ABSTRACT

Rice is the main food of the inhabitants of Sri Lanka. Hence, monitoring at every growth phase of a rice plant is important information fact for an estimate the yield prediction before harvest. It helps many aspects like strategic planning and decision making regarding the food security and facilitation of safe harvest storages. In Sri Lanka, a traditional “crop cutting survey” fails to forecast rice yield before harvest as this experiment is conducted during the harvest. However, previous research studies have concluded that satellite RS (Remote Sensing) as a potential tool for identifying the paddy stages and estimating and forecasting crop yields. In this study, used Sentinel-2A free RS data from 2015 to 2019 over Gampaha, Kiridiwita selected paddy area, for classifying the paddy stages (Germination, Vegetative phase, Reproductive phase and Ripening phase) and also build the model for selected area harvest predicting. For classifying the paddy stages, used image classification using Support Vector Machine(SVM) and Convolutional Neural Network(CNN) approach. Since Sentinel-2A carries the Multispectral Imager (MSI) then (B04, B03, B02), (B11, B08, B04), (B11, B08, B02) bands combinations are selected to build the image classification model for identifying the paddy crop stages. Among them (B11, B08, B02) bands combination shows a more accurate model with an average overall accuracy of 92% in SVM classification model and also CNN shows the same accuracy with data augmentation. For forecasting the paddy yield of the selected area, analysis the linear gradient and polynomial coefficients in order two of NDVI(Normalized Difference Vegetation Index) values of every Yala, Maha season in each year and map with the past harvest data and build the model. The gradient of the linear in temporal NDVI paddy seasonal data(Yala & Maha) vs past harvest data shows more correlation with other than Polynomial coefficients. So paddy yield forecasting selected model was in AOI. Once the paddy crop is identified as in Ripening phase, it shows the paddy yield forecasting result per acre. Therefore, this experiment has demonstrated that the seasonal rice yield can successfully be forecasted one month prior to harvest with considerably higher accuracy.

Keywords — paddy yield forecasting, paddy stage identification, Sentinel-2 images, NDVI, SVM

LIST OF ABBREVIATIONS

AOI - Area Of Interest

ML - Machine Learning

IR - Infrared

SWIR - Short Wave Infrared

RS - Remote Sensor

MODIS - Moderate Resolution Imaging Spectroradiometer

ESA -   European Space Agency

NDVI - Normalized Vegitation Index

AFRI - Aerosol free Vegitation Index

LAI - Leaf Area Index

SVM - Support Vector Machine

CNN - Convolutional Neural Network

DNN - Deep Neural Netwok

ANN - Artificial Neural Network

1A1 - One Against One

1AA - One Against All

# Introduction

## Background

Rice is the main food of the inhabitants of Sri Lanka. Paddy crops are cultivated as a wetland crop in all the districts. The total land devoted to paddy is estimated to be about 708,000 Hectares at present (Division, n.d.). There are two cultivation seasons namely Maha and Yala which are synonymous with two monsoons in Sri Lanka. According to Paddy Statistics of Department of the census and Statistic, the average yield of paddy estimated for 2017 Yala season was 83.2 bushels per net acre. It decreases 2.4 bushels per net acre compare to 2016 Yala Season and Paddy estimated 2017/2018 Maha season was 83.43 bushels per net acre. It increases 123.26 bushels per acre compare to previous Maha season. The aforementioned data infers the growth of the rice yield per unit area, and also reflects the development of the national economy. Therefore, precisely mapping the cultivation is and estimating the yield of paddy rice are both crucial for national food security and national development evaluation.

In Sri Lanka, Average paddy yield is estimated by a simple survey. This process, commonly known as “crop cutting survey”, was introduced by the Food and Agriculture Organization of the United Nations (FAO) in 1951. At present, a sample of 3,000 villages for the main season (Maha) and 2,000 villages for the second season (Yala) are selected to carry out experiments. (Division, n.d.). However, this method fails to predict rice yield before the harvest as the sampling is conducted during the harvest. As well as, this survey is usually time-consuming, subjective, costly and likely to significant errors because of the limited ground survey, resulting in poor rice yield estimations.

