BABEŞ-BOLYAI UNIVERSITY
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Machine learning models for weather nowcasting

PhD thesis summary

PhD student: Mihai Andrei
Scientific supervisor: Prof. dr. Czibula Gabriela

2021

Keywords: Weather nowcasting, Machine learning, deep learning, relational association rule mining, deep neural networks, self-organizing maps.

Contents

Thesis table of contents

List of publications

The ranking of publications was performed according to the CNATDCU (National Council for the Recognition of University Degrees, Diplomas and Certificates) standards applicable for doctoral students enrolled after October 1, 2018. All rankings are listed according to the classification of journals[1] and conferences[2] in Computer Science.

Publications in Web of Science - Science Citation Index Expanded

[CMA+21] Gabriela Czibula, **Andrei Mihai**, Alexandra-Ioana Albu, Istvan Czibula, Sorin Burcea, Abdelkader Mezghani. *AutoNowP: An approach using deep autoencoders for precipitation nowcasting based on radar echo prediction*. Mathematics, Special Issue on "Computational Optimizations for Machine Learning". 2021; in press. (**2020 IF=2.258**).

Rank A, 2 points.

[CMt21] Gabriela Czibula, **Andrei Mihai**, Eugen Mihuleţ. *NowDeepN: An ensemble of deep learning models for weather nowcasting based on radar products' values prediction*. Applied Sciences, Special Issue on"Applied Machine Learning". 2021; 11(1):125. (**2020 IF=2.679**).

Rank B, 4 points.

Publications in Web of Science, Conference Proceedings Citation Index

[CMC19b] Gabriela Czibula, **Andrei Mihai**, Istvan G. Czibula, *RadRAR: A relational association rule mining approach for nowcasting based on predicting radar products' values.* 24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES2020), Procedia Computer Science, Volume 176, 2020, pp. 300-309.

Rank B, 4 points.

[CMtT19] Gabriela Czibula, **Andrei Mihai**, Eugen Mihuleţ and Daniel Teodorovici. *Using self-organizing maps for unsupervised analysis of radar data for nowcasting purposes*. 23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, 2019, Procedia Computer Science Vol 159, (2019) pp. 48–57.

Rank B, 2 points.

[1] https://uefiscdi.ro/premierea-rezultatelor-cercetarii-articole
[2] http://portal.core.edu.au/conf-ranks/ Source CORE 2018

[CMC19a] Gabriela Czibula, **Andrei Mihai**, Liana Maria Crivei. *SPRAR: A novel relational association rule mining classification model applied for academic performance prediction.* International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES), 2019, Procedia Computer Science Vol 159, (2019) pp. 20–29

Rank B, 4 points.

[CAG19] Liana Maria Crivei, **Andrei Mihai**, Gabriela Czibula. *A study on applying relational association rule mining based classification for predicting the academic performance of students.* The 12th International Conference on Knowledge Science, Engineering and Management (KSEM), LNAI 11775, 2019, pp. 287-300.

Rank B, 4 points.

[MCt19] **Andrei Mihai**, Gabriela Czibula and Eugen Mihuleţ. *Analyzing Meteorological Data Using Unsupervised Learning Techniques.* 2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP), IEEE Computer Society Press, pp. 529 - 536. (**ISI Proceedings**).

Rank C, 2 points.

[Mih20] **Andrei Mihai**. *Using self-organizing maps as unsupervised learning models for meteorological data mining.* IEEE 13th International Symposium on Applied Computational Intelligence and Informatics, SACI 2020, Timioara, pp. 23–28

Rank C, 2 points.

[SCIM20] Ioana Angela Socaci, Gabriela Czibula, Vlad-Sebastian Ionescu, **Andrei Mihai**. *A deep learning technique for nowcasting based on radar products' values prediction.* IEEE 13th International Symposium on Applied Computational Intelligence and Informatics, SACI 2020, Timioara, pp. 117–122

Rank C, 1 point.

Publications score: 25 points.

Introduction

The main research domain of our doctoral thesis is *deep learning* applied in the field of meteorology. Our PhD thesis is entitled "Deep learning models for weather nowcasting" and aims at developing new machine learning models for improving weather nowcasting prediction.

Predicting weather, and particularly severe weather, is an important challenge both for the meteorological and machine learning researchers. The complexity and difficulty of the problem is mainly due to the chaotic character of the atmosphere and the implicit large set of meteorological information (radar, satellite or ground meteorological observations) which have to be analyzed by meteorologists. Thus, understanding the relationships between various meteorological parameters extracted from radar observations may be useful for providing additional comprehension about severe weather development and would help to identify situations when severe weather can occur.

The problem of issuing a nowcasting warning can be very difficult for meteorologists, since there is often an extremely large set of meteorological data (available in the form of radar, satellite or ground meteorological observations) which has to be analyzed in a very short period of time and the constraints imposed on a desirable solution are complex and not exactly known. Therefore, *Machine Learning (ML)* based methods (particularly the *supervised* ones) are necessary for obtaining effective solutions for the nowcasting problem. In addition, the *unsupervised* learning methods are useful for extracting accurate and meaningful patterns from the large amount of weather related data and to improve decision-making for high-impact weather.

Firstly, in order to better understand the data, we employed *unsupervised learning*. *Unsupervised learning* is a sub-field of ML dealing with algorithms which extract useful information from raw data without using labeled examples (as in supervised learning). For example unsupervised learning methods could organize data in some clusters based on some given similarity function, or provide an encoder and decoder with which to compress and decompress data of some type, or maybe extract some rules from the data given a structure of the rules. For this we relied on *self-organizing maps* (SOMs) [SK99], a type of unsupervised Artificial Neural Network (ANN).

Then, based on the information extracted, we created *supervised learning* models, for meteorological data prediction. *Supervised learning* is a subfield of machine learning, dealing with the task of approximating a mapping from some input domain to some output domain based on example input-output pairs. A *supervised learning algorithm* generalizes the training data, producing a function that, given an input can return a close enough approximation of the correct output. In order to create supervised learning models we mainly focused on *Neural Network* models. *Neural networks* have been modeled to be similar to complex webs of neurons. This morphology has been adopted in computer science, by building densely interconnected systems that have as building blocks basic units, that take as input a series of real-valued numbers and produce a single real-valued output [Mit97]. We have also experimented with creating supervised learning models using *Relational Association Rules*. *Relational Association Rules* represents an important data analysis and mining technique useful in multiple ML tasks as they are able to express different type of non-ordinal relations between data attributes.

Approached Problem

The main problem that was approached in our work is *nowcasting* in the domain of meteorology. The term "nowcasting" is derived from the contraction of "now forecasting", intended to mean forecasting of very short term events. *Weather nowcasting* is the short term analysis and forecast of weather patterns, generally for the next 0 to 6 hours, and is of major interest within the meteorological research.

As stated by the World Meteorological Organization (WMO) [WMO18] weather, and particularly severe weather, causes many natural disasters and is responsible for lots of damages and loss of life. Since the number and intensity of severe weather events is increasing in various regions of the world, the problem of forecasting such phenomena and issuing early warnings is nowadays one of the most popular topics in meteorology. Accurate weather nowcasting is a key element for issuing relevant early warnings.

While Numerical Weather Prediction (NWP) can be quite successfully used for general weather forecasting, for very short term predictions – nowcasting – it not as usable, as it uses accurate simulations of the physical equations governing the atmospheric model, thus needing a lot of time and computational power to make predictions [TSM21]. For this reason, most state-of-the art nowcasting systems used employ other methods – i.e. extrapolation of meteorological data [SCW+15]. This highlights one aspect that makes weather nowcasting such a complex topic: predictions need to be fast, so that early warnings can be issued as early as possible.

The other aspect that makes weather nowcasting a complex task is the sheer quantity of data available, that needs to be analysed to make good nowcasting predictions. First of all, there are many data sources, that all could be relevant. There are many meteorological satellites that generate data continuously, tracking data about elements such as temperature, winds or clouds; while on Earth there are terrestrial stations constantly gathering real-time data, from radar stations to surface-water gauging stations measuring rainfall and flooding. More recently, relevant data can be gathered from items such as solar panels, smart thermometers and smart online air conditioners. In many nowcasting systems radar data is used as the source for prediction. Even so, a radar gathers data from hundreds of thousands square kilometers, on many elevations and outputs tens of different products that each give some kind of information about the current weather. It is hard for operational meteorologists to analyse all this data, and usually they have a subset of most relevant elevations and products that they use for prediction and early warning issuing.

