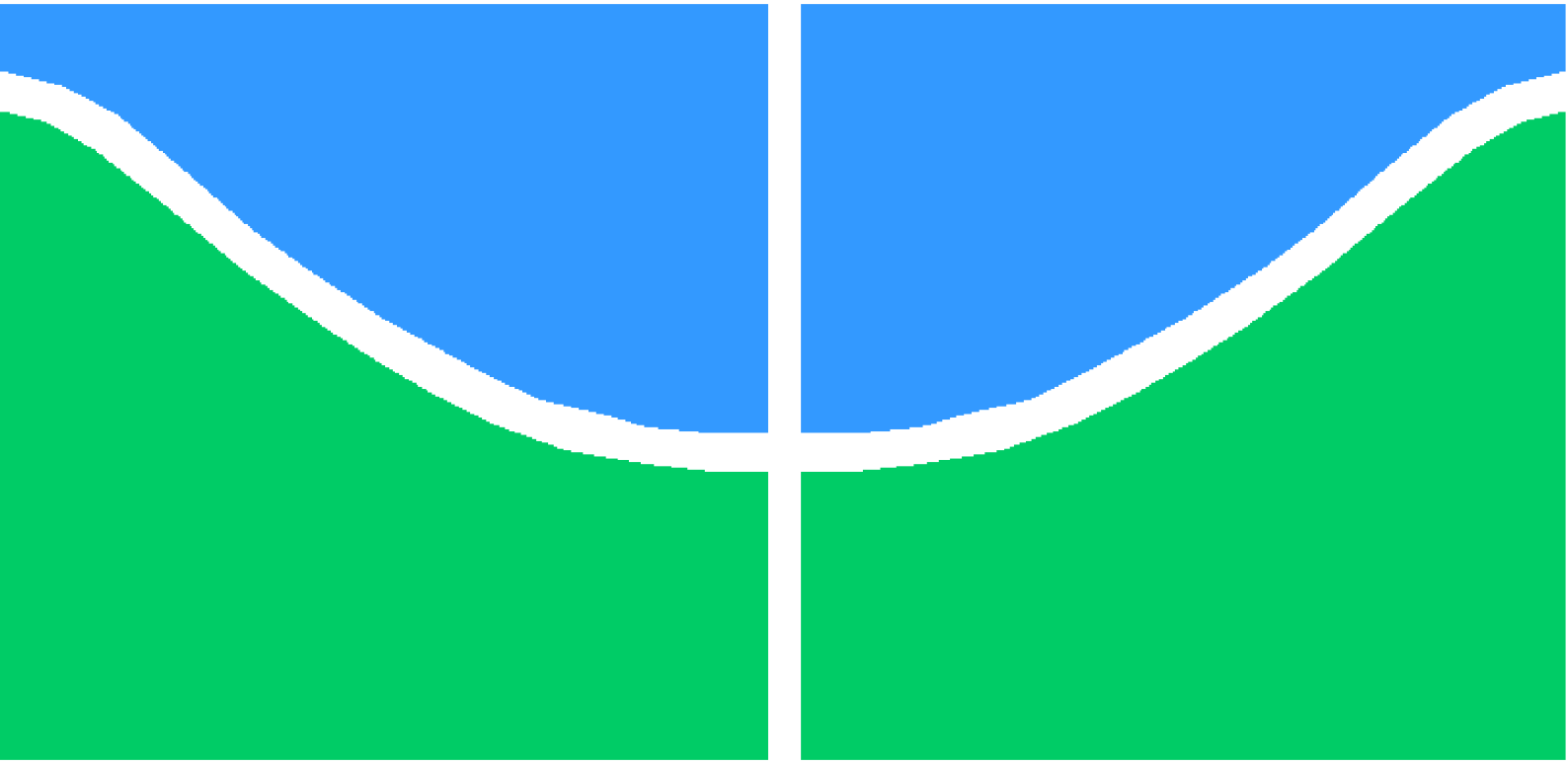


Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Aerospace Engineering

Convolutional Neural Network and Remote Sensing applied to sugarcane biomass prediction

Author: Arthur Borges Bringel Machado

Advisors: Dr. William Reis

Co-advisor: Dr. Vinicius Bufon

Co-advisor: Dr. Luciano Shiratsuchi

Brasilia, DF

2023

Arthur Borges Bringel Machado

Convolutional Neural Network and Remote Sensing applied to sugarcane biomass prediction

Undergraduate thesis submitted to the undergraduate course in Aerospace Engineering of the Universidade de Brasilia, as partial requirement for obtaining the Bachelor's Degree in Aerospace Engineering.

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Advisor: Dr. William Reis

Co-advisor: Dr. Vinicius Bufon

Co-advisor: Dr. Luciano Shiratsuchi

Brasilia, DF

2023

Arthur Borges Bringel Machado

Convolutional Neural Network and Remote Sensing applied to sugarcane biomass prediction

Undergraduate thesis submitted to the undergraduate course in Aerospace Engineering of the Universidade de Brasilia, as partial requirement for obtaining the Bachelor's Degree in Aerospace Engineering.

Thesis approved. Brasilia, DF, 27 de julho de 2023:

Dr. William Reis
Advisor

Dr. Vinicius Bufon
Co-advisor

Dr. Luciano Shiratsuchi
Co-advisor

Dr. Ronne Toledo
Guest 1

Brasilia, DF
2023

Abstract

The Brazilian sugarcane industry is investing in research to develop new technologies to improve crop management and sustainability, trying simultaneously to reduce cost and improve profitability. This work aims to assess the capacity of the model to predict average field biomass, as well as to assess within-field biomass spatial variability, by coupling satellite imagery (Sentinel-2) biomass estimates with machine learning techniques, specifically using Convolutional Neural Networks. The model was assessed in both goals, field-as-a-whole- and whithin-field-level- experiments. The results showed that the model predicted average field biomass with a mean absolute percentage error (MAPE) of 3.95%, average error (AE) of 0.30 tonne/ha, root-mean-square error (RMSE) of 3.48 tonne/ha, and mean absolute error (MAE) of 3.07 tonne/ha. Nevertheless, it performed low capacity to predict a biomass map based on its Gaussian normal distribution and scatter plot. The good performance of the maps' error compared to the reference maps are indicators that further studies can increase the model capacity of predicting biomass maps.

Key-words: Convolutional Neural Network, Sentinel-2, sugarcane biomass.

Resumo

A indústria de cana-de-açúcar no Brasil está investindo em pesquisas para desenvolver novas tecnologias para melhorar o manejo e a sustentabilidade da cultura, tentando simultaneamente reduzir custos de produção e aumentar sua rentabilidade. Este trabalho tem como objetivo avaliar a capacidade do modelo de prever a biomassa média do campo, bem como avaliar a variabilidade espacial da biomassa dentro do campo, acoplando estimativas de biomassa de imagens de satélite (Sentinel-2) com técnicas de aprendizado de máquina, especificamente usando Redes Neurais Convolucionais. O modelo foi avaliado em ambos os objetivos, experiências no campo como um todo e no nível do campo. Os resultados mostraram que o modelo previu a biomassa média do campo com um erro percentual absoluto médio (MAPE) de 3,95%, erro médio (AE) de 0,30 toneladas/ha, erro quadrático médio (RMSE) de 3,48 toneladas/ha e erro absoluto médio (MAE) de 3,07 toneladas/ha. No entanto, ele performou com uma baixa capacidade de previsão de mapa de biomassa com base na sua distribuição normal gaussiana e no seu gráfico de dispersão. O bom desempenho do erro dos mapas em comparação com os mapas de referência são indicadores de que estudos futuros podem aumentar a capacidade do modelo de prever mapas de biomassa.

Palavras-chaves: Rede Neural Convolucional, Sentinel-2, biomassa de cana-de-açúcar.

List of Figures

List of Tables

Contents

1 Introduction

Agriculture plays an important role in the Brazilian economy, contributing more than a quarter of the country's gross domestic product (GDP) in 2020, equivalent to almost R$2 trillion (CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL AND CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA, 2021). One of the main contributors to the GVP of agricultural production was sugarcane, accounting for over 8% (R$ 69.72 billion) in 2021 (CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL, 2021). Besides the production of sugar and biopolymers, sugarcane biofuel (ethanol) and bioelectricity contributes to Brazil having one of the cleanest energy production systems in the world. In 2020, around 48% of the energy matrix and 84% of the electricity in Brazil were renewable, compared to 12 and 31%, respectively, in the OCDE countries (MATOS, 2022).

In the last decades, sugarcane production in Brazil has shifted from the Atlantic Forest to the Cerrado biome, a tropical Savannah characterized with a higher water deficit challenge. In addition, climate changes have increased severe drought risk, shortening the rainy season and increasing frequency and intensity of dry spells within the rainy season, negatively impacting the efficiency of crop management practices, sugarcane biomass yield, profitability, and the sugarcane industry sustainability (ZEBALLOS et al., 2022).

Brazilian sugarcane industry is investing in research to develop new technologies to improve crop management and sustainability and to cope with climate challenges. Assessing the actual fields' biomass production, and the within-field spatial variability is key issue to improve cropping efficiency and sustainability.

Nowadays, sugarcane field biomass estimates are done by field observation by experienced agronomists, supported by satellite imagery biomass estimates, field sampling estimates, and fields historical data. Besides achieved absolute error within 3% (comunicação pessoal - Usina Alta Mogiana) of mean field biomass, it is an empiric method, largely dependent on the experience of the team, and highly demanding on time and financial resources. In addition, this method does not allow assessing the yield spatial variability within each field.

Alternatively, sugarcane yield monitors have been developed by the sugarcane harvesters industry. Yield monitors can estimate yield spatial variability, as well as, the mean biomass of each field, and are desirable and necessary tools to improve crop efficiency and spatial variability of the fields. However, as the data is assessed only when sugarcane is being harvested, not allowing within-season management adjustments, only being applied in the next growing season. In addition, due to the costs of yield monitors technology does not allow its adoption at the desired pace. Recent research has shown the potential of coupling satellite imagery biomass index with machine learning techniques to improve

field biomass estimates, as well as to assess within-field biomass spatial variability, early in the season. The most used machine learning technique to predict biomass integrating remote sensing is the Random Forest, although it only considers pixel information, and not its spatial pattern. However, Convolutional Neural Network applied to satellite imagery has the intrinsic characteristic of recognizing spatial patterns, with potential to improve sugarcane biomass prediction.

1.1 General objective

To analyze whether convolutional neural network and remote sensing can predict sugarcane biomass.

1.2 Specific objective

- To assess the model prediction capacity considering average field biomass scale.
- To assess the model prediction capacity considering the within-field biomass spatial variability.

1.3 Document structure

This document is divided in seven sections, with the introduction being the first one. The sections two and three describe the theoretical framework. In section two, a brief summary of the physiological basis of plant leaf is related with the nature of the light spectrum, which forms the core subject of remote sensing applied to agriculture. The following themes of this section present a structured knowledge of a remote sensing platform (satellite) and a type of sensor (multispectral), where some basic concepts are defined to let the reader understand their different aspects and their interrelationships. Afterward, vegetation indices are shortly presented, and a summary of Sentinel-2 and its MultiSpectral Instrument (MSI) is described, focusing on the remote sensing approach.

The fourth section is focused on the neural network technique, where the reader is presented to the artificial neural network (ANN) and to the convolutional neural network (CNN). In this section, the perceptron algorithm is described, as well as the gradient descent. The basic concepts of an ANN, such as the activation function and the loss function, are explained along with their impact and importance in the model. After, the CNN is explained and the concepts of the convolutional operator and the pooling process are introduced.

In the fifth section, a brief description of the methodology describes the main process that is used in the next phase of this project. The first section describes the study site and its climatological overview. After, it describes in general terms the processes that

will occur with the input data type that will train the convolutional neural network model. At least, the process timeline is shown in section five.

In section five, the results were displayed in terms of the model capacity to predict the average biomass of the tested field and to generate a predicted map. MAPE, AE, RMSE and MAE of the tested field were presented, furthermore, a Gaussian normal distribution plot, AE maps and a scatter plot of the predicted and reference values. These results were discussed in section six and conclusions about this study were made in section seven.

2 Remote sensing applied to agriculture

Remote sensing was first used by fixing cameras and sensors on kites, balloons, aircraft and even birds. The term by itself already denotes that it is related to some data acquisition without a direct contact with the targeted object. Nowadays, the applications of remote sensing are mostly related to satellite constellations and UAVs (BERNARDI, 2014), and this is given the rise of CubeSats and drones that turned costs to collect data more accessible. The conventional approach of modern agriculture, originated after the mechanization of land use, is to treat fields as a uniform area, which causes a lot of waste and makes an unsustainable use of the land. Otherwise, with the use of remote sensing, agriculture turns to be more specific and precise on its management processes, representing an improvement and a professionalization on its business model. The main idea behind this new approach is to treat small areas of the field as a separate management unit, but its affordability could only be worth it after the development of the mechanization capability of applying fertilizers and pesticides on a variable rate and the right tools to keep up with the spatial and temporal variability of the field. Spatial and temporal variability are key to remote sensing applications, as these are the main constraints of a project, specially when dealing with satellite imagery. Usually they will be called spatial and temporal resolution. Additionally, it is important to notice the sensor's limitations that will imply on the project's constraints, and these are the sensor's radiometric resolution and the sensor's spectral resolution. These four concepts will be better discussed along this chapter.

