Deep Learning Based Multi-temporal Crop Mapping Accounting for Sample Imbalance

MARGARRET ASHMITHA ANTONY RAVI
August, 2022

SUPERVISORS:
Mr Prabhakar Alok Verma, IIRS
Dr M. Belgiu, ITC

DEEP LEARNING BASED MULTI-TEMPORAL CROP MAPPING ACCOUNTING FOR SAMPLE IMBALANCE

MARGARRET ASHMITHA ANTONY RAVI
Enschede, The Netherlands, August, 2022

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Mr Prabhakar Alok Verma, IIRS
Dr M. Belgiu, ITC

THESIS ASSESSMENT BOARD:
Dr. C. Persello (Chair)
Dr. R.G. Garg (External Examiner, IIT Roorkee)

ABSTRACT

Accurate mapping of crop types is essential for solving problems of food security, crop inventory and to help farmers in making decisions regarding improving productions and managing agricultural practices. A variety of scientific methods starting from conventional supervised and unsupervised classifiers to machine and deep learning classifiers has been utilized to perform crop type classification from remote sensing images. In reality an agricultural field appears to have a lot of majority crops with a few scattered minority crops in a particular cropping season. Generally, crop classification studies ignore these minority crops or gather them together into a class called other crops and carry out classification. This study is aimed at providing a solution for the problem of class imbalance that occurs in deep learning based classification. Owing to the limited availability of labelled data, a 1D CNN sequential deep learning model is selected. The study area chosen is Jhansi District in the Bundelkhand Region of Uttar Pradesh in India. Sentinel 2 freely accessible optical data with a high spatial resolution of 10 m is chosen for the study. Taking into account the Rabi Season, 23 multi-temporal Sentinel 2 images covering Rabi crop season is considered. The majority crop class identified is wheat and the minority crop classes identified are mustard and chickpea. The Hyper-parameter of the model is optimized through Bayesian Optimization.

Two approaches are considered in this study to address the problem of class imbalance. They are algorithm level balancing techniques and data level balancing techniques. Cost sensitive learning is introduced in the 1D CNN model by making changes to the loss function. Three different loss functions such as categorical cross entropy, focal loss and class weighted loss are employed to arrive at one best loss function that can solve imbalance problem in crop classification. Secondly, data level balancing methods such as under sampling, oversampling and hybrid methods (combination of under sampling and oversampling) are applied to the training dataset and the 1D CNN model is trained and tested with these balanced dataset. The results are assessed based on accuracy metrics such as Overall Accuracy, F1 –score, Precision, and Recall. G-Mean score is preferred to assess the accuracy of individual classes as it is proved to be utilized in imbalanced classification problem.

It was observed that class weighted loss performed the best out of all loss functions with an overall accuracy of 70.37 % and a G-Mean score of 56.24%. When it comes to data level balancing techniques, under sampling outperformed oversampling as it lead to the generation new incorrect artificial samples. After carefully interpreting the results of both these techniques, and comparing those with sampling techniques algorithm level balancing with class weighted loss was chosen to be the best out of all the different methods incorporated to solve the imbalance in the classification problem.

Keywords: Crop Mapping, CNN, Time-Series, Deep Learning, Class Imbalance, Oversampling and under sampling, cost-sensitive learning.

ACKNOWLEDGEMENTS

In the first place, I wish to express my gratitude to the almighty for providing me with the wonderful opportunity to pursue a Master's degree in Geoinformatics at two of the best institutions in the world.

My sincere thanks go out to the two most wonderful supervisors, Mr. Prabhakar Alok Verma and Dr. Mariana Belgiu, who provided constant support, motivation, technical insights and constructive feeddbacks throughout the MSc research phase. The successful completion of my thesis would not have been possible without their assistance.

I am greateful to Dr. Sameer Saran, and course coordinators Dr. Kapil Oberai and Dr. Sander Oude Elberink for making the entire process of MSc program run smoothly.

I would also like to thank Dr. Alfred Stein and Dr. Claudio Persello for their critical comments and feedbacks. As well, I would like to extend my sincere thanks to Sina for his assistance with my thesis.

I am forever thankful to my family, Appa, Amma and my sister Adira for their blessings, emotional and moral support throughout the course of the MSc program.

Finally, I would like to acknowledge the contribution of International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) and Agriculture and Soils Department, IIRS in helping me by providing ground truth data needed to kick start my research.

.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS5

ANN- Artificial Neural Networks

AVHRR- Advanced Very High Resolution Radiometer

CNN- Convolutional Neural Network

DT – Decision Trees

DTW- Dynamic Time Warping

ESA-European Space Agency

EVI- Enhanced Vegetation Index

FL – Focal Loss

GEE – Google Earth Engine

MLP- Multi-Layer Perceptron

MODIS - Moderate Resolution Imaging Spectroradiometer

NDVI- Normalised Difference Vegetation Index

RF – Random Forest

ROS- Random Oversampling

RUS – Random Under sampling

SAR- Synthetic Aperture Radar

SMOTE- Synthetic Minority Oversampling Technique

SVM – Support vector Machine

1. INTRODUCTION

1.1. Research Background and Problem Statement

Agriculture is an important sector that aids in the production of food, fiber, and energy, which contributes significantly to a country's economy. In light of the growing global population, climate change, and Sustainable Development goals demand the agriculture industry to be effective and sustainable, to ensure ensuring food safety and security (Martos et al., 2021). Particularly, Goal 2 – Zero Hunger of the Sustainable Development Goals designed by the United Nations thrives on ending hunger and achieving food security while promoting sustainable agriculture by 2030. To accomplish this, accurate mapping and efficient monitoring of the agricultural landscape are essential. To sustain and enhance crop production, farmers and stakeholders must make informed decisions on many related activities(Z. Sun et al., 2020). Crop types maps and estimates of their extent provide essential information for agricultural monitoring and management (Inglada et al., 2016). In this context, crop type classification provides information about production, yield estimation, crop rotation, crop stress detection, crop damage measurement, agricultural activity monitoring and crop insurance (Gumma et al., 2020).

"In situ crop type mapping is labour-intensive time-consuming and expensive" (Adrian et al., 2021; Orynbaikyzy et al., 2019). Remote Sensing plays a vital role since it delivers consistent, repetitive data with a global coverage allowing for near-real-time monitoring of agricultural croplands with more efficient and scalable crop mapping. Satellite Remote Sensing is recognized as a reliable and trustworthy tool for crop species identification and area estimation since 1970 (Alabi et al., 2016). Earlier, coarser resolution (1 km) NDVI derived from the Advanced Very High Resolution Radiometer (AVHRR) was used for crop-related studies (Doraiswamy & Cook, 1995; Gallo & Flesch, 1989). Cropland mapping has made substantial use of the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on board the Terra and Aqua satellites, which provides nearly daily global coverage. (Y. Chen et al., 2018; Massey et al., 2017; Zhong et al., 2011). For sustainable agriculture production and food security, up to date cropland distribution maps that are precise, accurate, and have a high spatial resolution are essential. (Teluguntla et al., 2015; Thenkabail et al., 2012). With its historical dataset, Landsat aids in long-term crop mapping and change detection studies(Xu et al., 2018). A considerable number of crop mapping studies use Copernicus Sentinel 2, owing to its high spatial resolution of 10 meters and public availability of dataset (Hernandez et al., 2020; Immitzer et al., 2016; Pinjarla et al., 2019). Despite providing acceptable results, cloud cover and shadows leading to gaps in optical imagery are major drawbacks that substantially reduce the classification accuracy. In contrast to optical-only classification, Van Tricht et al., (2018) demonstrated that the combining all-weather radar and optical data produced more information thereby improving the accuracy of crop classification.

Furthermore, using SAR data that is rich in spatial and spectral information may help distinguish specific crop types that are difficult to identify with only optical images. (Inglada et al., 2016). Crop type mapping based on multi-temporal methods is preferable than single-date image analyses due to the dynamic nature of agricultural systems since the seasonality of crops is effectively captured when utilising multi-temporal images (Foerster et al., 2012).

MODIS-based NDVI time-series data in 250 m resolution has been extensively used in phenology-based research works (Massey et al., 2017). Zhong et al., (2011) adopts piecewise logistic function and asymmetric double-sigmoid function to extract phenological parameters from 250 m MODIS data to classify major crop types present in the San Joaquin Valley, California. Phenological metrics calculated from MODIS Normalised Difference Vegetation Index and Enhanced Vegetation Index are used by Yan et al., (2015) for mapping different vegetation cover types. High spatiotemporal resolution data acquired from MODIS and Landsat 8 NDVI fusion is used to map crop species and rotation types using Random Forest(RF) and Decision Tree(DT) Classifier (Li et al., 2021).

Because of its ability to generate high classification results with a small number of training samples and irregular satellite image time series, Dynamic Time Warping (DTW) has been effectively employed for crop mapping (Belgiu et al., 2020). Belgiu & Csillik, (2018) evaluated Time-Weighted DTW (TWDTW) method using Sentinel-2 time series and performed pixel-based and object-based classifications to map crop types. In this study, the object-based classification provided better accuracy when compared to the pixel-based approach. Based on Sentinel 2A/B multi-temporal data, winter wheat was identified using phenology time weighted dynamic time warping (PT-DTW) method and an overall classification accuracy of 89.90% was achieved (Dong et al., 2020).

Over the years, several shallow machine-learning classifiers like RF, DT, Support Vector Machines (SVM) and Neural Networks (NN) have been used by researchers for mapping complex heterogeneous agricultural landscapes(Li et al., 2021; Shelestov et al., 2017). A large number of studies compare the performance of one algorithm against the other and RF and SVM seem to provide better classification results (Fang et al., 2020; Feng et al., 2019). Machine Learning Shallow classifiers require optimal feature selection and exhaustive parameter tuning to reach high accuracy (Moreno-Revelo et al., 2021). Some of the limitations of shallow machine learning shallow classifiers are overcome by Deep Learning approaches. Deep Learning is a subset of Machine Learning which can be used to solve complex problems with a high precision and performance using a hierarchical representation of data employing different convolutions. Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) are the major DL methods.

Out of these, CNN has been widely utilized in Remote Sensing for crop classification problems (Adrian et al., 2021; Moreno-Revelo et al., 2021). Some of the most common architectures used in Deep Learning for crop classification are SegNet, 2D UNet and 3D UNet. Zhong et al., (2019) concluded that compared to long short-term memory (LSTM) of RNN family, which is widely used for time-series studies, by using one-dimensional convolutional (Conv1D) layers, a deep network model was able to produce a good classification result. Adding to this, the study recommends utilising a 3D spatiotemporal convolution model in future works to produce high accuracy. Although, a 3D CNN which would capture the spectral, temporal and spatial information of crops would beneficial for producing highly accurate crop classification, this method is limited to studies having training data with larger spatial coverage of the study area. In case of studies with limited training samples, 1D CNNs are preferred as they have demonstrated good performance(Wei et al., 2019).

Neglecting minority classes is a significant problem that many studies fail to address. Most common approach to resolve this problem is to group these minority crops into a class named 'other crops' and perform classification (Ibrahim et al., 2021; Zhao et al., 2019). Yet, these infrequent minority crops present in the fields are mapped with poor accuracy owing to lesser training samples. This problem of class imbalance has been addressed by incorporating sampling techniques such as under sampling and oversampling in Machine Learning studies for crop classification (Douzas et al., 2019; Waldner et al., 2019). Under sampling is the process of removing majority class instances whereas oversampling creates new minority class instances. Previous deep learning studies have tried to solve the problem of class imbalance using cost sensitive learning techniques and data augmentation techniques (Ando & Huang, 2017; Schaefer et al., 2021; Waldner et al., 2019). This research will focus on implementing crop classification using a 1D Convolutional Neural Network architecture capable to account for the imbalanced classification problem. In addition to this, the impact of oversampling and under sampling techniques for addressing class imbalance is also studied.

1.2. Research Objectives and Questions

The main objective of this research is to evaluate the performance of 1D CNN Deep learning architecture for crop type mapping. In this research, the following sub-objectives and the corresponding research questions will be addressed,

1) To design and implement a deep learning model capable of addressing the sample imbalance challenge in crop type mapping.

- What are the number of epochs, neurons, learning rate, dropout rate and batch size required to produce accurate classification results?
- What is the final loss and accuracy of the proposed deep learning model?

2) To assess the sensitivity of the deep learning model with different loss functions to sample distribution of crop types.

- Does the developed deep learning model increase the accuracy of minority classes?

3) To compare the developed deep learning model with the sample balancing techniques such as under sampling and oversampling.

- Which data sampling method is suitable for the current study and improves the accuracy of classification?
- To what extent does the developed deep learning model outperform the class balancing techniques?

1.3. Innovation Aimed at

The innovation in this research work is focussed on developing a novel deep learning based classification algorithm capable of addressing imbalanced class distribution problem in crop type mapping. Different cost-sensitive learning techniques are explored to arrive at the deep learning model with the best loss function which can account for the problem of class imbalance.

1.4. Structure of the Thesis

There are six chapters in this thesis. The rationale for the research and the problem statement are discussed in the first chapter, "Introduction." Along with the innovation used in the work, this chapter also covers the main research objective, sub objectives and relevant research questions. The details of earlier research works done for applications similar to this research are included in the second chapter, "Literature Review." The study region and datasets that were selected for this research are explained in depth in the third chapter, "Study Area and Datasets Used," along with the reasons of why they were selected. The fourth chapter, titled "Research Methodology," goes into great detail about this study's methodology. It also explains in depth on how different crop classes were split up and employed in this study. Additionally, the preparation of data followed by the implementation of Algorithm Level Balancing and Data Level Balancing techniques are discussed in detail in this Chapter. The fifth chapter, "Results and Discussions," discusses the significant results of the research. The final chapter, "Conclusions and Recommendations," presents the findings from the study's analysis along with future recommendations. This section also includes responses to each research question.

2. LITERATURE REVIEW

This chapter provides a description of relevant literature to establish the state of art and justification for the choice of method used in this research.

2.1. Crop Mapping using Multi temporal Remotesly Sensed Images

Remotely sensed time series images are being created at an extraordinary rate and scale from a variety of platforms, including Sentinel-2 (5-day frequency) and Landsat (16-day frequency). As a consequence, numerous research studies have been focussed on the objective of extracting meaningful information from these massive amounts of multi-temporal remote sensing images. Particularly in agricultural remote sensing studies, the advantage of multi-temporal images is employed to capture the seasonality of the crops. One of the most important feature of vegetation is its seasonality (Zhong et al., 2019). Vegetation phenology and plant growth dynamics retrieved from multi-temporal remote sensing images serves as an efficient tool for crop classification (Y. Xie et al., 2008). The seasonal evolution of plant activity through stages of dormancy, active growth, senescence, and back to dormancy is referred to as vegetation phenology (Penuelas et al., 2009; Richardson et al., 2013). Remote sensing derived vegetation phenology is often monitored using time series of vegetation indicators. (Cai et al., 2018)suggests that inclusion of temporal phenology information as well as uniformly distributed spatial training samples in the research domain increases classification performance. A wide range of studies use crop phenology derived from vegetation indices (NDVI, EVI) provided by satellites such as MODIS and Landsat for automatic crop mapping (Foerster et al., 2012; Li et al., 2021; Massey et al., 2017; Yan et al., 2015; Zhong et al., 2011).

2.2. Drawbacks of Machine Learning for Crop Classification

Shi et al., (2016) uses RF Classifier and investigates how providing different algorithmic parameters namely number of trees, number of features affects the performance of the RF algorithm for remote sensing image classification and concluded that these parameters had an impact of accuracy of the classifier and also the thematic map's accuracy. Feng et al., (2019) explored how distinct band information significantly helps in distinguishing different crop types and established the robustness of machine learning methods over traditional classification by comparing non-parametric classifiers (RF and SVM) with statistical-based maximum likelihood classifier. In this study, the size of the feature space was proven to impact the classification accuracy of SVM, and RF-based feature selection is suggested as a solution to this problem. Despite a little loss in accuracy, SVM that employed fewer features reduced uncertainty for several crops (Löw et al., 2013).Therefore, it is clear that selection of optimal features is essential in order to produce high classification accuracy with machine learning.

2.3. Deep Learning

Deep Learning which is an extension of classical machine learning is a type of representation learning that can extract hidden patterns from the data using hierarchical representation of data which constitutes layers of convolutions (Schmidhuber, 2015). This allows larger learning capabilities and thus higher performance and precision (Kamilaris & Prenafeta-Boldú, 2018). Also, deep learning has been demonstrated to be very effective in image classification and object detection. In the field of remote sensing, deep learning has opened up new possibilities through its application to computer vision and image analysis(Hoeser & Kuenzer, 2020; Huang et al., 2018; Kulathunga & Afanasyev, 2018; Zhu et al., 2017). "Deep learning differs from classic visual recognition methods in that it automatically learns features from enormous amounts of data rather than requiring feature engineering manually" (L. Zhang et al., 2016). This allows us to easily learn the best features for new specific problems without requiring high expertise and effort.

A deep neural network (DNN) is an artificial neural network (ANN) which has multiple layers of neurons between the input and output layers. In general, a neural network has three major layers, namely input layer which is the first layer of the neural network, hidden layer which has n number of layers constitutes the middle layer and output layer, is the last layer of the neural network. Many different types of neural networks are in practice depending on the type of application and the data used. Feed-forward neural network, Multi-Layer Perceptron (MLP), CNN, RNN, Modular neural network are some of the commonly used and popular deep neural networks. RNNs have demonstrated success in identifying and extracting temporal correlations.

2.4. Convolutional Neural Networks

CNN, also referred to as ConvNets, is a deep learning neural network intended for processing structured arrays of data, particularly images. CNNs have demonstrated high performance in image classification when compared to other types of deep learning models (Guo et al., 2017).CNN introduced by LeCun et al., (1989) made use of concepts developed by a Japanese researcher by the name of Kunihiko Fukushima Fukushima, (1988),who created a fundamental image recognition neural network. The convolutional layers in the CNN performs operations similar to image processing filters(Teoh & Rong, 2022). Convolutional neural network finds its applications in object detection (You Only Look Once (YOLO), Fast Region-based Convolutional Neural Network (R-CNN), Mask R-CNN, SSD) and image classification (AlexNet, VGG network, ResNet) Three types of CNN that are mainly used are 1D CNN, 2D CNN and 3D CNN. In 1D CNN, the input and output data is two dimensional array and the kernel moves in one direction. 1D CNN finds application in time-series. 2D CNN has three dimensional data with the kernel moving in two directions. It is used for image data problems. 3D CNN has four dimensional data and kernel moves in three dimensions (Shukla & Iriondo, 2022). Figure 1 depicts the essential layers of a CNN.

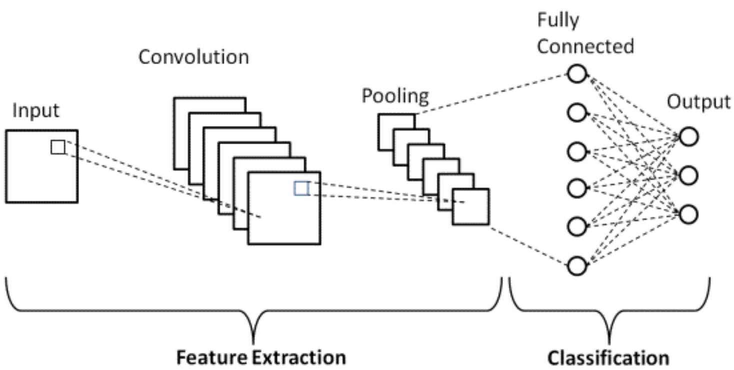

Figure 1 Basic Components of Convolutional Neural Network

2.5. Remote Sensing based Crop Mapping using Deep Learning

According to the survey done by Kamilaris & Prenafeta-Boldú, (2018), regarding the Deep Learning techniques and their applications for agriculture, conclusion is drawn that the research based on deep learning surpasses the traditional machine learning-based methods in terms of accuracy. As CNNs and RNNs extract sequential information, they are preferable to conventional ANNs in the context of image classification (Moreno-Revelo et al., 2021).

Zhong et al., (2019) used EVI time series to classify 13 summer crop categories in a very diverse irrigated agricultural system in California using different classifiers such as XGBoost, RF and, SVM in machine learning and, Long Short-Term Memory (LSTM) and 1D CNN in deep learning and established that 1D CNN architecture provided the best accuracy out of all the DL and non-deep learning models. In the same way, 1D CNN performed better than Recurrent Neural Network(RNN), Multi-Layer Perceptron (MLP), and machine learning methods like RF, SVM and XGBoost for crop mapping with optical SAR fused data(Van Tricht et al., 2018).

Convolutional long short-term (ConvLSTM) memory network and fully convolutional network (FCN) were combined by Teimouri et al., (2019) to identify different crops using multi-temporal Sentinel 1 SAR data and an overall accuracy of 86% was achieved. A 2D CNN multi-source multi-temporal crop mapping performed by Kussul et al., (2018) provides an accuracy greater than 90%. A 3D Convolutional Neural Network was used by (Ji et al., 2018) to classify crops in spatio-temporal remote sensing images and the results suggested that 3D CNN with proper structure was able to capture the crop growth dynamics effectively and outperformed 2D CNN where the temporal information is lost owing to its own mathematical restriction. The crop types in the US Corn Belt were mapped using a 3D FCN, accounting for the spatial-temporal relationship in satellite imageries (Mohammadi et al., 2021). In this research work, the employed IOU loss function learnt individual crop types better.

For application with insufficient training samples, the performance of compact 1D CNN is superior (Kiranyaz et al., 2021). 1D CNNs have several advantages in comparison to 2D CNNs and many other deep learning architectures, including lower computational complexity, a faster training process, a better capability of extracting relevant features from time-series and sequence data, and, most importantly, effective performance even on a small amount of data (Rala Cordeiro et al., 2021). In conclusion, this research adopts 1D CNN Deep learning architecture to perform crop classification taking into account the limited labelled data.

