MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

ANALISYS OF BURNED AREAS IN PARQUE DOS VEADEIROS USING WFI IMAGES

ANTONIO GOMES DE OLIVEIRA JUNIOR

Master's Dissertation Proposal for the Postgraduate Course in Applied Computing, supervised by Prof. Dr. Thales Körting and Dr. Pedro Andrade

URL of the original document:
<http://urlib.net/>

INPE
São José dos Campos
Maio de 2025

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

ANALISYS OF BURNED AREAS IN PARQUE DOS VEADEIROS USING WFI IMAGES

ANTONIO GOMES DE OLIVEIRA JUNIOR

Master's Dissertation Proposal for the Postgraduate Course in Applied Computing, supervised by Prof. Dr. Thales Körting and Dr. Pedro Andrade

URL of the original document:
<http://urlib.net/>

INPE
São José dos Campos
Maio de 2025

Cataloging in Publication Data

Sobrenome, Nomes.

Cutter ANALISYS OF BURNED AREAS IN PARQUE DOS VEADEIROS USING WFI IMAGES / Nome Completo do Autor1; Nome Completo do Autor2. – São José dos Campos : INPE, Maio de 2025.

xii + 25 p. ; ()

Dissertação ou Tese (Mestrado ou Doutorado em Nome do Curso) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, AAAA.

Orientador : José da Silva.

1. Palavra chave. 2. Palavra chave 3. Palavra chave. 4. Palavra chave. 5. Palavra chave I. Título.

CDU 000.000

Informar aqui sobre fontes financiadoras (a modificação desta linha deve ser feita no arquivo publicacao.tex).

ATENÇÃO! A FOLHA DE APROVAÇÃO SERÁ INCLUIDA POSTERIORMENTE.

Mestrado ou Doutorado em Nome do Curso

ABSTRACT

Wildfires are a significant environmental issue, especially in fire-prone regions such as the Brazilian Cerrado. Remote sensing data is crucial for burnt area mapping and monitoring, yielding valuable information for fire prevention and management. The objective of this study is to create machine learning models, more specifically Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) architectures, for the classification of burned areas from a time series data set gathered from the WFI sensor onboard the CBERS-4, CBERS-4A, and AMAZONIA-1 satellites. Based on spectral indices such as BAI, EVI, NDVI, and NDWI, we will construct classification models with the aim of determining the probability of burned area events in various areas. The process is proposed to involve the preprocessing of data sets, feature engineering, training, and testing based on precision, recall, and precision measures. The outcome is foreseen as a more robust classification model for burned area to improve wildland fire tracking and environmental management strategies.

Palavras-chave: Machine learning. Cerrado. Wildfirest. Remote Sensing. WFI.

ANÁLISE DE ÁREAS QUEIMADAS NO PARQUE DOS VEADEIROS USANDO IMAGENS WFI

RESUMO

Os incêndios florestais são um problema ambiental significativo, especialmente em regiões propensas a incêndios, como o Cerrado brasileiro. como o Cerrado brasileiro. Os dados de sensoriamento remoto são cruciais para o mapeamento e monitoramento de áreas queimadas e monitoramento de áreas queimadas, produzindo informações valiosas para a prevenção e o gerenciamento de incêndios. O objetivo deste estudo é criar modelos de aprendizado de máquina, mais especificamente Redes Neurais Recorrentes (RNN). (RNN) e arquiteturas de memória longa de curto prazo (LSTM), para a classificação de áreas queimadas a partir de um conjunto de dados de séries temporais coletados do sensor sensor WFI a bordo dos satélites CBERS-4, CBERS-4A e AMAZONIA-1. Com base em índices espectrais, como BAI, EVI, NDVI e NDWI, construiremos modelos de classificação com o objetivo de determinar o grau de queimadas. modelos de classificação com o objetivo de determinar a probabilidade de eventos de áreas queimadas em várias áreas. Propõe-se que o processo envolva o pré-processamento de conjuntos de dados, engenharia de recursos, treinamento e testes com base em medidas de precisão, recuperação e exatidão. e precisão. O resultado é previsto como um modelo de classificação mais robusto para a área queimada queimadas para melhorar o rastreamento de incêndios florestais e as estratégias de gerenciamento ambiental

Palavras-chave: Aprendizado de maquina. Cerrado. Queimada. Sensoriamento Remoto. WFI.

LIST OF FIGURES

Page

CONTENTS

Page

1 Introduction

1.1 Motivation

Wildfires have significant environmental, economic, and social impacts worldwide (KALOGIANNIDIS et al., 2023). The increasing frequency and intensity of these events demand enhanced monitoring and classification strategies. Satellite Remote Sensing provides continuous, large-scale observations that support the detection and analysis of burned areas (SZPAKOWSKI; JENSEN, 2019). The *Wide Field Imaging Camera* (WFI) sensor aboard the CBERS-4, CBERS-4A, and AMAZONIA-1 satellites offers high temporal resolution data, which can be effectively used to analyze fire dynamics and develop classification models (ARRUDA et al., 2024).

The Brazilian Cerrado is recognized as one of the world's most diverse savanna ecosystems. Wildfires are a natural component of this biome, playing a crucial ecological role by exerting a selective pressure that has let to the development of fire-resistant plant species through various morphological and physiological adaptations (NASCIMENTO, 2001).

Natural fire cycles in the Cerrado typically occur every three to six years, depending on specific regional factors (BARROSO; PIVELLO, 2000; JÚNIOR et al., 2014). However, despite this long-standing ecological relationship, the rising frequency of human-induced fires poses a serious threat to the ecosystems of the Cerrado (DURIGAN et al., 2020). These anthropogenic fires typically occur at the inappropriate times, persist for extremely long periods, and reach higher intensities, endangering the biodiversity of the region. The native flora of the Cerrado is not adapted to resist such frequent or intense fires, leading to severe ecological degradation (FIDELIS et al., 2018).

Deep learning techniques, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have shown remarkable success in processing time series data. By applying these methods to spectral indices derived from satellite imagery, we aim to develop a robust model capable of classifying the ocurrence of burned areas (ALENCAR et al., 2022). The results of this study may contribute to fire management initiatives, supporting decision-making processes related to environmental conservation and climate change mitigation.

