



UNIVERSITY OF SALENTO

DEPARTMENT OF INNOVATION
ENGINEERING

Computer Engineering Master Degree Program

MASTER'S DEGREE THESIS

in

HIGH PERFORMANCE COMPUTING

Leveraging Machine Learning for detecting Marine Heatwaves in the Mediterranean sea

Thesis supervisor:

Prof. Giovanni ALOISIO

Graduate student:

Andrea VEDRUCCIO

Thesis supervisor:

Eng. Gabriele ACCARINO, Ph.D.

Academic year 2021/2022

SSD: ING-INF/05

Contents

1	Introduction	2
1.1	Definition and classification of MHWs	2
1.2	Drivers	4
1.3	The most important events of the past 20 years	5
1.4	Impacts on natural, physical, and humans systems	8
2	Background	11
2.1	Convolutional Neural Networks	12
2.2	Convolutional Layer	13
2.3	Pooling Layer	15
2.4	Fully-connected Layer	16
2.5	Non-Linearity Layers	17
3	Related Work	18
4	Experimental Setups	23
4.1	Setup A	23
4.1.1	Climate Data	24
4.1.2	Data Preprocessing	26
4.2	Setup B	29
4.2.1	Climate Data	29
4.2.2	Data Preprocessing	31
4.3	Setup B.2	34
4.3.1	Climate Data	34
4.3.2	Data Preprocessing	34
4.4	Training, Testing and Validation Datasets	36
4.5	Deep learning architecture and procedure	36
4.5.1	Capsule Neural Network architecture	41
4.6	Hardware & Software Resources	43
4.7	Model Evaluations	43
4.7.1	Accuracy	43
4.7.2	Loss functions	44
4.7.3	Confusion Matrix	45

5	Results and Discussions	46
5.1	Architecture 1 - SETUP A	46
5.2	Architecture 2 - SETUP B	46
5.3	Architecture 3 - SETUP B.2	49
6	Conclusions and Future Work	51
7	Sitography	52

Abstract

Due to the impact of extreme heat waves on society and biodiversity, their study is a key challenge. Climate models can be used to predict their occurrence or likelihood. This thesis explores the use of machine learning architectures, trained using climate reanalysis data, as an alternative strategy to predict the occurrence of extreme marine heat waves (MHWs) and identify clusters in the Mediterranean Sea and beyond (Central Mediterranean sea, Western Mediterranean Sea, Eastern Mediterranean Sea, Adriatic Sea, Caspian Sea, Black Sea). Achieving this difficult goal requires addressing issues such as class size imbalance, which is inherently of rare events. Several convolutional neural network were trained, using 35 years of climate reanalysis data results from 1981 to 2016. Starting from different input drivers such as ,surface temperature (T2M), geopotential height fields at 500 hPa (GEO), sea level pressure (SLP), Net air-sea heat flux (NET), Incoming solar radiation (INC), Surface latent heat flux (LAT) , the convolutional networks trained in the different setups and architectures did not achieve significant performance in predicting the occurrence of extreme marine heat waves, but they have offered numerous points of interest and reflection to continue on this path in terms of the heterogeneity of useful data, the size and resolution of the input maps to be used (extending the domain of the Mediterranean Sea to North Africa and the northeast Atlantic, to include the strong influence on the emergence of MHWs of the African and Azores anticyclones, respectively), which drivers offer greater correlation between input and output, and how overfitting problems may arise.

Keywords: *Marine heat wave, Atmospheric dynamics, Extreme event, Prediction, Machine learning,*

Introduction

The study and prediction of extreme Marine Heat Waves (MHWs) is a key challenge for the scientific community because of the strong impact on society and biodiversity. MHWs, when ocean temperatures are extremely warm for an extended period of time, can have significant impacts on marine ecosystems, industries, and society.

Current physics-based weather prediction systems or climate models can be used to predict their occurrence or likelihood. This work explores the use of machine learning architectures, trained using climate reanalysis data, to classify MHW in the Mediterranean Sea.

1.1 Definition and classification of MHWs

Heat waves are known to occur in the atmosphere. However, heat waves can also occur in the seas and oceans and are known as marine heat waves or MHWs.

MHWs can occur in both summer and winter and are defined by differences with the expected temperatures for the location and time of year.

The definition of MHWs developed by [18] is used. A marine heat wave (MHW) occurs when seawater temperatures exceed a seasonally variable threshold (usually the 90th percentile) for at least 5 consecutive days based on a 30-year historical baseline period. Successive heat waves with intervals of 2 days or less are considered part of the same event.

The work "Categorizing and Naming Marine Heatwaves proposes"[17] proposes a categorization scheme for heat waves based on intensity (I), which is the sea surface temperature anomaly (SSTA) based on long-term climatology for a location.

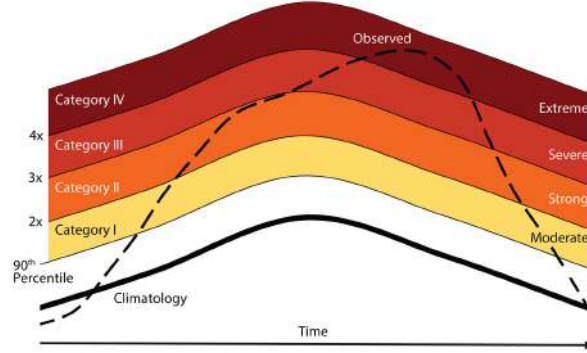


Figure 1.1: Categorization schematic for marine heatwaves (MHWs) showing the observed temperature time series (dashed line), the long-term regional climatology (bold line), and the 90th percentile climatology (thin line). Multiples of the 90th percentile difference ($2\times$ twice, $3\times$ three times, etc.) from the mean climatology value define each of the categories I–IV, with corresponding descriptors from moderate to extreme. This example peaked as a Category IV (extreme) MHW [17].

Thresholds defining MHW categories could be based on percentiles of the historical distribution of SST values in a region (e.g., 90th, 95th, 98th), which is consistent with the 90th percentile definition for a MHW .

However, unless the reference period is adjusted, this measure will saturate as the oceans warm further, meaning that all days in a year could be classified as heat wave days, as this definition is bounded by a maximum value at the 100th percentile of the observed distribution.

Therefore, heat wave categories, identified as described above, are based on multiples of the value represented by the local difference between the climatological mean and the climatological 90th percentile, which is the threshold used to identify heat waves. This anomaly varies by location and time of year. Multiples of this local difference will describe different categories of MHWs. Magnitude scale descriptors, defined as moderate ($1-2\times$, category I), strong ($2-3\times$, category II), severe ($3-4\times$, category III), and extreme ($> 4\times$, category IV), can be assigned to each point in space and time of an MHW event, based on the intensity (I) measure, as shown in the Figure 1.1.

1.2 Drivers

It is very complex to identify the drivers of MHW, which can be classified into teleconnection, local or regional processes. Two quantitative measurements of these drivers have been proposed to identify MHWs, mean sea surface temperature and sea surface temperature variability[19][32]. Teleconnection processes are very long-range interactions between different climate anomalies (thousands of kilometers)[14]. Local processes are those such as substance transport due to ocean mass motion, air-sea fluxes, the effect of wind on an ocean surface[20]; regional processes are periodic variations in climate that affect a particular area;

The scientific community using a variety of ocean temperature data has identified a major increase in marine heat waves over the past century. From 1925 to 2016, the duration of marine heatwaves and global average frequency increased by 17 % and 34% , respectively, resulting in a 54 percent increase in annual marine heatwave days globally[34].

The increase in average ocean temperature may largely explain this trend, suggesting that there will be a further increase in marine heatwaves in the coming years in the presence of continued global warming.

Marine heat waves are expected to continue to increase in duration, frequency and intensity under future climate change scenarios, increasing the threat to marine ecosystems and the socioeconomic activities that depend on them [12].

At a future warming of $+3.5^{\circ}\text{C}$ relative to pre-industrial levels, the number of marine heatwave days is expected to increase by at least 40 times. At this level of warming, marine heat waves reach intensities of $+2.5^{\circ}\text{C}$ and have a spatial extent more than 20 times greater than pre-industrial levels [11].

With high future emissions, by the end of the 21st century, much of the oceans could reach a permanent marine heat wave state, compared to a fixed pre-industrial threshold.

1.3 The most important events of the past 20 years

The most significant (MHWs) that have had severe impacts on marine ecosystems have occurred in the past 20 years.

The most important events occurred in the northern Mediterranean in 2003[41] (4°C warmer than average for 30 days, mass mortality of marine life in rocky reefs), along the coast of Western Australia in 2011[36] (over 3°C warmer than average for 60 days, algae, fish and sharks moved south), in the Northwest Atlantic in 2012[5] ($2\frac{1}{2}^{\circ}\text{C}$ warmer than average for 56 days, the largest recorded event, lobster fishing peaked early and led to economic tensions between Canada and the U.S.) in the northeast Pacific in 2013-2015 "The Blob"[2] ($2\frac{1}{2}^{\circ}\text{C}$ warmer than average for 226 days, caused seasonally warm weather in the Pacific Northwest of the U.S. and Canada), off southeastern Australia in 2015/16[33] (MHWs first occurred in the southeastern tropical Indian Ocean in November 2015, gradually emerging eastward until March 2016, when all waters from the Northwest Shelf to the Coral Sea were affected).

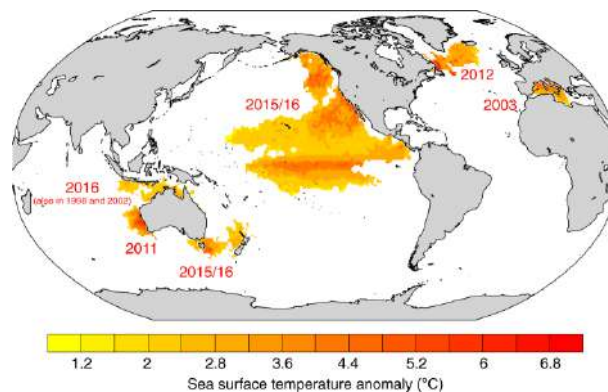


Figure 1.2: The map was developed using the NCAR command language. Major recent marine heat waves documented in the literature. The image shows the maximum sea surface temperature anomaly in regions where the temperature exceeds the 99th percentile, using the National Oceanic and Atmospheric Administration (NOAA) daily sea surface temperature dataset with optimal interpolation. The 99th percentile is calculated over the reference period 1982-2016 [12].

Large areas remained in a state of MHW for 3-4 months continuously, with maximum intensities exceeding 2°C), and in northern Australia in 2016 (Figure 1.2).

Currently in the Mediterranean Sea, temperatures in the Ligurian Sea and Gulf of Taranto are higher than usual, with levels nearly 5°C above average, according to recent satellite observations made of sea surface temperature (SST, <https://doi.org/10.48670/moi-00172>), provided by the Copernicus Marine Service.

In June 2022, two anticyclones brought boiling air from subtropical Africa to the Mediterranean Sea affecting these anomalous phenomena. For example, the Ligurian Sea experienced MHW conditions for 3 weeks before these decayed and recurred in mid-June, while the Gulf of Taranto reached nearly 5°C above average (Figure 1.3 - 1.4).

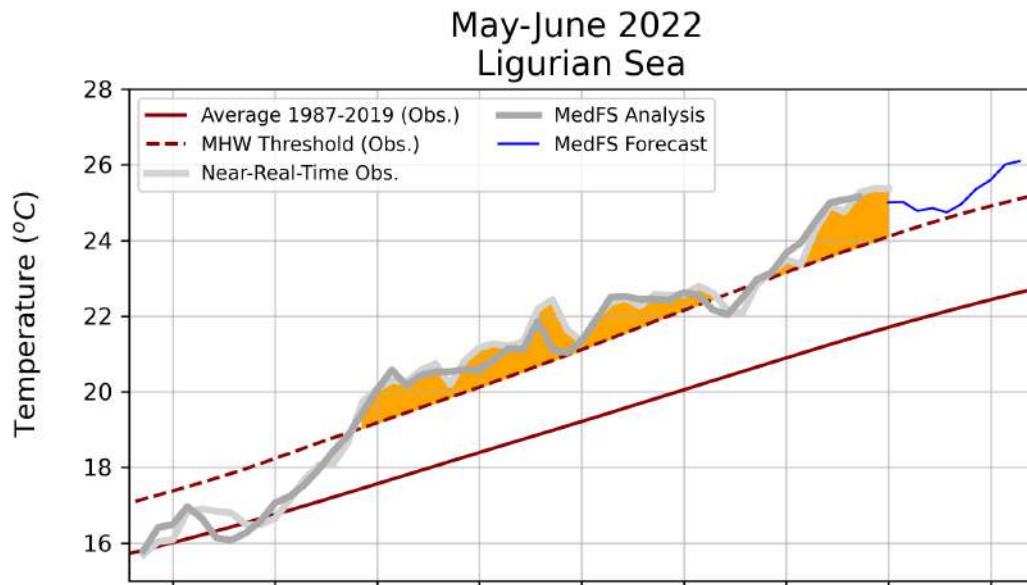


Figure 1.3: Ligurian Sea. Average temperatures and the MHW threshold (defined as the 90th percentile of temperatures for each date) are calculated from long-term observed ("obs.") satellite data. Near real-time satellite data are the most up-to-date data provided by the satellites themselves. Orange shading is used to highlight when the observed temperature is above the MHW threshold.

