DEGREE PROJECT IN CIVIL ENGINEERING AND URBAN MANAGEMENT,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2018

Spatio-temporal prediction of residential burglaries using convolutional LSTM neural networks

NOAH HOLM

EMIL PLYNNING

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ARCHITECTURE AND THE BUILT ENVIRONMENT

Abstract

The low amount solved residential burglary crimes calls for new and innovative methods in the prevention and investigation of the cases. There were 22 600 reported residential burglaries in Sweden 2017 but only four to five percent of these will ever be solved. There are many initiatives in both Sweden and abroad for decreasing the amount of occurring residential burglaries and one of the areas that are being tested is the use of prediction methods for more efficient preventive actions.

This thesis is an investigation of a potential method of prediction by using neural networks to identify areas that have a higher risk of burglaries on a daily basis. The model use reported burglaries to learn patterns in both space and time. The rationale for the existence of patterns is based on near repeat theories in criminology which states that after a burglary both the burgled victim and an area around that victim has an increased risk of additional burglaries. The work has been conducted in cooperation with the Swedish Police authority.

The machine learning is implemented with convolutional long short-term memory (LSTM) neural networks with max pooling in three dimensions that learn from ten years of residential burglary data (2007-2016) in a study area in Stockholm, Sweden. The model's accuracy is measured by performing predictions of burglaries during 2017 on a daily basis. It classifies cells in a 36×36 grid with 600 meter square grid cells as areas with elevated risk or not. By classifying 4% of all grid cells during the year as risk areas, 43% of all burglaries are correctly predicted. The performance of the model could potentially be improved by further configuration of the parameters of the neural network, along with a use of more data with factors that are correlated to burglaries, for instance weather. Consequently, further work in these areas could increase the accuracy.

The conclusion is that neural networks or machine learning in general could be a powerful and innovative tool for the Swedish Police authority to predict and moreover prevent certain crime. This thesis serves as a first prototype of how such a system could be implemented and used.

Keywords: crime prediction, crime forecasting, residential burglary, deep convolutional neural network, CNN, long short-term memory, LSTM, recurrent neural network

Preface

This M.Sc. thesis concludes our studies at the Degree Programme in Civil Engineering and Urban Management and the Master of Transport and Geoinformation Technology at KTH Royal Institute of Technology in Stockholm, Sweden. With an interest in both urban development and IT the topic of applying machine learning to a challenge of today's society was an exciting last semester of our studies.

The thesis has been conducted for the Swedish Police authority and we would like to thank everyone in the IT department's GIS group for the support they have given us throughout the work. Special thanks goes to Lisa Edman, our supervisor who's been assisting with small and larger matters to make this thesis possible. Additionally, we want to express our gratitude to those in the Police who took time to meet us and give their view on background theories, the current situation and how helpful these tools could be.

We would also like to express our gratitude to our academic supervisor Associate Professor Győző Gidofalvi at KTH for thoughts and guidance in the design and execution of the thesis. Last but not least, we want to thank Henry Eklind and Robin Eklind for their support in the initial development of the neural network.

Emil Plynning & Noah Holm
Stockholm, Sweden
June 2018

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

Very few residential burglary cases in Sweden are solved (Brå, 2018), and only four to five percent of burglaries lead to a person being convicted for the crime. During 2017 the reported number of residential burglaries was about 22 600 which according to these statistics would result in 21 470 unsolved residential burglary cases. Because of the low rate of solved burglaries and the issues the Police face when investigating residential burglary there is need for other methods to not only investigate and solve, but also prevent burglaries from happening.

According to Weisburd et al. (2016), five of ten identified factors explaining crime concentration are poor physical design (1), inadequate guardianship (2), repeat victimization (3), many targets (4) and hot products (5). While some of these factors are hard to control or unreasonable to change in a residential area, there are measures that may be taken. By targeting factors one and two with preventive actions the risk of crime can be decreased. Furthermore, with knowledge of which areas that are affected by factors three, four and five the possibility to influence factor one and two increases. Additionally, with increased knowledge of which places are affected by crime the possibility of making preventive actions increase.

In Sweden today there is a usage of methods for visualizing crime incidences and occasionally some low level analysis of the occurrences. However, for preventive methods to be effective the knowledge about where to target the preventive actions is needed. One such source of knowledge could be information regarding high risk areas for residential burglaries. In a personal interview in April 2018, Göran Landvall, national specialist on residential burglary, expressed that there could be a huge potential of a system that can help in the allocation of available resources. This thesis aims to research a method to identify these high risk areas with a prediction model using neural networks to help the Swedish police authorities and related parties to improve burglary prevention.

Because of the low amount of solved crimes such as residential burglaries, groups other than the police are and can be used to combat crime as well as increase perceived safety in local communities. Several examples of volunteer community support services that patrol areas of their city exist in both Sweden and abroad. The aim of these services are to strengthen safety and reduce crime. One of the newer examples is the Stockholm municipality Solna which are recruiting for their community patrol service during the work of this thesis (Sehlin, 2018; Solna Stad, 2018).

Some countries have come further than Sweden in the implementation of prediction software that is actively used by the police. There are several examples of decreased crime rates after the introduction of prediction methods. In Santa Clara County, California, the residential crime rate in a specific patrol area decreased by 23% after the introduction of a near repeat prediction method from 2010 to 2011 (Koehn, 2012).

The thesis is divided into 5 chapters where this chapter (1) describes the aims and research question, criminological theories that explain why crime prediction is possible and some of the methods used. Additionally, some theory about artificial neural networks is mentioned, as well as a study of related works. In Chapter 2 the methodology for selecting data, building and training the model is explained in detail and the results from the trained model are explained in Chapter 3. Chapter 4 is the authors' comments on the results and the methodology, as well as some comments regarding the limitations of this study. Lastly, a conclusion is drawn and further work is proposed in Chapter 5.

1.2 Aims and objectives

The aim of this thesis is to provide a method for predicting areas with an elevated risk of residential burglary. This information may be used when distributing resources to potentially reduce the amount of committed burglaries. The prediction method is based on a spatio-temporal neural network model used together with a geographical information system (GIS) to identify these areas.

1.3 Research Question

Is it possible to predict residential burglaries in the near future by using a neural network trained on data of previous burglaries? How could such a model be constructed and how accurate is it?

1.4 Criminological theories

To justify the possibility of predicting crime, this section introduces some fundamental criminological theories. Section 1.4.1 and 1.4.2 introduce theories that form the theoretical basis for the possibility to predict future crime by looking at patterns of past crime. The assumption that it is possible to reduce crime rate by predicting future crime is based on the routine activity theory in Section 1.4.3. A discussed methodology for reducing crime is introduced in Section 1.4.4.

1.4.1 Rational choice theory

One of the fundamental theories is the rational choice theory which today is used to explain parts of criminal behavior. The theory revolves around the criminal as a rational self-interested individual trying to maximize their own benefits (Gül, 2009). The rational choice theory is often a condition for the crime prevention applied today which this thesis aims to support. However, critics of the theory state that the viewpoint of the criminal as a rational individual is wrong or skewed, because criminals seldom have enough time, knowledge or resources to make a correct conclusion of the maximization of benefits of a certain action (Newman, Clarke & Shoham, 1997).

1.4.2 Near repeat

There is a strong correlation showing that victims of one crime are more likely to be victimized again in the near future. A Swedish study by Carlstedt (2001) shows that a store that has been victim of one burglary in a year is twice as likely to report an additional burglary compared to a store that had not been burgled. For burglaries that targets homes the risk is 13 times higher. The same study concludes that the risk of repeat victimization is greatest the first month after the initial crime.

The repeat victimization phenomena is not only applicable on the same victim in the case of burglaries, Bowers, Johnson and Pease (2004) shows that there is a heightened risk of burglary in an entire area after the initial event. All households within 400 meters of the burgled household are subject to a significantly increased risk of being burgled themselves for up to two months (Bowers, Johnson & Pease, 2004).