Paddy yield forecast a few months before the harvest can be of huge importance to take necessary management decision regarding seasonal crop productions. In addition, it is most important to strategic planning and decision making regarding the food security and facilitation of safe harvest storages. According to the Sri Lankan Department of Agriculture, in the 2015/2016 Maha and 2016 Yala seasons there were 2.5 and 1.3 million metric tons of rice produced, respectively. However, during these two seasons, it was recorded that farmers in many parts of the country waited days near the storage facilities to receive payment for their crop. In the end, they were disappointed over the failure to set up an adequate mechanism to purchase their harvest. On the other hand, in early 2017, the price of rice increased significantly due to the inadequate stores of rice within the country. These two recent incidents reflect the inadequate management and poor decision making of national rice production in Sri Lanka. The conditions of surplus in 2015/2016 and scarcity in 2017 were primarily due to the unavailability of a system to forecast the expected paddy yield by a considerable time before harvesting (T. L. Dammalag).

So, to overcome previously mentioned cases by proper import, export or rearrange storage facilities, taken based on reliable yield estimations as suggested by N.A. Noureldin (N.A. Noureldin, 2013). Recently, there are lunching various satellites like Landsat, Sentinel, Tera etc with particular remote sensors to crop monitoring. Whereas, remote sensing data have the potential to provide timely, systematic high quality spatial and accurate details about land feature including the environmental impact of the crop growth (Kogan, 2002). Using sentinel-2A satellite data, creating a machine learning model to identify the stage of the paddy crop and paddy yield prediction in the ripening phase in real-time.

## Motivation Facts

* Resourcefulness

In the early years, there was no opensource satellite to get the NDVI values and satellite images of AOI to researches to conduct this kind of research. But nowadays we have several opensource satellites and tools to archive the bands' details and analyse them. And also there are several kinds of research were conducted in the research area. Therefore, the availability of data has been a great motivation to increase the researches in the Machine Learning field.

* Helpfulness

In Sri Lanka, there is poor management of rice storage because there is no proper accurate mechanism to forecast the yield of rice before the harvest. This leads to directly national economy. Because paddy rice is both crucial for national food security and national development evaluation. To overcome the improper management of rice storage(export, import), this model will help.

* Popularity

Rice is the single most important crop occupying 34 percent (0.77 /million ha) of the total cultivated area in Sri Lanka (RRDI, n.d.). It is the largest cultivated crop in Sri Lanka

* Availability of Software and frameworks

Unlike the early days at present, there are lots of software and frameworks which support machine learning. It makes easier to focus on the research, more than worrying about how to implement it. MATLAB, OpenCV, Weka are good examples of software and Scikit-learn, Tensorflow, are good examples of frameworks which support machine learning.

## Objectives

* Identify the best bands combination of Sentinel-2A satellite images to paddy crop analysis and paddy crop classification(B11, B08, B04 / B11, B08, B02 / B04, B03, B02).
* Identify the cloud images from best bands combination.
* Identify the ripening phase of AOI.
* Analysis of the relationship between the linear or poly gradient of NDVI over the paddy season vs paddy yield of that paddy season(past data).

## Benefits of the model and web-platform

* When given the coordinates of some area can be classified the paddy stage.
* Can be classified the given area contain the cloud or not.
* Can get approximate yield prediction in AOI.
* Real-time processing with satellite images

# Literature Review

## Overview of the chapter

This chapter contains the knowledge gathered to conduct the research project. This Knowledge was gathered from studying theories of machine learning, studying existing methods for image classification, regression models and the ways for optimizing them and also how spectral bands affect the vegetation and clouds such as IR, red, blue, green, SWIR and different VI(vegetation indices) for creating a model to predict the harvest of paddy.

