, and Stone-Steel-Towers.

Pavia university image is captured using ROSIS-03 optical sensors in 0.43 to 0.86-micrometer wavelength range. The hyperspectral image contain a total of 115 spectral bands. About 12 number of noise bands are removed from the sensor data and remaining 103 spectral images of size 610×340 with 1.3 meters geometric resolution is used for experiments. Image ground truth is differentiated into 9 classes namely Asphalt, Meadows, Gravel, Trees, Painted Metal Sheets, Bare Soil, Bitumen, Self-Blocking Bricks, and Shadows.

Salinas dataset captured the Salinas Valley area of California using AVIRIS sensor of 224 bands. Similar to Indian pine dataset, 20 bands of water absorption area are removed from the original data, and the remaining 204 bands are used for the experiment. Each spectral image is of size 512×214 with a spatial resolution of 3.7 meters, and the ground truth has 16 class labels. The classes are Brocoli_green_weeds_1, Brocoli_green_weeds_2, Fallow, Fallow_rough_plow, Fallow_smooth, Stubble, Celery, Grapes_untrained, Soil_vinyard_develop, Corn_senesced_green_weeds, Lettuce_romaine _4wk, Lettuce_romaine _5wk, Lettuce_romaine _6wk, Lettuce_romaine _7wk, Vinyard _untrained, and Vinyard_vertical_trellis. The class labels, number of samples present in each dataset, and pseudo-colored image of ground truth are represented in Table 2.3.

Table 2.3: Ground truth image of various datasets, class labels and number of samples

Indian Pines (IP)

Class	Samples
Background	10776
Alfalfa	46
Corn-notill	1428
Corn-mintill	830
Corn	237
Grass-pasture	483
Grass-trees	730
Grass-pasture-mowed	28
Hay-windrowed	478
Oats	20
Soybean-notill	972
Soybean-mintill	2455
Soybean-clean	593
Wheat	205
Woods	1265
Buildings-Grass-Trees-Drives	386
Stone-Steel-Towers	93
Total Samples	21025

Pavia University (PU)

Class	Samples
Background	164624
Asphalt	6631
Meadows	18649
Gravel	2099
Trees	3064
Painted metal sheets	1345
Bare Soil	5029
Bitumen	1330
Self-Blocking Bricks	3682
Shadows	947
Total Samples	207400

Salinas (SA)

Class	Samples
Background	56975
Brocoli_green_weeds_1	2009
Brocoli_green_weeds_2	3726
Fallow	1976
Fallow_rough_plow	1394
Fallow_smooth	2678
Stubble	3959
Celery	3579
Grapes_untrained	11271
Soil_vinyard_develop	6203
Corn_senesced_green_weeds	3278
Lettuce_romaine_4wk	1068
Lettuce_romaine_5wk	1927
Lettuce_romaine_6wk	916
Lettuce_romaine_7wk	1070
Vinyard_untrained	7268
Vinyard_vertical_trellis	1807
Total Samples	111104

2.6.2 MODIS Vegetation Data

This research work focused on MODIS land data products and the data is collected from `https://modis.ornl.gov/globalsubset/`. This work collected Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and Two Band Extended Vegetation Index (EVI2) using MOD13Q1, MOD13A1, and MOD13A2. Other vegetation parameters such as Leaf Area Index (LAI), fraction of Photosynthetically Active Radiation (fPAR), and Gross primary Production (GPP) are collected from the MODIS products MOD15A2 and MOD17A2. The land cover mask for the study area is generated using MODIS land cover MCD12Q1.

2.6.3 Climatic Data

The research work uses the climatic parameters minimum temperature (min T), maximum temperature (max T), Rainfall, Wind, and Humidity to predict the yield. The climatic parameters for the study area are collected from the website `https://www.worldweatheronline.com`. Here, the temperature is measured in degree celsius($^{\circ}$ C) and the amount of rainfall in millimeters (mm). The measuring units of wind and humidity are kilometer per hour (km/h) and percentage (%), respectively

2.6.4 Crop Yield Data

The annual crop yield for Paddy in the study region is collected from The Indian Ministry of Agriculture & Farmers Welfare website (`http://agricoop.nic.in`). The website `http://www.ecostat.kerala.gov.in/index.php/agricultures` also provides a detailed report of agricultural statistics in Kerala per year. This research work uses the annual yield of Paddy from the year 2012 to 2019 in prediction model.

2.7 SUMMARY

This chapter critically reviewed the latest work done in dimensionality reduction and classification techniques in hyperspectral images. A detailed review of crop yield prediction models is also discussed in this chapter. The resources used section mentioned various hyperspectral datasets used for evaluating the first two objectives of the research work. The details and sources of MODIS data, crop yield data, and climatic data are

also highlighted in this chapter. The research gaps are listed, and the current research work's problem statement is also included in one section of the chapter. The proposed methodology for each of the objectives will explain in later chapters.

CHAPTER 3

BAND EXTRACTION FOR HYPERSPECTRAL IMAGES

3.1 INTRODUCTION

One of the most challenging problems faced by HSI processing is the curse of dimensionality. The presence of highly redundant data acquired by hyperspectral sensors leads to increased computational complexity and reduced classification performance. The literature survey section thoroughly describes various band selection and band extraction techniques. Here in this chapter, the main focus is the introduction of an efficient nonlinear band extraction technique. The main focus of the research problem is HSI classification, and it needs a better DR algorithm at its preliminary stage. Hence the first objective was to model a highly efficient DR technique.

The proposed nonlinear DR technique incorporates the advantages of random projection and manifold learning. The JL lemma (Johnson and Lindenstrauss 1984) behind random projection helps to reduce the dimensions of data while retaining its nonlinearity.

The contributions of the chapter are :

- A hybrid nonlinear DR technique for HSI preprocessing.

- Usage of Gaussian random projection preserves the nonlinear property of natural images.

- The performance of the DR techniques is compared against various available non-linear techniques in terms of classification performance and execution time.

The chapter is organized as follows: Section 3.1 introduces the band extraction technique. Section 3.2 briefs preliminary concepts associated with random projection and manifold learning. Section 3.3 describes the proposed band extraction technique along with its pseudocode. Section 3.4 details the experimental setup and analysis of results, and Section 3.5 summarizes the proposed method and its significance.

3.2 PRELIMINARIES

Hyperspectral image X is denoted as a 3-D cube of size $M \times N \times D$, where M and N are the spatial width and height of the image and D denotes the number of spectral bands. In dataset representation, each data $X = \{x_1, x_2, \ldots, x_S\} \in \mathbb{R}^{D \times S}$ belongs to some class c and S is the total number of pixels ($S = M * N$).

3.2.1 Random Projection

Random projection (Johnson and Lindenstrauss 1984; Venkatasubramanian and Wang 2011) is a dimensionality reduction technique, which maps the high dimensional data into a lower dimension by preserving the distance between data points. Suppose the input data X having size $n \times d$ random projection map X into a lower-dimensional data Y of size $n \times k$ using a random projection matrix $R^{d \times k}$. The computational complexity of random projection is $O(knd)$.

$$Y^{n \times k} = X^{n \times d} R^{d \times k} \tag{3.1}$$

The idea behind random projection is Johnson Lindenstrauss lemma (Johnson and Lindenstrauss 1984) : Suppose we have an arbitrary matrix $X \in \mathbb{R}^{n \times d}$. Given any $\epsilon > 0$, there is a mapping $f : R^d \to R^k$, for any $k \geq O\frac{logn}{\epsilon^2}$, such that, for any two rows $x_i, x_j \in X$, we have

$$(1 - \epsilon)||x_i - x_j||^2 \leq ||f(x_i) - f(x_j||)^2 \leq (1 + \epsilon)||x_i - x_j||^2 \tag{3.2}$$

Equation 3.2 states that while converting input data from d dimension to k dimension, its Euclidean distance is preserved with a factor of $1 \pm \epsilon$.

Gaussian random projection(GRP) is one of the most commonly used random pro-

jections. Here the random matrix R generated from Gaussian distribution, which satisfies Orthogonality and normality. For a $m \times n$ random matrix, the time complexity for random projection to k dimension is $O(mnk)$. Time complexity reduces to $O(snk)$ if the random matrix has only s number of non-zero elements. Thus Sparse random projection(SRP) evolved with a new random matrix R having more zero elements (Dasgupta and Gupta 2003; Johnson and Lindenstrauss 1984).

A simple format of Random matrix is known as sign matrix, its distribution is as follows

$$R_{ij} = \begin{cases} \frac{1}{\sqrt{k}}, & \text{p=1/2} \\ \frac{-1}{\sqrt{k}}, & \text{p=1/2} \end{cases} \tag{3.3}$$

where R_{ij} is each element in random matrix R and k is the desired dimension. Achlioptas (2001) proposed a sparse random matrix from this sign matrix, its distribution is:

$$R_{ij} = \frac{1}{\sqrt{k}} * \begin{cases} 1, & \text{p=1/6} \\ -1, & \text{p=1/6} \\ 0, & \text{p=2/3} \end{cases} \tag{3.4}$$

Choosing a measure for amount of sparsity in sparse matrix, Matousek designed a new matrix

$$R_{ij} = \sqrt{\frac{q}{k}} * \begin{cases} 1, & \text{p=1/6} \\ -1, & \text{p=1/6} \\ 0, & \text{p=2/3} \end{cases} \tag{3.5}$$

where q is the measure of amount of sparsity. Achlioptas generally choose $q = 3$ and construct the sparse random matrix as:

$$R_{ij} = \sqrt{\frac{3}{k}} * \begin{cases} 1, & \text{p=1/6} \\ -1, & \text{p=1/6} \\ 0, & \text{p=2/6} \end{cases} \tag{3.6}$$

In all the sparse random matrices, $\frac{2}{3}$ portion is filled with zero, thus it reduce the computational complexity of projection.

37

3.2.2 Kernel PCA

Kernel PCA is an extension PCA in nonlinear space, capable to capture the higher order statistics of data. Input is transformed into new feature space, where the discrimination power of each instant from other is very high. Convert data in $\mathbb{R}^D$ to feature space F by applying a nonlinear transformation ϕ on input data x.

$$\mathbb{R}^D \rightarrow F : x \rightarrow \phi(x) \tag{3.7}$$

Choosing the proper ϕ for calculating the kernel matrix in Kernel PCA is a challenging area of research. Gaussian kernel is used in this research work because hyperspectral images follows Gaussian distribution in most of the cases. The kernel matrix K is calculated as:

$$K_{ij} = exp(\frac{-1}{2\sigma^2}||x_i - x_j||) \tag{3.8}$$

where K_{ij} is each element in kernel matrix K, the value of σ depend on the dataset and x_i, x_j are each pixel values in the image. The data is centered using the equation 3.9

$$K_c = K - 1_N K - K 1_N + 1_N K 1_N \tag{3.9}$$

where $1_N = \frac{1}{N} * ones(N, N)$, K_c is centered kernel and N the total number of samples(pixels) in hyperspectral image. ones(m,n) means a $m \times n$ matrix having all the values will be one. Similar to linear PCA, kernel PCA follows eigen decomposition on the kernel matrix K_c and find k eigen vectors with largest eigen values. Eigen decomposition of $N \times N$ matrix is computationally complex when compared with linear PCA decomposition. Thus a group of representative pixel selection from each class of pixels will reduce the computational complexity.

3.2.3 Local Linear Embedding

Local linear embedding(LLE) is an unsupervised nonlinear dimensionality reduction techniques which transform high dimensional data to low dimension based on neighborhood preservation. In the first step of LLE, find neighbors of each data points using k-nearest neighbor algorithm. Each data point can be a linear combination of its neighbor. Solving the linear equations to find the weight matrix W in such a way that the cost function will be minimum. Cost function is represented as reconstruction error

$$E(W) = \sum_i |x_i - \sum_j W_{ij} x_j|^2 \tag{3.10}$$

In the above equation x_i is the data point, x_j is its neighbor and W_{ij} is the weight(contribution) of jth data point for i^{th} point reconstruction. The weight matrix must satisfy two constraints in solving this least square method. They are:

- $W_{ij}= 0$ if x_j not a neighbor of x_i

- $\sum_j W_{ij}= 1$, i.e, sum of each row in weight matrix must be 1.

The lower dimension data points are calculated using eigen decomposition of weight matrix. LLE transformation convert the weight matrix into a sparse matrix M, calculate smallest k eigen vectors of M and find the dimensionally reduced data using linear reconstruction technique. Here k is the desired dimension and M is created as:

$$M = (I - W)^{'} * (I - W) \tag{3.11}$$

, I is identity matrix. Hessian LLE is a variation of LLE, which works mainly for smooth manifolds and it is computationally complex than normal LLE. The parameter k for nearest neighbor estimation is chosen by cross validation.

3.2.4 Isomap

Isomap is a nonlinear dimensionality reduction technique, similar to LLE. In this technique, construct a neighborhood graph through finding the neighbors of each data points using k-nearest neighbor algorithm. The shortest distance between each pair of data points are calculated using Floyd- Warshall algorithm. This distance matrix is considered as the weight matrix W for further processing. Lower dimensional embedding of data points are done through classical multi scale embedding(MDS) technique. MDS calculate squared proximity matrix B from weight matrix W as:

$$W^{(2)} = [w_{ij}^2] \tag{3.12}$$

Apply double centering on this matrix and find B, find m eigen vectors correspond to largest eigen values and find low dimensional data using this eigen vectors.

3.2.5 Spatial spectral neighborhood graph(SSNG) technique

Most of the graph based nonlinear dimensionality reduction techniques focus on intrinsic graph construction. SSNG technique works on superpixel based segementation to

cluster the HSI pixels into different classes. Superpixels are esimated using principal components of input hyperspectral image. Next step in SSNG is the consruction of neighborhood graph. The distance between two superpixels are calculated using a new distance metric, which is formulated as:

$$d(x_i, x_j) = ||x_i^{spe} - x_j^{spe}||_2 + \lambda ||x_i^{spa} - x_j^{spa}||_2 \tag{3.13}$$

x_i^{spe}, x_i^{spa} are spectral and spatial pixel values and λ is a balancing coefficient. Then calculate the weight matrix W from the neighborhood graph and create a Laplacian matrix L using the following equation:

$$L = D - W \tag{3.14}$$

where D is a diagonal matrix in which diagonal elements are each row sum of weight matrix. Apply optimal linear transformation on Laplacian matix to map high dimensional data to low dimensional data.

3.3 PROPOSED BAND EXTRACTION TECHNIQUE

As discussed in the preliminary section, hyperspectral images look like a 3D cube of dimension $M \times N \times D$. This 3D image is converted to two- dimensional data of size $S \times D$ for the dimensionality reduction technique. The proposed DR technique is a hybrid band extraction technique that uses both linear and nonlinear methods.

Hyperspectral images holding nonlinear behavior, and hence nonlinear DR techniques are more relevant for preserving their nonlinear properties. However, direct implementation of nonlinear DR techniques is computationally expensive. The proposed DR technique is working in two phases. In the first phase, a linear dimensionality reduction technique converts the data from D dimension to an intermediate dimension k_1. The first phase of the proposed method using Random projection as linear DR techniques. The JL lemma behind random projection helps to reduce the distortion in data while mapping higher dimensions to lower dimensions. Hyperspectral images are mapped to lower dimension k_1 using two types of the random projection matrix: Gaussian random projection (GRP) and sparse random projection (SRP).

In phase two of the proposed DR technique, nonlinear dimensionality reduction techniques are applied on intermediate k_1 dimensional data. Dominant nonlinear meth-

ods such as KPCA, LLE, and Isomap have used the second phase. A new graph-based DR technique was also used to correlate the effect of random projection before the non-linear technique. Here the desired dimension is calculated by analyzing the cumulative eigenvalue. The nonlinear technique is repeated for different dimension (from 2 to 30) and find the minimum dimension value B whose cumulative eigenvalue greater than 95%. Algorithm 3.1 shows the pseudocode representation of proposed hybrid dimensionality reduction technique.

The output obtained from this proposed DR technique is a dimensionally reduced data of size $X^{S \times B}$. This data must be double-checked for classification accuracy before feeding it as an input to the HSI classification technique. Dimensionally reduced image is given into the k-nearest neighbor classifier. The performance of the proposed method is evaluated using overall classification accuracy, the area under curve, and execution time for classification. Classification accuracy and area under the curve are calculated using a 10-fold cross-validation technique. Entire data is divided into ten blocks, and classification is repeated ten times by considering one block as testing data and the remaining nine blocks as training at a time.

Algorithm 3.1: Hybrid Dimensionality Reduction Technique

Input: Dataset $X \in \mathbb{R}^{S \times D}$, D is dimension of input and S is number of samples
Output: Low dimensional dataset $X \in \mathbb{R}^{S \times B}$, B is desired dimension
1: Choose an intermediate dimension k_1 for Random projection
2: Create the random matrix $R^{D \times k_1}$ using GRP and SRP
3: Perform random projection and create intermediate data $X^{S \times k_1}$
4: **for** $i = 2$ to 30 **do**
5:     Perform nonlinear DR techniques to reduce data as $X^{S \times i}$
6:     Find percentage of cumulative eigen value for each i
7: **end for**
8: choose $B = min(i)$, which having cumulative eigen value $\geq 95\%$
9: **return** $X^{S \times B}$

3.4 RESULTS AND DISCUSSIONS

3.4.1 Experimental Setup

The proposed method is evaluated by experiments in universally available hyperspectral image datasets. All executions are done on Intel(R) Xeon(R) Silver 4114 CPU @ 2.24 GHz with a RAM of 196 GB under CentOS Linux release 7.4.1708 (Core) using python3 programming implementation. Three hyperspectral datasets are used to evaluate the performance of the proposed method and state-of-the-art techniques. They are Indian pines, Pavia University, and Salinas. These datasets and their dimentions, number of samples, and class lables are detailed in Section 2.6.1.

3.4.2 Performance Evaluation Measures

This section describes three evaluation measures, namely overall accuracy(OA), the area under ROC curve(AUC), and execution time(ET) used for comparing the performance of hybrid dimensionality reduction techniques with classic nonlinear techniques. A detailed explanation of evaluation measures as follows:

Overall accuracy: Overall accuracy is a measure that states how many samples are correctly mapped to their corresponding class. It is the most comfortable measure to find the performance of a classifier. Consider a dataset that has N number of samples and C class labels, then the confusion matrix M of classification is a square matrix of size $C \times C$. Overall accuracy is measured using the equation

$$OA = \frac{\sum_{i=1}^{C} M_{ii}}{N} \tag{3.15}$$

Diagonal elements M_{ii} in the confusion matrix is the number of samples correctly classified, and the overall accuracy of a classifier is always measured on a percentage scale.

Area under ROC curve: The area under the ROC curve or precisely area under the curve in classification performance evaluation is an aggregate measure over all possible classification thresholds. It was used when the model ranks random positives over random negatives. It measures how well the prediction ranked than its absolute value; therefore, AUC is scale-invariant. AUC value ranges from 0 to 1.

Execution Time: The time taken for the classification of different dimensional data is measured in seconds. This paper mainly focuses on reducing the computational complexity of nonlinear DR techniques. Thus ET places a significant role in this performance evaluation.

3.4.3 Result Analysis

As described in section 3.3, the proposed algorithm is a hybrid band extraction method. The proposed method first performs Random projection on the input datasets to reduce the dimension from D to k_1 and then performs various nonlinear dimensionality reduction techniques in this k_1 dimensional space.

In Random projection, the projection matrix may be either dense or sparse. According to the JL transform, the desired dimension depends on the number of samples. Considering the datasets mentioned above, $log(N)$ value for all the datasets is more significant than their original dimension D. Thus the desired size for the random projection matrix is chosen based on the value of ϵ. The desired dimension B was identified for different ϵ values. The B value obtained for the minimum ϵ is selected and found that the value is nearly equal to $D/2$ for all the datasets under evaluation. Therefore the value of B is set to $D/2$ and represents the data using B Gaussian mixture elements. Now compare the hybrid model using both dense and sparse projection results for performance evaluation.

The first phase of the proposed band extraction technique is to apply random projection to reduce the number of bands from D to $D/2$. The datasets IP, PU, SA having 200, 103, and 204 bands were used as the input for the band extraction phase. The random projection technique is carried out according to JL Lemma and ϵ value, and the number of bands in the input dataset reduced to 100, 52, and 102 for IP, PU, and SA. The random projected output is fed as an input to different nonlinear DR techniques. The cumulative eigenvalues for different hybrid band extraction techniques were computed and listed out in tables 3.1, 3.2, and 3.3

Table 3.1: Percentage of cumulative Eigen values of principal components in various DR techniques for IP Dataset

No of PCs	GRP+KPCA	GRP+LLE	GRP+ISOMAP	GRP+SSNG	SRP+KPCA	SRP+LLE	SRP+ISOMAP	SRP+SSNG
2	63.18	63.02	61.15	62.65	60.58	60.79	58.32	56.10
4	73.74	71.45	72.26	69.86	67.15	69.77	65.12	67.63
6	81.54	81.33	82.56	80.33	79.34	73.23	78.33	79.75
8	88.62	87.34	82.69	86.35	82.01	73.68	81.21	81.55
10	91.97	90.22	83.01	87.42	83.76	77.49	83.78	83.74
12	94.24	91.56	86.59	89.07	84.11	81.23	83.90	84.00
14	94.86	92.08	89.90	93.79	85.73	81.92	85.57	86.71
16	**95.92**	93.94	91.84	94.25	87.22	83.70	86.33	88.43
18	97.60	94.41	93.78	**95.47**	91.93	85.06	89.21	91.58
20	98.15	**95.11**	**95.23**	95.94	94.20	89.73	89.96	94.17
22	98.41	95.66	95.60	96.73	**95.17**	92.56	93.20	**95.03**
24	98.88	96.19	96.07	97.09	96.07	**95.34**	**95.32**	95.11
26	98.91	97.02	96.87	97.95	96.09	95.79	95.39	95.78
28	98.99	97.39	97.14	98.29	96.32	96.09	96.10	96.22
30	99.01	97.77	97.88	98.63	96.49	96.27	96.98	96.49

Table 3.2: Percentage of cumulative Eigen values of principal components in various DR techniques for PU Dataset

No of PCs	GRP+KPCA	GRP+LLE	GRP+ISOMAP	GRP+SSNG	SRP+KPCA	SRP+LLE	SRP+ISOMAP	SRP+SSNG
2	72.14	71.18	71.66	72.29	68.27	66.97	65.22	65.76
4	77.23	72.24	75.65	73.49	69.49	67.17	66.30	65.88
6	82.12	77.79	78.06	76.90	74.16	71.08	69.43	71.31
8	87.13	81.46	83.19	82.76	79.03	73.26	74.77	73.56
10	93.56	84.33	86.61	87.92	81.87	75.16	78.90	79.28
12	94.23	89.04	89.94	93.27	83.05	82.92	83.55	83.66
14	**95.39**	93.21	94.70	**95.02**	89.83	85.57	89.09	87.52
16	97.06	**95.11**	**95.32**	95.57	93.97	91.53	92.25	91.67
18	98.19	95.74	95.86	95.91	**95.23**	93.81	94.67	94.01
20	98.47	96.31	96.07	96.68	95.46	**95.36**	**95.21**	**95.07**
22	99.12	96.82	97.75	97.01	95.79	95.59	96.28	95.94
24	99.38	97.54	98.08	98.87	97.32	96.27	97.45	97.13
26	99.56	98.20	98.05	99.03	98.67	98.22	98.88	97.98
28	99.61	99.11	99.14	99.26	99.10	99.18	99.06	99.02
30	99.72	99.51	99.46	99.59	99.42	99.78	99.32	99.74

Table 3.3: Percentage of cumulative Eigen values of principal components in various DR techniques for SA Dataset

No of PCs	GRP+KPCA	GRP+LLE	GRP+ISOMAP	GRP+SSNG	SRP+KPCA	SRP+LLE	SRP+ISOMAP	SRP+SSNG
2	74.15	74.37	73.27	69.09	67.30	65.52	63.18	64.09
4	79.23	76.09	74.12	70.00	69.93	66.47	67.70	67.86
6	82.23	81.26	77.71	71.13	73.42	71.18	74.41	72.39
8	88.24	83.74	82.20	81.25	79.09	75.53	76.26	75.57
10	93.89	88.42	86.61	84.44	83.75	79.91	79.05	79.09
12	94.98	91.97	89.99	87.83	86.62	85.57	83.76	83.64
14	**95.69**	92.48	92.15	91.19	89.01	87.92	85.24	86.03
16	96.78	93.17	94.44	**95.13**	92.19	91.11	89.99	89.32
18	98.12	**95.13**	**95.04**	95.56	94.16	93.34	91.07	91.11
20	99.01	96.34	95.42	96.66	**95.36**	94.49	92.67	**95.28**
22	99.34	97.71	96.07	97.14	95.51	**95.47**	**95.38**	95.79
24	99.56	97.94	96.46	97.99	96.42	96.84	96.10	96.34
26	99.62	98.46	97.14	98.16	97.98	96.97	97.77	97.05
28	99.67	98.72	98.08	98.87	98.58	97.41	98.08	98.61
30	99.69	99.11	98.82	98.91	99.02	98.88	98.84	98.97

Table 3.1 lists the percentage of cumulative eigenvalues for IP datasets for various principal components. The analysis on the table indicates that the cumulative eigenvalue of GRP-based band extraction techniques is higher than that of SRP-based band extractions. GRP+KPCA produces more than 95 percent of cumulative eigenvalue for a minimum of 16 principal components. GRP+LLE and GRP+ISOMAP meet the threshold cumulative eigenvalue while considering 20 principal components. Although LLE, ISOMAPGRP, and SSNG work based on neighborhood graph concepts, GRP+SSNG performed better and meet threshold conditions at 18 principal components. While analyzing the table 3.1, it is found that SRP+KPCA and SRP+SSNG perform similarly and took 22 principal components to meet the criteria. Similarly, The performance of SRP+LLE and SRP+ISOMAP are alike and meets the criteria in 24 principal components.