Even more, meteorological institutes hold a large set of historical meteorological data, such as radar measurements, satellite data and historical meteorological observations. This historical data could give important insights for weather patterns and manifestations, which could help create better predictions. But the historical data is too large to be analysed by meteorologists, thus automated systems are needed. Data mining techniques are particularly suitable fur such tasks.

Taking these elements in consideration, we decided approaching the weather nowcasting problem from a machine learning perspective. Once a proper machine learning model is created and trained it can give fast predictions based on new data. A machine learning model is trained using existing data. In meteorology there is too much historical data to be analysed by hand, while in machine learning the more training data, the better the model. Therefore, machine learning seems a good fit for weather nowcasting as it can take advantage of the huge amount of existent historical meteorological data. Also, there are unsupervised machine learning methods that can be used to analyse the historical data and retrieve meaningful patterns and information.

In this work we aim creating machine learning models for weather nowcasting purposes. The goal was to create new models that could potentially be incorporated in either existing and already operational systems or new, state-of-the-art nowcasting systems. In order to do this we chose subset of meteorological data to use for the proof of concept of our models. We focused our efforts on radar data, and we used real historical data provided to us by the Romanian National Meteorological Administration. While the main focus was to create models for meteorological data prediction, we also created unsupervised models for data analysis, in order to extract information from the historical data.

Original Contributions

Our research was focused on two main directions: (1) investigating *unsupervised learning* models (such as *self-organizing maps* and *relational association rule mining*) for historical meteorological data analysis and information extraction; and (2) developing *supervised learning models* for weather nowcasting prediction. For (2) we focused on *deep learning* models, such as *deep neural networks*, *recurrent neural networks* and *convolutional neural networks*. And such, our results and main contributions are also separated in these two directions, presented in Chapters 2 and 3:

(1) **Unsupervised learning models for radar data analysis.**

Seeing the nature of the problem and the amount of historical data available, we considered that there would be interest in analysing the data to find patterns or meaningful information. To this end we proposed using *self-organizing maps (SOMs)* for unsupervised radar data analysis. We did these experiments with two goals in mind: devising useful SOM-based methods for radar data analysis and extracting some information from the data we had to use further in the modeling of the radar data predictor. Our results on this line of research were the following:

a) Our first experiment had the goal to uncover how the values of the radar products evolve between consecutive radar scans. The details and results of our work on this idea resulted by being published in [CMtT19]. The methodology and results are also detailed in Section 2.2. We have shown that in general the values of radar products change slowly over time, except some specific moments during severe events. We have also shown that the data is very similar in periods where there are no meteorological events and that, during significant events, one particularly noisy product (V – Velocity) cannot be ignored, as the data is not as well described without it.

b) In our next study, we analysed the change in radar product values at a much lower level: we searched for patterns in how values change, for one specific product, at one specific timestamp (moment). We chose to analyse one highly relevant product (R02 – Reflectivity at the second lowest elevation) at a timestamp in the middle of the severe weather event and one before the beginning of the severe event. The results of this work were published in [MCt19]. We describe the methodology and results in detail in Section 2.3 of this thesis. We have found empirical evidence that similar values for a radar product at a given moment are encoded in similar neighborhoods at previous time moments, thus showing that a meaningful relation exists between the value of the product at a moment and its neighbourhood at the previous moment, a relation that can be used by supervised algorithms for prediction. We have also shown that the same pattern holds both for normal and severe weather conditions, and also holds if we consider 1 or 5 previous moments, showing that there might be possible to create predictions only from one previous step (making training and predictions faster).

c) Since the previous experiment was done only on one specific product, we wanted to verify if the same patterns appear in other radar products we considered using for prediction.

Thus we extended the experiment to also study data from the lowest elevation, and also 2 other products (V – Velocity and VIL – Vertically integrated Liquid). The results of this extention of experiments were published in [Mih20]. In this thesis, the details and results are presented in Section 2.4. We have shown that the same relation and patterns appears for all the considered products and elevations.

(2) **Supervised Learning Models for weather nowcasting.**

Our second direction of research is towards developing new supervised models for weather nowcasting. The goal was to create new machine learning models that can be used for radar data prediction and validate them as proofs of concept. In order to validate the models, we used different measures, but the most relevant ones are *Root Mean Squared Error (RMSE)* for regression tasks and *Critical Success Index (CSI)* for classification tasks. *RMSE* is often used as a measure in literature for weather prediction and *CSI* is a meteorology specific measure for the predictions of whether there will be a meteorological event at a location. We also computed *RMSE* only for non-zero values, as those are the meteorologicaly relevant values, and zero values are much more abundant than non-zero values, thus skewing the results. During our research we developed the 3 following machine learning models:

a) **NowDeepN.** The first model we created was based on *deep neural networks*. The ideea was to predict the value of one radar product at a location based on the values of all products at the previous time step in a neighbourhood of that location. Since we predict multiple products, we have multiple networks for each product predicted – resulting in a model containing an ensemble of 13 networks (we are not using the "ensemble learning" paradigm). The *NowDeepN* model description and results were published in [CMt21]. They are also described in detail in Section 3.1. On testing data we obtained a *RMSE* of 2.25 $\pm$0.12 with zeros and of 5.93$\pm$0.14 on nonzero values. If we considered the value of 5 dbZ as a threshold for classification, we obtained a *CSI* of 0.64. Comparing to related work, the comparison is favorable for *NowDeepN* in 5 out of 7 cases.

b) **RadRAR.** This model is based on *Relational Association Rules (RARs)* mining. While initially using *Relational Association Rule (RAR)* mining as an unsupervised data mining tool, we later found a way to use the extracted rules for prediction, and ended by creating the *RadRAR* binary classifier model. As one of the drawbacks of *RARs* is that they are less scalable, *RadRAR* was trained and tested on a smaller geographical region than the other 2 models, and only considers one radar product (R01 – Reflectivity at the lowest elevation angle). Our work on this model is published in [CMC19b]. We described the details of this model in Section 3.2 in this thesis. Using a threshold of 35 dbZ, which is a meteorologicaly relevant thershold for product R01, we obtained a *CSI* of 0.56 $\pm$ 0.02, performing better in 8 out of 9 comparisons with related work and other classifiers.

c) **XNow.** The third model we developed is based on *deep convolutional networks*. This time, we started with the idea of predicting the entire region and all products at once, from the data at the previous moment. The model is heavily inspired by the *UNet* [RFB15] and *Xception* [Cho17] architectures, *XNow* actually being a modified version of the latter to work similarly as the former. This model was published in [SCIM20] and is presented in detail in Section 3.3 in this thesis. With the *XNow* model we obtained a *RMSE* of 1.85 $\pm$ 0.15 on data with zeros and 2.28 $\pm$ 0.17 on nonzero values. This is a very good result, being better than *NowDeepN* and marginally better than the best model we found in the literature, with a similar experiment design and purpose.

Thesis Structure

The rest of the thesis is organized as follows. In the first Chapter the theoretical background and the literature review is presented. In Section 1.1 we first present the type of radar data that we use and how it is collected and then we present our literature review. In the second part of the first chapter – Section 1.2 – we detail the theoretical basis necessary for the machine learning algorithms we used in our research.

In Chapter 2 we present our experiments using unsupervised machine learning methods – mainly Self-Organizing Maps – on radar data. Since all these experiments use the same data set, we first present in detail this data set in Section 2.1. The first experiment is described in Section 2.2, where we analyse the change of the radar products' values in time. Our second experiment using Self Organizing Maps is detailed in Section 2.3. Since our second experiment was done only one one of the radar products, at one elevation angle, we considered extending the experiment to multiple products and elevations. The results of this extension are presented in Section 2.4. We also introduced new evaluation measures to better interpret the results and efficiency of the SOM.