## Introduction

This research addresses this very important and urgent issue on how to predict the harvest through NDVI values in the third phase(ripening phase) of paddy rice using remote sensing images with support vector machines. The non-destructive nature and capabilities of large scale observation and on-site monitoring makes remote sensing technique being an ideal means used for implementation practices in rice production (Chwen-Min Yan, March 2007). (Sidik Mulyono M. I., 2012) proposed a rice yield prediction method based on land and airborne hyperspectral images. It involved laborious and high-cost campaigns both infield and airborne. It was aimed as a capacity building measure to anticipate the launching of the next generation of Hyperspectral spaceborne satellite (Hyperspectral Image Suite: HISUI). In the absence of hyperspectral satellite imagery, it is very difficult to extrapolate the application of our model and thus to extend our area of estimations. Whereas, the airborne campaign is quite expensive. Instead of waiting and relying on the availability of hyperspectral satellite images, we study the use of Sentinel-2A which is a kind of multispectral remote sensing satellite operated and developed by the ESA(European Space Agency). (Ferencz C., 2004) deduce with their research that Satellite Remote Sensing (RS) can be successfully applied for this particular research area. The most multispectral satellite system measure varies spectral bands including the visible to the mid-infrared region of the electromagnetic spectrum (J.K Shwetank, 2010). The main advantage of sentinel-2A is its direct broadcast capability, and anyone can constantly download this broadcast data free of charge. Moreover, its acquiring system almost covers all of the earth surfaces, and acquire every 4-5 days at the same place, so make it possible to apply for monitoring an object on the earth surface. It plays a vital role in broad applications such as the development of the prediction model for crop yields and classification of crops using multidimensional regression of NDVI and surface temperature (Paul C. Doraiswamy, 2007).

The analysis of Sentinel-2A data for paddy growth stages classification is not a trivial task due to many factors such as large spatial variability of each growth stage due to non-uniform plantation time, atmospheric effects and the curse of dimension in dealing with a larger number of frequency bands. (Tianxiang Zhang, 2017) develops a novel approach to analyze satellite remote sensing images, particularly Sentinel-2A satellite images using machine learning techniques. (Sidik Mulyono M. I.) address the problem of harvest time and area prediction through growth stages classification using MODIS remote sensing data with support vector machines.

Regarding with yield prediction model literature, (Rajitha.K) examines the inﬂuences of cirrus cloud on spectral indices of vegetation like NDVI and AFRI. The spectral absorption typically occurs from 670 to 780 nm wavelength range of the electromagnetic spectrum (P. Kempeneers, 2004). In this case, leaf chlorophyll has a strong absorption at 0.45 μm and 0.67 μm, and high reflectance at near-infrared (0.7–0.9 μm). Healthy plants have a high NDVI (normalized difference vegetation index) value because of their high reflectance of infrared light, and relatively low reflectance of the red spectrum (J Moore, 2003). The model is based on vegetation indices (VI) which could be collected from remote sensing satellite data. VI are optical measures of vegetation canopy ‘‘greenness’’. Some researches have observed that there is a linear relationship between rice yield and VI calculated based on satellite (Cheng Q., 2010) also shows the exponential relationship (WayanNuarsa I., 2011). They give a direct measure of photosynthetic potential resulting from the composite property of total leaf chlorophyll, leaf area, canopy cover, and structure. In this case, the NDVI value is closely related to crop yield. It has a direct correlation with LAI, biomass and vegetation cover (Xiao), (Nemani, 1989). Most researches have observed that correlation between NDVI and green biomass yield, therefore NDVI can be used to estimate yield before harvesting (M.S., 1997). A research was conducted in Indonesia using Landsat, found that the best age of a paddy yield prediction is 63 days after transplanting (WayanNuarsa I., 2011). Another research in Egypt concludes that the best age of paddy yield prediction is 90 days after transplanting (N.A. Noureldin, 2013). Another research also in Sri Lanka show the best age of yield prediction is 80 days after transplanting (T. L. Dammalag).