The idea behind the unitary treatment of the fields involves concepts that go all the way back to a single plant phenology and the nature of light, two important things to understand the reason why sensors see what they see and its relationship to the organism development. Besides any other plant process, photosynthesis is central to keep up not only a plant but the world's dynamics, sustaining the majority of lives on Earth, except some types of bacteria (JENSEN, 2014). Photosynthesis has been responsible to bringing more than 90 percent of the energy required to power automobiles, airplanes, ships, factories and electronic devices, since it has brought the energy of the Sun to plants

that were transformed into fossil fuels (coal and oil) by geological forces (JENSEN, 2014). It is the mechanism by which plants convert light energy into chemical energy (glucose) with the aid of a substance called chlorophyll (JENSEN, 2014). This substance is stored within the chloroplast, an organelle (small part of the cell surrounded by a membrane and with a specific function) of the plant cell, and it is also called one of the plant pigments (JENSEN, 2014). Pigments are molecules of a plant cell that were evolved to absorb specific wavelengths of light when illuminated (JENSEN, 2014). Other organelles in a plant cell contains other pigments, such as carotenes, xenthophyll, phycoerithrin, phycocyanin, anthrocyanin. Figure 1 illustrates the internal structures of a leaf, the mesophyll region, and one can see that either within the palisade parenchyma cells or within the spongy parenchyma cells the chloroplast organelle is present as well as its characteristic pigment.

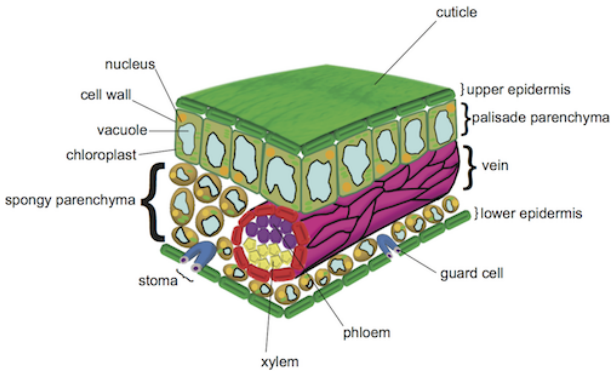


Figure 1: Structure of the leaf mesophyll.

A plant leaf has green coloration given the amount of chlorophyll inside it. Thus, the analysis of a plant's spectral behavior through sensors can indicate the abundance of this pigment in a sample collection and provide a way to assess the plant's development and growing potential. The wavelengths through which a healthy leaf absorbs the most light energy are the range between 0.43 – 0.45 μm and 0.65 – 0.66 μm, which includes wavelengths in the blue and red regions (as shown in Figure 2). This means that green and infrared (IR) wavelengths have a higher influence on the reflectance visible spectrum, because the incident sunlight that reaches the leaf is partially reflected, absorbed, and transmitted through it (JENSEN, 2014). The reflectance and transmittance curves behave as a mirror of each other (JENSEN, 2014), as can be seen by the Big Bluestem grass leaf absorption, reflectance, and transmittance in Figure 3. Moreover, a healthy plant green leaf's near-infrared (NIR) reflectance varies around 40 - 60 percent of its incident energy, while its transmittance remains between 45 - 50 percent (JENSEN, 2014). However, as these values are based on a single leaf, the plant's total reflectance may be higher due to the effect of multiple leaf layers. The NIR reflectance of the top leaf layer will be added to the reflected amount of the leaves below it, creating a phenomenon known as leaf additive reflectance (JENSEN, 2014).

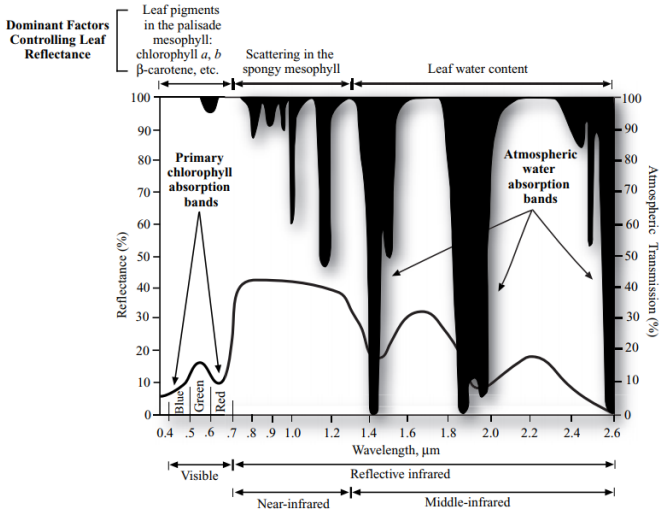


Figure 2: Chlorophyll absorption spectrum.

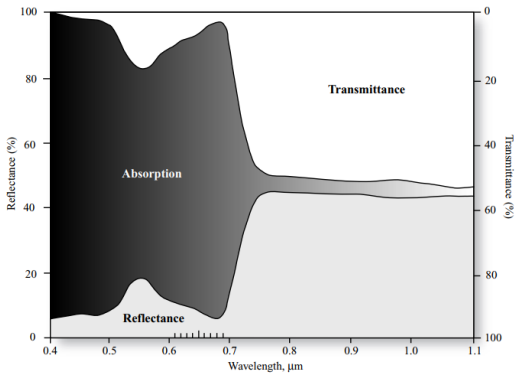


Figure 3: Absorption, reflectance and transmittance of Big Bluestem grass leaf.

2.1 Remote sensing platforms

According to (JAFARBIGLU; POURREZA, 2022), the platforms in which remote sensing is based are defined as vehicles carrying sensing devices to measure a target object data without interfering in its ambient. Satellites and aircraft are both common platforms, and each have its specific advantages and disadvantages. Usually, they will have either a good spatial coverage or a good spatial resolution, and the best choice to implement would depend on the target object characteristics. In agriculture, it is important to have in mind the goals of the application as well as the fields' constraints in order to manage data acquisition. Agricultural use of remote sensing aims to obtain mainly vegetation indices, requiring multispectral sensors mounted on the platforms, while the target field size turns out to be a relevant variable that can influence on the decision of which type of platform to use (JAFARBIGLU; POURREZA, 2022). For example, in large fields, it is necessary to have a spatial coverage with its matching proportion, so an unmanned aerial vehicle (UAV) could not reach its demands, otherwise, the contrary is also true. For this reason, it is important to first define the platforms' parameters, and the Figure 4 shows two characteristics that usually are associated with remote sensing applied to agriculture: altitude of the platform and its endurance.

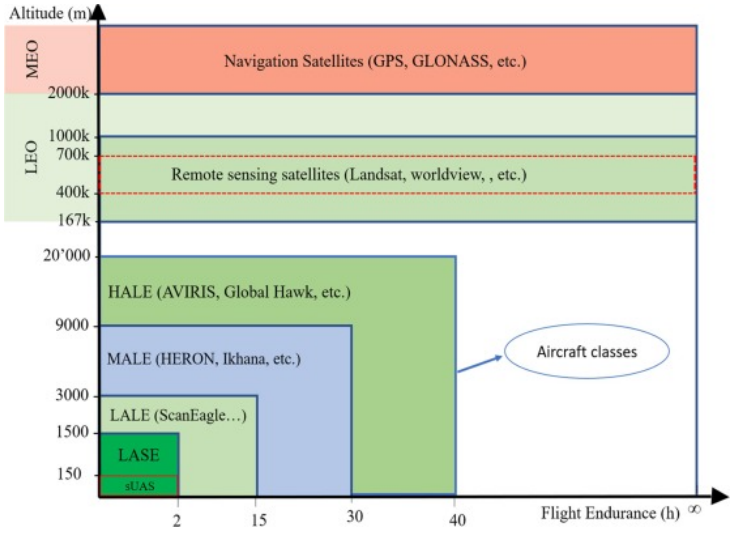


Figure 4: Remote sensing platforms' classification by altitude and autonomy.

Satellite platforms have longer flight endurance than unmanned aerial systems (UAS), but this characteristic has not so much importance for a common consumer of orbital products, since this information's only purpose is to indicate how often the platform should be set up and activated (JAFARBIGLU; POURREZA, 2022). For a comparison between drones on the other side, flight endurance is a real distinctive quality that can assure a mission long-term sustainability, given a maximum flight speed and a field size (JAFARBIGLU; POURREZA, 2022). Although flight endurance may not be so useful when it comes to satellites, there is another important temporal factor that applies specifically to it: temporal resolution.

2.1.1 Temporal resolution

Temporal resolution is a crucial aspect of remote sensing technology, particularly when applied to agriculture. It refers to the frequency at which data is acquired and how quickly it can be refreshed over time (JAFARBIGLU; POURREZA, 2022). In agriculture, the appropriate temporal resolution depends on the crop type, its growth cycle the nature of the problem (JAFARBIGLU; POURREZA, 2022).

Sugarcane and soybean are two significant crops in the agricultural industry. They both require high temporal resolution monitoring for efficient management. For instance, sugarcane has a growth cycle of approximately 12 months (MALL et al., 2016), and soybean takes approximately three to four months to reach maturity (CARNEIRO et al., 2020). As such, high temporal resolution monitoring is required for effective management of these crops.

For a sugarcane crop in India, the recommended temporal resolution would be two weeks or less if the phenological stage was to be assessed. It is given the phenological stages size that don't take shorter than two weeks to transition to the next one (MALL et al.,

2016). With this temporal resolution, it is possible to monitor the crop development, detect any abnormalities, and adjust the management practices accordingly. For a soybean crop in Brazil, a temporal resolution of approximately 10 days is sufficient for the phenological stage assessment. It is given its phenological stages transition size with less than 10 days (CARNEIRO et al., 2020). This resolution allows for monitoring of crop growth, detecting potential disease outbreaks and pests, and managing the application of fertilizers and irrigation.

The appropriate temporal resolution in remote sensing technology is crucial in providing timely and accurate data for effective crop management. The required temporal resolution varies among crops and their growth cycles (JAFARBIGLU; POURREZA, 2022). Understanding the temporal resolution requirements is essential for the successful application of remote sensing in agriculture (JAFARBIGLU; POURREZA, 2022).

2.1.2 Spatial resolution

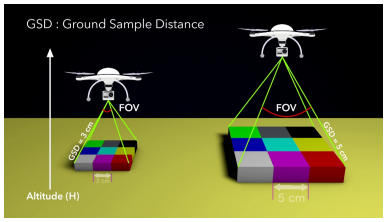
There is a spatial characteristic that satellites and drones share, and it is known as ground sample distance (GSD) or spatial resolution. It is a measure of the size of a pixel on the ground in a remote sensing image, and can be defined as the distance on the ground between the centers of two adjacent pixels in an image (ZHANG; MOORE, 2015). GSD is affected by the altitude of the sensor platform, the lens system, and the size of the sensor. The smaller the GSD, the higher the spatial resolution of the image.

The GSD is an important parameter because it determines the level of detail that can be resolved in an image. A smaller GSD means that smaller features can be seen in the image, and more detailed information can be extracted from it. For example, a GSD of 1 meter means that each pixel in the image represents an area of 1 square meter on the ground. A GSD of 0.5 meters means that each pixel in the image represents an area of 0.25 square meters on the ground, which is four times more detailed.

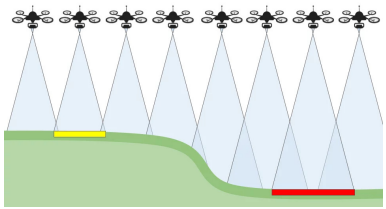
The variation of GSD within a mission is affected by changes in the altitude of the platform and the elevation of the terrain. As the platform changes altitude, the GSD will change accordingly Figure 5a. Similarly, if the platform remains at the same altitude but the terrain elevation changes, the GSD will also change Figure 5b.

2.2 Multispectral sensor

A remote sensing system initially identifies the electromagnetic radiation that emerges from the object of study and traverses through the earth's atmosphere. The detected energy is collected as an analog electrical signal that is subsequently transformed into a digital value by the process of analog-to-digital (A-to-D) conversion (WOLF et al., 2014; JENSEN, 2014). Radiometric and geometric preprocessing may be necessary



(a) Spatial resolution varying with altitude.



(b) Spatial resolution varying with ground elevation.