The three supervised machine learning models we have developed for weather nowcasting are described in Chapter 3. The first model proposed is *NowDeepN*, presented in Section 3.1, which is based on deep neural networks. The second model we developed, based on Relational Association Rule mining, is described in the Section 3.2 of this thesis. The last model we developed, *XNow*, based on convolutional neural networks, and more exactly on the *Xception* architecture, is presented in Section 3.3.

***

The author acknowledges financial support from the NO Grants 2014-2021, under Project contract no. 26/2020. The author would like to thank Eugen Mihuleţ from the National Meteorological Administration from Romania, for providing the meteorological data sets used in the experiments and for his useful feedback on the obtained results.

Chapter 1

Background

In this chapter we present the theoretical elements that will be used in the thesis. It is split into two parts: weather nowcasting, covering elements related to weather nowcasting and radar data; and machine earning methods, covering elements regarding machine learning methods considered in our research.

1.1 Weather nowcasting

According to a recent joint article from Nordic and Baltic countries [Swe18], climate change including extreme rain phenomena is expected. In consequence, there is an increasing need for accurate and early warning of severe weather events. As the number and intensity of severe meteorological phenomena increases, predicting them in due time to avoid disasters becomes highly demanding for meteorologists.

The division of weather prediction dealing with weather analysis and forecast for the next 0 to 6 hours is called *nowcasting* and plays an increasing role in crisis management and risk prevention. The problem of issuing a nowcasting warning is a difficult task for meteorologists, mainly because of the extremely large set of data which has to be analyzed in a short period of time. Therefore, ML based methods are useful for offering effective solutions for nowcasting by learning relevant patterns from the large amount of weather data and thus improving decision-making for high-impact weather. Most of the existing operational and semi-operational methods for nowcasting are using the extrapolation of radar data and algorithms mainly based on cell tracking.

The current thesis uses radar data provided by the WSR-98D weather radar [NOA18]. About every 6 minutes data is collected on a complete set of about 30 base and derived products, gathered over 7 different elevations. The base products are particle *reflectivity* (R), providing information on particle size and type, and *particle velocity* (V), containing information on particle motion. Both products are available for several elevation angles of the radar antenna, and for each time step a set of seven data products, R01-R07 and V01-V07, is delivered, each of them corresponding to a certain tilt of the antenna. Among the derived products, of particular interest for the study is VIL (vertically integrated liquid), an estimation of the total mass of precipitation above a certain unit of area.

1.2 Machine learning models

Supervised learning is a subfield of machine learning, dealing with the task of approximating a mapping from some input domain to some output domain based on example input-output pairs. Unsupervised learning is a subfield of ML dealing with algorithms which extract useful information from raw data without using labeled examples (as in supervised learning).

A *self-organizing map* (SOM) [SK99] is an unsupervised learning model, a type of ANN from the category of *competitive learning* networks. A SOM contains two layers: the input layer and the output layer. These layers are *densely* connected. Usually, a SOM is trained using the Kohonen algorithm [SK99]. While a SOM is tool for visualizing high dimensional data, it is also very effective clustering problems, data-mining tasks or classification [LO92]. The U-Matrix method [KK96] is usually used for visualizing a trained SOM.

Relational association rules (RARs) [SCC06] are an extension of *Association rules* (ARs), which are powerful data analysis and mining tools. RARs can discover various types of relationships between data attributes. Discovery of Relational Association Rules ($DRAR$) is the Apriori-like algorithm used for mining the interesting RARs from a dataset [CBC12].

Neural network learning methods provide a robust approach to approximating real-valued, discrete-valued or vector-valued target functions [Mit97]. Neural networks are suited for problems that deal with noisy, complex data, such as camera, microphone or sensor data. Their success is due to their similarity to effective biological systems, that are able to generalize and associate data that has not been explicitly trained upon during the training phase, and correlate that data to a class where it belongs.

Unlike classical neural networks, *Deep Neural Networks (DNNs)* contain multiple hidden layers and have a large number of parameters which makes them able to express complicated target functions, i.e. complex mappings between their input and outputs [SHK+14]. Nowadays, DNNs are powerful models in the ML literature applied for complex classification and regression problems from various domains.

Recurrent Neural Network (RNN) [HS13] are capable to express dynamic and temporal processes and to model sequential information. Due to their ability to model sequences, RNNs were successfully used to solve numerous tasks in which the input was organized in time steps, including: speech recognition [Lip15], [GWD14], image and video processing [WKS16], machine translation and sentiment analysis [CvMG+14]. A RNN contains at least one feed-back connection, so the activations can flow round in a loop [HS13]. That enables the networks to do temporal processing and learn sequences, like: performing sequence recognition/reproduction or temporal association/prediction.

A Long Short-Term Memory (Long Short-Term Memory network (LSTM)) [HS97] is a particular model of RNN allowing the unit activations to retain important information over a much longer period of time. In order to store an information for a longer period, a circuit for simulating a memory cell should be implemented.

Convolutional Neural Networks (Convolutional Neural Networks (CNNs)) [KSH12] are ANNs that receive as input multi-channel images. Since they are ANNs at the core, the ANNs concepts are valid for CNNs, too: they receive input, process it through the propagation function, then pass the result to an activation function, and finally produce an output.

The ideea of Convolutional Long Short-Term Memory network (ConvLSTM) is to introduce Convolution inside the LSTM cell. LSTM does not use spatial data, only temporal data, which is a great disadvantage for task that could take advantage of spatial data. Combination of LSTM and CNN were mentioned in 2014 by Donahue et al. in [DHG+14].

Chapter 2

New unsupervised learning models for meteorological data analysis

Predicting weather, and particularly severe weather, is an important challenge both for meteorological and machine learning researchers. The complexity and difficulty of the problem is mainly due to the chaotic character of the atmosphere and the implicit large set of meteorological information (radar, satellite or ground meteorological observations) which have to be analyzed by meteorologists. Thus, understanding the relationships between various meteorological parameters extracted from radar observations may be useful for providing additional comprehension about severe weather development and would help to identify situations when severe weather can occur.

In the following sections *self-organizing maps* are being explored as an unsupervised classification model for detecting patterns in radar data which are relevant in predicting short-term weather changes.

In the studies presented in this chapter real data provided by the National Meteorological Administration (NMA) was used. In Section 2.1 this data set is presented in detail.

All the elements presented in this chapter were also published in three original papers [CMtT19, MCt19, Mih20]. In the following we highlight the main original contributions presented in the chapter:

- The first experiment is presented in Section 2.2. With the main goal of analyzing how the values for the weather radar products are evolving between consecutive radar scans, we empirically show that in general there is a slow change in the values over time, except for the situations when when certain severe phenomena occur. The study conducted in Section 2.2 is an original work published in [CMtT19] and is aimed to provide a better insight regarding how the values of weather radar products are evolving in time both in calm and severe weather conditions, with the broader goal of using these findings for weather nowcasting.

- Section 2.3 introduces an alternative radar data model and aims at obtaining an empirical evidence that: (1) there are some patterns in the way the radar products' values transition from one time moment to another in both normal and severe weather conditions; and (2) that similar values for a product at a given moment are encoded in similar neighborhoods at previous time moments. The approach from this section was published in the original paper [MCt19].

- The previous experiment was, at first, tested on only one radar product. We then extended the scope of the study for other radar products with the aim of supporting the idea that the results are consistent on different radar data products. This extension is presented in Section 2.4 and the results were published in [Mih20].

The presentation from this chapter is based on the original papers [CMtT19, MCt19, Mih20].

2.1 Radar data set

For our experiments we use real radar data provided by NMA, the Romanian meteorological administration.

The data was provided by the WSR-98D weather radar [NOA18] located in Bobohalma, Romania. About every 6 minutes data is collected on a complete set of about 30 base and derived products, gathered over 7 different elevations. The most interesting products are particle *reflectivity* (R), *particle velocity* (V) and *vertically integrated liquid* (VIL). Products R and V are available for several elevation angles of the radar antenna and a set of seven data products, R01-R07 and V01-V07, is delivered, each of them corresponding to a certain tilt of the antenna. The data grid provided by the radar for the selected geographical area at a given time moment is fit to a matrix. The radar provides one data matrix for each of the meteorological products, and each matrix has 624 rows and 800 columns (i.e. $m = 800$ and $n = 624$).

