BABEŞ BOLYAI UNIVERSITY, CLUJ NAPOCA, ROMÂNIA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Automatic Detection and Classification of Atmospherical Fronts

– MIRPR report –

Ploscar Andreea Alina, Muscalagiu Anca Ioana
Informatica Engleza, group 935

2022-2023

Abstract

Automatic determination of fronts from atmospheric data serves as an important aid for weather forecasting, as atmospherical fronts provide strong indicators of certain meteorological characteristics. This paper proposes an application that detects and classifies air fronts in an automatic manner, using Convolutional Neural Networks for pattern recognition in images of type "synoptic map", which contain multiple weather parameters. The user can upload a synoptic map and the application informs him if in this image there is an existing atmospherical front on Romania's territory. If there is any front detected, then the algorithm will also classify it in one of three categories: cold, warm or mixed front.The data used to train and test this CNN model is collected from different meteorological stations around the world, between 2013 to 2020.

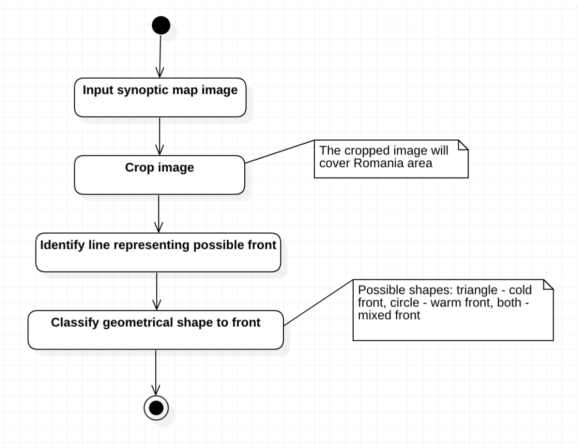

Figure 1: Work Flow

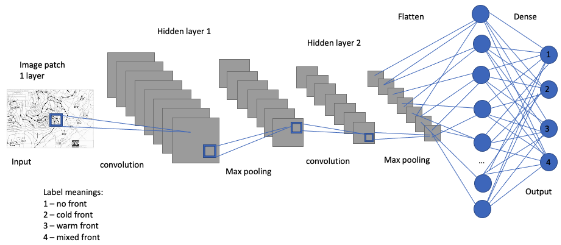

Figure 2: Initial Model

Contents

List of Figures

Chapter 1

Introduction

1.1 What? Why? How?

This paper aims to use Convolutional Neuronal Networks on synoptic maps collected from meteorological stations to determine the existence and category of fronts over Romania's territory in a specific moment. Front detection is a subject addressed very little in the past years and only for broader territories like Europe and America. Our model proposes to improve the accuracy of air fronts detection for smaller regions like Romania.

Detecting and classifying atmospherical fronts recently become an important topic not only in the meteorological field, but also in the medical field. The specialists have observed a certain correlation between the number of strokes and the particular fronts existent on a certain territory over a longer period of time. Therefore, developing an algorithm that can correctly and automatically determine and classify the type of front over a specific territory is an important step in researching the exact correlation of health risks and specific weather characteristics that persons suffering of a stroke were exposed to over a longer period of time, helping in preventing and reducing the risk of this condition occuring again.

1.2 Paper structure and original contribution(s)

The research presented in this paper advances the theory, design, and implementation of several particular models.

The main contribution of this report is to present an intelligent algorithm for solving the problem of determining and classifying atmospherical fronts.

The second contribution of this thesis consists of building an intuitive, easy-to-use and user friendly software application. Our aim is to build an algorithm that will help specialists in the meteorological field forecast different weather characteristics on smaller regions, more specifically on Romania's territory. The model proposed in this report aims to serve as a starting point for further reasearch on the corellation between weather attributes and stroke frequency in certain area.

Chapter 2

Scientific Problem

2.1 Problem definition

The application detects air fronts represented by means of geometric shapes (lines, semicircles, triangles) in synoptic maps with the help of Convolutional Neural Networks , then classify the fronts in one of 4 classes: no existing air front, warm, cold or mixed front.

The input data of the problem are represented by sets of "synoptic map" type images, collected from different weather stations, in which different meteorological characteristics are represented: air pressure, temperature and humidity, baric tendency, wind direction and speed. Thus, the application aims to detect the location of Romania from the analyzed weather maps, cut the images to contain only the territory of interest and process them to detect the atmospheric sources, represented by different geometric shapes.