1.2 Statement of Research Problem

The objective of this research is to develop a model to classify burnt areas using a time series from the WFI dataset, using Brazil's own satellites, to answer the following questions:

- How effectively can Deep Learning models classify burned areas using WFI time series data, which does not have access to the SWIR band often used in the the literature (KEY; BENSON, 2006)?
- What are the most relevant attributes of the data set that will be used in the training of the classification model? What are the best indices possible, among the available bands, to get an accurate burned area classification?
- Can temporal patterns in the time series data positively influence the classification results?

1.3 The connection with INPE's mission

The INPE's mission is:

"*Develop, operate and use space systems to advance science, technology and applications in the areas of outer space and the terrestrial environment, and offer innovative products and services for the benefit of Brazil*[1]."

This research aims to leverage the resources available from Brazil's own satellites to more accurately classify burnt areas in Cerrado, using our own resources to develop research that can impact environmental policies and monitoring of the area. INPE and other organizations, have some projects aimed at fire detection, prevention and classification that serves as inspiration for this project:

1.3.1 Queimadas Project

The Queimadas Program, which was developed by Brazil's National Institute for Space Research (INPE), has monitored vegetation fires within the national boundary since 1986. Satellite imagery detects thermal anomalies at the 3.7–4.1 µm band, which are associated with active fire outbreaks detection in almost real-time. These

[1]Source: Plano... (2022) Original Text: "Produzir ciência e tecnologia, operar sistemas, formar pessoas e oferecer produtos e serviços singulares e soluções inovadoras nas áreas do espaço exterior e do sistema terrestre, para o avanço e a difusão do conhecimento e o desenvolvimento sustentável, em benefício do Brasil e do mundo."

outbreaks represent specific locations with concentrated burning, allowing authorities to track fire activity with high spatial and temporal resolution using data from multiple satellites. This information is crucial for environmental management, research, and policy-making. By offering open-access fire data, the program supports efforts to analyze fire frequency, intensity, and patterns, helping to understand ecological impacts and guide mitigation strategies (GARCIA, 2020).

1.3.2 MapBiomas Project

To categorize burned areas throughout Brazil, the MapBiomas Fire Project uses a methodology that has been calibrated differently for each region. To increase the precision of the classification and detection of burned area, the nation is separated into 28 classification regions according to biomes and climatic factors. Burned pixels are identified by their low Normalized Burn Ratio (NBR) values, and annual mosaics from Landsat imagery (1985–2023) are used. Using spectral bands like red, near-infrared (NIR), and shortwave infrared (SWIR), a Multi-Layer Perceptron Neural Network (MLPN) is trained on manually labeled burned and unburned areas. Spatial filtering is used in post-processing to eliminate noise and fill in gaps, while masking is used to eliminate confusing land covers such as urban areas and water. Datasets like MODIS MCD64A1 and INPE fire hotspots are used to validate the results, allowing for consistent fire dynamics monitoring across Brazil biomes (ALENCAR et al., 2024).

2 Background

This chapter provides the theoretical foundation for this work. We first present the concept of artificial intelligence and machine learning, followed by a focus on neural networks. Then we describe the modeling of classification of burned areas using satellite image time series.

2.1 Artificial Intelligence and Machine Learning

The study of artificial intelligence (AI) began in the 1950s as a way to perform tasks traditionally handled better by humans, especially those lacking efficient computational solutions using conventional algorithms. "Intelligence"in the context of AI is characterized by a rational agent: to be rational, an agent must autonomously make decisions based on environmental perception and adapt to achieve optimal results (RUSSELL; NORVIG, 2004).

For a long time, AI was seen as a theoretical field with limited real-world application. In the 1970s, expert systems emerged to solve real-world problems, but these relied heavily on domain experts to encode knowledge using logical rules (MONARD; BARANAUSKAS, 2003).

As computational problems grew more complex, machine learning (ML) techniques emerged to reduce reliance on expert intervention, enabling systems to learn from past data (FACELI et al., 2011).

2.1.1 Key Concepts in Machine Learning

To understand this work, it is essential to define some key machine learning terms (MONARD; BARANAUSKAS, 2003):

- **Sample**: An instance in the dataset, such as a satellite observation.

- **Class**: The output label to be learned or classified, e.g., "burned"or "unburned"area.

- **Noise**: Errors or inconsistencies in the dataset, such as mislabeled images.

- **Bias**: The algorithm's tendency to favor one hypothesis over others.

2.1.2 Types of Machine Learning

These models learn by minimizing an objective function, typically a loss or error function (FACELI et al., 2011). ML algorithms can be categorized as supervised, unsupervised, or semi-supervised (FACELI et al., 2011), (WITTEN; FRANK, 2005):

- **Supervised Learning**: The model learns from labeled data. Classification of burned areas is an example, where each satellite image is labeled.
  - *Regression*: classify continuous values.
  - *Classification*: classify discrete classes.
- **Unsupervised Learning**: Finds patterns without labeled outputs (e.g., clustering vegetation types).
- **Semi-supervised Learning**: Uses a mix of labeled and unlabeled data.

2.1.3 Sampling, Training, and Validation

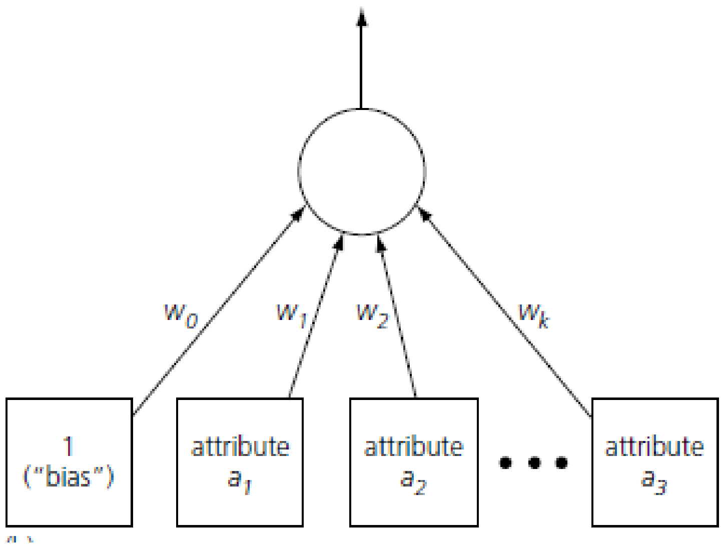
n supervised learning, datasets are typically split into training and testing sets. The algorithm is trained using the training set, a sample of the total dataset, and it attempts to predict the label of each element in the testing set based on its attributes. The results are then evaluated by comparing the predictions with the test set, using different metrics depending on the type of machine learning method used (MONARD; BARANAUSKAS, 2003).