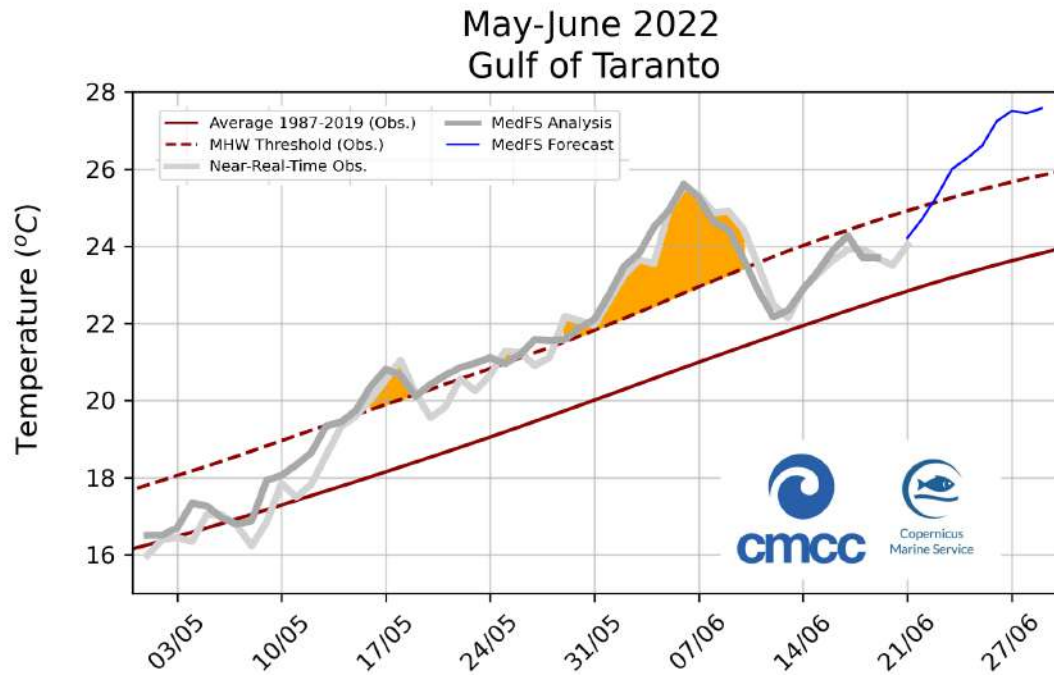


Figure 1.4: Gulf of Taranto. Average temperatures and the MHW threshold (defined as the 90th percentile of temperatures for each date) are calculated from long-term observed ("obs.") satellite data. Near real-time satellite data are the most up-to-date data provided by the satellites themselves. Orange shading is used to highlight when the observed temperature is above the MHW threshold.

The latest analysis (<https://www.cmcc.it/article/marine-heat-wave-in-the-mediterranean-observations-and-predictions>) from the Mediterranean Forecasting System (MedFS), operated by the Euro-Mediterranean Center on Climate Change (CMCC), predicts MHW will persist weakly in the Ligurian Sea but continue to intensify in the Gulf of Taranto. The marine heat wave will still cover most of the western basin, but will continue to spread into the Adriatic and Ionian Seas (Figure 1.5).

In May 2003, conditions were similar to those of today. July and August were followed by the most severe, long-lasting and ecologically devastating events ever recorded. Since then, the number, duration and intensity of storm surges have continued to increase throughout the Mediterranean. At CMCC, ongoing work under the FEVERSEA ESA Research Fellowship (Dr. Giulia Bonino) and the EuroSea project (Dr. Ronan McAdam) will continue to advance research in understanding and predicting MHWs in this region vital to economic activity and

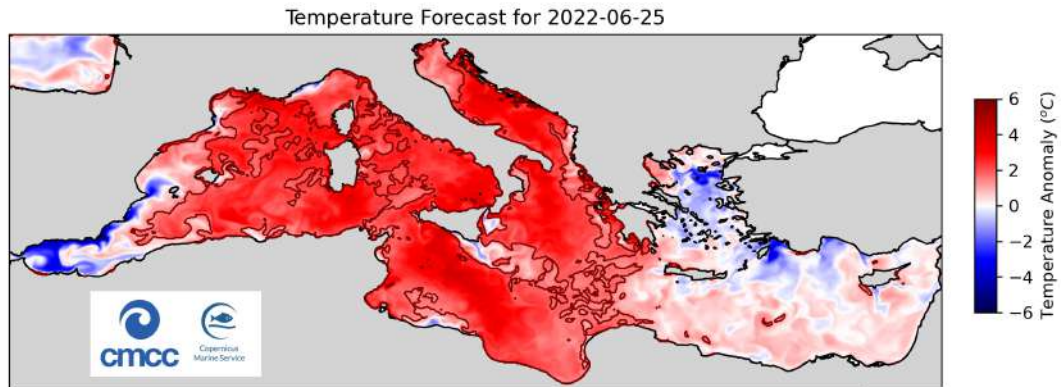


Figure 1.5: Anomaly is the difference between predicted and average temperatures (i.e., red indicates warmer than normal). The dark red outline highlights where the MHW threshold was exceeded for that day.

biodiversity.

1.4 Impacts on natural, physical, and humans systems

Biological, economic, social and atmospheric impacts in recent years are well documented in the literature.

Geographical shifts and increasing changes in the composition of entire species, harmful algal blooms, mass beaching of mammals and mass mortality of particular species [12][42], a strong impact on political and socioeconomic relations when it comes to aquaculture or important fish species, and disturbance of weather conditions on land through teleconnections that may persist for several weeks have been reported in recent years (Figure 1.6).

The 2011 MHW in Western Australia caused a reduction in the abundance of habitat-forming algae, with an entire regime shift in the temperate reef ecosystem and a southward shift in the distribution of tropical fish communities[43].

The MHW in the Northwest Atlantic in 2012 led to a drastic drop in seafood prices and heightened economic tensions between Canada and the US[30].

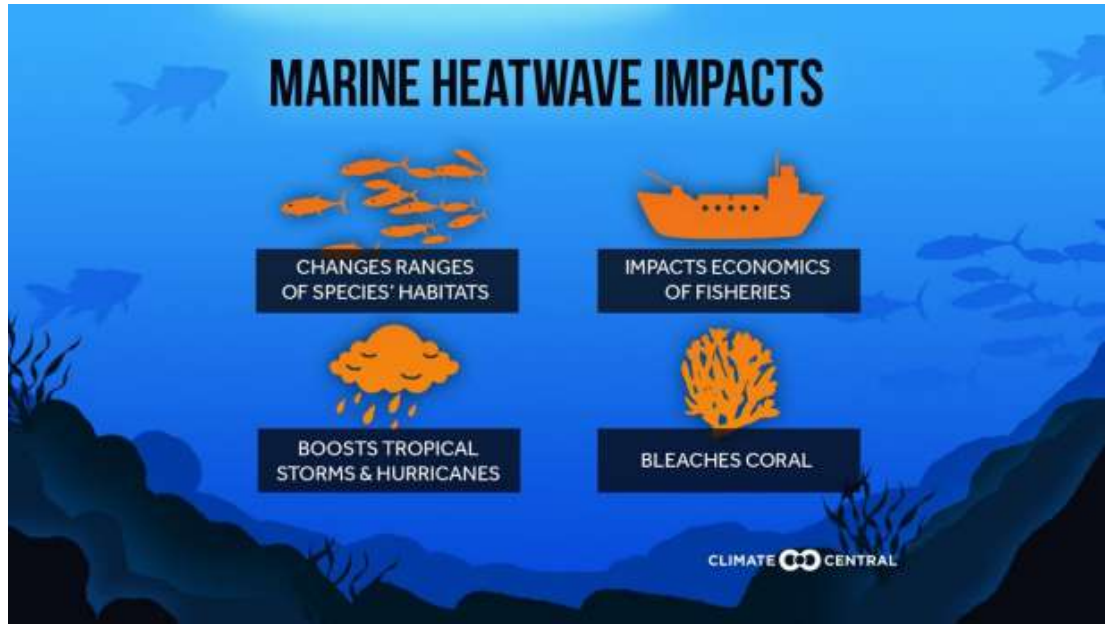


Figure 1.6: A wide range of negative impacts of MHWs

The anomalous high temperatures recorded in the northeast Pacific increased the probability of three consecutive dry winters in California during 2011-2014[7].

The "Blob" in the Northeast Pacific has caused increased mortality of whales and sea lions and the emergence of new marine species in the Northern California region, but not only that, the economic impact has been devastating by causing commercial and recreational fisheries to close, resulting in millions of dollars in losses to the fishing industries [3].

Another example is the MHW associated with El Niño 2015/16, which led to the third mass coral bleaching event in recorded history, with more than 90 percent of surveyed reefs on the Great Barrier Reef bleached in 2016[21][12].

Based on this problem, which is a source of study and interest to the scientific community, this work is concerned with predicting and recognizing ,using Machine Learning techniques, in which region a MHWs might occur in the Mediterranean Sea, having available temperature data from 1982 to 2016.

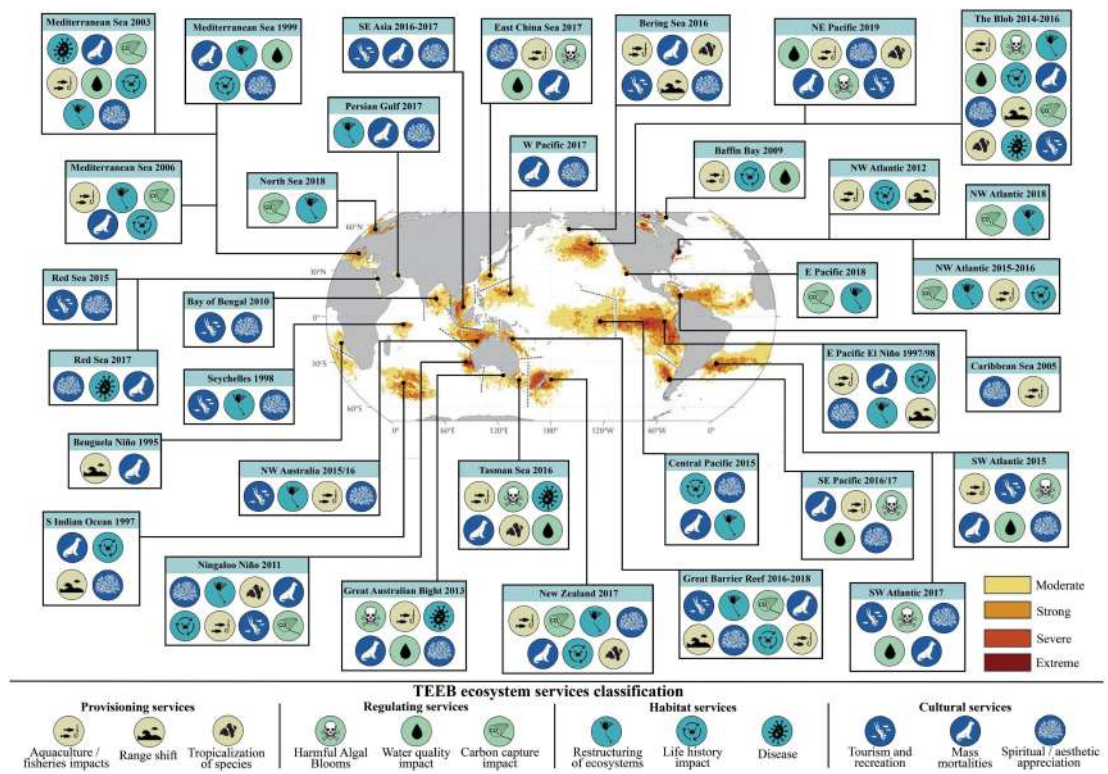


Figure 1.7: Socioeconomic impacts of major marine heatwave (MHW) events since 1995[40].

Background

Chapter 2

In recent years, many machine learning applications have been successfully developed, ranging from programs that learn to detect fraudulent e-mails, to facial recognition systems, to autonomous vehicles that learn to drive on public roads, to information filtering systems that learn users' browsing preferences.

The field of machine learning is concerned with the issue of building computer programs that improve automatically through experience. More precisely, following the definition of Tom M. Mitchell in [31]: “*A computer program is said to learn from experience (E) with respect to some class of tasks (T) and performance measure (P), if its performance at tasks in T , as measured by P , improves with experience E ”*

In machine learning, the ultimate goal is to develop a model that can extract patterns from existing data. A good machine learning model can effectively learn from a large amount of data and accurately identify patterns in the data. According to the task and the type of data used for training, that model can predict or estimate the chosen variable or variables, or indeed generate data as consistent as possible with the dataset, rendering it almost impossible to distinguish between a real sample and a generated one, which is especially useful in prediction tasks.

Deep learning is a type of machine learning and artificial intelligence (AI) that simulates the way humans acquire knowledge. Deep learning is a fundamental component of data science, which includes predictive modeling and statistics.

Deep learning, in its simplest form, can be considered a way to automate predictive analytics. While traditional machine learning algorithms are linear, deep learning algorithms are stacked in a hierarchy of increasing complexity and abstraction.

Hierarchy algorithms apply a nonlinear transformation to the input and use what is learned to create a mathematical model at the output. The iterations proceed until the output achieves a satisfactory level of accuracy.

The number of layers of processing through which data must pass is what inspired the name *deep*.

A type of advanced machine learning algorithm, known as an artificial neural network, underlies most deep learning models. As a result, deep learning can sometimes be referred to as Deep Neural Network. Neural networks come in several variants, including convolutional neural networks (CNNs), recurrent neural networks, feedforward neural networks, and artificial neural networks, and each has advantages for specific use cases. However, they all work in a rather similar way, feeding the data and letting the model figure out for itself whether it has taken the right interpretation on a particular component of the data.

The next section will explore the fundamentals of CNNs useful for solving problems in Climate Science.

2.1 Convolutional Neural Networks

A convolutional neural network (ConvNet/CNN) is a deep learning algorithm that is able to receive an input image, assign importance (learnable weights and biases) to various aspects of the image and be able to distinguish them from each other. The preprocessing needed in a CNN is much less than in other classification algorithms. While in primary methods filters are created by hand, with sufficient training, CNNs have the ability to learn these features.

The architecture of a ConvNet is similar to that of the connectivity model of neurons in the human brain and was derived from the organization of the visual cortex. Individual neurons respond to stimuli only in a narrow region of the visual field, known as the receptive field. A set of such fields overlaps to cover the entire visual area.

As mentioned above, CNNs are primarily based on the fact that the input will

be images. This causes the architecture to have to be set up to best meet the needs of handling a specific type of data. One of the main differences is that the neurons that make up the layers within the CNN consist of neurons organized in three dimensions, the spatial dimensionality of the input (height and width) and depth. In practice, given as input an image of size $32 \times 32 \times 3$ (height, width and depth), we will have a final output layer composed of a dimensionality of $1 \times 1 \times n$, where n represents the possible number of classes.

CNNs are composed of three types of levels:

1. Convolutional layer - determine the output of neurons connected to local regions of the input by calculating the scalar product between their weights and the region connected to the input volume. The rectified linear unit (ReLU) aims to apply an 'elementary' activation function, such as sigmoid [39], to the output of the activation produced by the previous layer.
2. Pooling layers - simply perform downsampling along the spatial dimensionality of the given input, further reducing the number of parameters within that activation.
3. Fully-Connected layers - attempt to produce class scores from the activations, to be used for classification. It is also suggested that ReLU be used between these layers, to improve performance.