The percentage of cumulative eigenvalues generated by different non-linear DR techniques for the PU dataset is given in table 3.2. Even though the SRP is computationally easier than GRP, the sparse nature of SRP leads to limited performance. Therefore, the GRP-based band extraction techniques meet the threshold criteria with minimal principal components for the PU dataset. GRP+KPCA and GRP+SSNG took only 14 principal components to achieve the required threshold, whereas GRP+LLE and GRP+ISOMAP took 16 principal components. All graph-based techniques on sparse projected data tend to perform similarly and meet the criteria at 20 number of principal components. SRP+KPCA achieved threshold criteria in 18 principal components.

Table 3.3 displays the percentage of cumulative eigenvalues for SA datasets for various principal components. The results obtained are very similar to that of the IP dataset because both datasets have spatially dependent pixel classes. The Gaussian random projected data meets the threshold criteria in the lower number of principal components compared to Sparse random projected techniques. 14, 18, 18, and 16 were the required principal components for GRP-based non-linear techniques: GRP+KPCA, GRP+LLE, GRP+ISOMAP, and GRP+SSNG. SRP+KPCA and SRP+SSNG required 20 principal components to meet the threshold, whereas the other two sparse projected techniques took 22 components.

Table 3.4: Evaluation parameters for various DR techniques in IP dataset

Technique	No of PCs	OA	AUC	ET
KPCA	22	85.79	0.71	85.79
LLE	18	75.79	0.72	102.32
ISOMAP	18	77.15	0.73	118.72
SSNG	16	**86.16**	0.72	97.43
GRP+KPCA	16	78.14	**0.75**	44.61
GRP+LLE	20	74.80	0.71	73.42
GRP+ISOMAP	20	74.37	0.71	79.27
GRP+SSNG	18	85.92	0.72	56.48
SRP+KPCA	22	65.06	0.62	**44. 54**
SRP+LLE	24	72.24	0.69	61.53
SRP+ISOMAP	24	73.25	0.70	62.33
SRP+SSNG	22	80.32	0.68	48.75

Table 3.4 , 3.5, 3.6 displays evaluation parameters for various state of the art non-linear dimensionality reduction techniques and the proposed hybrid techniques for the datasets IP, PU, and SA. Each table lists the number of principal components, OA, AUC, and ET, for each technique under study. The best values obtained for evaluation parameters are highlighted in bold letters. It is clear that the execution time for SRP-based techniques will always be lower than any other methods because of the sparse nature of the data. However, the values of OA and AUC will be the decisive factor in selecting an optimal DR method.

Table 3.4 represents the evaluation parameters obtained for the IP dataset when different DR techniques are applied. SRP+KPCA method was able to finish the DR in much lesser time(44.54), but they offered very low values for OA(65.06) and AUC(0.62). The data in the table indicate that SSNG produced a higher OA value but failed to produce better values for AUC and ET. Hence SSNG is not a satisfiable dimensionality reduction method. GRP+KPCA method was able to produce a better AUC value(0.75) with minimal execution time(44.61) using only 16 principal components.

The evaluation parameters obtained for the PU dataset are listed in table 3.5. Here also SRP+KPCA was able to finish in lesser time(40.82). However, they lack in classification performance. For this dataset, GRP+KPCA scored higher OA and AUC values with minimal running time considering 14 principal components. Even though SSNG

Table 3.5: Evaluation parameters for various DR techniques in PU dataset

Technique	No of PCs	OA	AUC	ET
KPCA	20	73.89	0.77	75.14
LLE	18	81.66	0.80	92.14
ISOMAP	16	84.13	0.79	91.46
SSNG	14	85.36	**0.81**	84.16
GRP+KPCA	14	**88.30**	**0.81**	50.53
GRP+LLE	16	79.35	0.74	62.18
GRP+ISOMAP	16	81.43	0.77	60.25
GRP+SSNG	14	83.48	0.79	67.48
SRP+KPCA	18	71.13	0.73	**40.82**
SRP+LLE	20	72.57	0.71	42.15
SRP+ISOMAP	20	73.59	0.75	42.67
SRP+SSNG	20	79.42	0.75	45.39

Table 3.6: Evaluation parameters for various DR techniques in SA dataset

Technique	No of PCs	OA	AUC	ET
KPCA	16	80.549	0.82	85.95
LLE	16	83.54	0.79	88.46
ISOMAP	16	85.52	0.79	88.37
SSNG	16	88.58	0.81	87.99
GRP+KPCA	14	83.16	**0.83**	56.23
GRP+LLE	18	82.34	0.77	61.43
GRP+ISOMAP	18	83.39	0.78	62.58
GRP+SSNG	16	**86.48**	0.80	67.31
SRP+KPCA	20	76.22	0.81	**48.19**
SRP+LLE	22	80.32	0.77	51.34
SRP+ISOMAP	22	81.38	0.76	53.48
SRP+SSNG	20	82.64	0.78	54.19

produced a similar AUC value(0.81) similar to GRP+KPCA, the execution time was much higher(84.16). Hence the best DR technique in this comparison is GRP+KPCA. Analyzing evaluation parameters for the SA dataset listed in table 3.6, GRP+SSNG yields a better overall accuracy(86.48); however, they fail to meet a good AUC and ET value. SRP+KPCA executes in a minimal time(48.19), they still lack classification performance. The better DR technique is GRP+KPCA, as they achieve higher classification performance within minimal time.

Table 3.7 depicts the comparison of various hybrid DR techniques for the IP dataset in terms of overall accuracy. Its graphical representation is shown in 3.1. In terms

Table 3.7: Comparison of various DR techniques in terms of OA for IP dataset

No of bands	GRP+KPCA	GRP+LLE	GRP+Isomap	GRP+SSNG	SRP+KPCA	SRP+LLE	SRP+Isomap	SRP+SSNG
2	55.95	55.62	52.12	67.52	55.67	55.66	53.89	70.99
4	60.86	59.91	57.21	70.72	59.74	59.87	56.4	72.59
6	64.44	60.72	60.77	74.34	64.97	62.45	60.54	73.89
8	65.03	62.15	61.12	77.24	64.7	63.79	62.38	77.9
10	67.31	64.6	63.8	80.29	64.68	64.45	64.87	77.12
12	70.07	65.06	66.61	83.58	63.46	65.04	66.91	79.48
14	75.14	68.94	69.14	85.35	64.72	69.13	70.69	79.31
16	78.14	69.33	69.22	85.25	64.14	69.76	70.23	78.83
18	78.16	71.13	70.13	85.92	63.53	71.13	72.9	79.11
20	77.2	74.8	74.37	85.67	64.64	71.8	72.38	79.74
22	77.79	74.09	74.23	85.35	65.06	72.06	72.13	80.32
24	77.94	73.1	73.45	84.47	65.95	72.24	73.25	80.38
26	77.95	73.15	74.69	83.45	63.13	71.47	73.9	80.65
28	78	74.18	74.57	85.46	64.33	72.43	73.28	79.93
30	76.09	74.15	74.12	85.51	64.65	72.97	72.38	79.12
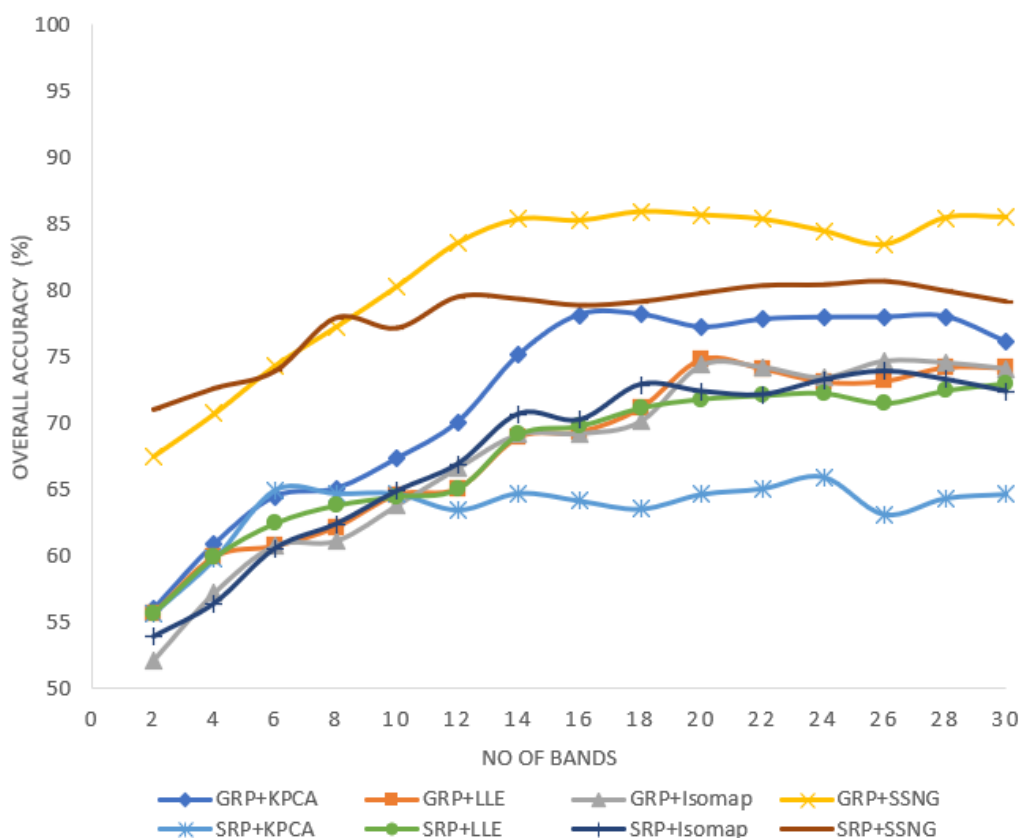
Figure 3.1: OA against number of bands for IP dataset using various DR techniques

of overall accuracy, the GRP+SSNG techniques show higher performance over other techniques. While analyzing the OA values of all techniques, it is found that all of them show a steady improvement up to 14 number of principal components, and then OA values tend to converge. The comparison of OA against the number of bands for the PU

Table 3.8: Comparison of various DR techniques in terms of OA for PU dataset

No of bands	GRP+KPCA	GRP+LLE	GRP+Isomap	GRP+SSNG	SRP+KPCA	SRP+LLE	SRP+Isomap	SRP+SSNG
2	73.96	74.23	73.01	81.11	64.61	65	68.25	65.13
4	74.49	76.91	74.81	81.72	65.37	67.35	65.13	68.34
6	75.54	78.24	76.75	82.21	66	69.76	69.11	67.45
8	78.58	78.15	77.34	82.98	66.23	70.15	69.92	70.9
10	82.21	78.47	79.74	82.46	67.78	70.24	70.87	71.12
12	85.53	79.11	80.34	83.23	68.35	71.12	71.21	71.95
14	88.3	79.32	80.32	83.48	68.63	71.35	71.94	73.31
16	87.81	79.35	81.43	83.99	69.22	72.56	71.23	76.83
18	87.45	79.01	80.92	83.35	71.13	72	72.49	78.11
20	88.23	79.17	81.22	83.67	71.07	72.57	73.59	79.42
22	88.74	78.33	81.88	82.45	70.45	72.11	73.13	79.56
24	87.01	77.1	81.14	82.64	71.5	71.25	73.25	79.22
26	88.95	78	80.69	82.81	71.45	72.87	72.9	79.55
28	88.1	77.12	80.98	83.46	71.9	71.53	73.28	79.35
30	88.09	78.49	81.87	83.51	70.64	71.97	73.38	78.12

dataset is represented in Table 3.8. Figure 3.2 shows the relationship between the number of bands and overall accuracy. The performance of GRP+KPCA dominates over other hybrid DR techniques in terms of OA. Even though GRP+SSNG performs better than other techniques for a lower number of bands, they lack performance improvement at higher stages.
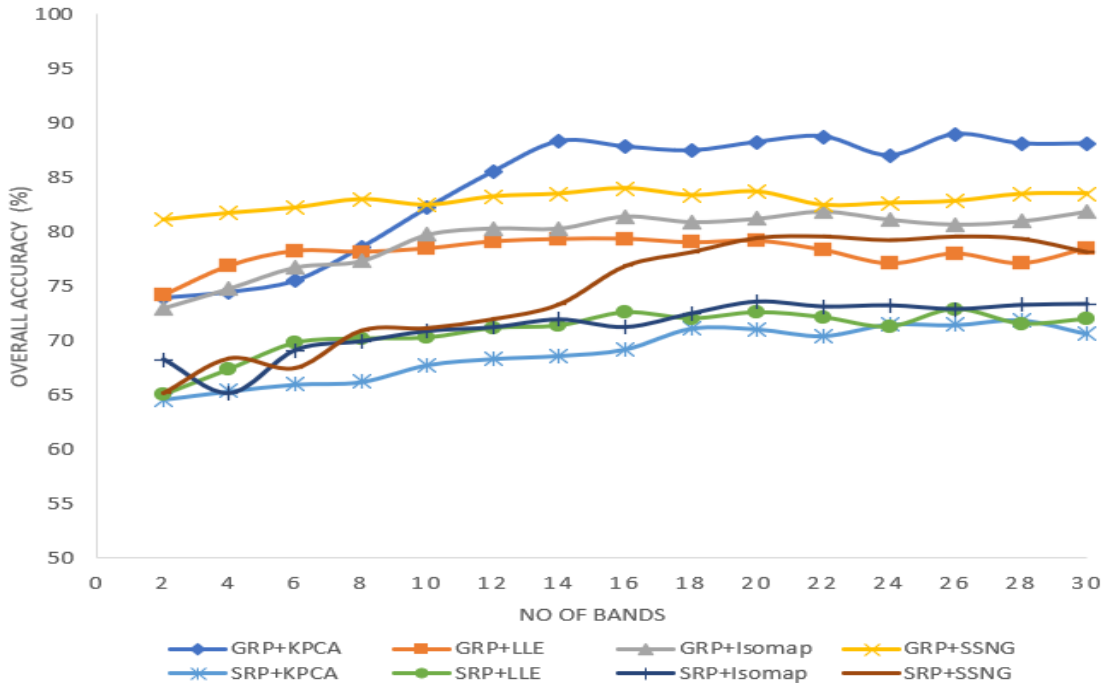
Figure 3.2: OA against number of bands for PU dataset using various DR techniques

The analysis of OA against the number of bands for the SA dataset is given in Table 3.9. The relationship between the number of bands and overall accuracy is pictured in

Table 3.9: Comparison of various DR techniques in terms of OA for SA dataset

No of bands	GRP+KPCA	GRP+LLE	GRP+Isomap	GRP+SSNG	SRP+KPCA	SRP+LLE	SRP+Isomap	SRP+SSNG
2	75	75.73	72.21	83.52	73.67	73.33	73.25	74.99
4	75.48	79.22	75.22	83.72	75.54	72.22	73.4	74.59
6	76.54	80.24	77.46	84.97	74.73	74.24	74.54	76.88
8	78.63	81.15	78.34	85.98	74.28	74.57	75.38	77.9
10	80.31	81.6	80.35	85.61	74.33	78.55	77.87	78.12
12	81.02	81.34	81.12	85.23	75.46	79.47	79.91	79.95
14	83.16	81.32	81.42	86.35	75.72	79.15	80.69	80.31
16	82.99	81.33	83.22	86.48	75.14	79.33	80.23	80.83
18	82.16	82.34	83.39	86.35	76.53	80.46	81.9	81.11
20	83.23	82.12	83.31	86.67	76.22	80.56	81.38	82.64
22	83.74	82.93	83.12	86.35	76.95	80.32	81.38	82.77
24	83.94	81.1	82.45	85.47	76.95	80.25	81.25	81.38
26	82.95	82.14	83.69	85.45	75.13	79.47	80.9	82.65
28	83	81.32	83.57	86.46	76.33	80.43	81.28	82.93
30	83.09	81.49	83.11	86.51	76.65	80.97	81.38	82.12
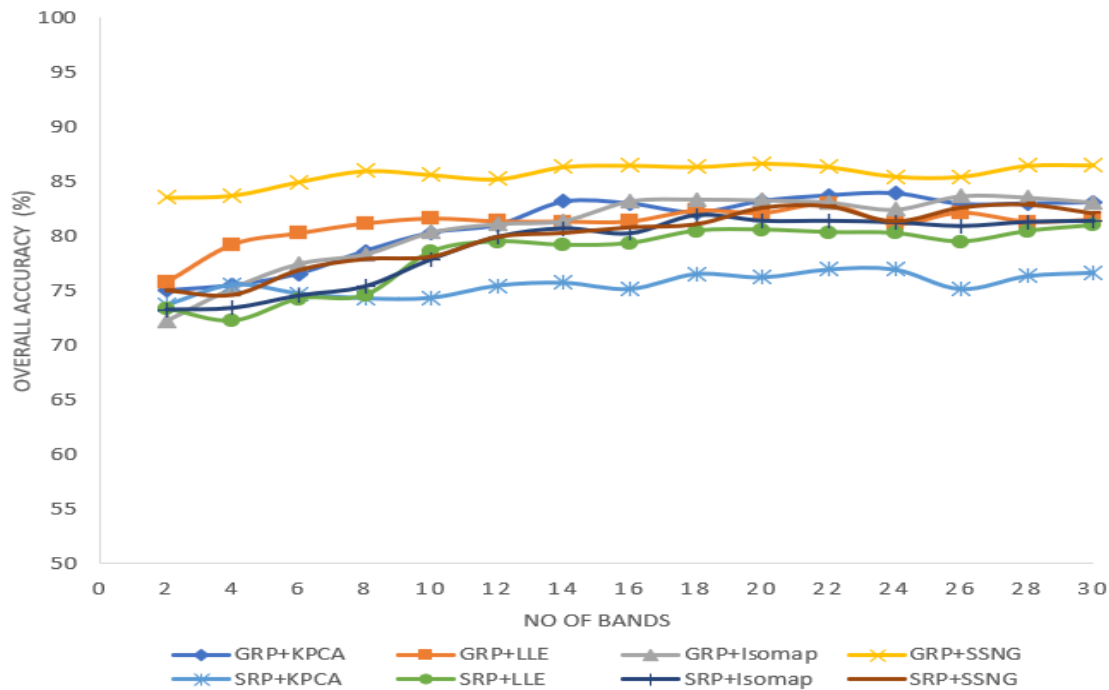
Figure 3.3: OA against number of bands for SA dataset using various DR techniques

Table 3.10: Comparison of various DR techniques in terms of AUC for IP dataset

No of bands	GRP+KPCA	GRP+LLE	GRP+Isomap	GRP+SSNG	SRP+KPCA	SRP+LLE	SRP+Isomap	SRP+SSNG
2	0.51	0.52	0.51	0.65	0.52	0.53	0.52	0.63
4	0.59	0.57	0.55	0.67	0.54	0.54	0.55	0.63
6	0.62	0.61	0.57	0.68	0.53	0.54	0.59	0.65
8	0.62	0.6	0.59	0.69	0.54	0.57	0.6	0.66
10	0.64	0.61	0.61	0.68	0.57	0.58	0.6	0.65
12	0.68	0.62	0.63	0.7	0.59	0.59	0.63	0.66
14	0.72	0.63	0.66	0.71	0.59	0.62	0.67	0.65
16	0.75	0.66	0.67	0.71	0.6	0.64	0.68	0.66
18	0.75	0.69	0.68	0.72	0.61	0.66	0.68	0.67
20	0.74	0.71	0.71	0.71	0.61	0.66	0.69	0.67
22	0.73	0.69	0.71	0.71	0.62	0.67	0.69	0.68
24	0.72	0.7	0.7	0.71	0.61	0.69	0.7	0.68
26	0.73	0.71	0.69	0.7	0.61	0.68	0.7	0.67
28	0.74	0.69	0.69	0.69	0.59	0.67	0.69	0.68
30	0.73	0.68	0.7	0.7	0.6	0.68	0.69	0.65
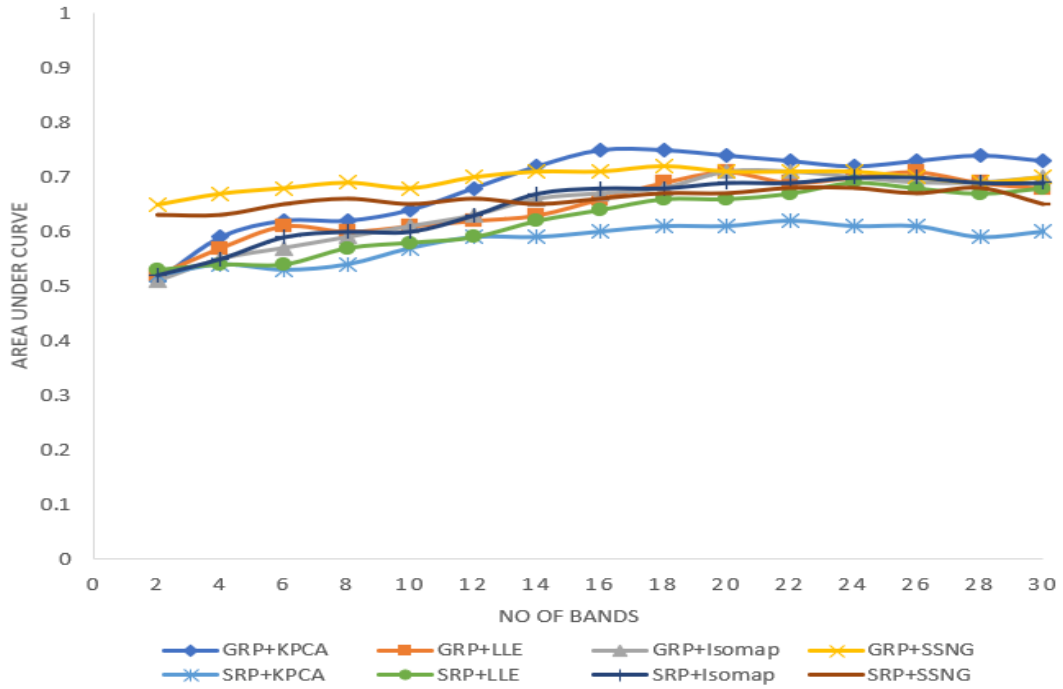
Figure 3.4: AUC against number of bands for IP dataset using various DR techniques

figure 3.3. Here also, GRP+SSNG outperformed the other techniques in terms of OA. Performance of all the techniques tends to converge at a higher number of bands(after 14).