## Image classification methods

There are several ways for image classification. Most of the classifiers are a maximum likelihood, minimum distance, decision tree, support vector machine and neural network. They are making a conclusive decision about the land cover class and require a training sample. In contrast, the clustering-based algorithm, such as K-NN, K-mean or ISODATA, are an unsupervised classifier, and fuzzy-set classifier is soft classification providing more information and potentially a more accurate result.

For identifying the stage of paddy crop, in this research, we use an image classification algorithm. Here represent some of the related work regarding RS image classification. Mainly for this research, we literature reviewed on SVM(Support Vector Machine) and CNN(Convolutional Neural Network) approach for multispectral image classification on multi-temporal data.

### SVM for RS Image Classification

#### Overview

SVMs are a recently new supervised classification technique for the land covering and mapping community. They have their fundamental in Statistical Learning Theory and have acquired prominence because they are robust, accurate and are effective even when using a small training dataset. Over the last two decades or so, RS has increasingly become a prime source of land cover information. As the name suggests, SVMs are essentially binary classifiers, however, they can be improved to handle the multiple classification tasks in remote sensing studies.

Support Vector Machines have their roots in Statistical Learning Theory (N, 1995). They have been broadly applied to machine vision fields to recognition things such as character, handwriting digit and text. (Joachims, 1998; N, 1995), and recently satellite image classification problems (Huang C., 2002). SVMs, like ANN (Artificial Neural Networks) and other nonparametric classifiers, have a reputation for being vigorous (Foody M. G., A Relative Evaluation of Multiclass Image Classification by Support Vector Machines, 2004a; Foody M. G., Toward Intelligent Training of Supervised Image Classifications: Directing Training Data Acquisition for SVM Classification, 2004b). SVMs function by nonlinearly projecting the training dataset in the input space to a feature space of higher dimension by use of a kernel function. This results in a linearly separable dataset which can be separated by a linear classifier. This process enables us to the classification of RS image datasets which are usually nonlinearly separable in the input space. In many cases, classification in high dimension feature spaces results can be over-fitting in the input space. By the way, in SVMs over-fitting is controlled by the principle called “structural risk minimization” (N, 1995). The heuristic risk of misclassification is reduced by maximizing the margin between the data points and the decision boundary (D, 2004 November). In practice, the tradeoff between these factors is managed through a margin of error parameter(usually called C parameter) that is tuned through cross-validation methods (D, 2004 November). The functions used to project the dataset from input space to feature space are sometimes called kernels machines (or kernel), examples of which include polynomial, quadratic functions and Gaussian (more commonly referred to as radial basis functions). Each function has unique parameters which have to be determined according to classification and also they are usually determined through a cross-validation process. A deeper mathematical thesis of SVMs can be found in Vapnik (1995), Campbell (2000) and Christianini (2002). SVM is basically designed for binary classification (Melgani F., 2004). But there exist strategies using that can be adopted to multiclass tasks associated with remote sensing studies.

#### SVM Multiclass Strategies

SVM classification is mainly a binary classification technique, which needs to be modified to handle the multiclass tasks in real-world problems such as getting the land cover information from satellite images. There are two of the common approaches to enable this adaptation which are the 1A1 and 1AA techniques. The 1AA technique represents the earliest and rapid SVM multiclass approach (Melgani F., 2004) and involves the division of an N class dataset into N two-class cases. Means that this will train one classifier per class in total N classifiers. For class ith,  it will assume ith labels as positive and the rest as negative. Say that the classes of interest in a satellite image include vegetation, water and built-up areas, the classification would be effected by classifying vegetation against non-vegetation areas i.e. (water and built-up areas) or water against non-water areas i.e. (vegetation and built-up areas). The 1A1 technique involves constructing a machine for each pair of classes resulting in N(N-1)/2 machines. Means that 1A1 have to train a separate classifier for each different pair of labels. This leads to N(N-1)/2 classifiers. From ML theory, it is known that the disadvantage the 1AA approach has over 1A1 is that its performance can be compromised due to unbalanced training datasets (Gualtieri J. A., 1998). However, 1A1 is much less sensitive to the problems of unbalanced datasets but is much more computationally expensive. Some researches concluded that whereas one can be certain of high classification results with the 1A1 approach, the 1AA yields approximately as good classification accuracies (Gidudu Anthony).