The day used as case study is the 5th of June 2017, a day with moderate atmospheric instability in the region, manifested through thunderstorms accompanied by heavy rain and medium-size hail. Concerning these phenomena, the National Meteorological Administration issued five severe weather warnings, code yellow. In the chosen geographical area, there were two distinct episodes with intense meteorological events in June 5, 2017: the first one between approximately 09:00 and 11:00 UTC, and the second one between approximately 12:00 and 17:00 UTC, with the most severe events taking place between 14:00 and 15:00 UTC.

The radar data used in our case study has been recorded between 00:04:04 UTC and 23:54:02 UTC. We have a total of 231 time stamps (i.e. $k = 231$), with time stamp 1 corresponding to 00:04:04 UTC. The most interesting time stamps are the ones in which there is data about the above mentioned meteorological events: the time stamps from 88 to 106 contain the data for the meteorological event from 09:00 to 11:00 and the time stamps from 117 to 165 contain the data for the meteorological event from 12:00 to 17:00. The data for the maximum values approximated to be between 14:00 to 15:00 are contained in the time stamps from 137 to 145.

The data gathered by the radar contains a special value that represents "No Data". This value is usually represented by -999 but we decided to replace it with 0 as in most cases this value refers to air particles with 0 reflectivity (i.e. no significant water droplets). 'No data' may also represent air volumes which have returned no signal, for example if a sector with high reflectivity is between the radar and the respective location. In this case, replacing it with 0 is also correct, since the entire region is obturated and the data is not relevant for the learning process. The radar data is prone to different type of errors, meteorological and technical, which implicitly are to be found in the output data matrix. Meteorological errors (e.g., the underestimation of a particle's reflectivity) are difficult to identify and eliminate, but some errors occurring during the data conversion have been identified and corrected. For example, the product V should only contain values from -33 to 33 but we found values of -100. In order to avoid introducing into our experiments the noise that these values represent, we decided to skip them in the unsupervised learning process. More exactly, during the training using the Kohonen algorithm, the erroneous values of -100 were omitted while computing the Euclidian distance between the input instances and the neurons from the map.

2.2 Analysis of radar data change over multiple timestamps

Self organizing maps are being explored in this section as an unsupervised classification model for detecting patterns in radar data which are relevant in predicting short-term weather changes. The approach introduced in this section is an original work published in [CMtT19]. With the main goal of analyzing how the values for the weather radar products are evolving between consecutive radar scans, we empirically show that generally these values are slowly modifying in time, excepting situations when certain severe phenomena occur. The study conducted in this section is aimed to provide a better insight regarding how the values of weather radar products are evolving in time both in calm and severe weather conditions, with the broader goal of using these findings for weather nowcasting.

We are assessing the usefulness of SOMs to unsupervisedly uncover the underlying structure in radar data, for analyzing how the values for several weather radar products are evolving between consecutive radar scans and for studying the relevance of the radar products in predicting short-term weather changes. Through several experiments performed on the real radar data provided by the NMA, we aim to obtain an empirical evidence that the radar meteorological products' values are generally smoothly changing in time in normal weather conditions, excepting situations when certain severe phenomena occur. In addition, we expect that SOMs are able to distinguish severe weather conditions using radar data.

We propose in the following a data model which will be further used in our experiments. The idea is to assign, at each time stamp, a vectorial representation to each 3D data grid provided by the radar. In this model, for a day d, a time stamp t_i^d $(1 \leq i \leq k)$ and a set $Prod$ of meteorological products, a data parallelepiped $P_{t_i{}^d}(m, n, Prod) = (p_{xyz})_{\substack{x=\overline{1,m} \\ y=\overline{1,n} \\ z=\overline{1,|Prod|}}}$ is constructed. In this parallelepiped, OX and OY axes represent the rows and columns from the radar data grid, and the depth axis OZ represents the meteorological products. For obtaining the vectorial representation for the data parallelepiped $P_{t_i{}^d}(m, n, Prod)$, it is liniarized.

Two data sets $D1$ and $D2$ are constructed for representing the radar data collected during the time stamps $t_1, t_2, \ldots, t_k$ using the data model previously introduced. The difference between $D1$ and $D2$ is given by the set of meteorological products used for representing the instances. In $D1$ the entire set of meteorological products provided by the radar (i.e. 24) is used, while $D2$ employs only 13 products: base *reflectivity* (R) of particles on six elevations, *velocity* (V) on six elevation and the estimated quantity of water (VIL) contained by a one square meter column of air.

For detecting the underlying structure of the data sets $D1$ and $D2$, the SOM model is applied for obtaining an unsupervised two-dimensional representation of the data sets.

As a preliminary step before applying the unsupervised SOM models, a statistical analysis was performed on the data set with the goal of analyzing the variation of the meteorological products on each time stamp.

For the SOM employed in the experiments we used our own implementation, without any third party libraries. For building the SOM, we used a torus topology [KTO+07].

The following parameters were used for the SOM: a configuration of 30x30 neurons on the map, 20000 training epochs and a learning rate of 0.1.

In our implementation, the lower values on the U-matrix are depicted as darker places, while higher places are marked as whiter regions. Accordingly, darker regions encode similar data instances, whilst whiter regions represent separation boundaries between the data clusters.

The results suggests that there might be some readable changes in the meteorological products almost 2 hours before the start of the event which might help to forecast the start of the phenomena.

Also, we concluded that R, V and VIL values may be useful for predicting meteorological events and that additional meteorological products (other than R, V and VIL) do not bring significant additional information about the phenomenon.

For assessing the relevance of V in usupervisedly uncovering meteorological events, we performed the first experiment using only R and VIL products, without considering V. Analysing the results we concluded that V is also relevant in detecting severe meteorological events and R and VIL measurements have to be used together with V for increasing the performance of the detection process.

The results show evidence that the values of the radar products clearly discriminate between calm weather and severe events. The SOM is also able to unsupervisedly detect these patterns using only the R, V and VIL products. This suggests the feasability of learning to predict (using R, V and VIL products) an entire data parallelepiped at a certain time based on data parallelepipeds at previous time moments.

This section presented a study towards applying SOMs as an unsupervised classification method for analyzing meteorological radar data and investigating the relevance of several meteorological products in detecting severe weather phenomena.Several experiments were performed, analyzing the results we obtained an empirical evidence that in normal weather condition the values of the meteorological products are smoothly changing in time, excepting situations when certain severe phenomena occur. Thus, meteorological events reflected in changes occurred in the values of several meteorological products are indeed detected by unsupervised learning algorithms.

2.3 Analysis of patterns in radar data transition between consecutive radar scans

The approach introduced in this section is an original work published in [MCt19].

The main goal of the approach introduced in this section is to better understand the relationships between the meteorological products extracted from one radar observations and some radar data observation from previous times, both in severe and normal weather conditions.

In this section we are investigating the ability of SOMs models to unsupervisedly learn meteorological relevant patterns, particularly in situations when severe meteorological events occurred. We are particularly focused on the patterns of meteorological data arising from one time moment to another. As a proof of concept, SOMs were used in the current study as an unsupervised learning tool for analyzing radar data recorded at the national level and used for weather nowcasting. Through experiments an empirical evidence that (1) there are some patterns in the way the radar products' values transition from one time moment to another in both normal and severe weather conditions and that (2) similar values for a product at a given moment are encoded in similar neighborhoods at previous time moments.

For computational modelling of the radar data, we propose a *cell-level* data model. In this model, we aim to assign to a certain cell (x,y) from the grid, at each time stamp $time$ a vectorial representation. This representation contains the data products' values from a neighboring area (subgrid) of a certain length surrounding the point (x,y), for a temporal window of length l before $time$: $time$-l, $time$-l+1,$\ldots$, $time$-1.

The experiments are aimed to analyze the extent to which SOMs are able to unsupervisedly uncover meteorological phenomena in radar data. The goal of the experiments is to test if similar values for a radar product at a given moment are encoded in similar neighborhoods at previous time moments. For a certain time stamp t, two data sets D and D' are constructed. The difference between D and D' is given by the length l considered for the temporal window: in D we are using a length l of

1, while D' considers a value for l greater than 1.

As a preliminary step before applying the unsupervised SOM model, a statistical analysis was performed on the data set, with the goal of analyzing the variation of the meteorological products on each time stamp.