According to the rational choice theory (Section 1.4.1), criminals consider their own maximum gain when selecting their victim. Carlstedt (2001) explains that the reason criminals often choose to target the same victim again is because the risk is lower since they have better knowledge about the neighborhood, the target building etc. and a lower effort is required.

1.4.3 Routine activity

The routine activity theory or approach is a subsection of crime opportunity theories that was first proposed by Cohen and Felson (1979). The theory turns focus away from the offender's characteristics such as poverty or unemployment. Instead, the occurrence of crime is described as an opportunity when the combination of the three elements; likely offenders, suitable targets and the absence of a capable guardian occur.

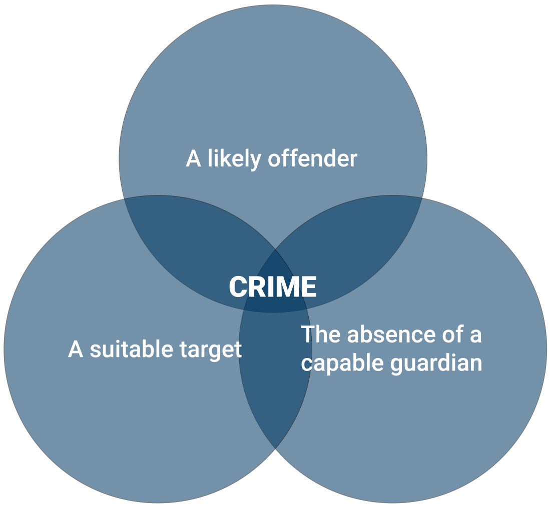

Figure 1.1: The three elements of crime outlined by the routine activity theory.

The approach of the routine activity theory and the argument of Cohen and Felson (1979) is that modern society's increase in human mobility has shifted routine activities away from people's homes to workplaces and other activities. This shift has together with a societal increase in socio-economic status increased the likelihood of the combination of the likely offenders, suitable targets and the absence of capable guardians. Suitable targets are not able to be changed and therefore crime prevention approaches must be turned to likely offenders or the provision of capable guardians. This thesis aims to provide more capable guardians by predicting risk locations and therefore possibly decreasing the amount of crime.

1.4.4 Crackdown theory

Police crackdowns are sudden activities which purpose are to decrease a specific crime incidence in a short period of time. Sherman (1990) outlines crackdowns into two main approaches; the the offense-specific and geographically focused. The offense-specific is a change in action to a specific crime where the police authority e.g. arrest offenders instead of fining them. The geographically focused approach is instead an allocation of resources to an area where there is a high number of incidences.

Crackdown approaches in policing has received much criticism for being ineffective and potentially abusive (Sherman, 1990; Scott, 2003). Critics claim that crackdowns only impact crime incidences for a limited amount of time, while being expensive and stealing resources from other areas and problems. Additionally, there is a risk that crime is displaced due to the crackdown, which only transfers the problem to another area. With an increased pressure on individual police personnel to enforce the law there is also risk of police abuse and potentially harming community relations. However, studies have been made that show that thoroughly planned and analyzed crackdowns may have longer effects and the potential to be a useful tool for police forces (Sherman, 1990; Scott, 2003).

1.5 Crime prediction methods

There is a variety of prediction methods used in predictive policing. Perry et al. (2013) introduced a categorization of predictive methods:

1. Methods for predicting crimes.
2. Methods for predicting offenders.
3. Methods for predicting perpetrators' identities.
4. Methods for predicting victims of crimes.

While there are methods for all these four categories, this section mainly aims to describe the first category. The method used in this thesis is an example of the first category.

A hotspot represents an area with a high crime concentration, relative to the distribution of crime across the whole region of interest. Chainey and Ratcliffe (2005) stress the importance of keeping this relative relationship in mind throughout the analysis. Additionally, they describe the identification of hotspots of criminal activity as one of the first steps when exploring patterns in more detail and trying to explain why certain areas suffer from more crime than others.

Several different methods have been used to map crime hotspots. Bowers, Johnson and Pease (2004) have implemented a prospective hotspot mapping technique that use a temporal weight in relation to their communicable risk. They apply greater weight to those crimes that have happened more recently and close to other similar events. They claim that this makes their methodology 30% more accurate than existing hotspot mapping techniques.

Regression methods employ a mathematical model to predict future crime by using previous crime incidences possibly together with other inputs such as number of homes in an area, number of unemployed persons etc. (Perry et al., 2013). The difference to hot spot mapping is this utilization of other variables.

Near repeat methods build on the theories described in Section 1.4.2 where algorithms use this foundation to identify high risk areas. Examples of such methods are the self-exciting point process modeling (Mohler et al., 2011) which builds upon seismological models. There are also specifically developed GIS softwares such as the Near repeat calculator by Jerry Ratcliffe from Temple University.

Chainey and Ratcliffe (2005) discuss the importance of understanding temporal changes in patterns when trying to predict, and ultimately attempting to reduce crime. The precision of a temporal analysis is determined by the temporal resolution which reflects the uncertainty of when an event occurred. They describe crime data as a reported event with usually both *from date* and *to date* which denotes the timespan. At some point in this timespan the event has occurred. Consequently, a large timespan often results in analytical difficulties but does not mean that the analyst cannot understand the patterns of criminal activity.

Recent advances in machine learning have motivated a lot of research in prediction methods for several fields such as meteorology, crime and traffic. Additional mentions of research related to this thesis is presented in Section 1.7.

1.6 Artificial neural networks

Research by Chainey and Ratcliffe (2005) into various crime prediction techniques (mainly regression) concludes that forecast methods which use retrospective data for the same month from the previous year to perform a prediction, have not been particularly accurate in the past. An approach which relies on pattern prediction based on a time series could very well outperform these regression techniques. Which is why this thesis evaluates the use of artificial neural networks (ANNs) instead of regression. The following chapter gives a short overview of the most important concepts and layers of artificial neural networks that are used in this thesis.

The usage of ANNs has exploded in the recent years as our increased computing power have enabled us to fully take advantage of their power. They have been successfully applied in many fields, the majority of which are pattern recognition and classification problems (Bishop, 1995). ANNs are computing systems that are inspired by the biological neural networks that exist in real life animal brains and many concepts within the field have been developed by studying the biological mechanisms of a brain (Bishop, 1995). Artificial neural networks "learn" tasks or patterns by studying examples of input data and the corresponding output data.

ANN models can be viewed as mathematical models trying to find the function for a given problem by learning from it. ANNs offer a general framework for representing non-linear mappings using several input and output variables (Bishop, 1995). The process of determining the values for the parameters of the ANN using a training data set is what is generally referred to as learning (Bishop, 1995). Learning in this sense implies using a set of observations and their truth or outcome to find the function that solves the task in the optimal way. In order to evaluate the learning progress, a loss function is required. The loss function indicates the distance from a particular solution to the optimal solution. Learning algorithms iteratively search through the solution space to find a function that has the smallest possible loss which should represent the most accurate solution by updating the networks weights after each epoch (iteration). ANNs improve their performance with each epoch of training until a certain point where they start learning the specific patterns of the training set instead of the general patterns. This is commonly referred to as overfitting.

1.6.1 Convolution layers

Convolutional neural networks is a subset of feed-forward neural networks that are designed to work with two-dimensional input data, usually in the form of an image.

Convolutional layers work by convolving small sized kernel across the image. Each unit in a layer receives inputs from a set of units located inside this kernel in the previous layer. This idea of connecting units to local neighborhood receptive fields was inspired by how the animal visual cortex works (Lecun et al., 1998).

With the aid of these local receptive fields, the neurons can typically identify elementary features, for instance; edges, endpoints and corners. Convolutional layers typically consist of multiple feature maps or filters, each composed with different weight vectors. This allows multiple features to be extracted at each location. These features are combined by the following layers, which are able to detect features of higher order. (Lecun et al., 1998)

The ability of multilayer convolutional networks to learn complex, high-dimensional, nonlinear mappings by training on large amounts of data makes them suitable for image recognition tasks. They are specifically designed to handle the variability of two-dimensional data and have been shown to outperform all other techniques. (Lecun et al., 1998)

1.6.2 Max pooling layers

Max pooling is a simple local operation which takes a $n \times m$ dimensional rectangle and extract the highest value found within this area and applies it to the current cell, it serves to filter out the most important features and by eliminating non-maximal values it also reduces computation for following layers.