2.6. Sample Imbalance Issue in Crop Classification

Class imbalance happens when there aren't enough samples of some classes to train classifiers, which causes the rare classes to have high error probabilities (Waldner et al., 2019). In day to day classification problems, class imbalance is a common issue, where a significant difference in the number of training samples across different classes causes learning algorithms to over-generalize for the classes in the majority (Ando & Huang, 2017). For the classifier to perform at its optimum, it is crucial that there should be no or very little imbalance in the training samples (Z. Liu & Sun, 2008). Figure 2 and 3 gives a representation of sampling techniques in a graphical and spatial context.

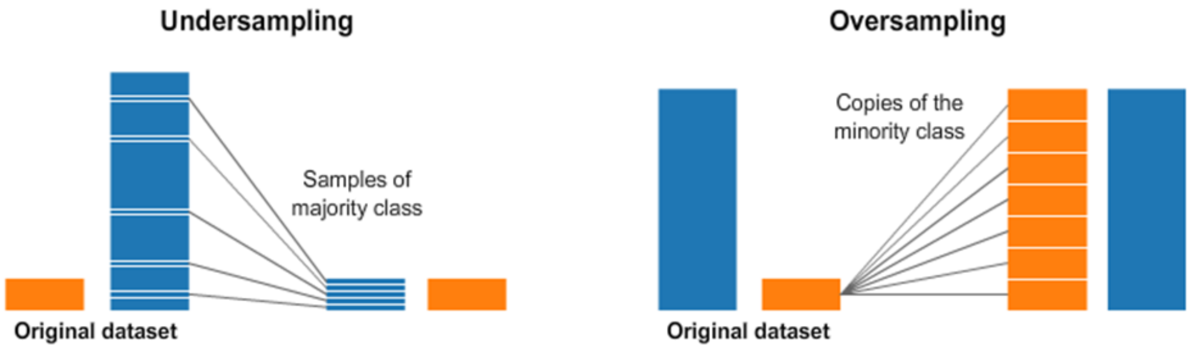

Figure 2 Representation of the concept of under sampling and oversampling. (Badr, 2019)

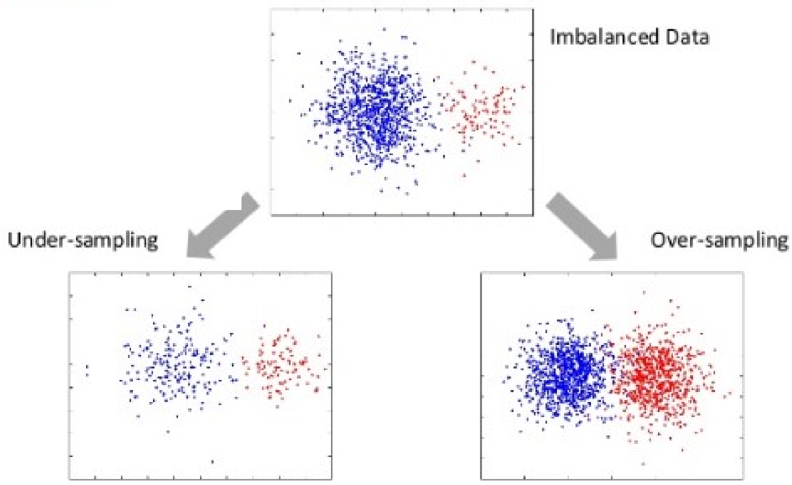

Figure 3 Spatial Depiction of under sampling and oversampling (Pilotti, 2020)

Oversampling is the most widely used data balancing technique and is proven to be robust (Ling & Li, 1998). **Random Oversampling** is the basic version of oversampling and it simply replicates randomly selected samples from the minority class ie. it generates duplicate minority class instances (Buda et al., 2017). Cluster-based oversampling clusters the majority and minority classes first using K-means algorithm and then oversampling is performed on each cluster independently (Jo & Japkowicz, 2004). This minimises both between-class and within-class imbalance. Japkowicz, (2000) used **random under sampling**, which is removes the majority class samples until their numbers are equal to those of the minority class. Under sampling is an excellent option for class imbalanced problems with large amount of data, such as millions of rows. Both random resampling methods have various short-comings. Random Oversampling can easily result in overfitting of model as duplicated samples are visited frequently (N Chawla et al., 2002). Additionally, "Over-sampling will cause an increased training time due to the increased size of the training set" (Nitesh Chawla et al., 2004). Overfitting occurs when a model fits too closely to the training data, that it becomes difficult to generalise to new data. On the other hand, random undersampling can potentially remove certain valuable information.

A novel strategy was developed by N Chawla et al., (2002) to address the overfitting problem and widen the decision zone of minority class occurrences. This method involved the generation of new synthetic samples by operating in the feature space instead of the data space. This method was called as **Synthetic Minority Oversampling Technique (SMOTE).** This was the first effort to add new training samples to the training data in order to enhance the data space and reduce the data distribution's sparsity. In short, along the path connecting a minority class samples to its K nearest neighbours, SMOTE creates new synthetic minority class samples. The k nearest neighbours is randomly chosen in accordance to the amount of over-sampling required (Nitesh Chawla, 2005). Figure 4 represents the working of SMOTE. This method greatly reduced overfitting. However, the surroundings of the artificial examples are ignored, and artificial examples may be produced in the regions adjacent to the majority samples, leading to misclassification. (Waldner et al., 2019).

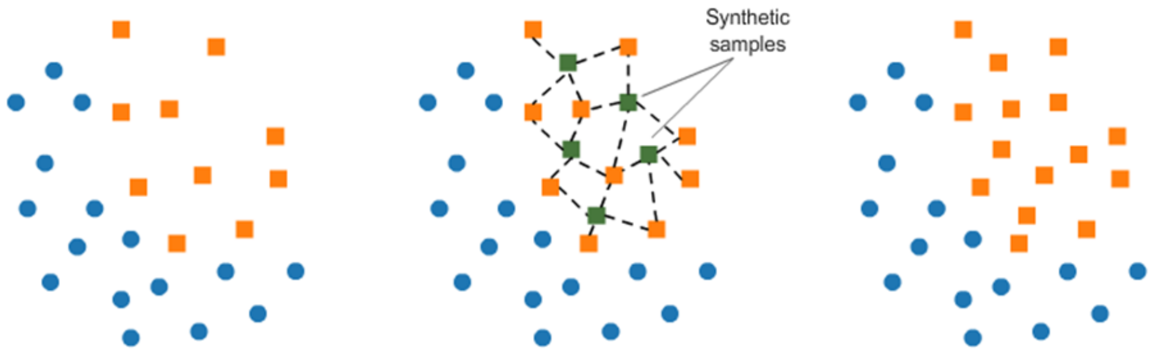

Figure 4 Synthetic Minority Oversampling Technique (SMOTE) (ALENCAR, 2018)

By introducing new minority samples in safe regions, Safe Level SMOTE aims to counteract SMOTE's shortcoming (Bunkhumpornpat et al., 2009). Each minority case is given a safe-level ratio based on the proportion of surrounding minority examples. When the safe-level ratio is close to zero, it is regarded to be noise, but when it is high, it is considered safe. Safe Level SMOTE achieves improved accuracy performance than SMOTE and Borderline-SMOTE by synthesising the minority cases more around a larger safe level (Bunkhumpornpat et al., 2009). The effect of Safe Level SMOTE in unbalanced data classification was studied by (Meidianingsih & Sartono, 2017)and they proved that safe-level SMOTE method showed a better performance than SMOTE method by using F- measure as accuracy metric.

ADAptive SYNthetic sampling (ADASYN) approach adaptively generates minority data samples according to the data distributions. The basic idea behind ADASYN is to employ a weighted distribution for each and every minority class examples based on their level of difficulty in learning, with more synthetic data created for minority class samples which are more difficult to learn than minority class examples that are easier to learn (He et al., 2008). The number of neighbours necessary to create minorities is not determined in ADASYN, which is a major distinction between ADASYN and SMOTE.

Tomek Links are pairs of very close instances but belonging to different classes(Tomek, 1976). Removing the instances of the majority class from each pair of Tomek Link increases the space between the two classes, making classification easier. Another option is cleaning where both instances are removed. Tk can be implemented by either of these methods. The working of Tomek Links in shown in the Figure 5.

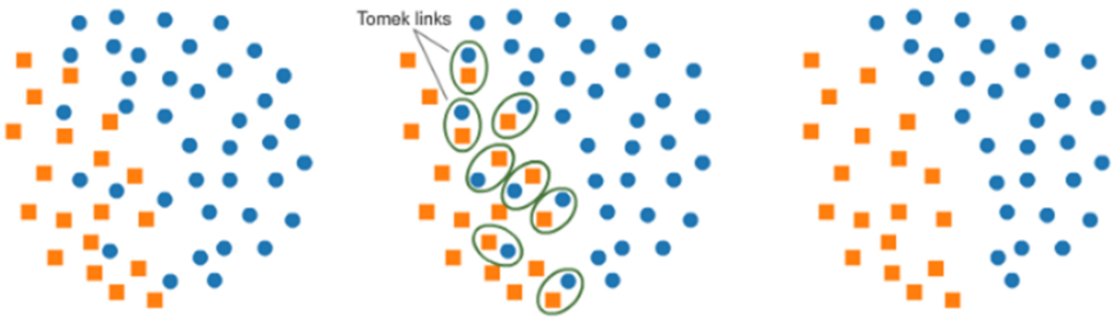

Figure 5 Tomek Links - An under sampling technique (ALENCAR, 2018)

Edited Nearest Neighbour (ENN) method was introduced by (Wilson, 1972). Most of the samples with class labels that are different from the vast majority of their k nearest neighbours are eliminated by ENN. The parameter K denotes the number of nearest neighbours and by default it takes the value of 3.

The combination of SMOTE and under-sampling performs better than plain under-sampling(Bowyer et al., 2011). A combination of SMOTE and Tomek Links employed by (Batista et al., 2003) gave promising classification results. Likewise, the work of (Batista et al., 2004) demonstrated that Smote + Tomek and Smote + ENN in particular for data sets with few minority samples, provided very good results in practice. This process where different sampling methods are combined is called as hybrid technique.

2.7. Algorithmic Level Balancing

Class imbalance is a frequent issue that has been thoroughly researched in traditional machine learning, but there is relatively little systematic study available in the context of deep learning (Buda et al., 2017).
In algorithmic level balancing methods, the imbalance issue is addressed in the training phase of the model (Boogaard et al., 2022). This is accomplished with the help of loss function. Loss function also referred to as cost function quantifies the deviation between the predicted class values from the model and the actual target class values (ground truth). In simple terms, it gives the prediction error of the neural network. During training, loss is used to learn the optimum weights in the neural network. The goal of the training process is minimizing the loss so that the accuracy of the model can be maximized.

In multi-class classification problems, **Cross Entropy loss** function is prominently utilized (Zhou et al., 2019). The majority of crop mapping techniques based on deep learning use categorical cross entropy loss to determine the crop types (Kattenborn et al., 2021). Errors of each class are given equal weights in cross entropy loss (Boogaard et al., 2022). In order to implement cost sensitive learning, focal loss and class weighted loss are used. These loss functions which provide an emphasis on minority classes are preferred in the context of classification problems with imbalanced training data. **Focal loss** (FL) is an improved and reshaped version of standard cross entropy loss which attempts to solve imbalance by reducing the weight of the loss assigned to well-classified samples at the same time increasing the weight of hard-misclassified samples (T.-Y. Lin et al., 2017). In FL, the weights are inversely proportional to the class share of the dataset (Boogaard et al., 2022). "In **Class Weighted loss**, per-class weights are computed for each class by which the focus of a learning algorithm is shifted towards minority classes and these computed weights are then used to increase the loss for minority classes and decrease the loss for majority classes" (Boogaard et al., 2022; Schiele et al., 2019).

3. STUDY AREA DESCRIPTION AND DATA ACQUISITION

3.1. Study Area Description

The study area chosen is Jhansi district located in the Bundelkhand region in central India. The Bundelkhand region covers seven districts of Uttar Pradesh (Jhansi, Chitrakoot, Lalitpur, Jalaun, Banda, Mahoba, and Hamirpur,) and six districts of Madhya Pradesh (Panna, Sagar, Tikamgarh, Chhatarpur, Damoh, and Datia). Jhansi, one of the districts of Uttar Pradesh state is located in the south western part of Bundelkhand. The geographical extent of the study area lies between 25°07' to 25°57' N latitudes and 78°10' to 79°25' E longitudes, covering a total area of 5,024 km2. Its borders are formed by the district of Jalaun to the north, Mahoba and Hamirpur to the east, Tikamgarh of Madhya Pradesh to the south, Lalitpur District to the southwest, and Datia and Bhind Districts of Madya Pradesh to the west. Pahuj, Betwa, and Dhasan are the three principal rivers that intersect or border the district. Figure 6 denotes the study area.

The climate in the region is sub-humid with four distinctive seasons, namely winter (January-February), pre-monsoon (March-May), monsoon (June-September), and post-monsoon (October-December) characterized by a hot dry summers and cold winters. Historical records show that the district receives 850 mm of rain on average annually, with the monsoon season accounting for roughly 91% of that total (Central Ground Water Board (CGWB), 2017). The region is mostly rain-fed, with less than 25% of cropland being used for double cropping (R. & Yadav, 2015). January is the coldest month of the year, with a mean daily lowest mean temperature of 9.2°C. A mean maximum temperature of 42.6 °C and a minimum temperature of 28.8 °C are recorded in May, making it the hottest month.

Presently, agricultural fields have replaced major parts of the study area which were covered by forest earlier. Approximately 62.4 percent of the district's land is used for agriculture (Kumar et al., 2021). According to the district profile, 30 percent of the land is fallow or barren. 80% of the people in the district live in villages and are relying on agriculture, cattle, and forest resources. (Singh et al., 2013). Chickpeas, wheat, paddy, sorghum, barley, lentil, maize, sesame, mustard, groundnut, soybean, urad, moong, peas, vegetables, and fruits are some of the most important crops cultivated in the Bundelkhand region(Gupta et al., 2014). Contrary to other croplands, the percentage of Rabi crops (69%) is higher than the Kharif crops (31%). Major crops grown in the Rabi season are wheat and gram and the sowing of these crops is influenced by October rainfall (Pandey et al., 2021). “Under normal rainfall years, cereals account for 54.6 % of food grain output, followed by pulses (32.4%), oilseeds (8 %), sugarcane (0.2 %), and other crops (4.8 %)” (R. & Yadav, 2015). Therefore, it is highly indispensable to map the different crop types present in the study area.

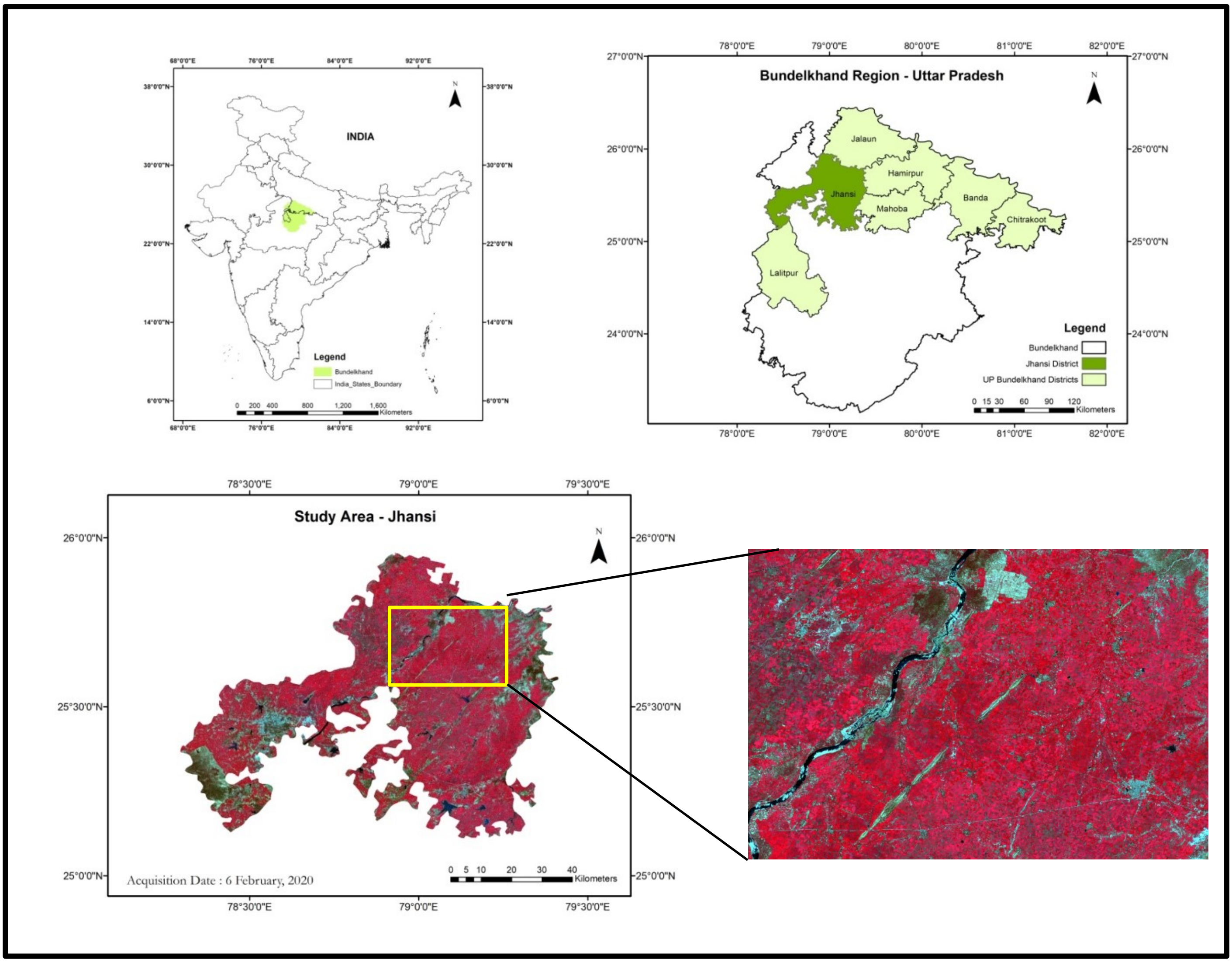

Figure 6 Study Area Map denoting a) Study Area Location in India b) Location of the Study Area in the Bundelkhand Region located in the state of Uttar Pradesh c) Sentinel 2 Image of Jhansi District d) Patch from the Study Area used for prediction.

3.2. Satellite Data

This research will utilize multispectral, wide-swath data from the Sentinel-2 satellite constellation. Sentinel 2 optical data is chosen as it has shown high capability in mapping crop types, evaluation of the state and change of vegetation and it has also been successfully applied in various crop classification studies (Belgiu & Csillik, 2018; Immitzer et al., 2016; Van Tricht et al., 2018).

The European Space Agency (ESA) originally launched Sentinel-2A in June 2015 as a part of the Copernicus Earth observation programme of the European Union. Subsequently, in March 2017, Sentinel -2B was launched. These 2 satellites together provide a high temporal resolution of 5 day**s** at the equator. The Sentinel 2 Multispectral Instrument (MSI) collects data in 13 spectral bands, visible and NIR at 10 meters, red edge and SWIR at 20 meters, and atmospheric bands at 60 meters spatial resolution. ESA provides Sentinel 2 data products at two levels: orthorectified top-of-atmosphere reflectance (Level 1C) and orthorectified atmospherically corrected surface reflectance (Level 2 A). This research will take advantage of the analysis ready Level-2A optical data (with bottom of atmosphere reflectance) available in Google Earth Engine (GEE) cloud computing platform. Sentinel 2 spectral bands with the bands selected for the study highlighted in bold is shown in the Table1.

Table 1 Spectral Bands of Sentinel 2 Image (European Space Agency)

Band and Description	Central Wavelength (µm)	Spatial Resolution (m)
B1 Coastal Aerosol	0.443	60
B2 Blue	**0.490**	**10**
B3 Green	**0.560**	**10**
B4 Red	**0.560**	**10**
B5 Vegetation Red Edge	0.705	20
B6 Vegetation Red Edge	0.740	20
B7 Vegetation Red Edge	0.783	20
B8 NIR	**0.842**	**10**
B8A Vegetation Red Edge	0.865	20
B9 Water Vapour	0.945	60
B10 SWIR - Cirrus	1.375	60
B11 SWIR	1.375	60
B12 SWIR	2.190	20

The size of landholdings not only impacts agricultural productivity but also affects the accuracy of satellite-generated spatial maps (Gumma et al., 2020). Marginal land holdings (less than 1 hectare) accounted for 40% of all holdings in Madhya Pradesh, Bundelkhand. Small and Semi medium holdings of 1-2 hectares and 2-4 hectares respectively make up the majority of Bundelkhand, whereas medium and large holdings (greater than 10 hectares) represent 30% to 45% of the total area of all holdings across

Bundelkhand districts. Considering the landholding size, Sentinel 2 data with a spatial resolution of 10 meters is chosen to capture the high variability of crop types in the croplands of Jhansi District.

3.3. Ground Truth Data / Reference Data

Vector point data provided by International Crop Research Institute for Semi-Arid Tropics (ICRISAT) and Agriculture and Soil Department, Indian Institute of Remote Sensing(IIRS) are combined together to form the ground truth data. The data is collected during the years 2019- 2020. Finally, this dataset is divided into training, validation and testing sets to perform crop type classification.

4. RESEARCH METHODOLOGY

4.1. Overall Methodology

The flowchart of the overall methodology adopted in the research is denoted in the Figure 7. This research makes use of multi-temporal satellite imagery that captures the growth phase of crops to map the crop types present in the study area. Most importantly, a major problem ignored in classification – class imbalance is also addressed. The methodology adopted is divided into three subsections namely, i) Data Pre-Processing, ii) Addressing Class Imbalance problem and, iii) Accuracy Assessment.