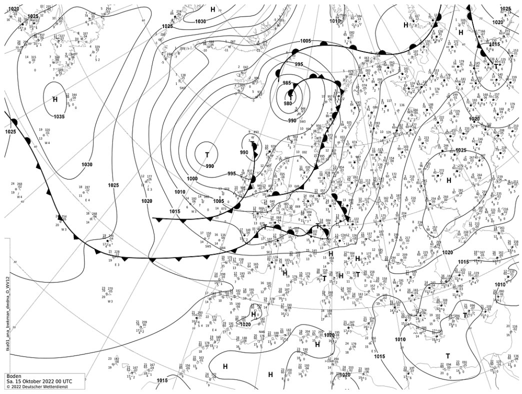

Figure 2.1: Synoptic Map

The atmospheric front (or air front) represents the transition between two air masses different in density or temperature. Their contact can cause radical weather changes, such as precipitation, temperature or pressure variations. In synoptic type weather maps, the two air masses are delimited by continuous lines, and the type of air front is determined by different geometric shapes: semicircle for warm fronts, triangles for cold fronts and alternating triangles and semicircles for mixed fronts. The type of the front is determined by the dominating type of air mass: cold or warm.

Figure 2.2: Cold Front

Figure 2.3: Warm Front

Figure 2.4: Stationary Front

Figure 2.5: Occluded Front

The output for this model is represented by one of the following classes: No air front, Cold, Warm or Mixed front. For the purpose of this paper both the Stationary and the Occluded fronts are considered as Mixed fronts. The labels used for this classifications are: 1 - no front, 2 - cold front, 3 - warm front, 4 - mixed front.

Chapter 3

State of the art - Related work

The detection and classification of air fronts is usually addressed manually, but the demand for an automatic approach increased along with the dataset volume. There have been multiple approaches in creating an automatic model, but the most relevant work for this study is S. Niebler et al.: Automatic detection and classification of fronts, 2022.

This paper uses labeled data from two weather services: the North American National Weather Service and the Deutscher Wetterdiens. The study introduces a deep neural network to detect and classify fronts from multi-level ERA5 reanalysis data, putting an emphasis on simplifying the post processing step. With a supervised learning approach, the method used is a CNN that automatically learns atmospheric features that correspond to the existence of a weather front. For each spatial grid point the algorithm predicts a probability distribution, the likelihood of the point belonging in one of the five classes: warm front, cold front, occlusion, stationary front or background.

In comparison with this paper, our study focuses on one particular area from the eastern Europe, covering Romania. Thus, the output of our algorithm contains a single result, one of the four classes. The occluded and the stationary front mentioned in this paper are considered as mixed fronts in our model because identifying both triangles and circles in out image would lower the precision of classifying both classes, information that is not relevant to us. Another important distinction is the presence of at most one front in our images, because the covered area is smaller, therefore there is no possibility for wrongly classifying mixed front with alternating cold and warm fronts.

Chapter 4

Investigated approach

The approach in this paper involves a few steps of preprocessing of the input data. The initial images received from the weather stations cover the Europe continent, containing multiple or even all types of fronts, making the result of the classification meaningless. In order to have the input data which can be classified in one predominant class, the synoptic maps were splitted into 16 equally sized tiles. This method lowers the chance of an image containing multiple fronts, as it covert a smaller area. The resulting images were prelabeled and organized in training set, validation set and test set, each one of them being classified further in one of the 4 output classes.

The algorithm used is based on the classical model for image recognition through supervised learning, making use of Convolution Neural Networks with multiple convolution, pooling and dropout layers. The CNN is organized in multiple 2D convolution layers, taking into consideration the format of the input image (black and white) which is represented with only one slice, average pooling and dropout layers, which help prevent overfitting. The last step of the neural network is the flattening phase, outputting the probability of the input belonging to each of the 4 classes. For the first layer the number of input channels is 1, because the algorithm uses black and white images. After the 2d convolution is applied, the layer performs in addition a batch normalization and a rectified linear unit function (ReLU). All of the convolutional and pooling layers form the sequential layer, which is followed by the flattening phase.

The training phase starts with no accuracy, slightly increasing over the epochs. In each epoch the algorithm trains the model and iterates over batches of images loaded from the train dataset. The classes are predicted using images from the current batch and the output is compared to the actual labels. The loss function used is cross entropy and it backpropagates the loss into the network. This allows the parameters to be adjusted according to the computed gradients. After the iteration, the learning rate is adjusted, decreasing with the current epoch. The train loss and accuracy are measured using the predictions and the actual labels for all images from each batch. The testing phase evaluates the model on the test dataset and saved the model if a greater accuracy is obtained.