For the SOM [SK99] employed in the experiments we used our own implementation, without any third party libraries. For building the SOM, we used a torus topology [KTO+07]. In our implementation, the lower values on the U-matrix [KK96] are depicted as darker places, while higher places are marked as whiter regions. Accordingly, darker regions encode similar data instances, whilst whiter regions represent separation boundaries between the clusters. Through this experiment, we would expect the SOM to unsupervisedly detect a relationship between the value of a certain product for a grid cell c at a certain time stamp t and its vectorial representation using the proposed *cell-level* data model at time stamps preceding t.

The analysis of the results led to the conclusion that higher values for $R02$ can be predicted from previous time stamps, irrespective of the temporal window length or if there are meteorological events present or not. Furthermore a good enough estimate of the actual value of $R02$ is predictable. This result implies the fact that patterns can be learned from the data and supervised learning for the prediction of $R02$ is feasible for the purpose of nowcasting.

As a conclusion of our study, SOMs are able to unsupervisedly uncover in radar data hidden patterns which are relevant from a meteorological perspective. The findings of our study suggest promising results in applying predictive supervised learning models for weather nowcasting using radar data.

2.4 Extension of analysis for multiple radar products

In the previous section (Section 2.3) we focused our experiments on only one radar product, R02.

In this work, published in [Mih20], we are extending and further analyzing the ability of SOMs to encode and extract from radar data patterns regarding how the radar products are transitioning from one time moment to another, analysing four other radar products, empirically showing that our previous results can be generalised for the most used radar products in nowcasting.

Before building the SOM model, a *data cleaning* step is first applied to the radar data. The goal of this preprocessing step is to correct the erroneous values provided by the radar. Erroneous values represent values that are outside the bounds for one product (e.g. a value of 75 for R01, which normally should be between 0 and 65). We correct this values using an estimation algorithm that estimates the correct value from the values of the points in the 13x13 neighbourhood around the erroneous values.

The main ML model used is the SOM with a 2D map. Each instance of the input data is a *cell-level* vector, the same as the one used in the previous section. In this work we extend that experiment to investigate if the hypothesis that similar values at a time moment are encoded in similar neighborhoods at previous moments still hold for other radar data products ($R01$, $V01$, $V02$ and VIL). Two different number of previous timestamps are used in our experiments: only *one* previous timestamp and *five* previous timestamp. Therefore we have eight experimental results – for each of the four radar products ($R01$, $V01$, $V02$ and VIL) we have two results, one with one previous timestamp and one with five previous timestamps.

For evaluating the quality of the SOM mapping we are introducing an evaluation measure ASE (*Average Similarity Error*) which measures how similar are the values of a certain radar product which are mapped on similar regions of the SOM. We introduce this error in order to measure how different is the mapping on one neuron from the ideal mapping.

The ASE measure will have values between 0 and 1, where 0 means that all interesting neurons have labels of equal value (which is ideal), and 1 means that all interesting neurons have labels from both of the extremes in equal amount. Therefore, smaller values for ASE indicate a better mapping, from a meteorological viewpoint.

For obtaining a better insight on the structure of the radar data, we have decided to strengthen the constraints imposed for an *interesting neuron*. We decided to do this because we observed that there were many neurons which had many 0 labels and one or very few non-zero labels, but very close to 0. Therefore we have a secondary *Average Similarity Error* (ASE') which only considers neurons that contain only non-zero values.

We employed in our experiments our own implementation of SOM, which was build using a 2D lattice having a torus topology [KTO+07]. The U-matrix method is used to visualize the resulting map, where lower values in the U-matrix are depicted as darker regions, while whiter regions depict higher values.

By analyzing the resulting U-matrices of the 8 experiments conducted we observed that they are consistent with the results obtained in our previous work [MCt19]. This means that similar combinations of values of the radar data products at previous timestamps are correlated to similar values of product values at the current timestamp, for all of the studied products, leading to the conclusion that the studied radar products can be predicted for future timestamps from the values of the radar products at previous time stamps.

For each of these resulted SOM maps, we have computed the measures introduced previously. ASE is very low for all the resulted maps. All values are below 0.05, with the exception of R01, which are below 0.1. An error lower than 0.05 means that the labels mapped for one single neuron have, on average, differences no bigger than 5% of the maximum difference possible for that label. This means that the labels mapped to one neuron are very similar to one another, which is desirable.

The ASE' measure is very similar to the ASE measure, with the only difference being the neurons on which the errors are measured. ASE' is measured using much fewer neurons, a third to a sixth of the number of the neurons used by ASE. Nevertheless, ASE' is not much bigger than ASE. Overall, the values of the measures ASE and ASE' are quite promising, supporting the interpretation of the maps that similar labels are mapped to similar regions.

Using only 1 previous timestamp or multiple previous timestamps for training does not seem to impact the result in a significant way. Using 5 previous timestamps the numbers of neurons used (both N and N') was lower for all experiments, but the measures were not impacted, as they are very similar.

Chapter 3

Contributions in developing deep learning models for weather nowcasting

The second purpose of our research was to create novel models for weather nowcasting *prediction*. More exactly, we created *supervised* machine learning models that predict radar echo for one time moment based on the previous time moments, and validated the models. This chapter present these *supervised* machine learning models we have developed and the experiments we have performed with these new models.

All the elements presented in this chapter were also published in three original papers: [CMt21, CMC19b, SCIM20]. Our original contributions presented in this chapter are the following:

- In Section 3.1 we present our first model, *NowDeepN*, published in [CMt21]. This model was based on *deep neural networks*. The ideea was to predict the value of one radar product at a location based on the values of all products at the previous time step in a neighbourhood of that location. Since we predict multiple products, we have multiple networks for each product predicted On testing data we obtained a *RMSE* of 2.25 ±0.12 with zeros and of 5.93±0.14 on nonzero values. If we considered the value of 5 dbZ as a threshold for classification, we obtained a *CSI* of 0.64. Comparing to related work, the comparison is favorable for *NowDeepN* in 5 out of 7 cases.

- Section 3.2 describes our nex model, *RadRAR*, based on *Relational Association Rules (RARs)* mining, model and experiment published in our paper [CMC19b]. As one of the drawbacks of *RARs* is that they are less scalable, *RadRAR* was trained and tested on a smaller geographical region than the other 2 models, and only considers one radar product (R01 – Reflectivity at the lowest elevation angle). Using a threshold of 35 dbZ, we obtained a *CSI* of 0.56 ± 0.02, performing better in 8 out of 9 comparisons with related work and other classifiers.

- The last model we developed is *XNow* is presented in Section 3.3. This model is based on *deep convolutional networks* and was published in [SCIM20].We started with the idea of predicting the entire region and all products at once, from the data at the previous moment. The model is heavily inspired by the *UNet* [RFB15] and *Xception* [Cho17] architectures. With the *XNow* model we obtained a *RMSE* of 1.85 ± 0.15 on data with zeros and 2.28 ± 0.17 on nonzero values. This is a very good result, being better than *NowDeepN* and marginally better than the best model we found in the literature, with a similar experiment design and purpose.

The presentation from this chapter is based on the original papers [CMt21, CMC19b, SCIM20].

3.1 NowDeepN: An approach for nowcasting prediction using deep neural networks

With the goal of helping meteorologists in analysing radar data for issuing nowcasting warnings, we introduced in our original paper [CMt21] a supervised learning model $NowDeepN$ based on an ensemble of *deep neural network* regressors for predicting the values for radar meteorological products which may be used for weather nowcasting. The presentation of the model in this thesis is based on our published work [CMt21].

As a proof of concept, $NowDeepN$ is proposed for learning to approximate a function between past values of the radar products extracted from radar observations and their future values. Experiments were performed on real radar data provided by the Romanian National Meteorological Administration and collected on the Central Transylvania region.

For *NowDeepN* we use the same data model introduced in Section 2.3.

The radar data is prone to different type of errors, meteorological and technical, which implicitly are to be found in the output data matrix. For reducing the noise that the invalid values represent, a *data cleaning* step is proposed. The underlying idea behind the cleaning step is to replace the invalid values on a certain point (i, j) with the weighted average of values from a neighborhood of length 13 surrounding the point. The weight associated to a certain neighbor of the point is inverse proportional to the Euclidian distance between the neighbor and the point, such that the closest neighbors' values have more importance in estimating the value of point.