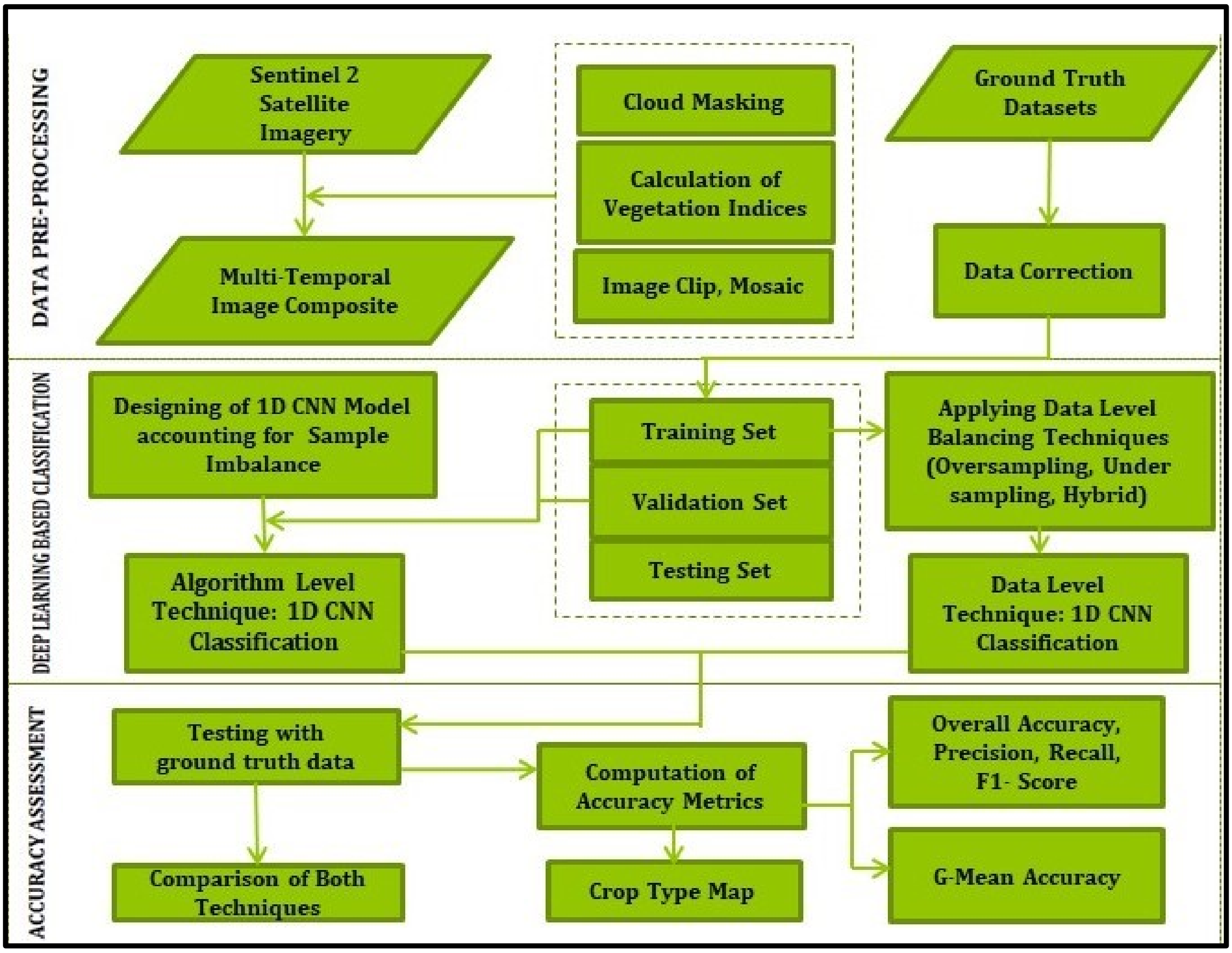

Figure 7 Overall Methodology

Pre-processing of the satellite imagery and reference data used is discussed in the section 4.2. Section 4.3 briefly discusses the CNN deep learning model utilized. A 1D CNN deep learning model is used and the hyper parameters of which are optimized by selecting the ones yielding high accuracy. The problem of imbalance is addressed by incorporating algorithm-level balancing techniques discussed in the Section 4.4 and data level balancing techniques discussed in the Section 4.5. Usage of various loss functions are attempted to solve the class imbalance issue. Generally, deep learning models use cross entropy loss function. Besides, other loss functions such as focal loss and class weight loss are incorporated in the 1D CNN deep learning model and classification of the crop types are carried out. The results of the three loss functions are compared by means of accuracy metrics to arrive at the one with the best accuracy which classifies both the minority and majority classes without any bias. Section 4.6 describes the different accuracy metrics used in detail. Different accuracy measures such as Overall Accuracy, Precision, Recall and F1 score are used. Also, G mean accuracy is preferred as it gives a true picture of individual class's accuracy and performs better for classification problems with minority classes. The accuracy of each individual class are computed and compared against each other. Next, data level balancing techniques discussed in the Section 4.5 are explored to address imbalanced class problem at data level. Three main approaches followed are: Oversampling, under sampling and hybrid methods. Nine different data level balancing techniques are used and the results of classification for each of these methods are compared in terms of accuracy metrics. Finally, the results of algorithm level balancing and data level balancing methods are compared against each and the method that performs better in classifying the majority and minority classes accurately is picked out.

4.2. Data Preparation and Preprocessing

4.2.1. Satellite Data Pre-processing

Imageries from November, 2019 to April, 2020 were selected considering the growing season of the crops from the crop calendar. The percentage of cloud cover given was 50 and cloud masking was applied. Additionally, cloud masking generates regions with no data which could not be interpolated from previous and next day images. Because continuous cloud-free images were not available, temporal interpolation was not feasible. Added to these shortcomings, the NDVI trend also showed spikes and dips due to presence of clouds. Most of the crops selected had sowing season starting from November. Each and every image from the image collection was examined for spatial coverage of the study area with less cloud cover. Images with high cloud cover were ignored. For the entire time period, there were 23 images in total covering the entire study area. The Figure 8 shows the distribution of Sentinel 2 images over the selected months. To cover the full study region, 6 images are needed. Therefore, dates with less than 6 images are not considered for the study.

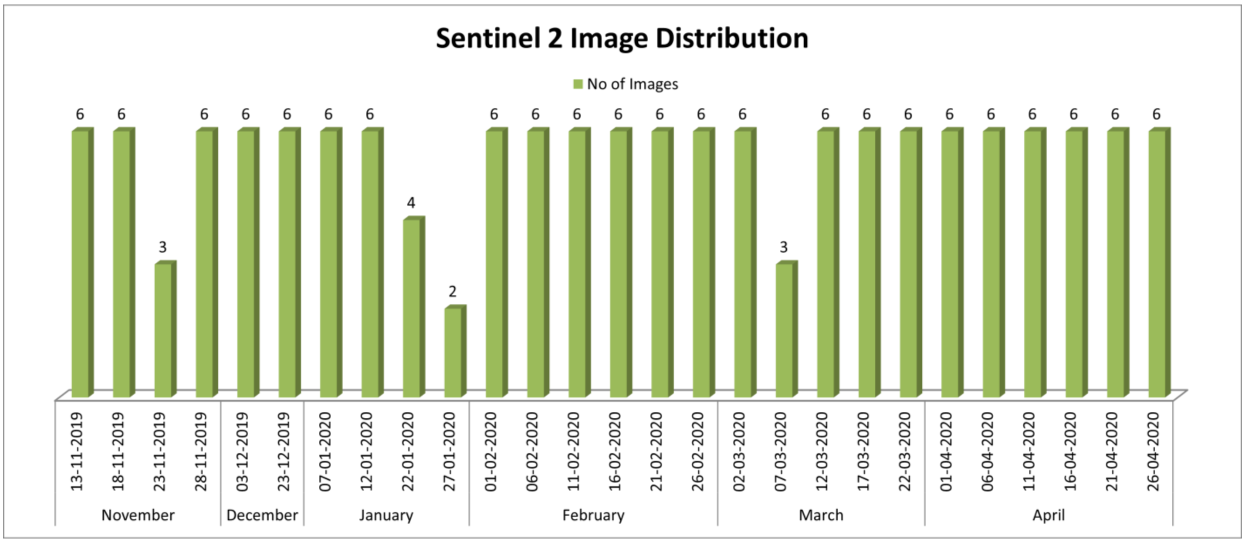

Figure 8 Distribution of Sentinel 2 Imageries over the study period (November 2019 - April 2020)

Most deep learning studies on crop classification are based on Red(R), Green(G), Blue(B), and Near Infrared(NIR) spectral bands (Ji et al., 2018; Z. Sun et al., 2020; Zhong et al., 2019). Accordingly, the Sentinel 2 bands B2, B3, B4, and B8 which denote R, G, B and NIR are selected as suitable bands and the other bands such as cloud and aerosol are dropped due to their coarse resolution and unsuitability to crop related studies. Normalised Difference Vegetation Index (NDVI) is calculated for all the 23 images from November to April using the formula

$$\boldsymbol{NDVI = \frac{NIR - Red}{NIR + Red}} \tag{1}$$

This newly computed NDVI is also stacked with the other bands. Finally, each image in the study period contains 5 bands namely R, G, B, NIR and NDVI with a spatial resolution of 10 meters.

4.2.2. Ground Truth Data Preparation and Cleaning

Initially, the ground truth dataset for the study area was in the form of vector points which included mixed cropping class also. Mixed cropping is dropped as it may create confusion within crop class identification and finally eight land cover classes which included three crop classes were selected. Class barren and fallows were merged into one class named Barren/Fallows as they depicted the same NDVI trend. The finalised classes are Wheat, Mustard, Chickpea, Forest, Scrubland, Barren/Fallows, Settlements and Water. More ground data points for classes such as water, forest, settlements are added by manually by visual interpretation. The final ground truth points overlaid in the study area is represented in the Figure 9.

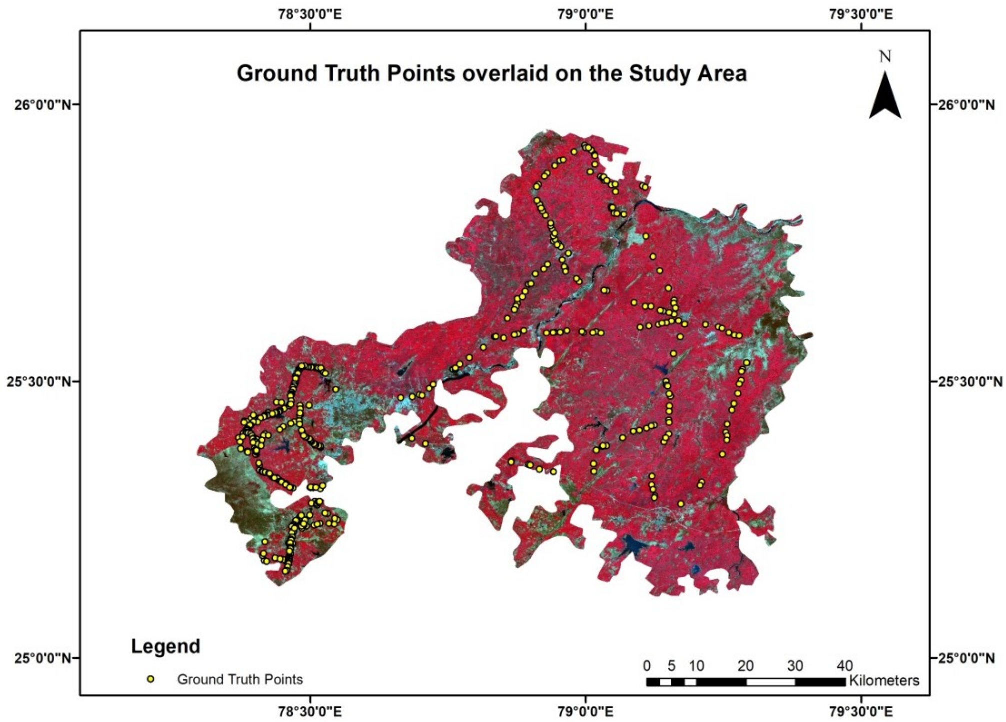

Figure 9 Distribution of ground truth points over the study area

Now, polygons are created by digitizing the ground features for every class. For this delineation, Google Earth imagery and the Sentinel 2 satellite imagery of the same time of the year are chosen as reference. The reason for polygon creation is to improve the number of pixels for each class. Finally, there are 310 numbers of polygons in total. The distribution of ground truth polygon samples for each class is given in the Table 2. From the ground truth sample count, it can be inferred that there is a high imbalance between different classes.

Table 2 Class wise sample count of polygon ground truths

Class Name	Sample Count
Wheat	93
Mustard	33
Chickpea	25
Forest	30
Scrubland	24
Barren/Fallows	26
Settlements	43
Water	36

In order to be fed into the deep learning model, the sample dataset has to be divided into training, validation, and testing sets. As the name suggests, training data is used for training of the deep learning model whereas the validation set aids in improving the model's accuracy by fine tuning the model after every epoch during the process of model training. The main reason for dividing the dataset into a validation set is to prevent the model from overfitting, which occurs when the model gets exceptionally good at identifying samples in the training set but struggles to generalise to new data and make accurate classifications. Finally, the test set which is a separate dataset, gives the accuracy the model after the completion of training process by testing the correctness of the predicted samples with the help of the metrics such as accuracy, precision, recall etc. The model should never be built using the test set. Out of the three, the training set forms the largest ratio because the model should have sufficient data to learn. These three datasets should be completely spatially independent of each other, and distributed over the study site.

Remote sensing data naturally contains spatial autocorrelation (Griffith & Chun, 2016; Wulder & Boots, 1998). The tendency for nearby locations or places to share similar values is called spatial autocorrelation. In the context of Remote Sensing, neighbouring pixels are more similar than faraway ones. If this is ignored during the split of training and testing sets, it can result in overestimation of generalisation capabilities of the model (Karasiak et al., 2022). This can be solved by using a data splitting strategy that guarantees spatial isolation in between training and test sets. One possible solution to this problem is by separating the training and testing datasets spatially during the data-splitting process. To solve the issue of spatial autocorrelation, one option is to spatially stratify the object samples, assuring spatial independence across the training and test sets (Cánovas-García et al., 2017; Inglada et al., 2017). Sampling based on objects is preferable to sampling based on pixels. Figure 10 provides an overview of the train test split. The optimum split ratio depends upon the type of application, the model used and the dimension of data. In this study, the sample data (delineated polygons) is divided into training, validation and test set in the ratio of 70:30 (train: test) taking the spatial distribution of the samples into account. The unique values of each data set are checked to confirm if the training, validation and testing sets include all the class types. Table 3 represents the number of polygons in the train, validation and test sets.

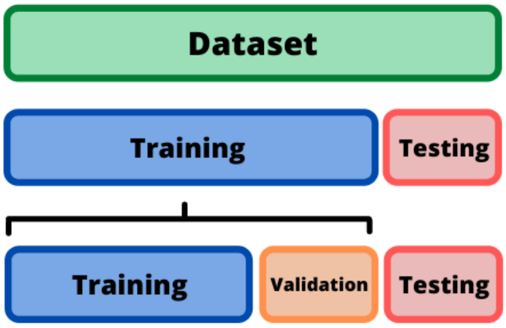

Figure 10 Train, Validation and Test Split (*How to Train and Test Data Like a Pro*, 2021)

Table 3 Per Class Distribution of Ground Truth Polygons in Training, validation and Testing Sets

Class	Training	Validation	Testing	Total
Wheat	50	9	34	93
Mustard	17	6	10	23
Chickpea	13	3	9	25
Forest	18	2	6	26
Scrubland	11	5	8	24
Barren/Fallows	17	5	4	26
Settlements	24	8	12	44
Water	10	2	3	15
Total	160	40	86	286

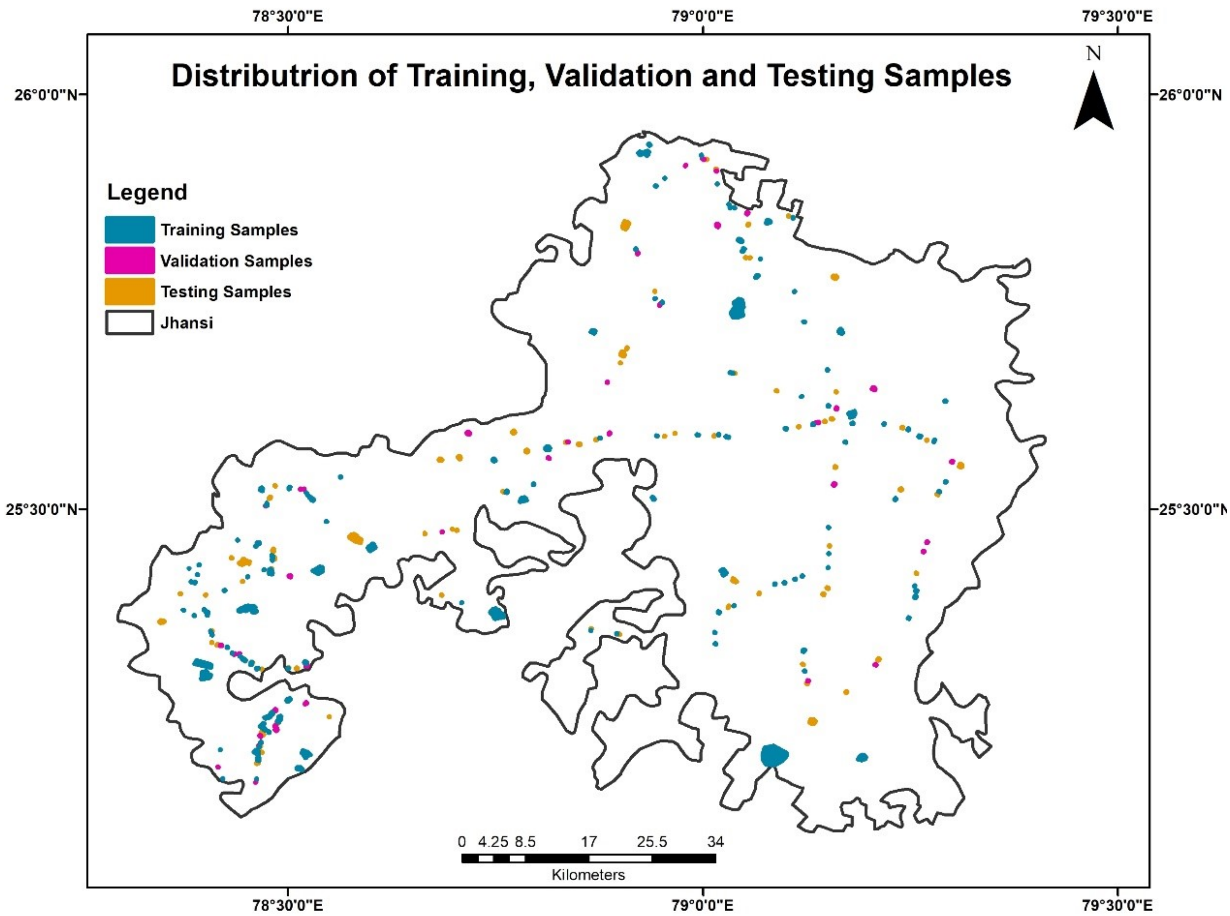

Figure 11 Spatial distributions of training, validation and testing set in the study area

The polygon samples are rasterised using the polygon to raster tool provided under the Conversion tool of Arc Tool Box. Land cover classes are given as the value field and cell size of the new raster takes the pixel size of the Sentinel 2 image. The spatial distribution of training, validation and testing labels is shown in the Figure 11. The pixel distribution of each class is given in the Table 4. Observing the distribution of this dataset one can infer that it is highly imbalanced and, the majority crop class is wheat with a pixel count of 4139 and Mustard and Chickpea are the minority crop classes with pixel count 1546 and 1824 respectively. Besides Barren/Fallows, a non-crop class is also a minority class.

Table 4 Number of pixels per class used in the training set

Class	Train Set
Wheat	4139
Mustard	1546
Chickpea	1824
Forest	4776
Scrubland	3196
Barren/Fallows	1374
Settlements	3678
Water	2695

4.2.3. NDVI trend for different crop types

In general, the quality of the satellite observations is often influenced by the presence of clouds and atmospheric aerosols. This results in the generation of images undesirable for environmental remote sensing studies. Therefore, the acquisition of good quality multi-temporal remote sensing data decides the generation of an NDVI time series of high quality which in turn is responsible for accurate crop type mapping(X. Liu et al., 2022; X. Zhang et al., 2022). In practice, these images with high cloud cover can be either discarded or reconstructed using the temporal filters for remote sensing classification(X. Zhang et al., 2022). Savitzky-Golay (SG) filter designed by (J. Chen et al., 2004) is based on the notion that NDVI time-series exhibit annual cycles of vegetation growth and decline. It is a simple but robust temporal filtering method which smoothens out noise in NDVI time-series. According to the study's findings, this technique is more efficient in producing NDVI time-series of excellent quality. SG filter is employed in several crop mapping studies (Amin et al., 2022; Htitiou et al., 2021; L. Pan et al., 2021).

The NDVI trend for each crop is created by randomly selecting 50 ground truth samples and then, SG filter is applied to remove the noise and generate a smooth temporal profiles. The three different crops are wheat, mustard and chickpea. Wheat has sowing season in the mid of November to first week of December and its harvesting period during the last week of March to first week of April. For Mustard, the sowing season is first week of November and the harvesting season is last week of February. For Chickpea, the sowing and harvesting period are from 1st week of November Last week of November and 1st week of March to Last week of March respectively. This crop calendar information is found in accordance to the computed NDVI trend. Quality check of the ground truth samples is performed with the help of the shape and maximum value of NDVI trends to figure out the wrong samples and remove them. The NDVI time series trends for the three major crop classes and all the nine land cover classes considered are shown in the Figure 12 and Figure 13 respectively. Table 4 shows the growing seasons of the three different crops considered in this study.