2.1.4 Machine Learning Models

2.1.5 Perceptron and Multi-layer Perceptrons (MLP)

Neural networks are the name of a set of machine learning techniques based on optimization, inspired by the structure and functioning of the human nervous system. The perceptron is the simplest type of neural network, consisting of a single layer of neurons. Each neuron has a threshold activation function that returns an output depending on the input. It iterates over the training set elements until it can adjust the weights of the neurons to generate a hyperplane that solves the problem. Despite its simplicity, a perceptron network performs well in classifying linear problems—that is, problems whose outcomes can be classified using a hyperplane (WITTEN; FRANK, 2005). Figure 2.1 shows a visual representation of the perceptron.

Figure 2.1 - Perceptron.



Source: (WITTEN; FRANK, 2005)

The perceptron network is trained using a supervised error correction algorithm. During network training, the weights w for an object x_i are adjusted according to the Equation 2.1 (FACELI et al., 2011):

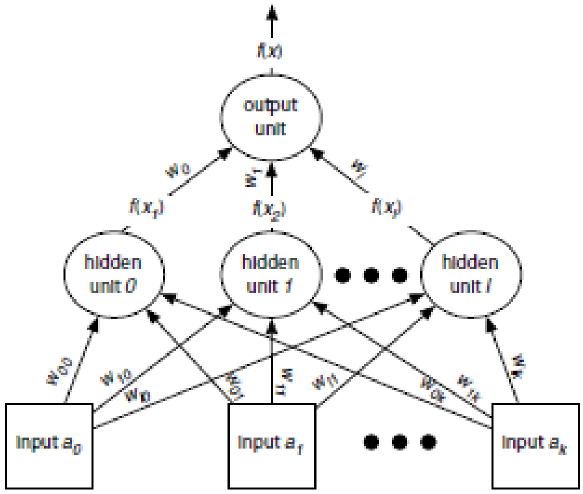
$$w_j(t+1) = w_j(t) + \eta a_i^j (y_i + f(a_i)) \tag{2.1}$$

Where $w_j(t)$ is the weight of the jth input connection at time instant t, η is a learning rate, x_j^i is the value of attribute j of input value x_i, and $f(x)$ is the output produced by the network at time instant t and y_i is the correct output for the network according to the label of x_i.

To address non-linear problems with neural networks, it is necessary to create one or more intermediate layers of hidden units called multi-layer perceptrons (MLP). These networks are typically fully connected, meaning that each perceptron in one layer is connected to all perceptrons in the next layer. In this case, it is also necessary to use non-linear activation functions, such as the sigmoid activation function (FACELI et al., 2011). Figure 2.2 shows a representation of a multi-layer perceptron network with one hidden layer.

A sigmoid function, used as the activation function in the MLP model, is defined

Figure 2.2 - Multi-layer perceptron with one hidden layer.



Source: (WITTEN; FRANK, 2005)

by:

$$f(x) = \frac{1}{1 + e^{(-x)}} \tag{2.2}$$

Finally, if we analyze an MLP with a single hidden layer, the output of the hidden layer can be represented as: $H = \phi(XW_{xh} + b_h)$, where X is a set of n examples with d input values each, ϕ is the hidden layer activation function, h is the number of hidden units in a hidden layer, W_{xh} are the weights applied to the input values, and b_h is the bias value.

The formula used to calculate the output value O for the output layer can be defined as: $O = HW_{hq} + b_q$, where W_{hq} are the weights and b_q is the bias value of the output layer (ZHANG et al., 2021).

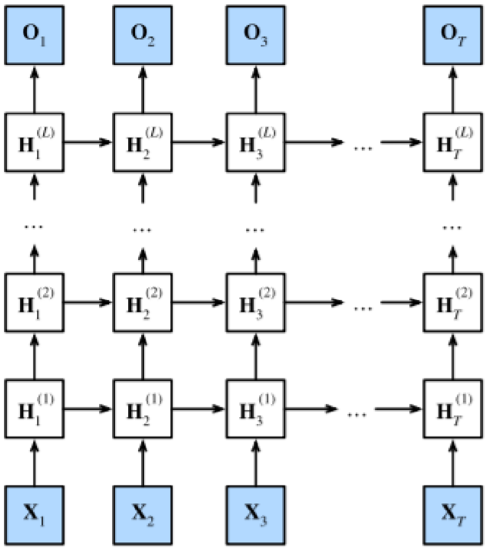
2.1.6 Deep Learning

Deep learning models, such as MLPs with many hidden layers, learn hierarchical representations of data. This enables them to capture complex patterns necessary for tasks like burned area detection in satellite imagery (GOODFELLOW et al., 2016).

2.1.7 Recurrent Neural Networks (RNN)

RNNs, or recurrent neural networks, are neural networks that, in addition to hidden layers, also have hidden states. Hidden layers are layers between the input and output layers. Hidden states, however, are input values that can be used by neurons during a time step, computed using data from previous time steps. LSTMs are a specific type of RNN (ZHANG et al., 2021).

Figure 2.3 - Architecture of a deep RNN.



Source: (ZHANG et al., 2021)

In Figure 2.3, the hidden layers are represented by the layers H_i, and the hidden states are shown as the horizontal arrows between neurons in the same hidden layer.

LSTM models, a type of RNN model, solve three weaknesses of standard RNN models: (ZHANG et al., 2021)

- Introduce a way to "remember"past data when it is relevant;
- Introduce a way to "forget"past data when it is no longer relevant;
- Introduce a way to "reset"memory to forget data that has become irrelevant.

2.1.8 Long Short-Term Memory (LSTM)

The design of LSTMs is inspired by logic gates in a computer. The LSTM network introduces a memory cell, designed to record additional information (ZHANG et al., 2021). This allows the neural network cells to have "short-term memory"of past states to help in the learning process. This memory can also be cleared and weighted during training by the network itself.