Figure 5: Different spatial resolutions for a photography of the University of Maryland campus.

to improve the interpretability of the digital remotely sensed data (WOLF et al., 2014; JENSEN, 2014). The data may then be enhanced for subsequent human visual analysis or processed further using digital image processing algorithms (WOLF et al., 2014; JENSEN, 2014). The information extracted from visual or digital image processing can provide valuable insight into the biophysical and land-cover characteristics of the area of interest.

Digital remote sensor data is typically stored as a matrix or array of numbers, where each digital value is located at a specific row and column in the matrix (WOLF et al., 2014; JENSEN, 2014). A pixel is defined as the smallest nondivisible element of a digital image, and each pixel at a specific row and column has an original brightness value (BV) associated with it (as showed in Figure 6) (WOLF et al., 2014; JENSEN, 2014). The dataset may consist of n individual bands of multispectral or hyperspectral imagery (WOLF et al., 2014; JENSEN, 2014). Therefore, it is possible to identify the brightness value of a particular pixel in the dataset by specifying its row, column, and band coordinate. It is important to note that the n bands are all geometrically linked to one another, allowing for the accurate identification of features across multiple bands (WOLF et al., 2014; JENSEN, 2014).

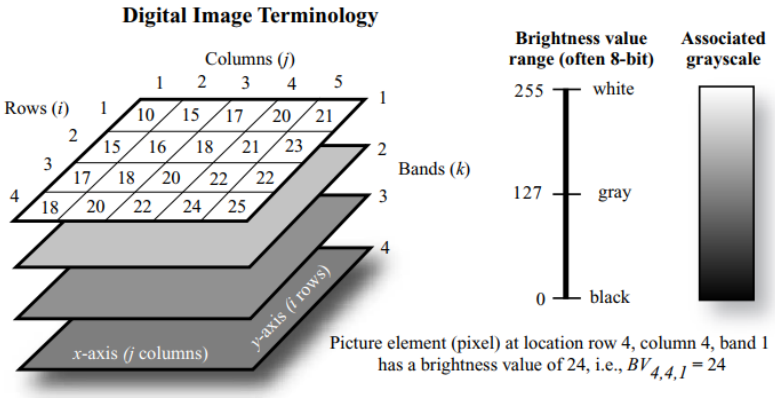


Figure 6: Digital image multiband matrix.

A multispectral sensor can have several configurations that allow it to capture the electromagnetic waves and covert it to a digital matrix, but it will generally have

a group of filters and a component that detects the waves' energy and converts it to electrical signal through components made of semiconductor materials, also known as the sensor's detectors (WOLF et al., 2014; JENSEN, 2014). The number of detectors and its arrangement are key to constraint the digital image resolution, because each pixel in the image will be somehow related to a detector or a set of them (WOLF et al., 2014; JENSEN, 2014).

2.2.1 Spectral Resolution

Spectral resolution in remote sensing refers to the ability of a sensor to resolve fine spectral features in the electromagnetic spectrum (WOLF et al., 2014). It is determined by the number of spectral bands and their spectral width or resolution (WOLF et al., 2014). High spectral resolution sensors can distinguish between narrow spectral features, while low spectral resolution sensors provide a broader view of the electromagnetic spectrum with fewer, wider bands (WOLF et al., 2014).

Spectral resolution plays a critical role in remote sensing applications, particularly in those related to vegetation, soil, and water (JAFARBIGLU; POURREZA, 2022). Fine spectral resolution enables the identification and discrimination of various plant species, plant stress, and the identification of soil and water properties (JAFARBIGLU; POURREZA, 2022). Therefore, it is important to select the appropriate spectral resolution according to the specific objectives of the remote sensing application. For example, to monitor crop health, high spectral resolution is desirable as it enables the detection of subtle changes in plant pigments and biochemicals that are indicative of plant stress (JAFARBIGLU; POURREZA, 2022). On the other hand, if the objective is to assess broad-scale land cover changes, a coarser spectral resolution may be sufficient.

Spectral resolution can also affect the accuracy of image classification and mapping (WOLF et al., 2014). The higher the spectral resolution, the greater the potential to distinguish between different surface features, leading to more accurate classification results (WOLF et al., 2014). However, high spectral resolution also requires more computational resources and can result in data storage issues, especially for large-scale applications (WOLF et al., 2014).

2.2.2 Radiometric Resolution

Radiometric resolution is a critical characteristic of digital imagery that refers to the ability of an imaging sensor to distinguish small differences in the energy level of incoming radiation (WOLF et al., 2014; JENSEN, 2014). It is a measure of the number of quantization levels used to digitize the incoming analog signal, which determines the number of different brightness or color values that can be represented in a digital image (WOLF et al., 2014; JENSEN, 2014).

Quantization, a subcategory of radiometric resolution, is the process of converting the amplitude of the original electromagnetic energy into a number of discrete levels (WOLF et al., 2014; JENSEN, 2014). The greater the number of quantization levels, the more accurate the digital representation of the analog signal, and the more details that can be discerned in the image (WOLF et al., 2014; JENSEN, 2014). For example, an image with 8-bit quantization has 256 different levels of brightness or color, while a 10-bit image has 1024 levels, and a 12-bit image has 4096 levels.

2.3 Vegetation Indices

In remote sensing and Earth observation studies, the need to map the ground according to vegetation, soil, water and its content ended up with vegetation indices (VI), a way to measure plant development using meaningful and dimensionless numbers (BANNARI et al., 1995). This number is the outcome of a mathematical expression using different spectral band reflectances, thus depending on the bands and equations it will point to different information about the plant (BANNARI et al., 1995). Despite prior efforts to acquire ground information from raw spectral bands, simple ratio (SR) was proposed in 1972 and is considered to be the first true vegetation index (JENSEN, 2014), given its attempt to point to the chlorophyll band signature - the inverse relationship between red and near-infrared reflectance in a healthy green vegetation (Figure 2). Its associated mathematical expression is:

$$\mathrm{SR} = \frac{\rho_{NIR}}{\rho_R}, \tag{1}$$

where ρ_{NIR} and ρ_R are the reflectance of near-infrared and red bands, respectively.

The Normalized Difference Vegetation Index (NDVI) it was proposed by Dr. John W. Rouse and is one of the most used indices in agriculture in order to follow up with crop development. The idea behind NDVI proposal was to find a way to document the seasonal vegetation changes quantitatively (ROUSE et al., 1974). Its main advantage in comparison with simple ratio is its better accuracy in measuring relative greenness when comparing different locations and different cycles (ROUSE et al., 1974). The index is calculated using the following formula:

$$\mathrm{NDVI} = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R} \tag{2}$$

There are many vegetation indices and some of them are improvements of the others, in which better filters and corrections on the original ones were applied. Table 1 summarizes a few of the most widely adopted vegetation indices.

Vegetation Index	Equation	Reference
Simple Ratio (SR)	$SR = \frac{\rho_{NIR}}{\rho_R}$	(BIRTH; MCVEY, 1968)
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{\rho_{NIR}-\rho_R}{\rho_{NIR}+\rho_R}$	(ROUSE et al., 1974)
Enhanced Vegetation Index (EVI)	$EVI = \frac{\rho_{NIR}-\rho_R}{\rho_{NIR}+6\rho_R-7.5\rho_B+1}$	(HUETE et al., 2002)
Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = \frac{\rho_{NIR}-\rho_G}{\rho_{NIR}+\rho_G}$	(GITELSON et al., 1995)

Table 1: Selected remote sensing vegetation indices.

2.4 Sentinel-2

The European Space Agency launched the first Sentinel-2 satellite in 2013 as part of the Global Monitoring for Environment and Security (GMES) program. The mission complies with the systematic global coverage of land surfaces from 56°S to 84°N, a high revisit frequency of every five days, a high spatial resolution (10 m, 20 m and 60 m), 13 bands in the visual and near-infrared (VNIR) and short-wave IR (SWIR) parts of the light spectrum and a wide field of view of 290 km. In order to accomplish the high revisit frequency, two identical satellites operate simultaneously in a polar sun-synchronous orbit, Sentinel-2A and Sentinel-2B (FLETCHER; AGENCY., 2012).

Sentinel-2 carries the MultiSpectral Instrument (MSI), a pushbroom based sensor that digitize the observation data into a 12 bits image. It is made up of a Three-Mirror Anastigmat (TMA) telecentric telescope (Figure 7) in which the focal planes consist of 12 VNIR and SWIR detectors in two separate staggered rows (Figure 8). Before reaching the focal planes, the reflected light passes through a beam splitter that provide the separation of the VNIR and SWIR channels. Then, the spectral separation occurs in the stripe filters mounted on top of each detector, creating a linear array of data that will form a two dimensional image for each band with the satellite movement (Figure 9) - this is the reason why it is called a pushbroom sensor (FLETCHER; AGENCY., 2012).

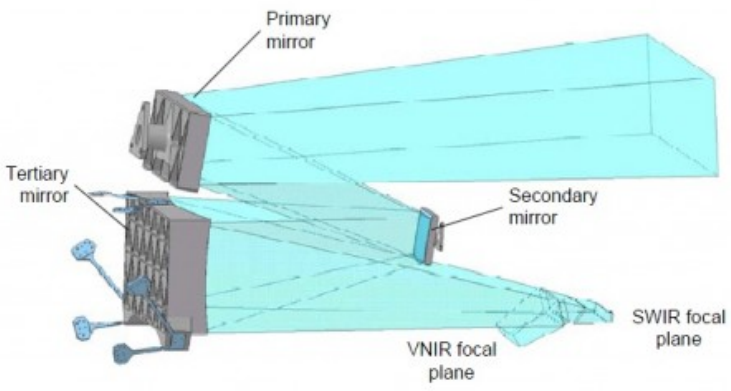


Figure 7: Three-Mirror Anastigmat telecentric telescope of MSI.

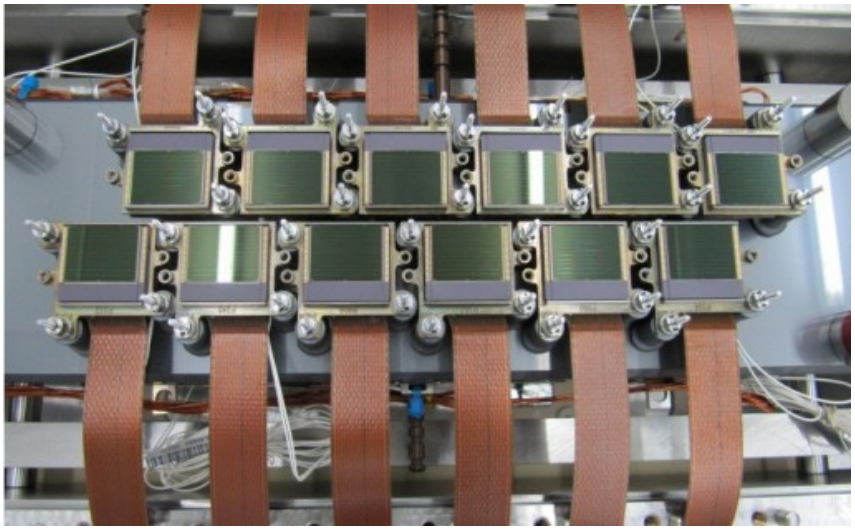


Figure 8: Staggered configuration of one of the 12 detectors of one of the focal planes in MSI.

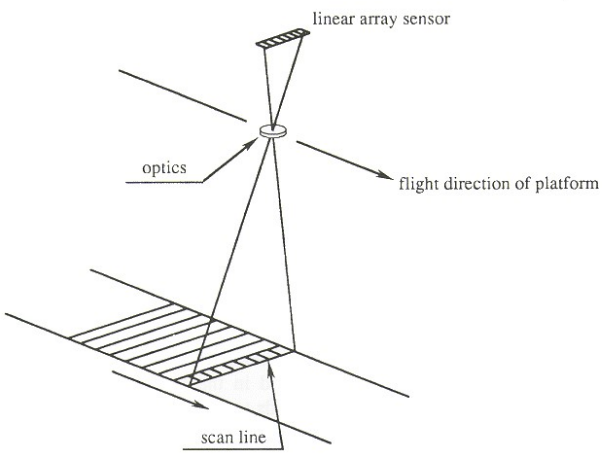


Figure 9: Linear Array "Pushbroom" sensor configuration.

The specifications of each band number of MSI are shown in Table 2. Each band is characterized by its central wavelength and bandwidth, indicating that a given band captures data observations from a range that encompasses the central wavelength minus half the bandwidth to the central wavelength plus half the bandwidth.