#### The mathematical background behind the SVM

SVM is a two-dimensional description of the optimal surface evolved from the linearly separable case, the basic idea can be used in figure 1. Figure 1 shows the basic idea. Two types are separated by H without errors. H1 and H2 are places which pass the recent point of H. The distance of H1 and H2 was called a class interval. Optimal separating surface is not only to ensure error-free separation of the two types of samples but also called the largest class interval.

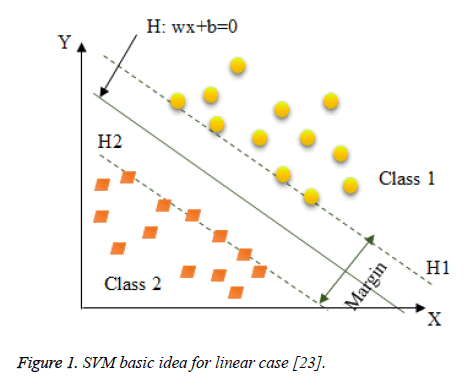


Figure 1: Optimal separating surface

In pattern recognition, the linear discriminate function in the n-dimensional space:

The classification hyperplane equation can be written . Linear separable case, the discriminate function was normalized. So that all training samples are met , even to meet away from the classification of the surface of the sample , so the class interval equivalent to , thus making the interval on the equivalent to or . Make a classification of the surface of all samples correctly classified, it is necessary to meet:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Satisfy the above equation (1) and make the smallest classification surface is the optimal classification surface. A point on the hyperplane are called support vectors, they support the optimal classification surface (Deng N., 2009).

##### Linearly Separable

According to the above, the optimal separating surface can be expressed as the following constrained optimization problem, Seeking the minimum of the following the function:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

We can define the Lagrange function as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Where is a coefficient of Lagrange. Problem is seeking the minimum of and to the Lagrange fn. and seek partial differential and make them equal to zero, the original problem can be transformed into the following simple dual problem:

the constraints

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

|  |  |  |
| --- | --- | --- |
|  | ,n | (5) |

The maximum value of solving the following function about :

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

If is the Optimal solution,

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

The optimal weight coefficient vector of the surface is a linear combination of the training sample vector. According to the Kuhn-Tucker conditions, the optimal solution must also meet the following conditions:

|  |  |  |
| --- | --- | --- |
|  |  |  |

For most samples will be zero, the value of zero correspondings to the type with equality (1) are support vector, they are usually all the samples part. To solve the above problems, the optimal Category function is

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

sgn() is the sign function.

##### Linear Inseparable Problem

The main idea of the linearly inseparable problem, the support vector machine is the input vector is mapped to a high-dimensional feature vector space and constructs the optimal separating surface in the feature space. When the training set of linear nontime-sharing, some training samples can not satisfy condition (1). Conditions can be amended to add a slack variable .

The constraints are:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

When misclassification occurs, , is training set the upper bound of the number of misclassified. It can be used as a description of the degree of the training set is misclassified.

This makes it two goals: is large as possible. is as small as possible.

These two goals into one goal, the introduction of the penalty parameter C as the right combination of these two target weight, under the constraints of the conditions (9) and the minimum of the following functions:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

Where: C is the penalty parameter. It actually plays a role in the control of the degree of right or wrong sub-sample of punishment to achieve a compromise between the proportion of misclassified samples and the complexity of the algorithm. Its optimal solution to the saddle point of Lagrange function is defined as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

|  |  |  |
| --- | --- | --- |
|  |  |  |

Where and is all Lagrange multiplier vector. The specific method for solving Linear separable is very similar, but the condition is changed (N., 2011; Zhang Y., 2011; Zhang, 2012)