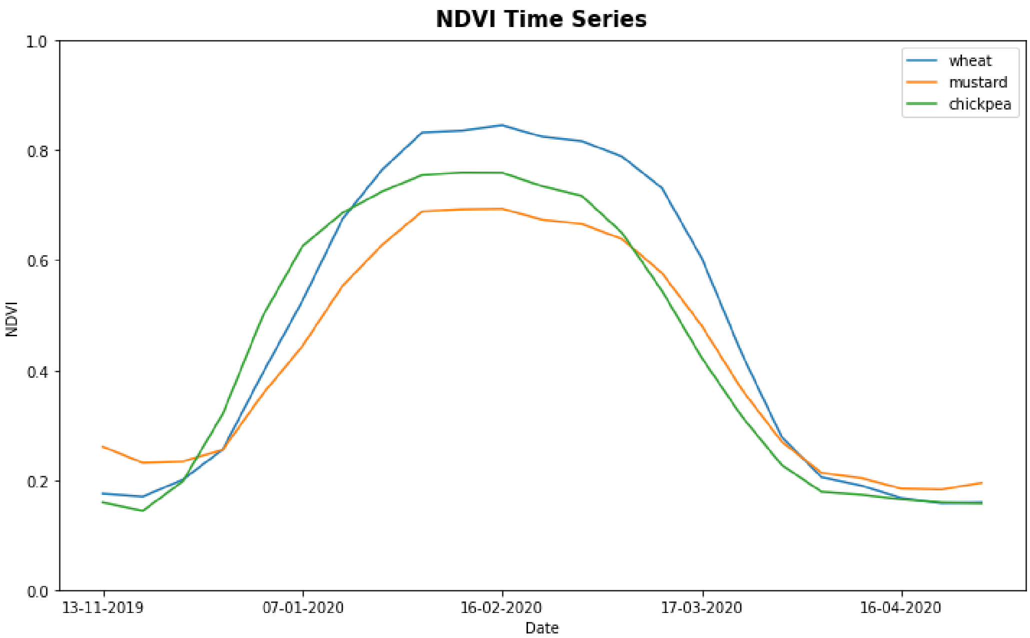

Figure 12 Normalised Difference vegetation Index Time Series Trends for the identified crop classes - Wheat, Mustard and Chickpea

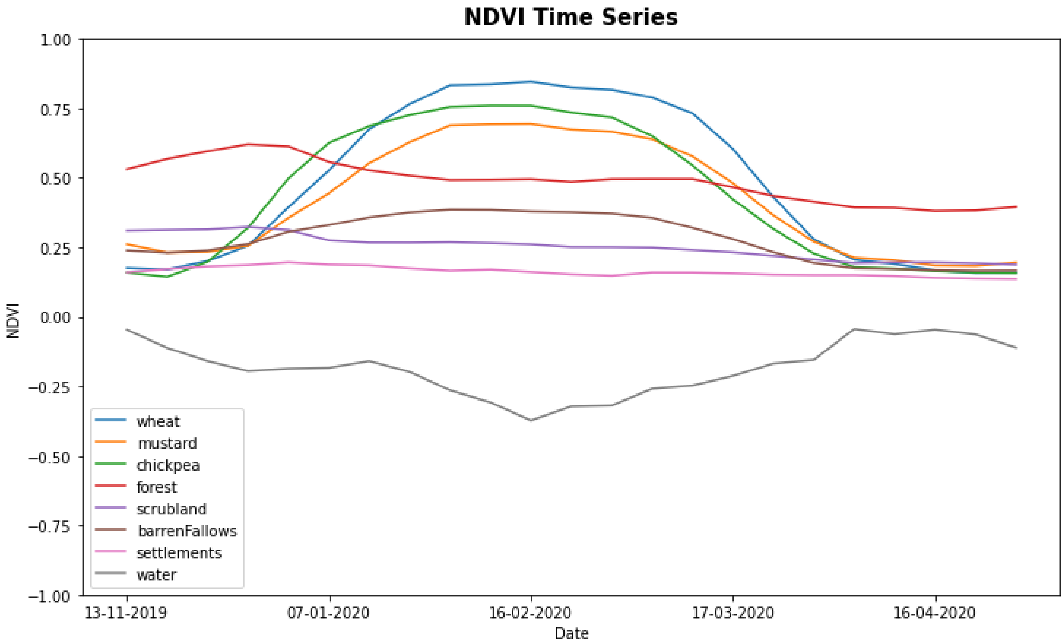

Figure 13 Normalised Difference vegetation Index Time Series Trends for the 8 classes identified in the study

Table 5 Sowing and Harvesting Period of Crops – Wheat, Mustard and Chickpea

Source: (*Crop Calendar PMFBY*, 2018)

Crop Class	**Sowing Period**		**Harvesting Period**	
	From	**To**	**From**	**To**
Wheat	2nd Fortnight of November	1st Fortnight of December	Last week of March	1st week of April
Mustard	1st week of October	1st week of November	Last week of February	Last week of March
Chickpea	1st week of November	Last week of November	1st week of March	Last week of March

The three major component of a Convolutional neural network are convolutional layer, pooling layer and fully connected layer. The design of the convolutional layer and pooling layer, the activation function, the loss function, and the regularisation are the main areas where convolutional neural network optimization is focused.

4.2.4. Convolutional Layers

The core building block and the first layer of a CNN is Convolutional layer. This layer utilises convolutions (or filters or kernels) to slide across the input image while looking for patterns in the image by computing the dot product between the kernel and the input. The depth of convolutional layer is defined by the number of filters used. The size of the filter depends upon the specified application and should always be less than the size of the input image. As a result of convolution operation, the image is abstracted to a feature map (or activation map). This feature map is fed into the next layer to learn more features from the input image. The input feature maps are first convolved with a learnt kernel in order to produce a new feature, and the outcomes are then fed into a nonlinear activation function (Guo et al., 2017). The mathematical context of convolution operation is shown in the Figure 14.

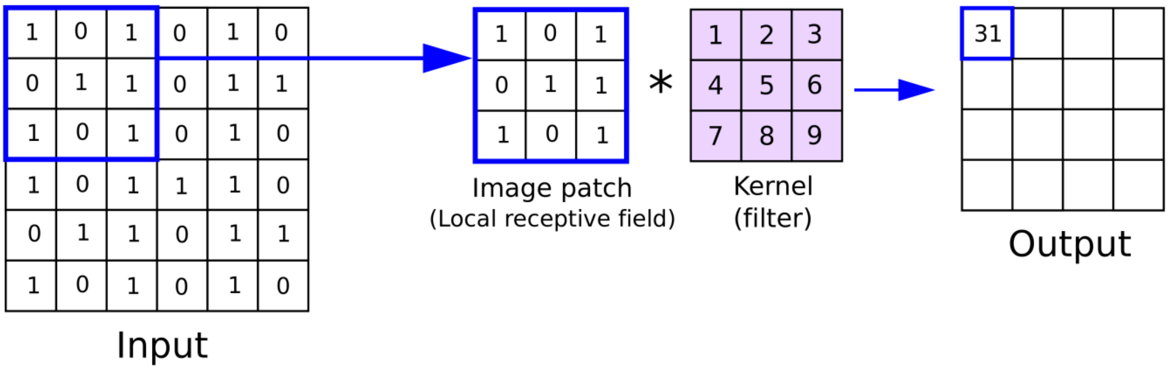

Figure 14 Representation of the Convolution operation in CNN(Reynolds, 2019)

4.2.5. Activation Function

By introducing non-linearity in the neural network, activation function improves the capability of neural networks to represent complex non-linear mapping (X. Pan et al., 2018). The typical activation functions are Sigmoid, Tanh and ReLU. Sigmoid activation is generally used for logistic regression problems. It ranges from 0 to 1. Tanh is similar to sigmoid but it ranges from -1 to 1. Out of these, Rectified Linear Unit (ReLU), proposed by Nair & Hinton, (2010) which performs a threshold operation on each input component, has demonstrated promising performance and results in various deep-learning research works (Y. Liu et al., 2017). ReLU activation function is expressed as $g(z) = \max(0, z)$. The most commonly used activation function is ReLU and it is used in this research. Widely used activation functions in deep learning are represented in the Figure 15. Softmax activation is used in the final layer in multiclass classification problems to map the outputs to the number of classes.

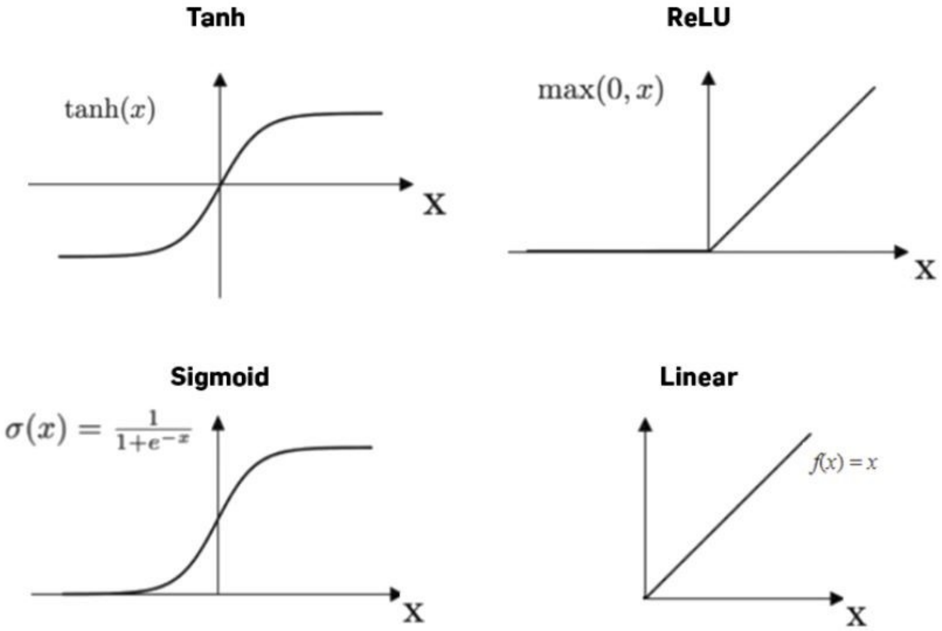

Figure 15 Different types of Activation Functions used in Deep Learning

4.2.6. Pooling Layers

As a result of multiple convolutions on input data, the dimensionality of data increases. To address this, pooling layer is used which employs down sampling strategy. The pooling layer is positioned amidst two convolution layers. Accordingly, it reduces the size of the feature map during the training phase, which minimizes the computational complexity while preserving discriminant information and tackles the overfitting issue (Mohammadimanesh et al., 2019; Yao et al., 2017). Well-known pooling layers include maximum, minimum, and average functions. In deep learning research, the max-pooling function has demonstrated stability and effectiveness and outperforms other pooling strategies (Mohammadimanesh et al., 2019; Rizaldy, 2018). It uses a rectangular window of size 2×2 moves on top of the feature map and outputs the highest value within the window. The working of max pooling operation is shown in the Figure 16.

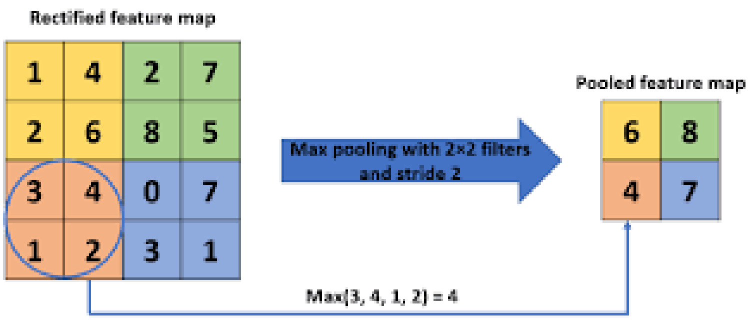

Figure 16 Representation of 2×2 Max Pooling with Stride of 2 (Gholamalinejad & Khosravi, 2020)

4.2.7. Dropout

Connecting all the features to the fully connected layer may result in overfitting of the training dataset. When a given model performs so well on training data that it has a negative effect on the model's performance when applied to new data, this is known as overfitting. In order to overcome this, dropout layer is utilized in the Convolutional Neural Network. "One of the most effective and popular regularization techniques is dropout" (Srivastava et al., 2014). When dropout is applied, a fraction of neurons are eliminated from the neural network during the training process. For example, a dropout of 0.2, results in 20% of the nodes in the neural network to be dropped out randomly. Dropout's primary objective is to prevent the model from picking up statistical noise from the data by purposefully adding stochasticity to the training process. (Srivastava et al., 2014).

4.2.8. Fully Connected Layers

The last layer is the fully connected layer or dense layer. This layer takes the flattened input from the previous layers and classifies the input data into required number of output classes. Softmax regression is frequently used for classification because it produces an effective probability distribution of outputs (Guo et al., 2017) .

4.3. 1D CNN Deep Learning Classfier

A multi-temporal stack of the study area is prepared by stacking all 23 images, each with 5 bands. Now, the resulting image consists of 115 bands (23×5) in total. The training labels raster is overlaid on the satellite image and pixel values that fall with the samples are extracted and stored in the form of CSV (Comma Separated Values) file. This process is repeated for validation and test sets also. Any NAN values in the file are dropped. Finally, the array is of shape $23{,}223 \times 115$, which forms the input of the CNN model. The layers of 1D CNN is shown in the Figure 17.

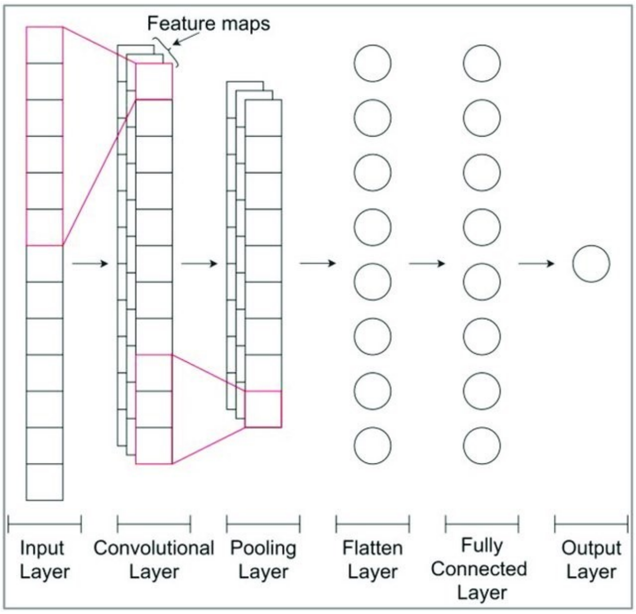

Figure 17 A visual representation of 1D CNN made up of input layer, a convolutional layer, max pooling layer, flatten, fully connected layer, and final output layer (Nisa & Kuan, 2021)

A 1D CNN is designed and executed for crop type classification with imbalanced classes. The CNN model is constructed using Keras. Keras is a high-level, compact uncomplicated Python library that is built on top of the TensorFlow framework in Python. Creation of neural network layers, shape and mathematical information can be easily defined using Keras library.

The main justifications for choosing 1D CNN are its shallow architecture with excellent learning capabilities, lower computational complexity than 2D CNN, and superior performance demonstration on scenarios with limited labelled data (Kiranyaz et al., 2021). Pixel wise classification of multispectral satellite imagery utilizes 1D operation. Therefore, the deep learning CNN model which uses this operation is addressed as 1D CNN. The architecture used in the model is 'Sequential'. The layers in the 1D CNN are arranged as follows: input layer, convolutional layer, pooling layer, flatten, dense layer and the final output layer. Hierarchical feature extraction is performed by the convolutional layer, the pooling layer, while the fully connected layer which follows these layers perform the role of a classifier by generating the predicted probabilities for each class sample in the input data. In reality, CNN is the dot product of the local regions of the input and trainable weights, allowing for effective matrix multiplication (Sidike et al., 2019). The architecture of the 1D CNN model adopted in this research is shown in the Figure 18.

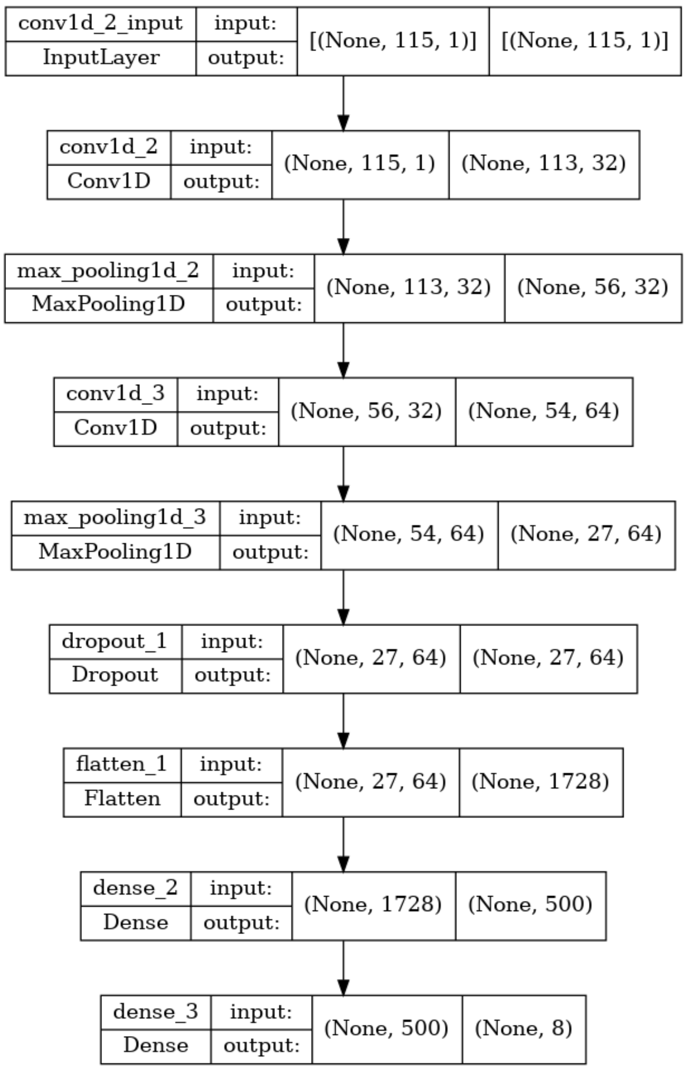

Figure 18 Architecture of the proposed 1D CNN Model

The designed model has input layer of shape (115, 1) which is equal to the number of input bands. Two convolutional layers with 32 and 64 filters are used. The filter size is chosen as 3×3. The Activation function used is ReLU. 2×2 Max Pooling with a stride of 2 is applied after every convolution layer. A flatten layer is placed following this. Next, a fully connected layer is placed. Finally, the last layer is the output layer that contains neurons equal to the number of classes which is 8 in this case. One hot encoding is done on the labels as it improves the classification accuracy. The optimizer used is Adam which is an enhanced version of stochastic gradient descent. Once the CNN architecture has been established, forward- and backward-propagation training is used to iteratively update the trainable parameters to minimise the difference between the predicted output and the actual output. The number of neurons, learning rate, dropout rate and number of epochs are chosen by performing hyper-parameter optimization.

4.4. Algorithm Level Balancing

Categorical Cross Entropy Loss Function In modern deep neural network algorithms, cross entropy loss also known as logarithmic loss is universally used for training classification related tasks(Hui & Belkin, 2021). The categorical cross entropy loss function is preferred for multiclass classification problems as it yields high accuracy (Andreieva & Shvai, 2021). With classification problems having multiple class and the truth labels one hot encoded, categorical cross entropy loss is used. Theoretically, a perfect model results in a cross entropy loss of 0. As the predicted probability moves away from the actual probability, cross-entropy loss increases. Cross entropy loss severely penalises predictions that are confident but actually wrong. The Cross Entropy loss is given by the formula

$$\text{Cross Entropy Loss} = -\sum_{i=1}^{n} t_i \log(p_i) \tag{2}$$

Here, t_i is the actual class values whereas p_i denotes the softmax probability of class i

Since each data instance is given equal weight by the traditional cross-entropy based loss function, the network oversees classes with fewer observations. In classification or segmentation tasks when there is a class imbalance, cross entropy loss is therefore inappropriate (Fernando & Tsokos, 2022).

Focal loss is derived from cross entropy loss. By applying a adjusting term on cross entropy loss, focal loss weighs down easy examples assuming that the contribution of these samples to the total loss is less even if they are high in number and focuses on the training of hard examples or the minority classes (T. Lin et al., 2020).

$$\text{Focal Loss, } FL(p_i) = -(1-p_i)^{\gamma} \log(p_i) \tag{3}$$

Where, p_i is the estimated probability of the model using ground truth and γ is a focussing parameter.

γ is always greater than or equal to 0. When $\gamma = 0$, the focal loss is equivalent to cross entropy loss. The contribution of difficult samples to the overall loss can be increased by increasing the focusing parameter γ. The focussing parameter $\gamma = 2$ is found to work best in the study by T. Lin et al., (2020).
In practice, α-balanced variant of the focal loss is used as it generates slightly higher accuracy compared to the non-α-balanced form. α is the term introduced to handle the imbalanced classes.

$$FL(p_i) = -\alpha(1 - p_i)^{\gamma}\log(p_i) \qquad \textbf{(4)}$$

Both γ and α, are hyper-parameters. It should be noted that it is necessary to select values for these two hyper-parameters together. In practice, as γ increases, α should be slightly lowered ($\gamma = 2$, $\alpha = 0.25$ works best) (T. Lin et al., 2020). Fig 19 shows the focal loss curve.