	S2A		S2B		
Band Number	**Central wavelength (nm)**	**Bandwidth (nm)**	**Central wavelength (nm)**	**Bandwidth (nm)**	**Spatial resolution (m)**
1	443.9	27	442.3	45	60
2	496.6	98	492.1	98	10
3	560.0	45	559	46	10
4	664.5	38	665	39	10
5	703.9	19	703.8	20	20
6	740.2	18	739.1	18	20
7	782.5	28	779.7	28	20
8	835.1	145	833	133	10
8a	864.8	33	864	32	20
9	945.0	26	943.2	27	60
10	1373.5	75	1376.9	76	60
11	1613.7	143	1610.4	141	20
12	2202.4	242	2185.7	238	20

Table 2: Sentinel-2 spectral bands.

The MultiSpectral Instrument sensor is very useful to agricultural applications given the fact that it covers red and near-infrared bands with the band numbers 4 and 8 respectively. Figure 10 shows the use of Equation (2) to generate a NDVI image of the three fields in the Alta Mogiana sugarcane farm on September 1, 2022.

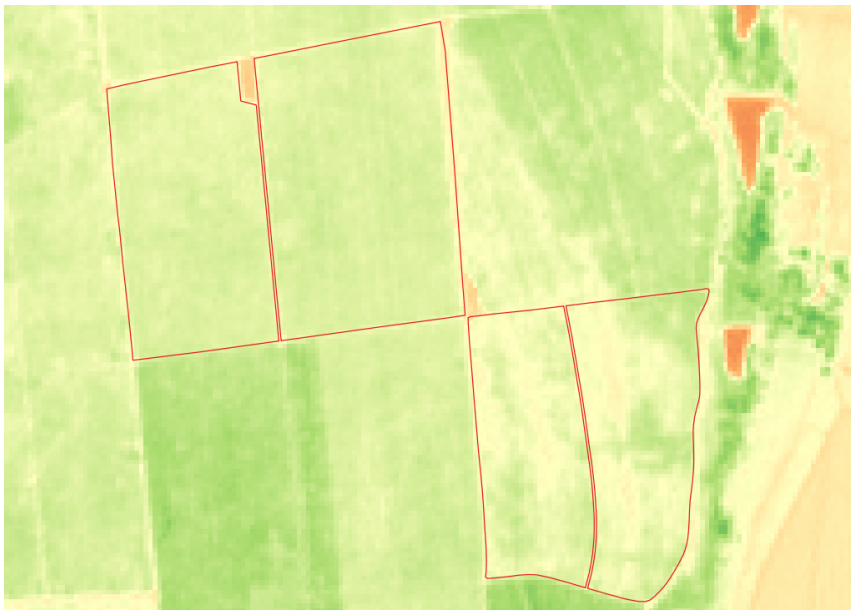


Figure 10: NDVI of Alta Mogiana sugarcane farm with three fields highlighted.

3 Neural Network

Convolutional Neural Networks (CNN) are a specialized type of artificial neural network (ANN) that have revolutionized computer vision applications. CNN can recognize patterns in images and assign them to specific categories with high accuracy. It is used for object detection to identify the presence and location of objects within an image, semantic segmentation to assign a label to each pixel in an image based on context, and other tasks requiring the learning of complex patterns from data. CNN is particularly effective for tasks that require high-level understanding of complex patterns in the data (ALZUBAIDI et al., 2021).

3.1 Artificial Neural Network

In order to understand how a convolutional neural network works, it is important to set some baseline concepts about an artificial neural network, given the fact that the CNN applies ANN in a different arrangement. For this reason, a perceptron will be introduced here, given its wide use and simple model of classification. It is a structure that models a neuron using a weighted sum of many inputs (dendrites) and produce a single output (the axon), as illustrated in Figure 11. Its representation consists of a vector of weights related to each input and a bias weight (CHARNIAK; EUGENE, 2018).

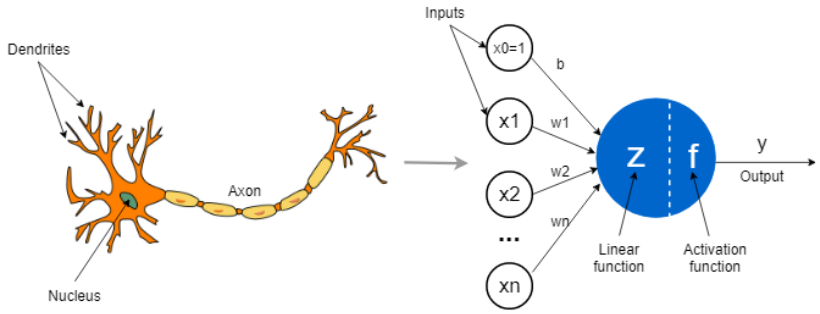


Figure 11: Perceptron model compared with a neuron illustration.

The bias weight and the vector of input weights constitute the parameters of the perceptron (identified by the letter ϕ), and they are built in an iterating process that compares the expected output for a training set of input data with the result of each iteration output (CHARNIAK; EUGENE, 2018). The possible outputs of a perceptron algorithm are based on a function, also known as activation function. The iterating process is called training, in which, for every iteration, a dot product between the input vector ($\vec{x}$) and the weight vector ($\vec{w}$) will be summed with the bias, then passed to the activation function to associate its result with the expected output (CHARNIAK; EUGENE, 2018).

The training process is an iteration in which it is desired to minimize the error between the output and the actual output value. This process is made using a relatively

simple technique called Gradient Descent (CHARNIAK; EUGENE, 2018). The basic idea of this technique is to modify the parameters of a perceptron according to the rule:

$$\Delta\phi = -\mathcal{L}\frac{\partial L}{\partial \phi} \tag{3}$$

where $\mathcal{L}$ is the learning rate, a real number that indicates how much should a parameter be changed at a given time, and L is the loss function, a function from an outcome to measure how bad it is for the problem (GOODFELLOW; BENGIO; COURVILLE, 2016; CHARNIAK; EUGENE, 2018). The loss function is part of the training design, so it should be adjusted to the type of problem it addresses to (GOODFELLOW; BENGIO; COURVILLE, 2016; CHARNIAK; EUGENE, 2018). Table 3 summarizes some of the most used loss functions associated to different scenarios, and it can be noticed that all presented functions are differentiable, and that is a criteria stated by Equation (3).

Problem type	**Loss function**
Binary classification	Binary cross-entropy (CHARNIAK; EUGENE, 2018)
Multiclass classification	Cross-entropy (CHARNIAK; EUGENE, 2018)
Non-linear regression	Mean squared error (CHARNIAK; EUGENE, 2018)

Table 3: Common loss functions associated to different neural network problems.

In neural networks with more than one layer of perceptrons (denoting the "deep" in the "deep learning" nomenclature), in which each layer feeds the next, the training process happens in an iterative forward and backward pass (CHARNIAK; EUGENE, 2018). At first, the network is initialized with random values for the parameters of each layer, then, in the forward pass, the activation functions are computed for the entire network as well as the loss function (CHARNIAK; EUGENE, 2018). Afterward, the back-propagation process computes the gradient loss for the entire training data set and takes its average to apply in Equation (3) to set up the parameters change which will be used in the next iteration step, and then the forward pass starts again beginning of a new cycle (CHARNIAK; EUGENE, 2018). These iteration steps keep going until the loss function is minimized to a lower value, concluding the Gradient Descend technique. There is a common variation of this technique, that updates the parameters every certain amount of data training set (also known as batch size), which denote the Stochastic Gradient Descend (CHARNIAK; EUGENE, 2018).

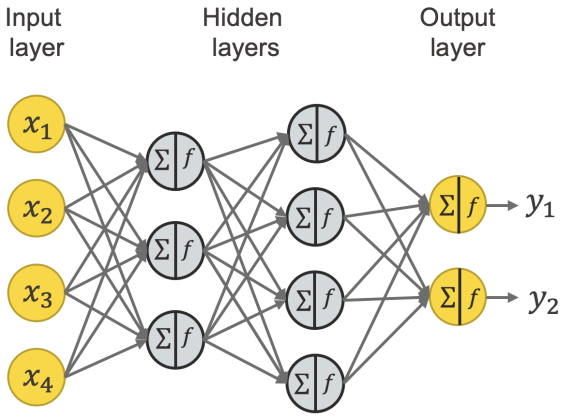


Figure 12: Deep neural network with two hidden layers. Each node has a weighted sum represented by the Greek letter Σ and an activation function represented by the letter f.

The activation function is a key point to determine what type of problem the neural network is set up. For a binary classification - used to determine whether the input is part of a class or not -, a Heaviside function could be used, however, it might be of great use to apply a sigmoid function or even a rectified linear unit function (relu) when the ANN is to classify inputs into different classes. Sigmoid and relu functions can be seen in Figure 13. When a neural network is to be applied in a multiclass problem, there should be one perceptron for each class in the output layer (CHARNIAK; EUGENE, 2018). The perceptron with the highest output value is the right class to describe the input (CHARNIAK; EUGENE, 2018). The reasons for choosing an appropriate activation function are:

- Non-linearity: Activation functions introduce non-linearity to the output of a neuron, which is essential for capturing complex patterns and relationships in the data. Without non-linearity, the ANN would be limited to performing linear transformations, which may not be sufficient to model complex data distributions (CHARNIAK; EUGENE, 2018).
- Gradient Descent: The activation function affects the gradient descent algorithm used to optimize the weights of the network. Different activation functions have different gradients, which can impact how quickly the network converges to an optimal solution during training (CHARNIAK; EUGENE, 2018).
- Vanishing Gradient Problem: Some activation functions, such as the sigmoid and hyperbolic tangent functions, can suffer from the vanishing gradient problem. This means that as the network becomes deeper, the gradients can become very small, making it difficult for the network to learn (GOODFELLOW; BENGIO; COURVILLE, 2016).

- Output range: It should be noted that an activation function must have a wide range of values to avoid problems related to bad performance, given its representation range to fit the problem that will cause the network to struggle (CHARNIAK; EUGENE, 2018).

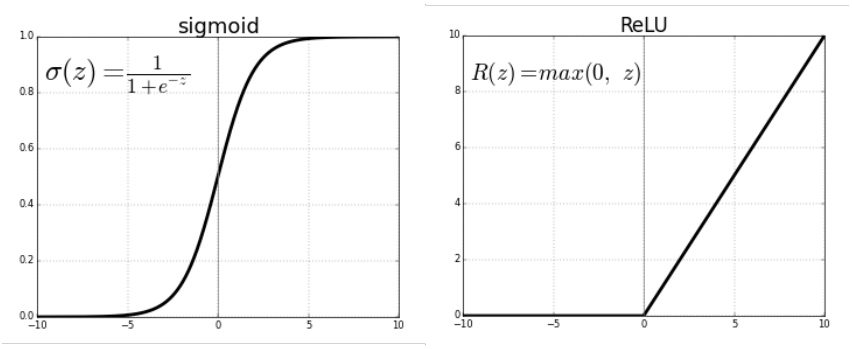


Figure 13: Graphical representation of sigmoid and relu activation functions.

3.2 Convolutional Neural Network

Convolutional neural networks are a specialized type of neural network designed to process data with a grid-like structure, such as images or audio signals. Unlike fully connected neural networks, where every neuron in one layer is connected to every neuron in the next layer, CNN has neurons that are only connected to a small region of the input (CHARNIAK; EUGENE, 2018). By exploiting local spatial correlations in the input data, CNN can learn representations that are efficient and robust to variations in the input.

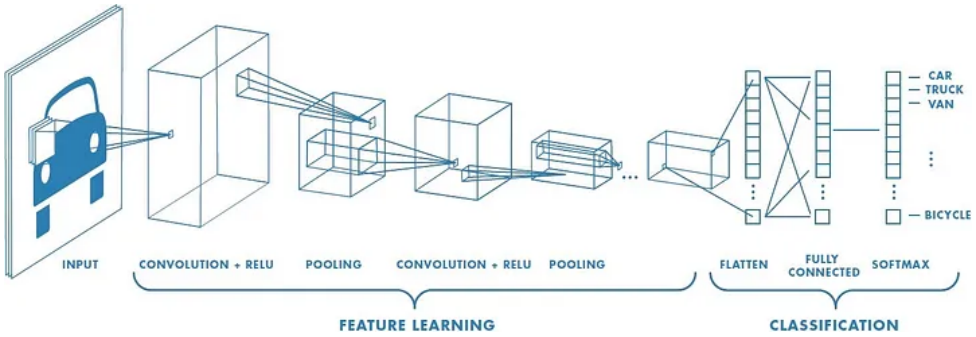


Figure 14: A convolutional neural network that classifies images into different automobiles. The shown input image of the training set is a car. A batch of the input is convoluted with a set of filters/kernels to generate a set of feature maps that will be passed as an argument to an activation function. Each element of the feature maps will be computed and activated, resulting in a new set that will pass through a process of pooling that minimizes noise values and the set dimensions. Afterward, the last layer output set turns into the next layer input set, and the process happens again. It is important to notice that the illustration represents a thinner and wider box at the beginning layer and ends up with a thick and narrow box in the last layer of feature learning. This illustrates that after each convolution with the filters/kernels the set increases and after the pooling the set magnitude stays the same and each pooled output becomes either the same size or narrower.