### Grid Search Method | meka danna methods walata

Grid search approach is an alternative for finding the best model parameter to predict unlabelled data accurately (M. Ataei, 2004). The method is known as a systematic method for the best parameter values, each of which should be tested by setting some prediction values. The Method then displays the score to be chosen for each parameter value. This approach is applicable when the amount needed for each of the separate variables is known to be within the limits specified by upper or lower limits (M. Ataei, 2004). The Hsu, Chang and Lin (C.W. Hsu, 2003) search method shows that the best value in two parameters, which are, in this case, c and y, is the grid search method and the cross-validation. The research investigates different pairs of (C, y) and is selected for the most detailed cross-validation. The experiments show two explanations of why the method for grid search has been chosen

1. The likelihood of feeling unsafe by using methods to avoid a thorough search of approximation parameters.
2. In order to find good Grid Search parameters, the computational time required by advanced methods is not much greater because only two parameters are usable.

### Deep learning theory

(Hinton G., 2006) specifically proposed the principle of deep learning in 2006. Deep learning's basic purpose is to set up a deep neural network to replicate the human brain's learning and interpretation process. Compared to traditional machine learning theories, the most significant difference in deep learning is the focus on supervised learning from the large data set by multi-layer neuron organization. Several deep learning architectures such as Deep Belief Networks (DBN) (Hinton, Osindero, & Teh, 2006), Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) (. Graves, et al., 2009) have been applied in recent years to fields such as computer vision (Ciresan, Meier, & Schmidhuber, 2012), speech recognition, natural language processing, audio recognition, and bioinformatics, and have been shown to deliver state-of-the-art results in these fields.

### CNN for RS image classification

CNN has obtained remarkable results in the classification of images, recognition and other vision tasks in deep learning techniques, and has the highest score in many visual databases such as ImageNet, Pattern Analysis, Statistical Modeling and Computational Learning Visual Microsoft Popular Objects in Context (MS-COCO) and Object Classes (PASCAL VOC). The basic structure of the standard CNN for image classification is stacks of "convolutional-pooling" layers as multi-scale extractors of features and subsequent numbers of fully connected layers as classifiers. A lot of research has appeared in recent years on CNN-based remote sensing image analysis. (Nguyen, 2013) proposed a satellite image classification method using a network of five layers and achieved a classification accuracy of more than 75%. For long-term route planning, (Wang, 2015) used a three-layer CNN structure and Finite State Machine (FSM) for road network extraction. (Castelluccio, Poggi, Sansone, & Verdoliva, 2015) explored the use of CNNs to identify remote sensing scenes semantically. (Hu, Xia, Hu, & Zhang, 2015) used a pre-trained CNN model to classify various scenes from high-resolution remote sensing imagery. (Zhou, Newsam, Li, & Shao, 2016) used CNN architecture as a high-resolution remote sensing image recovery (HRRSIR) deep-function extractor. A CNN-based architecture was suggested by (Mnih, 2013) to learn contextual features for aerial picture labelling on a large scale. The model creates a complex classification patch instead of outputting a single value image group. (Lagkvist, Kiselev, Alirezaie, & Loutﬁ, 2016) proposed a novel method of classification of remote sensing imaging based on CNN's for five groups (vegetation, soil, path, building and water), outperforming existing classification approaches. In addition to the CNN family approach, (Yuan, Lin, & Wang, 2016) used the Stacked AutoEncoder classifier for a classification experiment using the Manifold Ranking based Salient Band selection.

### CNN

CNN is considered to be one of the best techniques for studying image information and have shown state-of-the-art findings on image recognition, segmentation, identification and retrieval related tasks (Ciresan, Meier, & Schmidhuber, 2012). In industry, companies such as AT&T, Google, Microsoft, Facebook and NEC have developed active research groups to explore new CNN architectures (L. Deng, 2013). At present, most of the frontrunners of image processing competitions use deep CNN-based models.