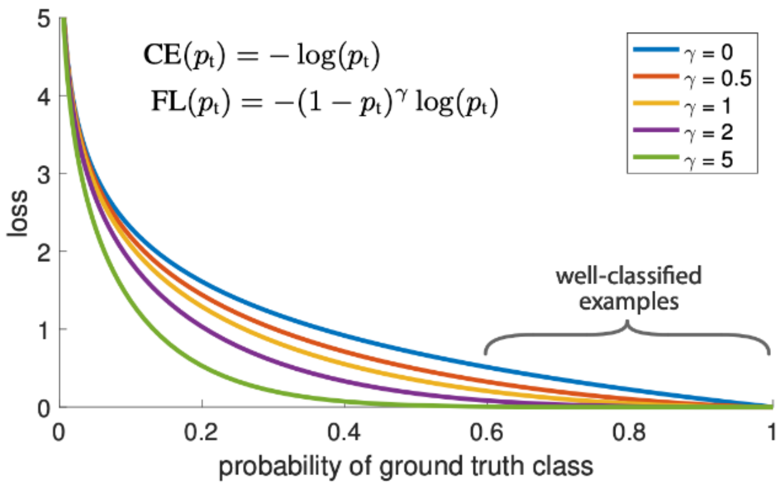

Figure 19 Influence of focussing parameter γ on accuracy and loss(T. Lin et al., 2020)

Another approach to handle imbalance is by using, class weighted loss function. In this method, the imbalance is addresses by providing different weights to the majority and minority classes in the training phase of the deep learning model. The basic concept is to penalize the minority class for misclassifying itself by increasing class weight while simultaneously decreasing class weight for the majority class. In an imbalanced classification problem, the gradients are multiplied by a weight that is distinct to each class. The class weights are computed using the formula,

$$\text{Weight of class, } w_j = \frac{n\ samples}{n\ classes \times n\ samples_j} \qquad \textbf{(5)}$$

Where, w_j denotes the weight for individual class (j represents class),
n samples is the total number of training samples in the entire dataset,
n classes is the number of unique classes and,
n samples $_j$ is the number of samples in the respective class j

The weights are given between 0 and 1. Suppose in a binary classification problem, if we give the minority class a weight of n, the majority class will receive a weight of 1-n. The weight of the minority class will always be higher than the weights of majority class.

Class Weighted Loss, adopted by Schiele et al., (2019) the cross-entropy loss is weighed with a factor equal to the ratio between the number of samples in the majority class and the total number of classes considered and it is given by the equation,

$$\text{Class Weighted Loss, } L_{\alpha} = \frac{n_{max}}{n_i} \log(p_i) \tag{6}$$

Where, n_{max} is the total number of majority class samples, n_i is the total number of classes present and p_i is the estimated probability. By incorporating class weighted loss in model training, the minority classes also show an equal impact on the loss function.

4.5. Data Level Balancing Techniques

In order to perform, oversampling and under sampling techniques, the training raster label with imbalanced classes is converted to points using the "raster to point" conversion tool of ArcMap. Then, satellite imagery of 1st February 2020 is selected. This particular imagery is chosen for extracting spectral values considering the less cloud cover with clear atmospheric conditions and peak growth stage of crops. Using the "extract multi values to points" tool, the band values are extracted for all the data points from this satellite imagery. Now the different bands namely R, G, B, NIR, and NDVI along with latitude and longitude are considered as the predictor/ explanatory/ independent variable. The variable which is measured or tested in an experiment is known as the dependent/ response variable. The class label is taken as the response variable. These 23,223 training points form the data for performing class balancing. It should be taken care to perform data balancing techniques on training dataset alone. The testing dataset should be left untouched. The sample count for initial imbalanced training class is depicted in the bar chart in Figure 20.

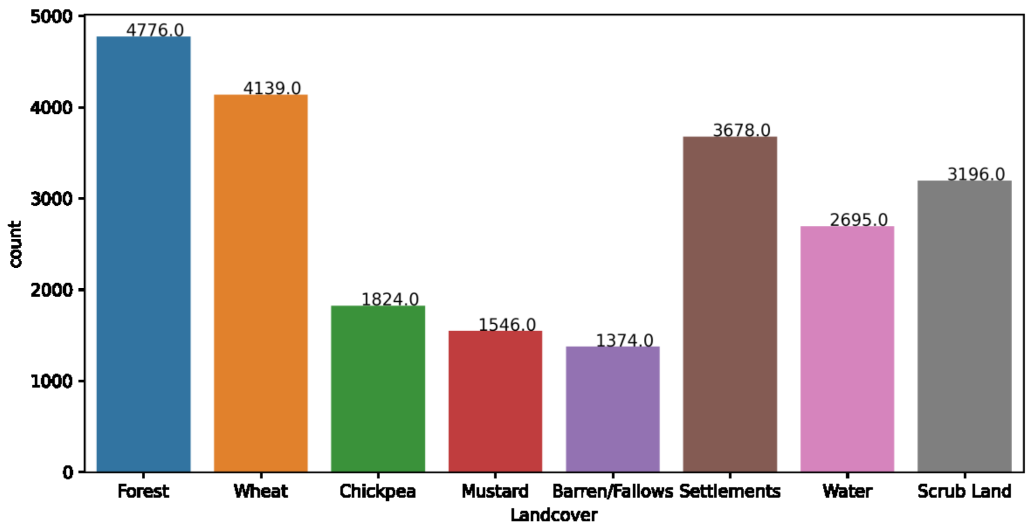

Figure 20 Distribution of sample count in imbalanced training dataset

Imbalanced-learn (*imblearn*) is an open source python library which is equipped with tools for handling class imbalance problem in classification. This is a part of *sklearn* library. The main aim of data level balancing techniques is to equalize the number of samples per class either by decreasing majority class instances or increasing minority class instances. In oversampling, methods such as Random Oversampling, Smote and Safe Level Smote are employed. Under sampling methods that are adopted in this study are, Random Under sampling, Tomek Links and Edited Nearest Neighbour (ENN). Hybrid approaches/ ensemble methods (combination of oversampling and under sampling) such as Smote + Tomek, Smote + ENN, Safe Level Smote + Tomek are applied.

The working of SMOTE algorithm is shown in the Figure 21. Assume that class A is the minority class and class B is the majority class. The k-nearest neighbours for each x observation in class A are then located, and new synthetic instances are created along the line connecting the minority instance x to its k-nearest neighbours (Elhassan & Al-Mohanna, 2017).

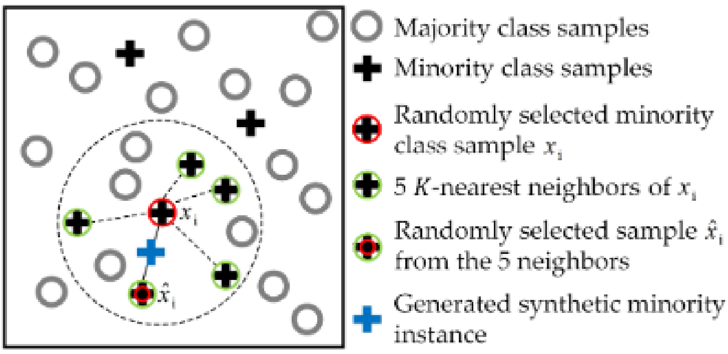

Figure 21 Basic principle of Synthetic Minority oversampling Technique(SMOTE)(Ma et al., 2019)

New synthetic instances in SMOTE are created using the formula presented in (Elhassan & Al-Mohanna, 2017),

$$S_{syn} = r\,(S_{kNN} - S_f) + S_f \tag{7}$$

where, $S_{syn\ denote}$ the synthetic samples generated,

S_f is the sample data selected in the minority class,

S_{kNN} is the k-nearest neighbour that is considered from 1 of 3 and

r is the random constant between 0 and 1

By using the nearest neighbour criterion, **Tomek link** eliminates instances of the majority class that are closer to the minority class (Sawangarreerak & Thanathamathee, 2020). For easy understanding of Tomek(T) link working, if x_a and x_b are the majority and minority classes respectively, then the distance between them $d(x_a, x_b)$ is called as tomek link with the condition that that no other class x_z such that $d(x_a, x_z) < d(x_a, x_b)$ or $d(x_b, x_z) < d(x_a, x_b)$ exits. "If any two instances are T-Link, then either one of them is a noise example or both examples are situated on the class boundaries" (Elhassan & Al-Mohanna, 2017).

In **Edited Nearest Neighbour**, for an instance x, its k nearest neighbours (generally, k=3) are considered, and the sample instance with class labels different from majority of the k nearest neighbours is removed.

4.6. Accuracy Assessment Metrics

In the final stage, the accuracy of the crop type map is assessed with the help of the test set adopting different accuracy measures. Typically, in classification analysis, a confusion matrix is used to assess the classifier and the resulting map. A typical confusion matrix is represented in the Figure 22. Before talking about accuracy metrics, one needs to understand the concepts of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). An outcome where the model properly predicted the positive class is referred to as TP. Similarly, a TN is an outcome for which the model correctly predicts the negative class. FP is when the model predicts the positive class incorrectly whereas FN is an outcome where the model predicts the negative class wrongly. For simplification, a binary classification with classes positive and negative will have a confusion matrix denoted as

		Predicted Labels	
		Positive	Negative
True Labels	Positive	True Positive(TP)	False Negative(FN)
	Negative	False Positive(FP)	True Negative(TN)

Figure 22 Confusion Matrix depicting the terms True Positive, True Negative, False Positive and False Negative

Overall Accuracy (OA) is the most commonly used accuracy measure that gives the percentage of correctly classified pixels in the output map.

$$\text{OA} = \frac{\text{No of correctly classified samples}}{Total\ Number\ of\ Samples} \times 100 \qquad \textbf{(8)}$$

$$\text{OA}= \frac{TN+TP}{TN+FP+TP+FN} \qquad \textbf{(9)}$$

Precision is defined as the ratio of the positives that are predicted correctly by the model over total positive samples.

$$\text{Precision}= \frac{TP}{TP+FP} \qquad \textbf{(10)}$$

Recall also known as Sensitivity is the ability of model for correct classification of true positives whereas **Specificity** is the ability of the model to correctly classify true negatives.

$$\text{Recall}= \frac{TP}{TP+FN} \tag{11}$$

$$\text{Specificity}= \frac{TN}{TN+FP} \tag{12}$$

F1-score is given by the weighted average of precision and recall. The best score of F1 is 1 and worst is 0. F1 score is a better metric for uneven class distribution.

$$\text{F1 score} = 2 \times \frac{Precision\times Recall}{Precision+ Recall} \tag{13}$$

G- Mean Accuracy

OA gives more attention on the majority classes providing less importance to the minority classes. To overcome this, it will be advantageous to use G-Mean Accuracy, which offers a more accurate assessment in terms of imbalanced learning. Y. Sun et al., (2006). The balance between classification performances on both the majority and minority classes is measured by the Geometric Mean (G-Mean) accuracy metric(Akosa, 2017). This accuracy measure equally assesses the classification performance of each individual class.For multiple class scenarios, G-mean Accuracy is defined as,

$$\text{G-Mean Accuracy} = \left(\prod_{i=1}^{j} PA_i\right)^{\frac{1}{j}} \tag{14}$$

where, j denotes the number of classes and PA i is the Producer Accuracy of class i. PA measures the likelihood of a reference pixel being classified correctly for a specific class. It can also be denoted as

$$G - Mean \ = \sqrt{Sensitivity \ \times Specificity} \tag{15}$$

The accuracy of crop type maps is validated using the testing samples with the help of OA and G-Mean Accuracy. The Accuracy from the developed deep learning model is compared against the accuracy yielded by adoption of data balancing methods to arrive at the best method for crop type classification. Additional accuracy metrics such as precision, recall, and F1-score will also be computed for individual crop type classes.

4.7. Hyperparameter Optimization

Generally trial and error method is practiced for the setting model's hyper-parameters. In this research, Bayesian Optimization is the method chosen for hyper-parameter optimization. Since it is known that the learning rate shouldn't be either too small or too large, the optimal learning rate is found by giving different learning rates from 0.01 to 1. Dropout rates considered are from 0.1 to 0.5. The number of epochs considered for model training is given from 5 to 120. And finally, batch sizes between 50 and 1000 are considered. The model is run several times by applying Bayesian optimization choosing different combinations of hyper-parameter values mentioned above and the best parameters are the parameters of the model which yields the highest accuracy and low loss values. The loss and accuracy curves of training and validation dataset also play an important role in deciding the best hyper-parameters. Rectified linear unit (ReLu) is chosen as the Activation function because it is the most popular and efficient activation functions that provides faster CNN training and is employed in several crop classification studies (Kussul et al., 2018; Mazzia et al., 2020; S. Wang et al., 2020; B. Xie et al., 2019). Softmax activation is used in the final fully connected layer. The hyper parameter optimization is done for the CNN model with focal loss and class weighted loss as they have shown better accuracy than cross entropy loss (discussed in Section 5.2.2 and 5.2.3). The optimal hyper-parameters obtained are shown in the Table 6.

Table 6 Best Hyper-parameters for 1D CNN model

Hyper-parameters	**1D CNN with Focal Loss**	**1D CNN with class weighted loss**
Learning Rate	0.01	0.01
Dropout Rate	0.1	0.1
No of Epochs	33	97
Batch Size	741	952

In all there cases, the hyper-parameter tuning gave a number of around 500 neurons (533 for Cross Entropy loss, 460 for Focal Loss and 522 for Class Weighted Loss) in the dense layer to be the best. Therefore 500 neurons were fixed in the dense layer present before the final output layer.

5. RESULTS AND DISCUSSIONS

5.1. Comparison of results obtained by employing different loss functions

5.1.1. Categorical Cross Entropy Loss Function

First, the 1D CNN model is trained using categorical cross entropy loss function. The confusion matrix obtained is shown in the Figure 23. The Overall Accuracy obtained is 63.54 %. The classes, forest, water, settlements being the majority classes are classified with better accuracy compared to the other classes. There seems a great confusion between the three crop classes. A good number of samples are classified correctly as wheat class opposed to the poorly classified minority crop classes such as mustard and chickpea. Barren/fallows, one of the minority classes also seems to have been largely confused with scrubland. A possible solution to this misclassification problem would be to increase the number of training samples. As expected, categorical cross entropy did not account for the imbalance in the dataset and it is not suitable for classification problems with highly imbalanced data. But, without increasing the training samples can the classification of individual classes be improved? A possible approach is to apply different loss functions.

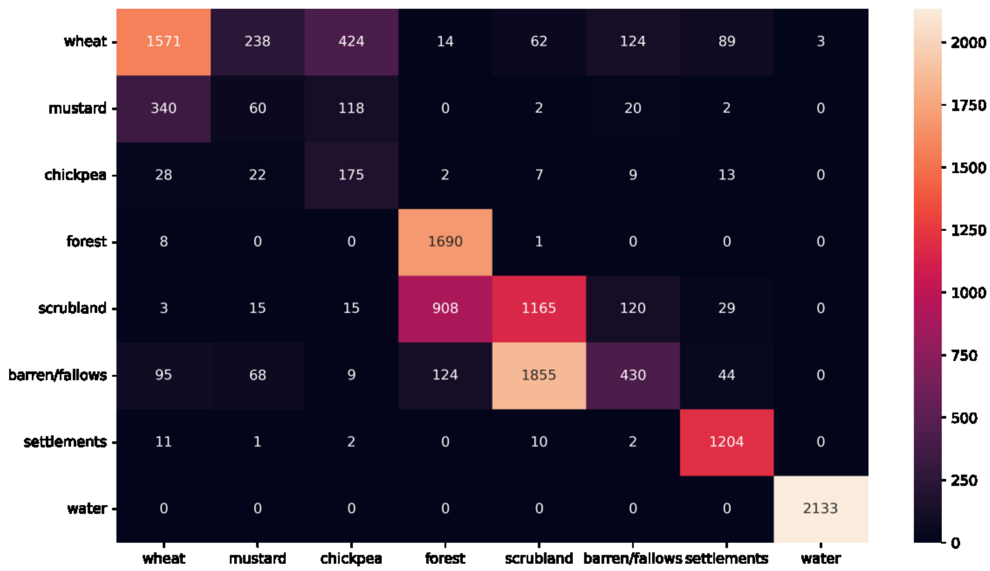

Figure 23 Confusion Matrix obtained through cross entropy loss function.

5.1.2. Focal Loss Function

Now, the cross entropy loss function is replaced by focal loss function. A gamma value of 2 and alpha value of 0.25 is set and the 1D CNN model is compiled with focal loss function. An improved overall accuracy of 66.57% is obtained the confusion matrix for which is shown in the Figure24. The classification of the minority class barren/fallows has improved while there is no significant improvement in the classification of other minority crop classes. There is large amount of confusion of the minority crop classes with the majority crop class, wheat. As Focal loss function increases the contribution of minority class to updating of weights, the contribution of the minority class with least number of samples Barren/Fallows has increased which in turn is the reason for better classification of this class. The other two minority crops mustard and chickpea are not classified as good as the Barren/Fallows class even after the application of focal loss. Almost similar temporal trend might be a possible reason for the poor classific

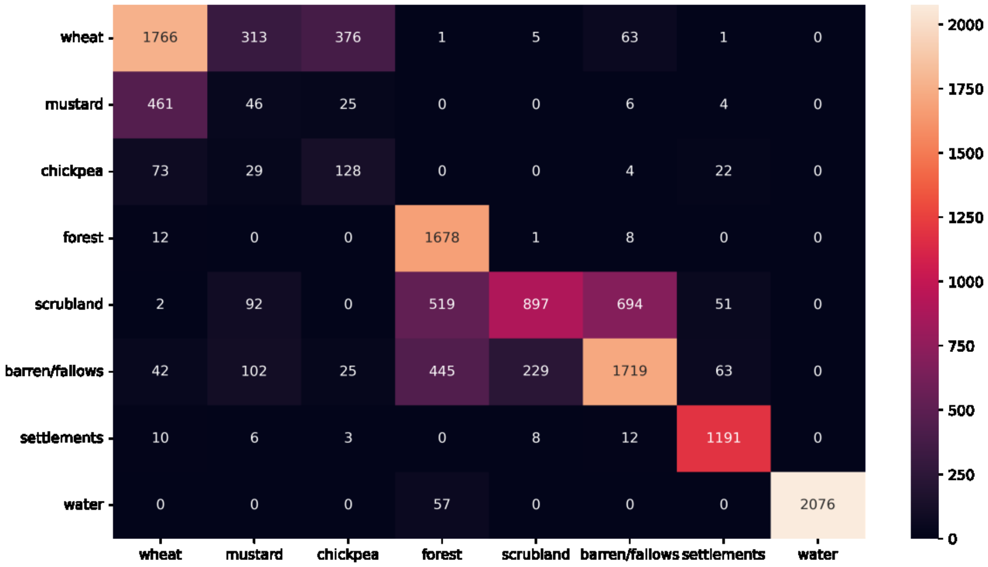

Figure 24 Confusion Matrix obtained through Focal Loss function.

5.1.3. Class Weighted Loss Function

Larger weight values are assigned to minority crop classes. Mustard, Chickpea and Barren/Fallows being the minority classes, have weight values of 1.87, 1.59, and 2.11 respectively. Wheat, the majority crop class, has a weight value of 0.70. The class weights obtained by computation are shown in the Table 7. In cross entropy, each class is weighed by the respective weights computed and the CNN model is trained. Class weighted loss improved the overall accuracy of the CNN model even more, providing a testing accuracy of 70.37 which is higher than the focal loss. In terms of Overall Accuracy, Class Weighted loss function is picked up as the best out of all three loss functions used in this research. The confusion matrix of the 1D CNN model trained by providing class weight values is shown in the Figure 25. Observing the confusion matrix, one can infer that the crop classes Mustard and Chickpea show improved classification compared with the other two CNN models discussed earlier.

Table 7 Class weights for each individual class.

Class Name	Weights
Wheat	0.70
Mustard	1.87
Chickpea	1.59
Forest	0.60
Scrubland	0.90
Barren/Fallows	2.11
Settlements	0.78
Water	1.07

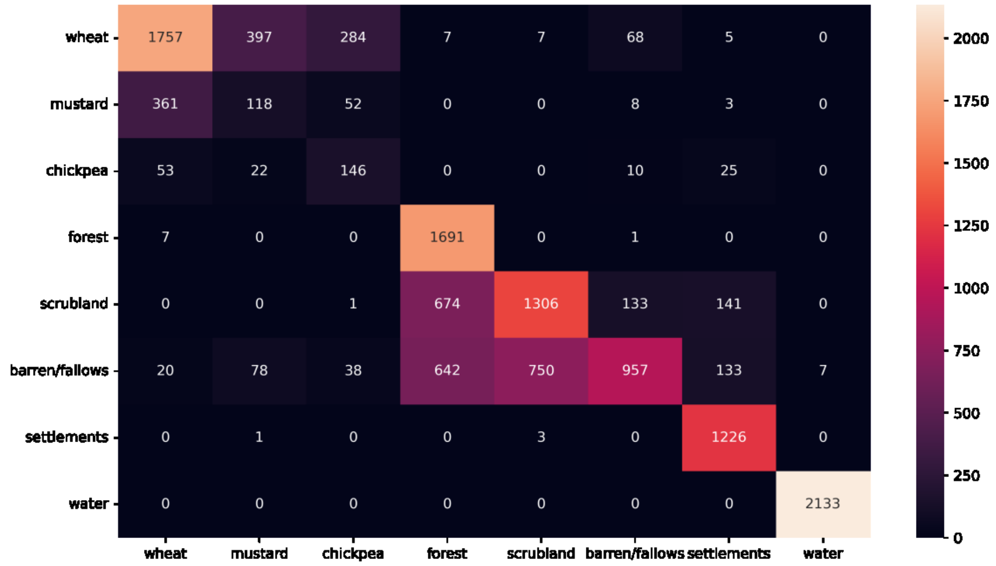

Figure 25 Confusion Matrix obtained through Class Weighted Loss function.