The key idea behind CNN is to use a set of learnable filters or kernels that are convolved with the input to produce a set of feature maps (shown in Figure 14). Each feature map represents the output of a particular filter at every spatial location in the input. By stacking multiple convolutional layers, a CNN can learn increasingly complex features that capture different levels of abstraction (GOODFELLOW; BENGIO; COURVILLE, 2016; CHARNIAK; EUGENE, 2018). In addition to convolutional layers, CNN also typically includes pooling layers, which downsample the feature maps by taking the maximum or average value over a small local region (GOODFELLOW; BENGIO; COURVILLE, 2016).

CNN usually end with one or more fully connected layers, which perform a classification or regression task. These layers take the output of the convolutional and pooling layers, and each neuron's input is a flattened version of the feature maps, rather than the raw input data (GOODFELLOW; BENGIO; COURVILLE, 2016). CNN has proven to be effective for a wide range of computer vision and audio processing tasks. They are powerful and flexible neural networks that can learn increasingly complex features from grid-like input data by exploiting local spatial correlations (GOODFELLOW; BENGIO; COURVILLE, 2016; CHARNIAK; EUGENE, 2018).

3.2.1 The convolution operation

The convolution operation is a mathematical operation that is performed between the input image and a set of filters to produce feature maps (GOODFELLOW; BENGIO;

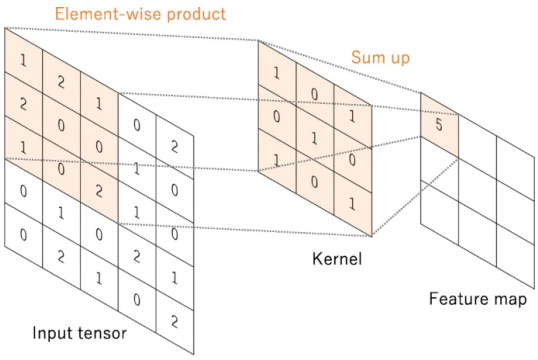
COURVILLE, 2016; CHARNIAK; EUGENE, 2018). Mathematically, the convolution operation can be represented as follows:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) \tag{4}$$

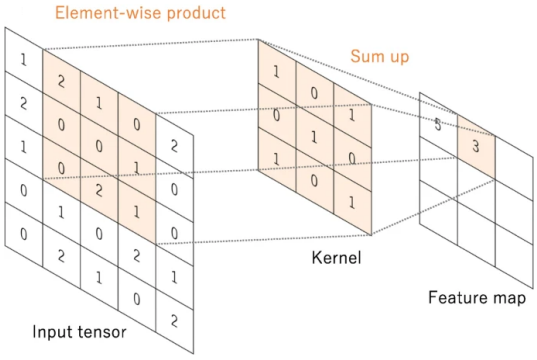
where I is the input image, K is the convolutional kernel or filter, and S is the resulting feature map. The symbol $*$ represents the convolution operation, and (i, j) represents the spatial coordinates of a pixel in the feature map.

The convolution operation is performed by sliding the kernel over the input image, computing the dot product between the kernel and the corresponding region of the input image at each position, and storing the result in the corresponding pixel of the feature map (GOODFELLOW; BENGIO; COURVILLE, 2016). The kernel is a small matrix of learnable parameters that are updated during the training process to learn different features in the input image (GOODFELLOW; BENGIO; COURVILLE, 2016; CHARNIAK; EUGENE, 2018). The convolution operation is illustrated in Figure 15a, Figure 15b and Figure 15c with a single input and a single kernel, resulting in a single feature map. In practical applications, both input and kernel are a set of multiple units, resulting in a set of feature maps - the magnitude of the feature map set is equal to the magnitude of the input multiplied to the magnitude of the kernel set.

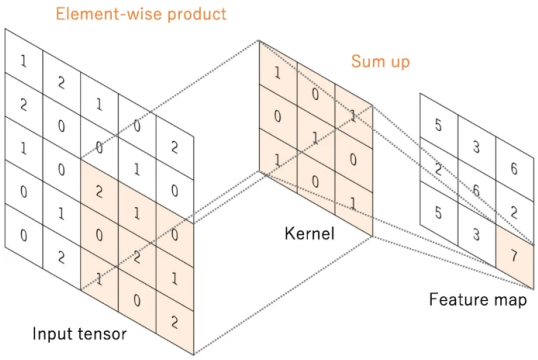
The size of the feature map is determined by the size of the input image, the size of the kernel, and the stride and padding used during the convolution operation (CHARNIAK; EUGENE, 2018). The stride determines the step size at which the kernel moves across the input image (as shown in Figure 15d), while padding is used to add zeros around the edges of the input image to preserve its spatial dimensions. The output feature map has a smaller spatial dimension than the input image, but a greater number of channels, representing the different learned features (CHARNIAK; EUGENE, 2018).



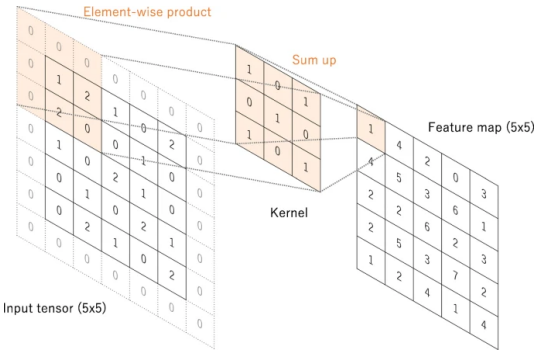
(a) Convolution operation to generate element $(1, 1)$ of the feature map.



(b) Convolution operation to generate element $(1, 2)$ of the feature map.



(c) Convolution operation to generate element $(9, 9)$ of the feature map.

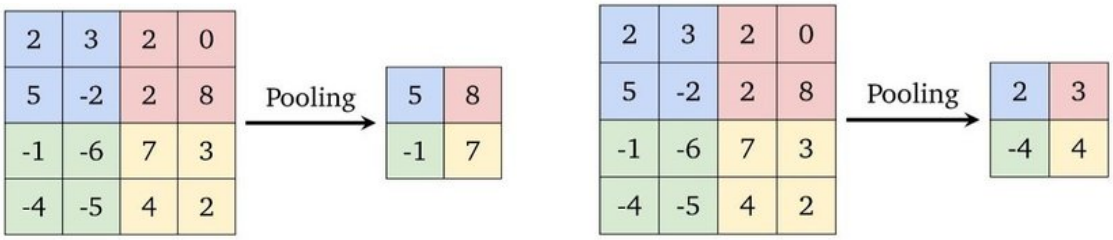


(d) Convolution operation with padding to generate a feature map with the same dimensions as the input.

Figure 15: Convolution operation between the input (input tensor) and one kernel. The stride here is 1, meaning that the kernel matrix will slide only one column of the input.

3.2.2 Pooling

Pooling is a downsampling operation that is typically applied after each convolutional layer to reduce the spatial dimensions of the feature maps and improve computational efficiency (GOODFELLOW; BENGIO; COURVILLE, 2016). The most common type of pooling operation used in convolutional neural networks is max pooling, which selects the maximum value from a small region of the feature map and discards the rest. During the max pooling operation, the kernel is moved across the input feature map with a certain stride, and for each position (i, j) in the output feature map, the maximum value within the local receptive field is selected and stored in the corresponding position of the output feature map (GOODFELLOW; BENGIO; COURVILLE, 2016). This process results in a downsampled version of the input feature map, with reduced spatial dimensions (as shown in Figure 16a).



(a) Max pooling.

(b) Average pooling.

Figure 16: Two pooling algorithms being applied to the same feature map.

Other types of pooling operations, such as average pooling or L2 pooling, can also be used in CNN (GOODFELLOW; BENGIO; COURVILLE, 2016). Average pooling computes the average value within each local receptive field (shown in Figure 16b), while L2 pooling computes the root-mean-square value. These pooling operations have the effect of reducing the spatial dimensions of the feature maps, while also introducing some degree of a phenomenon called translational invariance (GOODFELLOW; BENGIO; COURVILLE, 2016). Translational invariance is a property that allows the model to recognize the same pattern in different locations of the input image (GOODFELLOW; BENGIO; COURVILLE, 2016). This property is especially important for computer vision tasks, where objects can appear in different positions, orientations, and scales (GOODFELLOW; BENGIO; COURVILLE, 2016).

4 Material and methods

In order to analyze whether convolutional neural network and remote sensing can predict sugarcane biomass, a comparison of predicted data from the studied model against reference data were conducted.

The sugarcane harvester's yield monitor data were adopted as reference data. Despite the fact that yield monitor data from sugarcane harvesters are also estimated, and not an actual field biomass data, it is the only source of fine scale and high density biomass data available in sugar mills.

The model prediction capability was carried out in two spatial scales. The first was the whole field scale, aiming to estimate the average field biomass. And the second and finer scale aimed to estimate the within-field spatial variability of the biomass, that would allow creating biomass maps of sugarcane fields.

4.1 Study site

The data used in this project were collected from three fields (18, 46 and 53) of Alta Mogiana sugar mill, a sugarcane mill located in Sao Joaquim da Barra, in the state of Sao

Field	Start date of the season	End date of the season	CGDD at harvest	Cultivar	Reference biomass (tonne/ha)
18	Oct. 22, 2021	Sep. 24, 2022	1653	CTC 4	97.6
46	Oct. 22, 2021	Sep. 21, 2022	1644	CTC 4	77.8
53	Oct. 6, 2021	Sep. 19, 2022	1702	SP 80-3280	82.8

Table 4: Fields' season information.

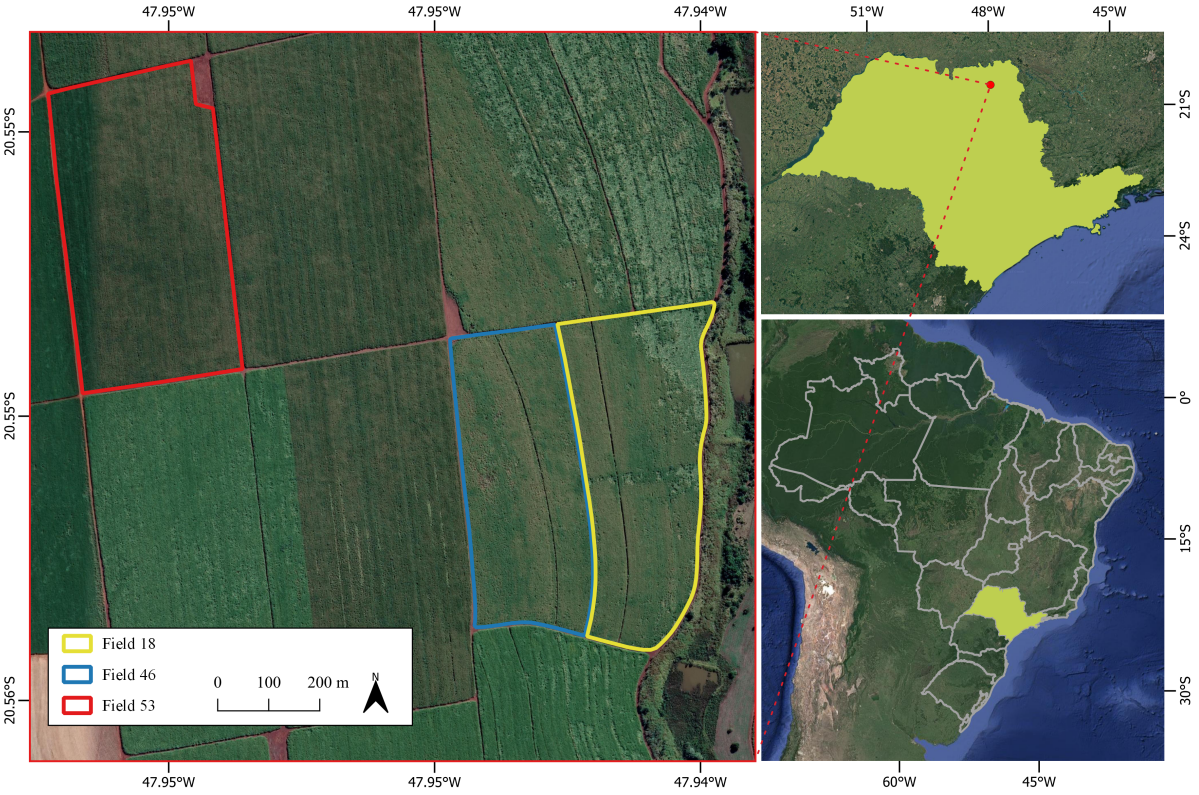


Figure 17: Fields used to train and test the deep learning model.