When these layers are stacked, a CNN architecture is formed (Figure 2.1).

Using this simple transformation method, CNNs are able to transform the original input layer by layer using convolutional and downsampling techniques to produce class scores for classification and regression goals.

2.2 Convolutional Layer

The convolution layer is the central block of the CNN. It carries most of the computational load of the network.

This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters, known as the kernel, and the other matrix is

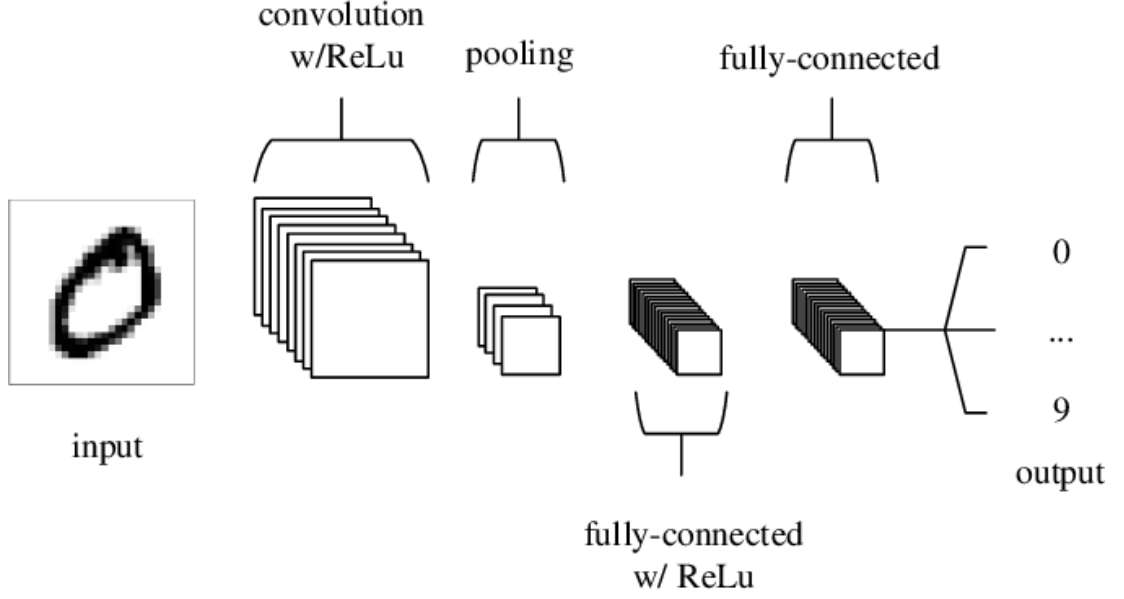


Figure 2.1: An simple CNN architecture, comprised of just five layers.[35]

the narrow portion of the receptive field. The kernel is spatially smaller than an image, but it is deeper. This means that if the image consists of three channels (RGB), the height and width of the kernel will be spatially small, but the depth extends to all three channels (Figure 2.2).

During the forward pass, the kernel runs along the height and width of the image, generating the image representation of that receptive region. This results in a two-dimensional representation of the image, known as the activation map, which provides the kernel's response at each spatial location in the image. The scrolling dimension of the kernel is called the stride.

If we have an input of size $H \times W \times D$ and D out number of kernels with a spatial dimension F with stride S and amount of padding P , then the size of the output volume can be calculated by the following formula:

$$\sqrt{W_{out}} = ((W - F + 2P)/S) + 1$$

This will result in an output volume of size $W_{out} \times W_{out} \times D_{out}$.

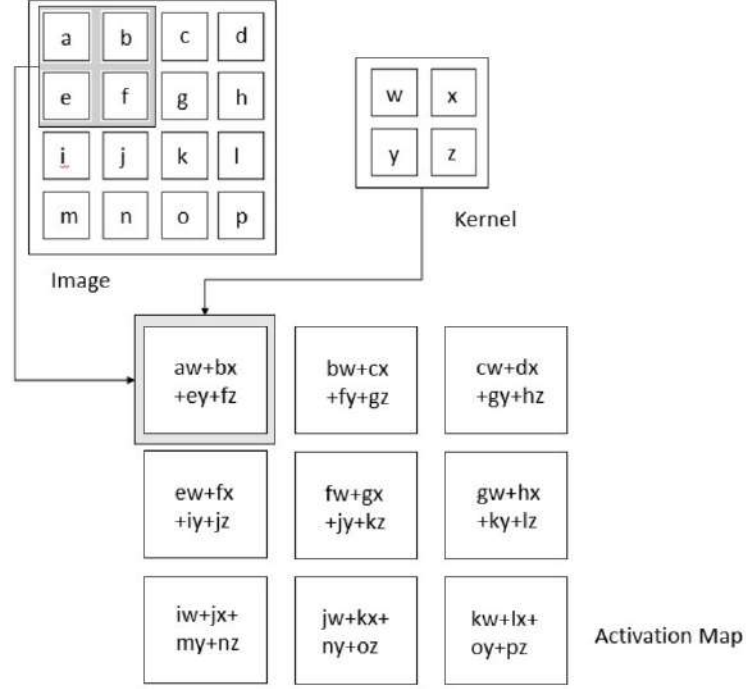


Figure 2.2: Illustration of Convolution Operation

2.3 Pooling Layer

Pooling levels aim to gradually reduce the dimensionality of the representation, and thus to further reduce the number of parameters and computational complexity of the model (Figure 2.3).

There are several pooling functions, such as rectangular neighborhood, L2 norm, rectangular neighborhood average, and a weighted average based on distance from the center pixel [27]. The most widely used process is max pooling, which reports the maximum neighborhood output.

If we have an activation map of size $W \times W \times D$, a pooling kernel of spatial size F and a stride S , the size of the output volume can be determined by the following formula:

$$\sqrt{W_{out}} = ((W - F)/S) + 1$$

This will result in an output volume of size $W_{out} \times W_{out} \times D$.

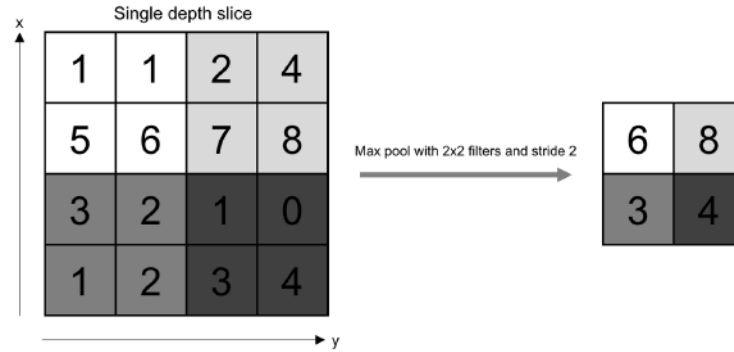


Figure 2.3: Illustration of Max Pooling Operation with 2x2 filters and stride 2

2.4 Fully-connected Layer

The fully connected layer contains neurons that are directly connected to neurons in the two adjacent layers, without being connected to any layer within them (Figure 2.4). This is why it can be computed as usual by a matrix multiplication followed by a bias effect.

The FC layer helps to map the representation between the input and the output.

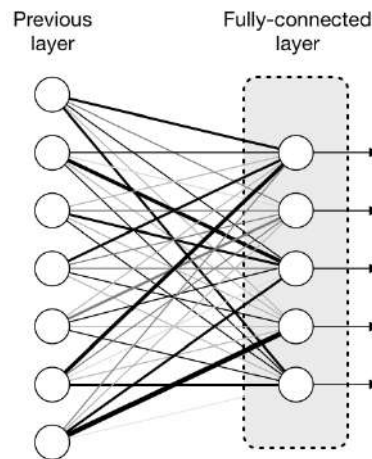


Figure 2.4: Illustration of Fully-connected Layer

2.5 Non-Linearity Layers

Because convolution is a linear operation and images are far from linear, nonlinearity layers are often placed directly after the convolutional layer to introduce nonlinearity into the activation map. The most used are:

- Sigmoid - The sigmoid nonlinearity has the mathematical form:

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$$

It takes a real-valued number and crushes it into a range between 0 and 1. If the local gradient becomes very small, in backpropagation [16] the gradient becomes zero. Instead, if the data arriving at the neuron are always positive, the sigmoid output will be either all negative all positive, resulting in a "zig-zag" dynamics of gradient updates for weight.

- Tanh - Tanh crushes a real-valued number in the interval $[-1, 1]$. Like sigmoid, the activation saturates, but unlike sigmoid neurons, its output is centered on zero.
- ReLU - The rectified linear unit (ReLU) has become widely used in the last decade. It calculates the function $f(k) = \max(0, k)$. The activation is simply a threshold at zero. Compared with sigmoid and tanh, ReLU is more reliable and speeds up convergence by 5-6 times. One drawback is that ReLU can be weak during training. A large gradient running through it can update it such that the neuron is no longer updated. However, we can overcome this problem by setting an appropriate learning rate [1].

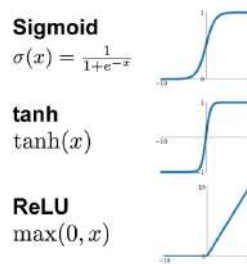


Figure 2.5: Activation Functions Graph

Related Work

Machine learning has been used for decades in weather forecasting and climate science in general, for physical analysis, data assimilation and post-processing and more.

In recent years, considerable success for prediction purposes has been achieved using deep neural networks[38][8].

In [44] elementary weather prediction models are developed using deep convolutional neural networks trained on past weather data to predict one or two fundamental weather fields on a Northern Hemisphere grid without any explicit knowledge of physical processes(the proposed neural networks produce realistic 14-day weather forecasts despite having no explicit knowledge of atmospheric physics).

In the literature, deep learning has been used for severe weather risk assessment and prediction of extreme weather events. In [29] shows how the application of AI techniques, particularly deep learning, together with a physical understanding of the environment, improves forecasting capabilities for several types of high-impact weather phenomena such as severe thunderstorms, tornadoes, and hurricanes.

In [23], it is shown that time-series-centered Long Short-Term Memory neural networks are valid for temperature prediction. So far, no machine learning system has been employed to evaluate long-term heat waves. From a forecasting perspective, compared with forecasting intraday heat waves, this is a more difficult task, by virtue of the fact that one should be able to make a forecast on the sum of the number of days anticipating the heat wave and its duration. In addition, the phenomenology should be very different.

An important contribution is made by the paper[22], in which deep learning is

used to predict the future occurrence of extreme and long-lasting heat waves. The learning performed exploits 1,000 years of climate model outputs. From this data, the algorithm predicts, from 500 hPa geopotential height and surface temperature data, the occurrence or non-occurrence of a long-lasting heat wave starting within t days (Figure 3.1).

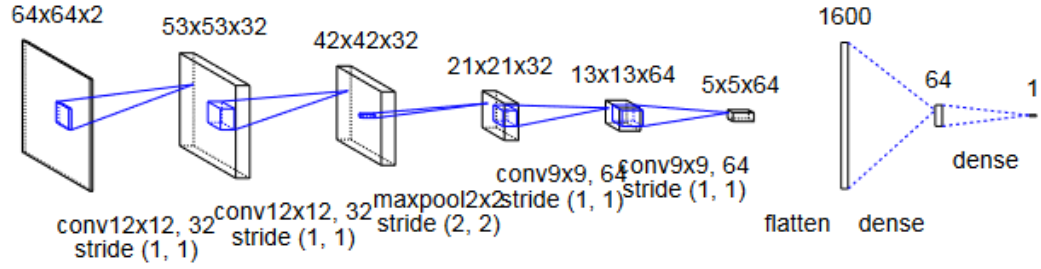


Figure 3.1: CNN-based heatwave predictor architecture. The output is associated with the probability of occurrence of an heatwave in the upcoming t days.[22]

The proposed machine learning prediction algorithm should be useful in the future as a key element in solving the three main scientific challenges of extreme climate studies: the lack of historical data, the problem of model sampling, and the study of model bias. More specifically, the present work has specific objectives: first, to advance machine learning methodology to study rare heat waves of long duration; second, to carry out the first study on the predictability of these extreme events (Figure 3.2).

The closest related work to this case study is certainly Analog "Forecasting of Extreme-Causing Weather Patterns Using Deep Learning"[4] in which two deep learning techniques are used, for anomalous T2m (air temperature at 2 meters above the surface) pattern recognition: the convolutional neural network (ConvNet)[26][13] and a more advanced method, the capsule neural network (CapsNet)[37].

The paper uses data from the Community Large-ensemble Project (LENS)[25], which consists of a 40-member ensemble of fully coupled atmosphere-ocean-ice simulations of the Community Earth System Model Version 1 (CESM1) with

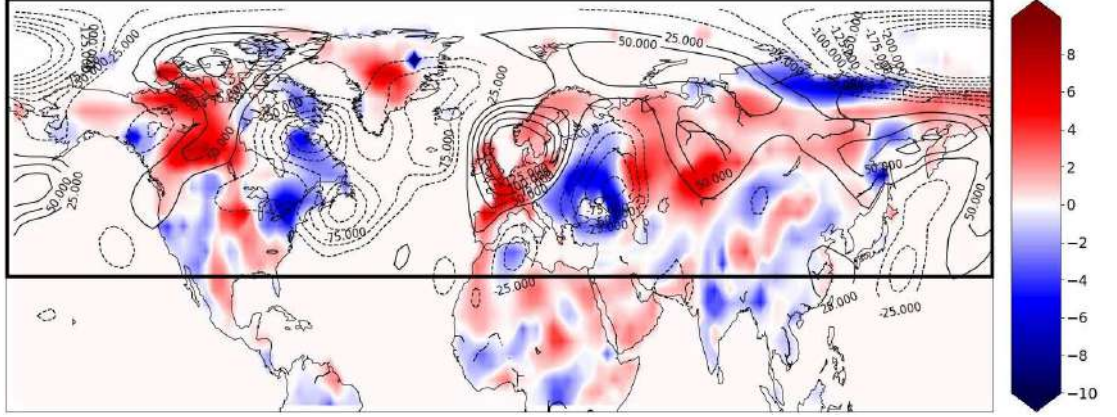


Figure 3.2: Snapshot of the surface temperature surface fluctuations (T_s fluctuations, according to the color bar, in Kelvin) and of the geopotential height at 500mb (Zg fluctuations, contours) over the Northern Hemisphere. This snapshot is taken on July 20th on a arbitrary year of the PlaSim simulation. The spatial resolution is 64×128 (latitude \times longitude). The thin contour lines, representing the anomaly of Zg, are in meters. The thick black contour delimits the zone that is used for prediction by the machine learning procedure[22].

historical radiative forcing 1920-2005.