5.1.4. Analysis of Best performing Loss Function in terms of solving class imbalance: Comparison of different Accuracy Metrics

As we know that, overall accuracy does not account for imbalanced classes and doesn't provide accuracy of individual classes, the classification results are evaluated and compared in the Table 8 using other accuracy metrics such as Precision, Recall and F1 score. Generally, values of precision and recall more than 0.50 are regarded as satisfactory. Low precision values of mustard and chickpea indicates the presence of large number of false positives. Similarly, low value of recall was observed for the minority crop classes signifying the presence of more false negatives which means that the pixels were not being classified into their relevant classes.

Table 8 Comparison of Precision, Recall and F1 score accuracy metric for individual classes using three different loss functions

		Precision	
Class	**Categorical Cross Entropy**	**Focal Loss**	**Class Weighted Loss**
Wheat	0.76	0.81	0.80
Mustard	0.15	0.11	0.19
Chickpea	0.24	0.29	0.28
Forest	0.62	0.74	0.56
Scrubland	0.38	0.42	0.63
Barren/Fallows	0.61	0.55	0.81
Settlements	0.87	0.92	0.80
Water	1.00	1.00	1.00

		Recall	
Wheat	0.62	0.63	0.70
Mustard	0.11	0.13	0.22
Chickpea	0.68	0.54	0.57
Forest	0.99	1.00	1.00
Scrubland	0.52	0.60	0.58
Barren/Fallows	0.16	0.25	0.36
Settlements	0.98	0.97	1.00
Water	1.00	1.00	1.00
		F1 Score	
Wheat	0.69	0.71	0.74
Mustard	0.13	0.12	0.20
Chickpea	0.35	0.38	0.38
Forest	0.76	0.85	0.72
Scrubland	0.43	0.49	0.60
Barren/Fallows	0.26	0.35	0.50
Settlements	0.92	0.95	0.89
Water	1.00	1.00	1.00

Consequently, a poor F1 score was observed for the minority crop classes also. The Class weighted loss generated a F1 score of 0.74, 0.20 and 0.38 for the crop classes' wheat, mustard and chickpea respectively, which is the best out of the three models. Comparison of Overall Accuracy and G-Mean Accuracy for the three different loss functions used is shown in the Figure 26.

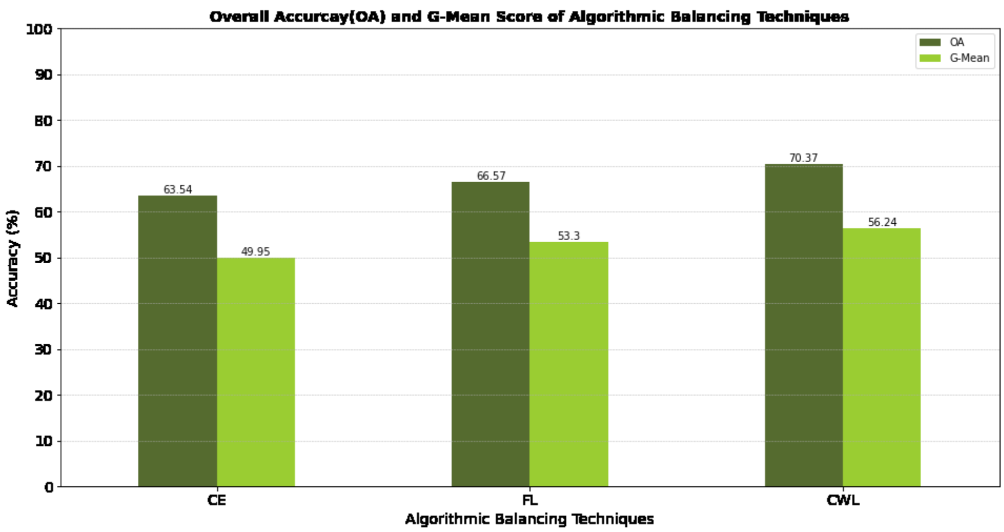

Figure 26 Overall Accuracy and G-Mean Score of different Algorithmic Balancing Techniques

The results of G Mean Accuracy, which provides a better evaluation metric for imbalanced classification problem is shown in the Table 9. By comparing the G- Mean Accuracy of the three loss functions, it can be seen that Class Weighted loss performs the best. In other two cases, there is a bias in accuracy with high G mean values for the majority class and lower value of G Mean for minority classes. But, in class weighted loss, the accuracy of individual class is not biased towards majority classes. The three crop classes have a G Mean Accuracy of 0.81, 0.45 and 0.74 respectively which falls in an acceptable range. The minority class Mustard has a G Mean of 0.45 which is low compared to all classes. Looking at the overall G-Mean Accuracy provided by all three loss functions, class weighted loss has an accuracy of 56.24 % which is the highest of all whereas cross entropy has a G-Mean which is lesser than 50 %.

Table 9 Comparison of Geometric Mean accuracy metric for individual classes using three different loss functions

Class	Categorical Cross Entropy	Focal Loss	Class Weighted Loss
Wheat	0.77	0.77	0.81
Mustard	0.32	0.35	0.45
Chickpea	0.80	0.72	0.74
Forest	0.95	0.97	0.93
Scrubland	0.65	0.70	0.73
Barren/Fallows	0.39	0.49	0.59
Settlements	0.98	0.97	0.98
Water	0.99	1	0.99
Overall G-Mean Accuracy	0.4995	0.5330	**0.5624**

The model accuracy and loss plots for the training and validation datasets while using three different loss functions are shown in the Figure 27. The focal loss function took the least number of epochs (No of epochs: 33, provided by Bayesian optimization) to reach high training accuracy. Training with more epochs lead to drop in the training accuracy and subsequently the testing accuracy. Therefore, only fewer epochs were used. In cross entropy loss function, the training accuracy steadily increases and after around 10 epochs, the value of accuracy doesn't fluctuate much. High fluctuations in the validation loss and accuracy denote that the validation set doesn't represent the whole dataset properly. The lower number of validation samples provided is the reason for this. All the three models have good training accuracy of greater than 0.95. When it comes to validation accuracy, class weighted loss is the highest with validation accuracy reaching the highest value of 0.80.

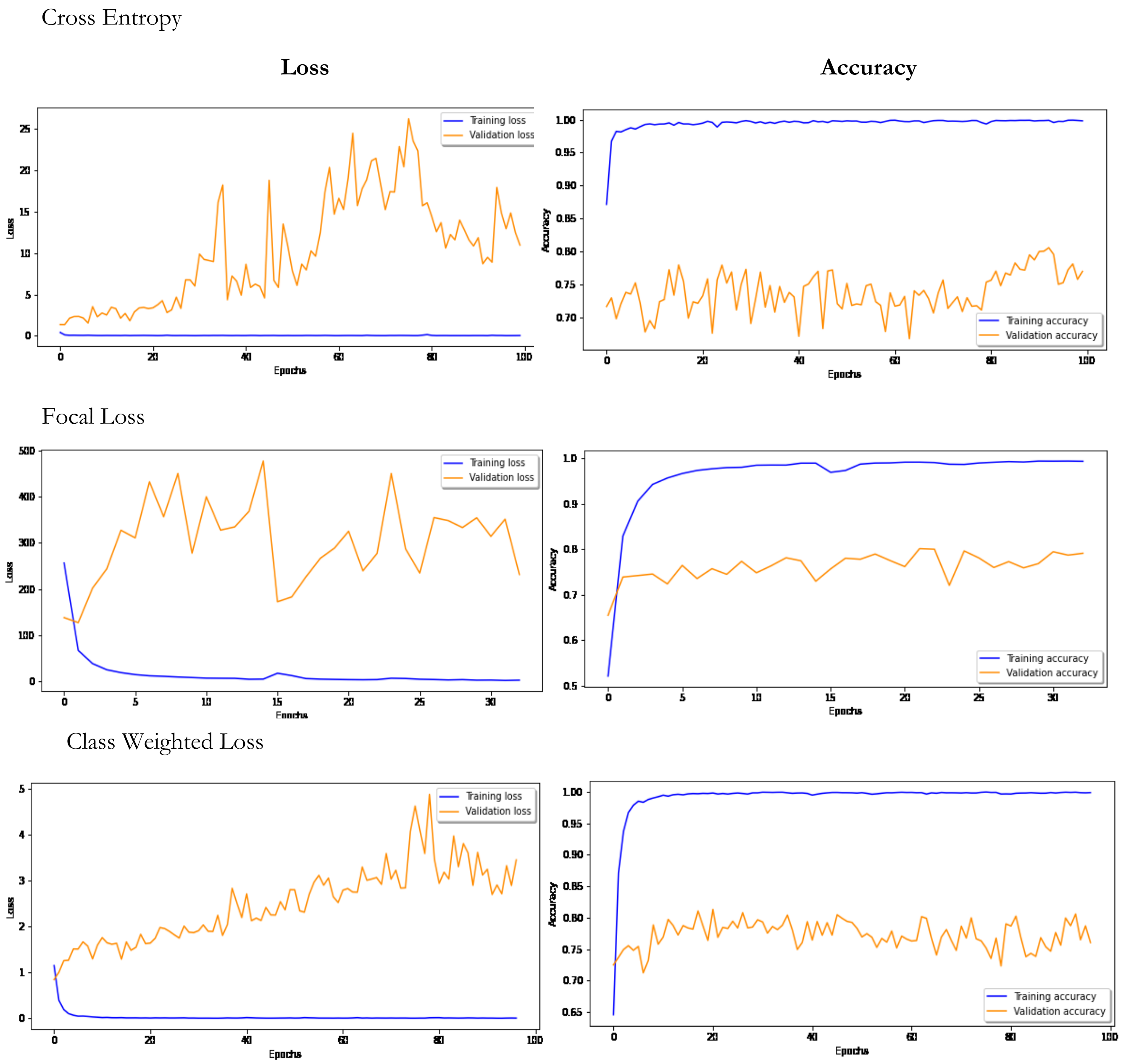

Figure 27 Training and validation curves of accuracy and loss for Cross Entropy loss, Focal loss and Class Weighted loss

5.2. Visualization of Classification results/ Qualitative Analysis

A smaller patch of size 3423, 2100 is selected from the study area. It was checked if all classes were present in this region and prediction was done. Owing to the large size (12453, 9405, 115) of the multi-temporal image, which was almost 50 GB in memory, only a subset of the study area was selected for prediction. The predicted map with different classes is shown in the Figure 28

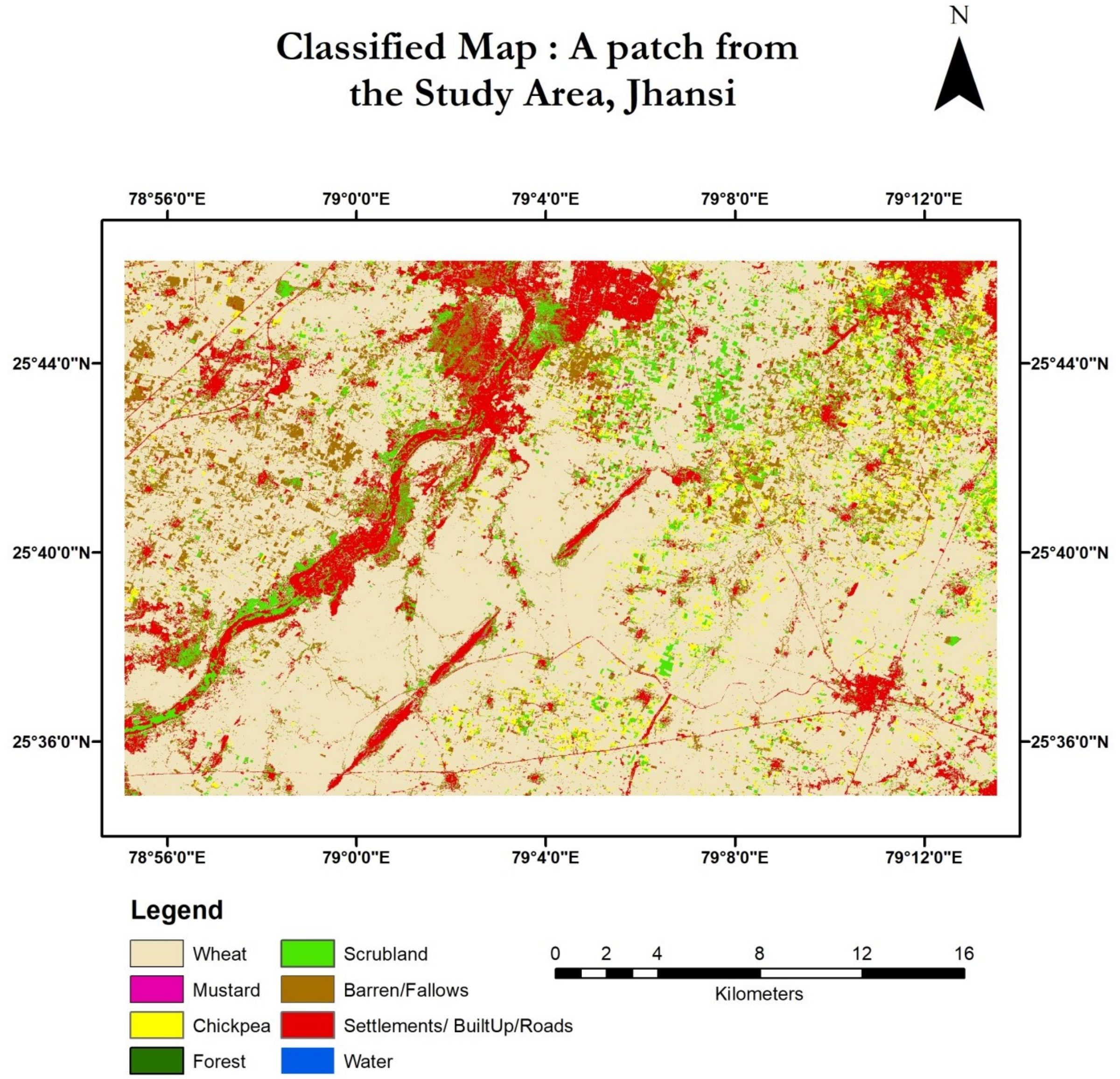

Figure 28 Classification Map obtained through 1D CNN for a patch in the Study Area- Jhansi

5.3. Results of Class balancing

The overall accuracies and G-Mean accuracies for all nine data level class balancing techniques tested in this study is shown in the Figure 29. The under sampling techniques attempted are Random under sampling (RUS), Tomek Links and ENN. Oversampling techniques attempted are Random Oversampling (ROS), SMOTE and Safe Level SMOTE (SL SMOTE). And, finally, SMOTE + RUS, SMOTE+ Tomek Links and Smote+ ENN are the hybrid sampling techniques attempted. Each of these techniques tries to balance out the number of samples in each class and attempts to reduce the imbalance between the classes.

In general, in this research, under sampling techniques provided improved accuracy compared to oversampling and hybrid data balancing approaches. Oversampling which has proven to be a better resampling technique in most of the studies didn't perform well in this case. It didn't provide acceptable results. The accuracy of classification by oversampling technique dropped below 60 % with overall

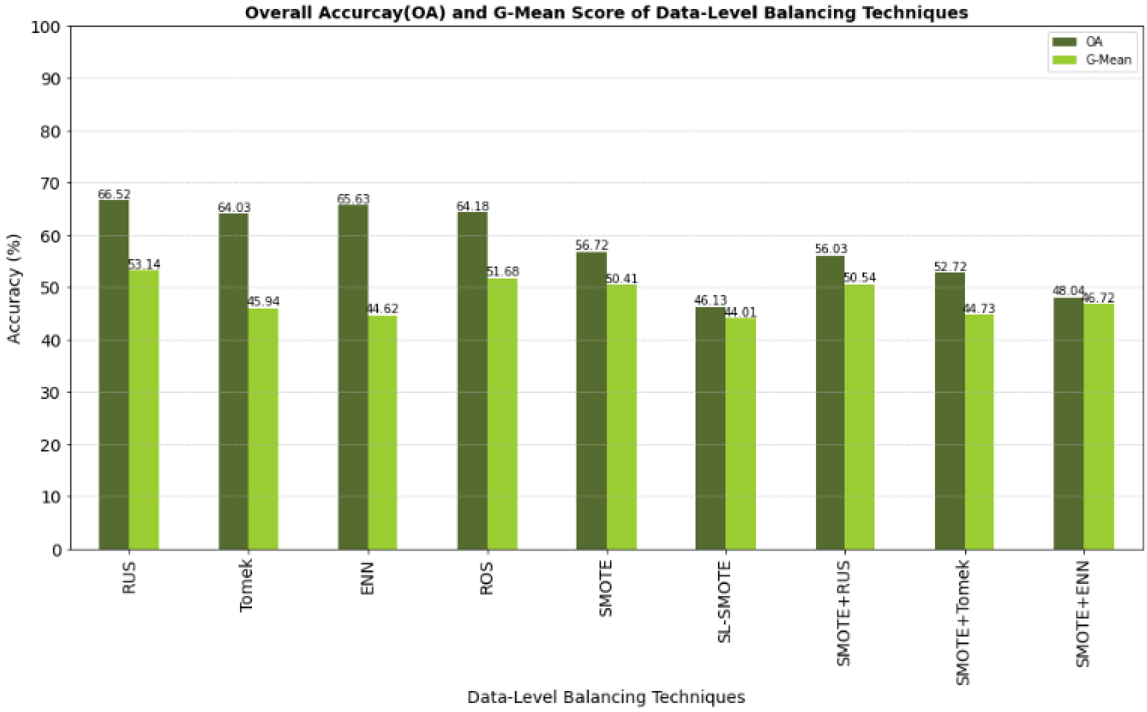

Figure 29 Overall Accuracy and G-Mean Score of different Data Level Balancing Techniques

The main reason for this is the generation of new synthetic minority class instances at wrong geographical locations. For example, let's consider SMOTE, which is an oversampling technique. In implementation, let's say SMOTE has randomly selected a minority class instance in the feature space, which is a mustard crop filed. Now, it looks for k=3 nearest neighbours of minority class instance: mustard. Finally, new artificial instances of class mustard are generated along the line joining the instance of interest and the k nearest neighbours. Almost similar reflectance values for the crop classes are a possible reason which might have caused confusion during the generation of new artificial crop instances. For the construction of feature space the R, G, B, NIR and NDVI values of a single satellite imagery which captures the peak growth condition of crops is considered. Instead of this, feature space construction from multi-temporal images might produce information about the length of growing season as well as the temporal profile of crops. Adoption of this technique would be a possible approach to overcome the false synthetic sample generation caused by considering single satellite imagery. Similarly, hybrid sampling approaches which are a combination of oversampling and under sampling techniques did not perform as expected due to the aforementioned reason. Some possible reasons for the failure of SMOTE are blind selection of neighbourhood, creation of overlapping artificial samples which adds more noise to the data. (Jiang et al., 2021). Also, SMOTE is not very effective for high dimensions data used in the research.

The incorrect artificial instances generated for minority class mustard is shown in the Figure 30. We can clearly observe that newly generated samples have been mixed with forest and scrubland class. In the same way, Figure 31 shows the newly generated chickpea instances. The NDVI trends of these instances do not resemble agricultural crops. Figure 32 denotes the false samples generated for class barren/fallow. In this case, some of the new artificial samples represent the trend of agricultural crops. Therefore, we can confirm that the drop in accuracy of the CNN classification model with oversampled training data is due to the feeding of incorrect input samples to the model. The correct temporal profiles of the three target crops can be referred from the Figure 15 under Methodology Section 4.2.3.