The relationship between AUC and the number of bands in the proposed hybrid algorithms is analyzed using cross-validation. GRP+SSNG performed better than other techniques for the lower number of bands for the IP dataset. While in higher bands, GRP+KPCA outperformed and retained its AUC almost similar after 14 bands. Tabular

Table 3.11: Comparison of various DR techniques in terms of AUC for PU dataset

No of bands	GRP+KPCA	GRP+LLE	GRP+Isomap	GRP+SSNG	SRP+KPCA	SRP+LLE	SRP+Isomap	SRP+SSNG
2	0.72	0.71	0.72	0.72	0.67	0.65	0.65	0.63
4	0.73	0.71	0.74	0.73	0.67	0.64	0.65	0.64
6	0.75	0.73	0.75	0.73	0.68	0.65	0.66	0.67
8	0.76	0.72	0.75	0.76	0.69	0.66	0.66	0.67
10	0.78	0.72	0.76	0.76	0.69	0.66	0.69	0.68
12	0.78	0.73	0.75	0.77	0.7	0.68	0.7	0.7
14	0.81	0.73	0.76	0.79	0.7	0.69	0.69	0.69
16	0.8	0.74	0.77	0.78	0.71	0.7	0.72	0.73
18	0.81	0.74	0.75	0.78	0.73	0.7	0.73	0.74
20	0.81	0.74	0.76	0.77	0.73	0.71	0.75	0.75
22	0.81	0.72	0.76	0.78	0.72	0.71	0.75	0.75
24	0.81	0.73	0.76	0.77	0.72	0.7	0.73	0.75
26	0.8	0.73	0.75	0.79	0.73	0.69	0.73	0.74
28	0.8	0.73	0.76	0.78	0.71	0.71	0.74	0.73
30	0.81	0.72	0.75	0.78	0.72	0.71	0.74	0.74
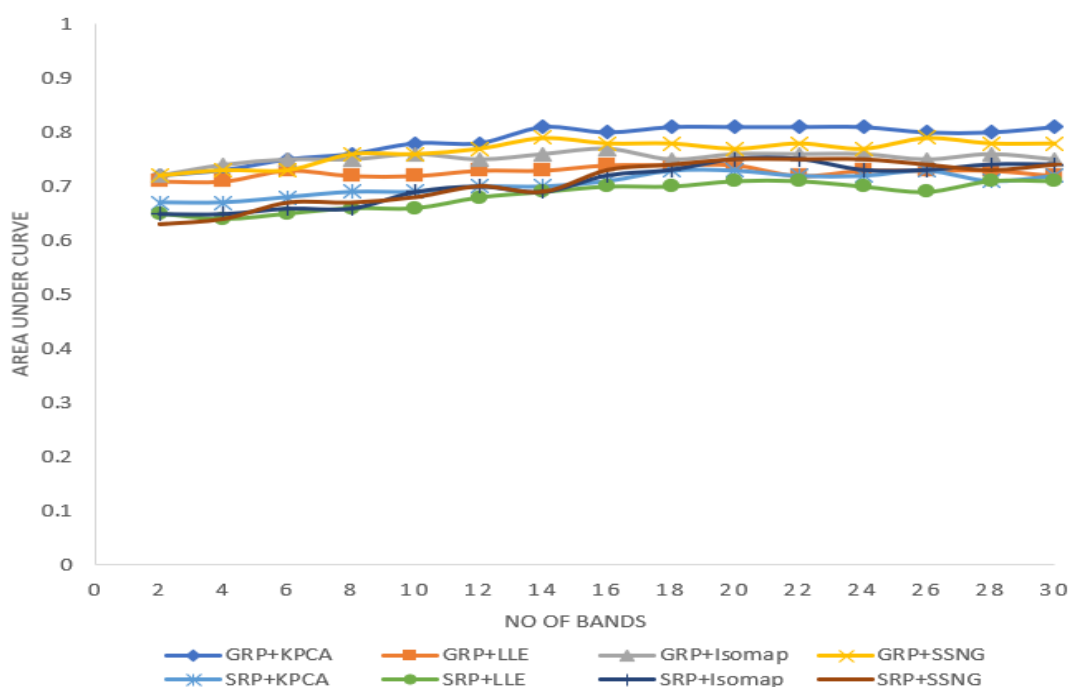
Figure 3.5: AUC against number of bands for PU dataset using various DR techniques

and graphical representation of the number of bands versus AUC for IP dataset is shown in table 3.10 and figure 3.4.

The comparison of various hybrid DR techniques in terms of AUC for the PU dataset is shown in table 3.11. AUC for GRP+KPCA and GRP+SSNG shows a similar pattern of performance for the PU dataset. GRP+KPCA improves its performance after eight bands and shows a slight increase than other techniques. The graphical representation of this comparison is picturized in figure 3.5.

Table 3.12: Comparison of various DR techniques in terms of AUC for SA dataset

No of bands	GRP+KPCA	GRP+LLE	GRP+Isomap	GRP+SSNG	SRP+KPCA	SRP+LLE	SRP+Isomap	SRP+SSNG
2	0.72	0.71	0.71	0.75	0.72	0.65	0.72	0.74
4	0.74	0.73	0.71	0.76	0.74	0.68	0.72	0.74
6	0.74	0.73	0.72	0.76	0.74	0.68	0.73	0.74
8	0.79	0.74	0.74	0.76	0.75	0.7	0.72	0.75
10	0.78	0.74	0.73	0.77	0.75	0.72	0.74	0.77
12	0.81	0.76	0.76	0.79	0.77	0.71	0.75	0.76
14	0.83	0.75	0.76	0.79	0.77	0.73	0.74	0.76
16	0.82	0.76	0.77	0.8	0.78	0.73	0.74	0.77
18	0.8	0.77	0.78	0.8	0.78	0.76	0.75	0.77
20	0.81	0.77	0.78	0.8	0.81	0.75	0.75	0.78
22	0.82	0.77	0.77	0.79	0.81	0.77	0.76	0.77
24	0.83	0.75	0.77	0.79	0.8	0.77	0.76	0.78
26	0.83	0.76	0.77	0.78	0.8	0.76	0.76	0.78
28	0.82	0.77	0.76	0.8	0.8	0.77	0.74	0.78
30	0.81	0.77	0.78	0.8	0.8	0.77	0.75	0.76
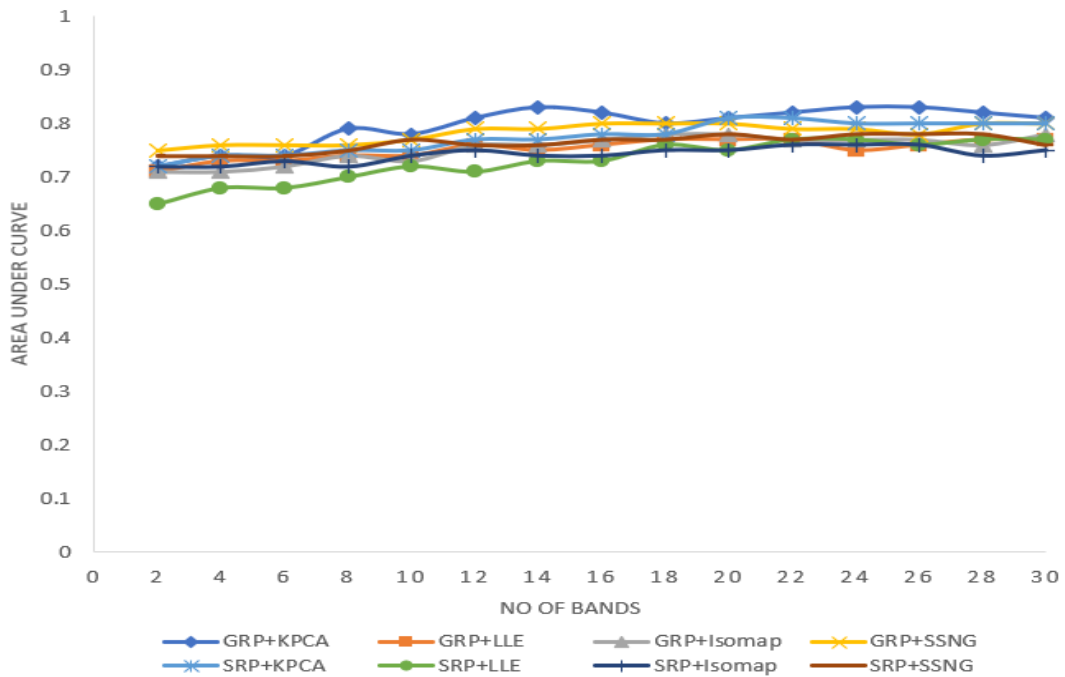
Figure 3.6: AUC against number of bands for SA dataset using various DR techniques

The effect of AUC for the number of bands in the SA dataset is almost similar to that of the IP dataset. Here, GRP+SSNG gives a higher value in lower bands and then GRP+KPCA overshoots. For 10 and 16 band-reduced data, both show the same performance, and in mid and higher bands, GRP+KPCA works better. Table 3.12 and figure 3.6 are the representation of AUC against number of bands for SA dataset.

| (a) IP groundtruth | (b) KPCA | (c) GRP+KPCA | (d) SRP+KPCA |
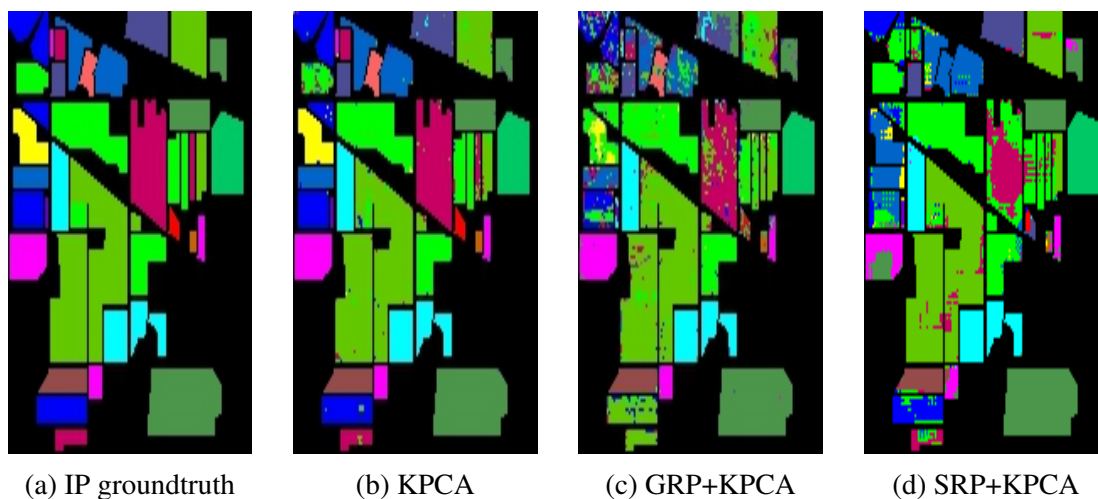
Figure 3.7: Classification map for Indian pines dataset

| (a) PU groundtruth | (b) KPCA | (c) GRP+KPCA | (d) SRP+KPCA |
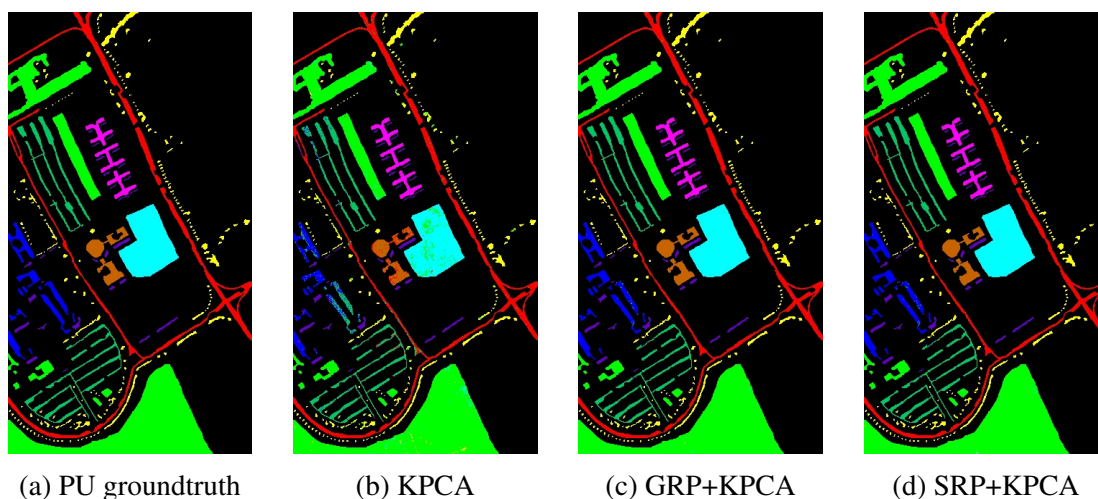
Figure 3.8: Classification map for Pavia University dataset

Figure 3.7, 3.8, and 3.9 picturizes the classification maps obtained for IP, PU, and SA dataset while KPCA, GRP+KPCA, and SRP+KPCA applied. GRP+KPCA produced a better result in terms of OA and AUC, and the same technique was able to complete nonlinear dimensionality reduction within a lesser time.

| (a) SA groundtruth | (b) KPCA | (c) GRP+KPCA | (d) SRP+KPCA |
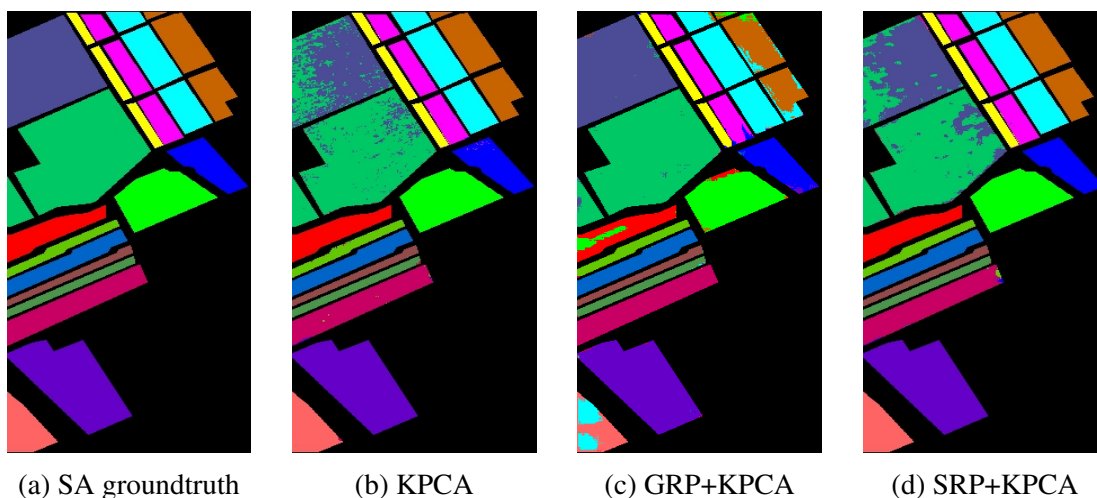
Figure 3.9: Classification map for Salinas dataset

3.5 SUMMARY

This chapter mainly focused on the curse of dimensionality issue in hyperspectral images. A hybrid dimensionality reduction technique is introduced as a solution to reduce the computational complexity of state-of-the-art nonlinear band extraction techniques. This DR technique uses the advantage of the random projection concept in band extraction. The application of Gaussian random projection before the nonlinear technique thereby reduces the computational complexity.

The performance of the proposed technique is evaluated against other renowned DR techniques. Various nonlinear DR techniques KPCA, LLE, Isomap, and SSNG, are tried over gaussian and sparse random projection, and its classification performance and execution time are evaluated. The research work concluded that GRP+KPCA is the best hybrid DR technique for the given group of datasets regarding classification performance and execution time. All the datasets give maximum performance in dimensionally reduced data having 14 or more dimensions. This data is used as the input for the HSI classification model introducing in chapter 4.

CHAPTER 4

HYBRID CNN BASED HSI CLASSIFICATION

4.1 INTRODUCTION

Classification of hyperspectral images is a challenging task. Accurate mapping of individual pixels in the hyperspectral images is highly desirable for various applications such as crop area identification, land change detection, drought monitoring, etc. Chapter 2 discussed different available classification techniques used for HSI classification. Deep learning-based classification techniques' evolution boosts performance and accuracy compared to conventional machine learning techniques. The methods that use spatial and spectral features for classification outperform those using either one of these features. The curse of dimensionality retard the performance of HSI classification. Therefore all recent classification techniques use dimensionality reduction before classification.

The proposed Hybrid CNN classification model is designed to incorporate both spatial and spectral features to maximize classification accuracy. The technique introduces a dedicated module for multiscale feature extraction.

The contributions of the chapter are :

- Multiscale spatial and spectral feature extraction using 3D-CNN with different window size.

- The usage of 2D-CNN on top of 3D-CNN to reduce the computational complexity and learn more spatial features.

- The proposed Hybrid CNN model extract more spatiospectral features and increase the classification accuracy when compared with existing deep learning models.

The chapter is organized as follows: Section 4.1 is for introduction, Section 4.2 briefs the convolutional neural network and its major layers. Section 4.3 describes the proposed hybrid CNN model and its block diagram, and Section 4.4 details the experimental setup. Section 4.5 presents the results and its detailed analysis using various evaluation parameters, and Section 4.6 summarizes the proposed method and its significance.

4.2 CONVOLUTIONAL NEURAL NETWORKS

Deep learning techniques for image classification is prevalent recently. The basics of CNN is highly related to the structure of the human visual system. CNN is a multilayer neural network consist of a convolution layer, pooling layer, and fully connected layer. CNN uses local connections to extract spatial features from 2-D images. CNN also provides a facility of weight-share mechanism to adjust the parameters that define the network's performance. Typically, convolution layer is the input layer of a CNN model, and this layer performs the convolution of image patches with a set of kernels. Convolution is a dot product operation between two matrices, namely receptive field and kernel (learnable parameters). Generally, the kernel is spatially smaller than that of input data, and the kernel is sliding over the receptive field and produces a feature map of the input data. The pooling layer comes next, and this layer is intended to perform the minimization of the extracted feature map. The pooling layer helps to gather more comprehensive and abstract features from the output of the convolution layer. It reduces the spatial dimension of the feature map and replaces a set of outputs into a single value based on the statistics of its nearby feature values. The most commonly used pooling technique is max pooling, and it replaces a set of features by its maximum. The next layer in a CNN architecture is the fully connected layers. The fully connected layer is a multilayer perceptron in which all neurons are connected to every succeeding layer neuron. This layer is used to map features into the output. The detailed descriptions of

these layers are given below.

Convolutional Layers: Convolution layers are the receptive fields in the general CNN architecture. Convolution layers extract the feature maps by convolving the input image blocks with several learnable filters. A mathematical representation of the operations in convolutions layers will deliver further clarity. The individual neuron's output for inputs x is calculated as

$$y = f(w * x + b) \tag{4.1}$$

Here w is the filter weight, and b is bias. $f(.)$ stands for the nonlinear activation applied on a weighted sum of input.

Consider an image input X having dimensions $m \times n \times d$. Here $m \times n$ refers to the spatial resolution, and d denotes the number of bands in the input. Consider x_i as the ith feature map of the input$X$. The convolution layer output y_j is given as

$$y_j = \sum_{i=1}^{d} f(x_i * w_j + b_j), j = 1, 2, \ldots k \tag{4.2}$$

The weight w_j and bias b_j represent the j th kernel among the total k kernels, $*$denotes the convolution operator, and $f()$ denotes the activation function.

2D-CNN model convolves the input data using a two-dimensional kernel before activation. This helps to capture spatial features from the input image. Reframing the equation (4.1) for 2D-convolution output of each neuron as

$$y_{mn} = f(\sum_{r} \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} k_{ij} x_{(i+m)(j+n)} + b_{mn}) \tag{4.3}$$

Here y_{mn} is the feature extracted at position (m,n), k is the 2D-convolution kernel of size $h \times w$. In the case of a 2D image, this convolution process is performed over all the feature maps r in the receptive field and took the sum of all values for nonlinear activation. This process repeated for all layers in the case of hyperdimensional data. When the data is three dimensional (e.g., video, color images, hyperspectral or multispectral images), these have both spatial and spectral or temporal dimensions. 2D convolution fails for these input data because it captures only spatial features. 3D-CNN is the modified form of 2D-CNN models, which perform three-dimensional convolution instead of 2D convolution in 2D-CNN. 3D convolution helps to extract both spatial and spectral features from three-dimensional spectral images. The feature extracted from 3D-CNN

model is:

$$y_{mnp} = f(\sum_{r}\sum_{i=0}^{h-1}\sum_{j=0}^{w-1}\sum_{l=0}^{b-1} k_{ijl}x_{(i+m)(j+n)(l+p)} + b_{mnp}) \tag{4.4}$$

Here the kernel k is three dimensional and features calculated by performing 3D convolution over 3D input data. The most commonly used activation function f used in convolutional neural network is ReLU (Rectified Linear Unit) (Krizhevsky et al. 2012). ReLU is faster than other activation functions while using gradient descent techniques for training. The equation for ReLU is

$$ReLU(y) = max(0, y) \tag{4.5}$$

Pooling Layers: Pooling layers are usually followed by convolution layers because the output of convolution layers is mostly redundant. The pooling layer performs some specific operations to reduce the feature redundancy. The introduction of a pooling layer helps to reduce the spatial size of the obtained feature map. The reduction in the spatial size of the feature map limits the number of parameters and reduces the complexity of the CNN. The pooling layer outputs a more abstract feature map. The average pooling operations z for a $n \times n$ windows size neighbour (S) is given below.

$$z = \frac{1}{F}\sum_{(i,j)\in S} x_{ij} \tag{4.6}$$

where the number of elements in S denoted by F, and the activation values for (i, j) is given by x_{ij}.