The main idea of the LSTM architecture is that cycles in a neuron act as feedback paths, and the weights of these paths can be adjusted. By making these weights controllable by another hidden unit, the time intervals for which data is remembered can be dynamically adjusted depending on the input (BENGIO YOSHUA; COURVILLE, 2016).

To control the behavior of each memory cell—how information is stored and manipulated—a set of three gates is used: the input gate, forget gate, and output gate. The inputs to an LSTM cell are the input value at the current time step X_t and the hidden state from the previous time step H_{t-1}. These inputs are processed by fully connected layers with sigmoid activation functions to compute the values of the gates. As a result, the outputs of these gates range between 0 and 1 (ZHANG et al., 2021).

- **Forget Gate:** Decides how much of the previous memory cell state should be kept. It determines what information to discard from the cell state.
  - Inputs X_t and H_{t-1} are multiplied by weight matrices, added to biases, and passed through a sigmoid activation.
  - The resulting values are multiplied by the previous cell state C_{t-1} to determine how much of it to retain.
- **Input Gate:** Determines what new information should be stored in the memory cell.
  - Inputs X_t and H_{t-1} go through a similar process using sigmoid and tanh functions.
  - The result defines how much of the candidate values generated by tanh will be added to the memory state.
- **Output Gate:** Defines what part of the memory cell will be output as the new hidden state H_t.
  - Applies sigmoid to the weighted inputs X_t and H_{t-1}, and tanh to the new cell state C_t.
  - These are multiplied to form the new hidden state H_t.

The cell state C_t is updated using:

$$C_t = F_t \cdot C_{t-1} + I_t \cdot \tilde{C}_t$$

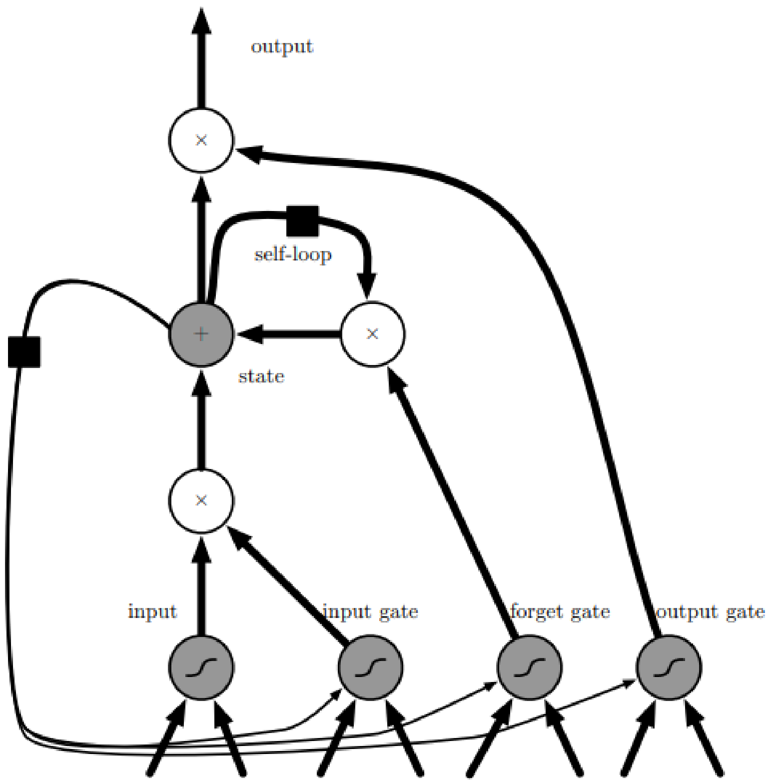
where F_t is the forget gate output, I_t is the input gate output, and $\tilde{C}_t$ is the candidate memory.

Then, the hidden state H_t is given by:

$$H_t = O_t \cdot \tanh(C_t)$$

where O_t is the output gate value.

Figure 2.4 - Diagram of LSTM.



The diagram shown in Figure 2.4 shows how the final output value of an LSTM is calculated. In the diagram, the squares mean a delay of one time step. The input is handled by a perceptron with a sigmoidal activation function at the input, and the result of the perceptron can be written to the state cell if allowed by the activation function. Another observation that can be obtained is the linear cycle of the state cell with weight controlled by the *forget gate*. In the end, the output is defined by the value in the memory state together with the value of the output gate, which can disable the output.

In general, LSTM artificial neural networks have performed well in time series processing, speech recognition, handwritten text recognition and automated translation.

2.1.9 Convolutional Neural Networks (CNN)

CNNs, or Convolutional Neural Networks, are effective for image processing tasks. They detect spatial hierarchies of features and are commonly used in Remote Sensing for classifying satellite images, including burned area detection (BENGIO YOSHUA; COURVILLE, 2016). CNNs are a typw of feed-forward neural network, meaning that the connections between nodes do not form cycles, as in recurrent networks. CNNs are specialized in processing data with a grid-like topology, such as images or time series, and have shown excellent results in solving problems of this nature.

Images have many statistical properties that are invariant to translation. For instance, a photo of a cat remains a photo of a cat if it is shifted one pixel to the right. CNNs exploit these properties by sharing parameters across different spatial locations. That is, the same hidden unit with the same weights can be used to process various regions of the input.

This means the same pig detector can identify a pig whether it appears in column i or column $i + 1$ of the image. This parameter sharing drastically reduces the number of model parameters and allows the network to scale up without requiring a proportionate increase in training data. As a result, CNNs can effectively incorporate domain knowledge into the architecture to aid the learning process (BENGIO YOSHUA; COURVILLE, 2016).

CNNs leverage the concept of spatial invariance to learn useful representations with fewer parameters. According to (ZHANG et al., 2021), some principles that guide the design of neural architectures for computer vision are:

- In early layers, the network should respond similarly to the same pattern in the input set, regardless of where it appears in the image. This principle is called translation invariance.

- Early network layers should focus on local regions, without considering the entire image. This is the locality principle. Eventually, these local insights can be combined to make image-wide predictions.

To achieve this, CNNs apply a mathematical operation called convolution (Equation 2.3) in at least one of their layers. The convolution process involves:

a) Passing input values to the network.

b) Applying a kernel function to the inputs.