The regression problem we are focusing on is the following: to predict a sequence of values for a set $Prod$ of radar products at a given time moment t on a certain location (i,j) on the map, considering the values for the neighboring locations of (i,j) at time moment t-1. $NowDeepN$ uses an ensemble of DNNs for learning to predict the values of the radar products from the set $Prod$ based on their historical values. The ensemble consists of np DNNs ($np = |Prod|$), one DNN for each radar product.

One of the difficulties regarding the regression problem previously formulated is that the training data sets are highly *imbalanced*. More specifically, there are a lot of training instances labeled with zero (i.e. $y_k = 0$) corresponding to points on the map without specific weather events and a much smaller number of instances with a non-zero label (i.e. corresponding to a severe meteorological phenomenon). The imbalanced nature of the data may lead to a regressor which is biased to predict zero values, as the majority of the training examples used for building the regressor were zero-labeled.

For assessing the performance of $NowDeepN$, a *cross-validation* testing methodology is applied on each of the data sets. The data sets are randomly splitted in 5 folds. Subsequently, 4 folds will be used for training and the remaining fold for testing and this is repeated for each fold (5 times).

For each training-testing split, two evaluation measures are used and computed for each training-testing split: *Root mean squared error* (RMSE) and *Normalized root mean squared error* (NRMSE) [HK06]. The RMSE computes the square root of the average of squared errors obtained for the testing instances. The NRMSE represents the normalized RMSE, obtained by dividing the RMSE value to the range of the output and is usually expressed as a percentage. For a more precise evaluation of the results, the values for the evaluation measures (RMSE and NRMSE) are also computed for the non zero-labeled instances (RMSE$_{non-zero}$, NRMSE$_{non-zero}$).

The data set used in the *NowDeepN* experiments is the same as the one presented in 2.1.

In order to estimate the impact of the data cleaning step, we analyzed the data set before and after cleaning. The observations we made from this analysis lead us to the hypothesis that the cleaning step would impact the overall performance of $NowdDeepN$, and this should be visible at least at lower degrees of elevations for V.

For the DNNs used in our experiments, the implementation from the Keras deep learning API [Ker18] using the Tensorflow neural networks framework was employed. The code is publicly available at [CMt21]. Given the fact that our data was quite high-dimensional we needed a relatively complex neural network. These networks were trained for 30 epochs using 1024 instances in a training batch.

We intend to analyze how correlated are our computational findings with the meteorological evidence. In order to allow an easier interpretation of the results from a meteorological perspective, we computed the *Mean of Absolute Errors* for all instances (MAE), as well as only for the non-zero labeled instances (MAE$_{non-zero}$). We obtained an average NRMSE of less than 4% for the R products, which would entail a close resemblance between the predicted data and the real data, resemblance. From a meteorological point of view, the MAE for both all and non-zero instances is a satisfactory one, meaning that the predicted value is on the same level or on a neighbouring level on the product value scale.

In order to empirically validate the hypothesis that the cleaning step improves the predictive performance of $NowDeepN$, we have evaluated the model trained on the uncleaned data set, using the same methodology.

Comparing the results, we observed an improvement in the predictive performance of $NowDeepN$ achieved on the cleaned data. For determining the significance of the features, we are comparing the results of $NowDeepN$ using the original set of features with those obtained by applying $NowDeepN$ after the prior application of a feature extraction step. Two feature extractors were applied on the original set of features, for reducing the dimensionality of the input data: a sparse denoising AE and the PCA algorithm. Comparing the results with those obtained without applying a feature extraction step we observed an improvement in the predictive performance of $NowDeepN$ achieved without a prior feature extraction step. The relevance of the features is validated by the fact that a dimensionality reduction technique (AE/PCA) applied prior to the classification using $NowDeepN$ does not improve the learning performance.

We started the comparison between $NowDeepN$ and related work by comparing our model to a simple baseline model, the *linear regression* (LR). For an exact comparison, the data model used for $NowDeepN$ was used for the LR model as well. By applying the LR on the dataset an overall RMSE for the non-zero values (RMSE$_{non-zero}$) of 6.094 was obtained.

We found four approaches having similar goal to our paper, that of predicting the future values of the radar products' values based on their historical values. The approaches from the literature which are the most similar to ours are those proposed by Yan Ji [Ji17], Han et al. [HSZ$^+$17, HSZ19] and Yan et al. [YJM$^+$20].

The results reveals that overall, in **71%** of the cases (5 out of 7 comparisons), the comparison is favorable to $NowDeepN$. Our proposal is outperformed only by the work of Yan Ji [Ji17] which reported a better HR and a maximum RMSE slightly better than ours.

Tran and Song [TS19] tackled the precipitation nowcasting problem from a computer vision perspective, by applying certain thresholds on the reflectivity values (5/20/40 dBZ). The comparative results highlight that $NowDeepN$ obtained better results than the model proposed by Tran and Song [TS19] in **77.7%** of the cases (7 out of 9 comparisons). We note the good performance of $NowDeepN$ at higher values for the reflectivity threshold, which indicate the ability of our model to detect moderate and heavy precipitation and medium and large hail.

We introduced in this section a supervised learning based regression model $NowDeepN$ which used an ensemble of *deep artificial neural network* for predicting the values for meteorological products at a certain time moment based on their historical values. $NowDeepN$ was intended to be a proof of concept for the feasibility of learning to approximate a function between past values of the radar products extracted from radar observations and their future values.

3.2 $RadRAR$: A relational association rule mining approach for nowcasting based on predicting radar products' values

Relational Association Rules (RARs) [SCC06] extend the classical *association rules* by capturing relationships between values of attributes characterizing a data set. In our original paper [CMC19b] we are investigating, as a proof of concept, the suitability of applying RAR mining for distinguishing between severe and normal weather conditions, with the aim of using these predictions for nowcasting. In addition, we aim to point out the relevance of the RARs mined from radar data, from a meteorological viewpoint. Thus, we are proposing a new one-class classifier, named $RadRAR$ (*Radar products' values prediction using Relational Association Rules*) for convective storms nowcasting based on radar data.

The radar data used in our experiments is provided by the WSR-98D weather radar [NOA18]. In the current study we are focusing on a single meteorological product, namely $R01$. We decided to select R01, as it is one of the most relevant radar products used by operational meteorologists for issuing nowcasting warnings.

Accordingly, we assign to each location l from the analysed map (data grid) at timestamp t a high-dimensional vector whose elements are the values of R01 for the locations situated in a neighborhood of l at timestamp t-1. We note that the label of the d^2-dimensional instance previously described is the value of R01 for the geographical point l at timestamp t.

We note that a value of 13 has been selected for the diameter of the neighbourhood, since it represents about 5 kilometers in the physical world and this distance commonly determines small gradients of the meteorological parameters. The reflectivity values above a certain *threshold* (35 dBZ is generally used [HSZ19, DW93]) are indications about potential moderate to heavy storms occurrences. Thus, we are diving the data set D in two classes of instances: the *positive* class (also denoted as "+") represents the instances labeled with R01 values higher than 35, while the *negative* class (denoted as "-") represents the instances labeled with R01 values less or equal to 35. Thus we have 2 data sets, D_+ and D_- for positive and, respectively, negative data.

We propose $RadRAR$, a one-class classifier which is trained on D_- and will learn to predict, based on the neighborhood of a certain location at time t, whether the radar echo value at time t+1 will be higher than 35dBZ. The prediction is based on estimating the probability p_- that a certain 169-dimensional instance belongs to the "-" class.