Chapter 5

Methodology

5.1 Base Method

From an experimental point of view, we approached this problem using a dense CNN with the following architecture :

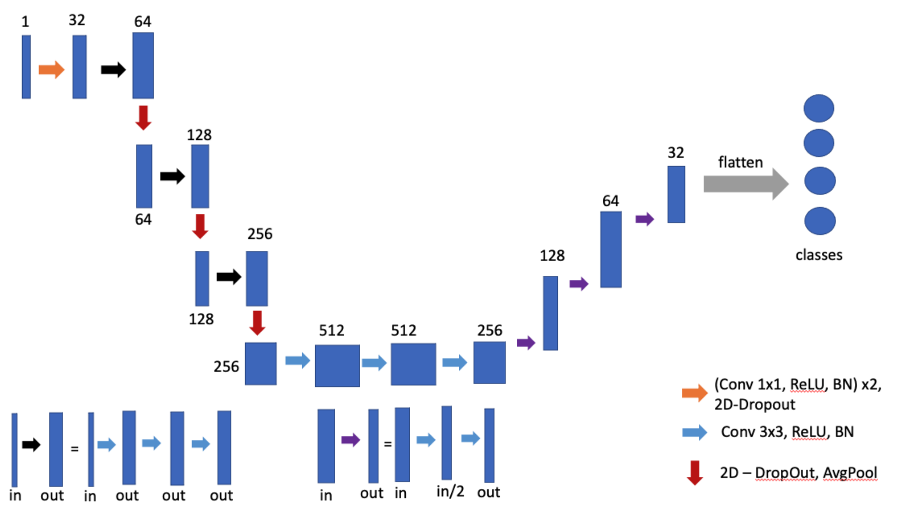

Figure 5.1: CNN model

The structure is inspired by Niebler in [1], being simplified for our use case with front classification in black and white photos. Some of the layers that we removed include the upsampling layers and their following convulsion layers.

In order to evaluate our method we are measuring the number of correctly classified fronts, computing the overall and by class accuracy. We also compute a confusion matrix to better visualize the performance of our CNN model. We display this matrix as a heatmap and we plot the evolution with each epoch of the overall accuracy.

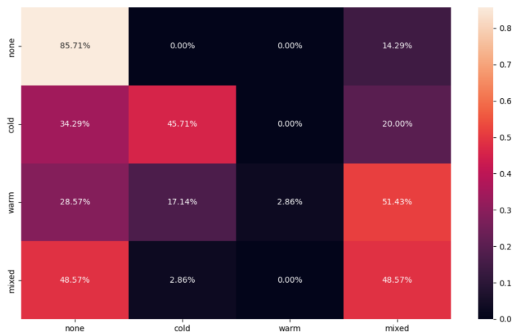

Figure 5.2: Confusion matrix as a Heatmap

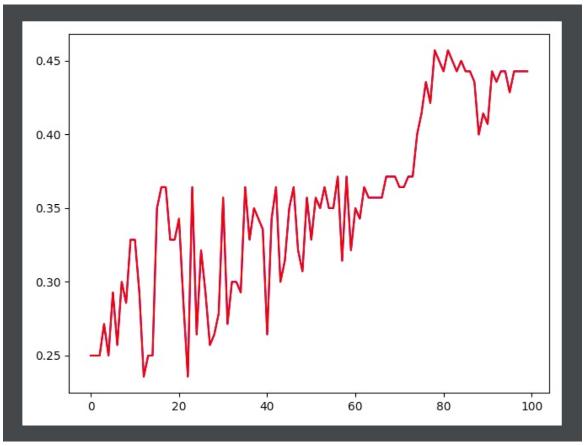

Figure 5.3: Accuracy evolution

This paper tests the hypothesis that atmospheric fronts covering a small area (less that 350000 sq km) can be classified by a CNN with an accuracy over 80%. Our experiments starts with loading the data, being split in train and validation data. These two are shuffled into batches of 32 images, in order to maximize the probability of having each type of front in all batches to aquire knowledge about all fronts after processing each batch. For each epoch we measure the train loss and train and test accuracy. If the test accuracy is better the best accuracy measured so far we save the model and upgrade the best one. We train the model over 50 epochs, saving intermediary best models.