If the stride of the max pooling is set to larger than 1, the dimensions of the input data will be reduced and parts of the activators will be ignored. This can be used to reduce overfitting and also improve training times. Since it provides additional robustness to position, max pooling is a way of reducing the dimensionality of intermediate representations without losing much of the spatial relationships.

Max pooling can also be applied in three dimensions which works in the same way with an added depth. 3D max pooling can serve to reduce temporal dimensionality and or extract the most relevant features from a temporal 2D sequence.

1.6.3 Long short-term memory networks

A recurrent neural network is a class of artificial neural networks which can be used to identify dynamic temporal behavior from a time sequence of data. In difference to ordinary neural networks, recurrent neural networks have an internal state or memory which allows them to process and learn from sequences of data.

Ordinary recurrent neural networks suffers from two problems; vanishing and exploding gradient, which make them very hard to use efficiently in practice. A way of overcoming this problem is to use Long short-term memory (LSTM) networks which are able to counter these problems. LSTMs contain a gated cell where information can be stored in, written to, or read. What is stored in the cell is controlled via gates that open and close. The behaviour of the gates is dependent on the signals they receive, they block or pass on information based on its strength. This selective behavior is what enables LSTMs to store both long- and short-term information. (Hochreiter & Schmidhuber, 1997)

1.6.4 Convolutional LSTM layers

Convolutional LSTM (ConvLSTM) networks are able to solve spatio-temporal sequence forecasting problems by combining the techniques of regular convolutional networks and adding the temporal aspect through an LSTM architecture. Stacking multiple ConvLSTM layers has proved efficient for more general spatio-temporal sequence forecasting problems (Shi et al., 2015).

A distinguishing feature of the ConvLSTM created by Shi et al. (2015) is that the 3D tensors last two dimensions are spatial dimensions (rows and columns) and the first dimension is the temporal dimension. The inputs and states can be viewed as vectors standing on a spatial grid. The ConvLSTM determines the future state of a certain cell in the grid by the inputs and past states of its local neighbors.

One distinction and advantage of this design is that all input and output elements are 3D tensors which preserves the spatial information whilst still using a recurrent perspective. Shi et al. (2015) use a network with multiple stacked ConvLSTM layers which has proven useful in forecasting complex dynamical systems.

1.6.5 Cross entropy loss

Cross-entropy loss, shown in Equation 1.1, measures the performance of a model with the output value ranging between 0 and 1. It increases as the predicted value diverges from the true value. Cross-entropy loss penalizes both false negative and false positive errors, but especially those predications that are both confident and wrong. $\mathbf{Y}$ is the true value of the outcome, $\mathbf{Y} \in \mathbb{Z}\{0, 1\}$. $\mathbf{X}$ is the predicted outcome, $\mathbf{X} \in \mathbb{R}\{0, 1\}$.

$$loss = \mathbf{Y} \cdot -ln(sigmoid(\mathbf{X})) + (1 - \mathbf{Y}) \cdot -ln(1 - sigmoid(\mathbf{X})) \tag{1.1}$$

Weighted cross-entropy loss, seen in Equation 1.2, works in the exact same way as Equation 1.1 with the addition of a multiplicative coefficient weight $w_p \in \mathbb{R}$. This weight regulates the loss of predictions by punishing false negatives and false positives in an asymmetric way by multiplying the loss for all $\mathbf{Y} = 1$. $w_p < 1$ decreases the amount of false positives, $w_p > 1$ decreases the amount of false negatives.

$$loss = w_p\mathbf{Y} \cdot -ln(sigmoid(\mathbf{X})) + (1 - \mathbf{Y}) \cdot -ln(1 - sigmoid(\mathbf{X})) \tag{1.2}$$

1.7 Related works

Similar research has been conducted to utilize recent developments of machine learning methods for prediction of crime and other more or less similar phenomena. Tran et al. (2015) use a similar 3D convolutional network architecture for prediction of spatio-temporal features in video segments. Although not directly applied on crime data the difficulties and methodology of spatio-temporal prediction remains, and many similarities can be seen in the architecture of this work. In Zhuang et al. (2017) a spatio-temporal neural network is used to forecast crime hot spots in a grid of 600 feet cells. Additional findings in the article is that the LSTM model, which is utilized in this thesis, performs better than more traditional recurrent neural network architectures. Duan et al. (2017) also apply convolutional neural networks for crime prediction in a grid of 470×380 meter cells. Using an asymmetrical grid cell structure seems odd for convolutional operations. Despite this, their use of several convolutional layers in a spatio-temporal crime network achieve high precision crime forecasting. Wang et al. (2017) uses a deep learning predictor ST-ResNet (Zhang, Zheng & Qi, 2016) which uses a residual deep learning neural network for predicting crime in a grid structure over Los Angeles. One difference in the data pre-processing is that the authors conduct another time-decision method where the crime's start time is considered as the time of crime. The related works studied are similar to this thesis in terms of aggregating crime points to a grid for easier interpretation in the deep learning algorithms. The extensive use of convolutional neural networks is another similarity, although many related works uses a deeper and more time consuming architectures than this thesis.

Additional research has been conducted in prediction methods not building on deep learning software but instead complex algorithms to create prediction functions. Yu et al. (2014) uses the self-developed Cluster-Confidence-Rate-Boosting algorithm to stitch local crime patterns into a global pattern set. A test on residential burglary data aggregated to their 800 meter cell grid shows an achieved high accuracy. Flores et al. (2015) conducts spatio-temporal crime prediction with a probabilistic risk function together with georeferenced public services such as banks, libraries etc.

2 Methodology

This chapter presents the methodology of data processing and the implementation of the neural network for predicting burglaries. It starts with explaining the data used and the preprocessing steps that were performed before using a convolutional LSTM network to compute potential risk areas for residential burglaries.

2.1 Software and frameworks

Several software and frameworks were used in this thesis. In the pre-processing steps, data integration software FME was used to filter and reformat the data into the desired format. The Python package ArcPy was used to aggregate point data and move it between spatial format and non-spatial matrices for the neural network. The neural network was implemented using the Python based neural networks API Keras with Tensorflow as backend. For final visualizations ArcGIS Pro was used together with ArcGIS Portal.

2.2 Burglary data

The data used in this thesis were reported residential burglaries from the Swedish Police authority. The model processes only time and place. The spatio-temporal autocorrelation in accordance to the Knox test as described by Kulldorff and Hjalmars (1999) was computed using a distance threshold of 1800 m and a temporal threshold of 16 days. The resulting z-score of 12.47 verifies that there is a distinct pattern for when and where burglaries occur. The Know test was performed on the data remaining after the filtering described in Section 2.4.1. There were no consideration of other environmental factors such as weather, socio-economic factors or public buildings.

Reported burglaries between 2007 and 2017 was selected from the Swedish Police authority system based on six crime codes. This amounted to 116 756 burglaries in Stockholm County and 77 300 in the study area. Brå (2006) states that the reported amount of burglaries truthfully reflects the actual amount of burglaries occurring, therefore it is considered that the data used is comprehensive and reflects the reality accurately. The selected crime codes, their corresponding description and share of the total number of crimes used follows in 2.1 below.

Table 2.1: Crime codes used in the analysis.

Theft through break in	Crime code	%
From basement, attic	0825	17
From holiday house	0826	1
From house/townhouse (attempt)	0857	6
From apartment (attempt)	0874	13
From house/townhouse (accomplished)	9801	22
From apartment (accomplished)	9802	41

2.3 Study area

The study area was a 21.6×21.6 kilometer square covering Stockholm's central and suburban parts. The area includes both apartment areas, suburban housing and townhouses. This area contains 66% of all reported burglaries in Stockholm County between 2007 and 2017.