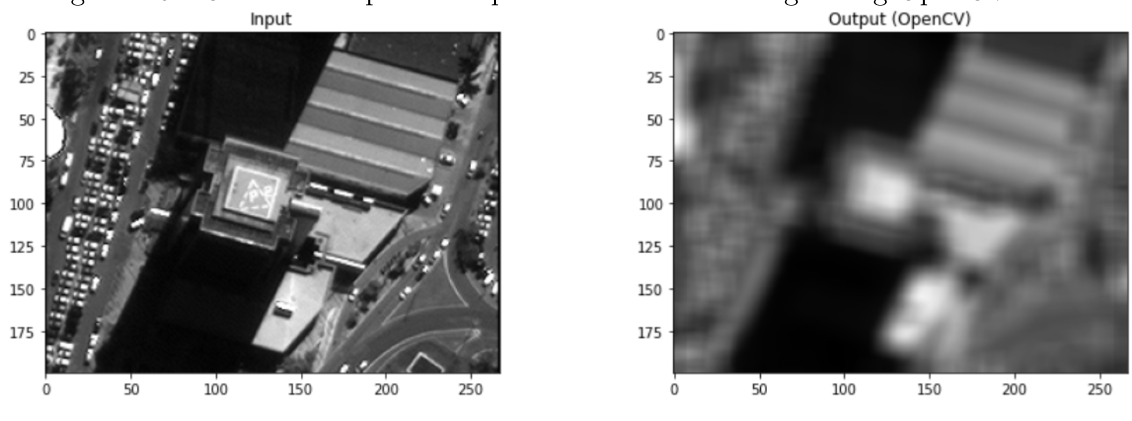
c) Producing a feature map as output.

The resulting feature map is then passed to a neural network. It represents a set of patterns recognized in the input attributes and often significantly reduces the dimensionality of the input, improving the network's efficiency (BENGIO YOSHUA; COURVILLE, 2016).

$$S(i,j) = \sum_m \sum_n I(m,n) * K(i-m, j-n) \tag{2.3}$$

Where I is the input, K is the kernel function, and S is the output feature map. This convolution process is inspired by how the human brain processes images—initial layers recognize simple patterns, and sequent layers build on those to recognize more complex structures until the image is interpreted as a combination of learned features. An example of this process is shown in Figure 2.5:

Figure 2.5 - Convolution process representation on an image using OpenCV.



2.2 Remote Sensing and Time Series Analysis for Burned Area Mapping

Remote Sensing is extensively used in burned area detection and monitoring, owing to the capacity of satellites for frequent and broad coverage. The WFI sensor acquires spectral data in blue, green, red, and near-infrared wavelengths, from which vegetation and fire-related indices can be calculated. A major limitation in burned area mapping from WFI data is the lack of the Shortwave Infrared (SWIR) band, precluding the utilization of the Normalized Burn Ratio (NBR) index—a very common spectral index for fire mapping. The SWIR band plays an essential role in the discrimination between burned and unburned vegetation since it is sensitive to vegetation moisture content changes and charred material (KEY; BENSON, 2006). In the absence of this spectral band, surrogate approaches must use visible and near-infrared spectral indices, including the Burned Area Index (BAI) and the Normalized Difference Vegetation Index (NDVI). These indices have decreased performance in discriminating burned surfaces from some types of soil or water bodies, and the classification becomes problematic, necessitating extra validation procedures (MARTÍN et al., 2006).

2.2.1 Time Series Analysis in Fire Monitoring

Time series data analysis allows for the determination of patterns between fire incidents for time periods. Several spectral indices, including the Normalized Difference Vegetation Index (NDVI), Burned Area Index (BAI), and Normalized Difference Water Index (NDWI), are widely utilized for burned area detection and vegetation recovery assessment following fire incidents (LIU et al., 2021).

2.2.2 Trend, Seasonality, and Cycle

To facilitate discussion on time series, it is helpful to define terms such as trend, seasonality, and cycles (HYNDMAN; ATHANASOPOULOS, 2018).

- Trend: A trend exists when there is a long-term increase or decrease in the time series data.
- Seasonality: A seasonal pattern occurs when the time series is affected by seasonal factors, for example, specific months in a year—that recur at a known regular frequency.
- Cycle: A cycle is observed when the data show rises and falls over a period of time that does not follow a fixed frequency. For instance, changes due

to economic crises.

- Outliers and Spurious Data: Outliers can be defined as instances or observations with extreme values compared to the majority of the dataset. One possible cause of outliers is spurious data—i.e., data entered incorrectly or with errors, which is not uncommon when dealing with real-world data.

2.2.3 Machine Learning and Deep Learning in Fire classification

Machine learning approaches have been extensively used in burned area classification in Remote Sensing research. Both global fire map data, like the MODIS MCD64A1 product, and regional programs, such as MapBiomas, have offered comprehensive insights on fire cycles by employing state-of-the-art machine learning algorithms. Machine Learning and Deep Learning for Fire classification Traditional classification techniques such as Random Forest have been used for burnt area mapping (CHUVIECO et al., 2018) (ALENCAR et al., 2022). Deep learning techniques such as RNNs and LSTMs can process sequential data and detect trends and may be suitable to model temporal dependencies in wildfire analysis. Similar papers, where ML techniques and satellite images are used for burned area mapping include:

2.2.3.1 Evaluation Metrics

Accuracy alone may be misleading for imbalanced data (e.g., most areas are unburned). Common metrics include:

- **Accuracy**: Proportion of classifyed burned areas that are correct.
- **Recall**: Proportion of actual burned areas correctly identified.
- **F1-score**: Harmonic mean of precision and recall.
- **Confusion Matrix**: Summarizes classification outcomes.

2.2.3.2 Burned Area Detection in the Brazilian Amazon using Spectral Indices and GEOBIA

The study "Burned Area Detection in the Brazilian Amazon using Spectral Indices and GEOBIA"from Penha et al. (2020) proposed a methodology to detect burned areas in the Brazilian Amazon using the integration of spectral indices and Geographic Object-Based Image Analysis (GEOBIA). Using medium-resolution data from Landsat-8 OLI and Sentinel-2A MSI images, the authors evaluated nine spectral indices to distinguish between burned and unburned regions and concluded that the Burned Area Index (BAI) was the most effective index. For more precise classification, a GEOBIA model segmented the imagery into objects and applied rule-based classification, improving accuracy—especially for small burned areas (<1 km^2) that pixel-based approaches tend to overlook.