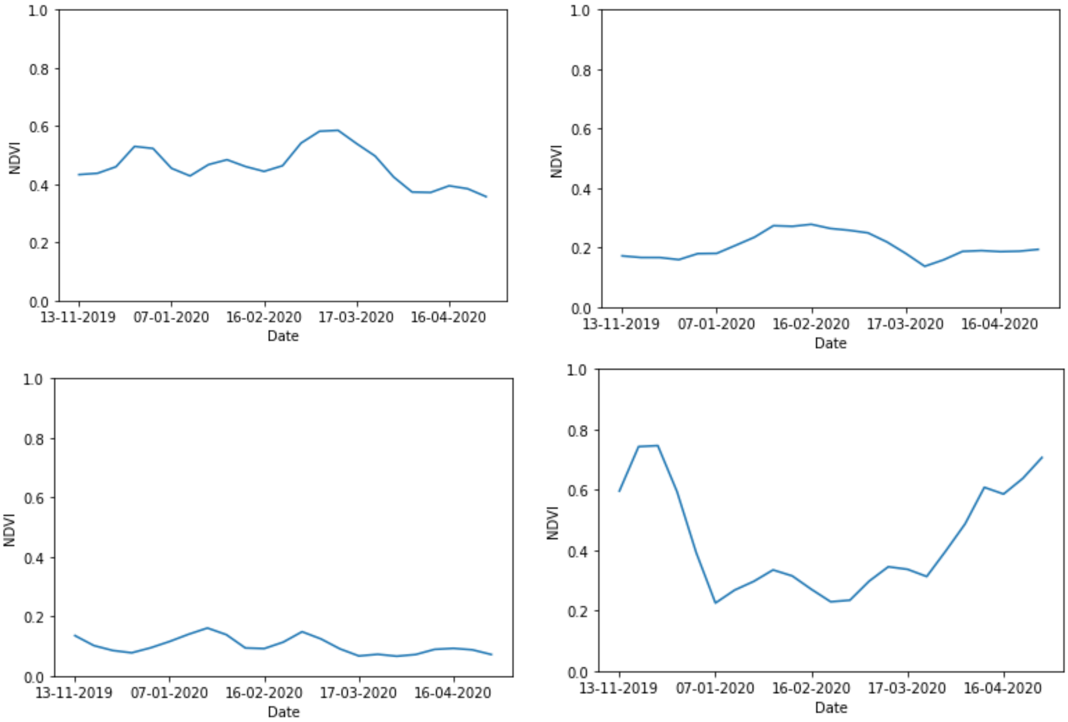

Figure 30 Incorrect artificial samples of Mustard class generated through SMOTE oversampling technique

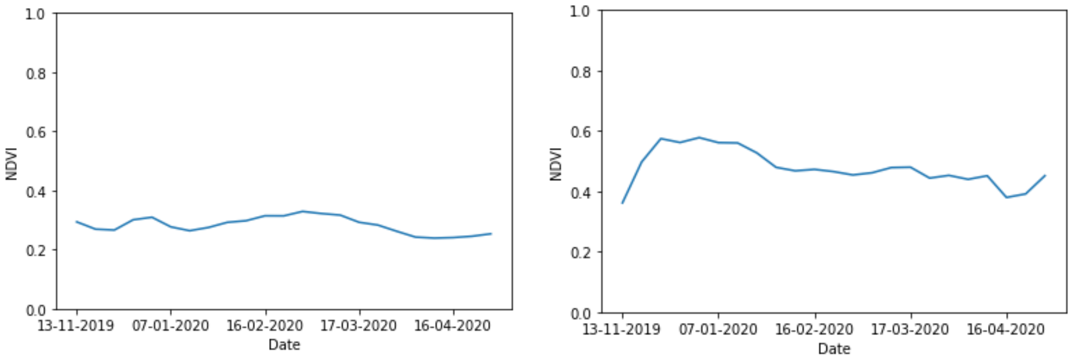

Figure 31 Incorrect artificial samples of Chickpea class generated through SMOTE oversampling technique

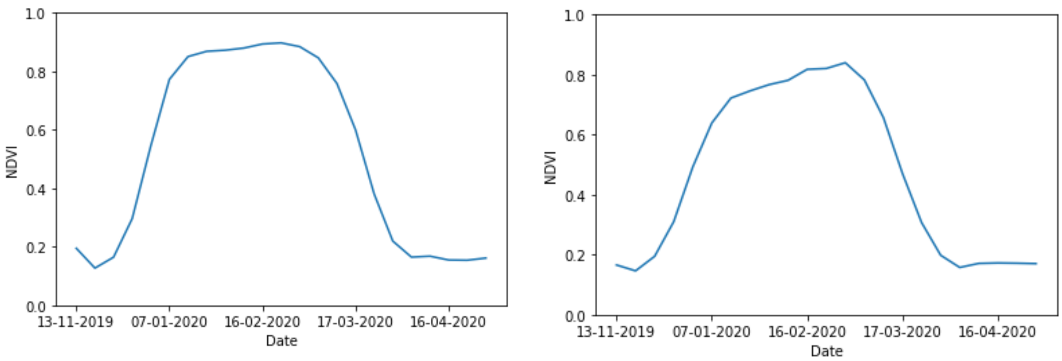

Figure 32 Incorrect artificial samples of Barren/Fallows class generated through SMOTE oversampling technique

Even though class samples with potential information were removed in under sampling, it performed better than oversampling as under sampling doesn't involve the generation of new artificial samples some of which could possibly be false. Training the CNN model with balanced out classes from under sampling, resulted in an enhanced overall accuracy in all the three under sampling techniques. The G-Mean Accuracy obtained by using Tomek Link (0.45) and ENN (0.49) did not show any significant improvement. In contrast, classification results of Random under sampling showed an enhanced G-Mean accuracy of 0.5314 which is slightly higher than the one obtained from imbalanced classes (0.4995).

5.4. Data Level Balancing Techniques vs Algorithmic Balancing Techniques

By analysing the accuracy metrics, mainly G-Mean Accuracy of obtained through the classification results of three algorithmic level balancing techniques (cross entropy loss, focal loss and class weighted loss) and nine data level balancing techniques utilized in this research, it is evident that algorithmic level balancing methods have performed better with class weighted loss resulting with the highest accuracy. Algorithmic level data balancing techniques enhance classification results by adjusting the loss function, without disturbing or making changes to the training data. Therefore, the results provided by these techniques are trustworthy as opposed to the results generated by data level balancing techniques which includes the generation of artificial samples which came out to be false in this research.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusion

With the intent to find a solution for the issue of sample imbalance in crop type mapping, the capabilities of algorithmic level balancing techniques as well as data level balancing techniques were explored in this research. With the help of multi-temporal satellite imagery, the phenological growth information was captured using NDVI. Due to smaller training dataset availability, 1D CNN sequential deep learning model was chosen as it is known to provide better accuracy. For the first objective, Bayesian optimization method was opted to find the best hyper-parameters of the deep learning model. The number of epochs, neurons, learning rate, dropout rate and batch size were found out using this method. Now, in order to solve the problem of class imbalance, two techniques were adopted.

To address the second objective, Algorithmic Balancing techniques were used. These techniques attempts to solve imbalance issue by introducing loss functions which will try to reduce the imbalance in the dataset. Different loss functions such as categorical cross entropy, focal loss and class weighted loss were employed to solve the issue of class imbalance. Focal loss weighs down the class samples that could be easily learnt whereas Class weighted loss emphasises learning of the minority classes by introducing weights for each and every class. The best model is picked out by assessing the accuracy metrics such as Overall Accuracy, Precision, Recall and F1 score. The final conclusion is drawn with the help of G-Mean Score, which serves as the best accuracy metric for problems with class imbalance. Out of all three 1D CNN deep learning model, the model with the class weighted loss function showed the best performance with an overall accuracy of 70.37%, and a G-Mean Score of 56.24%. The G-Mean score for the individual crop classes from this model were 0.81 for wheat, 0.45 for mustard and 0.74 for chickpea.

For the third objective, Data level class balancing was performed with under sampling, oversampling and hybrid techniques. After performing balancing in the dataset, 1D CNN model with cross entropy loss is trained and the classification results are obtained. From the results obtained, it is clear that oversampling and hybrid sampling techniques failed to address class imbalance issue due to the generation of false artificial samples and the accuracy of the models that utilized these data for training went down. Out of all the data sampling techniques, random under sampling, which randomly removes majority class samples to balance out the data performed the best with an overall accuracy of 66.52% and G-Mean Score of 53.14%.

In conclusion, it is evident from the classification results that simple data level resampling techniques such as oversampling and under sampling techniques did not show any significant advancement in accuracy for imbalanced classification problem discussed in this research and just made the existing problem worse. Hence, these techniques are not advisable. In the contrary, modifying the learning of the deep learning model by exploring different loss functions provided better classification results which is highlighted by the enhanced G-Mean Score. Therefore, algorithmic balancing is regarded as a better method of addressing imbalance in crop classification studies as opposed to data level balancing techniques.

6.2. Answer to the reseach questions

Question 1: What are the number of epochs, batch size, learning and dropout rate required to produce accurate classification results?

Answer: As discussed in the Section 5.1, all the hyper-parameters of the designed CNN model are optimised through Bayesian hyper-parameter optimization technique. Running 25 iterations with the range values provided for each hyper-parameter, the best ones are selected. For the best 1D CNN model which utilized class weighted loss function, the optimized parameters are as follows: learning rate of 0.01 with Adam optimizer, dropout rate of 0.1, batch size of 952, 97 epochs and 500 neurons in the dense layer prior to output layer.

Question 2: What is the final loss and accuracy of the proposed deep learning model?

Answer: The 1D CNN model with class weighted loss yields a final overall accuracy of 70.37 % and an over G-mean score of 60.56 % which is the highest out of all methods of data balancing experimented in this research. As the validation data samples provided were less, the training and validation loss curves showed a small gap and did not overlap.

Question 3: Does the developed deep learning model increase the accuracy of minority classes?

Answer: Running the classification model with Cross entropy loss resulted in an Overall Accuracy of 63.54 % and an overall G-Mean score of 49.95 %. Replacing cross entropy loss with class weighted loss function, by training the CNN model with individual class weights, resulted in an enhanced overall accuracy and G-Mean accuracy of 70.37% and 56.24% respectively. There a change of 6.83% in terms of overall accuracy. A 6.29 % change in terms of G-Mean Accuracy can be regarded as significant improvement of classification of, minority classes.

Question 4: Which data level sampling method is suitable for the current study and improves the accuracy of classification?
Answer: In comparison with the oversampling and hybrid sampling (oversampling + under sampling), under sampling technique performed better. Random under sampling method performed the best out of all the sampling methods experimented in this research in terms of G-Mean accuracy.

Question 5: To what extent does the developed deep learning model outperform the class balancing techniques?

Answer: The 1D CNN deep learning model with class weighted loss provides good classification results (Overall Accuracy: 70.37%, G-Mean Accuracy: 56.24%) which is higher than the best data level class balancing technique (Overall Accuracy: 66.52%, G-Mean Accuracy: 53.14%). The difference in accuracies between both these methods is a notable percentage of 3.85% in Overall accuracy and 3.1 % in G-Mean Accuracy.

6.3. Short comings of the Research, Recommendations and Future Work

1. The major short coming of this research was the unavailability of essential training data. This in turn led to the adoption of very small validation and testing datasets. One possible approach to overcome this problem would be conduct exhaustive field campaign. Another approach would be to employ transfer learning. A number of studies have applied transfer learning in CNN for crop classification problems and has shown notable results (Nowakowski et al., 2021; Ribani & Marengoni, 2019; L. Wang et al., 2020; W. Zhang et al., 2020). Transfer learning was not possible in this research due to the unavailability of pertained model with the same crop classes.

2. A 1D CNN model was used in this research which had the spectral and temporal information captured from the time series images. The presence of lesser number of training data had an impact on the selection of CNN model to be used for classification. The 3D CNN which is spatio-temporal deep learning model that captures the spatial, spectral and temporal information of crops couldn't be used for crop classification in this case to due to the lesser count and poor spatial coverage of labelled data. A 3D U-Net architecture with Intersection over union loss function is successfully used for learning the individual crop types by (Mohammadi et al., 2021). The capability of 3D CNN model for addressing class imbalance could be explored.

3. The reference data set had ground truths in the form of point data. To increase the sample count for training, the individual fields and other land covers had to be digitized by visual interpretation with Google earth and the Sentinel 2 satellite imagery as reference. This process was time consuming and was the primary challenge in data preparation.

4. In the temporal window of November2019 to April 2020, the dates with high cloud cover were dropped from the study. Although, CNN model produced acceptable results with the context of addressing class imbalance, some valuable temporal information might have been lost due to this technique adopted. Cloud gap filling couldn't be performed due to the lack of continues temporal images with good atmospheric conditions. A potential solution to overcome this problem would be to adopt fusion of Optical and SAR data. Compared to standalone optical data, optical SAR fused data provides robust classification results (Adrian et al., 2021; Riedel et al., 2007; L. Wang et al., 2020).

5. In this study, under sampling removed led to the removal critical information whereas oversampling generated wrong synthetic instances and led to overfitting. Therefore data balancing techniques didn't perform as expected and did not improve the accuracy of classification and failed to provide meaningful solution to the imbalanced classification problem.

LIST OF REFERENCES

Adrian, J., Sagan, V., & Maimaitijiang, M. (2021). Sentinel SAR-optical fusion for crop type mapping using deep learning and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *175*, 215–235. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2021.02.018

Akosa, J. S. (2017). *Predictive Accuracy : A Misleading Performance Measure for Highly Imbalanced Data*.

Alabi, T., Haertel, M., & Adekola-Chiejile, S. (2016). *Investigating the Use of High Resolution Multi-spectral Satellite Imagery for Crop Mapping in Nigeria - Crop and Landuse Classification using WorldView-3 High Resolution Multispectral Imagery and LANDSAT8 Data*. https://doi.org/10.5220/0005767301090120

ALENCAR, R. (2018). *Resampling strategies for imbalanced datasets*. https://www.kaggle.com/code/rafjaa/resampling-strategies-for-imbalanced-datasets/notebook

Amin, E., Belda, S., Pipia, L., Szantoi, Z., El Baroudy, A., Moreno, J., & Verrelst, J. (2022). Multi-Season Phenology Mapping of Nile Delta Croplands Using Time Series of Sentinel-2 and Landsat 8 Green LAI. *Remote Sensing*, *14*(8). https://doi.org/10.3390/rs14081812

Ando, S., & Huang, C. (2017). *Deep Over-sampling Framework for Classifying Imbalanced Data*.

Andreieva, V., & Shvai, N. (2021). Generalization of Cross-Entropy Loss Function for Image Classification. *Mohyla Mathematical Journal*, *3*, 3–10. https://doi.org/10.18523/2617-7080320203-10

Badr, W. (2019). *Having an Imbalanced Dataset? Here Is How You Can Fix It*. Towards Data Science. https://towardsdatascience.com/having-an-imbalanced-dataset-here-is-how-you-can-solve-it-1640568947eb

Batista, G., Bazzan, A., & Monard, M.-C. (2003). Balancing Training Data for Automated Annotation of Keywords: a Case Study. In *the Proc. Of Workshop on Bioinformatics*.

Batista, G., Prati, R., & Monard, M.-C. (2004). A Study of the Behavior of Several Methods for Balancing machine Learning Training Data. *SIGKDD Explorations*, *6*, 20–29. https://doi.org/10.1145/1007730.1007735

Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, *204*, 509–523. https://doi.org/https://doi.org/10.1016/j.rse.2017.10.005

Belgiu, M., Zhou, Y., Marshall, M., & Stein, A. (2020). DYNAMIC TIME WARPING FOR CROPS MAPPING. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B3-2*, 947–951. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-947-2020

Boogaard, F. P., van Henten, E. J., & Kootstra, G. (2022). Improved Point-Cloud Segmentation for Plant Phenotyping Through Class-Dependent Sampling of Training Data to Battle Class Imbalance. *Frontiers in Plant Science*, *13*, 838190. https://doi.org/10.3389/fpls.2022.838190

Bowyer, K. W., Chawla, N. V, Hall, L. O., & Kegelmeyer, W. P. (2011). {SMOTE:} Synthetic Minority Over-sampling Technique. *CoRR*, *abs/1106.1*. http://arxiv.org/abs/1106.1813

Buda, M., Maki, A., & Mazurowski, M. (2017). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*. https://doi.org/10.1016/j.neunet.2018.07.011

Bunkhumpornpat, C., Sinapiromsaran, K., & Lursinsap, C. (2009). *Safe-Level-SMOTE: Safe-Level-Synthetic Minority Over-Sampling TEchnique for Handling the Class Imbalanced Problem BT - Advances in Knowledge Discovery and Data Mining* (T. Theeramunkong, B. Kijsirikul, N. Cercone, & T.-B. Ho (Eds.); pp. 475–482). Springer Berlin Heidelberg.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., & Li, Z. (2018). A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, *210*, 35–47. https://doi.org/https://doi.org/10.1016/j.rse.2018.02.045

Cánovas-García, F., Alonso, F., Castillo, F., & Oñate-Valdivieso, F. (2017). Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery. *Computers & Geosciences*, *103*. https://doi.org/10.1016/j.cageo.2017.02.012

Central Ground Water Board (CGWB). (2017). Aquifer mapping and ground water management plan of Jhansi District, Uttar Pradesh. Northern Region, Lucknow: Central Ground Water Board. In *Central Ground Water Board (CGWB)*. http://cgwb.gov.in/AQM/NAQUIM_REPORT/UP/Jhansi.pdf

Chawla, N, Bowyer, K., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.*, *16*, 321–357.

Chawla, Nitesh. (2005). Data Mining for Imbalanced Datasets: An Overview. In *Data Mining and Knowledge Discovery Handbook, ISBN 978-0-387-09822-7. Springer Science+Business Media, LLC, 2010, p. 875* (Vol.

5, pp. 853–867). https://doi.org/10.1007/0-387-25465-X_40

Chawla, Nitesh, Japkowicz, N., & Kołcz, A. (2004). Editorial: Special Issue on Learning from Imbalanced Data Sets. *SIGKDD Explorations*, *6*, 1–6. https://doi.org/10.1145/1007730.1007733

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment*, *91*(3), 332–344. https://doi.org/https://doi.org/10.1016/j.rse.2004.03.014

Chen, Y., Lu, D., Moran, E., Batistella, M., Dutra, L. V., Sanches, I. D., da Silva, R. F. B., Huang, J., Luiz, A. J. B., & de Oliveira, M. A. F. (2018). Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *International Journal of Applied Earth Observation and Geoinformation*, *69*, 133–147. https://doi.org/https://doi.org/10.1016/j.jag.2018.03.005

Crop Calendar PMFBY. (2018). Pradhan Mantri Fasal Bima Yojana; PMFBY. https://pmfby.gov.in/pdf/New_Crop_Calendar_20.09.18.pdf

Dong, Q., Chen, X., Chen, J., Zhang, C., Liu, L., Cao, X., Zang, Y., Zhu, X., & Cui, X. (2020). Mapping Winter Wheat in North China Using Sentinel 2A/B Data: A Method Based on Phenology-Time Weighted Dynamic Time Warping. In *Remote Sensing* (Vol. 12, Issue 8). https://doi.org/10.3390/rs12081274

Doraiswamy, P. C., & Cook, P. W. (1995). Spring Wheat Yield Assessment Using NOAA AVHRR Data. *Canadian Journal of Remote Sensing*, *21*(1), 43–51. https://doi.org/10.1080/07038992.1995.10874595

Douzas, G., Bacao, F., Fonseca, J., & Khudinyan, M. (2019). Imbalanced learning in land cover classification: Improving minority classes' prediction accuracy using the geometric SMOTE algorithm. *Remote Sensing*, *11*(24). https://doi.org/10.3390/rs11243040

Elhassan, A. A. M., & Al-Mohanna. (2017). *Classification of Imbalance Data using Tomek Link (T-Link) Combined with Random Under-sampling (RUS) as a Data Reduction Method.*

Fang, P., Zhang, X., Wei, P., Wang, Y., Zhang, H., Liu, F., & Zhao, J. (2020). The Classification Performance and Mechanism of Machine Learning Algorithms in Winter Wheat Mapping Using Sentinel-2 10 m Resolution Imagery. In *Applied Sciences* (Vol. 10, Issue 15). https://doi.org/10.3390/app10155075

Feng, S., Zhao, J., Liu, T., Zhang, H., Zhang, Z., & Guo, X. (2019). Crop Type Identification and Mapping Using Machine Learning Algorithms and Sentinel-2 Time Series Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(9), 3295–3306. https://doi.org/10.1109/JSTARS.2019.2922469

Fernando, K. R. M., & Tsokos, C. P. (2022). Dynamically Weighted Balanced Loss: Class Imbalanced Learning and Confidence Calibration of Deep Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(7), 2940–2951. https://doi.org/10.1109/TNNLS.2020.3047335

Foerster, S., Kaden, K., Förster, M., & Itzerott, S. (2012). Crop type mapping using spectral-temporal profiles and phenological information. *Computers and Electronics in Agriculture*, *89*, 30–40. https://doi.org/10.1016/j.compag.2012.07.015

Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks*, *1*(2), 119–130. https://doi.org/https://doi.org/10.1016/0893-6080(88)90014-7

Gallo, K. P., & Flesch, T. K. (1989). Large-area crop monitoring with the NOAA AVHRR: Estimating the silking stage of corn development. *Remote Sensing of Environment*, *27*(1), 73–80. https://doi.org/https://doi.org/10.1016/0034-4257(89)90038-2

Gholamalinejad, H., & Khosravi, H. (2020). *Pooling Methods in Deep Neural Networks, a Review.*

Griffith, D. A., & Chun, Y. (2016). Spatial Autocorrelation and Uncertainty Associated with Remotely-Sensed Data. In *Remote Sensing* (Vol. 8, Issue 7). https://doi.org/10.3390/rs8070535

Gumma, M., Tummala, K., Dixit, S., Collivignarelli, F., Holecz, F., Kolli, N., & Whitbread, A. (2020). Crop type identification and spatial mapping using Sentinel 2 satellite data with focus on field-level information. *Geocarto International*, 1–14. https://doi.org/10.1080/10106049.2020.1805029

Guo, T., Dong, J., Li, H., & Gao, Y. (2017). Simple convolutional neural network on image classification. *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, 721–724. https://doi.org/10.1109/ICBDA.2017.8078730

He, H., Bai, Y., Garcia, E., & Li, S. (2008). ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. In *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN.2008.4633969

Hernandez, I., Benevides, P., Costa, H., & Caetano, M. (2020). EXPLORING SENTINEL-2 FOR LAND COVER AND CROP MAPPING IN PORTUGAL. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B3-2*, 83–89.

https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-83-2020
Hoeser, T., & Kuenzer, C. (2020). Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sensing*, *12*. https://doi.org/10.3390/rs12101667
How to Train and Test Data Like a Pro. (2021). Sdsclub.Com. https://sdsclub.com/how-to-train-and-test-data-like-a-pro/
Htitiou, A., Boudhar, A., Lebrini, Y., Lionboui, H., Chehbouni, A., & Benabdelouahab, T. (2021). Classification and status monitoring of agricultural crops in central Morocco: A synergistic combination of OBIA approach and fused Landsat-Sentinel-2 data. *Journal of Applied Remote Sensing*, *15*(1). https://doi.org/10.1117/1.JRS.15.014504
Huang, B., Lu, K., Audeberr, N., Khalel, A., Tarabalka, Y., Malof, J., Boulch, A., Saux, B. Le, Collins, L., Bradbury, K., Lefèvre, S., & El-Saban, M. (2018). Large-Scale Semantic Classification: Outcome of the First Year of Inria Aerial Image Labeling Benchmark. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 6947–6950. https://doi.org/10.1109/IGARSS.2018.8518525
Hui, L., & Belkin, M. (2021). Evaluation of Neural Architectures Trained with Square Loss vs Cross-Entropy in Classification Tasks. *ArXiv*, *abs/2006.0*.
Ibrahim, E. S., Rufin, P., Nill, L., Kamali, B., Nendel, C., & Hostert, P. (2021). Mapping crop types and cropping systems in nigeria with sentinel-2 imagery. *Remote Sensing*, *13*(17). https://doi.org/10.3390/rs13173523
Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. In *Remote Sensing* (Vol. 8, Issue 3). https://doi.org/10.3390/rs8030166
Inglada, J., Vincent, A., Arias, M., & Marais-Sicre, C. (2016). Improved Early Crop Type Identification By Joint Use of High Temporal Resolution SAR And Optical Image Time Series. In *Remote Sensing* (Vol. 8, Issue 5). https://doi.org/10.3390/rs8050362
Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., & Rodes, I. (2017). Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing*, *9*, 95. https://doi.org/10.3390/rs9010095
Japkowicz, N. (2000). The Class Imbalance Problem: Significance and Strategies. *Proceedings of the 2000 International Conference on Artificial Intelligence ICAI*.
Ji, S., Zhang, C., Xu, A., Shi, Y., & Duan, Y. (2018). 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. In *Remote Sensing* (Vol. 10, Issue 1). https://doi.org/10.3390/rs10010075
Jiang, Z., Pan, T., Zhang, C., & Yang, J. (2021). A new oversampling method based on the classification contribution degree. *Symmetry*, *13*(2), 1–13. https://doi.org/10.3390/sym13020194
Jo, D. T., & Japkowicz, N. (2004). Class imbalances versus small disjuncts. *SIGKDD Explorations*, *6*, 40–49. https://doi.org/10.1145/1007730.1007737
Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. https://doi.org/https://doi.org/10.1016/j.compag.2018.02.016
Karasiak, N., Dejoux, J.-F., Monteil, C., & Sheeren, D. (2022). Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*, *111*(7), 2715–2740. https://doi.org/10.1007/s10994-021-05972-1
Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*, 24–49. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2020.12.010
Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. (2021). 1D Convolutional Neural Networks and Applications: A Survey. *Mechanical Systems and Signal Processing*, *151*. https://doi.org/10.1016/j.ymssp.2020.107398
Kulathunga, G., & Afanasyev, I. (2018). *Deep Learning Approach for Building Detection in Satellite Multispectral Imagery*.
Kumar, P., Dobriyal, M., Kale, A., & Pandey, A. K. (2021). Temporal dynamics change of land use/land cover in Jhansi district of Uttar Pradesh over past 20 years using LANDSAT TM, ETM+ and OLI sensors. *Remote Sensing Applications: Society and Environment*, *23*, 100579. https://doi.org/https://doi.org/10.1016/j.rsase.2021.100579
Kussul, N., Shelestov, A., Lavreniuk, M., Yailymov, B., Kolotii, A., Yailymova, H., Skakun, S., & Shumilo, L. (2018). DEEP LEARNING APPROACH FOR CROP MAPPING ON BIG DATA. In *Space research in Ukraine* (pp. 79–84). Kyiv : Akademperiodyka.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten Digit Recognition with a Back-Propagation Network. In D. Touretzky (Ed.), *Advances in Neural Information Processing Systems* (Vol. 2). Morgan-Kaufmann. https://proceedings.neurips.cc/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf

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

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). *Focal Loss for Dense Object Detection*. https://doi.org/10.1109/ICCV.2017.324

Lin, T., Goyal, P., Girshick, R., He, K., & Dollár, P. (2020). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826

Ling, C. X., & Li, C. (1998). Data Mining for Direct Marketing: Problems and Solutions. *KDD*.