2.3.1 Study area grid

A square area of grid cells was created to enclose the study area. A size of 600×600 meters per grid cell was deemed appropriate based on Bowers, Johnson and Pease (2004) findings that the area of significantly increased risk of being burgled as a consequence of a previous burglary is all homes within a 400 meter radius. The grid was generated orthogonal to the north and east axis, in the end it consisted of 36×36 cells covering the entire study area. One grid per day during the study period was generated to be able to store the individual events per day. The grid is shown in the results section, see Figure 3.5.

A uniform grid structure where every cell except the edge cells have eight neighbors is required in order to perform the convolutional operations that the neural network use and rely on.

2.4 Pre-processing steps

Some necessary steps were performed to prepare data for the machine learning algorithm. Firstly, the burglary data was filtered and processed (Section 2.4.1) and secondly the processed burglaries were aggregated to the study area grid (Section 2.4.2).

2.4.1 Burglary data processing

The residential burglary data was processed to structure the information in the desired format and filter based on attributes before aggregating the point to the created grid of the study area (Section 2.4.2). The difference in the reported start and end time for the crime was calculated and crimes with a difference larger than 24 hours were discarded resulting in a reduced amount of data points from 77 300 to 42 681 (55%). This action almost halved the amount of data but was deemed necessary to reduce the amount of data points with uncertain temporal resolution since the predictions are performed on a daily basis. Burglaries with start and end times on two different dates, got their date determined as the day where the start or end time was furthest from midnight. Most attributes were discarded and the only attribute that was used further in the analysis was the decided date of crime. In Table 2.1, the distribution of crime codes in the used data is shown. The majority of crime incidences are accomplished crimes in residents (63%), although some are reported as attempts (19%) and the rest are from basements and attics connected to residents.

2.4.2 Aggregation of burglaries to grid

In order to perform convolutional machine learning the point data had to be transformed to raster format. This was done by aggregating points contained by a grid cell to that cell.

The aggregation was date based which means that before actually aggregating the points one column was created per date to store the information corresponding to that day. Then all points were aggregated cell-wise, storing the number of points aggregated per date in the corresponding date column. In order to keep the data normalized between $0, 1$ grid cells with more than one incident the same day were set to 1. When studying the the cells that contain a burglary for the entire study period, 7.22% of all these cells have two occurring burglaries the same day and 1.20% have three or more burglaries occurring the same day. Because of the low incidence of cells with more than one crime per day this binary normalization was used rather than using a linear or distribution scaling. The resulting data format can be seen in table 2.2.

Table 2.2: The attribute table for one year of data post aggregation. A value of 1 indicated that there were an incident in the corresponding cell that date.

CELL ID	2015-01-01	2015-01-02	...	2015-12-31
1	0	0	...	0
2	0	0	...	0
3	0	1	...	1
4	1	0	...	0
5	0	0	...	0

The grids were then exported from shape-files to text files in a matrix format, where each individual file corresponds to one date and the burglaries that occurred that date. These matrices is the final format for the input data. The neural network model trains on sequences of these matrices. Worth noting is that the resulting matrices are extremely sparse due to burglaries being fairly uncommon when comparing to the full amount of cells. On average 12.15 cells per frame are marked as a burglary when looking at the data for 2017, this means that on average only 0.9% of the cells are non-zero.

2.4.3 Sequencing input data

Recurrent neural networks train on sequences of data. In our case the model is trained with a chronological sequence of 16 matrices starting at a date t and is then given the outcome matrix Y at date $t + 16$, in order to learn the most probable outcome given a series of frames with incidents. The sequence length of 16 was chosen for two reasons. The first reason is based on research by Johnson and Bowers (2004) where the first month after a burglary is identified as the highest risk for neighbors to experience repeat victimization (see more on this topic in Section 1.4.2). The reason for choosing two weeks instead of a month is because the risk is even higher the first two weeks, but also since a short sequence implies lower training times. The second reason for picking a sequence length of 16 is because $16 = 2^4$ which means that the output after four 3D max pooling layers is a single matrix. This is further explained in Section 2.5.1.

Table 2.3: Illustration of a sequence of input matrices.

Sequence of incident matrices				**Outcome matrix**
2015-01-01	...	2015-01-16	$\rightarrow$	2015-01-17
$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}$	...	$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}$	$\rightarrow$	$\begin{bmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \dots & y_{mn} \end{bmatrix}$

In order to improve training times a stride of four is used when building sequences for training. This means that if one sequence is started the 1st of January and ends the 16th of January, the next sequence will start on the 5th of January and end the 20th of January. With a sequence length of 16 days this stride ensures that each frame is included four times in one training epoch, instead of 16 times with a stride of one. This greatly reduces training times whilst still making sure all data occurs the same amount of times which is important to avoid bias for particular data points.

2.5 Neural network model

The model trained in this thesis uses a combination of multiple layers. The functionality of these layers are described in Section 1.6, the following sub-sections describes the composition of layers and their order. The composition of the model was decided through an iterative trial and error phase trying to find the combination of layers and parameters giving the best result when evaluated as described in Section 2.7.2.

2.5.1 Model layers

The model consists of four ConvLSTM-layers where all of the ConvLSTM-layers are followed by a Maxpooling3D-layer and the final layer before the ouput is a regular convolutional layer. The model architecture can be seen in Figure 2.1.