The classification process we propose takes place in two phases: *training* and *testing*. During the training, a classification model consisting of a set of interesting RARs from the set D_- will be built, and during testing, the model built during the training will be applied for deciding the class ("+" or "-") for a testing instance unseen during

For evaluating the performance of the $RadRAR$ model, it is tested on data sets containing both *positive* and *negative* instances which are completely disjoint from the training data set. For a testing data set, the *confusion matrix* consisting of four values is computed: True Positives – TP, True Negatives – TN, False Positives – FP and False Negatives – FN. As evaluation measures, we are using four measures computed based on the values from the confusion matrix, used in supervised learning for assessing the performance of binary classifiers: *sensitivity* or *probability of detection* (**POD** = $\frac{TP}{TP+FN}$), *specificity* or true negative rate (**Spec** = $\frac{TN}{TN+FP}$), *false alarm rate* (**FAR** = $\frac{FP}{TP+FP}$) and *Area Under the ROC Curve* (**AUC** = $\frac{\textbf{POD}+\textbf{Spec}}{2}$). Additionally, we also consider the *Critical success index* (**CSI**) measure which is usually used for convective storms nowcasting based on radar data –

CSI=$\frac{TP}{TP+FN+FP}$. All the previously mentioned evaluation measures range in [0, 1]. Excepting FAR which has to be minimized, higher values for all other evaluation measures indicate better classifiers.

The case study used in our experiments is the radar data provided by the radar for the 5th of June 2017, a day with moderate atmospheric instability manifested through thunderstorms accompanied by heavy rain and medium-size hail. In the area from the central Transylvania region there were two distinct episodes with intense meteorological events in June 5, 2017. We restrict ourselves to a small size experiment, as our aim is to establish a proof of concept for the relevance of using RARs for nowcasting based on radar data. The data used for training $RadRAR$ is collected at approximately 14:37 UTC (in the middle of the severe event). The data sets D_+ and D_- collected from the raw radar data consist of 1321 and 19991 instances, respectively.

In our experiments, two possible relations between the features' values are considered in the mining process: $\mathcal{R} = \{\leq, \geq\}$. After the relations were defined, the set RAR_- of interesting relational association rules were discovered from D_-.

The approaches from the literature which are the most similar to ours are those proposed by Yan Ji [Ji17] and Han et al. [HSZ+17, HSZ19]. For better highlighting the effectiveness of $RadRAR$ as an anomaly detector, we replaced it with an *autoencoder* (AE). The AE had been built using the Keras framework in Python, with Tensorflow backend.

Analysing the results we observed that our $RadRAR$ proposal provides better results for the evaluation measures in 8 out 9 comparisons. From the results we may conclude that the RARs uncovered within the radar data are effective for predicting if the radar echo values are higher than 35dBZ, obtaining performances which are generally better than the results from the literature [HSZ19, HSZ+17].

As a proof of concept, we introduced in this section a novel one-class classification model $RadRAR$ based on uncovering interesting *relational association rules* for estimating if the radar echo values will be higher than 35dBZ. Thus, based on the predicted values, the approach is useful for discriminating between normal and stormy weather conditions. Real radar data provided by the Romanian National Meteorological Administration have been used for assessing the performance of $RadRAR$.

3.3 $XNow$: A convolutional deep learning technique for nowcasting based on radar products' values prediction

We introduced in our original paper [SCIM20], a *convolutional neural network* model $XNow$ for short-term prediction of radar data by adapting the Xception architecture [Cho17] mainly used in the literature for image processing. Experiments performed on real radar data highlight that the proposed deep learning model is able to accurately predict the value for the radar data at a certain time moment in a certain geographical region, based on their historical values.

The exported radar data is stored as two-dimensional matrix (grid) in which each point correspond to a geographical location and contains the value of a radar product at a given time moment. Thus, a sequence of matrices is available, each matrix corresponding to a certain time stamp t and a certain meteorological product p (e.g. R01).

The next step applied before building the $XNow$ deep learning model is to apply a preprocessing step on the sequences S_t for correcting some erroneous values recorded by the radar. In order to avoid these errors we decided to replace them with an estimation. The estimation is a weighted average of the values in a neighbourhood (a 13 by 13 matrix with the point to estimate at the center) where the weight is the Euclidean distance between the neighbour and the point.

The target function in our learning problem is the mapping f such that for a certain data grid G_t, $XNow$ will have to provide an estimation of the 3D data grid G_{t+1} containing the radar products' values at time $t + 1$. For achieving the learning task, $XNow$ model will be obtained by training an adapted Xception architecture. A training sample is in the form (G_t, G_{t+1}).

The original Xception architecture abstracts within each layer its input, such that in the end we get a compact representation of it from which a single value, representing the prediction, is obtained. Nevertheless, our goal is to reconstruct the original input, similar to the behavior of an encoder-decoder architecture, while still preserving the effectiveness of Xception as a convolutional neural network. In this regard, we consider a slightly modified version of the classical version, by substituting its final layers.

We will use 70% of the data set for training $XNow$, 20% for the model validation and the remaining of 10% will be further used for testing. For evaluating the performance of $XNow$, the RMSE value is computed over the samples of the previously gathered test data. The radar data used in our experiments contains a lot of zero-valued data points, consequently, the RMSE value for the non-zero values, denoted by **RMSE_nonzero**, will be provided, as well.

The experiments were performed using data provided by the Romanian NMA and represents the data collected by the radar in 10 days. The days were selected such that in some days there were significant meteorological events while in some other days there were almost no meteorological activity and some in between, so as to resemble as much as possible to typical summer weather. The data comes from a radar situated in central Transylvania and provides data for a large area.

The results revealed that RMSE values are slightly higher for the non-zero values than for all the values. However, the value of **2.282** obtained for the average **RMSE_nonzero**, when normalized we obtain a *normalized root mean squared error* of about 3%, highlighting a very good performance of $XNow$.

For better highlighting the effectiveness of $XNow$ (i.e. the enhanced Xception model), experiments were performed using the classical Xception architecture, as well. The average RMSE values computed over the multiple runs of $XNow$ were computed. Analysing the results we observed that $XNow$ provides better RMSE values than the classical Xception architecture on the cleaned data - we observe an almost 2 times improvement on the average **RMSE_nonzero**. Moreover, there is a small *standard deviation* of the RMSE values over the multiple runs and this conducts to a small CI, highlighting the stability of the $XNow$ model.

The approach from the literature which is the most similar to ours is that proposed by Yan Ji [Ji17]. The exact RMSE values obtained in estimating the values for R are not provided, but only the *hit rate* defined as the percentage of cases in which the absolute error is less or equal to 5. The minimum, maximum and average *hit rate* values are reported by Yan Ji. Starting from the provided *hit rates*, we deduced the inferior limits of the range of the RMSE values. This limit is quite low (being obtained when all values were exactly predicted) and thus it is hard to deduce an accurate approximation of the RMSE values. Our best $XNow$ model obtained a better performance than the ANN proposed by Yan Ji [Ji17].

We have introduced in this section a *convolutional neural network* model $XNow$ for predicting, in a supervised learning manner, the future values of the radar products, with the aim of assisting meteorologists in decision making processes (e.g. providing nowcasting warnings). Experiments were performed on real radar data provided by the Romanian National Meteorological Administration. For highlighting the effectiveness of $XNow$, it was compared to the classical Xception architecture and the obtained results were also compared to the current performance of existing solutions. An average *normalized root mean squared error* less than 3% was obtained, highlighting a very good performance of the $XNow$ regressor.

Conclusions

The aim of our PhD research, as per the title of this thesis, was to develop new machine learning models, both supervised and unsupervised, to be used in weather nowcasting contexts.

For the unsupervised part of our research we have chosen the Self Organizing Map (SOM) model to study. We developed two data models for the radar data and techniques to apply the SOM on the data. The first model is based on the goal of uncovering how the radar products evolve during consecutive radar scans. By interpreting the resulting U-matrices we have shown that radar product values change slowly over time, except some specific moments related to severe weather phenomena. Our second data model was based on the goal of studying the relationship between the value of one radar product at a location at a time moment and the values of the radar products in a neighbourhood of that location in previous time moments. Interpreting the SOM results with this data model we have shown that for similar values of one product, the neighbourhoods at previous time moments are similar. We have also created an evaluation measure – the Average Similarity Error – that shows that the results of our SOM experiments are significant.