From this dataset, surface air temperature, measured as temperature at 2 m above the ground (T2m), is used, and geopotential height at 500 mb (Z500), to predict in which region of the North American continent in the subtropical and midlatitude regions between 30°N and 60°N heat wave events may occur.

CapsNets are trained on midtropospheric large-scale circulation patterns (Z500) labeled 0–4 depending on the existence and geographical region of surface temperature extremes over North America several days ahead (Figure 3.3).

Trained nets predict the occurrence/region of heat waves, only using Z500, with accuracies (recalls) of 69-45% (77-48%) or 62-41% (73-47%) 1-5 days in advance. Also using surface temperature, the accuracies (recalls) with CapsNets increase to about 80% (88%). In both contexts, CapsNets outperform more traditional techniques such as convolutional neural networks and logistic regression, and their accuracy is minimally affected by the reduction in training set size.

The results highlight the potential of multivariate data-driven frameworks for more accurate and rapid prediction of extreme weather phenomena, which can augment the goals of numerical weather prediction in ensuring timely warning.

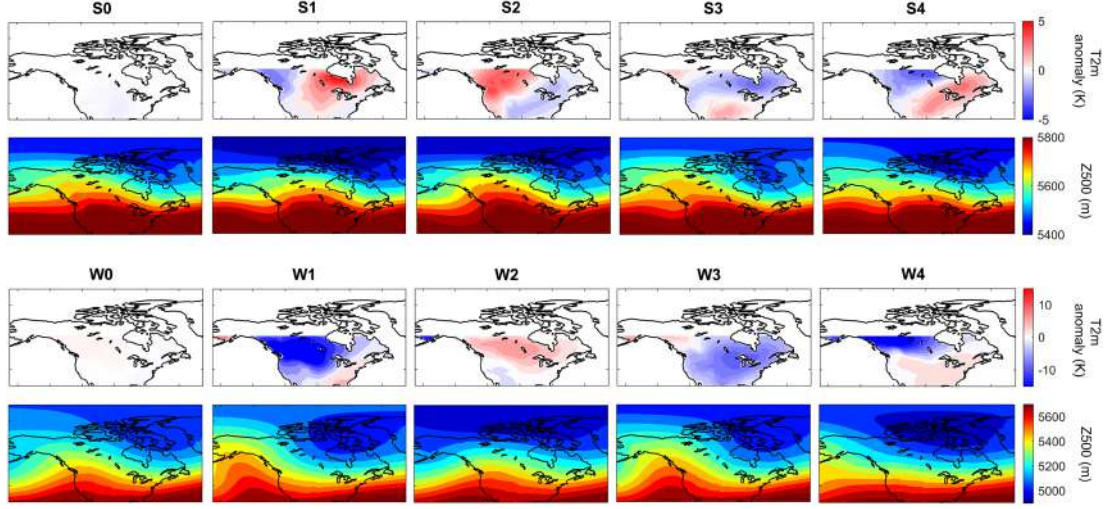


Figure 3.3: Cluster centers of T2m anomalies at the onsets and Z500 patterns of 3 days earlier. The top (bottom) two rows correspond to summers (winters). S0 (W0) shows the average of T2m and Z500 patterns from days with no heat wave (cold spell). S1–S4 and W1–W4 are obtained from K-means clustering the anomalous T2m patterns at onsets into four classes, which roughly separates the extreme events into four geographical regions: Northern Canada (S1), Western United States-Canada (S2), Southern United States (S3), and Eastern United States-Canada (S4) in summers and Northwest United States-Canada (W1), Alaska (W2), Northeast United States-Canada (W3), and Northern Canada (W4) in winters. Rows 1 and 3 show the cluster centers, while rows 2 and 4 show the average of Z500 patterns 3 days before the onsets for each cluster[4].

Figure 3.4 illustrates the data-driven structure for cold wave prediction based on Z500 data from the previous 3 days. Using the impact-based self-labeling procedure, Z500 patterns are classified as W0, W1, W2, W3, or W4, depending on the T2m cluster index of previous 3 days. The boxes above depict examples of T2m patterns at the beginning and the corresponding Z500 patterns (from 3 days earlier) for each cluster. During training, only Z500 patterns and their labels are entered into the CapsNet.

When trained, the CapsNet can predict, from a given Z500 pattern, the indices of T2m clusters from 3 days later, thus predicting the occurrence and geographic region of cold spells. For a test case such as the one shown, a cold wave was predicted in W3 3 days later.

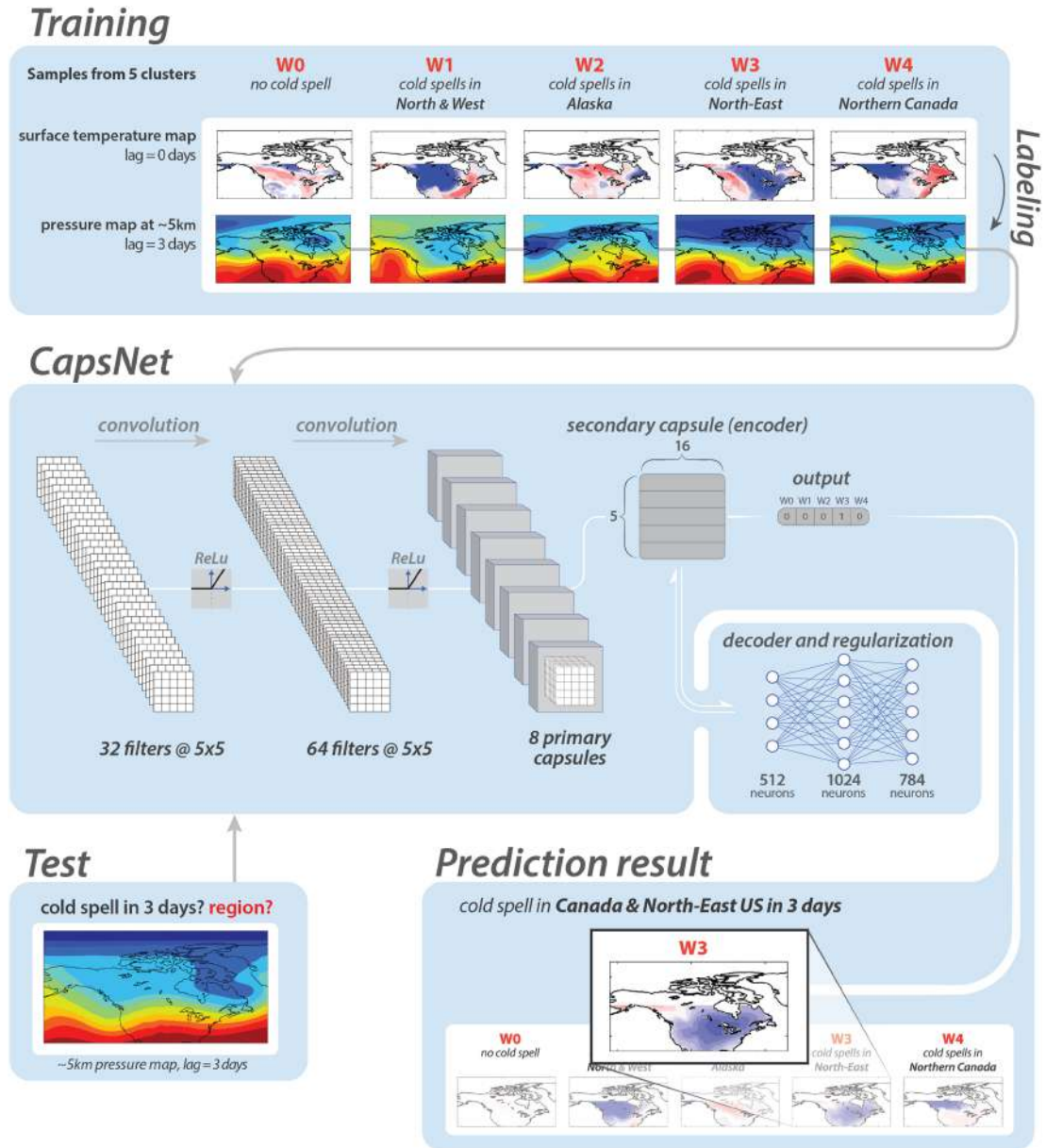


Figure 3.4: Illustration of the data-driven structure for cold wave prediction based on Z500 data from 3 days before [4].

Note that for winters, Z500 models over a larger domain extending from the Pacific Ocean to 145°E are included in the CapsNets, but a smaller domain is shown in this figure for a more accurate representation.

Chapter

4

Experimental Setups

This work is done in collaboration with the Ocean Modeling and Data Assimilation (ODA) and Advanced Scientific Computation divisions of the Mediterranean Center on Climate Change (CMCC).

4.1 Setup A

The objective is to predict the presence or absence of MHW in the Mediterranean region by evaluating the conditions of six input factors over a broader domain that also includes the eastern Atlantic and northern Europe, and classify the occurrence of a MHW into one of four classes from C0 to C3: C0 refers to the "no occurrence" class, C1 to C3 identify the occurrence of a MHW in one of three classes (Figure 4.1).

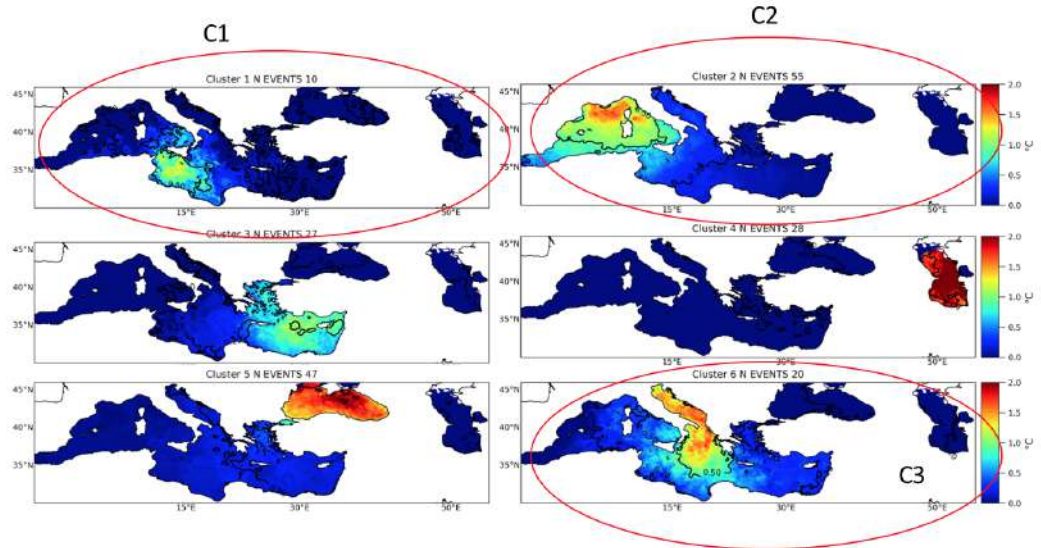


Figure 4.1: The obtained cluster patterns are depicted by averaging the mean intensity of all event maps corresponding to each cluster. The contours represent the average anomaly of SST (Sea Surface Temperature).

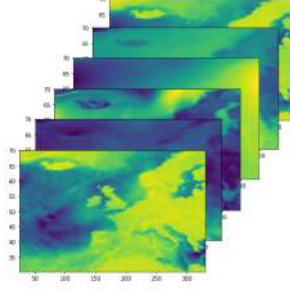


Figure 4.2: Example of six input variables stacked in an input volume for the date 2008-12-04.

4.1.1 Climate Data

The data used in the present work are collected from the ERA5 global reanalysis dataset and cover the period from 1981 to 2016, with average daily amounts, totaling about 12000 samples, and covering the domain from -30.00° E to 24.75° E longitude and 70.00° N to 30.25° N latitude with a geographic resolution of $0.25^\circ \times 0.25^\circ$ for a size of 160x220 pixels (Figure 4.2).

The climate variables to be evaluated during inference consist of the following six drivers:

1. T2M - Temperature at a height of 2 meters.
2. WIND - Wind speed at a height of 10 meters.
3. LAT - Surface latent heat flux (the flow of heat from the earth's surface to the atmosphere).
4. INC - Incoming solar radiation (calculated as the sum of downward surface solar radiation and downward surface solar thermal radiation).
5. SLP - Pressure at mean sea level.
6. NET - Net air-sea heat flux (obtained as the sum of net surface solar radiation, net surface heat radiation, surface sensible heat flux and surface latent heat flux).

The data are harvests in NetCDF (Network Common Data Form) a file format for collecting multidimensional scientific data , called variables, such as temperature, wind direction and speed, humidity, pressure, and each of these variables can be displayed according to a dimension, such as time (Figure 4.3).



Figure 4.3: Examples of netCDF data: Left—Data structure; Right—Temperature at a height of 2 meters.

The target data, provided by CMCC, are structured in a comma-separated values (CSV) text file with a specific format that allows the data to be saved in a structured table (Table 4.1).

The structured table of the target data have the following attributes:

- ID - Identifier of the MHW.
- C0 - No occurrence of MHW.
- C1 - Occurrence in the southern region of the Mediterranean Sea.
- C2 - Occurrence in the northern region of the Mediterranean Sea.
- C3 - Occurrence in the western region of the Mediterranean Sea.
- DATE - Reference date.

These data were acquired over a period of 35 years, from 1981 to 2016. The data are acquired approximately on a daily basis since there are not 365 (or 366 for leap years) records for all years, also there may be multiple records in reference

to the same one (due to the reason that for a specific date, multiple MHW may have occurred in different regions). Thus, the total samples are 7929 for the target data.

Target data					
ID	C0	C1	C2	C3	DATE
9.0	0.0	0.0	0.0	1.0	1984-11-08
9.0	0.0	0.0	0.0	1.0	1984-11-09
9.0	0.0	0.0	0.0	1.0	1984-11-16
38.0	1.0	0.0	0.0	0.0	1995-03-31
59.0	0.0	0.0	1.0	0.0	2000-05-12
59.0	0.0	0.0	1.0	0.0	2000-05-15

Table 4.1: Example of Target data. "1.0" indicates the presence of the occurrence, "0.0" otherwise.