Fully Connected Layers: The fully connected layers receive the flattened output of the pooling layer. In fully connected layers, there exist a connection between all layer neurons to every succeeding layer neuron. The presence of such profound connections helps to extract more in-depth features. This layer maps the generated feature maps to outputs. The fully connected layer is given as

$$Y' = \sum_{i=1}^{C} f(WX' + b) \tag{4.7}$$

X, Y denotes the input and output, whereas W, b denotes the weight and bias of convolutional layers. Softmax is another activation function generally used in the last layer of a deep learning model for classification. Softmax produces the probability distribution of all the samples, and its sum is equal to one.

Different optimization techniques used in CNN for updating the weights using training data. The most commonly used optimization is gradient descent technique. Adam (adaptive momentum estimation) is another optimizer generally used for nonconvex problems (Kingma and Ba 2014). Adam maintains a separate learning rate for each weight value in the network based on the first and second moment of gradient. The parameters used in adam optimizer are learning rate (lr), beta1, and beta2. beta1 and beta2 are the exponential decay of the first and second moments, respectively.

4.3 PROPOSED CLASSIFICATION MODEL

The proposed classification model uses a hybrid DR technique (discussed in chapter 3), which is the combination of linear Gaussian Random Projection(GRP) and a nonlinear form of PCA(Kernel PCA). The computational complexity of the nonlinear technique diminished by applying GRP on input image X and thereby reducing the number of bands from D to $\lfloor D/2 \rfloor$. Then perform KPCA on this intermediate result and X's size change to $M \times N \times B$. Here the spatial dimension of input image retained after DR while the spectral dimension reduced from D to B, the desired number of bands. Ground truth of the input image Y is converted using one-hot encoding and represented as $Y = (y_1, y_2, \ldots, y_C)$, where C denotes the number of classes present in the input image.

Input image X is divided into a total of $M * N$ overlapping 3D patches of size $w \times w \times B$ by considering each pixel x_i as the central pixel of the patch where w denote the window size. 3D-CNN convolve each image patch using the 3D convolution kernel and extract the features having both spatial and spectral characteristics. A nonlinear activation function is applied to this convolved features. In the proposed model, spatial and spectral features are captured from each image patch using various window size, and these features fed into 2D-CNN for more spatial feature extraction. This hybrid model reduces the computational complexity of 3D-CNN as well as the number of parameters required for tuning the model.

The flow diagram of the proposed hybrid model is shown in Fig 4.1. In block-I, the proposed method consider three window size($w_1 = 15, w_2 = 13, w_3 = 11$) at a time
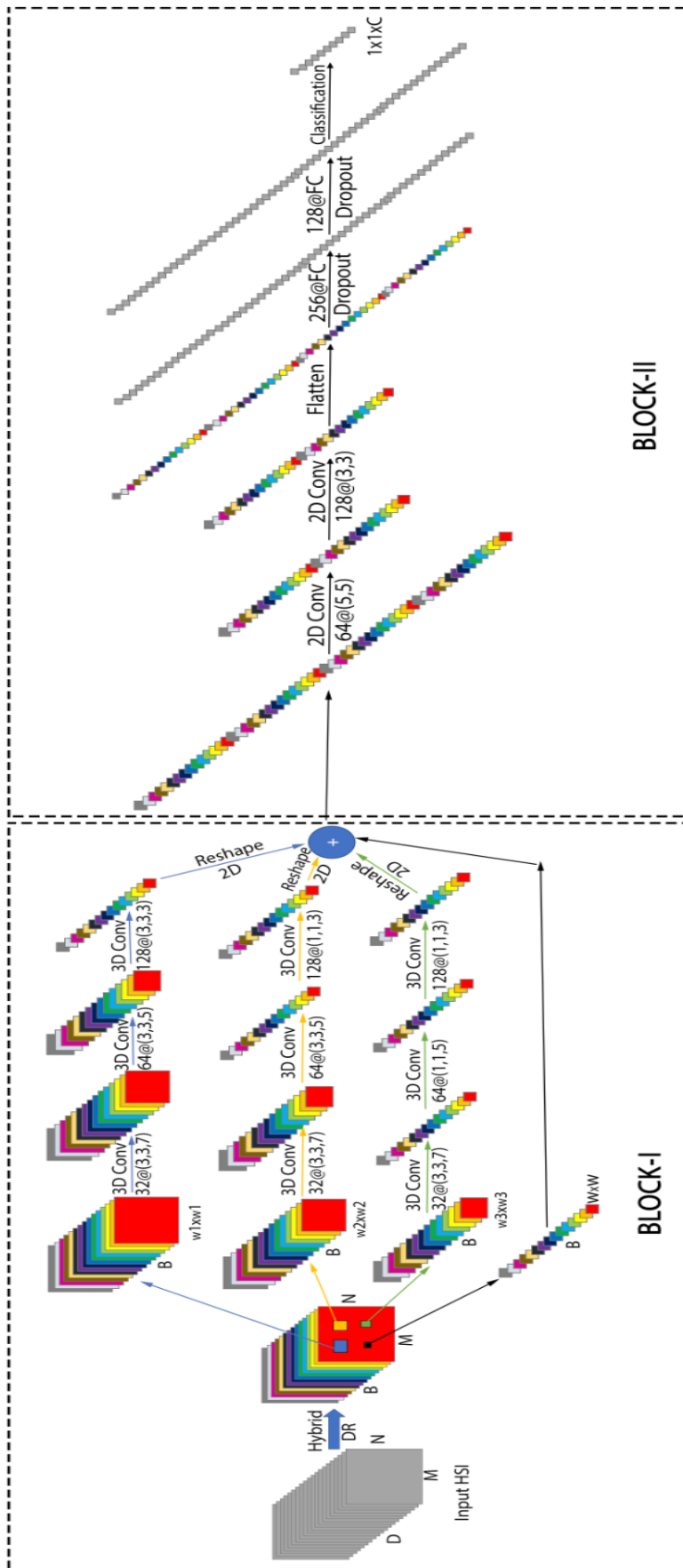
Figure 4.1: Proposed hybrid deep learning model for hyperspectral image classification

and create image patches. Thus we get $M * N$ image patches of size $w_i \times w_i \times B$ for each window size w_i. The convolution kernel for w_1 is $32 \times 3 \times 3 \times 7 \times 1$, $64 \times 3 \times 3 \times 5 \times 32$, $128 \times 3 \times 3 \times 3 \times 64$ for each of the convolution layer consecutively. Here $128 \times 3 \times 3 \times 3 \times 64$ means applying 128 3D- kernel of size $3 \times 3 \times 3$ on 64 image feature maps extracted from previous layer. Similarly $32 \times 3 \times 3 \times 7 \times 1$, $64 \times 3 \times 3 \times 5 \times 32$, $128 \times 1 \times 1 \times 3 \times 64$ and $32 \times 3 \times 3 \times 7 \times 1$, $64 \times 1 \times 1 \times 5 \times 32$, $128 \times 1 \times 1 \times 3 \times 64$ are the kernel size for w_2 and w_3 in respective convolution layers as shown in Fig 4.1. Valid padding is used in all the convolutional layers, thus after each convolution the dimension of feature map reduced. If the image patch of size $w \times w \times B$ is convolved using a three dimensional kernel of size $i \times j \times j$, then the size of feature map is:

$$(w - i + 1, w - j + 1, w - k + 1) \tag{4.8}$$

The feature map dimension of all window size coincide into a unique size of $128 \times W \times W \times BB$ where $W = 9$ and $BB = B - 12$ are the window size and number of bands of the feature map after three consecutive 3D-convolution.

The block-II of the proposed classification model consist of 2D- CNN and fully connected neural network. Outputs from previous blocks are reshaped into 2D maps of size $W \times W \times DD$, where $DD = 128 * BB$. The input image is now divided into image patches of size $W \times W \times B$ and concatenate the three 2D-feature maps in spectral dimension with this 2D image patch. Thus the feature map size for a 2D-convolution layer is $W \times W \times (B + DD)$. Two 2D-convolution layers are used in the proposed model with kernel size $64 \times 5 \times 5 \times (B + DD)$ and $128 \times 3 \times 3 \times 64$. Now the output of 2D-block contains a sufficient number of features extracted by considering both spatial and spectral properties of the input image without much loss in spectral features in less execution time. The feature learning phase of the proposed model executed in less time without much loss in spectral features.

The proposed model avoid pooling layer in CNN to maintain maximum spectral and spatial information. The extracted features are flattened and fed into dense fully connected layers for classification. Two fully connected layers are used in this model with 256 and 128 neurons in each layer. A dropout of 0.2% is applied in each fully

Table 4.1: Summary of the proposed model on Indian Pines dataset

Layer (type)	Output shape	# Parameters	Connected to
input_1 (InputLayer)	(15, 15, 15, 1)	0	
conv3d_1 (Conv3D)	(13, 13, 9, 32)	2048	input_1
conv3d_2 (Conv3D)	(11, 11, 5, 64)	92224	conv3d_1
conv3d_3 (Conv3D)	(9, 9, 3, 128)	221312	conv3d_2
op1 (Reshape)	(9, 9, 384)	0	conv3d_3
input_2 (InputLayer)	(13, 13, 15, 1)	0	
conv3d_4 (Conv3D)	(11, 11, 9, 32)	2048	input_2
conv3d_5 (Conv3D)	(9, 9, 5, 64)	92224	conv3d_4
conv3d_6 (Conv3D)	(9, 9, 3, 128)	24704	conv3d_5
op2 (Reshape)	(9, 9, 384)	0	conv3d_6
input_3 (InputLayer)	(11, 11, 15, 1)	0	
conv3d_7 (Conv3D)	(9, 9, 9, 32)	2048	input_3
conv3d_8 (Conv3D)	(9, 9, 5, 64)	10304	conv3d_7
conv3d_9 (Conv3D)	(9, 9, 3, 128)	24704	conv3d_8
op3(Reshape)	(9, 9, 384)	0	conv3d_9
input_4 (InputLayer)	(9, 9, 15)	0	
concatenate_1 (Con-catenate)	(9, 9, 1167)	0	op1, op2, op3, input_4
conv2d_1 (Conv2D)	(5, 5, 64)	1867264	concatenate_1
conv2d_2 (Conv2D)	(3, 3, 128)	73856	conv2d_1
flatten_1 (Flatten)	(1152)	0	conv2d_2
dense_1 (Dense)	(256)	295168	flatten_1
dropout_1 (Dropout)	(256)	0	dense_2
dense_2 (Dense)	(128)	32896	flatten_1
dropout_2 (Dropout)	(128)	0	dense_2
dense_3(Dense)	(16)	2064	dropout_2
Total parameters : 2,742,864			

connected layer to avoid overfitting. The output dimension and number of parameters used in each layer for the proposed model is shown in Table 4.1. ReLU activation function is used in all convolution layers and softmax is used for classification. Adam optimizer with loss function categorical-crossentropy having learning rate 0.0001 and decay $1e − 06$ used for optimatization. The training process repeats for 100 epochs of batch size 100 without any batch normalization and augmentation.

4.4 EXPERIMENTAL SETUP

Four hyperspectral datasets are used to evaluate the performance of proposed model and state of the art techniques. They are Indian pines(IP), Pavia University(PU), Salinas(SA) and Houston dataset2013 (IEEE GRSS Data Fusion Contest (Pacifici et al. 2013)).

The accuracy of CNN model depend on the sampling scheme (Lange et al. 2018) and the selection of training and testing samples may interlace when we choose the samples randomly. Hence, we selected the Houston dataset, in which the training and testing samples belongs to different area independent of spatial view. The Houston University dataset is captured by Compact Airborne Spectrographic Imager (CASI) having spatial resolution 2.5 meters in 2012 (Pacifici et al. 2013). The dataset consist of 144 bands of size 349×1905 and 15 class labels. The classes are Healthy grass, Stressed grass, Synthetic grass, Trees, Soil, Water, Residential,Commercial,Road, Highway, Railway, Parking Lot 1, Parking Lot 2, Tennis Court and Running Track.

4.5 RESULTS AND DISCUSSIONS

The availability of ground truth for hyperspectral images is very less. Therefore, the usage of a supervised deep learning model is not so adequate for hyperspectral image classification. However, our model is a supervised technique, and it uses the minimum amount of data for training the model. The IP, PU, and SA dataset were randomly divided into 20% training set, 10% validation set, and remaining 70% for testing, whereas the Houston dataset has specific training and validation samples given by IEEE GRSS Data Fusion Contest. The proposed method is compared against six state -of- the- art techniques starts from conventional SVM (Melgani and Bruzzone 2004) classifier to new deep learning techniques such as 2-D CNN (Chen et al. 2016), 3D-CNN (Chen et al. 2016), SSUN (Xu et al. 2018), SSRN (Zhong et al. 2018) and HybridSN (Roy et al. 2020).

4.5.1 Evaluation Parameters

To evaluate the classification performance of the proposed model against other state-of-the-art methods, we have considered the parameters Overall Accuracy (OA), Average Accuracy (AA), and Kappa statistics (K).

OA is the fraction of samples that are correctly classified and total test samples. The confusion matrix obtained from the classification is a square matrix of size $C \times C$, C is the number of class labels present in the dataset. The equation for calculating OA from the confusion matrix is

$$OA = \sum_{i=1}^{C} \frac{t_{ii}}{t} \tag{4.9}$$

t_{ii} denotes the number of correctly classified samples for class i, and t denotes total test samples. t_{ii} is obtained from the diagonal elements of the confusion matrix.

AA denotes the mean of class-wise accuracy, where class-wise accuracy is obtained from the confusion matrix as

$$CA_i = \frac{t_{ii}}{\sum_{j=1}^{C} t_{ij}} \tag{4.10}$$

CA_i is the class-wise accuracy for class i and t_{ij} denotes the number of samples of class i classified into class j, i.e., CA is the fraction of correctly classified samples for class i and the total number of test samples in the same class. Average accuracy is defined as

$$AA = \frac{\sum_{i=1}^{C} CA_i}{C} \tag{4.11}$$

The row and column sum of the confusion matrix represents the chance of agreement of classified result and the chance of agreement of actual result. The kappa statistic **?** is the difference between these two values. K value lies in between -1 and 1, and classification accuracy improves as K tends to 1. Let t_{i+} is the sum of row elements ($\sum_{j=1}^{C} t_{ij}$) and t_{+i} is the sum of column elements ($\sum_{i=1}^{C} t_{ij}$) in confusion matrix for class i. Kappa statistic value is calculated as

$$K = \frac{t \sum_{i=1}^{C} t_{ii} - \sum_{i=1}^{C} t_{i+} t_{+i}}{n^2 - \sum_{i=1}^{C} t_{i+} t_{+i}} \tag{4.12}$$

4.5.2 Result Analysis

The support vector machine (Melgani and Bruzzone 2004) is the most popular pattern recognition method used for multiclass classification. SVM provides high classifica-

tion accuracy, and its sensitivity is less concerning the number of training samples. Therefore, this work chooses SVM with radial basis function as one of the existing classification models for result comparison. The next state-of-the-art technique chooses for performance comparison is a 2D-convolutional neural network. 2D-CNN (Chen et al. 2016) model uses PCA as a preprocessing technique and chooses the first principal component. Input to the 2D-CNN model is $K \times K$ neighborhood of each pixel and extract the relevant features after each convolution and pooling. These learned features are fed into a linear regression classifier and find the class labels for each pixel. 2D-CNN models extract only local spatial features and to compare the proposed model with the spatio-spectral feature extracted technique, 3D-CNN (Chen et al. 2016) chose as another state-of-the-art model. 3D-CNN model chooses $K \times K \times B$ neighborhood for each pixel and performs three-dimensional convolution and pooling. Both 2d-CNN and 3D-CNN models use ReLU as activation function and mini-batch update technique for updating weights. The use of the dropout technique mitigates overfitting in both models.

Spectral-spatial unified network (SSUN) is a deep learning model used as another progressive model for comparison. SSUN integrates a spectral feature extraction, spatial feature extraction model, and a classifier into a unified network. Spectral features are extracted using an LSTM, and a multiscale CNN extracts spatial features. LSTM uses the number of time steps as 3 for all the datasets, and mini-batch optimization of batch 64 is used. The classification layer is a fully connected layer, which used 128 number of neurons. SSRN is similar to the 3D-CNN model, and it uses batch normalization and a 50% dropout for regularization of the training process. Training is continued up to 200 epochs with a batch size of 16, and residual blocks are used to avoid the decrease in classification accuracy with an increase of convolutional layers. HybridSN model is another spatiospectral feature-based classification technique used for result analysis with the proposed technique. HybridSN uses PCA as a dimensionality reduction technique and a combination of 3D-CNN and 2D-CNN used for feature extraction. The model used adam optimizer for training with a batch size of 256, and the training continues for 100 epochs.

Execution time is another parameter choose to compare the proposed method with existing technologies. The reduction of computational complexity is a factor in the design of deep learning models. Hybrid DR techniques reduce the number of computations and thus reduce the training time in feature extraction and classification. Tabular representation of evaluation parameters OA, AA, and K for different classification methods on IP, PU and SA datasets are listed in Table 4.2. Table 4.3 shows the training time and testing time for the state-of-the-art technique and proposed model on IP, PU and SA datsets. Table 4.4 list out the evaluation parameters and execution time for Houston dataset. Here training time is measured in minutes and testing time in seconds. Execution in dedicated GPU reduce the execution time and process the data faster than CPU.

Indian pines dataset shows the very lowest accuracy in SVM based classification. The average accuracy of SVM classification is less than 80%. i.e., the number of misclassifications is very high for each class. For the 2D-CNN model, 27×27 neighborhoods of each pixel in the first principal component is taken as input. Three convolution layers of kernel size 4×4, 5×5, and 4×4 with 32, 64, and 128 filters are used in this model. The first two layers perform pooling (2×2), and the second and third layers perform a 50% dropout for regularization. The average accuracy of the 2D-CNN model is better than that of the conventional SVM model, and it increased by almost 7%. Since the 2D-CNN model using only one principal component, the training time is lesser than that of SVM. 3D-CNN model also has three convolution layer and two pooling layers. The convolution kernel size of each layer is $4 \times 4 \times 32$, $5 \times 5 \times 32$ and $4 \times 4 \times 32$ with 128, 192 and 256 filters. 2×2 is performed after the first and second layers of convolution with a 50% drop out after the second and third layers. The training time of the 3D-CNN model is very high compared with all the techniques because this model is not performing any DR technique and used $27 \times 27 \times 200$ image patches for feature extraction. The classification accuracy of SSUN and SSRN are better than those of all the previous technique. These methods extract both spatial and spectral features, and the use of residual block in SSRN increases the OA value 1% more than that of SSUN. The training and testing time of SSRN is more than double of SSUN. The proposed

Table 4.2: Comparison of classification evaluation parameters for IP, PU and SA datasets using different methods

Methods	IP Dataset			PU Dataset			SA Dataset		
	OA (%)	AA (%)	K (%)	OA (%)	AA (%)	K(%)	OA (%)	AA (%)	K (%)
SVM	85.30±2.8	79.03±2.7	83.10±3.2	94.34±0.2	92.98±0.4	92.50±0.7	92.95±0.3	94.60±2.3	92.11±0.2
2D-CNN	89.48±0.2	86.14±0.8	87.96±0.5	97.86±0.2	96.55±0.0	97.16±0.5	97.38±0.0	98.84±0.1	97.08±0.1
3D-CNN	91.10±0.4	91.58±0.2	89.98±0.5	96.53±0.1	97.57±1.3	95.51±0.2	93.96±0.2	97.01±0.6	93.32±0.5
SSUN	98.40±0.3	98.23±0.3	98.14±0.4	99.46±0.3	99.28±0.3	99.26±0.4	99.83±0.1	99.21±0.2	99.75±0.3
SSRN	99.19±0.3	98.93±0.6	99.07±0.3	99.90±0.0	99.91±0.0	99.87±0.0	99.98±0.1	99.97±0.0	99.97±0.1
HybridSN	99.21±0.1	98.54±0.2	99.11±0.1	99.93±0.0	99.83±0.0	99.91±0.0	99.99±0.0	99.99±0.0	99.99±0.0
Proposed Method	99.80±0.0	99.71±0.1	99.75±0.1	99.99±0.0	99.98±0.0	99.99±0.0	100±0.0	100±0.0	100±0.0

Table 4.3: Training time(min) and testing time(sec) for IP, PU, SA datasets for different HSI classification techniques

Methods	IP Dataset		PU Dataset		SA Dataset	
	Training time(min)	Testing time(s)	Training time(min)	Testing time(s)	Training time(min)	Testing time(s)
SVM	3.5	1.5	10.2	5.1	12.7	7.2
2D-CNN	1.27	1	2.5	1.3	2.7	1.7
3D-CNN	74.24	11	112.32	23.12	122.15	27
SSUN	22.1	6	33.12	4	29.12	4
SSRN	56	12	86	18	89	18.24
HybridSN	69	16	52	20	68	17
Proposed Method	32	1	43.1	11	45.35	12

Table 4.4: Classifcation evaluation parameters, training and testing time for Houston dataset

Methods	OA	AA	K	Training time	Testing time
SVM	65.09	66.34	62.31	24.5	2.1
2D- CNN	88.60	81.25	87.28	5.8	1.3
3D-CNN	89.35	85.62	88.13	274.26	2.6
SSUN	91.04	85.25	90.02	130.24	2.1
SSRN	95.24	92.23	94.70	191.33	2.78
HybridSN	98.13	95.54	96.18	73.57	1.89
Proposed Method	99.12	98.29	99.11	51.48	1.2

method gives maximum OA, AA value when compared with all the techniques. The training time is less than that of other spatiospectral feature-based deep learning models. Fig 4.2 shows the ground truth and classification map of the Indian pines dataset. The map of the proposed method is almost similar to that of ground truth, i.e., the amount of misclassification is very less.

Pavia University dataset is spatially larger than that of the IP dataset; the number of training and testing samples high. PU dataset also uses a 27×27 neighborhood similar to IP in the 2D-CNN model. The number of convolution layers, kernel size, and the number of filters are the same for this dataset. The OA, AA values are more than 95%, and execution time was less than that of the SVM model for the same dataset. The 3D-CNN used for the PU dataset has a small variation from the 3D- CNN model used for the IP dataset. In this model, the number of filters used in each convolution layer is 32, 64, and 128. All other state-of-the-art techniques used similar models for every dataset with the same batch size and epochs. The difference in OA, AA, and K for every spatiospectral feature-based technique is minute. The training and testing time is less for the proposed method is less than that of other methods. The classification map of the Pavia University dataset is shown in Fig 4.3.