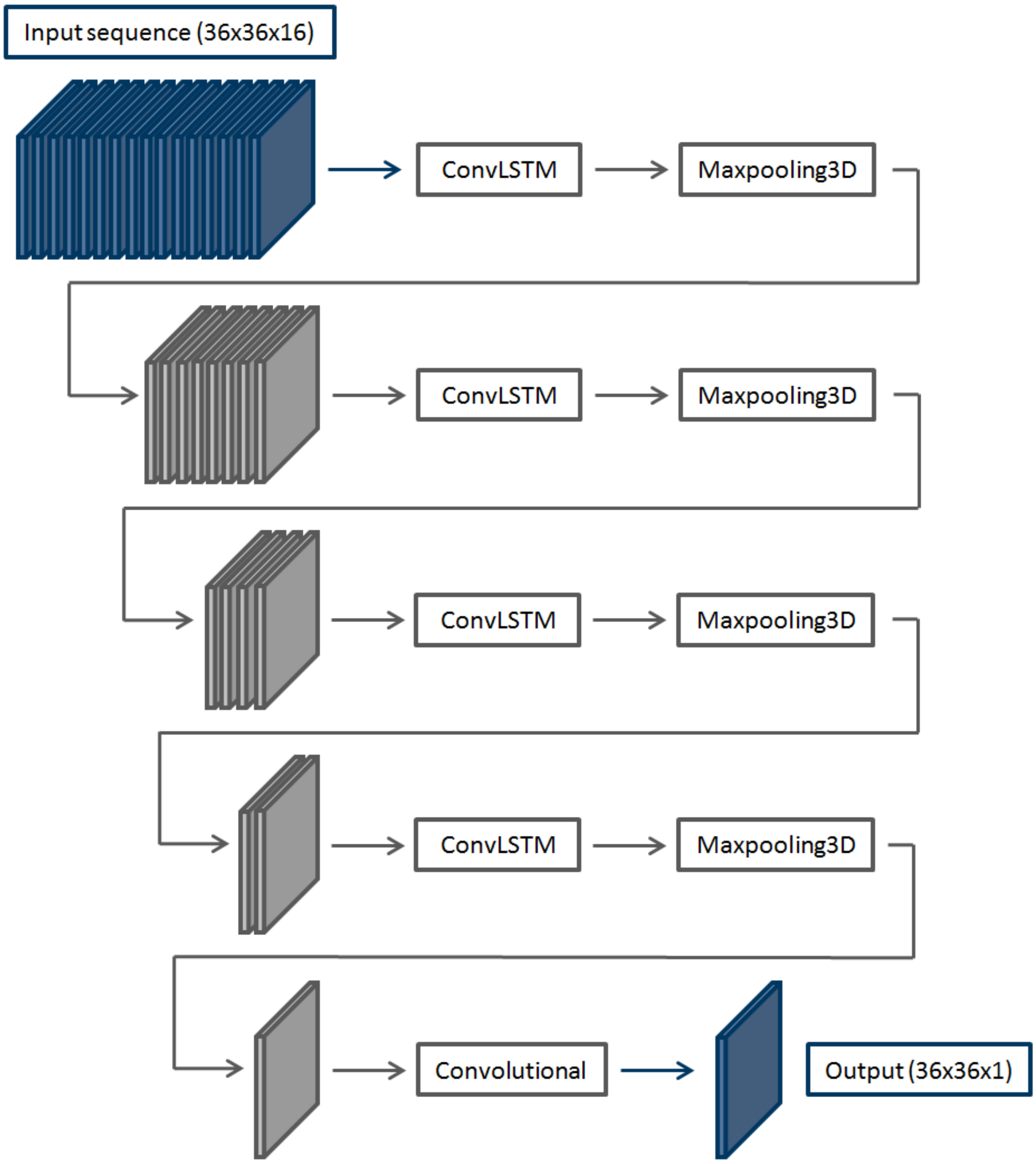

Figure 2.1: The model architecture and sequence depth following each max pooling layer.

The ConvLSTM-layers are the layers performing the spatio-temporal pattern recognition. They all have a kernel size of $(3,3)$ and use padding in order to retain the dimensions of the data after each pass since the kernel otherwise would strip the edges each time if padding is not used. The ConvLSTM-layers are followed by max pooling in order to reduce training time and to improve their ability to handle sparse data.

The max pooling layers have a kernel size and a stride of $(1, 1, 2)$. This design means that the max pooling layer extracts the most significant features from two after each other following frames and combines them into one, this process halves the amount of sequences after each layer. This means that the first ConvLSTM-layer receives a sequence of length 16, the second layer a sequence of length 8, the third 4, the fourth 2, and the last one receives a flat matrix. The speed at which max pooling is applied affects how the neural network generalizes from the training data and this method of slowly stacking max pooling with small sizes is a way of tackling sparse data whilst retaining spatio-temporal relations (Graham, 2014).

The final layer in the model is a regular Convolutional3D-layer with a kernel size of $(3, 3, 1)$ which outputs a single frame serving as the prediction.

2.5.2 Loss function

False negatives are a lot worse than false positives when it comes to crime prediction. The trade-off between false positives versus false negatives can be handled to some extent by using an asymmetric loss function. Our model uses a weighted cross-entropy loss function, see Equation 1.2, which reduced the loss of a false positive. This means that the training does not punish the model as hard when making false predictions and therefore increasing the predicted values.

This weighting of the loss function also helps to prevent the predictions from converging towards 0. With an unweighted loss function the prediction would quickly approach 0 for most cells since this is by far the most common cell value in our sparse data and therefore the most favorable prediction in most cases.

2.6 Training

The model was trained with sequences generated using the method described in Section 2.4.3. All data from 2007-01-01 to 2016-12-31 was used to train the model, resulting in a total of 909 unique sequences which were shuffled before being used as training data. The reason for using shuffled training data is to focus the model on giving an accurate prediction based only on the input sequence as its decision basis.

A validation ratio of 20% was used during training, hence the model was trained on 727 sequences and validated on 182 sequences. The model was trained for 100 epochs using a batch size of 40.

The number of epochs used to train the model was decided by examining the loss-plot and stopping the training once the validation loss had reached its minimum, see Figure 2.2. Beyond this point the model starts to overfit. The figure indicates that this is around epoch 100 which is why the final model was trained for 100 epochs. The dotted lines show the loss from a continued training-session over an additional 25 epochs. This is used to indicate a reference mark for when overfitting starts to occur and how the loss develops after continued training.

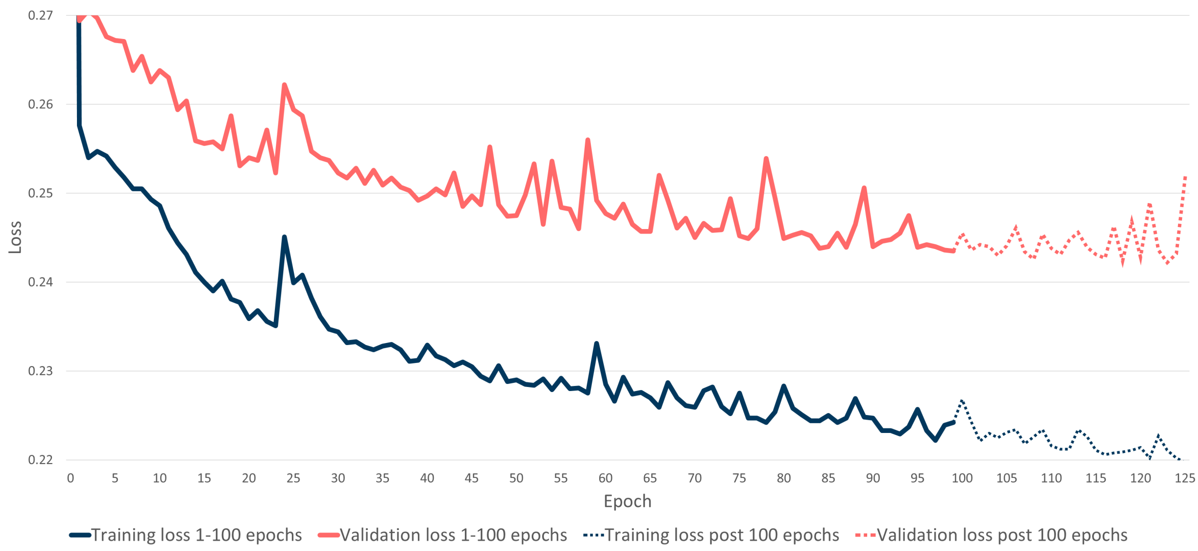

Figure 2.2: The training and validation loss of the model.

2.7 Prediction

Data from every day of 2017 is used to test and evaluate the results of the model when performing predictions on previously unseen data.

The prediction output generated from the model is a matrix with the same format as the input data. Each cell-value corresponds to the predicted cell-value for the upcoming frame (at time $t + 1$), the prediction is made using all the frames between $t - 16$ and t. A high value indicates a high predicted risk of that cell being the location of a burglary, a low value indicates a low predicted risk.

A percentile classification is used in order to determine which cells to classify as risk-areas. The boundary value for risk classification is determined by the $96th$ percentile. This percentile boundary is computed using by calculating the $96th$ percentile for all projected frames of 2017 combined. All cells with predicted values below this threshold are classified as ordinary risk areas, all cells with values above or equal to this threshold are classified as areas with elevated risk. Using an average percentile boundary for the entire year ensures that the model will classify more cells during times of high burglary rate since more projections will exceed the classification threshold, and fewer cells during times of less crime.

The choice of using the $96th$ percentile as classification boundary was based on maximizing the number of correct cells and having a high accuracy with the cells that are actually classified as risk cells. Maximizing the accuracy implies that the number of classified cells is minimized, which is important since classifying to many risk areas would make it impossible for the police to target and patrol these areas. More specifically, the $96th$ percentile was identified as the optimal threshold by multiplying the normalized percentage of correctly classified cells with the normalized accuracy and selecting the highest result. The performance of the model performing predictions with different percentile thresholds is presented in Section 3.1.1.