Paulo. The crop stage for field 53 is first ratoon, while fields 18 and 46 are fourth ratoon. This means that since the plantation, field 53 was harvested two times and the other two were harvested five times. Table 4 shows details about each field's harvest period and plant cultivar. The mean annual precipitation, minimum and maximum temperature in the region are 1651 mm, 18°C and 28°C, respectively, according to the National Institute of Meteorology (INMET) historical weather data. The soil type is Oxisol, according to the United States Department of Agriculture (USDA) soil taxonomy. Figure 17 shows the location of the three fields.

The reference values of field biomass was acquired by filtering the data from the sugarcane harvester yield monitor.

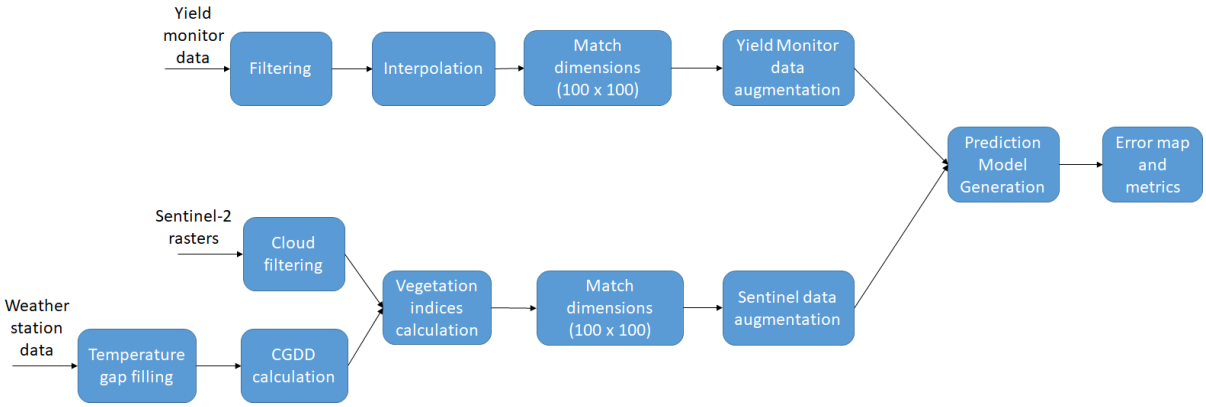


Figure 18: Modeling framework flowchart.

4.2 Sentinel-2 and yield monitor data

The deep learning model was developed to predict a biomass map of a sugarcane field with accumulated growing degree days varying from 1600 and 1700. Weather data, Sentinel-2 satellite imagery and yield monitor data from the target fields were used for model training and test. The deep learning model had both Sentinel-2 imagery and weather data combined as its single input while yield maps to elaborate the labels - the desired output. In order to train the model, two datasets were prepared in a preprocess that sets a data structure pattern for the input and label. Figure 18 illustrates the main steps of the modeling framework.

The model used as input data the preprocessed Sentinel-2 images. The Sentinel-2 images were downloaded using Google Earth Engine API, specifying the cumulative growing degree days (CGDD) range corresponding to 549 through 910 for the harvest period of each field, excluding the ones with cloud covering over the studied fields. The CGDD range was selected based on (LOFTON et al., 2012) and (MAIA; BUFON; LEãO, 2023) conclusions about the ideal time to predict sugarcane yield using vegetation indices and the availability of clear images. Three sets of images were collected, containing 5 images for field 18, 5 images for field 46 and 4 images for field 53.

The response variable of the model - biomass -, was derived from filtered yield monitor data. Yield monitor data is a vector file containing a set of points in which the coordinates indicate the location of the harvested biomass. Each point contains attributes concerning the time of collection, associated yield, associated trash (leaves and other wasted parts of the sugarcane). Hence, the yield and trash of each point were summed to create the attribute biomass, constituting the label variable used in the model whether for the training or testing process. The number of yield monitor biomass data points for the fields 18, 46 and 53 were 60798, 57999 and 86435 respectively, while the area of each respective field is 14.75 ha, 13.19 ha and 18.88 ha.

4.3 Yield monitor filtering

Although the yield monitor data was used as the reference value for the prediction process, it still has some discrepancies that do not represent reality as harvester maneuvering and the monitor biomass sensing system still face some early development stage challenges. An exploratory error assessment of the yield monitor estimates, compared to hand harvested field biomass in selected points (data not shown) indicates a mean absolute percentage error (MAPE), average error (AE), root mean square error (RMSE), and the mean absolute error (MAE) of 5.08 %, 1.52 tonne/ha, 38.29 tonne/ha and 5.09 tonne/ha, respectively. Therefore, in order to be used as reference data, a filtering process was adopted. The values of biomass reached values with magnitudes as high as 2424 tonne/ha, an unreasonable number for biomass in sugarcane fields. A process of filtering outliers and yield map error removal was made using Menegatti e Molin (2004) methodology. This method was applied to the yield monitor vector file through Microsoft Excel functions.

4.4 Weather data and growing degree days

Cumulative growing degree days (CGDD) were included in the input, adding a correlation between the plant's electromagnetic reflectance and its related phenological growing stage. Growing degree days are a measure of the accumulated heat energy available for plant growth during a specific period. It provides an estimate of how favorable the temperature conditions are for plant development.

Daily minimum and maximum temperatures of each day from the beginning through the end of the harvest period were gathered in order to calculate daily CGDD of each field. The weather data was downloaded from the INMET station of São Joaquim da Barra, the closest one to the fields. The dates that no data was collected were filled with data from other stations close to the fields.

After the weather data was prepared, the GDD and CGDD of each field was calculated using the following equations, as proposed by Teruel, Barbieri e Ferraro (1997):

$$GDD = \begin{cases} \frac{T_{\max}+T_{\min}}{2} - T_b, & \text{for } T_{\min} > T_b \\ \frac{(T_{\max}-T_b)^2}{2(T_{\max}-T_{\min})}, & \text{for } T_{\min} \leq T_b \end{cases}, \tag{5}$$

where T_{max} is the maximum temperature of the day, T_{min} is the minimum temperature of the day and T_b is the basal temperature, which is 18°C for sugarcane in Brazil (TERUEL; BARBIERI; FERRARO, 1997). The cumulative growing degree days in turn is just the sum of all GDD from the start of the season up to the current date.

4.5 Vegetation indices calculation

Images from Sentinel-2 were downloaded with values discretized as digital numbers (DN), then the pixels of each band were divided by 10000 in order to have reflectance values ranging from 0.0001 through 1 (GASCON et al., 2017). The spectral signatures GNDVI and EVI shown in Table 1 were calculated and incorporated to the input dataset, as well as the bands Red and Blue. The GNDVI and EVI were both divided by the cumulative growing degree days to create new indices regarding the date that the image was taken and the start of the field's harvest season. The addition of these two bands to the input image aimed an improved relationship of the images with plant phenological growing stage, as observed by Lofton et al. (2012).

4.6 Yield monitor interpolation

The georeferenced point data from the yield monitor was submitted to an interpolation in order to have a grid with georeferenced cells matching the clipped field's from Sentinel-2. Given the density of biomass points, Inverse Distance Weighting (IDW) interpolation method was used, a simple bidimensional interpolation method that takes in consideration that the data is spatially dependent, though it is not true for biomass in a field. The interpolation was made using a standalone code in Python that followed the equation (LU; WONG, 2008):

$$\hat{z} = \sum_{i=1}^{n} \frac{z_i \, d_i^n}{d_i^p}, \tag{6}$$

where $\hat{z}$ is the interpolated value, z_i is the value of the i-th point of the dataset, d_i is the distance between the estimated value and the i-th point of the dataset and p is the power of the interpolation, in which 2 was used with no specific criteria. This equation was used in every new cell of the grid to be generated but the cells that represent the field's surrounding bare soil. For the bare soil values that corresponded to areas out of the field, a biomass of 0 tonne/ha was forcibly set.

4.7 Field and bare soil data extraction

Once the Sentinel-2 images were downloaded, the area of interest of the study for each image was extracted using the sugarcane field's Shapefile Polygon provided by the sugar mill. Additionally, three bare soil pixels were selected from each image in order to provide a zero biomass reflectance reference.

4.8 Input and label data creation

Further, new images were generated from the clipped parcels including only the extracted three fields and the three bare soil reference pixels. The generated image had

100 pixels of width by 100 pixels of height, containing the parcels of clipped fields over a homogeneous background for each band. The value of the background in each band is the average value of the clipped bare soil in that respective band. The background of the generated input had different values not only for each band, but it varied according to the satellite captured reflectance of the soil, and to the CGDD of the field on that date for the fifth and sixth band.

4.9 Data augmentation

Two-dimensional convolution neural network models are a sequence of convolution operations between an input and a set of filters that recognize spatial patterns. Therefore, if it is trained to recognize images where the target object is in the same relative position and orientation, it will be biased to extract information from that same position and orientation. This issue is called spatial invariance, and this issue is faced in this model using a data augmentation strategy of varying the position and orientation of the fields on each image (WILK et al., 2018). New inputs and labels were generated with the orientation varying each 20° in each position, and the horizontal and vertical position varying each 5 pixels. It was noticeable that the pixels' arrangement of the fields did not match appropriately in a grid image for every orientation, thus, this rearrangement was made using linear interpolation. After the augmentation process, the total number of input data went from a total of 14 to 725 for field 18, 880 for field 46 and 564 for field 53.

4.10 Deep Learning model development

The deep learning model was developed using the Keras package for Python. The structure upon which each layer was built is the Sequential class, an easy way of grouping stacks of layers (neural networks, convolutional neural networks, and pooling) to the model.

The datasets of the input and the label were indexed to each other according to the field position and orientation of each image. For example, input and label with the field occupying the lower right side of the image and rotated 90 degrees from its center must match. The two datasets were then sorted and transformed into a Python data structure called Numpy Array - it turns the amount of memory to represent the same dataset shorter and increases the training process's performance. After, the two datasets were split into training and test sets, in order to separate the final model evaluation with unbiased results. For the training set, the fields 18 and 53 images were used, while the field 53 made up the test set.

The architecture of the model analyzed in this work was reached after the progressive improvement of primitive versions of deep learning models. It started with simple

neural networks up to the architecture presented in Figure 19. In the beginning of the process, the input values are normalized according to equation:

$$Normilized_{i,j} = \frac{input_{i,j} - meanBands_{i,j}}{\sqrt{varBands_{i,j}}}, \tag{7}$$

where $Normilized_i$, $input_{i,j}$, $meanBands_{i,j}$ and $varBands_{i,j}$ are the cell i, j values of the normalized input, of the input itself, of the mean of all bands and of the variance of all bands. This normalization process was important to improve the training process performance and stability given the similarities of the scale of input values. Without normalization, the model could inherit a specific band (feature) dominance due to its larger scale.

Furthermore, each cell passes through a neural network layer that returns two values. This layer captured information purely about band signatures. After, the grid was passed through a convolutional layer with 24 filters and kernel size of 3 by 3 pixels, resulting in a 98 by 98 pixels with 24 bands (features) each. Then, this new grid passed through a 2-dimension pooling, reducing even more its height and width (16 by 16 pixels). The pooling process aimed to smooth the grid spatial variation. The dimension of the output must be the same as the label's, thus, the previous grid was flattened and passed through another neural network to increase the width and height back to 100 by 100 pixels as well as the model complexity, meaning that it will capture more information.

4.11 Training and Statistical Evaluation of Models

In order to improve the model weights at each epoch (number of iterations in which the deep learning weights are adjusted), the MAE function was chosen to be the Loss function. Based on the result of the forward pass Loss function, the weights were adjusted using the optimizer RMSEProp from the Keras library. The model observed metrics were calculated to the predicted and the actual interpolated biomass map. They were used in the decision of the amount of iterations to successively adjust the weights of the model. After no more alteration of the Loss value and metrics, the training process was interrupted and finalized. The adopted metrics were the MAE and the RMSE of the cells corresponding exclusively to the field location, not to the whole 100 by 100 pixels image. The following equations stand for MAE and RMSE, the model metrics.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

The training process was set up to have the following parameters: training input and label, number of epochs and the validation split. In order to assess the accuracy or

error of each epoch, part of the training dataset was randomly selected to be used only for training and the other part was selected to assess the epoch metrics. Therefore, the model fitting was made with 20% of the dataset separated exclusively for validating the epochs with unbiased results of each epoch. At least, the testing process ran using data from field 46, and the MAPE, AE, RMSE, MAE, were adopted in order to assess the prediction effectiveness of the model in both targeted, field- and whithin-field spatial scales.