4.1.2 Data Preprocessing

Due to the lack of some daily samples for several years, the imbalance between the classes of occurrence (C0, C1, C2, C3 and their possible combinations) and the phenomenon of climate warming, which significantly influences the data for the past decades, it was necessary to carry out careful preprocessing of the data, which can solve all the problems that may arise from the factors listed above.

4.1.2.1 Target Data

First, it was planned to merge the occurrences of MHWs with the same time reference, so as to have only one record per date, to be associated with the *time* dimension of the input drivers (Table 4.2).

This process generated a set of new classes, the result of combining the existing classes. Therefore, new classes were assigned to the target dates, following the scheme in Table 4.3.

The next step was to check the imbalance of the 8 classes. The presence of an extremely unbalanced distribution of the response variable can lead the learning process to be biased, because the model tends to focus on the prevailing class and

Original data					
ID	C0	C1	C2	C3	DATE
45.0	0.0	0.0	0.0	1.0	1986-02-13
13.0	0.0	1.0	0.0	0.0	1986-02-13
Preprocessing data					
ID	C0	C1	C2	C3	DATE
58.0	0.0	1.0	0.0	1.0	1986-02-13

Table 4.2: Example of merging occurrences of MHws with the same time reference before and after preprocessing

C0	C1	C2	C3	New Class
1.0	0.0	0.0	0.0	0
0.0	1.0	0.0	0.0	1
1.0	1.0	0.0	0.0	1
0.0	0.0	1.0	0.0	2
1.0	0.0	1.0	0.0	2
0.0	0.0	0.0	1.0	3
1.0	0.0	0.0	1.0	3
0.0	1.0	1.0	0.0	4
1.0	1.0	1.0	0.0	4
0.0	1.0	0.0	1.0	5
1.0	1.0	0.0	1.0	5
0.0	0.0	1.0	1.0	6
1.0	0.0	1.0	1.0	6
0.0	1.0	1.0	1.0	7
1.0	1.0	1.0	1.0	7

Table 4.3: Method of assigning new classes

ignore rare events.

The results obtained are shown in Table 4.4.

Given the data, it was preferred in this first analysis to exclude records with mixed occurrences of MHW (4,5,6,7) and limit the study to the 4 major classes (0,1,2,3).

The imbalance problem is evident between classes 1 and 2. Therefore, it was decided to align the four classes using the undersampling technique. The best simple undersampling algorithm is random undersampling [10].

Class	Occurrences
0	1098
1	375
2	4080
3	1186
4	133
5	3
6	216
7	21

Table 4.4: Occurrences by class - Setup A

This is a nonheuristic algorithm that attempts to balance the target distributions by randomly eliminating instances of the majority class (Figure 4.4).



Figure 4.4: Undersampling technique.

This solution can eliminate valuable data that may be essential for classification models, but it is advisable in the presence of a large amount of data[28].

In this case, all four classes were aligned at 375 samples, for a total of 1,500 samples.

4.1.2.2 Input Data

For the input data, it was decided to use all six drivers for an initial analysis. On the *time* dimension, of the NetCDF data structure, a precise selection was made, using only the gionarlier maps to which a record corresponds in the target data for the same date, consequently eliminating maps for days not in the target data. Therefore, the input data are 1500 with a shape of (HxWxC) 160x220x6.

4.2 Setup B

The objective is to predict the presence or absence of MHW in the Mediterranean region by evaluating the conditions of four input factors over a broader domain that also includes the eastern Pacific and northern Europe, and classify the occurrence of a MHW into one of four classes from C0 to C6: C0 refers to the "no occurrence" class, C1 to C6 identify the occurrence of a MHW in one of six classes (Figure 4.5).

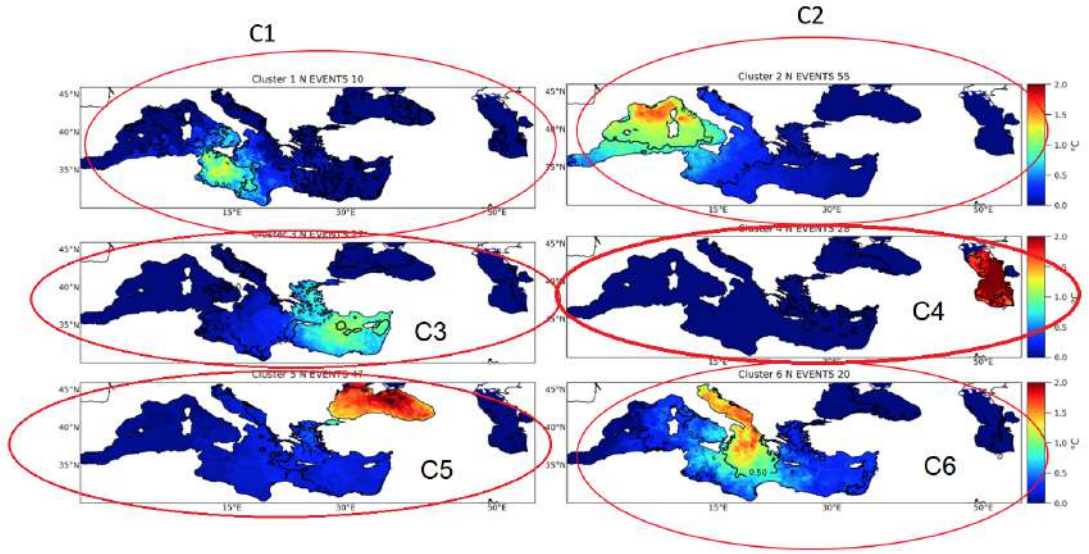


Figure 4.5: The obtained cluster patterns are depicted by averaging the mean intensity of all event maps corresponding to each cluster. The contours represent the average anomaly of SST (Sea Surface Temperature).

4.2.1 Climate Data

The data used in the present work are collected from the ERA5 global reanalysis dataset and cover the period from 1981 to 2016, with average daily amounts, totaling about 12000 samples, and covering the domain from -50.00° E to 80.00° E longitude and 70.00° N to 10.00° N latitude with a geographic resolution of $0.25^{\circ} \times 0.25^{\circ}$ for a pixel size of 241×521 (Figure 4.6).

The climate variables to be evaluated during inference consist of the following four drivers:

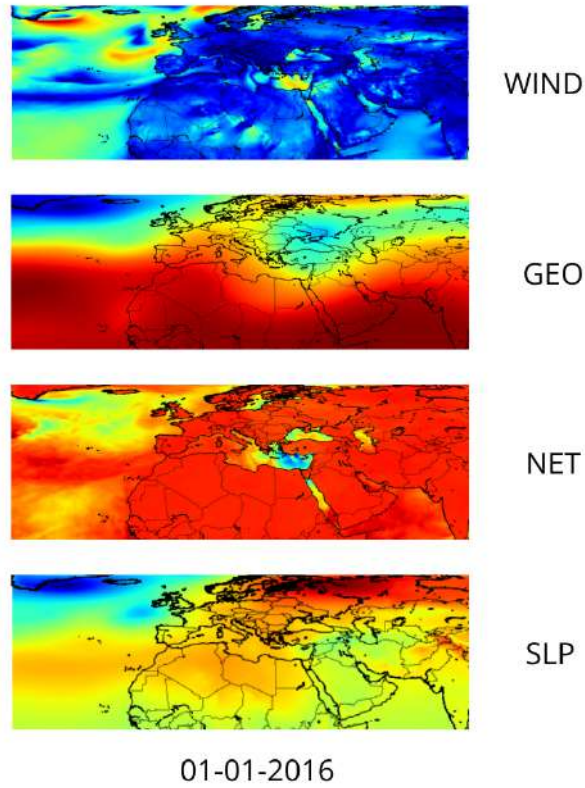


Figure 4.6: Example of four input variables stacked in an input volume for the date 01-01-2016.

1. WIND - Wind speed at a height of 10 meters.
2. GEO - Geopotential height approximates the actual height of a pressure surface above mean sea-level. Therefore, a geopotential height observation represents the height of the pressure surface on which the observation was taken.
3. SLP - Pressure at mean sea level.
4. NET - Net air-sea heat flux (obtained as the sum of net surface solar radiation, net surface heat radiation, surface sensible heat flux and surface latent heat flux).

These maps are also in NetCDF format (Figure 4.3).

The target data, provided by CMCC, are structured in a comma-separated values (CSV) text file with a specific format that allows the data to be saved in a structured table (Table 4.5).

The structured table of the target data have the following attributes:

- ID - Identifier of the MHW.
- C0 - No occurrence of MHW.
- C1 - Occurrence in Central Mediterranean sea.
- C2 - Occurrence in Western Mediterranean Sea.
- C3 - Occurrence in Eastern Mediterranean Sea.
- C4 - Occurrence in Caspian Sea.
- C5 - Occurrence in Black Sea.
- C6 - Occurrence in Adriatic Sea.
- DATE - Reference date.

These data were acquired over a period of 35 years, from 1981 to 2016. The data are acquired approximately on a daily basis since there are not 365 (or 366 for leap years) records for all years, also there may be multiple records in reference to the same one (due to the reason that for a specific date, multiple MHW may have occurred in different regions). Thus, the total samples are 15560 for the target data.

4.2.2 Data Preprocessing

Due to the lack of some daily samples for several years, the imbalance between the classes of occurrence (C0, C1, C2, C3, C4, C5, C6 and their possible combinations) and the phenomenon of climate warming, which significantly influences the data for the past decades, it was necessary to carry out careful preprocessing of the data, which can solve all the problems that may arise from the factors listed above.

Target data								
ID	C0	C1	C2	C3	C4	C5	C6	DATE
9.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1984-11-08
9.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1984-11-09
9.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1984-11-16
38.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1995-03-31
59.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	2000-05-12
59.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	2000-05-15

Table 4.5: Example of Target data. "1.0" indicates the presence of the occurrence, "0.0" otherwise.

4.2.2.1 Target Data

First, it was considered to make unique occurrences of MHW with the same time reference, randomly eliminating occurrences with the same temporal reference, so as to maintain only one record per date, to be associated with the temporal dimension of input drivers (Table 4.6).

Original data								
ID	C0	C1	C2	C3	C4	C5	C6	DATE
45.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1986-02-13
13.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1986-02-13
Preprocessing data								
ID	C0	C1	C2	C3	C4	C5	C6	DATE
45.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1986-02-13

Table 4.6: MHws with the same time reference before and after preprocessing

This process generated a new dataset with 10206 samples, distributed as shown by Table 4.7.

The next step was to check the imbalance of the 7 classes. The problem of imbalance is evident between class 1 and all the others. Therefore, it was decided to align the 7 classes by again using the Random Under-sampling technique.

In this case, all 7 classes were aligned at 334 samples, for a total of 2338 samples.

Class	Occurrences
0	995
1	334
2	3512
3	1397
4	1082
5	2057
6	829

Table 4.7: Occurrences by class - Setup B

4.2.2.2 Input Data

For the input data, it was decided to use all 4 drivers for an initial analysis. On the *time* dimension, of the NetCDF data structure, a precise selection was made, using only the gionarlier maps to which a record corresponds in the target data for the same date, consequently eliminating maps for days not in the target data. Therefore, the input data are 2338 with a shape of (HxWxC) 241x521x4.

4.3 Setup B.2

This configuration is a variation of configuration B, in which the selection process of section "4.2.2.1 Target Data" has been modified.

4.3.1 Climate Data

See Section 4.2.1

4.3.2 Data Preprocessing

See Section 4.2.2

4.3.2.1 Target Data

First, it was considered to make unique occurrences of MHW with the same time reference, randomly eliminating occurrences with the same temporal reference, so as to maintain only one record per date, to be associated with the temporal dimension of input drivers, as already done in Section 4.2.2.1 (Table 4.6).

This time, before balancing the 7 classes, classes 1 and 2 were merged , creating a new class referring to the West-Central Mediterranean Sea. The next step was to check the imbalance of the new 6 classes distributed as shown in Table 4.8.

Class	Occurrences
0	995
1-2	3846
3	1397
4	1082
5	2057
6	829

Table 4.8: Occurrences by class - Setup B.2

it was decided to align the 6 classes again using the random undersampling technique. In this case, all 6 classes were aligned to 829 samples, for a total of 4974 samples.

4.3.2.2 Input Data

For the input data, it was decided to use all 4 drivers for an initial analysis. On the *time* dimension, of the NetCDF data structure, a precise selection was made, using only the gionarlier maps to which a record corresponds in the target data for the same date, consequently eliminating maps for days not in the target data. Therefore, the input data are 4974 with a shape of (HxWxC) 241x521x4.

4.4 Training, Testing and Validation Datasets

Splitting into Training, Testing and Validation datasets was done by having possibly the same number of samples for each class (e.g., 10 samples of class 0, 10 samples of class 1, etc.).

4.5 Deep learning architecture and procedure

Choosing the architecture to be used was not easy,so several approaches were evaluated to build a model suitable for the task. The main objective of this work is to predict the occurrence of MHWs on the same day as the input (without any time lag between input and output) and to verify that the drivers provide useful information to perform the prediction,both in image synthesis and classification tasks. For an initial analysis, CNNs present in the state of the art were taken as a starting point,adapting them to the use case of this work.

Heatwave prediction is performed as a supervised classification problem using CNN-based deep learning architectures shown in Figure 4.7, 4.9, 4.10. The choice of CNN instead of classical ML methods stems from the large data size: other methods would require engineering arbitrary features.