2.7.1 Prediction statistics

When evaluating the result of a prediction against an actual outcome six different classifications are used for the statistics. They are as follows:

1. Correct predictions
   - **Neutral -** if the cell was classified as ordinary risk and no incident occurred it is counted as a neutral cell.
   - **Correct -** if the cell was classified as an elevated risk cell and a burglary occurred in the cell, it is counted as a correct prediction.
2. Lesser errors
   - **False positive with neighboring burglary -** this class is assigned to a cell if it was classified as an elevated risk cell and no burglary occurred in the cell, but there was a burglary in a neighboring cell.
   - **False negative with neighboring positive cell -** this class is assigned to a cell if it was classified as a cell with ordinary risk but a burglary occurred in this cell and a neighboring cell was classified as an elevated risk cell.
3. Severe errors

- **False positive -** if the cell was classified as an elevated risk cell and no incidents occurred in the cell or its neighbors it is classified as a false positive.
- **False negative -** if the cell was classified as an ordinary risk cell and the cell was subject to a burglary and no neighboring cells were classified as cells with an elevated risk the cell is classified as a false negative.

When examining statistics of a model's predictions, two different sets of the data are evaluated. In the first subset only the cells of each frame that were victims of a burglary are considered. This is referred to as the **Victim-cells subset**. In the Victim-cells subset three of the classes are included (*Correct*, *False negative with neighboring positive cell and False negative*). In the other set we consider all cells of each frame. This set is referred to as the **Complete-cells subset**, in this set all classes are included.

2.7.2 Prediction evaluation

When evaluating the performance of a model's prediction capability we mainly consider the statistics measured for the Victim-cells subset. The objective is to minimize the number of *False negatives* whilst keeping the amount of *Correct* cells as high as possible in relation to the amount of *False negative with neighboring positive cell*.

2.7.3 Generating risk forecast maps

Two different approaches are used to generate risk forecast maps. The first alternative is to use a five-grade color classification scheme. The class boundaries for this scheme is based on the average percentile bounds for all projections made 2017. The second alternative is to use a binary classification where only the $96th$ percentile is classified as cells with elevated risk and all other cells are classified as ordinary risk.

3 Results

This chapter describes the output from the convolutional neural network and the results after classification of this output. In Section 3.1 the output from the neural network is explained. The classification results from the output is further explained in Section 3.1.1.

3.1 Model prediction output

The model outputs a prediction matrix with the same dimensions as the input matrix. Each cell value in the output matrix corresponds to the most probable outcome according to the model. One has to keep in mind that these cell values are computed based on the biased loss function. When performing predictions on all days of 2017 the model output has the following average statistics per output frame:

- Cell mean value: 0.0656
- Median: 0.0439
- Standard deviation of cell values: 0.0643
- 96th percentile boundary value: 0.2381

The distribution of cell values computed from the predictions of 12 individual frames can be seen in Figure 3.1. Since the cell value distribution is different each month (see Figure 3.6) one frame per month was used when computing the distribution. Each frame corresponds to the prediction output the 20th that month.

From Figure 3.1 it is clearly visible that the majority of all cells are given a prediction close to 0 and below 0.10 whilst a small part of the cells are given a high prediction value. This would indicate that the model is able to identify hot-spot-like patterns in a restricted parts of the study area and giving high predictions to a limited set of cells only. It also implies that areas with a low frequency of burglaries are given a very low prediction value. This behavior will make the model unlikely to predict outliers but increases the accuracy in areas where burglaries occur frequently.

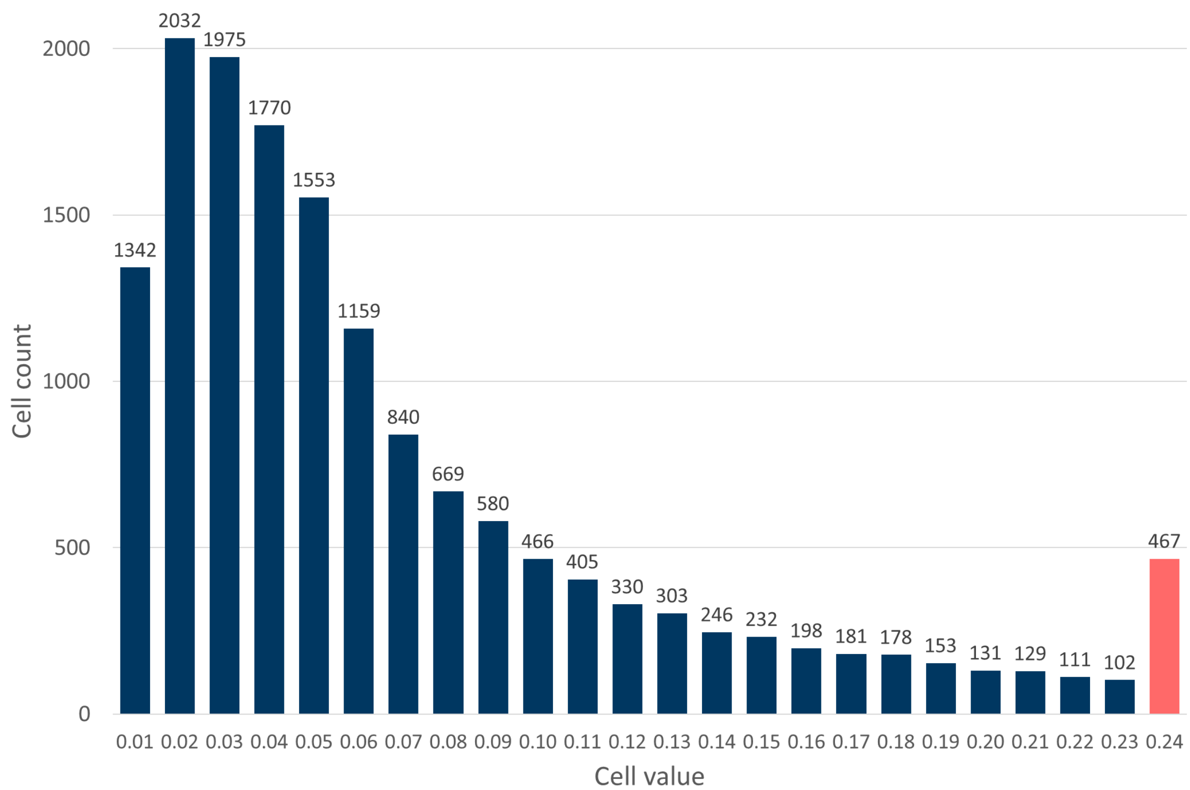

Figure 3.1: The distribution of predicted cell values computed from one frame per month 2017. Each frame was taken the 20th the corresponding month. The last bar aggregates all values larger than or equal to the $96th$ percentile.

3.1.1 Prediction statistics

Figure 3.2 shows the performance of the model using different percentile thresholds. It is clear that the model is able to predict a large amount of the occurring burglaries when the classification is based on the lower percentile thresholds. The prediction rate decrease from 74.35% in the *85th* percentile to 42.61% in the $96th$ percentile. The accuracy on the other hand increase as the percentile increases, especially after the *95th* percentile the increase in accuracy per percentile is substantial. The increased accuracy is crucial since it is important to minimize the amount of classified risk cells.