The classification map of the Salinas dataset is shown in Fig 4.4. Comparing the classification accuracy of 2D-CNN and 3D-CNN for the SA dataset in the table, it is found that OA, AA, and K value for 2D-CNN is better than that of 3D-CNN. The proposed model gave 100% accuracy for the Salinas dataset, and its classification map is precisely similar to that of ground truth. The training and testing time is less than
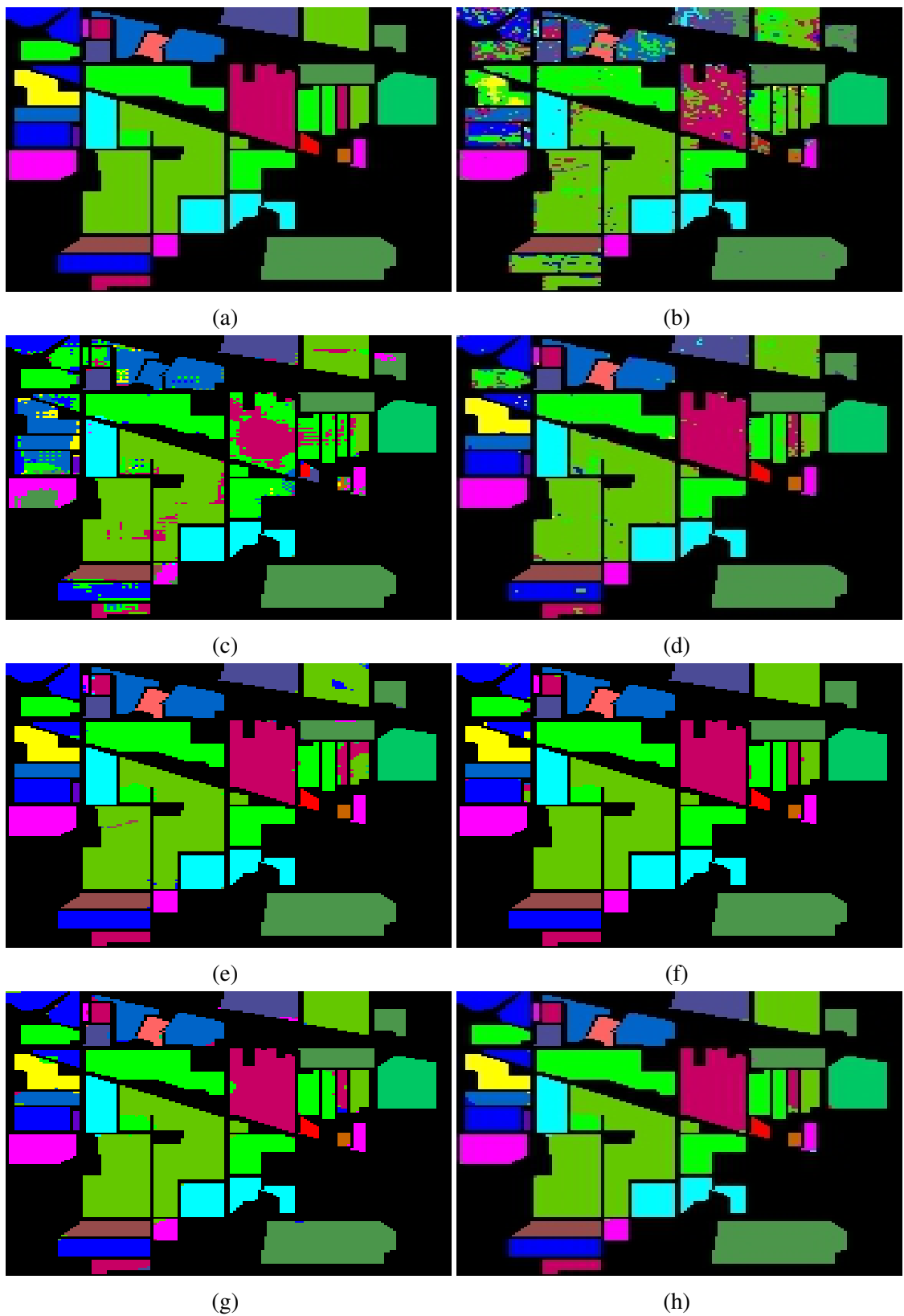
Figure 4.2: Classification map for Indian Pines dataset (a)Ground truth, (b)SVM, (c)2D-CNN, (d)3D-CNN, (e)SSUN, (f)SSRN, (g)HybridSN, and (h)Proposed
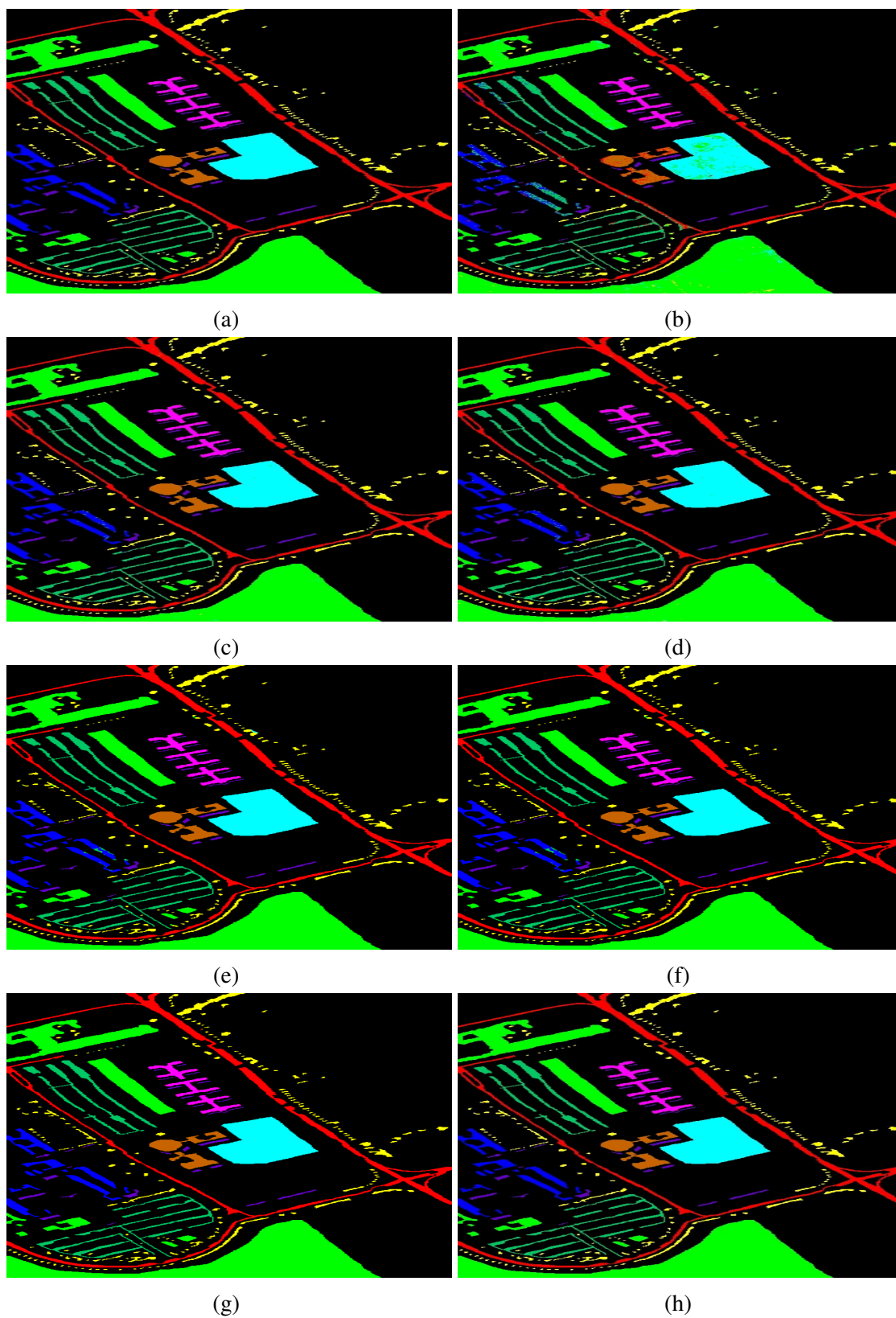
Figure 4.3: Classification map for Pavia University dataset: (a)Ground truth, (b)SVM, (c)2D-CNN, (d)3D-CNN, (e)SSUN, (f)SSRN, (g)HybridSN, and (h)Proposed

that of HybridSN. All the three dataset uses 25×25 neighborhood for convolution in HybridSN model. However, the number of extracted bands after the DR technique is different for each dataset. IP uses 30 principal components, and PU and SA use the first 15 principal components for feature extraction. The proposed model uses the same number of principal components for every dataset and size of the receptive field, the number of filters and filter size are similar.

The classification map of the Houston dataset is shown in Fig 4.5. To check the proposed model is to work better for spatially independent training and validation samples, the samples are selected from the IEEE GRSS Data Fusion Contest and carried out state-of-the-art techniques and proposed technique on these samples. The overall accuracy of 2D-CNN and 3D-CNN models is almost similar in the Houston dataset. The classification accuracy of recurrent neural network models SSUN and SSRN are increased, while the training time taken for these models is high. The classification evaluation parameters and execution time are better for the proposed model than all other techniques, and its classification map is similar to its ground truth with very less amount of misclassification. Analyzing table 4.2, table 4.3, table 4.4 and classification map for all the datasets, it is concluded that the proposed model has higher classification result in lesser training and testing time.

4.5.3 Effect of Limited Training Samples

The availability of labeled samples for hyperspectral images is very less, and its manual labeling is expensive. Therefore it is necessary to check the performance of the proposed model with a limited number of samples. According to the research paper (Chen et al. 2018), the training samples are selected from each class using the criteria $min(d_i * 20\%, 200)$ where d_i is the number of samples in each class. The remaining samples are taken to test the model. Table 4.5 shows the evaluation parameters for different classification methods on the four datasets using the limited number of training samples selected using the above criteria. As per analyzing the table 4.5, it is found that the proposed model performs better than that of all other techniques in a limited amount of training samples.
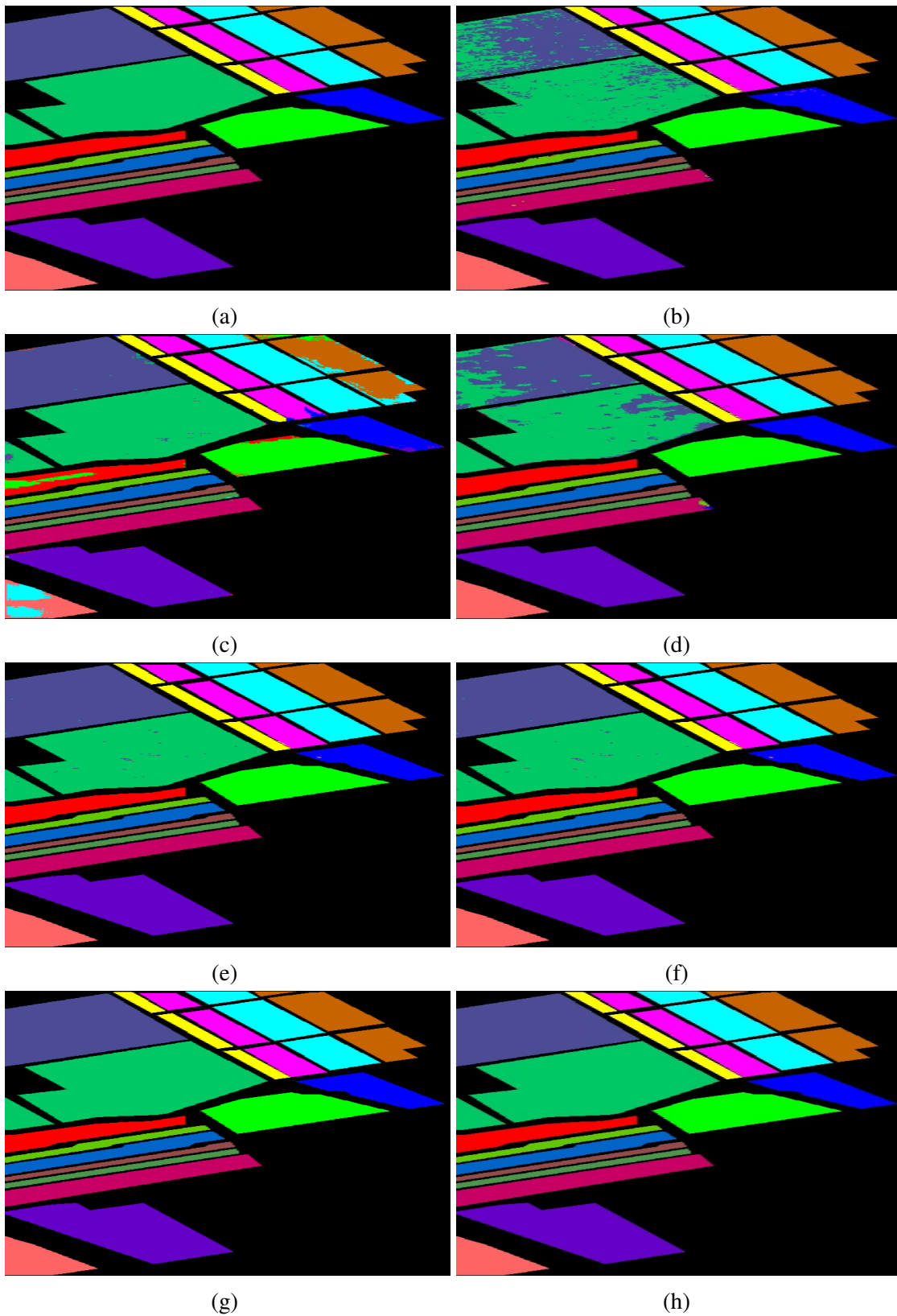
(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figure 4.4: Classification map for Salias dataset: (a)Ground truth, (b)SVM, (c)2D-CNN, (d)3D-CNN, (e)SSUN, (f)SSRN, (g)HybridSN, and (h)Proposed
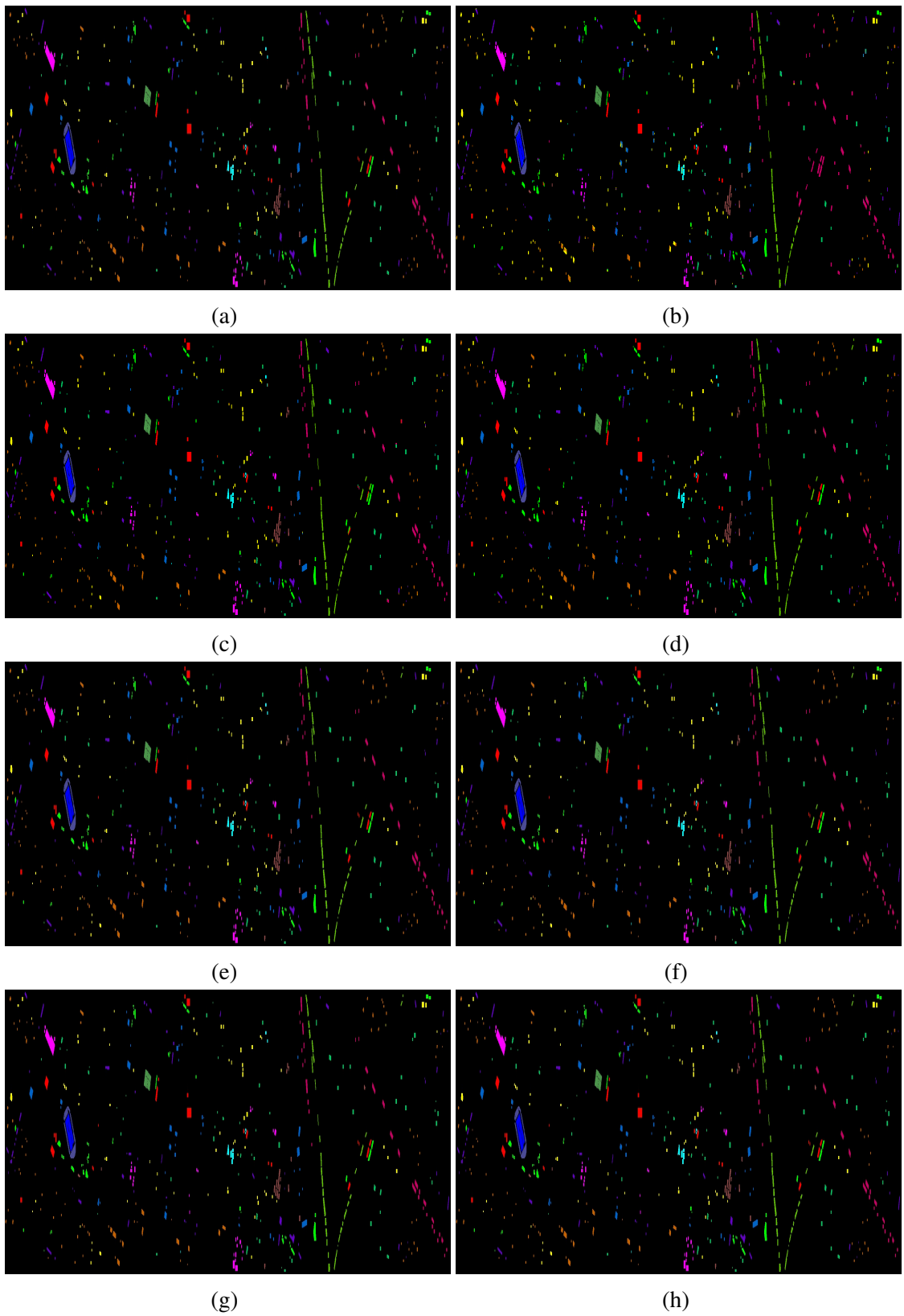
Figure 4.5: Classification map for Houston dataset: (a)Ground truth, (b)SVM, (c)2D-CNN, (d)3D-CNN, (e)SSUN, (f)SSRN, (g)HybridSN, and (h)Proposed

Table 4.5: Comparison of classification evaluation parameters for all datasets using different CNN methods with limited samples

Methods	IP Dataset			PU Dataset			SA Dataset			Houston Dataset		
	OA	AA	K	OA	AA	K	OA	AA	K	OA	AA	K
2D-CNN	86.81±0.5	85.37±0.3	85.48±0.3	95.25±0.6	94.46±0.3	95.11±0.3	97.38±0.0	96.26±0.3	96.12±0.3	87.36±0.2	84.46±0.3	86.36±0.5
3D-CNN	90.49±0.3	88.39±0.4	89.17±0.3	95.73±0.2	96.43±0.3	94.26±0.4	92.37±0.4	92.32±0.6	92.19±0.4	87.45±0.2	85.49±0.3	86.43±0.1
SSUN	95.27±0.3	94.29±0.3	95.14±0.3	97.49±0.2	97.22±0.5	97.24±0.7	98.39±0.2	98.58±0.3	98.72±0.2	90.35±0.2	88.42±0.2	89.37±0.2
SSRN	97.43±0.4	96.38±0.2	98.17±0.2	97.38±0.3	98.41±0.4	98.17±0.0	98.27±0.4	98.43±0.2	98.12±0.2	94.31±0.2	92.17±0.3	92.43±0.2
HybridSN	98.14±0.3	97.32±0.4	98.15±0.2	98.13±0.0	98.04±0.2	98.11±0.2	98.02±0.3	98.14±0.3	98.26±0.2	97.34±0.3	96.46±0.5	96.28±0.2
Proposed Method	99.17±0.2	98.11±0.1	98.75±0.1	99.12±0.2	98.87±0.3	98.45±0.2	98.12±0.2	99.12±0.3	99.35±0.5	98.99±0.3	98.22±0.2	98.90±0.2

Table 4.6: Overall accuracy obtained for various datasets when considering different window sizes in proposed method

Window size			OA			
w1	w2	w3	IP	PU	SA	Houston
13	11	9	98.13	98.72	99.1	98.01
15	**13**	**11**	**99.8**	**99.99**	**100**	**99.12**
17	15	13	99.73	99.94	99.52	98.45
19	17	15	99.66	99.95	99.84	98.74
21	19	17	99.14	99.23	99.43	98.45

78

If the number of training samples decreases, the large number of parameters in deep learning networks leads to overfitting, and the classification accuracy decreases. Since the proposed model has a lesser number of training parameters than other CNN models, all parameters are correctly optimized in a lesser number of samples and avoid overfitting. Thus the proposed method has better performance in lesser samples.

4.5.4 Effect of Input Window Size

The classification accuracy of CNN models depends on the window size. If the patch size is too large, the patch may contain pixels with different target classes, and it reduces the classification accuracy. If we choose smaller window size, the interclass diversity in samples reduces and lead to misclassification. Table 4.6 shows the overall accuracy of various datasets in the proposed method in different window sizes. From table 4.6 it is found that for the set of window sizes $w_1 = 15, w_2 = 13, w_3 = 11$ give maximum overall accuracy for all the datasets.

4.6 SUMMARY

This chapter presented a solution for HSI classification with improved classification accuracy in less computational time. The proposed hybrid CNN model analyzed both spatial and spectral features for HSI classification. All the existing spatio-spectral classification models lag in multiscale feature analysis. However, the introduction of a dedicated multiscale feature analysis module that extracts spatio-spectral features in different window sizes helped improving classification accuracy. The model uses a combination of 3D-2D CNN compared to existing 3D models. This Hybrid CNN nature helped to minimize the number of learnable parameters effectively and significantly reduce overall execution time.

The performance of the proposed Hybrid CNN model is evaluated against the existing machine learning and deep learning techniques. The dataset chosen for analyzing the model performance is heterogeneous, having vegetations, buildings, roads, and other human-made object classes. The model exhibited better classification performance for all types of classes. A classification technique needs to retain its performance even for spatially independent data. The Houston dataset (IEEE GRSS Data Fusion

Contest Samples) is one such spatially independent dataset. The model was able to produce higher accuracy in such challenging scenarios also. The model demonstrated excellent performance even in the presence of limited training samples.

CHAPTER 5

TCN BASED CROP YIELD PREDICTION

5.1   INTRODUCTION

Global food production depends heavily on accurate crop yield prediction systems. Accurate and timely prediction of crop yield aids in ensuring national and global food security. Hense crop yield prediction has a high socio-economic impact. Field surveys were the only mode of data collection for crop yield prediction in the early days. Recent advancements in the area of remote sensing greatly aid ahead of time crop yield prediction by covering a large geographical area with significantly less workforce.

The literature study conducted in crop yield prediction revealed that the predicted yield heavily depends on climatic and vegetation parameters. The continuous monitoring of these parameters during the entire crop cycle is necessary for accurate yield prediction, as discussed in Chapter 2. The deep learning-based prediction models, such as LSTM, RNN, etc., have better performance than conventional machine learning techniques. However, all the existing deep learning models fail to carry out multivariate time-based data analysis; also, the models have to increase their receptive field to analyze the data for a considerable period.

The proposed prediction technique implements the principles of dilated convolutions in the temporal convolutional network (TCN) to counterbalance the drawbacks of existing prediction models.

The contributions of the chapter are :

- Proposed a crop yield prediction model that accounts for multivariate time-based analysis of vegetation and climatic data.

- A reduction in prediction time complexity is achieved by using dilated convolution in TCN blocks.

- The impact of variations in vegetation and climatic parameters on yield are studied.

The chapter is organized as follows: Section 5.1 is for introduction, Section 5.2 briefs the various vegetation parameters used in this research and gives an overview of a basic TCN module. Section 5.3 describes the proposed TCN based prediction model and an explanation of various materials and data used for prediction. A diagrammatic representation of the proposed model is included in this section. Section 5.4 presents the results and its detailed analysis using various evaluation parameters, and Section 5.5 highlights the success story of our model, and the chapter is concluded in section 5.6.