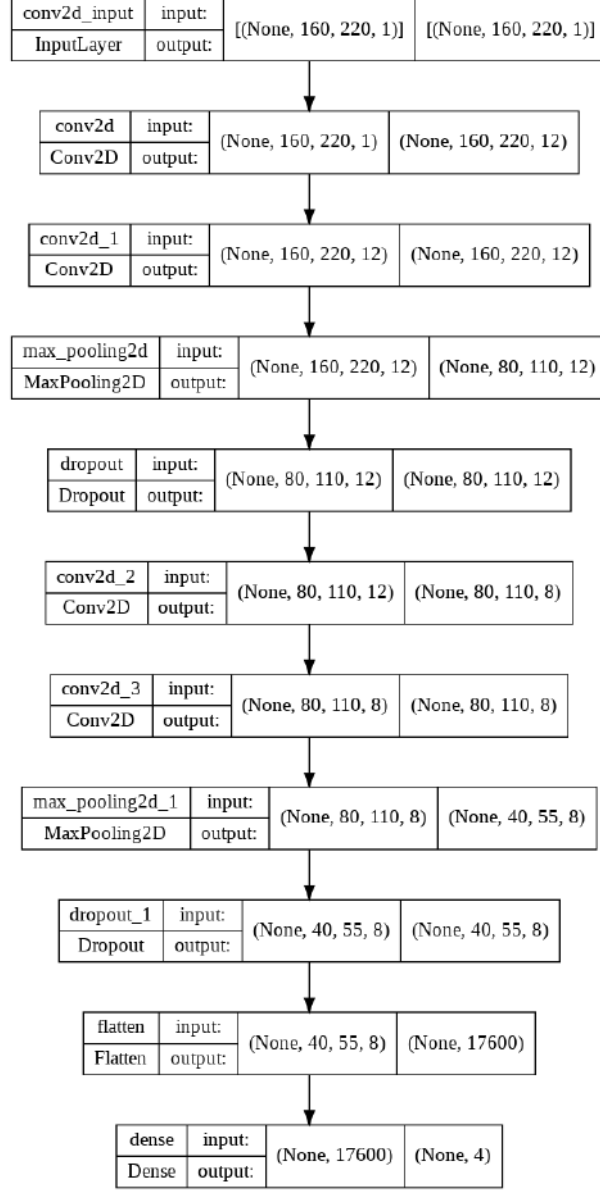


Figure 4.7: CNN-based heatwave predictor Architecture 1 - SETUP A.

In the proposed architecture, the first two convolutional layers have filters of size 12×12 and ReLU activation features. They are followed by a Maxpool layer so that the data size is divided by 2×2 with spatial dropout. The next two convolutional layers have filters of size 8×8 and ReLU activation features and are also followed by a spatial Dropout. Finally, a Flattening and a Fully-connected layer followed by a Softmax activation function [9] producing 4 outputs with values between 0 and 1, i.e., the probability of belonging to one of the 4 classes. The

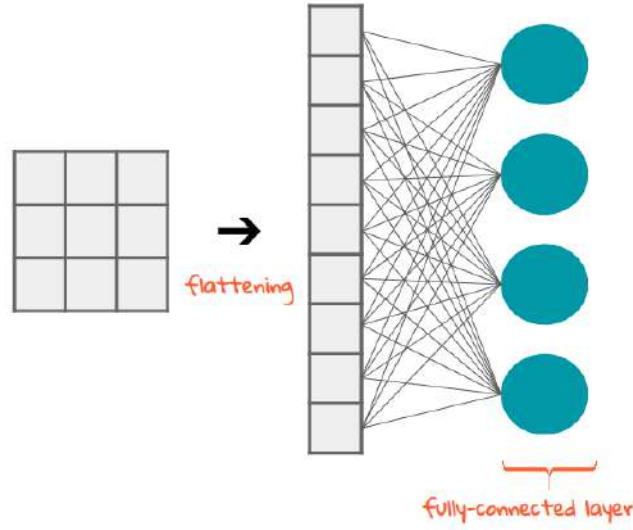


Figure 4.8: Illustration of the Flattening operation.

higher probability will be the prediction of the occurrence of MHW.

Flattening involves converting the data into a 1-dimensional matrix to be included in the next layer. The output of the convolutional layers is flattened to create a single long feature vector. This is connected to the final classification model, which is called the fully connected layer. In other words, we put all the pixel data in a single row and create connections to the final layer (Figure 4.8).

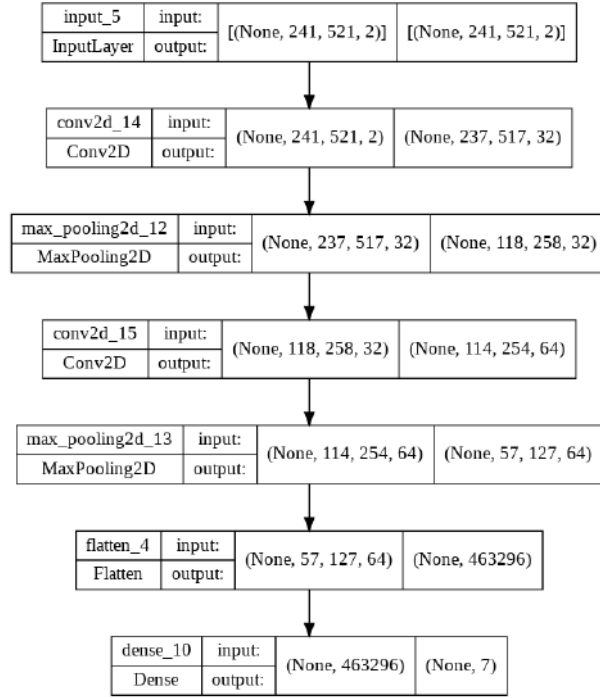


Figure 4.9: CNN-based heatwave predictor Architecture 2 - SETUP B.

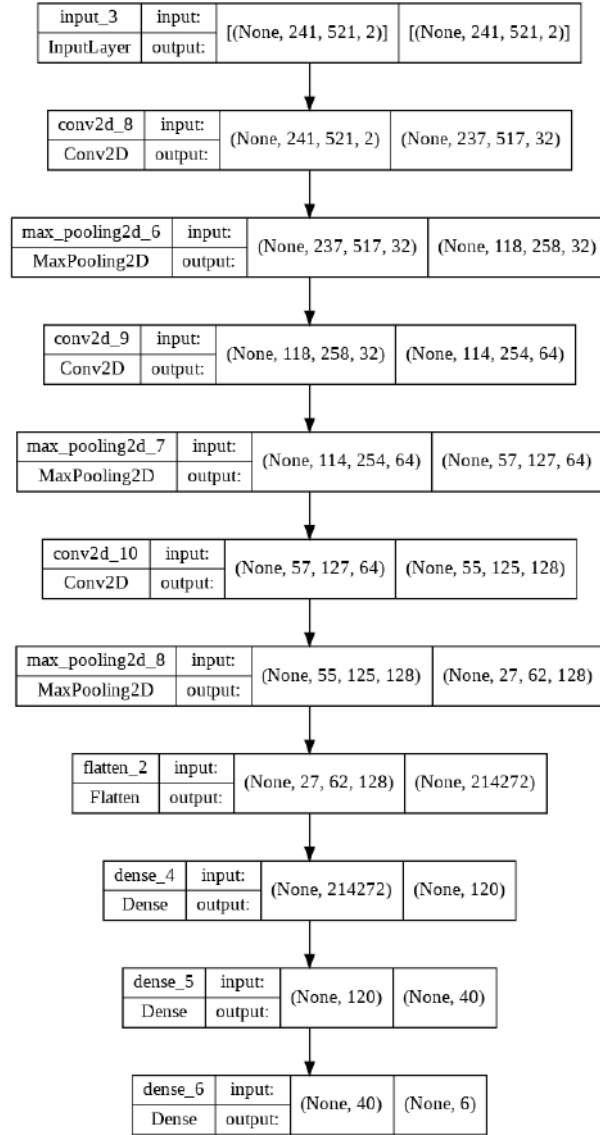


Figure 4.10: CNN-based heatwave predictor Architecture 3 - SETUP B.2.

4.5.1 Capsule Neural Network architecture

A capsule neural network (CapsNet) is a machine learning approach that represents a type of artificial neural network (ANN) that can be used to better model hierarchical relationships.

The concept is to add structures called "capsules" to a convolutional neural network (CNN) and reuse the output of several of those capsules to form more stable representations (with regard to various perturbations) for the higher capsules.

The Figure 4.11 shows an architecture of CapsNet. The architecture is shallow, with only two convolutional layers and one fully connected layer. Conv1 has 256 9x9 convolutional kernels with a stride of 1 and ReLU activation. This layer converts pixel intensities into local feature detector activities that are then used as inputs to the primary capsules [37].

Primary capsules are the lowest level of multidimensional entities, and from the perspective of inverse graphics, the activation of primary capsules corresponds to the inversion of the rendering process. This is a very different kind of computation than putting instantiated parts together to create familiar whole entities, which is what the capsules are designed to do. The second level (PrimaryCapsules) is a convolutional capsule level with 32 channels of 8D convolutional capsules (i.e., each primary capsule contains 8 convolutional units with a kernel of 9x9 and a stride of 2)[37].

Each primary capsule output views the outputs of all 256X 81 Conv1 units whose receptive fields stack on top of the capsule center position. In total PrimaryCapsules has [32x6x6] capsule outputs and each capsule in the [6x 6] grid shares its weights with each other. PrimaryCapsules can be seen as a convolution layer with Equation 1 as the nonlinearity block. nonlinearity. The final level (DigitCaps) has one 16D capsule per digit class and each of these capsules receives input from all capsules in the level below[37].

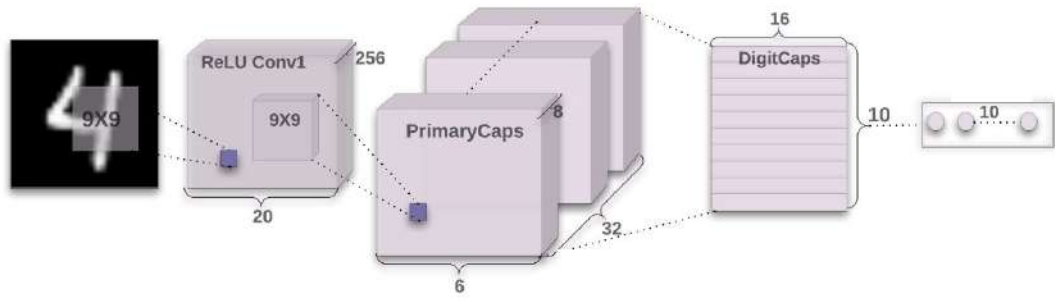


Figure 4.11: CapsNet with 3 layers. The length of the activity vector of each capsule in DigitCaps layer indicates presence of an instance of each class and is used to calculate the classification loss.

4.6 Hardware & Software Resources

Network testing and evaluation was performed on the Marconi100 (M100) GPU cluster hosted by CINECA. Each of the 980 nodes is equipped with 2 IBM Power9 CPUs and 4 NVIDIA Volta GPUs with Nvlink 2.0, 16 GB and 256 GB RAM [36]. The system runs on Red Hat Enterprise Linux Server version 8.1. The entire algorithm and models were developed using Python 3.8, TF 2.2 and Keras 2.4. These frameworks enabled rapid development of the architecture described above and allowed the 4 GPUs of an M100 node to be leveraged for experimentation, training and testing of the models. This was achieved through TensorFlow’s Distributed Training through Mirrored strategy, which synchronously trains multiple replicas of the model on different GPUs of the machine and combines the results into a final replica.

4.7 Model Evaluations

Model evaluation is the step-by-step process of using different evaluation metrics to understand the performances of a machine learning model. Model evaluation is relevant to assessing the effectiveness of a model during the initial stages of research and also plays a role in model monitoring.

4.7.1 Accuracy

Accuracy is a metric that generally describes how the model performs across all classes. It is useful when all classes are of equal importance. It is calculated as the ratio between the number of correct predictions to the total number of predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

4.7.2 Loss functions

Loss functions are used to determine the error between the output of our algorithms and the given target value. The loss function expresses how far off the mark our computed output is. Loss functions are used while training perceptrons and neural networks by influencing how their weights are updated. The larger the loss is, the larger the update. By minimizing the loss, the model's accuracy is maximized. However, the tradeoff between size of update and minimal loss must be evaluated in these machine learning applications.

Categorical Cross-Entropy loss was used in this work, it is a Softmax activation plus a Cross-Entropy loss. If this loss is used, a CNN is trained to produce a probability on the C classes for each image. It is used for multiclass classification (Figure 4.12).

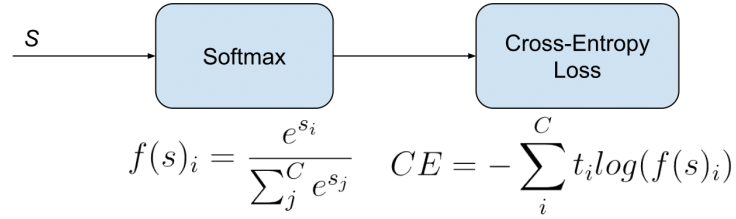


Figure 4.12: Softmax Cross-Entropy Pipeline.

In the specific (and usual) case of Multi-Class classification the labels are one-hot, so only the positive class C_p keeps its term in the loss. There is only one element of the Target vector t which is not zero $t_i = t_p$. So discarding the elements of the summation which are zero due to target labels, we can write:

$$CE = -\log \left(\frac{e^{s_p}}{\sum_j^C e^{s_j}} \right)$$

Figure 4.13: Cross-Entropy loss

where S_p is the CNN score for the positive class.

Defined the loss, now we'll have to compute its gradient respect to the output neurons of the CNN in order to backpropagate it through the net and optimize

the defined loss function tuning the net parameters. So we need to compute the gradient of CE Loss respect each CNN class score in s . The loss terms coming from the negative classes are zero. However, the loss gradient respect those negative classes is not cancelled, since the Softmax of the positive class also depends on the negative classes scores.

4.7.3 Confusion Matrix

A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix. The confusion matrix shows the ways in which your classification model is confused when it makes predictions. It gives you insight not only into the errors being made by your classifier but more importantly the types of errors that are being made. It is this breakdown that overcomes the limitation of using classification accuracy alone.

Chapter

5

Results and Discussions

Useful results were not obtained in all the many models and approaches tested, but numerous insights of interest were produced. This is certainly attributable to the strong imbalance in the presence of MHW, given the rarity of the event. It is possible that the correlation between inputs and cluster labels is very low, causing the model to perform worse than random in predicting classes.

5.1 Architecture 1 - SETUP A

The results of Setup A using the CNN described in Figure 4.7 did not produce relevant results. The network, due to the few heterogeneous data available, perfectly learned the occurrences of MHWs, often associating with the same event on close days, not recognizing the same occurrence for distinct events, occurring on different days or years. The mere presence of the T2M driver, as a strongly correlated driver, and the size of the small domain (160x220), were not enough, for the model to find a relevant pattern that would allow it to predict correctly, with some accuracy, the occurrences of MHWs (Figure 5.1). On the test dataset, the model has an accuracy of 64%.