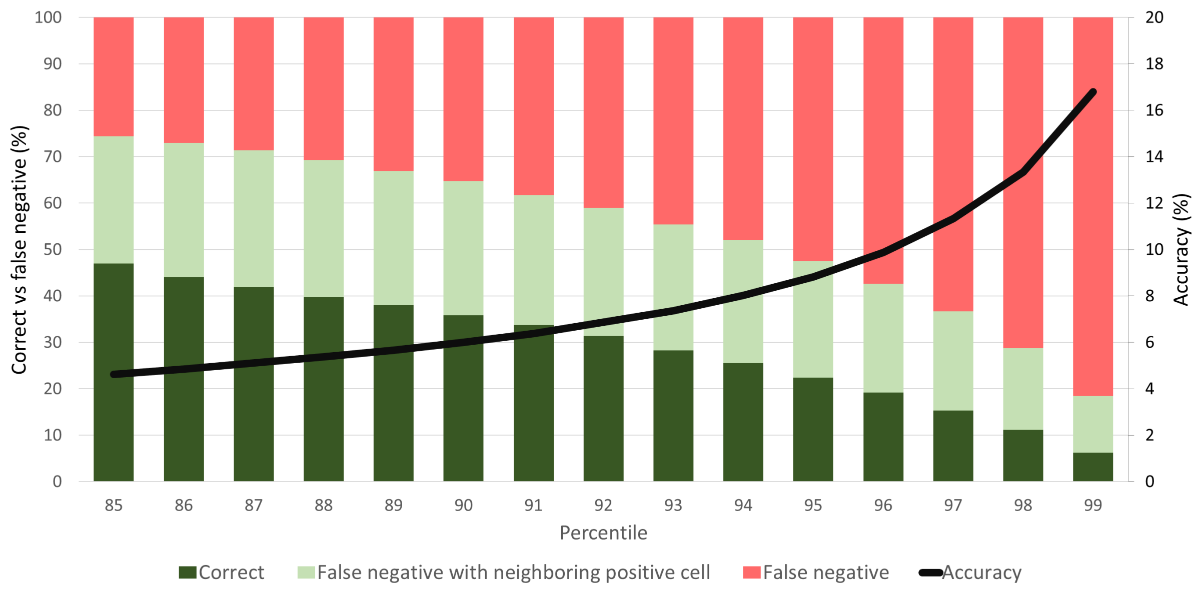

Figure 3.2: Model prediction statistics based on the percentile threshold. Left axis, the percentage of correctly projected cells and the percentage of false negatives. Right axis, the accuracy of the predicted cells (i.e., how many of the predicted cells are correct).

The class distribution for all classes using different percentile thresholds can be seen in Figure 3.3. It is evident that the most rapidly subsiding class is false positive which coincides with the increasing accuracy shown in Figure 3.2.

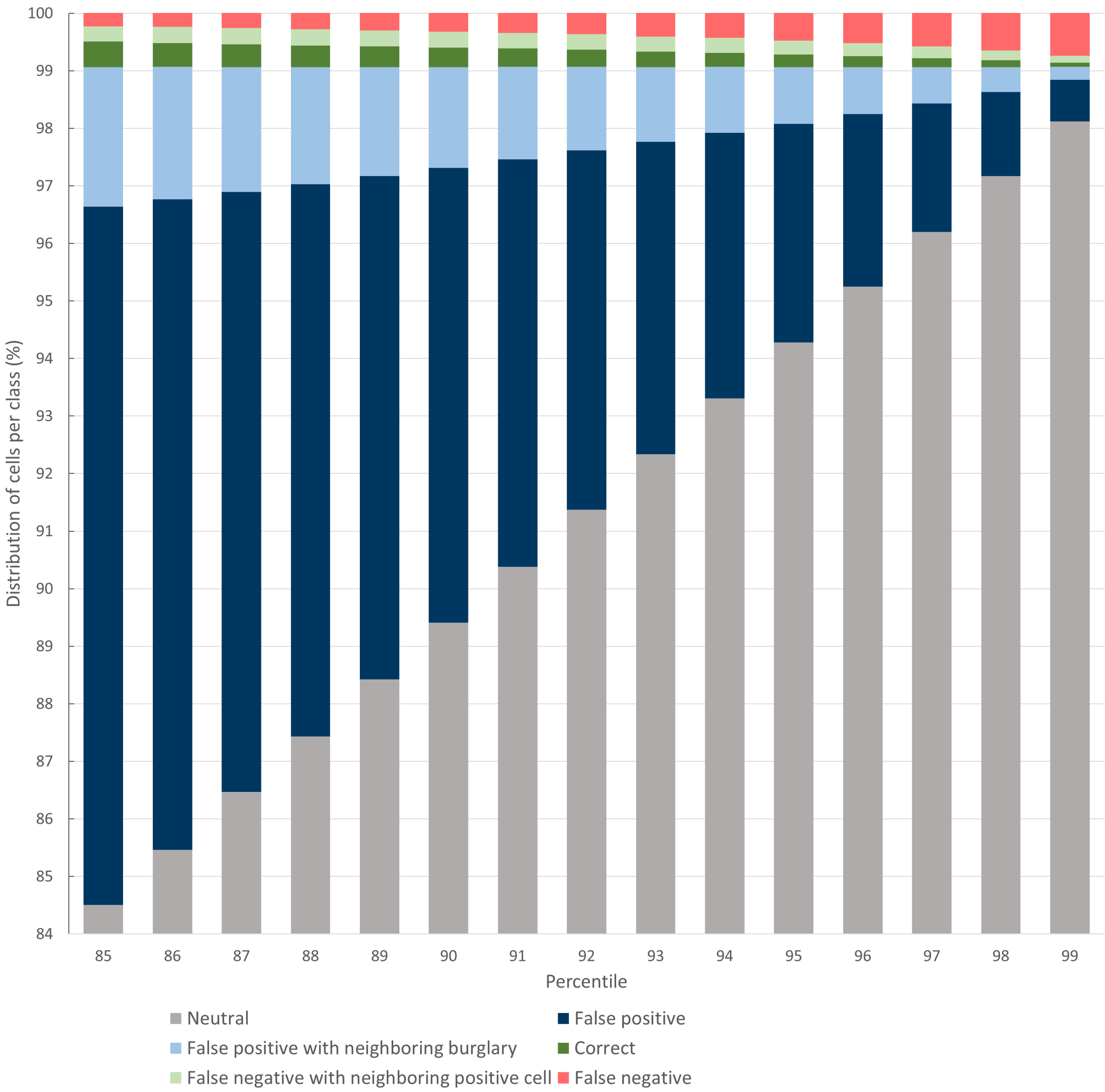

Figure 3.3: Model prediction class distribution based on the percentile threshold.

3.1.2 Prediction statistics for the 96th percentile

Figure 3.4 shows the class distribution for both the complete cells subset and the victim's only subset in more detail for the $96th$ percentile only. The model is able to predict 42.61% of all burglaries for the entire year of 2017 whilst only classifying on average 4% of all cells in the study area as risk cells.

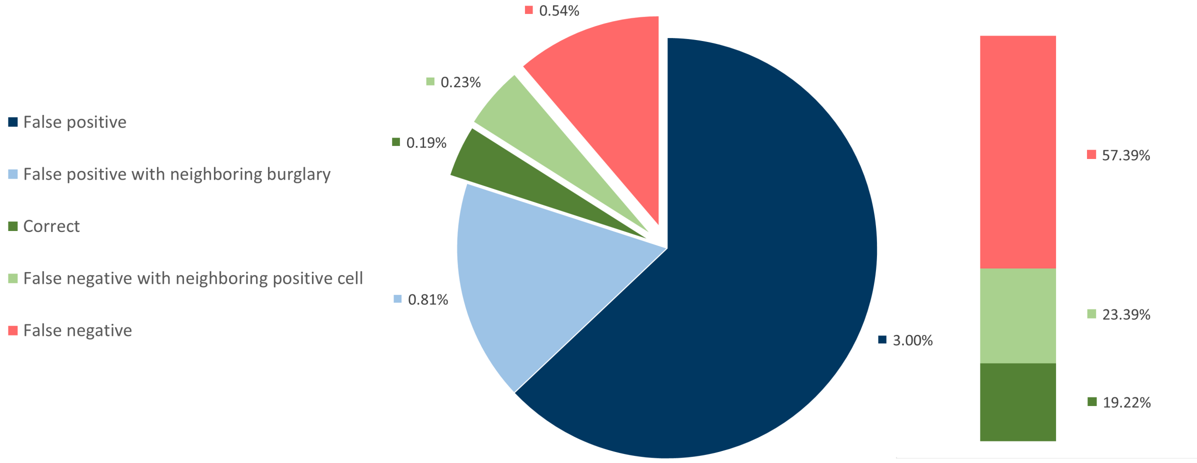

Figure 3.4: Model prediction class distribution for the $96th$ percentile. The pie chart shows the class distribution for the complete cells subset and the bar chart shows the class distribution for the victim cells subset.