The results show the feasibility of using GEOBIA and medium-resolution satellite imagery for operational fire monitoring. By incorporating object-level spatial features and post-fire spectral response, the method is a more stable alternative to traditional threshold-based classification. The research highlights the need for tailoring burned area detection methods to local vegetation processes and image conditions, particularly in complex and cloud-prone environments like the Amazon rainforest.

2.2.3.3 Burned Area Mapping in the Brazilian Savanna Using a One-Class Support Vector Machine Trained by Active Fires

The paper "Burned Area Mapping in the Brazilian Savanna Using a One-Class Support Vector Machine Trained by Active Fires"from Pereira et al. (2019) suggested a new approach to map burned areas within the Brazilian savanna (Cerrado) with a One-Class Support Vector Machine (OC-SVM) trained only with active fire data. This approach employs MODIS satellite imagery for burned area identification, avoiding the issue of insufficiently labeled datasets in fire-risk zones. With its focus on active fire pixels as the training dataset, the OC-SVM is able to effectively distinguish between burned areas and the rest of the landscape with minimal ground truth data.

The study demonstrated that the OC-SVM approach was more accurate in burned area detection than traditional thresholding algorithms. The outcome also cross-verifies with the reference data, and the model's performance was indicated at various environmental conditions. In this paper, the importance of machine learning approaches, with specific interest in one-class classifiers, is also emphasized as a tool to enhance burned area detection and monitoring using satellite imagery, particularly in fire-prone ecosystems like the Cerrado.

3 Methods

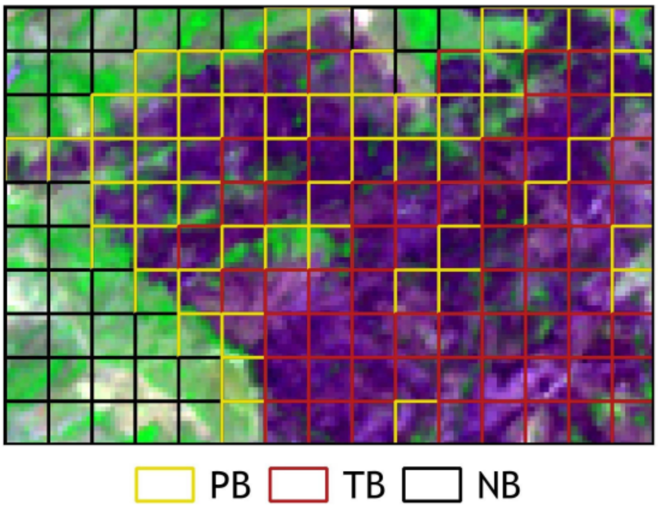
3.1 Dataset

The dataset used in this study is derived from the paper "A multi-temporal dataset for mapping burned areas in the Brazilian Cerrado using time series of remote sensing imager"from Oliveira and KÖrting (2025), this paper proposes a multi-temporal dataset to track burned patches in the Brazilian Cerrado, specifically in Chapada dos Veadeiros National Park, using time series data from satellite images. The dataset contains years 2020 to 2022 and merges images from the WFI sensor aboard the CBERS-4, CBERS-4A, and AMAZONIA-1 satellites. The temporal density of the data varies across years and months due to the integration of images from multiple satellite platforms and the exclusion of scenes affected by cloud cover. As a result, certain years and months contain significantly more observations than others. These images have four spectral bands (blue, green, red, and near-infrared) and five spectral indices (BAI, EVI, GEMI, NDVI, NDWI), which were assembled into a uniform grid of 500 m $\times$ 500 m cells along with the classification for each, manually assigned by a specialist by analyzing each satelite image (OLIVEIRA; KÖRTING, 2025). 235 cloud-free images were used after preprocessing, and the dataset was then enhanced further through atmospheric correction, co-registration, computation of spectral indices, and duplication to compensate for non-uniform sampling in time.

The author, them used the dataset to train a Random Forest (RF) algorithm using three different strategies: (1) binary classification where totally burned (TB) areas are labeled as burned and partially burned (PB) and non-burned (NB) as unburned; (2) grouping both TB and PB as burned; and (3) using all three classes separately. The RF classifier was optimized using GridSearchCV and validated through 200 Monte Carlo simulations per year. Results indicated that the best-performing strategy was to train with all three classes and then reclassify TB as burned while grouping PB and NB as non-burned (the fifth validation approach). This yielded the highest Intersection over Union (IoU) scores and consistent performance across all years, particularly in 2021.

Each instance in the dataset classified as Totally Burned (TB), Partially Burned (PB) or Not Burned (NB) as in the Figure 3.1:

Figure 3.1 - Burned Region Example.



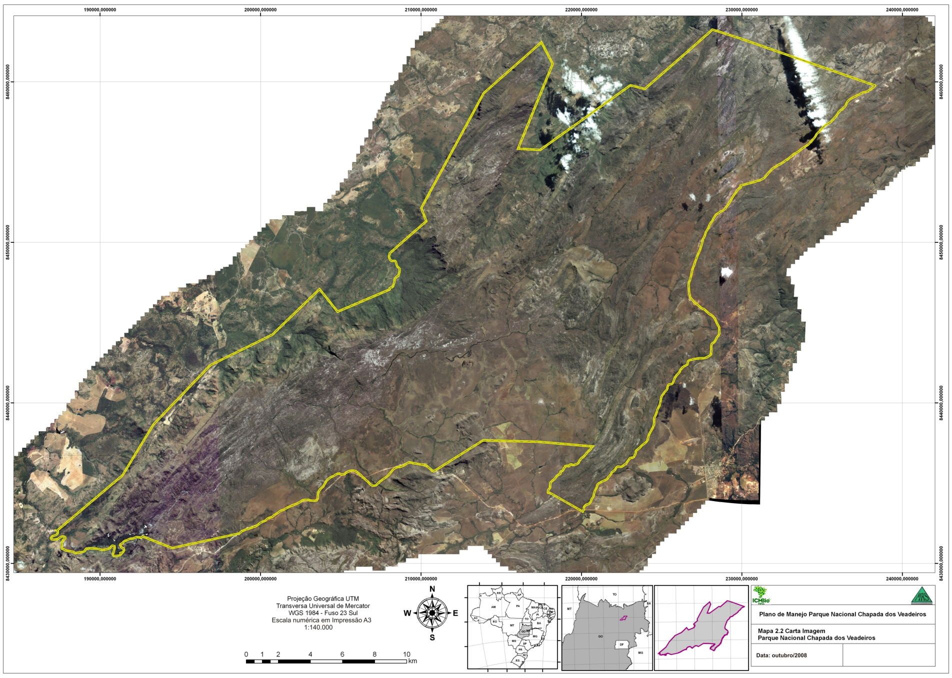
Source: : Oliveira and Körting (2025)