5.2 Architecture 2 - SETUP B

Therefore, following the literature, it was decided to expand and diversify the dataset by moving to "Setup B," containing another highly impactful driver such as geopotential height (GEO), and a larger analysis domain (241x521). In the literature, the GEO driver, together with the T2M driver, has given good results as already seen in Section 3. Whereas, the large domain, can improve the study, as it includes the whole area of northern Africa occupied by the Sahara

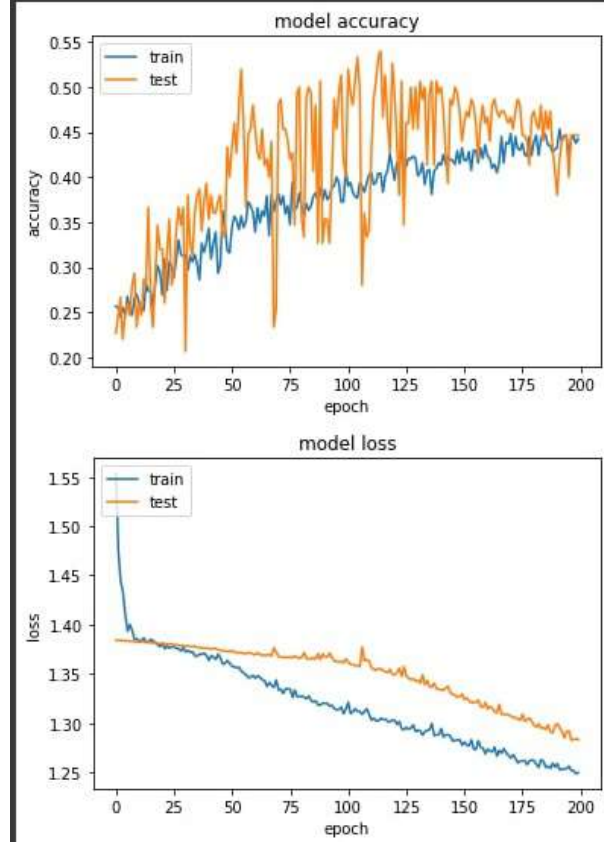


Figure 5.1: Report Model Accuracy - Model Loss for Setup A

Desert, characterized by the African subtropical anticyclone. Its characteristics are mainly identifiable in the high geopotential values, especially during the summer months.

The northward ascent of the African subtropical anticyclone to the Mediterranean results in the formation of a high pressure promontory. The warm currents from the Sahara, in addition to bringing a significant increase in above-average temperatures, as they cross the Mediterranean gradually become loaded with moisture, giving rise to sultry hot conditions along coastal areas.

During the summer season, the stabilizing presence of the African subtropical anticyclone, which is always accompanied by very high geopotential height values, is responsible for periods of widespread atmospheric stability with above-average temperatures, strongly influencing the generation of MHWs throughout

the Mediterranean Sea. Changing the input drivers, target classes, and the use of a larger domain produced results different from the previous ones, which can be considered a good starting point.

The results of Setup B with the CNN described in Figure 4.9, and the trend of "accuracy" and "loss" during training, show an overfitting problem [15]. In fact, it can be seen from Figure 5.2 that learning on the training dataset, leads the loss of the validation dataset to diverge.

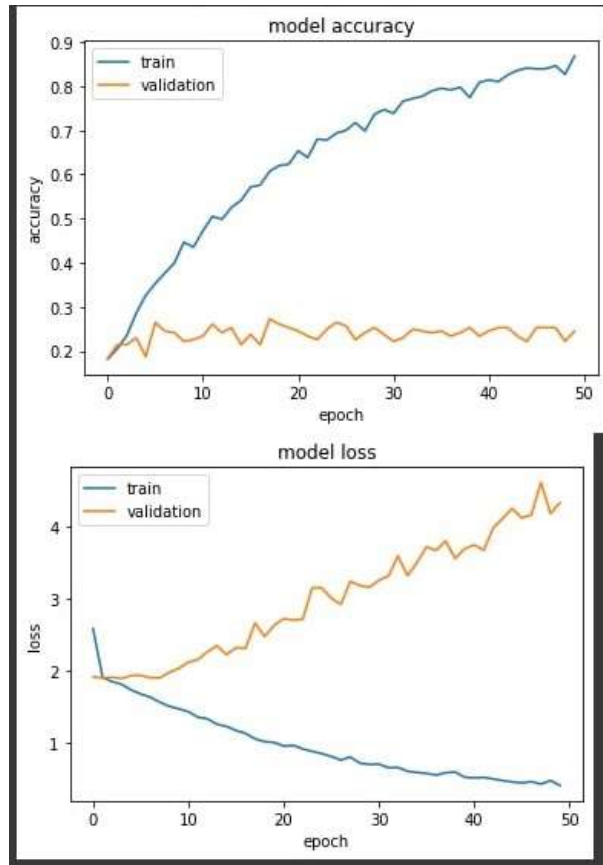


Figure 5.2: Report Model Accuracy - Model Loss for Setup B

So the model is perfectly learning only the training dataset and the more it learns, the more it errs on data it has never seen before. For this reason and others already mentioned in Section 4.3, we went to Setup B.2 with a different CCN, illustrated in Figure 4.10

5.3 Architecture 3 - SETUP B.2

The goal of the configurations in this section is to eliminate model overfitting by acting directly on the CNN architecture. To verify the overfitting problem, the confusion matrix was produced, using the training dataset (Figure 5.3)

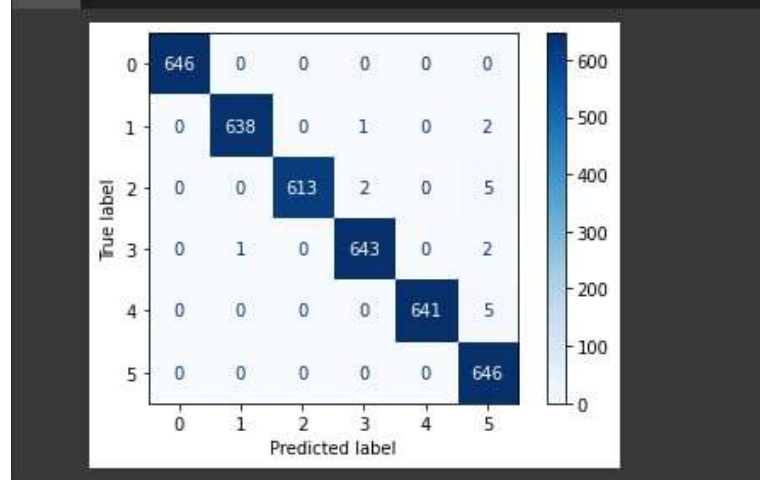


Figure 5.3: Confusion Matrix - Training dataset

First, the L2 regularization [6] was considered. This is a form of regression that constrains/adjusts or restricts coefficient estimates toward zero. In other words, this technique discourages learning a more complex or flexible model to avoid the risk of overfitting.

A simple relation for linear regression looks as follows. Here Y represents the learned relationship and β represents the coefficient estimates for the different variables or predictors (X). The fitting procedure involves a loss function, known as the sum of residual squares or RSS. The coefficients are chosen so as to minimize this loss function (Figure 5.4).

$$RSS = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2.$$

Figure 5.4: Residual Sum of Squares

Now, this will adjust the coefficients based on your training data. If there is noise in the training data, then the estimated coefficients won't generalize well to the future data. This is where regularization comes in and shrinks or regularizes these learned estimates towards zero.

Next, an additional convolutional layer was added before the flatten and 2 appropriately sized FC layers, so as to further balance the model parameters and avoid "bottlenecks". The results obtained in the fit phase are shown in Figure 5.5

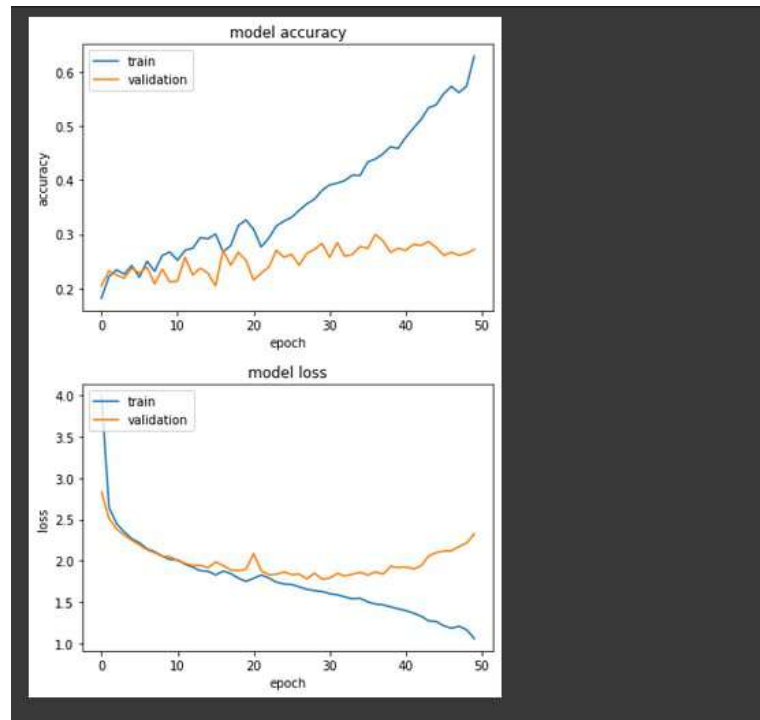


Figure 5.5: Report Model Accuracy - Model Loss for Setup B.2

It is evident how the Setups and Architectures used, produce an overfitting problem, which must be managed appropriately, working on the architecture so as to reduce the imbalance present. The resize of the input maps, which allows to further reduce the dimension of the activation volume, the right choice of layers and filter sizes, can produce relevant results that are not achieved with this first classification approach.

Conclusions and Future Work

In conclusion, the approaches studied in this paper failed to solve this difficult problem, but they helped to understand what issues need to be overcome and what may be the right solutions to do so. To try to explore this type of task further, other models could be implemented and adapted to the case study, such as including short-term memory layers (LSTM) in the network[24] to account for a sequence of days preceding the occurrence of MHWs on a single day or using a CapsNet, as seen in Section 3. Regarding the data, it is clear from this work and the works in the literature that the best drivers for making better prediction and classification of MHWs are temperature (T2M) the geopotential (GEO). Regarding the state of the art of the methods developed previously, it can be seen that the type of information used in the most successful experiments is based on SSTAs, which already identify anomalous events, or SSTs, which are directly related to the event MHWs and thus explicitly correlated. In contrast, the drivers provided may suffer from low correlation between inputs and outputs. Moreover, all these models and experiments show that the ML is very dependent on the type of data it is given. It is critical to balance the data available, use drivers with high correlations between input and output, use maps with the same domain and resolution, review the architecture of the CNN to be used so as not to create imbalances and not to have underfitting/overfitting problems. The lack of at least one of these requirements can be extremely detrimental to the training of any ML algorithm.

Sitography

Chapter

7

1. https://en.wikipedia.org/wiki/Capsule_neural_network
2. <https://arxiv.org/abs/1805.11195>
3. <https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-convolutional-neural-network-3607be47480>
4. <https://medium.datadriveninvestor.com/setting-up-google-colab-for-deeplearning-experiments-53de394ae688?gi=e3301a56b81d>
5. <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>
6. <https://discuss.tensorflow.org/t/create-capsnet-with-tensorflow-and-keras/6730>
7. <https://medium.com/@tecokids.monastir/fully-connected-layer-with-dynamic-input-shape-70c869ae71af>
8. https://it.frwiki.wiki/wiki/Sigmo%C3%AFde_%28math%C3%A9matiques%29
9. <https://medium.com/ml-cheat-sheet/understanding-non-linear-activation-functions-in-neural-networks-152f5e101eeb>
10. <https://towardsdatascience.com/visualizing-the-non-linearity-of-neural-networks-c55b2a14ad7a>
11. <https://proceedings.mlr.press/v51/lee16a.html>
12. <https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939>
13. <https://www.nature.com/articles/s43017-020-0068-4>

14. <https://www.cmcc.it/article/marine-heat-wave-in-the-mediterranean-observations-and-predictions>
15. https://sciencebrief.org/uploads/reviews/ScienceBrief_Review_MARINE_HEATWAVE_Oct2021.pdf
16. <https://www.frontiersin.org/articles/10.3389/fclim.2021.801217/full>
17. <https://www.sciencedirect.com/science/article/abs/pii/S0031320317304120>
18. <https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network>
19. <https://towardsdatascience.com/5-techniques-to-work-with-imbalanced-data-in-machine-learning-80836d45d30c>
20. <https://ieeexplore.ieee.org/document/9078901>
21. <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>
22. <https://pro.arcgis.com/en/pro-app/2.8/help/data/multidimensional/what-is-netcdf-data.htm>
23. <https://www.unidata.ucar.edu/software/netcdf/>
24. <https://climatedataguide.ucar.edu/variables/atmosphere/geopotential-height>
25. <https://climatedataguide.ucar.edu/climate-data/era40>
26. <https://www.sciencedaily.com/releases/2020/02/200204112518.htm>
27. <https://www.esri.com/en-us/arcgis-marketplace/listing/products/51fef6806f8a4339b07d59ef794bf716>
28. <https://www.science.org/content/article/scienceshot-predicting-heat-waves>
29. <https://journals.ametsoc.org/view/journals/bams/98/10/bams-d-16-0123.1.xml>
30. <https://gmd.copernicus.org/articles/12/2797/2019/>

31. <https://gmd.copernicus.org/articles/11/3999/2018/>
32. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001705>
33. <https://pure.uhi.ac.uk/en/publications/socioeconomic-impacts-of-marine-heatwaves-global-issues-and-oppor>
34. <https://www.nature.com/articles/s41598-020-75445-3>
35. <https://www.nature.com/articles/s41467-018-03732-9>
36. <https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017JC013326>
37. <http://www.marineheatwaves.org/all-about-mhws.html>

Bibliography

- [1] Nii O Attoh-Okine. Analysis of learning rate and momentum term in back-propagation neural network algorithm trained to predict pavement performance. *Advances in engineering software*, 30(4):291–302, 1999.
- [2] NA Bond, MF Cronin, H Freeland, and N Mantua. Causes and impacts of the 2014 warm anomaly in the ne pacific. *geophysical research letters*42: 3414-3420, 2015.
- [3] Letícia M Cavole, Alyssa M Demko, Rachel E Diner, Ashlyn Giddings, Irina Koester, Camille MLS Pagniello, May-Linn Paulsen, Arturo Ramirez-Valdez, Sarah M Schwenck, Nicole K Yen, et al. Biological impacts of the 2013–2015 warm-water anomaly in the northeast pacific: winners, losers, and the future. *Oceanography*, 29(2):273–285, 2016.
- [4] Ashesh Chattopadhyay, Ebrahim Nabizadeh, and Pedram Hassanzadeh. Analog forecasting of extreme-causing weather patterns using deep learning. *Journal of Advances in Modeling Earth Systems*, 02 2020.
- [5] Ke Chen, Glen G Gawarkiewicz, Steven J Lentz, and John M Bane. Diagnosing the warming of the northeastern us coastal ocean in 2012: A linkage between the atmospheric jet stream variability and ocean response. *Journal of Geophysical Research: Oceans*, 119(1):218–227, 2014.
- [6] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. L2 regularization for learning kernels. *arXiv preprint arXiv:1205.2653*, 2012.
- [7] Emanuele Di Lorenzo and Nathan Mantua. Multi-year persistence of the 2014/15 north pacific marine heatwave. *Nature Climate Change*, 6(11):1042–1047, 2016.