5.2 PRELIMINARIES

5.2.1 Vegetation Indices

Vegetation indices (VI) measure the growth, health, vegetation cover either qualitatively or quantitatively with remote sensing images. The red, green, and blue bands of the visible spectrum, ultraviolet, near, and mid-wave infrared, are the major multispectral bands used to calculate VI. Since the sensors, platforms, and instruments used to capture remote sensing images vary, the algorithm or expression to measure VI is not uniform (Batten 1998). Apart from health and growth, VI is used to measure protein content in the plant, land-water content, pigmentation, etc. This section familiarizes the VI used in the proposed prediction model and why they were chosen for vegetation analysis.

Normalized Difference Vegetation Index (NDVI):   In the early days, Red (R) and near-infrared (NIR) reflectances are commonly used to measure vegetation, and most of the vegetation indices are calculated using these two bands only (Pearson and Miller 1972). Ratio vegetation index (RVI) and vegetation index number(VIN) are some of

the older VI and is calculated as

$$RVI = \frac{R}{NIR} \tag{5.1}$$

$$VIN = \frac{NIR}{R} \tag{5.2}$$

In order to increase the discriminative power of former VI and avoid the sensitivity towards atmospheric parameters, a new vegetation index was later introduced and named as normalized difference vegetation index (NDVI). NDVI is the most familiar vegetation index, and its value ranges between 0 and 1. The equation to measure NDVI is given in **??**. Here R and NIR are the mean reflectances of red and near-infrared spectral bands. NDVI is also affected by atmospheric disturbances, clouds, and other climatic factors (Rouse et al. 1974).

$$NDVI = \frac{NIR - R}{NIR + R} \tag{5.3}$$

Enhanced Vegetation Index (EVI): EVI is an optimized vegetation index designed to improve vegetation and sensitivity monitoring by de-coupling atmospheric signals and other background disturbances. EVI is computed as:

$$EVI = G * \frac{NIR - R}{NIR + C1 * R - C2 * B + L} \tag{5.4}$$

Here NIR, R, and B are the atmospherically corrected or partially corrected surface reflectance of near-infrared, red and blue bands, respectively. L represents the nonlinear canopy adjustment parameter, C1 and C2 denote the aerosol resistance terms. According to MODIS EVI algorithm the values for each coefficients are L=1, C1 = 6, C2 = 7.5, and G (gain factor) = 2.5.

Two Band Extended Vegetation Index (EVI2): The blue band has a lesser wavelength, and it cannot travel for a considerable distance. Therefore, the blue band has less reflected energy and the signal-to-noise ratio, leading to an EVI calculation error. Another challenge is that some of the sensors are not designed to capture the blue band. Hence an optimized vegetation index with two bands is mandatory, and a two-band extended vegetation index (EVI2) is introduced; it is calculated as

$$EVI2 = 2.5 * \frac{NIR - R}{NIR + 2.4 * R + 1} \tag{5.5}$$

Leaf Area Index (LAI): Leaf area index is the total leaf area of the crop canopy. It is calculated as leaf area per unit ground area. Since LAI is a ratio of two areas, it is a

dimensionless quantity. LAI is the best indicator for crop productivity and is commonly used to measure photosynthesis rate, evapotranspiration, and so forth (Jensen 2007; Zheng and Moskal 2009). Based on disciplines and applications, researchers were using different types of LAIs and a set of various LAI is listed in Table 5.1.

Table 5.1: Different LAI definitions

Name	Definition	Application
Total Leaf Area Index (To-LAI)	One side leaf photosynthetic tissue area per unit ground area	Wide leaf measurement
Projected Leaf Area Index (PLAI)	Area of horizontal shadow cast above a horizontal leaf	Maximum area of leaf from overhead orbital view
Silhouette Leaf Area Index (SLAI)	Area of leaves inclined to horizontal surfaces	Identify shape of leaves
Effective Leaf Area Index (ELAI)	One half of the total area of light intercepted by leaves per unit horizontal ground surface area	Describe radiation inteception
True Leaf Area Index (TLAI)	One half the total green leaf area per unit horizontal ground surface area	Characterize radiation regimes

Photosynthetically Active Radiation (PAR): Plants use only a portion of sunlight for photosynthesis. The spectral range of radiation actively used for photosynthesis is called photosynthetically active radiation (PAR). PAR depends on daily radiation, sunlight hour, and so on. Yield is positively correlated with PAR and photosynthesis rate. For multispectral images, a fraction of PAR, named fPAR used to analyze the crop and yield (Jiang et al. 2004).

Gross Primary Production (GPP): Gross primary production (GPP) is the quantity of biomass produced in photosynthesis per unit time. GPP is also defined as the conversion rate of carbon to yield in the ecosystem without considering the plant's respiration.

5.2.2 Temporal Convolutional Networks (TCN)

Time-based analysis is a challenging area in most engineering and scientific applications such as stock market, weather forecasting, population analysis, etc. Any two points in a time series data are measured within an equal interval of time and are arranged in time order. Therefore the data points are discrete values measured in continuous time. Time series prediction uses historical data to predict future observations. The future data may either be more related to its recent time-based data or long-ago observations.

The recurrent neural network is one of the deep learning models designed for time series analysis and prediction. However, parallelism is not achievable in RNN, and repeated usage of the same operations in long sequence data leads to vanishing gradient problem. LSTM and gated recurrent unit (GRU) are designed to overcome these difficulties of RNN. The usage of convolutional networks in sequence data created a significant improvement in its analysis and forecasting. Adapting the idea of convolutional network and convolution of current timestamp data with past timestamp data in previous layer (van den Oord et al. 2016) , a new family of neural network architecture for sequence data named temporal convolutional network (TCN) or temporal convolutional neural network (TCNN) is introduced (Bai et al. 2018; Lea et al. 2016). TCN is designed based on two principles:

1. The model can take an input sequence of any length and produce an output sequence of same length (similar to recurrent neural network design)

2. The model is using causal convolution to avoid information leakage

Causal convolution means the value at time t is dependent only on the previous timestamp results and not on future timestamps. TCN uses one-dimensional (1D) data to achieve the first principle mentioned above, and all the hidden layers in the network are designed with an equal number of neurons. Therefore, causal convolution recalls the data in previous layers linearly with an increase in depth of the network. This upgradation results in the receptive field's linear growth and needs higher network depth to cover all input dependencies. TCN employs dilated convolution to avoid this drawback,

which increases the receptive field exponentially and covers entire input dependencies in lesser network depth.

Dilated Convolution: Dilated convolution covers a large receptive area by skipping certain steps of input time series. For an input sequence $X \in \mathbb{R}^T$ and a filter $h : \{0, 1, \ldots k - 1\} \to \mathbb{R}$, dilated convolution is defined as

$$H(x) = \sum_{i=0}^{k-1} h_i X_{s-d.i} \tag{5.6}$$

Here k denotes the size of the filter, and d is the dilating factor. Since the receptive field grows exponentially in each layer of the network, $d = 2^\nu$ where ν is the level of the network.
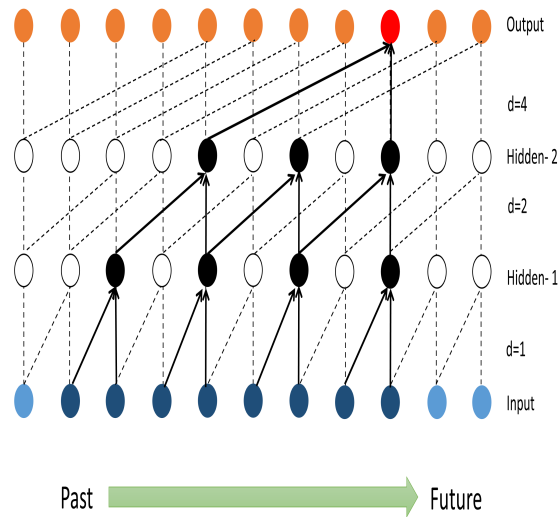

Figure 5.1: Dilated Causal Convolution with filter size=2

Consider figure 5.1, it shows the dilated convolution of an input sequence with a filter H of size $k = 2$. In level zero of the network, i.e., $\nu = 0$ the dilating factor

86

$d = 2^\nu = 2^0 = 1.$

$$H(x) = \sum_{i=0}^{2-1} h_i X_{s-1.i}$$
$$= \sum_{i=0}^{1} h_i X_{s-i} \qquad (5.7)$$
$$= h_0 X_s + h_1 X_{s-1}$$

By analyzing equation 5.7, it is found that the output of each neuron in hidden layer 1 (level zero) is the convolution of the filter with the input in the current timestamp and previous timestamp. Similarly while considering the hidden layer 2 (level $\nu = 1$) and $d = 2^1 = 2$

$$H(x) = \sum_{i=0}^{2-1} h_i X_{s-2.i}$$
$$= h_0 X_s + h_1 X_{s-2} \qquad (5.8)$$

i.e., the output of the neuron at hidden layer 2 is the result of the convolution of the filter with the previous layer neuron at the same timestamp and second to the same timestamp. This process will repeat for each successive hidden layer with a backshift in time as s-1, s-2, s-4, s-8, and so on. Therefore dilation is introducing a fixed time step in every two filter maps. The receptive field in TCN can be increased in two ways; either increase the filter size k or increase the dilation factor d by introducing more hidden layers.

Residual Blocks: Residual units are the building blocks of temporal neural networks. Residual blocks contain a series of residual connections, which perform a series of transformations on input and add it to the input. A residual block can have several dilated convolution layers and activation layers based on the application (Guirguis et al. 2021). Batch normalization and dropout are possible in each residual block to optimize the network. The output o of each residual block can be represented as:

$$o = Activation(F(X) + X) \qquad (5.9)$$

where F is a set of transformations and X denotes the input. ReLU is the standard activation function used in most of the TCN designs.

5.3 MATERIALS AND METHODS

5.3.1 Study Area

The study area chosen for testing the proposed TCN based crop yield prediction model is Kuttanad, the rice bowl of Kerala state in India. The Kuttanad region is spread across three districts of Kerala, namely Kottayam, Alappuzha, and Pathanamthitta. Kuttanad is familiar for its geographical feature as it is the lowest altitude land in India. The majority of paddy land areas are below sea level, and it is one of the critical places where paddy cultivation is carried out below sea level among the entire world. The Kuttanad region is further categorized into Lower Kuttanad, Upper Kuttanad, and North Kuttanad. This research work chose a $20.5km \times 20.5km$ area @ 9.537103, 76.447742 in the Upper Kuttanad region for analyzing the proposed TCN based prediction model. Figure 5.2 shows the map of the study area in the Upper Kuttanad region.

Figure 5.2: Study area in Upper Kuttanad region

5.3.2 Crop data

Paddy is the common crop in the study area; therefore, the prediction model predicts rice yield. Kerala has three seasons of Paddy cultivation, namely Virippu, Mundakan, and Puncha. Virippu(I crop) is also known as Autumn seasonal, starts its sowing time in April or May and plans its harvesting from September to October. Mundakan begins in September or October and ends in December or January. Therefore Mundakan is also known as Winter seasonal or II crop. Puncha or Summer harvest starts its sowing after Mundakan, and harvesting ends before May. Puncha is the III crop while considering

88

a crop year from May to April. Since Kuttanad is below sea level, Puncha or Summer seasonal has high significance in this region. The Puncha in the Upper Kuttanad region starts in October and ends in February. Therefore all crop-related data collection for the study area is significant between October and February. The repositories from which the annual crop yield data are mentioned in section 2.6.4

5.3.3 Climatic data

The correlation between climatic parameters and the yield of the crop is critical. Therefore, the collection of climatic data for the prediction model is significant. The proposed research work uses the climatic parameters minimum temperature (min T), maximum temperature (max T), Rainfall, Wind, and Humidity to predict the yield. The source of climatic data and its measures are discussed in section 2.6.3.

5.3.4 MODIS Data

The land cover mask for the study area and various vegetation parameters are collected from various MODIS products as mentioned in section 2.6.3. The land cover of study region consists of 17 classes, out of which 11 types are vegetation classes. This work merges classes 12 (cropland) and 14 (cropland/ natural vegetation) for crop mask and uses this area to calculate vegetation parameters. The classification map for the study region is shown in Figure 5.3. All MODIS data is resampled into 250m resolution, and the same spatial resolution scale is used for yield prediction. Since MODIS collects data only for clear pixels without any shadows, the average values for a given crop mask are better for yield prediction.

5.3.5 TCN Based Yield Estimation Model

The proposed TCN model uses three TCN blocks to identify the long-term dependency of vegetation indices and climatic parameters for rice yield prediction. Figure 5.4 shows the diagrammatic representation of the proposed TCN based prediction model. All TCN blocks in the proposed architecture use Rectified Linear Unit (ReLU) activation function. Instead of directly feeding feature maps to fully connected layers, the proposed model uses Global Average Pooling and Global Max Pooling layers. The introduc-
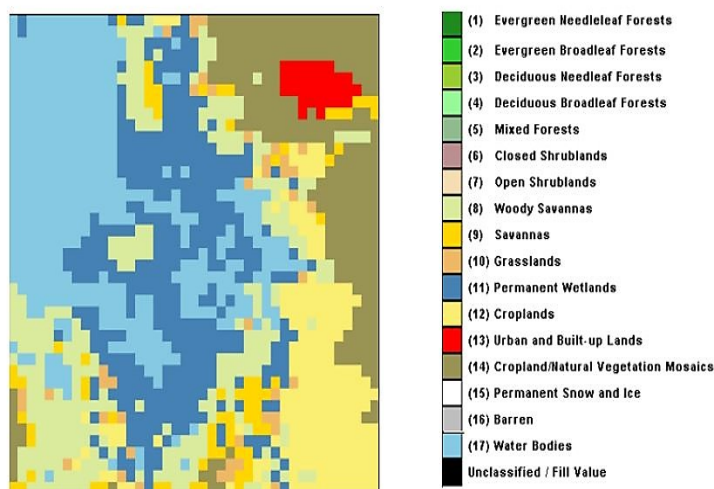
Figure 5.3: Classification map of study area

tion of these pooling layers contributes to the accurate identification of dependencies between feature maps and yield. The feature maps from global average pooling and global max-pooling are concatenated and passed to the following dense layer having 16 neurons. The overfitting problem is solved by the use of 0.1% dropout. The last layer of the proposed model performs rice yield prediction using the *linear* activation function. Model optimization is carried out using *adam* optimizer with loss function mean square error (mse). The proposed TCN model completed training in 500 epochs with minimal error.

Six vegetation parameters (NDVI, EVI, EVI2, LAI, fPAR, GPP), five climatic parameters (min T, max T, Wind, Rainfall, Humidity), yield measured in kilogram per hectare (kg/ha) are passed as the input to TCN block. Each TCN block comprises two 1D- convolution layers and a dilated convolution layer. The dilated convolution layer is designed to have two hidden layers apart from the input and output layer (i.e., the dilation parameter is set as [1,2,4]). A dropout of 0.1% is also applied to handle overfitting in the TCN block. The first TCN block of the proposed model uses 128 filters for convolution, and the number of filters reduced to 64 and 32 for succeeding TCN blocks. The convolution kernel size is set to 4 for all blocks. Since TCN uses the residual connection, the output of each TCN block is calculated as the sum of input and transformed input as mentioned in equation 5.9. The flow diagram of the proposed TCN block is
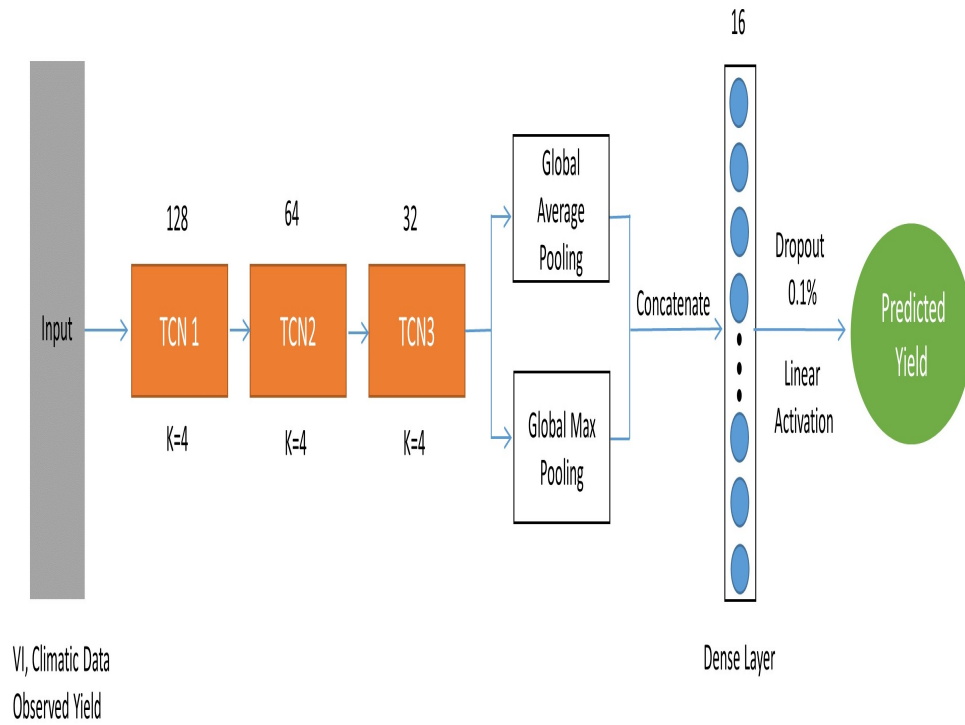
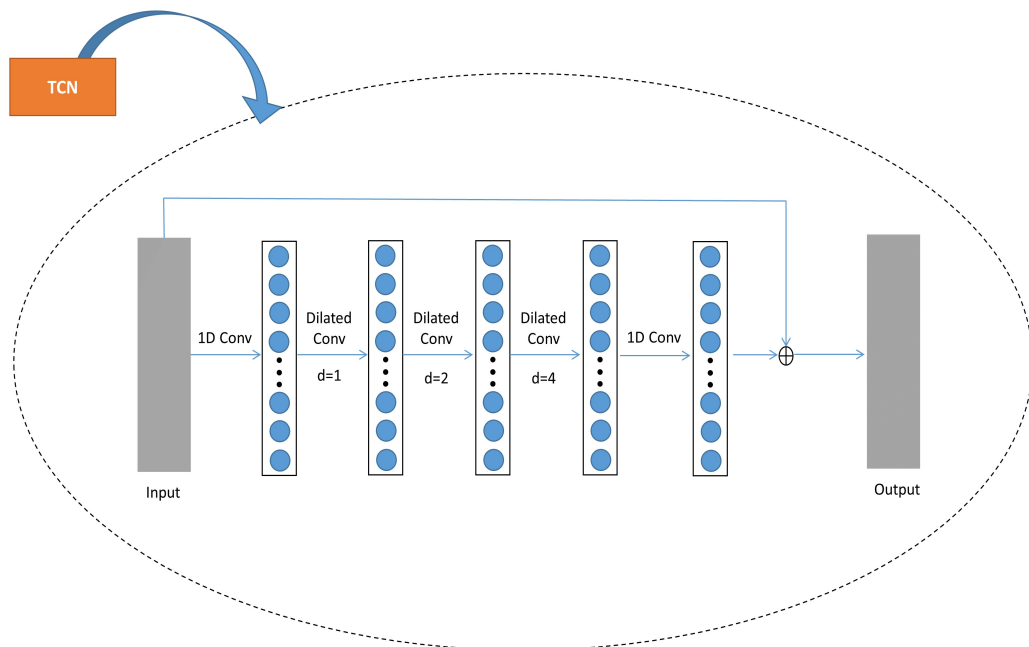Figure 5.4: Block diagram of proposed TCN based prediction model

Figure 5.5: Architecture of each TCN block

shown in Figure 5.5.

91

5.4 RESULTS AND DISCUSSIONS

The proposed model is implemented in python and executed on Intel(R) Xeon(R) Silver 4114 CPU @ 2.24 GHz paired with 196 GB RAM. The experimental study is conducted for a short duration variety of paddy whose life span lies between 105 to 120 days. Therefore, 19 weeks of vegetation and climatic parameters are collected from the study area for each agricultural crop year between 2012 and 2019.We collected various yield predicting parameters from October to February and used these data for prediction. In February, paddy is in the ripening stage of the crop cycle. For any rice variety, the ripening stage's length is similar, therefore choosing this stage is more accurate.

5.4.1 Experimental Results

The proposed yield prediction model is a supervised deep neural network utilizing the features of temporal convolution. Input samples are divided into 70% training set, 20% validation set, and remaining for testing. Various machine learning and deep learning based prediction models such as regression (Johnson et al. 2016; Son et al. 2013), deep neural network (Khaki and Wang 2019), LSTM (Schwalbert et al. 2020), CNN-LSTM (Sun et al. 2019) and CNN-RNN (Khaki et al. 2020) are compared against the proposed TCN based model.

Evaluation Parameters: The model is evaluated using the parameters coefficient of determination(R^2), mean absolute error (MAE), and root mean square error (RMSE).

MAE is the average of the difference between the predicted value and observed value, and is calculated as

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |P_i - O_i| \tag{5.10}$$

P_i is the predicted value, O_i is the observed value, and N is the total number of samples used for prediction.

RMSE provides an estimate of difference between predicted values and observed values, and is given by

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)^2} \tag{5.11}$$

Coefficient of Determination (R^2) is another measure to find the relationship between predicted values and observed values. It is calculated using the equation 5.12

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(P_i - \overline{O})^2} \qquad (5.12)$$

Here $\overline{O}$ denotes the mean of observed values over N samples. The value of R^2 falls in a range $-\infty$ to 1. A desirable prediction model must have minimal MAE and RMSE with maximum R^2.

5.4.2 Result Analysis

Regression is the most commonly used prediction technique for forecasting crop yield using remote sensing data. Regression models use combinations of NDVI, EVI, and LAI for predicting yield. Since yield depends on more than one parameter, a multiple linear regression (MLR) model is highly desirable. The works inJohnson et al. (2016) and Son et al. (2013) used multiple linear regression models to predict yield from MODIS NDVI-EVI, MODIS EVI-LAI, respectively. DNN based model in Khaki and Wang (2019) used genotype, soil, and climatic parameters to predict maize yield. This DNN model is optimized using adam optimizer with a learning rate of 0.03% and employed L2 regularization to avoid overfitting. The above model results indicated that the effect of genotype in crop yield is minimal compared to environmental factors.

LSTM based prediction model is chosen as the next state-of-the-art model to compare the proposed technique's performance because it helps in the temporal analysis. The model predicts the yield by analyzing NDVI, EVI, land surface temperature (LST), and precipitation. The integration of time series data with conventional static data reduced the error in prediction (Schwalbert et al. 2020). However, the accurate selection of yield prediction features is a crucial challenge and can be solved by introducing CNN. The CNN-LSTM (Sun et al. 2019) model used CNN for feature selection and LSTM to perform temporal analysis on relevant features. CNN-RNN (Khaki et al. 2020) is another hybrid prediction model chosen to examine the potential of the proposed TCN architecture. Although the CNN performs feature selection in both CNN-LSTM and CNN-RNN, the latter employs independent convolution networks for weather and soil data.

The usage of dilation convolution in the proposed prediction model helps to increase the network's receptive field with fewer learnable parameters. Hence this work compares the execution time of the training and testing phase of the proposed method with other yield prediction models. Table 5.2 lists the evaluation parameters MAE, RMSE, and R^2 in the training and validation phase for the state-of-the-art and proposed prediction model. The execution time taken for training and testing various prediction models is shown in table 5.3. The training time is measured in minutes and testing time in seconds.