Our research on the supervised learning part of the project culminated with the development of three new machine learning models for weather nowcasting: *NowDeepN*, *RadRAR* and *XNow*. We developed *NowDeepN*, an ensemble of 13 neural networks, each predicting a different radar product at a specific location based on all the radar products in a neighbourhood of that location. We have shown that *NowdeepN* performs quite well, compared to other models in the literature, the comparison being favorable to our model in 5 out of 7 cases. When developing *RadRAR* we had the goal to classify if the value of a radar product will be above or below a threshold. *RadRAR* first learns rules from data, separated between the 2 classes then, based on the mined rules, it can predict whether the value of R01 at a location is above or below the 35 dBZ threshold, based on the values of R01 in a neighbourhood of that location at the previous time step. We have shown that *RadRAR* is quite performant at this task, comparing favorably with other models from the literature. The last supervised machine learning model we developed was *XNow*. *XNow* is capable to predict all the data for one time step based on the data at the previous time step. We have empirically shown that the model has very good results, slightly outperforming the other models in the literature.

In the future we aim at continuing the development of these models. For *RadRAR* we envision making improvements to the rules mining algorithm and optimizing the data from which to extract the rules and number of rules. For continuing our efforts for creating better machine learning models for weather nowcasting, our main focus will be on the *XNow* model, as it had the best results out of the three models and also it is the most scalable. So, in future research projects, we plan to extend the XNow model to be able to predict for more than one time step in the future. We also envision using multiple previous time steps and increasing the training data from days to weeks or months .

References

[CAG19] Liana-Maria Crivei, Mihai Andrei, and Czibula Gabriela. A study on applying relational association rule mining based classification for predicting the academic performance of students. In *KSEM 2019 : The 12th International Conference on Knowledge Science, Engineering and Management, LNAI 11775*, pages 287–300, 2019.

[CBC12] Gabriela Czibula, Maria-Iuliana Bocicor, and Istvan Gergely Czibula. Promoter sequences prediction using relational association rule mining. *Evolutionary Bioinformatics*, 8:181–196, 04 2012.

[Cho17] F. Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807, 2017.

[CMA+21] Gabriela Czibula, Andrei Mihai, Alexandra-Ioana ALbu, Istvan Czibula, Sorin Burcea, and Abdelkader Mezghani. $autonowp$: An approach using deep autoencoders for precipitation nowcasting based on radar echo prediction. *Mathematics*, in press, 2021.

[CMC19a] G. Czibula, A. Mihai, and L.M. Crivei. A novel relational association rule mining classification model applied for academic performance prediction. page accepted for publication. Elsevier, 2019.

[CMC19b] G. Czibula, A. Mihai, and I.G. Czibula. Radrar: A relational association rule mining approach for nowcasting based on predicting radar products' values. pages 300–309. Elsevier, 2019.

[CMt21] Gabriela Czibula, Andrei Mihai, and Eugen Mihuleţ. $nowdeepn$: An ensamble of deep learning models for weather nowcasting based on radar products' values prediction. *Applied Sciences*, 11:121, 2021.

[CMtT19] Gabriela Czibula, Andrei Mihai, Eugen Mihuleţ, and Daniel Teodorovici. Using self-organizing maps for unsupervised analysis of radar data for nowcasting purposes. *Procedia Computer Science*, 159:48–57, 2019. 23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2019).

[CvMG+14] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics.

[DHG+14] Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. *CoRR*, abs/1411.4389, 2014.

[DW93] M. Dixon and G. Wiener. TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting - A radar-based methodology. *Journal of Atmospheric and Oceanic Technology*, 10(6):785–797, 1993.

[GWD14] Alex Graves, Greg Wayne, and Ivo Danihelka. Neural turing machines. *CoRR*, abs/1410.5401, 2014.

[HK06] Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679 – 688, 2006.

[HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997.

[HS13] Michiel Hermans and Benjamin Schrauwen. Training and analysing deep recurrent neural networks. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 190–198. Curran Associates, Inc., 2013.

[HSZ+17] Lei Han, Juanzhen Sun, Wei Zhang, Yuanyuan Xiu, Hailei Feng, and Yinjing Lin. A machine learning nowcasting method based on real-time reanalysis data. *Journal of Geophysical Research: Atmospheres*, 122(7):4038–4051, 2017.

[HSZ19] Lei Han, Juanzhen Sun, and Wei Zhang. Convolutional Neural Network for Convective Storm Nowcasting Using 3D Doppler Weather Radar Data. *arXiv e-prints*, page arXiv:1911.06185, Nov 2019.

[Ji17] Yan Ji. Short-term precipitation prediction based on a neural network. In *3rd International Conference on Artificial Intelligence and Industrial Engineering*, AIIE 2017, pages 246–251, 2017.

[Ker18] Keras. The Python Deep Learning library, 2018. https://keras.io/.

[KK96] S. Kaski and T. Kohonen. Exploratory data analysis by the self-organizing map: Structures of welfare and poverty in the world. In *Neural Networks in Financial Engineering. Proceedings of the Third International Conference on Neural Networks in the Capital Markets*, pages 498–507. World Scientific, 1996.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 1097–1105. Curran Associates Inc., 2012.

[KTO+07] Peter K. Kihato, Heizo Tokutaka, Masaaki Ohkita, Kikuo Fujimura, Kazuhiko Kotani, Yoichi Kurozawa, and Yoshio Maniwa. Spherical and torus som approaches to metabolic syndrome evaluation. In Masumi Ishikawa, Kenji Doya, Hiroyuki Miyamoto, and Takeshi Yamakawa, editors, *ICONIP (2)*, volume 4985 of *Lecture Notes in Computer Science*, pages 274–284. Springer, 2007.

[Lip15] Zachary Chase Lipton. A critical review of recurrent neural networks for sequence learning. *CoRR*, abs/1506.00019, 2015.

[LO92] J. Lampinen and E. Oja. Clustering properties of hierarchical self-organizing maps. *Journal of Mathematical Imaging and Vision*, 2(3):261–272, 1992.

[MCt19] Andrei Mihai, Gabriela Czibula, and Eugen Mihuleţ. Analyzing meteorological data using unsupervised learning techniques. In *ICCP 2019: IEEE 15th International Conference on Intelligent Computer Communication and Processing*, pages 1–8. IEEE Computer Society, 2019.

[Mih20] Andrei Mihai. Using self-organizing maps as unsupervised learning models for meteorological data mining. In *IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI 2020)*, pages 23–28. IEEE Hungary Section, 2020.

[Mit97] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.

[NOA18] NOAA's National Weather Service. Radar Operations Center. NEXRAD Technical Information, 2018. https://www.roc.noaa.gov/WSR88D/Engineering/NEXRADTechInfo.aspx.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[SCC06] Gabriela Serban, Alina Campan, and Istvan Gergely Czibula. A programming interface for finding relational association rules. *International Journal of Computers, Communications & Control*, I(S.):439–444, June 2006.

[SCIM20] Ioana Angela Socaci, Gabriela Czibula, Vlad-Sebastian Ionescu, and Andrei Mihai. $xnow$: A deep learning technique for nowcasting based on radar products' values prediction. In *IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI 2020)*, pages 117–122. IEEE Hungary Section, 2020.

[SCW+15] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, pages 802–810, Cambridge, MA, USA, 2015. MIT Press.

[SHK+14] Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[SK99] Panu Somervuo and Teuvo Kohonen. Self-organizing maps and learning vector quantization for feature sequences. *Neural Processing Letters*, 10:151–159, 1999.

[Swe18] Swedish Meteorological and Hydrological Institute. Cooperation is a must for adaptation to and mitigation of climate change, 2018. https://www.smhi.se/en/news-archive/.

[TS19] Quang-Khai Tran and Sa-kwang Song. Computer vision in precipitation nowcasting: Applying image quality assessment metrics for training deep neural networks. *Atmosphere*, 10(5):244, May 2019.

[TSM21] Kevin Trebing, Tomasz Staczyk, and Siamak Mehrkanoon. Smaat-unet: Precipitation nowcasting using a small attention-unet architecture. *Pattern Recognition Letters*, 145:178–186, 2021.

[WKS16] Michael Wand, Jan Koutník, and Jürgen Schmidhuber. Lipreading with long short-term memory. *CoRR*, abs/1601.08188, 2016.

[WMO18] WMO - World Meteorological Organisation. Weather Climate Water, 2018. https://www.wmo.int.

[YJM+20] Qing Yan, Fuxin Ji, Kaichao Miao, Qi Wu, Yi Xia, and Teng Li. Convolutional residual-attention: A deep learning approach for precipitation nowcasting. *Advances in Meteorology*, 2020:6484812, Feb 2020.