During this master dissertation, the student proposes to calibrate a random forest model and other deep learning models on this dataset using different techniques and try to optimize the results.

3.1.1 Study Area: Chapada dos Veadeiros National Park

The Chapada dos Veadeiros National Park, located in the state of Goiás, Brazil, is part of the Cerrado biome, one of the most biodiverse savanna ecosystems in the world. Covering approximately 240,000 hectares, the park is characterized by its rugged terrain, high-altitude plateaus, seasonal climate, and diverse vegetation types ranging from grasslands to dense shrublands and gallery forests. Due to its ecological importance and susceptibility to seasonal wildfires, the park represents a critical area for monitoring fire dynamics. Its well-documented fire history and diverse landscape make it an ideal region for applying machine learning techniques, such as neural networks, to classify and analyze burned areas using satellite imagery.

Figure 3.2 - Map of study area.



Source: Instituto Chico Mendes

3.2 Modeling Burned Area Classification

The task involves classifying each pixel or region in satellite images based on whether it corresponds to a burned area. This is framed as a supervised classification problem, where labeled satellite images (e.g., from fire incident maps or expert annotation) are used to train the model.

Key steps include:

a) Data preprocessing (e.g., resampling, normalization);

b) Baseline model training for comparison (Random forest with spectral indices);

c) Researching and selecting best indices for the models according to available data;

d) Model training and calibration using CNNs, LSTMs, or hybrid models;

e) Model evaluation using metrics like accuracy, precision, recall, and F1-score.

4 Schedule of Activities

This section presents the schedule of activities to be carried out during the master's program. The activities include the main steps required to develop the research, from literature review to dissertation defense.

The schedule is shown in Table 4.1, covering a 12-month period from March 2025 (month 3) to February 2026 (month 2). The activities labeled from **a** to **g** correspond to the following:

a) Review of the literature: researching similar projects, techniques, and vegetation indices;

b) Dataset pre-processing;

c) Testing and calibrating a baseline Random Forest model;

d) Testing Deep Learning models such as LSTM and RNN;

e) Compilation of results and writing/submission of a scientific paper;

f) Writing of the dissertation;

g) Dissertation defense.

Table 4.1 - Schedule of activities.

Activities	Months (2025–2026)											
	3	4	5	6	7	8	9	10	11	12	1	2
a												
b												
c												
d												
e												
f												
g												

REFERENCES

ALENCAR, A. A.; CONCIANI, D. E.; ROSA, E. R.; MARTIN, E. V.; ANDRADE, G.; HASENACK, H.; MARTENEXEN, L. F. M.; RIBEIRO, J. P. F. M.; SHIMBO, J.; ROSA, M.; DIAS, M.; CRUSCO, N.; SANTOS, N.; MONTEIRO, N. C.; DUVERGER, S. G.; AZEVEDO, T.; PIONTEKOWSKI, V. J.; ARRUDA, V. L. D. S.; SILVA, W. V. D.; ROCHA, W. D. F. **MapBiomas Fire Brazil Collection 3: Annual Burned Area Maps of Brazil (1985-2023). Algorithm Theoretical Basis Document (ATBD)**. MapBiomas Data, 2024. Available from: <<https://data.mapbiomas.org/citation?persistentId=doi:10.58053/MapBiomas/0KJBRA>>. 3

ALENCAR, A. A. C.; ARRUDA, V. L. S.; SILVA, W. V. d.; CONCIANI, D. E.; COSTA, D. P.; CRUSCO, N.; DUVERGER, S. G.; FERREIRA, N. C.; FRANCA-ROCHA, W.; HASENACK, H.; MARTENEXEN, L. F. M.; PIONTEKOWSKI, V. J.; RIBEIRO, N. V.; ROSA, E. R.; ROSA, M. R.; SANTOS, S. M. B. dos; SHIMBO, J. Z.; VéLEZ-MARTIN, E. Long-term landsat-based monthly burned area dataset for the brazilian biomes using deep learning. **Remote Sensing**, v. 14, n. 11, 2022. ISSN 2072-4292. Available from: <<https://www.mdpi.com/2072-4292/14/11/2510>>. 1, 15

ARRUDA, V. L. da S.; ALENCAR, A. A. C.; JúNIOR, O. A. de C.; RIBEIRO, F. de F.; ARRUDA, F. V. de; CONCIANI, D. E.; SILVA, W. V. da; SHIMBO, J. Z. Assessing four decades of fire behavior dynamics in the cerrado biome (1985 to 2022). **Fire Ecology**, v. 20, n. 1, p. 64, Jul 2024. ISSN 1933-9747. Available from: <<https://doi.org/10.1186/s42408-024-00298-4>>. 1

BARROSO, M.; PIVELLO, V. Lightning fires in a brazilian savanna national park: Rethinking management strategies. **Environ Management**, v. 26, p. 675–684, 01 2000. 1

BENGIO YOSHUA; COURVILLE, A. G. I. J. **Deep learning: adaptive computation and machine learning**. The MIT Press, 2016. (Adaptive Computation and Machine Learning series). ISBN 0262035618,9780262035613. Available from: <<http://gen.lib.rus.ec/book/index.php?md5=f3d21ce4de6685496ac5a07721b98d26>>. 10, 12, 13