Table 5.2: Comparison of evaluation parameters among various prediction models

Method	Training MAE	Training RMSE	Training $R^2(\%)$	Validation MAE	Validation RMSE	Validation $R^2(\%)$
MLR (NDVI-EVI)	6.38	6.51	77.78	9.32	9.49	64.74
MLR (EVI-LAI)	4.21	4.39	90.92	11.98	12.78	40.82
DNN	4.15	4.35	90.36	5.89	6.25	72.07
LSTM	3.91	4.13	90.89	3.85	4.11	91.11
CNN-LSTM	3.5	3.84	93.22	3.44	3.76	93.25
CNN-RNN	3.01	3.08	95.35	4.11	4.32	87.08
Proposed Model	0.97	1.17	97.89	0.91	0.97	98.03

Table 5.3: Training time(min) and Testing time(sec) among various prediction models

Method	Training time (min)	Testing time (sec)
MLR (NDVI-EVI)	3.6	2.5
MLR (EVI-LAI)	3.4	2
DNN	6.76	4
LSTM	7.6	4.1
CNN-LSTM	9	4.2
CNN-RNN	11	5
Proposed Model	2.3	2

Analysis on Table 5.2 indicates that multiple linear regression based models are underperforming and produce higher error rates with lesser R^2 values. The regression model using EVI-LAI combination performs better than a model with NDVI-EVI combination in terms of MAE, RMSE, and R^2 during training. A good fit model must have the capability to fit both training and validation data very well. In terms of validation performance, the model using NDVI-EVI outperformed the other. Even though the
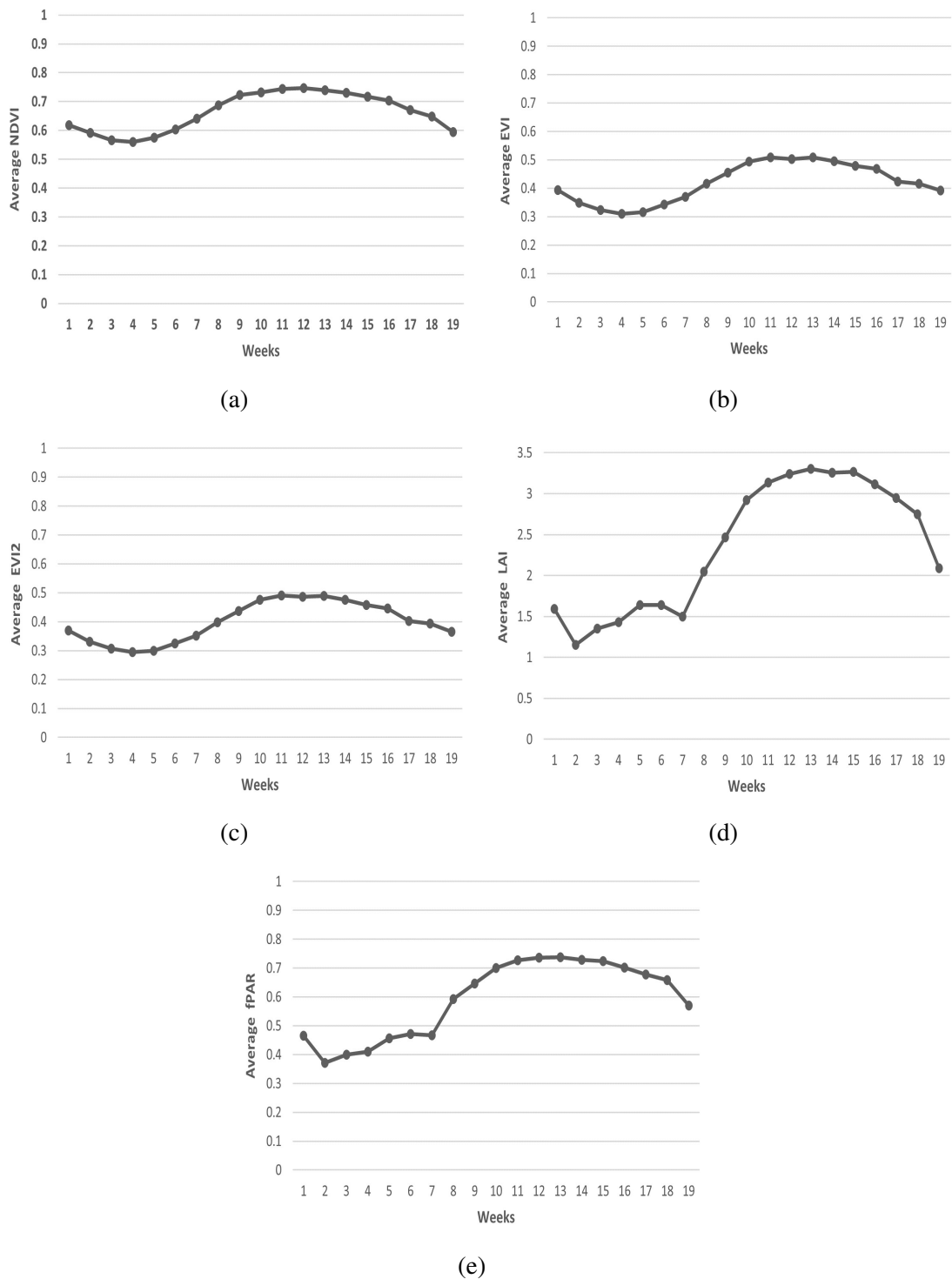
(a)

(b)

(c)

(d)

(e)

Figure 5.6: The effect of vegetation idices on rice measured for 19 weeks of crop cycle (a)Average NDVI vs Weeks, (b)Average EVI vs Weeks, (c)Average EVI2 vs Weeks, (d)Average LAI vs Weeks, (e)Average fPAR vs Weeks

prediction performance of DNN improved over conventional regression methods, they fail in validation performance. i.e., the DNN model produced higher validation MAE, RMSE, and lesser R^2compared to their training stage. The residual characteristics of LSTM help to improve the prediction capability compared to the DNN model. LSTM prediction model achieves a validation performance boost up to 20% over its former techniques. The CNN-RNN model produced lesser prediction errors than the CNN-LSTM model during the training phase, but it is found that the model is overfitting. The proposed TCN based prediction model outperforms all the other state-of-the-art techniques in both the training and validation phase. The very minimal MAE and RMSE values of the proposed model indicate that the model is reliable and capable of accurate early yield prediction.

While comparing the training time taken for various prediction models listed in Table 5.3, it is found that the multiple linear regression model takes less than 4 minutes for both MLR techniques. An increase in the number of intermediate layers in the DNN model increases its train and test phase time. Prior feature extraction and feature selection phase using CNN raise the training time of the hybrid models CNN-LSTM and CNN-RNN compared to other conventional techniques. Independent convolution models for soil and weather data in CNN-RNN increase the number of trainable parameters than CNN-LSTM based prediction model, and hence the training and testing time of CNN-RNN is higher than that of CNN-LSTM. The proposed TCN based prediction model is computationally less complex than other state-of-the-art techniques.

Figure 5.6 shows the variation in values of various vegetation indices for 19 weeks of the rice crop cycle. For a short-duration paddy, the life cycle can be divided into three stages: the vegetative stage, the reproductive stage, and the ripening stage. The vegetative stage lasts up to nine weeks of the life cycle, and the next four weeks (up to 13 week), the plant is in the reproductive stage, i.e., paddy started flowering after nine weeks. In the last 5-6 weeks, the plant is in a mature state and starts grain filling. The distribution of NDVI, EVI, and EVI2 are almost similar throughout the crop cycle. Their values are reducing slightly up to weeks 3 to 4 (transplanting and tillering stage) and increasing in the vegetative stage. The value of these three vegetation indices is
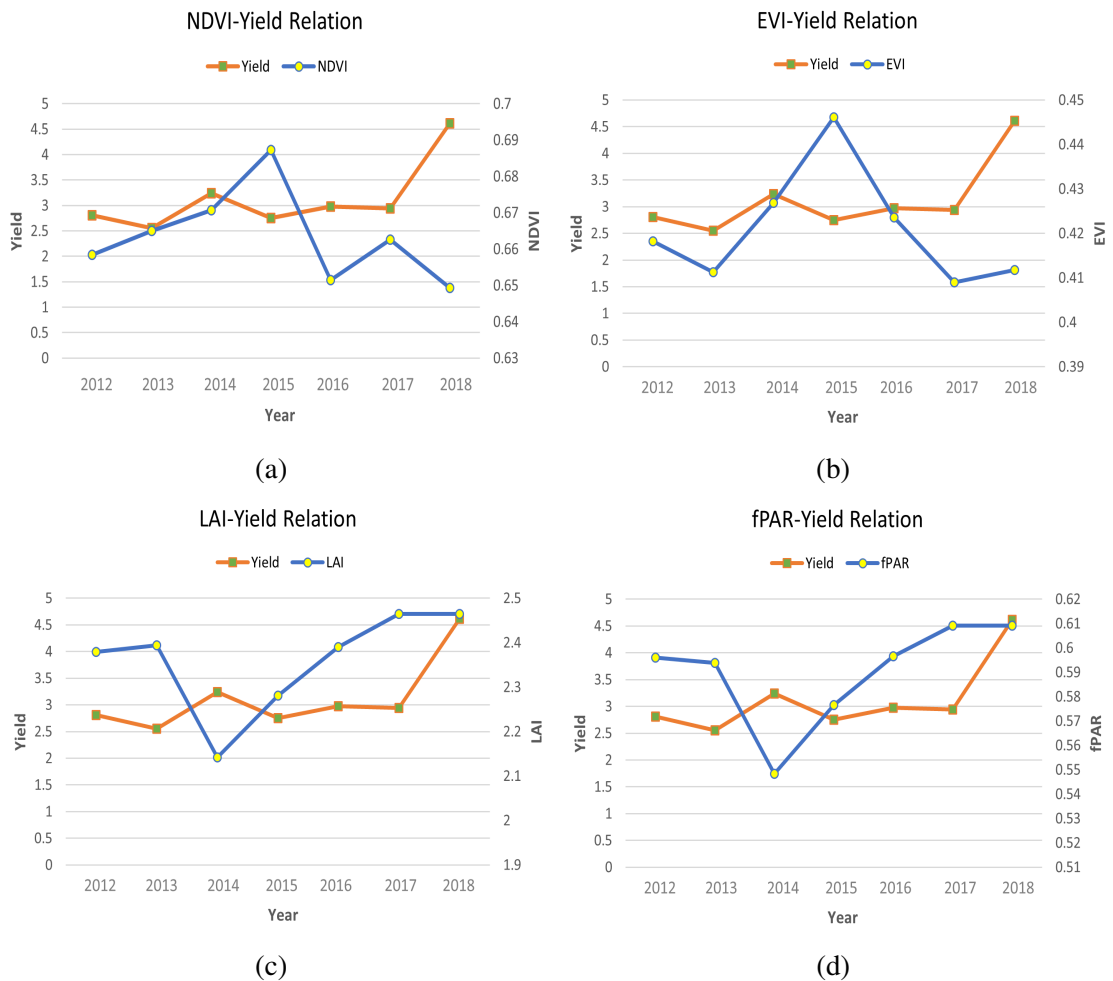
Figure 5.7: The effect of vegetation idices on rice yield (a)Average NDVI vs Yield, (b)Average EVI vs Yield, (c)Average LAI vs Yield, (d)Average fPAR vs Yield

steady-state in the paddy's reproductive stage and stated decreasing slowly in harvesting time. Similarly, the distribution of LAI and fPAR are same as shown in figure 5.6d and 5.6e. LAI and fPAR have an abrupt increase in paddy's reproductive stage, and it continues to increase around half of the reproductive stage.

The effect of vegetation indices in rice yield from 2012 to 2018 is picturized in Figure 5.7. The yield is represented in Tons per hectare in all subgraphs of Figure 5.7 and Figure 5.8. Although the vegetation parameters NDVI and EVI are directly proportional to yield, the study found that the relationship is not satisfied for the agricultural year 2015. Analysis of subgraphs 5.7c and 5.7d indicates that LAI and fPAR contribute to yield similarly. An increase in LAI naturally increases the photosynthetic rate and

97

(a)

(b)

(c)

(d)

(e)

Figure 5.8: The effect of climatic parameters on rice yield: (a)Average min T vs Yield, (b)Average max T vs Yield, (c)Average Wind vs Yield, (d)Average Rain vs Yield, (e)Average Humidity vs Yield

thereby results in a higher yield. The contradiction in the relationship between yield and vegetation parameters seen in Figure 5.7 is concrete evidence of climatic factors' effect on yield.

Figure 5.8 illustrates the effect of climatic factors on yield. Each subgraph plots the average value of climatic parameters for each agricultural year. The clarifications for the reduction in yield during 2015 is obtained from subgraphs 5.8b and 5.8d. Even though the vegetation parameters are favorable for high yield during 2015, the heavy rain and higher daytime temperature resulted in crop damage and less productivity. A consistent wind rate of 7-9 km/hr is beneficial for paddy growth and yield as per subgraph 5.8c. In 2018, the rainfall and temperature were higher compared to other agricultural years. However, the presence of higher LAI and fPAR result in higher productivity even in adverse climates.

5.5 SUCCESS STORY OF THE WORK

The forecast report of summer paddy for the year 2019-2020 by The Indian Ministry of Agriculture & Farmers Welfare (`http://agricoop.nic.in`) shows that the predicted yield for the Upper Kuttanad region is 4.37 ton per hectare. This prediction was conducted with the help of climatic data, vegetation data, and field surveys. The proposed TCN model predicted a relatively similar yield of 4.34 tons per hectare using remote sensing data and climatic parameters only. The study opens the door to accurate automated crop yield forecasting from high-resolution remote sensing data with minimal field survey and workforce.

5.6 SUMMARY

This chapter introduced a TCN based crop yield prediction model that outperformed the existing models. The model was able to perform the desired multivariate time-based analysis for accurate prediction efficiently. The application of dilated convolutions in TCN layers significantly reduced the receptive field size of the network and thereby reduced the prediction time.

The performance of the proposed TCN model is evaluated against the existing ma-

chine learning and deep learning prediction techniques. The chosen study area cultivates paddy as the principal crop, and the prediction model attempted to predict the summer rice yield. The research work identified the effect of climatic and vegetation parameters on yield. The predicted result of yield for the study area was significantly close to the Indian Ministry of Agriculture & Farmers Welfare forecast.

CHAPTER 6

CONCLUSIONS AND FUTURE SCOPE

Food security is a significant concern of the world population. Precision agriculture made a massive impact in ensuring the availability of food. Early yield prediction is one of the solutions used in precision agriculture for food security. The crop yield is dependent on many factors such as vegetation indices, climatic data, soil conditions, and so on. Therefore temporal analysis of these parameters and continuous monitoring of cultivation land is necessary for accurate yield prediction. The field survey was a solution for this data collection in the early days. However, satellite images for data analysis help reduce human resources, and it significantly covers large areas accurately.

A thorough literature study is carried on various satellite image-based yield prediction models and found that spectral images such as multispectral, hyperspectral images are widely used. The contiguous bands in hyperspectral images help in discriminating objects in the earth's surface clearly. Nevertheless, the redundancy in bands and the curse of dimensionality are significant concerns in hyperspectral image analysis. This research work focused on hyperspectral image-based crop classification as one of the objectives in an account of these concerns.

To design the best classification model for hyperspectral images, it is essential to perform a pre-processing step to extract optimal bands. Since most of the band extraction techniques are linear and nonlinear extraction techniques are computationally complex, it is advantageous to design a new computationally more accessible nonlinear dimensionality reduction technique for HSI. The proposed band extraction uses the

random projection technique to reduce computational overhead in the nonlinear DR methods. Gaussian random projection is applied prior to kernel PCA and reduces the number of bands by half. The GRP and KPCA combination extracted the optimal bands from input HSI and fed them to the classification model for further processing.

The second work of this thesis focused on developing a deep learning-based classification model for HSI. The existing deep learning models that use spatial and spectral features are classified better than a model that uses either of the features. Therefore, our model also focused on spatio-spectral feature extraction using convolutional neural networks. The multiscale spatiospectral feature-based hybrid CNN overcomes the drawbacks of existing deep learning models and produces a promising improvement in results. The hybridization of 3D-CNN and 2D-CNN in the proposed model helped to reduce the computational complexity too.

The following contribution of this research work is to design a crop yield prediction model using vegetation parameters and climatic data. The Upper Kuttanad region in Kerala state was chosen as the study region and predicted the rice yield for summer cultivation. The multivariate time-based analysis of yield predicting parameters is done with the help of a temporal convolution network. The dilated convolution module in the proposed model helps to find the dependencies of parameters without increasing the receptive field.

6.1 FUTURE SCOPE

As the techniques proposed in this thesis are performed better than existing models for HSI classification and yield prediction, this area is still in its infant stage. Therefore, there is significant scope for future works. Further research directions exist in this area are:

- The existing band extraction techniques are not so efficient for hyperspectral image analysis. Nonlinear DR techniques are still lagging in nonlinearity preserving. Hence design a best band extraction technique using optimization techniques is a future topic for research.

- The availability of ground truth for hyperspectral images is significantly less. Hence, designing an unsupervised classification model or using transfer learning in hyperspectral image classification requires more attention in the future.

- Yield of a crop depends on many environmental and vegetation parameters. This study uses multispectral images to collect the vegetation parameters. While parameter collections using hyperspectral images and other high-resolution satellite images improve the prediction results.

- Most of the yield prediction models are designed for a particular crop in a specific area. Hence develop a general yield prediction model using various dependency parameters is still a new path.

In conclusion, this dissertation proposes a new deep learning model for hyperspectral image classification. The classification model uses the band extraction technique as preprocessing step to solve the curse of dimensionality in HSI. The combination of GRP and kernel PCA is used as the band extraction technique, and random projection helps to reduce the computational overhead of KPCA. A new yield prediction using vegetation and climatic parameters is proposed for rice yield in the Upper Kuttanad region. The model uses dilated convolution technique in TCN for multivariate time-based analysis.

BIBLIOGRAPHY

Abou El-Magd, I. and El-Zeiny, A. (2014). "Quantitative hyperspectral analysis for characterization of the coastal water from damietta to port said, egypt." *The Egyptian Journal of Remote Sensing and Space Science*, 17(1), 61–76.

Achlioptas, D. (2001). "Database-friendly random projections." In *Proceedings of the Twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '01, Association for Computing Machinery, New York, NY, USA, 274–281.

Alganci, U., Sertel, E., Ozdogan, M. and Ormeci, C. (2013). "Parcel-Level Identification of Crop Types Using Different Classification Algorithms and Multi-Resolution Imagery in Southeastern Turkey." *Photogrammetric Engineering & Remote Sensing*, 84(11), 1053–1065.

Bachmann, C. M., Ainsworth, T. L. and Fusina, R. A. (2005). "Exploiting manifold geometry in hyperspectral imagery." *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), 441–454.

Bai, S., Kolter, J. Z. and Koltun, V. (2018). "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling." *CoRR*, abs/1803.01271.

Batten, G. D. (1998). "Plant analysis using near infrared reflectance spectroscopy: The potential and the limitations." *Australian Journal of Experimental Agriculture*, 38(7), 697–706.

Belkin, M. and Niyogi, P. (2003). "Laplacian eigenmaps for dimensionality reduction and data representation." *Neural Computation*, 15(6), 1373–1396.

Bhojani, S. H. and Bhatt, N. (2020). "Wheat crop yield prediction using new activation functions in neural network." *Neural Computing and Applications*, 32(17), 13941–13951.

Boitt, M. (2014). "Using hyperspectral data to identify crops in a cultivated agricultural landscape - a case study of taita hills, kenya." *Journal of Earth Science and climate change*, 5, 232–236.

Bose, P., Kasabov, N. K., Bruzzone, L. and Hartono, R. N. (2016). "Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series." *IEEE Transactions on Geoscience and Remote Sensing*, 54(11), 6563–6573.

Bourennane, S. and Fossati, C. (2015). "Dimensionality reduction and coloured noise removal from hyperspectral images." *Remote Sensing Letters*, 6(11), 854–863.

Brownlee, J. (2019). *Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python*, Machine Learning Mastery.

Bruzzone, L., Liu, S., Bovolo, F. and Du, P. (2016). "Change detection in multitemporal hyperspectral images." In *Multitemporal Remote Sensing: Methods and Applications*, Springer International Publishing, Cham, 63–88.

Burgan, R. E. and Hartford, R. A. (1993). "Monitoring vegetation greenness with satellite data." Technical report.

Camps-Valls, G., Gómez-Chova, L., Muñoz, J., Vila-Francés, J. and Calpe, J. (2006). "Composite kernels for hyperspectral image classification." *Geoscience and Remote Sensing Letters, IEEE*, 3, 93 – 97.

Chein-I Chang and Qian Du (2004). "Estimation of number of spectrally distinct signal sources in hyperspectral imagery." *IEEE Transactions on Geoscience and Remote Sensing*, 42(3), 608–619.

Chen, C., Jiang, F., Yang, C., Rho, S., Shen, W., Liu, S. and Liu, Z. (2018). "Hyperspectral classification based on spectral–spatial convolutional neural networks." *Engineering Applications of Artificial Intelligence*, 68, 165–171.

Chen, P., Jiao, L., Liu, F., Zhao, J., Zhiqiang, Z. and Liu, S. (2016). "Semi-supervised double sparse graphs based discriminant analysis for dimensionality reduction." *Pattern Recognition*, 61, 361–378.

Chen, Y., Jiang, H., Li, C., Jia, X. and Ghamisi, P. (2016). "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks." *IEEE Transactions on Geoscience and Remote Sensing*, 54(10), 6232–6251.

Chen, Y., Lee, W. S., Gan, H., Peres, N., Fraisse, C., Zhang, Y. and He, Y. (2019). "Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages." *Remote Sensing*, 11(13).

Chen, Y., Lin, Z., Zhao, X., Wang, G. and Gu, Y. (2014). "Deep learning-based classification of hyperspectral data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), 2094–2107.

Council, N. R. (1997). *Precision Agriculture in the 21st Century: Geospatial and Information Technologies in Crop Management*, The National Academies Press, Washington, DC.

Dasgupta, S. and Gupta, A. (2003). "An elementary proof of a theorem of johnson and lindenstrauss." *Random Struct. Algorithms*, 22(1), 60–65.

Datta, A., Ghosh, S. and Ghosh, A. (2014). "Band elimination of hyperspectral imagery using partitioned band image correlation and capacitory discrimination." *International Journal of Remote Sensing*, 35(2), 554–577.

Datta, A., Ghosh, S. and Ghosh, A. (2017). "Unsupervised band extraction for hyperspectral images using clustering and kernel principal component analysis." *International Journal of Remote Sensing*, 38(3), 850–873.

Fauvel, M., Chanussot, J. and Benediktsson, J. A. (2012). "A spatial-spectral kernel-based approach for the classification of remote-sensing images." *Pattern Recogn.*, 45(1), 381–392.

Feng, F., Li, W., Du, Q. and Zhang, B. (2017). "Dimensionality reduction of hyperspectral image with graph-based discriminant analysis considering spectral similarity." *Remote Sensing*, 9(4).

Fonseca, L. M. G., Namikawa, L. M. and Castejon, E. F. (2009). "Digital Image Processing in Remote Sensing." *2009 Tutorials of the XXII Brazilian Symposium on Computer Graphics and Image Processing*, (C), 59–71.