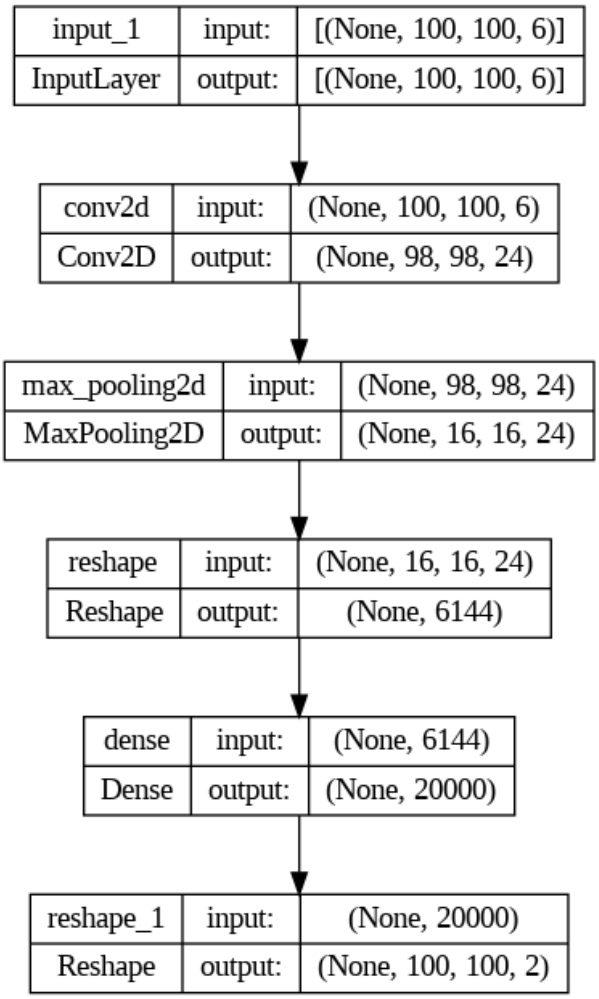


Figure 19: Deep learning architecture.

5 Results

5.1 Model prediction of average field biomass

For the training fields 18 and 53, with reference average biomass of 97.6 and 82.8 tonne/ha, respectively, the model predicted 97.63 and 82.80 tonne/ha. For the testing field 46, with reference average biomass of 77.8 tonne/ha, the developed model estimated 78.1 tonne/ha.

The MAPE, AE, RMSE, and MAE, presented in Table 5, were adopted in order

MAPE	AE	RMSE	MAE
3.95	0.30	3.48	3.07

Table 5: Prediction errors of the tested field.

to assess the prediction effectiveness of the model. The developed modeling framework tested in field 46 delivered a MAPE of 3.95% of total sugarcane field biomass, smaller than the 5.09% from the harvester yield monitor. AE of 0.30 tonne/ha, points out to an overestimation of the modeling framework relative to the field biomass, in the same way as the yield monitor AE of 1.52 tonne/ha. The RMSE of 3.48 tonne/ha, indicates that the modeling framework error around the average is lower than the yield monitor error stability, which presented an RMSE of 38.29 tonne/ha. Further, the MAE of 3.07 tonne/ha is lower than the yield monitor MAE of 5.09 tonne/ha. Due to the proximity of RMSE and MAE, the modeling framework presence of outliers is smaller in comparison with the yield monitor measurements.

5.2 Model prediction of within-field biomass spatial variability

The within-field biomass spatial variability prediction counted with more than 30% of the pixels with error higher than 28.42 tonne/ha. The error distribution of all pixels from every predicted map is shown in Figure 20. The red curve is the Gaussian normal distribution function that fits to the error histogram, in which the vertical dashed lines represent the multiples of the standard deviation position along the x-axis and the vertical solid line at 0.30 tonne/ha is where the average value of all pixels is located.

The Gaussian distribution of the pixels indicates almost equal amounts of under-estimated and overestimated biomass predictions. The negative and positive standard deviation interval around the average indicates the error in 68% of the pixels below 28.42 tonne/ha. By the percentage distribution of pixels, it is observed that around 8.5% of the pixels had an error less than 2.5 tonne/ha from the reference biomass equivalent pixel, and around 25% had an error less than 7.5 tonne/ha.

Figure 21 shows the reference and predicted biomass maps for the field 46 and Figure 22 shows the different average error maps of each rotation from the test dataset. In general, the model underestimated one corner of the test field extending to the adjacent side. In the maps with rotation of 0 degree and 280 degrees, this effect can be better seen although it is observed in almost all images. The error maps also diverged as rotation increased, meaning it did not present a pattern for the images of the same field. In contrast, images with the same rotation but different positions on the input grid did present similar error maps.

In general, pixels that overestimated the biomass were compensated by others underestimating it. This can be observed by the scatter plot in Figure 23, showing a certain

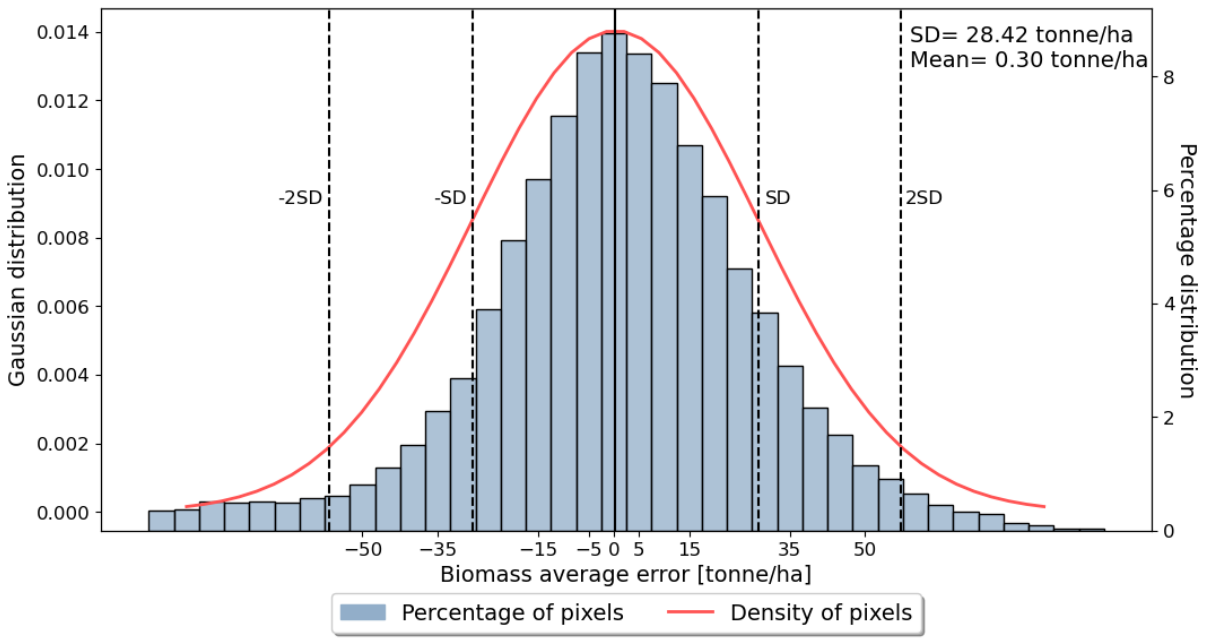


Figure 20: Gaussian normal distribution and percentage distribution of the average error of the pixels.

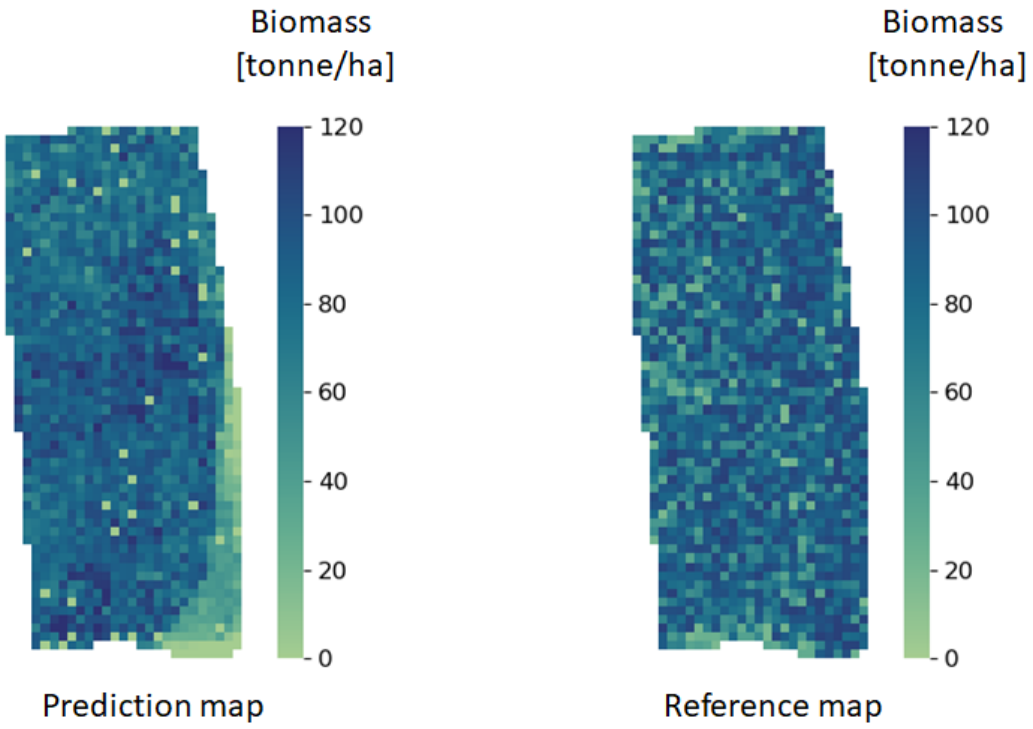


Figure 21: Predicted and reference biomass maps.

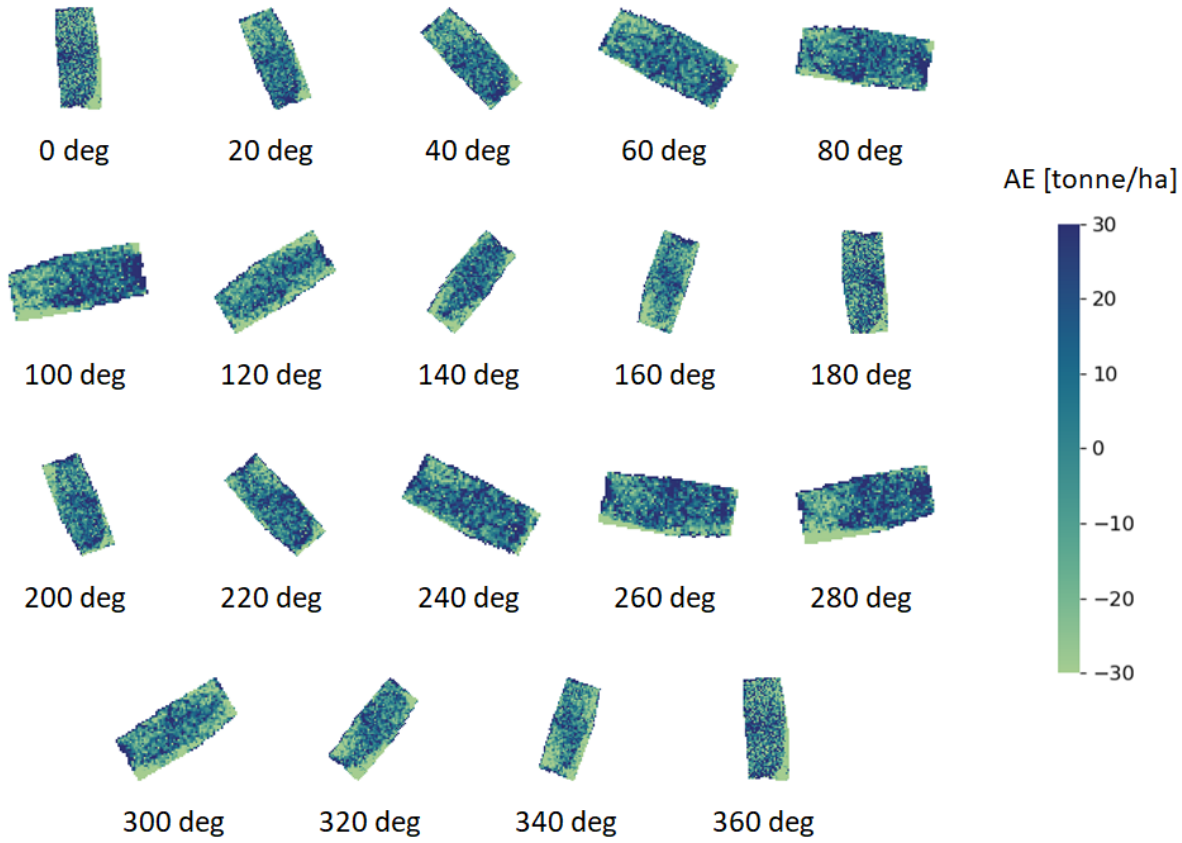


Figure 22: Average error maps from fields at different rotations.

balance in the number of pixels above and below the red line. Pixels from the reference biomass map presented values from 15 tonne/ha to beyond 115 tonne/ha, while predicted biomass was concentrated in a window between 60 and 100 tonne/ha. Further, there were pixels identified by the predicted map as bare soil - without biomass, in opposition to the reference biomass map that shows a biomass value for these pixels.