In comparison with other articles, because the aim of our experiment was classification and not detection of the fronts [1], we were able to measure the performance of our algorithm numerically, as described above, through the number of correctly classified pictures, making use of all the classic performance assessment methods used in the ML field.

To avoid overfitting to one class, we have used equally sized datasets from each category for the training and testing data. In order to increase the performance of our CNN in real life situations we have used both simple, easily identifiable fronts and more complex pictures illustrating multiple types of fronts, having a predominant one. This makes our datasets more realistic and relevant.

5.2 Improvements

5.2.1 Version 2

The structure for our second version is a simplification of our initial model, which was based on Niebler's model for detection of fronts. Since our task only focuses on classification of fronts, without the need of their location, the base model network was too complex for our use-case in terms of number of convulsions applied to our image. Therefore some of the initial layers that we removed include the convulsion layers that formatted the image up to 256 and 512 channels, because we don't need the same granularity in classifying the images as it was needed in detecting the front locations. In addition to removing the convulsion layers, we used a different type of pooling, maximum pooling, which, in comparison to the average we used before, brings out the features of the image in order for our CNN to more easily identify the front lines within the background.

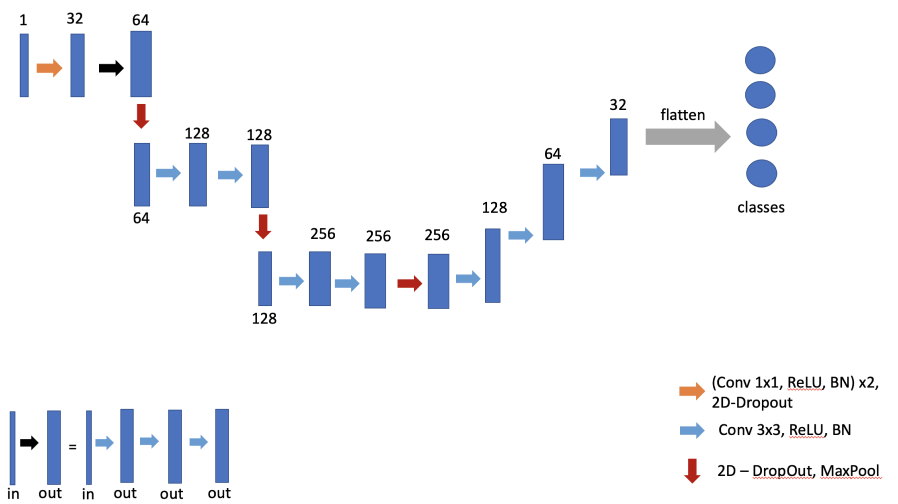

Figure 5.4: CNN model 2

The improvement in performance of our model can be identified through our confusion matrix and accuracy plot, which shows better accuracy of front identification for each class in comparison to our base version:

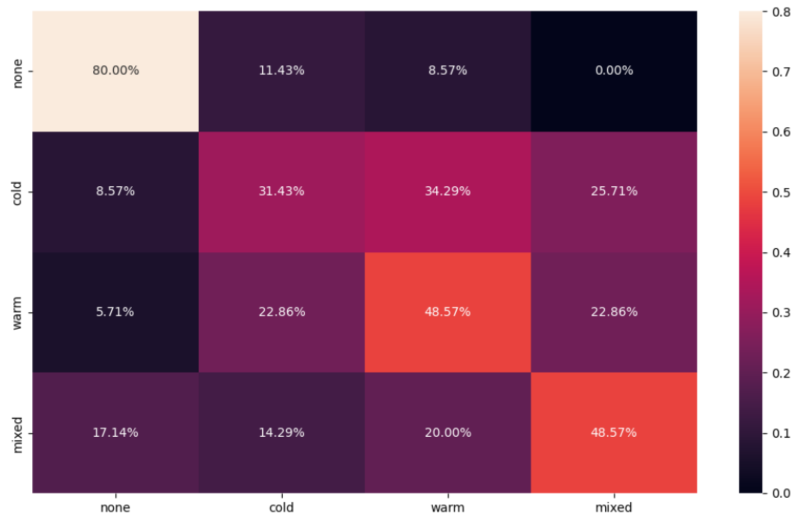

Figure 5.5: CNN model

5.2.2 Data

The consistent datasets were downloaded from Wetterzentrale, a German weather service which provides synoptic maps daily. The format of the map is presented in the picture below:

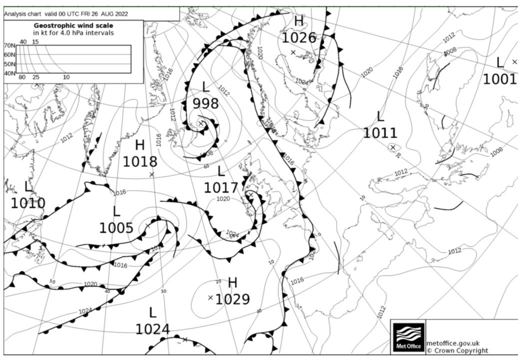

Figure 5.6: Synoptic Map

The images were split in 16 tiles, with the purpose of isolating fewer types of fronts in one picture. Although this process reduces the quality of the images, this does not affect the purpose of the study. The tiles were manually labelled into one of the 4 categories by the most predominant front, if there were multiple ones present in the picture. To generate more data, we augmented the data by rotating the tiles with 90, 180 respectively 270 degrees, thus generating another 3 front lines with different directions. Using this method, we gathered roughly 200 pictures per class from which 15% were used for validation purposes.

5.2.3 Results

5.2.3.1 Base Method

The best model we obtained during the training phase had an accuracy of 0.457% and was saved during the 49 epoch. We provide the accuracy overall and the accuracy per class evolution in the following plots:

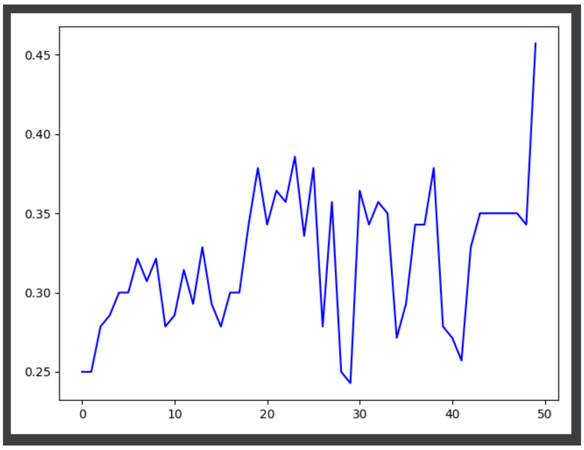

Figure 5.7: Overall Accuracy Evolution

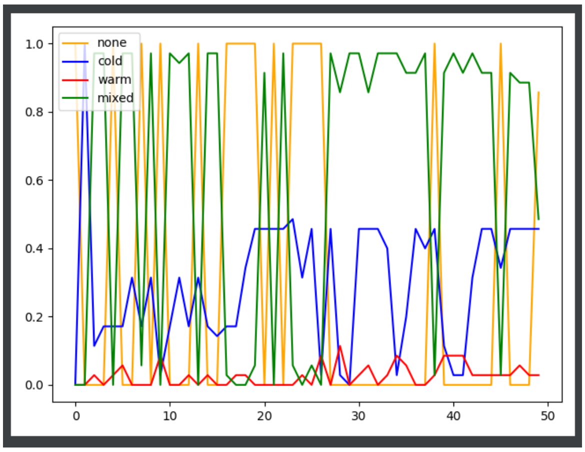

Figure 5.8: Accuracy Per Class Evolution

Chapter 6

Conclusion and future work

The main conclusion that can be drawn is that Automatic Front Detection can be a very efficient solution for the issues we currently face in weather prediction and medical analysis. Our study demonstrates that a supervised learning model, even with a relatively small training dataset, can accurately classify the types of fronts present on a small area from a synoptic map. One of the most important strengths of our research paper is that our experiment paves the path for further research and for discovery of new approaches in the domain, which is currently very little represented in the research world. Considering that ,to our knowledge, our research is one of the first to approach front classification on small areas, we have encountered a few shortcomings in our experiments. One of the most relevant weaknesses to our approach is the training dataset, which is very hard to obtain and label, because they are not publicly available in larger quantities. In addition to this, the synoptic maps might not be consistent in the future, depending on how metheorologists decide to represent the fronts. In conclusion, there is definitely room for improvement for our model, that consists one of the first models that tries to solve this problem. Two possible directions that could increase the perforance are larger, more qualitative trainig datasets and more efficient processors that could support more epochs of training for it.

Bibliography

[1] Stefan Niebler, Annette Miltenberger, Bertil Schmidt, and Peter Spichtinger. Automated detection and classification of synoptic-scale fronts from atmospheric data grids. *Weather and Climate Dynamics*, 3(1):113–137, 2022.