CHUVIECO, E.; LIZUNDIA-LOIOLA, J.; PETTINARI, M. L.; RAMO, R.; PADILLA, M.; TANSEY, K.; MOUILLOT, F.; LAURENT, P.; STORM, T.; HEIL, A.; PLUMMER, S. Generation and analysis of a new global burned area product based on modis 250 m reflectance bands and thermal anomalies. **Earth System Science Data**, v. 10, n. 4, p. 2015–2031, 2018. Available from: <<https://essd.copernicus.org/articles/10/2015/2018/>>. 15

DURIGAN, G.; PILON, N. A. L.; ABREU, R. C. R.; HOFFMANN, W. A.; MARTINS, M.; FIORILLO, B. F.; ANTUNES, A. Z.; CARMIGNOTTO, A. P.;

MARAVALHAS, J. B.; VIEIRA, J.; VASCONCELOS, H. L. No net loss of species diversity after prescribed fires in the brazilian savanna. **Frontiers in Forests and Global Change**, v. 3, 2020. ISSN 2624-893X. Available from: <<https://www.frontiersin.org/journals/forests-and-global-change/articles/10.3389/ffgc.2020.00013>>. 1

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. d. L. F. d. **Inteligência artificial: uma abordagem de aprendizado de máquina**. [S.l.]: LTC, 2011. 5, 6, 7

FIDELIS, A.; ALVARADO, S. T.; BARRADAS, A. C. S.; PIVELLO, V. R. The year 2017: Megafires and management in the cerrado. **Fire**, v. 1, n. 3, 2018. ISSN 2571-6255. Available from: <<https://www.mdpi.com/2571-6255/1/3/49>>. 1

GARCIA, M. P. **How to monitor forest fires — revistapesquisa.fapesp.br**. 2020. <https://revistapesquisa.fapesp.br/en/how-to-monitor-forest-fires-2/>. [Accessed 23-05-2025]. 3

GOODFELLOW, I. J.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MA, USA: MIT Press, 2016. <http://www.deeplearningbook.org>. 8

HYNDMAN, R.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2nd. ed. Australia: OTexts, 2018. 14

JÚNIOR, A. C. P.; OLIVEIRA, S. L. J.; PEREIRA, J. M. C.; TURKMAN, M. A. A. Modelling fire frequency in a cerrado savanna protected area. **PLOS ONE**, Public Library of Science, v. 9, n. 7, p. 1–11, 07 2014. Available from: <<https://doi.org/10.1371/journal.pone.0102380>>. 1

KALOGIANNIDIS, S.; CHATZITHEODORIDIS, F.; KALFAS, D.; PATITSA, C.; PAPAGRIGORIOU, A. Socio-psychological, economic and environmental effects of forest fires. **Fire**, v. 6, n. 7, 2023. ISSN 2571-6255. Available from: <<https://www.mdpi.com/2571-6255/6/7/280>>. 1

KEY, C.; BENSON, N. Landscape assessment: Ground measure of severity, the composite burn index; and remote sensing of severity, the normalized burn ratio. In: ______. [S.l.: s.n.], 2006. p. LA 1–51. 2, 14

LIU, J.; MAEDA, E. E.; WANG, D.; HEISKANEN, J. Sensitivity of spectral indices on burned area detection using landsat time series in savannas of southern burkina faso. **Remote Sensing**, v. 13, n. 13, 2021. ISSN 2072-4292. Available from: <<https://www.mdpi.com/2072-4292/13/13/2492>>. 14

MARTÍN, M.; GóMEZ, I.; CHUVIECO, E. Burnt area index (baim) for burned area discrimination at regional scale using modis data. **Forest Ecology and Management - FOREST ECOL MANAGE**, v. 234, 11 2006. 14

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. In: **Sistemas Inteligentes Fundamentos e Aplicações**. 1. ed. Barueri-SP: Manole Ltda, 2003. p. 89–114. ISBN 85-204-168. 5, 6

NASCIMENTO, I. V. Cerrado: o fogo como agente ecológico. **Territ. Rev. Port. Riscos Prev. Segur.**, Coimbra University Press, n. 8, p. 25–35, sep. 2001. 1

OLIVEIRA, A. C. de; KÖRTING, T. S. **WFI data for mapping burned areas**. [S.l.]: Science Data Bank, feb. 2025. 17

PENHA, T.; KöRTING, T.; FONSECA, L.; SILVA-JUNIOR, C.; PLETSCH, M.; ANDERSON, L.; MORELLI, F. Burned area detection in the brazilian amazon using spectral indices and geobia. **Revista Brasileira de Cartografia**, v. 72, 07 2020. 15

PEREIRA, A. A.; PEREIRA, J. M. C.; LIBONATI, R.; OOM, D.; SETZER, A. W.; MORELLI, F.; MACHADO-SILVA, F.; CARVALHO, L. M. T. de. Burned area mapping in the brazilian savanna using a one-class support vector machine trained by active fires. **Remote Sensing**, v. 11, n. 3, p. 257, 2019. Available from: <<https://www.mdpi.com/2072-4292/11/3/257>>. 16

PLANO Diretor 2022–2026. São José dos Campos, Brasil, 2022. Texto final consolidado pelo Comitê de Coordenação do Planejamento Estratégico (CCPE), com contribuições do Grupo de Planejamento Estratégico (GPE). Available from: <<https://www.gov.br/inpe/pt-br/central-de-conteudo/publicacoes/repositorio-de-arquivos/plano-diretor-2022-2026.pdf>>. 2

RUSSELL, S.; NORVIG, P. **Inteligência artificial**. CAMPUS - RJ, 2004. ISBN 9788535211771. Available from: <<https://books.google.com.br/books?id=wBMvAAAACAAJ>>. 5

SZPAKOWSKI, D. M.; JENSEN, J. L. R. A review of the applications of remote sensing in fire ecology. **Remote Sensing**, v. 11, n. 22, 2019. ISSN 2072-4292. Available from: <<https://www.mdpi.com/2072-4292/11/22/2638>>. 1

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005. ISBN 0120884070. 6, 7, 8

ZHANG, A.; LIPTON, Z. C.; LI, M.; SMOLA, A. J. Dive into deep learning. **CoRR**, abs/2106.11342, 2021. Available from: <<https://arxiv.org/abs/2106.11342>>. 8, 9, 10, 12