Many improvements have been made to CNN's learning approach and infrastructure to make CNN applicable to large and complex problems. Such inventions can be defined as refining conditions, regularizing, reformulating systems, etc. Nevertheless, after AlexNet's outstanding success on the ImageNet dataset (A. Krizhevsky, 2012), it is noted that CNN-based applications were prevalent. As a result, major developments have been proposed in CNN since 2012 and were mainly due to processing unit redesign and new block design. Furthermore, Zeiler and Fergus (Fergus, 2013) have introduced the concept of a layer-wise representation of functionality that has changed the trend towards the abstraction of features at low spatial resolution in deep architectures such as VGG (Zisserman, 2015). On the other hand, the Google group has introduced an interesting idea of splitting, transforming, and merging, and the corresponding block is known as the inception block. For the very first time, the invention block provided the idea of branching within a framework, which enables features to be abstracted from various spatial scales (K. He, 2015). The concept of skip connections introduced by ResNet for the training of deep CNNs became famous in 2015 and was subsequently used by most of the following networks, such as Inception-ResNet, WideResNet, ResNext, etc. (Komodaki, 2015).

### Basic CNN Components

Today, CNN is considered to be the most widely used ML technique, particularly in vision-related applications. CNN recently demonstrated state-of-the-art results in various ML applications. Because of CNN possesses both a successful extraction capability and a great ability to discriminate, it is therefore used in the ML system; it is mostly used for extraction and classification of features.

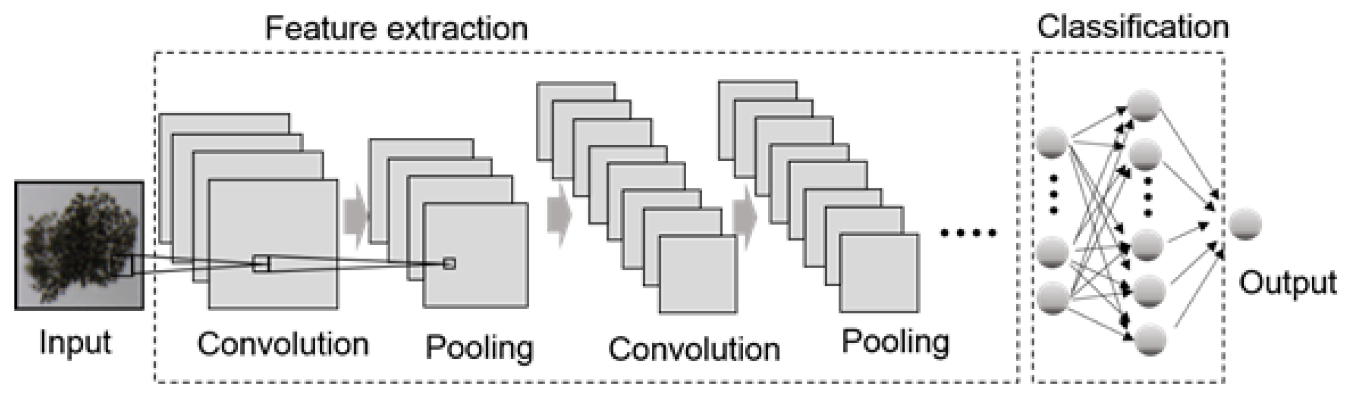


Figure 2: An example of the design of a CNN.

Figure 2 illustrates the CNN architecture generally consists of alternating layers of convolution and pooling accompanied by one or more fully connected layers at the top. Sometimes Fully connected layer is replaced by a global average pooling layer. In addition to the several learning stages, the CNN output optimisation is also assisted by various regulatory units such as batch normalisation and dropout (Bouvrie, 2006). The configuration of CNN components plays a key role in designing new architectures and thus achieving enhanced performance. This section discusses briefly the role of these components in CNN architecture.