6 Discussion

The model capacity to predict average field biomass was assertive given the deep learning metrics used to improve the accuracy. Mean squared error is the loss function based on average values of the whole field, thus, they were the main target of each iteration during training. Furthermore, the decision to end the training process was based on the map metrics that calculated the map mean errors MAE and RMSE, therefore only after they were stabilized. The selected bands also had an important influence in the prediction given the correlation between GNDVI and yield, a biomass component, as well as the index sensitivity to dense vegetation (RAHMAN; ROBSON, 2016), and the sensitivity of EVI to detect vegetation biomass variation (MAIA; BUFON; LEãO, 2023). Another factor that contributed to the accuracy of the model to predict average field biomass

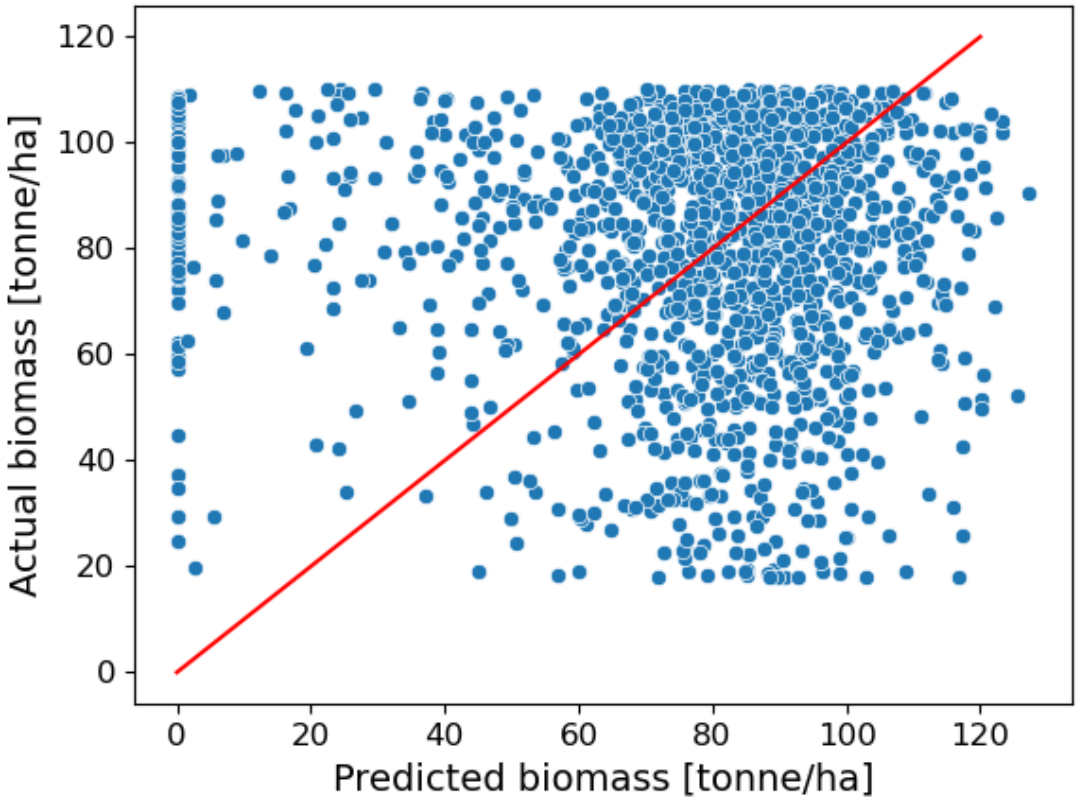


Figure 23: Scatter plot of the predicted and reference biomass in each pixel.

was the use of vegetation indices normalized the field thermal sum CGDD, adding more correlation to yield accumulation over thermal time (LOFTON et al., 2012). The reason why the model slightly overestimated the reference average biomass was merely given the numerical and stochastic nature of the model. Considering the field average reference biomass of 77.8 tonne/ha, an AE value of 0.30 tonne/ha of the prediction model could be considered acceptable, especially when compared to the AE value of 1.52 tonne/ha of the yield monitor estimates for the same field.

The larger AE for the within-field biomass spatial variability prediction was probably impacted by a dataset with small rotational variance. This is mainly noted given the lack of pattern in the error maps when the field data was submitted to rotation. The Loss function was processing the image as a whole and not pixel wise, thus, this may have contributed to the lack of correlation between each pixel of one input and its respective pixel of the output. Nevertheless, the Loss value was stabilized when the training ended, indicating underfitting. It means that the model could not capture the complexity of the nature of the processed data. A possible reason for it was the insufficiency of the dataset volume richness (MATHRANI et al., 2021), thus, the input rotation process was not enough to increase the dataset diversity in order to make a generalistic model. Besides, by observing the predicted error maps, the found pattern of a corner strip with

high underestimation indicates that the model was trying to identify the same shape of field 18 into field 46. Hence, this model flaw may have contributed to disturb the spatial variability correlation between predicted and actual biomass maps.

The scatter plot shows an upper and lower threshold that does not correspond to reality and contributed to the error related to the spatial variability of the prediction. This threshold was probably imposed by the yield monitor filtering method that removed outliers but also removed actual low, or even null biomass data. It could be considered a weakness of the method when applied to sugarcane, compared to grain crops, in which the methodology was developed. The reason for it was possibly the fact that sugarcane harvesters harvest each row at a pass and grain harvesters harvest multiple rows at a pass. Therefore, registering low or zero biomass in sugarcane rather than in grain harvester yield maps.

7 Conclusion

In conclusion, the model effectively predicted average biomass in a sugarcane field. The selection of appropriate bands, such as GNDVI and EVI, improved accuracy by correlating with yield and detecting vegetation biomass variation. Normalizing the vegetation indices through field thermal sum CGDD further enhanced prediction. The slight overestimation of reference average field biomass was due to the model's numerical and stochastic nature.

The limited rotational variance in the dataset contributed to the large error for the within-field biomass spatial variability prediction. The lack of patterns in error maps and the underfitting indicated the model's inability to capture pixel-wise correlation between predicted and actual data, and information complexity. Insufficient dataset richness and inadequate diversity hindered the development of a generalizable model. The presence of a corner strip with high underestimation disrupted the spatial variability correlation.

The unrealistic upper and lower threshold in the scatter plot affected the prediction's spatial variability prediction. This threshold, probably imposed by the yield monitor filtering method, was not optimized for sugarcane crop.

Finally, while the model demonstrated good predictive capabilities for average field biomass, limitations in the dataset's rotational variance and the presence of an unrealistic threshold influenced the accuracy and spatial variability of the predictions resulting in a bad quality prediction map in comparison with a yield monitor map. Further improvements are necessary to address these limitations and develop a more robust model that can better capture the complexities of sugarcane crop. Some of the improvements could be the application of other augmentation strategies, like zooming and noiseing, the increase of the dataset rotation to finer angles and around other axis, the development of

a sugarcane yield monitor filtering method, and even increasing the model complexity by relating a Loss function with each pixel.

Bibliography

ALZUBAIDI, L. et al. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, Springer Science and Business Media Deutschland GmbH, v. 8, 12 2021. ISSN 21961115. 24

BANNARI, A. et al. A review of vegetation indices. *Remote Sensing Reviews*, v. 13, p. 95–120, 1995. ISSN 02757257. 20

BERNARDI, A. C. C. et al. *Agricultura de precisão: resultados de um novo olhar*. [S.l.]: Embrapa, 2014. 58-73 p. ISBN 9788570353528. 13

BIRTH, G. S.; MCVEY, G. R. Measuring the color of growing turf with a reflectance spectrophotometer. 1968. 21

CARNEIRO, F. M. et al. Comparison between vegetation indices for detecting spatial and temporal variabilities in soybean crop using canopy sensors. *Precision Agriculture*, Springer, v. 21, p. 979–1007, 10 2020. ISSN 15731618. 16, 17

CHARNIAK; EUGENE. *Introduction to Deep Learning*. 2018. 24, 25, 26, 27, 28, 29

CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL. *VBP É Projetado Em R$ 1,20 Trilhão Em 2021*. [S.l.], 2021. Disponível em: <https://www.cnabrasil.org.br/storage/arquivos/dtec-VBP_outubro-16nov2021-v2.pdf>. 11

CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL AND CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA. *PIB Do Agronegócio Alcança Participação De 26,6% No Pib Brasileiro Em 2020*. [S.l.], 2021. Disponível em: <https://cnabrasil.org.br/storage/arquivos/sut.pib_dez_2020.9mar2021.pdf>. 11

FLETCHER, K.; AGENCY., E. S. *Sentinel-2 : ESA's optical high-resolution mission for GMES operational services*. [S.l.: s.n.], 2012. 70 p. ISBN 9789292214197. 21

GASCON, F. et al. Copernicus sentinel-2a calibration and products validation status. *Remote Sensing*, MDPI AG, v. 9, 6 2017. ISSN 20724292. 35

GITELSON, A. A. et al. Use of a green channel in remote sensing of global vegetation from eos-modis. *REMOTE SENS. ENVIRON*, ©Elsevier Science Inc, v. 58, p. 10010, 1995. 21

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. 2016. 25, 26, 28, 29, 30, 31

HUETE, A. et al. Overview of the radiometric and biophysical performance of the modis vegetation indices. 2002. Disponível em: <www.elsevier.com/locate/rse>. 21

JAFARBIGLU, H.; POURREZA, A. A comprehensive review of remote sensing platforms, sensors, and applications in nut crops. *Computers and Electronics in Agriculture*, Elsevier B.V., v. 197, 6 2022. ISSN 01681699. 15, 16, 17, 19

JENSEN, J. R. *Remote sensing of the environment : an earth resource perspective*. [S.l.]: Pearson, 2014. 614 p. ISBN 1292021705. 13, 14, 17, 18, 19, 20

LOFTON, J. et al. Estimating sugarcane yield potential using an in-season determination of normalized difference vegetative index. *Sensors (Switzerland)*, v. 12, p. 7529–7547, 6 2012. ISSN 14248220. 33, 35, 42

LU, G. Y.; WONG, D. W. An adaptive inverse-distance weighting spatial interpolation technique. *Computers and Geosciences*, v. 34, p. 1044–1055, 9 2008. ISSN 00983004. 35

MAIA, F. C. de O.; BUFON, V. B.; LEãO, T. P. Vegetation indices as a tool for mapping sugarcane management zones. *Precision Agriculture*, Springer, v. 24, p. 213–234, 2 2023. ISSN 15731618. 33, 41

MALL, R. K. et al. Managing impact of extreme weather events in sugarcane in different agro-climatic zones of uttar pradesh. v. 67, p. 233–250, 2016. 16, 17

MATHRANI, A. et al. Perspectives on the challenges of generalizability, transparency and ethics in predictive learning analytics. *Computers and Education Open*, Elsevier BV, v. 2, p. 100060, 12 2021. ISSN 26665573. 42

MATOS, R. A. da S. *Balanço Energético Nacional*. 2022. 11

MENEGATTI, L. A. A.; MOLIN, J. P. Remoção de erros em mapas de produtividade via filtragem de dados brutos. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.8, p. 126–134, 2004. Disponível em: <http://www.agriambi.com.br>. 34

RAHMAN, M. M.; ROBSON, A. J. A novel approach for sugarcane yield prediction using landsat time series imagery: A case study on bundaberg region. *Advances in Remote Sensing*, Scientific Research Publishing, Inc,, v. 05, p. 93–102, 2016. ISSN 2169-267X. 41

ROUSE, J. W. et al. Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. 1974. 20, 21

TERUEL, D. A.; BARBIERI, V.; FERRARO, L. *Sugarcane leaf area index modeling under different soil water conditions*. 1997. 33-44 p. 34

WILK, M. van der et al. Learning invariances using the marginal likelihood. In: . [S.l.]: Curran Associates Inc., 2018. p. 9960–9970. 36

WOLF, P. R. et al. *Elements of Photogrammetry with Applications in GIS Fourth Edition*. [s.n.], 2014. ISBN 9780071761116. Disponível em: <www.mhprofessional.com.> 17, 18, 19, 20

ZEBALLOS, J. C. et al. Drought risk assessment of sugarcane-based electricity generation in the rio dos patos basin, brazil. *Sustainability (Switzerland)*, MDPI, v. 14, 5 2022. ISSN 20711050. 11

ZHANG, Z.; MOORE, J. C. *Remote Sensing*. Elsevier, 2015. 111-124 p. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/B9780128000663000048>. 17