Ghosh, J. K. and Somvanshi, A. (2008). "Fractal-based dimensionality reduction of hyperspectral images." *Journal of the Indian Society of Remote Sensing*, 36(3), 235–241.

Gomez-Chova, L., Calpe, J., Camps-Valls, G., Martin, J. D., Soria, E., Vila, J., Alonso-Chorda, L. and Moreno, J. (2003). "Semi-supervised classification method for hyperspectral remote sensing images." In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*, volume 3, 1776–1778.

Green, A. A., Berman, M., Switzer, P. and Craig, M. D. (1988). "A transformation for ordering multispectral data in terms of image quality with implications for noise removal." *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 65–74.

Guirguis, K., Schorn, C., Guntoro, A., Abdulatif, S. and Yang, B. (2021). "Seld-tcn: Sound event localization & detection via temporal convolutional networks." *2020 28th European Signal Processing Conference (EUSIPCO)*.

Hadoux, X., Gorretta, N. and Rabatel, G. (2012). "Weeds-wheat discrimination using hyperspectral imagery." In *CIGR-Ageng 2012. International Conference on Agricultural Engineering*, Valencia, Spain, 8–12.

Ham, J., Yangchi Chen, Crawford, M. M. and Ghosh, J. (2005). "Investigation of the random forest framework for classification of hyperspectral data." *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), 492–501.

Hu, W., Huang, Y., Wei, L., Zhang, F. and Li, H. (2015). "Deep convolutional neural networks for hyperspectral image classification." *Journal of Sensors*, 2015, 258619.

Hu, W.-S., Li, H.-C., Pan, L., Li, W., Tao, R. and Du, Q. (2019). "Feature extraction and classification based on spatial-spectral convlstm neural network for hyperspectral images.").

Huang, H., Luo, F., Liu, J. and Yang, Y. (2015). "Dimensionality reduction of hyperspectral images based on sparse discriminant manifold embedding." *ISPRS Journal of Photogrammetry and Remote Sensing*, 106, 42–54.

Jensen, J. (2007). *Remote Sensing of the Environment: An Earth Resource Perspective.*, Prentice Hall, 2 ed edition.

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K.-M., Gerber, J. S., Reddy, V. R. and Kim, S.-H. (2016). "Random forests for global and regional crop yield predictions." *PLOS ONE*, 11(6), 1–15.

Ji, S., Zhang, C., Xu, A., Shi, Y. and Duan, Y. (2018). "3d convolutional neural networks for crop classification with multi-temporal remote sensing images." *Remote Sensing*, 10(1).

Jiang, D., Yang, X., Clinton, N. and Wang, N. (2004). "An artificial neural network model for estimating crop yields using remotely sensed information." *International Journal of Remote Sensing*, 25(9), 1723–1732.

Jiang, H., Hu, H., Zhong, R., Xu, J., Xu, J., Huang, J., Wang, S., Ying, Y. and Lin, T. (2020). "A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level." *Global Change Biology*, 26(3), 1754–1766.

Jiao, L., Liang, M., Chen, H., Yang, S., Liu, H. and Cao, X. (2017). "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification." *IEEE Transactions on Geoscience and Remote Sensing*, 55(10), 5585–5599.

Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A. and Bédard, F. (2016). "Crop yield forecasting on the canadian prairies by remotely sensed vegetation in-

dices and machine learning methods." *Agricultural and Forest Meteorology*, 218-219, 74 – 84.

Johnson, W. and Lindenstrauss, J. (1984). "Extensions of Lipschitz mappings into a Hilbert space." In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemporary Mathematics*, American Mathematical Society, 189–206.

Junying, S. and Ning, S. (2008). "A Dimensionality reduction algorithm of hyper spectral image based on fract analysis." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII(Part B7), 297–302.

Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C. and Anderson, M. (2020). "Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest." *Environmental Research Letters*, 15(6), 064005.

Khaki, S. and Wang, L. (2019). "Crop yield prediction using deep neural networks." *Frontiers in Plant Science*, 10, 621.

Khaki, S., Wang, L. and Archontoulis, S. V. (2020). "A cnn-rnn framework for crop yield prediction." *Frontiers in Plant Science*, 10, 1750.

Khodr, J. and Younes, R. (2011). "Dimensionality reduction on hyperspectral images: A comparative review based on artificial datas." In *2011 4th International Congress on Image and Signal Processing*, volume 4, 1875–1883.

Kingma, D. P. and Ba, J. (2014). "Adam: A method for stochastic optimization.")cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

Koonsanit, K., Jaruskulchai, C. and Eiumnoh, A. (2012). "Band Selection for Dimension Reduction in Hyper Spectral Image Using Integrated Information Gain and Principal Components Analysis Technique." *International Journal of Machine Learning and Computing*, 2(3), 248–251.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks." In Pereira, F., Burges, C. J. C., Bottou, L. and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc., 1097–1105.

Lange, J., Cavallaro, G., Götz, M., Erlingsson, E. and Riedel, M. (2018). "The influence of sampling methods on pixel-wise hyperspectral image classification with 3d convolutional neural networks." In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2087–2090.

Laparra, V., Malo, J. and Camps-Valls, G. (2015). "Dimensionality reduction via regression in hyperspectral imagery." *IEEE Journal of Selected Topics in Signal Processing*, 9(6), 1026–1036.

Lea, C., Vidal, R., Reiter, A. and Hager, G. D. (2016). "Temporal convolutional networks: A unified approach to action segmentation." *CoRR*, abs/1608.08242.

Li, D., Wang, X. and Cheng, Y. (2019). "Spatial-spectral neighbour graph for dimensionality reduction of hyperspectral image classification." *International Journal of Remote Sensing*, 40(11), 4361–4383.

Li, J., Xi, B., Li, Y., Du, Q. and Wang, K. (2018). "Hyperspectral classification based on texture feature enhancement and deep belief networks." *Remote Sensing*, 10(3).

Li, W., Feng, F., Li, H. and Du, Q. (2018). "Discriminant analysis-based dimension reduction for hyperspectral image classification: A survey of the most recent advances and an experimental comparison of different techniques." *IEEE Geoscience and Remote Sensing Magazine*, 6(1), 15–34.

Li, W., Wu, G., Zhang, F. and Du, Q. (2017). "Hyperspectral image classification using deep pixel-pair features." *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), 844–853.

Li, Y., Xie, W. and Li, H. (2017a). "Hyperspectral image reconstruction by deep convolutional neural network for classification." *Pattern Recognition*, 63, 371–383.

Li, Y., Zhang, H. and Shen, Q. (2017b). "Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network." *Remote Sensing*, 9(1).

Liu, S., Du, Q. and Tong, X. (2017). "Band selection for change detection from hyperspectral images." In Velez-Reyes, M. and Messinger, D. W., editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIII*, volume 10198, SPIE, International Society for Optics and Photonics, 283 – 289.

Liu, X. and Bo, Y. (2015). "Object-based crop species classification based on the combination of airborne hyperspectral images and lidar data." *Remote Sensing*, 7(1), 922–950.

Lodha, S. P. and Kamlapur, S. M. (2014). "Dimensionality reduction techniques for Hyperspectral Images." *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, 3(10), 1–5.

Luo, G., Chen, G., Tian, L., Qin, K. and Qian, S.-E. (2016). "Minimum noise fraction versus principal component analysis as a preprocessing step for hyperspectral imagery denoising." *Canadian Journal of Remote Sensing*, 42(2), 106–116.

Ly, N. H., Du, Q. and Fowler, J. E. (2014). "Sparse graph-based discriminant analysis for hyperspectral imagery." *IEEE Transactions on Geoscience and Remote Sensing*, 52(7), 3872–3884.

Ma, X., Geng, J. and Wang, H. (2015). "Hyperspectral image classification via contextual deep learning." *EURASIP Journal on Image and Video Processing*, 2015(1), 20.

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F. and Fritschi, F. B. (2020). "Soybean yield prediction from uav using multimodal data fusion and deep learning." *Remote Sensing of Environment*, 237, 111599.

Martínez, A. M. and Kak, A. C. (2001). "Pca versus lda." *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(2), 228–233.

Mei, S., Ji, J., Geng, Y., Zhang, Z., Li, X. and Du, Q. (2019). "Unsupervised spatial–spectral feature learning by 3d convolutional autoencoder for hyperspectral classification." *IEEE Transactions on Geoscience and Remote Sensing*, 57(9), 6808–6820.

Melgani, F. and Bruzzone, L. (2004). "Classification of hyperspectral remote sensing images with support vector machines." *Geoscience and Remote Sensing, IEEE Transactions on*, 42, 1778 – 1790.

Mohan, A., Sapiro, G. and Bosch, E. (2007). "Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images." *IEEE Geoscience and Remote Sensing Letters*, 4(2), 206–210.

Moharana, S. and Dutta, S. (2014). "Hyperspectral remote sensing of paddy crop using insitu measurement and clustering technique." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL8, 845–851.

Mughees, A., Ali, A. and Tao, L. (2017). "Hyperspectral image classification via shape-adaptive deep learning." In *2017 IEEE International Conference on Image Processing (ICIP)*, 375–379.

Murinto and PA, N. R. D. (2017). "Dimensionality reduction using hybrid support vector machine and discriminant independent component analysis for hyperspectral image." *International Journal of Advanced Computer Science and Applications*, 8(11).

Nevavuori, P., Narra, N. and Lipping, T. (2019). "Crop yield prediction with deep convolutional neural networks." *Computers and Electronics in Agriculture*, 163, 104859.

Nhaila, H., Merzouqi, M., Sarhrouni, E. and Hammouch, A. (2015). "Hyperspectral images classification and dimensionality reduction using homogeneity feature and mutual information." In *2015 Intelligent Systems and Computer Vision (ISCV)*, 1–5.

Pacifici, F., Du, Q. and Prasad, S. (2013). "Report on the 2013 ieee grss data fusion contest: Fusion of hyperspectral and lidar data [technical committees]." *IEEE Geoscience and Remote Sensing Magazine*, 1(3), 36–38.

Pan, B., Shi, Z. and Xu, X. (2018). "Mugnet: Deep learning for hyperspectral image classification using limited samples." *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 108–119. Deep Learning RS Data.

Paoletti, M., Haut, J., Plaza, J. and Plaza, A. (2018). "A new deep convolutional neural network for fast hyperspectral image classification." *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 120–147. Deep Learning RS Data.

Paoletti, M. E., Haut, J. M., Fernandez-Beltran, R., Plaza, J., Plaza, A., Li, J. and Pla, F. (2019a). "Capsule networks for hyperspectral image classification." *IEEE Transactions on Geoscience and Remote Sensing*, 57(4), 2145–2160.

Paoletti, M. E., Haut, J. M., Fernandez-Beltran, R., Plaza, J., Plaza, A. J. and Pla, F. (2019b). "Deep pyramidal residual networks for spectral–spatial hyperspectral image classification." *IEEE Transactions on Geoscience and Remote Sensing*, 57(2), 740–754.

Pearson, R. L. and Miller, L. D. (1972). "Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie, Pawnee National Grasslands, Colorado.." In *U.S. International Biological Program., & International Symposium on Remote Sensing of Environment. (1972)*, 2–6.

Phillips, R. D., Watson, L., Blinn, C. E. and Wynne, R. (2008). "An adaptive noise reduction technique for improving the utility of hyperspectral data." .

Prasad, S. and Bruce, L. M. (2008). "Limitations of principal components analysis for hyperspectral target recognition." *IEEE Geosci. Remote Sensing Lett.*, 5(4), 625–629.

Rahnemoonfar, M. and Sheppard, C. (2017). "Real-time yield estimation based on deep learning." In Thomasson, J. A., McKee, M. and Moorhead, R. J., editors, *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II*, volume 10218, SPIE, International Society for Optics and Photonics, 59 – 65.

Rao, N. R. (2008). "Development of a crop - specific spectral library and discrimination of various agricultural crop varieties using hyperspectral imagery." *International Journal of Remote Sensing*, 29(1), 131–144.

Rouse, J. W., Hass, R. H., Schell, J. A., Deering, D. W. and Harlan, J. C. (1974). "Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation." Technical report, Texas A & M University, College Station , Texas.

Roweis, S. T. and Saul, L. K. (2000). "Nonlinear dimensionality reduction by locally linear embedding.." *Science (New York, N.Y.)*, 290(5500), 2323–2326.

Roy, S. K., Krishna, G., Dubey, S. R. and Chaudhuri, B. B. (2020). "Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification." *IEEE Geoscience and Remote Sensing Letters*, 17(2), 277–281.

Santara, A., Mani, K., Hatwar, P., Singh, A., Garg, A., Padia, K. and Mitra, P. (2017). "Bass net: Band-adaptive spectral-spatial feature learning neural network for hyperspectral image classification." *IEEE Transactions on Geoscience and Remote Sensing*, 55(9), 5293–5301.

Saravi, B., Nejadhashemi, A. P. and Tang, B. (2020). "Quantitative model of irrigation effect on maize yield by deep neural network." *Neural Computing and Applications*, 32(14), 10679–10692.

Sateesh, B. and Sridhar, S. S. (2014). "Classification of Hyper Spectral Images using the Unsupervised Technique." *International Journal of Business Management and Economic Research(IJBMER)*, 5(1), 2507–2510.

Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P. and Ciampitti, I. A. (2020). "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil." *Agricultural and Forest Meteorology*, 284, 107886.

Shekoofa, A., Emam, Y., Shekoufa, N., Ebrahimi, M. and Ebrahimie, E. (2014). "Determining the most important physiological and agronomic traits contributing to maize grain yield through machine learning algorithms: A new avenue in intelligent agriculture." *PLOS ONE*, 9(5), 1–9.

Shidnal, S., Latte, M. V. and Kapoor, A. (2019). "Crop yield prediction: two-tiered machine learning model approach." *International Journal of Information Technology*.

Son, N. T., Chen, C. F., Chen, C. R., Chang, L. Y., Duc, H. N. and Nguyen, L. D. (2013). "Prediction of rice crop yield using modis evi-lai data in the mekong delta, vietnam." *International Journal of Remote Sensing*, 34(20), 7275–7292.

Sun, J., Di, L., Sun, Z., Shen, Y. and Lai, Z. (2019). "County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model." *Sensors*, 19(20), 4363.

Sun, W. and Du, Q. (2018). "Graph-Regularized Fast and Robust Principal Component Analysis for Hyperspectral Band Selection." *IEEE Transactions on Geoscience and Remote Sensing*, 56(6).

Tarabalka, Y., Fauvel, M., Chanussot, J. and Benediktsson, J. (2010). "Svm- and mrf-based method for accurate classification of hyperspectral images." *Geoscience and Remote Sensing Letters, IEEE*, 7, 736 – 740.

Terliksiz, A. S. and Altýlar, D. T. (2019). "Use of deep neural networks for crop yield prediction: A case study of soybean yield in lauderdale county, alabama, usa." In *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–4.

Thenkabail, P. S., Smith, R. B. and De Pauw, E. (2000). "Hyperspectral vegetation indices and their relationships with agricultural crop characteristics." *Remote Sensing of Environment*, 71(2), 158–182.

THENKABAIL, P. S., WARD, A. D. and LYON, J. G. (1994). "Landsat-5 thematic mapper models of soybean and corn crop characteristics." *International Journal of Remote Sensing*, 15(1), 49–61.

Thenkabail P.S., Smith R.B., P. E. (2000). "Hyperspectral Vegetation Indices for Determining Agricultural Crop Characteristics." *Remote Sensing of Environment*, 1–37.

Theodoridis, S. and Koutroumbas, K. (2006). *Pattern Recognition, Third Edition*, Academic Press, Inc., Orlando, FL, USA.

Tsouli Fathi, M., Ezziyyani, M., Ezziyyani, M. and El Mamoune, S. (2020). "Crop yield prediction using deep learning in mediterranean region." In Ezziyyani, M., editor,

Advanced Intelligent Systems for Sustainable Development (AI2SD'2019), Springer International Publishing, Cham, 106–114.

van den Oord, A., Kalchbrenner, N. and Kavukcuoglu, K. (2016). "Pixel recurrent neural networks." *CoRR*, abs/1601.06759.

van Klompenburg, T., Kassahun, A. and Catal, C. (2020). "Crop yield prediction using machine learning: A systematic literature review." *Computers and Electronics in Agriculture*, 177, 105709.

Venkatasubramanian, S. and Wang, Q. (2011). "The johnson-lindenstrauss transform: An empirical study." In *Proceedings of the Meeting on Algorithm Engineering & Expermiments*, ALENEX '11, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 164–173.

Wan, Y.-q., Fan, Y.-h. and Jin, M.-s. (2021). "Application of hyperspectral remote sensing for supplementary investigation of polymetallic deposits in huaniushan ore region, northwestern china." *Scientific Reports*, 11(1), 440.

Wang, A. X., Tran, C., Desai, N., Lobell, D. and Ermon, S. (2018). "Deep transfer learning for crop yield prediction with remote sensing data." In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS '18, Association for Computing Machinery, New York, NY, USA.

Wang, J. and Chang, C. I. (2006). "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis." *IEEE Transactions on Geoscience and Remote Sensing*, 44(6), 1586–1600.

Wang, S. and Wang, C. (2015). "Research on dimension reduction method for hyperspectral remote sensing image based on global mixture coordination factor analysis." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7/W4, 159–167.

Wang, X., Huang, J., Feng, Q. and Yin, D. (2020a). "Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of china with deep learning approaches." *Remote Sensing*, 12(11).

117

Wang, Y., Zhang, Z., Feng, L., Du, Q. and Runge, T. (2020b). "Combining multi-source data and machine learning approaches to predict winter wheat yield in the conterminous united states." *Remote Sensing*, 12(8).

Wu, H. and Prasad, S. (2018). "Semi-supervised dimensionality reduction of hyperspectral imagery using pseudo-labels." *Pattern Recognition*, 74, 212–224.

Xu, X., Gao, P., Zhu, X., Guo, W., Ding, J., Li, C., Zhu, M. and Wu, X. (2019). "Design of an integrated climatic assessment indicator (icai) for wheat production: A case study in jiangsu province, china." *Ecological Indicators*, 101, 943 – 953.

Xu, Y., Zhang, L., Du, B. and Zhang, F. (2018). "Spectral–spatial unified networks for hyperspectral image classification." *IEEE Transactions on Geoscience and Remote Sensing*, 56(10), 5893–5909.

Yalcin, H. (2019). "An approximation for a relative crop yield estimate from field images using deep learning." In *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–6.

Yang, Q., Shi, L., Han, J., Zha, Y. and Zhu, P. (2019). "Deep convolutional neural networks for rice grain yield estimation at the ripening stage using uav-based remotely sensed images." *Field Crops Research*, 235, 142 – 153.

Yu, S., Jia, S. and Xu, C. (2017). "Convolutional neural networks for hyperspectral image classification." *Neurocomputing*, 219, 88–98.

Zhang, L., Zhang, Z., Luo, Y., Cao, J. and Tao, F. (2020). "Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in china using machine learning approaches." *Remote Sensing*, 12(1).

Zhang, M., Hendley, P., Drost, D., O'Neill, M. and Ustin, S. (1999). *Corn and Soybean Yield Indicators Using Remotely Sensed Vegetation Index*, 1475–1481. John Wiley & Sons, Ltd.

Zhang, X., Sun, Y., Shang, K., Zhang, L. and Wang, S. (2016). "Crop classification based on feature band set construction and object-oriented approach using hyper-

spectral images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9), 4117–4128.

Zhao, W. and Du, S. (2016). "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach." *IEEE Transactions on Geoscience and Remote Sensing*, 54(8), 4544–4554.

Zheng, G. and Moskal, L. M. (2009). "Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors." *Sensors*, 9(4), 2719–2745.

Zhong, Z., Li, J., Luo, Z. and Chapman, M. (2018). "Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework." *IEEE Transactions on Geoscience and Remote Sensing*, 56(2), 847–858.

Zhou, F., Hang, R., Liu, Q. and Yuan, X. (2019). "Hyperspectral image classification using spectral-spatial lstms." *Neurocomputing*, 328, 39–47. Chinese Conference on Computer Vision 2017.

PUBLICATIONS

JOURNAL PAPERS

1. Mohan, A., Venkatesan M (2020). HybridCNN based hyperspectral image classification using multiscale spatiospectral features. Infrared Physics & Technology, Volume 108, 103326, ISSN 1350-4495. (DOI: `https://doi.org/10.1016/j.infrared.2020.103326.`, URL: `https://www.sciencedirect.com/science/article/pii/S1350449519310485`) (Impact Factor: 2.379)

2. Mohan, A., Meenakshi Sundaram, V(2020). V3O2: hybrid deep learning model for hyperspectral image classification using vanilla-3D and octave-2D convolution. J Real-Time Image Proc (2020). (DOI: `https://doi.org/10.1007/s11554-020-00966-z`, URL: `https://link.springer.com/article/10.1007/s11554-020-00966-z`) (Impact Factor: 1.968)

3. Mohan, A., Venkatesan M (2021). Multivariate Time Based Rice Crop Yield Prediction Using MODIS Data, *IEEE Transactions on Geoscience and Remote Sensing* (Paper submitted on February 21, 2021 and status is "Under Review").

CONFERENCE PAPERS

1. Mohan A., Venkatesan M. (2020) Spatial Data-Based Prediction Models for Crop Yield Analysis: A Systematic Review. In: Venkata Krishna P., Obaidat M. (eds) Emerging Research in Data Engineering Systems and Computer Communications. Advances in Intelligent Systems and Computing, vol 1054. Springer, Singapore. (DOI: `https://doi.org/10.1007/978-981-15-0135-7_33`, URL: `https://link.springer.com/chapter/10.1007/978-981-15-0135-7_33`)

2. Mohan A., Venkatesan M. (2020) Hybrid Dimensionality Reduction Technique for Hyperspectral Images Using Random Projection and Manifold Learning. In: Lu W., Zhu K.Q. (eds) Trends and Applications in Knowledge Discovery and Data Mining. PAKDD 2020. Lecture Notes in Computer Science, vol 12237. Springer, Cham. (DOI: `https://doi.org/10.1007/978-3-030-60470-7_12`, URL: `https://link.springer.com/chapter/10.1007/978-3-030-60470-7_12`)

3. Mohan A., Venkatesan M. (2020) Spatiospectral Feature Extraction and Classification of Hyperspectral Images Using 3D-CNN+ ConvLSTM Model. In: Goel N., Hasan S., Kalaichelvi V. (eds) Modelling, Simulation and Intelligent Computing. MoSICom 2020. Lecture Notes in Electrical Engineering, vol 659. Springer, Singapore. (DOI: `https://doi.org/10.1007/978-981-15-4775-1_18`, URL: `https://link.springer.com/chapter/10.1007/978-981-15-4775-1_18`)

BIODATA

Name: ALKHA MOHAN

Date of Birth: 25th May 1990

Gender: Female

Marital Status: Married

Father's Name: S Mohan

Mother's Name: N Ambika

Address: Vilayil House
Thekkemuri
East-Kallada (P.O)
Kollam
Kerala-691502

E-mail: mohan.alkha@gmail.com

Mobile: +91 9746922087

Qualification: B.Tech in Computer Science & Engineering

M.Tech in Computer Science & Engineering
 - Image Processing

(College of Engineering, Karunagappally, CUSAT)

Areas of Interest: Data Science, Image Processing