#### Convolutional Layer

The convolutional layer consists of a collection of kernels of convolution (each neuron functions as a kernel). Such kernels are bound to a small area of the image called a field of reception. It operates by splitting the image into small blocks (receptive fields) and combining them with a specific set of weights (multiplying filter elements with corresponding receptive field elements) (Bouvrie, 2006). The process of Convolution will convey the following:

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

Where the input image is represented by , x, y displays the spatial localization, and represents the th convolutionary kernel of the Kth row. The division of the image into small blocks helps to extract locally correlated pixel values. Also known as the feature motif is this locally aggregated information. A different set of image features are extracted with the same set of weights by sliding the convolution kernel over the entire image. This convolution operation 7 weight sharing feature makes the CNN parameter more efficient than fully connected networks. Convolution operation can also be categorized into different types based on filter type and size, padding type, and convolution direction (Bouvrie, 2006). Furthermore, if the kernel is symmetric, it becomes a correlation operation (Ian Goodfellow, 2015).

#### Pooling Layer

Feature motifs that can occur at different locations in the image as the production of a convolution cycle. Once characteristics are extracted, their exact location becomes less important as long as their approximate position is preserved relative to others. It is an interesting local process to pool or downsample like convolution. It summarizes similar information in the receptive field neighbourhood and generates the dominant answer in this local area (Bouvrie, 2006).

|  |  |  |
| --- | --- | --- |
|  |  | (12) |

Equation (12) indicates the pooling operation in which represents the th output feature map, shows the th input feature map, whereas determines the form of the pooling operation. Using the pooling process helps generate a combination of features invariant to translation changes and slight distortions (D. Scherer, 2010). Reduced functional map size to invariant feature set governs not only network complexity but also allows generalization by overfitting reduction. CNN (K. He, Spatial pyramid pooling in deep convolutional networks for visual recognition, 2015) uses various types of pooling methods, such as max, average, L2, overlap, spatial pyramid pooling etc.

#### Activation Function

Activation mechanism acts as a tool for decision making and helps to understand a complex pattern. It can speed up the learning process by choosing an appropriate activation function. The activation function is defined in equation (13) for a converted feature map.

|  |  |  |
| --- | --- | --- |
|  |  | (13) |

In the above equation, is an output of a convolution operation assigned to the activation function; adding non-linearity and returning a transformed output for th layer. Numerous activation functions such as sigmoid, tanh, maxout, ReLU, and variants of ReLU such as leaky ReLU, ELU, and PReLU (al, 2018; B. Xu, 2015; LeCun, 2007) are used in literature to inculcate nonlinear combination of features. ReLU and its variants are however preferred over other activations as it helps to overcome the problem of the vanishing gradient (Hochreiter, 1998).

#### Batch Normalization

Batch normalization is used within feature maps to address issues related to internal covariance change. The internal covariance shift is a change in the distribution of values of hidden units that delays the convergence (forcing the learning rate to a low value) and requires careful parameter initialization. Batch normalization for a transformed feature map is shown in equation (14).

|  |  |  |
| --- | --- | --- |
|  |  | (14) |

In equation (14), represents a standardized feature map, are the input feature map, and mean and variance of a feature map respectively for a mini-batch. Batch normalization unifies the map value distribution by taking it to zero mean and unit variance (Bouvrie, 2006). It also smoothes the gradient flow and serves as a controlling factor, thereby helping to improve the network's generalization.

#### Dropout

Dropout implements network regularisation, which eventually increases generalisation by automatically skipping a certain likelihood of some units or connections. Multiple connections in NNs that know a non-linear relationship are often co-adapted, leading to overfitting (G. E. Hinton, 2012). A random drop in some links or units creates multiple thin network architectures, and finally, one representative network with small weights is chosen. This selected architecture is then considered to be an approximation of all the networks proposed (N. Srivastava, 2014).

#### Fully Connected Layer

A fully connected layer is typically used for classification purposes at the end of the network. It is a global operation, unlike pooling and convolution. It takes data from the previous layer and the output of all the previous layers is evaluated globally (K. He, Going deeper with convolutions, 2015). This allows a non-linear combination of selected features used to classify data (Wang W. R., 2016).