Liu, X., Ji, L., Zhang, C., & Liu, Y. (2022). A method for reconstructing NDVI time-series based on envelope detection and the Savitzky-Golay filter. *International Journal of Digital Earth*, *15*(1), 553–584. https://doi.org/10.1080/17538947.2022.2044397

Liu, Y., Minh Nguyen, D., Deligiannis, N., Ding, W., & Munteanu, A. (2017). Hourglass-ShapeNetwork Based Semantic Segmentation for High Resolution Aerial Imagery. In *Remote Sensing* (Vol. 9, Issue 6). https://doi.org/10.3390/rs9060522

Liu, Z., & Sun, Z. (2008). *Active one-class classification of remote sensing image*. *7285*. https://doi.org/10.1117/12.816115

Löw, F., Michel, U., Dech, S., & Conrad, C. (2013). Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. *ISPRS Journal of Photogrammetry and Remote Sensing*, *85*, 102–119. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2013.08.007

Ma, H., Huang, W., Jing, Y., Yang, C., Han, L., Dong, Y., Ye, H., Shi, Y., Zheng, Q., Liu, L., & Ruan, C. (2019). Integrating growth and environmental parameters to discriminate powdery mildew and aphid ofwinter wheat using bi-temporal Landsat-8 imagery. *Remote Sensing*, *11*(7). https://doi.org/10.3390/RS11070846

Martos, V., Ahmad, A., Cartujo, P., & Ordoñez, J. (2021). Ensuring Agricultural Sustainability through Remote Sensing in the Era of Agriculture 5.0. In *Applied Sciences* (Vol. 11, Issue 13). https://doi.org/10.3390/app11135911

Massey, R., Sankey, T., Congalton, R., Yadav, K., Thenkabail, P., Ozdogan, M., & Sánchez Meador, A. (2017). MODIS phenology-derived, multi-year distribution of conterminous U.S. crop types. *Remote Sensing of Environment*, *198*, 490–503. https://doi.org/10.1016/j.rse.2017.06.033

Mazzia, V., Khaliq, A., & Chiaberge, M. (2020). Improvement in land cover and crop classification based on temporal features learning from Sentinel-2 data using recurrent-Convolutional Neural Network (R-CNN). *Applied Sciences (Switzerland)*, *10*(1). https://doi.org/10.3390/app10010238

Meidianingsih, Q., & Sartono, B. (2017). *The Study of Safe-Level SMOTE Method in Unbalanced Data Classification*.

Mohammadi, S., Belgiu, M., & Stein, A. (2021). *3D Fully Convolutional Neural Networks with Intersection over Union Loss for Crop Mapping from Multi-Temporal Satellite Images*. https://doi.org/10.1109/IGARSS47720.2021.9554573

Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Gill, E., & Molinier, M. (2019). A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS Journal of Photogrammetry and Remote Sensing*, *151*. https://doi.org/10.1016/j.isprsjprs.2019.03.015

Moreno-Revelo, M. Y., Guachi-Guachi, L., Gómez-Mendoza, J. B., Revelo-Fuelagán, J., & Peluffo-Ordóñez, D. H. (2021). Enhanced Convolutional-Neural-Network Architecture for Crop Classification. In *Applied Sciences* (Vol. 11, Issue 9). https://doi.org/10.3390/app11094292

Nair, V., & Hinton, G. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. In *Proceedings of ICML* (Vol. 27).

Nisa, E. C., & Kuan, Y.-D. (2021). Comparative assessment to predict and forecast water-cooled chiller power consumption using machine learning and deep learning algorithms. *Sustainability (Switzerland)*, *13*(2), 1–18. https://doi.org/10.3390/su13020744

Nowakowski, A., Mrziglod, J., Spiller, D., Bonifacio, R., Ferrari, I., Mathieu, P. P., Garcia-Herranz, M., &

Kim, D.-H. (2021). Crop type mapping by using transfer learning. *International Journal of Applied Earth Observation and Geoinformation*, *98*. https://doi.org/10.1016/j.jag.2021.102313

Orynbaikyzy, A., Gessner, U., & Conrad, C. (2019). Crop type classification using a combination of optical and radar remote sensing data: a review. *International Journal of Remote Sensing*, *40*(17), 6553–6595. https://doi.org/10.1080/01431161.2019.1569791

Pan, L., Xia, H., Wang, R., Niu, W., Tian, H., & Qin, Y. (2021). Mapping of the winter crop planting areas in Huaihe River Basin based on Google Earth Engine . *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, *37*(18), 211–218. https://doi.org/10.11975/j.issn.1002-6819.2021.18.025

Pan, X., Gao, L., Marinoni, A., Zhang, B., Yang, F., & Gamba, P. (2018). Semantic Labeling of High Resolution Aerial Imagery and LiDAR Data with Fine Segmentation Network. In *Remote Sensing* (Vol. 10, Issue 5). https://doi.org/10.3390/rs10050743

Penuelas, J., Rutishauser, T., & Filella, I. (2009). Phenology Feedbacks on Climate Change. *Science (New York, N.Y.)*, *324*, 887–888. https://doi.org/10.1126/science.1173004

Pilotti, G. (2020). *Oversampling and Undersampling: ADASYN vs ENN*. Quantyca. https://medium.com/quantyca/oversampling-and-undersampling-adasyn-vs-enn-60828a58db39

Pinjarla, B., Gumma, M., S, J., & Roy, P. (2019). *High resolution Sentinel-2 crop type mapping 2018- 2019 A case study in Ahmednagar district*. https://doi.org/10.13140/RG.2.2.10138.39362

R., P., & Yadav, K. (2015). Monitoring spatio-temporal pattern of drought stress using integrated drought index over Bundelkhand region, India. *Natural Hazards*, *77*, 663–677.

Rala Cordeiro, J., Raimundo, A., Postolache, O., & Sebastião, P. (2021). Neural Architecture Search for 1D CNNs—Different Approaches Tests and Measurements. In *Sensors* (Vol. 21, Issue 23). https://doi.org/10.3390/s21237990

Reynolds, A. H. (2019). *Convolutional Neural Networks (CNNs)*. https://anhreynolds.com/blogs/cnn.html

Ribani, R., & Marengoni, M. (2019). A Survey of Transfer Learning for Convolutional Neural Networks. *Proceedings - 32nd Conference on Graphics, Patterns and Images Tutorials, SIBGRAPI-T 2019*, 47–57. https://doi.org/10.1109/SIBGRAPI-T.2019.00010

Richardson, A., Keenan, T., Migliavacca, M., Ryu, Y., Sonnentag, O., & Toomey, M. (2013). Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, *169*, 156–173. https://doi.org/10.1016/j.agrformet.2012.09.012

Riedel, T., Thiel, C., & Schmullius, C. (2007). Fusion of optical and SAR satellite data for improved land cover mapping in agricultural areas. *European Space Agency, (Special Publication) ESA SP*, *SP-636*. https://www.scopus.com/inward/record.uri?eid=2-s2.0-36549034860&partnerID=40&md5=22291c3cc3a329ff372ed575e61576ca

Rizaldy, A. (2018). *Deep learning-based DTM extraction from LIDAR point cloud*. http://essay.utwente.nl/85872/

Sawangarreerak, S., & Thanathamathee, P. (2020). Random forest with sampling techniques for handling imbalanced prediction of university student depression. *Information (Switzerland)*, *11*(11), 1–13. https://doi.org/10.3390/info11110519

Schaefer, J., Milling, M., Schuller, B. W., Bauer, B., Brunner, J. O., Traidl-Hoffmann, C., & Damialis, A. (2021). Towards automatic airborne pollen monitoring: From commercial devices to operational by mitigating class-imbalance in a deep learning approach. *Science of The Total Environment*, *796*, 148932. https://doi.org/https://doi.org/10.1016/j.scitotenv.2021.148932

Schiele, J., Rabe, F., Schmitt, M., Glaser, M., Häring, F., Brunner, J. O., Bauer, B., Schuller, B., Traidl-Hoffmann, C., & Damialis, A. (2019). Automated Classification of Airborne Pollen using Neural Networks. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4474–4478. https://doi.org/10.1109/EMBC.2019.8856910

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/https://doi.org/10.1016/j.neunet.2014.09.003

Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., & Skakun, S. (2017). Large scale crop classification using Google earth engine platform. *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 3696–3699. https://doi.org/10.1109/IGARSS.2017.8127801

Shi, D., Yang, & Xiaojun. (2016). An Assessment of Algorithmic Parameters Affecting Image Classification Accuracy by Random Forests. *Photogrammetric Engineering & Remote Sensing*, *82*, 407–417. https://doi.org/10.14358/PERS.82.6.407

Shukla, P., & Iriondo, R. (2022). *Main Types of Neural Networks and its Applications — Tutorial*. TowardsAI. https://towardsai.net/p/machine-learning/main-types-of-neural-networks-and-its-applications-

tutorial-734480d7ec8e
Sidike, P., Sagan, V., Maimaitijiang, M., Maimaitiyiming, M., Shakoor, N., Burken, J., Mockler, T., & Fritschi, F. B. (2019). dPEN: deep Progressively Expanded Network for mapping heterogeneous agricultural landscape using WorldView-3 satellite imagery. *Remote Sensing of Environment*, *221*, 756–772. https://doi.org/10.1016/j.rse.2018.11.031
Singh, A. K., Raj, B., Tiwari, A. K., & Mahato, M. K. (2013). Evaluation of hydrogeochemical processes and groundwater quality in the Jhansi district of Bundelkhand region, India. *Environmental Earth Sciences*, *70*(3), 1225–1247. https://doi.org/10.1007/s12665-012-2209-7
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.
Sun, Y., Kamel, M. S., & Wang, Y. (2006). Boosting for Learning Multiple Classes with Imbalanced Class Distribution. *Sixth International Conference on Data Mining (ICDM'06)*, 592–602. https://doi.org/10.1109/ICDM.2006.29
Sun, Z., Di, L., Fang, H., & Burgess, A. (2020). Deep Learning Classification for Crop Types in North Dakota. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *PP*, 1. https://doi.org/10.1109/JSTARS.2020.2990104
Teimouri, N., Dyrmann, M., & Jørgensen, R. (2019). A Novel Spatio-Temporal FCN-LSTM Network for Recognizing Various Crop Types Using Multi-Temporal Radar Images. *Remote Sensing*, *11*, 990. https://doi.org/10.3390/rs11080990
Teluguntla, P., Thenkabail, P., Xiong, Gumma, M., Chandra, G., Cristina, M., Ozdogan, M., Congalton, R., Tilton, J., Sankey, T., Massey, Phalke, A., & Yadav, K. (2015). *Global Food Security Support Analysis Data at Nominal 1 km (GFSAD1km) Derived from Remote Sensing in Support of Food Security in the Twenty-First Century: Current Achievements and Future Possibilities, Chapter 6* (pp. 131–160).
Teoh, T., & Rong, Z. (2022). *Convolutional Neural Networks* (pp. 261–275). https://doi.org/10.1007/978-981-16-8615-3_16
Thenkabail, P., Knox, J., Ozdogan, M., Gumma, M., Congalton, R., Wu, Z., You, S., Cristina, M., Chandra, G., Nagler, P., Finkral, A., Marshall, M., & Mariotto, I. (2012). Assessing future risks to agricultural productivity, water resources and food security: How can remote sensing help? *Photogrammetric Engineering and Remote Sensing*, *78*, 773–782.
Tomek, I. (1976). Two Modifications of CNN. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-6*(11), 769–772. https://doi.org/10.1109/TSMC.1976.4309452
Van Tricht, K., Gobin, A., Gilliams, S., & Piccard, I. (2018). Synergistic Use of Radar Sentinel-1 and Optical Sentinel-2 Imagery for Crop Mapping: A Case Study for Belgium. In *Remote Sensing* (Vol. 10, Issue 10). https://doi.org/10.3390/rs10101642
Waldner, F., Chen, Y., Lawes, R., & Hochman, Z. (2019). Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods. *Remote Sensing of Environment*, *233*, 111375. https://doi.org/https://doi.org/10.1016/j.rse.2019.111375
Wang, L., Liu, M., Liu, X., Wu, L., Wan, P., & Wu, C. (2020). Pretrained convolutional neural network for classifying rice-cropping systems based on spatial and spectral trajectories of Sentinel-2 time series. *Journal of Applied Remote Sensing*, *14*(1). https://doi.org/10.1117/1.JRS.14.014506
Wang, S., Di Tommaso, S., Faulkner, J., Friedel, T., Kennepohl, A., Strey, R., & Lobell, D. B. (2020). Mapping Crop Types in Southeast India with Smartphone Crowdsourcing and Deep Learning. In *Remote Sensing* (Vol. 12, Issue 18). https://doi.org/10.3390/rs12182957
Wei, X., Yu, X., Liu, B., & Zhi, L. (2019). Convolutional neural networks and local binary patterns for hyperspectral image classification. *European Journal of Remote Sensing*, *52*(1), 448–462. https://doi.org/10.1080/22797254.2019.1634980
Wilson, D. L. (1972). Asymptotic Properties of Nearest Neighbor Rules Using Edited Data. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-2*(3), 408–421. https://doi.org/10.1109/TSMC.1972.4309137
Wulder, M., & Boots, B. (1998). Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *International Journal of Remote Sensing*, *19*(11), 2223–2231. https://doi.org/10.1080/014311698214983
Xie, B., Zhang, H. K., & Xue, J. (2019). Deep convolutional neural network for mapping smallholder agriculture using high spatial resolution satellite image. *Sensors (Switzerland)*, *19*(10). https://doi.org/10.3390/s19102398
Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of Plant*

Ecology, *1*(1), 9–23. https://doi.org/10.1093/jpe/rtm005
Xu, Y., Yu, L., Zhao, F. R., Cai, X., Zhao, J., Lu, H., & Gong, P. (2018). Tracking annual cropland changes from 1984 to 2016 using time-series Landsat images with a change-detection and post-classification approach: Experiments from three sites in Africa. *Remote Sensing of Environment*, *218*, 13–31. https://doi.org/https://doi.org/10.1016/j.rse.2018.09.008
Yan, E., Wang, G., Lin, H., Xia, C., & Sun, H. (2015). Phenology-based classification of vegetation cover types in Northeast China using MODIS NDVI and EVI time series. *International Journal of Remote Sensing*, *36*, 489–512. https://doi.org/10.1080/01431161.2014.999167
Yao, C., Zhang, Y., & Liu, H. (2017). APPLICATION OF CONVOLUTIONAL NEURAL NETWORK IN CLASSIFICATION OF HIGH RESOLUTION AGRICULTURAL REMOTE SENSING IMAGES. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2/W7*, 989–992. https://doi.org/10.5194/isprs-archives-XLII-2-W7-989-2017
Zhang, L., Xia, G.-S., Wu, T., Lin, L., & Tai, X. C. (2016). Deep Learning for Remote Sensing Image Understanding. *Journal of Sensors*, *2016*, 7954154. https://doi.org/10.1155/2016/7954154
Zhang, W., Liu, H., Wu, W., Zhan, L., & Wei, J. (2020). Mapping rice paddy based on machine learning with sentinel-2 multi-temporal data: Model comparison and transferability. *Remote Sensing*, *12*(10). https://doi.org/10.3390/rs12101620
Zhang, X., Cai, Z., Yang, J., Wang, C., Wei, H., Hu, Q., & Xu, B. (2022). Impacts of temporal smoothing methods on crop type identification . *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, *38*(4), 215–224. https://doi.org/10.11975/j.issn.1002-6819.2022.04.025
Zhao, L., Shi, Y., Liu, B., Hovis, C., Duan, Y., & Shi, Z. (2019). Finer classification of crops by fusing UAV images and sentinel-2A data. *Remote Sensing*, *11*(24). https://doi.org/10.3390/rs11243012
Zhong, L., Hawkins, T., Biging, G., & Gong, P. (2011). A phenology-based approach to map crop types in the San Joaquin Valley, California. *International Journal of Remote Sensing*, *32*(22), 7777–7804. https://doi.org/10.1080/01431161.2010.527397
Zhong, L., Hu, L., & Zhou, H. (2019). Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, *221*, 430–443. https://doi.org/https://doi.org/10.1016/j.rse.2018.11.032
Zhou, Y., Wang, X., Zhang, M., Zhu, J., Zheng, R., & Wu, Q. (2019). MPCE: A Maximum Probability Based Cross Entropy Loss Function for Neural Network Classification. *IEEE Access*, *PP*, 1. https://doi.org/10.1109/ACCESS.2019.2946264
Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307

APPENDIX

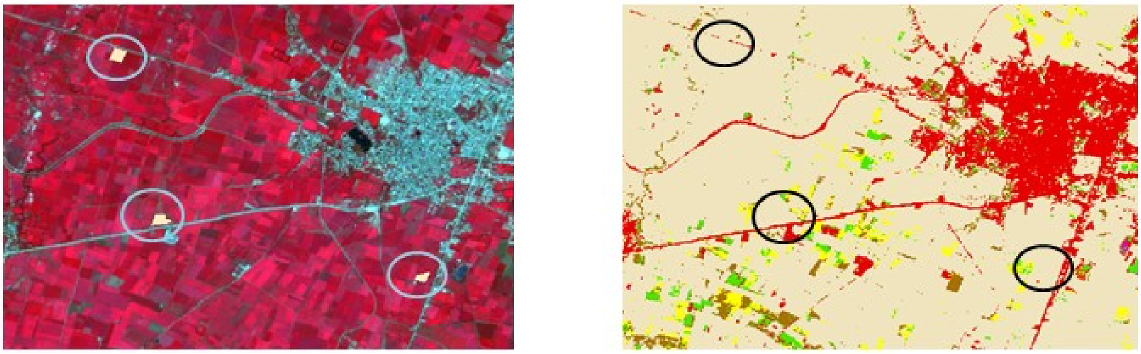

Figure 33 Identification of Wheat Samples in Classified Map

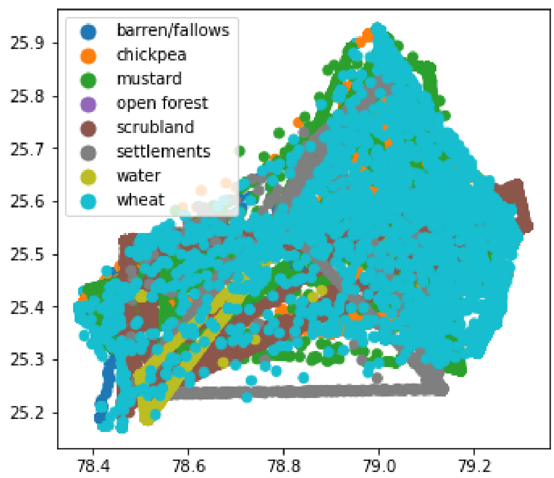

Figure 34 Result of SMOTE