-
- [8] P. D. Dueben and P. Bauer. Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11(10):3999–4009, 2018.
 - [9] Rob A Dunne and Norm A Campbell. On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function. In *Proc. 8th Aust. Conf. on the Neural Networks, Melbourne*, volume 181, page 185. Citeseer, 1997.
 - [10] Andrew Estabrooks and Nathalie Japkowicz. A mixture-of-experts framework for learning from imbalanced data sets. 2189:34–43, 09 2001.
 - [11] Thomas L Frölicher, Erich M Fischer, and Nicolas Gruber. Marine heatwaves under global warming. *Nature*, 560(7718):360–364, 2018.
 - [12] Thomas L Frölicher and Charlotte Laufkötter. Emerging risks from marine heat waves. *Nature communications*, 9(1):1–4, 2018.
 - [13] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
 - [14] Daifang Gu and S. G. H. Philander. Interdecadal climate fluctuations that depend on exchanges between the tropics and extratropics. *Science*, 275(5301):805–807, 1997.
 - [15] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.
 - [16] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier, 1992.
 - [17] Alistair Hobday, Eric Oliver, Alex Gupta, Jessica Benthuyssen, Michael Burrows, Markus Donat, Neil Holbrook, Pippa Moore, Mads Thomsen, Thomas Wernberg, and Dan Smale. Categorizing and naming marine heatwaves. *Oceanography*, 31, 06 2018.
 - [18] Alistair J Hobday, Lisa V Alexander, Sarah E Perkins, Dan A Smale, Sandra C Straub, Eric CJ Oliver, Jessica A Benthuyssen, Michael T Burrows, Markus G Donat, Ming Feng, et al. A hierarchical approach to defining marine heatwaves. *Progress in Oceanography*, 141:227–238, 2016.

-
- [19] Neil Holbrook, Hillary Scannell, Alexander Sen Gupta, Jessica Benthuisen, Ming Feng, Eric Oliver, Lisa Alexander, Michael Burrows, Markus Donat, Alistair Hobday, Pippa Moore, Sarah Perkins-Kirkpatrick, Dan Smale, Sandra Straub, and Thomas Wernberg. A global assessment of marine heatwaves and their drivers. *Nature Communications*, 10, 06 2019.
- [20] Neil J Holbrook, Hillary A Scannell, Alexander Sen Gupta, Jessica A Benthuisen, Ming Feng, Eric CJ Oliver, Lisa V Alexander, Michael T Burrows, Markus G Donat, Alistair J Hobday, et al. A global assessment of marine heatwaves and their drivers. *Nature Communications*, 10(1):1–13, 2019.
- [21] Terry P Hughes, James T Kerry, Mariana Álvarez-Noriega, Jorge G Álvarez-Romero, Kristen D Anderson, Andrew H Baird, Russell C Babcock, Maria Beger, David R Bellwood, Ray Berkelmans, et al. Global warming and recurrent mass bleaching of corals. *Nature*, 543(7645):373–377, 2017.
- [22] Valérien Jacques-Dumas, Francesco Ragone, Freddy Bouchet, Pierre Borgnat, and Patrice Abry. Deep learning based extreme heatwave forecast. 03 2021.
- [23] Zahra Karevan and Johan Suykens. Transductive lstm for time-series prediction: An application to weather forecasting. *Neural Networks*, 125, 01 2020.
- [24] Jasmeen Kaur and Sukhendu Das. Future frame prediction of a video sequence. 08 2020.
- [25] Jennifer E Kay, Clara Deser, A Phillips, A Mai, Cecile Hannay, Gary Strand, Julie Michelle Arblaster, SC Bates, Gokhan Danabasoglu, James Edwards, et al. The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8):1333–1349, 2015.
- [26] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [27] Chen-Yu Lee, Patrick W. Gallagher, and Zhuowen Tu. Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. In

- Arthur Gretton and Christian C. Robert, editors, *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 464–472, Cadiz, Spain, 09–11 May 2016. PMLR.
- [28] Inderjeet Mani and I Zhang. knn approach to unbalanced data distributions: a case study involving information extraction. 126:1–7, 2003.
- [29] Amy McGovern, Kimberly Elmore, David Gagne, Sue Haupt, Christopher Karstens, Ryan Lagerquist, Travis Smith, and John Williams. Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bulletin of the American Meteorological Society*, 98, 03 2017.
- [30] Katherine E Mills, Andrew J Pershing, Curtis J Brown, Yong Chen, Fu-Sung Chiang, Daniel S Holland, Sigrid Lehuta, Janet A Nye, Jenny C Sun, Andrew C Thomas, et al. Fisheries management in a changing climate: lessons from the 2012 ocean heat wave in the northwest atlantic. *Oceanography*, 26(2):191–195, 2013.
- [31] Tom M Mitchell and Tom M Mitchell. *Machine learning*, volume 1. McGraw-hill New York, 1997.
- [32] Eric Oliver, Markus Donat, Michael Burrows, Pippa Moore, Dan Smale, Lisa Alexander, Jessica Benthuisen, Ming Feng, Alex Gupta, Alistair Hobday, Neil Holbrook, Sarah Perkins-Kirkpatrick, Hillary Scannell, Sandra Straub, and Thomas Wernberg. Longer and more frequent marine heatwaves over the past century. *Nature Communications*, 9, 04 2018.
- [33] Eric CJ Oliver, Jessica A Benthuisen, Nathaniel L Bindoff, Alistair J Hobday, Neil J Holbrook, Craig N Mundy, and Sarah E Perkins-Kirkpatrick. The unprecedented 2015/16 tasman sea marine heatwave. *Nature communications*, 8(1):1–12, 2017.
- [34] Eric CJ Oliver, Markus G Donat, Michael T Burrows, Pippa J Moore, Dan A Smale, Lisa V Alexander, Jessica A Benthuisen, Ming Feng, Alex Sen Gupta, Alistair J Hobday, et al. Longer and more frequent marine heatwaves over the past century. *Nature communications*, 9(1):1–12, 2018.

-
- [35] Keiron O’Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.
- [36] Alan F Pearce and Ming Feng. The rise and fall of the “marine heat wave” off western australia during the summer of 2010/2011. *Journal of Marine Systems*, 111:139–156, 2013.
- [37] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. *Advances in neural information processing systems*, 30, 2017.
- [38] S. Scher and G. Messori. Weather and climate forecasting with neural networks: using general circulation models (gcms) with different complexity as a study ground. *Geoscientific Model Development*, 12(7):2797–2809, 2019.
- [39] Sagar Sharma, Simone Sharma, and Anidhya Athaiya. Activation functions in neural networks. *towards data science*, 6(12):310–316, 2017.
- [40] Kathryn E Smith, Michael T Burrows, Alistair J Hobday, Alex Sen Gupta, Pippa J Moore, Mads Thomsen, Thomas Wernberg, and Dan A Smale. Socioeconomic impacts of marine heatwaves: Global issues and opportunities. *Science*, 374(6566):eabj3593, 2021.
- [41] S Sparnocchia, ME Schiano, P Picco, R Bozzano, and A Cappelletti. The anomalous warming of summer 2003 in the surface layer of the central ligurian sea (western mediterranean). In *Annales Geophysicae*, volume 24, pages 443–452. Copernicus GmbH, 2006.
- [42] Caroline C Ummenhofer and Gerald A Meehl. Extreme weather and climate events with ecological relevance: a review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1723):20160135, 2017.
- [43] Thomas Wernberg, Scott Bennett, Russell C Babcock, Thibaut De Bettignies, Katherine Cure, Martial Depczynski, Francois Dufois, Jane Fromont, Christopher J Fulton, Renae K Hovey, et al. Climate-driven regime shift of a temperate marine ecosystem. *Science*, 353(6295):169–172, 2016.
- [44] Jonathan Weyn, Dale Durran, and Rich Caruana. Can machines learn to predict weather? using deep learning to predict gridded 500-hpa geopotential

height from historical weather data. *Journal of Advances in Modeling Earth Systems*, 11, 08 2019.

List of Figures

1.1	Categorization schematic for marine heatwaves (MHWs) showing the observed temperature time series (dashed line), the long-term regional climatology (bold line), and the 90th percentile climatology (thin line). Multiples of the 90th percentile difference ($2\times$ twice, $3\times$ three times, etc.) from the mean climatology value define each of the categories I–IV, with corresponding descriptors from moderate to extreme. This example peaked as a Category IV (extreme) MHW [17].	3
1.2	The map was developed using the NCAR command language. Major recent marine heat waves documented in the literature. The image shows the maximum sea surface temperature anomaly in regions where the temperature exceeds the 99th percentile, using the National Oceanic and Atmospheric Administration (NOAA) daily sea surface temperature dataset with optimal interpolation. The 99th percentile is calculated over the reference period 1982-2016 [12].	5
1.3	Ligurian Sea. Average temperatures and the MHW threshold (defined as the 90th percentile of temperatures for each date) are calculated from long-term observed ("obs.") satellite data. Near real-time satellite data are the most up-to-date data provided by the satellites themselves. Orange shading is used to highlight when the observed temperature is above the MHW threshold.	6

1.4	Gulf of Taranto. Average temperatures and the MHW threshold (defined as the 90th percentile of temperatures for each date) are calculated from long-term observed ("obs.") satellite data. Near real-time satellite data are the most up-to-date data provided by the satellites themselves. Orange shading is used to highlight when the observed temperature is above the MHW threshold.	7
1.5	Anomaly is the difference between predicted and average temperatures (i.e., red indicates warmer than normal). The dark red outline highlights where the MHW threshold was exceeded for that day. .	8
1.6	A wide range of negative impacts of MHWs	9
1.7	Socioeconomic impacts of major marine heatwave (MHW) events since 1995[40].	10
2.1	An simple CNN architecture, comprised of just five layers.[35] . .	14
2.2	Illustration of Convolution Operation	15
2.3	Illustration of Max Pooling Operation with 2x2 filters and stride 2	16
2.4	Illustration of Fully-connected Layer	16
2.5	Activation Functions Graph	17
3.1	CNN-based heatwave predictor architecture. The output is associated with the probability of occurrence of an heatwave in the upcoming t days.[22]	19
3.2	Snapshot of the surface temperature surface fluctuations (Ts fluctuations, according to the color bar, in Kelvin) and of the geopotential height at 500mbar (Zg fluctuations, contours) over the Northern Hemisphere. This snapshot is taken on July 20th on a arbitrary year of the PlaSim simulation. The spatial resolution is 64×128 (latitude \times longitude). The thin contour lines, representing the anomaly of Zg, are in meters. The thick black contour delimits the zone that is used for prediction by the machine learning procedure[22]. . . .	20

3.3	Cluster centers of T2m anomalies at the onsets and Z500 patterns of 3 days earlier. The top (bottom) two rows correspond to summers (winters). S0 (W0) shows the average of T2m and Z500 patterns from days with no heat wave (cold spell). S1–S4 and W1–W4 are obtained from K-means clustering the anomalous T2m patterns at onsets into four classes, which roughly separates the extreme events into four geographical regions: Northern Canada (S1), Western United States-Canada (S2), Southern United States (S3), and Eastern United States-Canada (S4) in summers and Northwest United States-Canada (W1), Alaska (W2), Northeast United States-Canada (W3), and Northern Canada (W4) in winters. Rows 1 and 3 show the cluster centers, while rows 2 and 4 show the average of Z500 patterns 3 days before the onsets for each cluster[4].	21
3.4	Illustration of the data-driven structure for cold wave prediction based on Z500 data from 3 days before [4].	22
4.1	The obtained cluster patterns are depicted by averaging the mean intensity of all event maps corresponding to each cluster. The contours represent the average anomaly of SST (Sea Surface Temperature).	23
4.2	Example of six input variables stacked in an input volume for the date 2008-12-04.	24
4.3	Examples of netCDF data: Left—Data structure; Right—Temperature at a height of 2 meters.	25
4.4	Undersampling technique.	28
4.5	The obtained cluster patterns are depicted by averaging the mean intensity of all event maps corresponding to each cluster. The contours represent the average anomaly of SST (Sea Surface Temperature).	29
4.6	Example of four input variables stacked in an input volume for the date 01-01-2016.	30
4.7	CNN-based heatwave predictor Architecture 1 - SETUP A.	37
4.8	Illustration of the Flatting operation.	38
4.9	CNN-based heatwave predictor Architecture 2 - SETUP B.	39
4.10	CNN-based heatwave predictor Architecture 3 - SETUP B.2.	40

4.11 CapsNet with 3 layers. The length of the activity vector of each capsule in DigitCaps layer indicates presence of an instance of each class and is used to calculate the classification loss.	42
4.12 Softmax Cross-Entropy Pipeline.	44
4.13 Cross-Entropy loss	44
5.1 Report Model Accuracy - Model Loss for Setup A	47
5.2 Report Model Accuracy - Model Loss for Setup B	48
5.3 Confusion Matrix - Training dataset	49
5.4 Residual Sum of Squares	49
5.5 Report Model Accuracy - Model Loss for Setup B.2	50

List of Tables

4.1	Example of Target data. "1.0" indicates the presence of the occurrence, "0.0" otherwise.	26
4.2	Example of merging occurrences of MHws with the same time reference before and after preprocessing	27
4.3	Method of assigning new classes	27
4.4	Occurrences by class - Setup A	28
4.5	Example of Target data. "1.0" indicates the presence of the occurrence, "0.0" otherwise.	32
4.6	MHws with the same time reference before and after preprocessing	32
4.7	Occurrences by class - Setup B	33
4.8	Occurrences by class - Setup B.2	34