A visualization of the output that has been used to calculate the statistics presented in Sections 3.1.1 and 3.4 can be seen in Figure 3.5. The map presents data from a day in mid-November which is burglary high season, see Figure 3.6 for a visualization of seasonal trends. The map plots the classified cells together with the burglaries that occurred the same day and should serve to clarify how the neighborhood based classification turns out in practice.

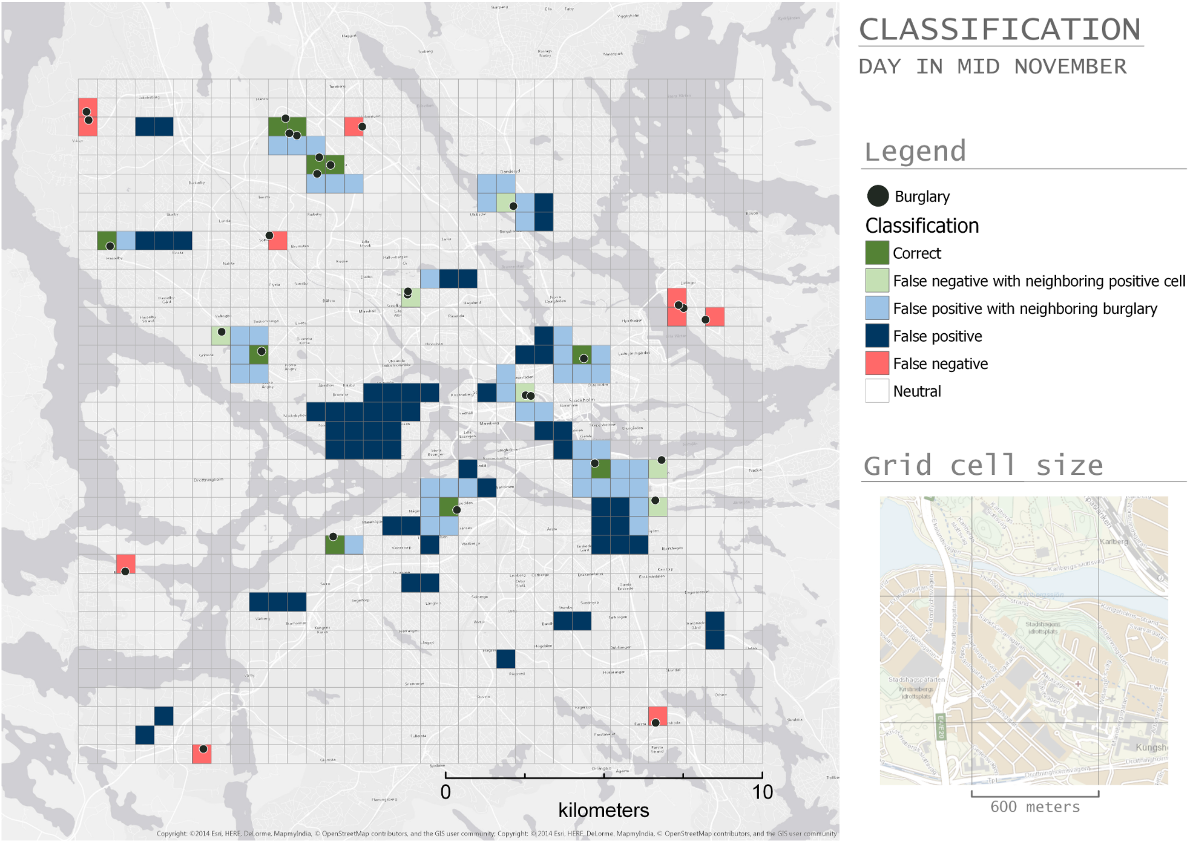

Figure 3.5: Classification of prediction results for a single day in November using the $96th$ percentile as threshold.

3.1.3 Risk cell forecasts per day

A clear temporal correlation between the number of burglaries and the number of projected risk cells per day can be identified when studying at this pattern over a longer time series. The Pearson correlation coefficient between these two series is 0.51 and the correlation of the 5 period moving averages is 0.79. This is illustrated in Figure 3.6 which shows a plot of the number of projected risk cells against the number of burglaries that occurred and the constant 4% of the cell count for the entire year of 2017. The constant 4% line indicates what the average amount of projections should be since the $96th$ percentile is used as the boundary. This indicates that the model can adapt to seasonal patterns in increased crime rate even though the events themselves have never been seen before by the model.

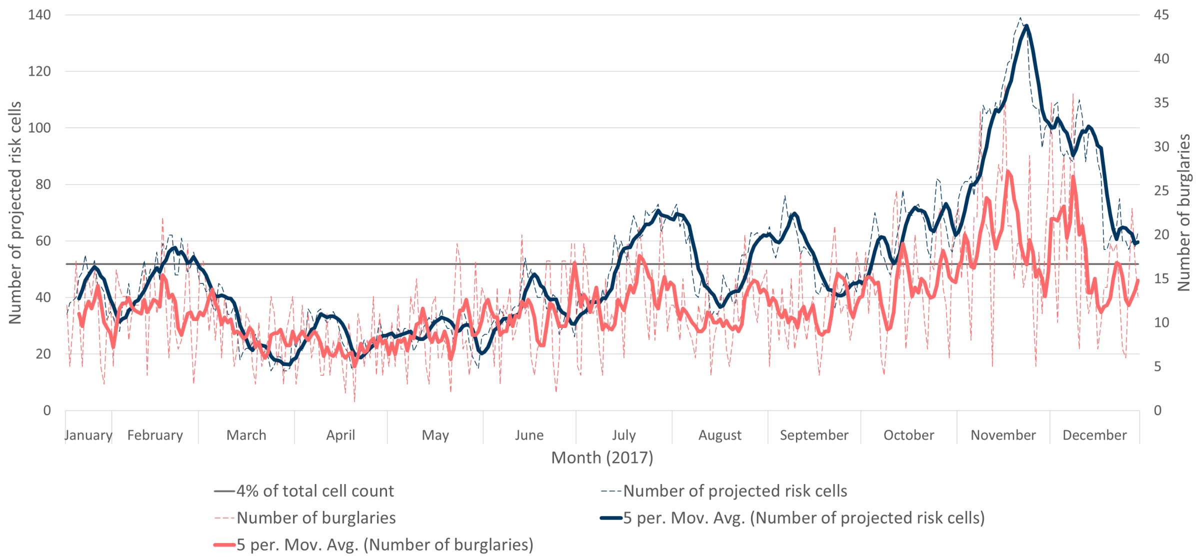

Figure 3.6: The number of cells classified as risk cells and the number of burglaries per day. The projected risk cells and the constant 4% line follows the left axis and the number of burglaries follows the right axis.

When observing the moving average of the number of burglaries and the number of predicted cells it is clear that the model is lagging behind the actual outcome by a few days. This is because the model use a sequence of previous data to compute the outcome of the next day which means that periods of high and low crime rate will stick with the model until the entire sequence contains new data.

3.1.4 Daily risk area map

Göran Landvall who is the Swedish Police's national specialist on residential burglary and responsible for the national strategy fighting burglaries clearly stated during his interview that it is of great importance that the output from the model is easy to interpret and understand for both regular officers and all other possible audiences. With this in mind, there are two different alternatives for visualizing the daily output risk prediction map.

The first alternative is to use the projected cell values and a five-grade color classification scheme based on the combined percentile bounds for all projections made 2017 as described in the first part of Section 2.7.3. The second alternative is to use a binary classification map where only the $96th$ percentile is classified as areas with elevated risk and all other areas are classified as ordinary risk, as described in the second part of Section 2.7.3.

Two example risk maps using the first alternative can be seen in Figure 3.7 and 3.8. Two risk maps for the same dates but instead using the second mapping alternative can be seen in Figure 3.9 and 3.10. Both maps use the same percentile classification boundaries but different dates, the map in March corresponds to typical output during burglary low season and the map in November corresponds to output during the peak of burglary high season.

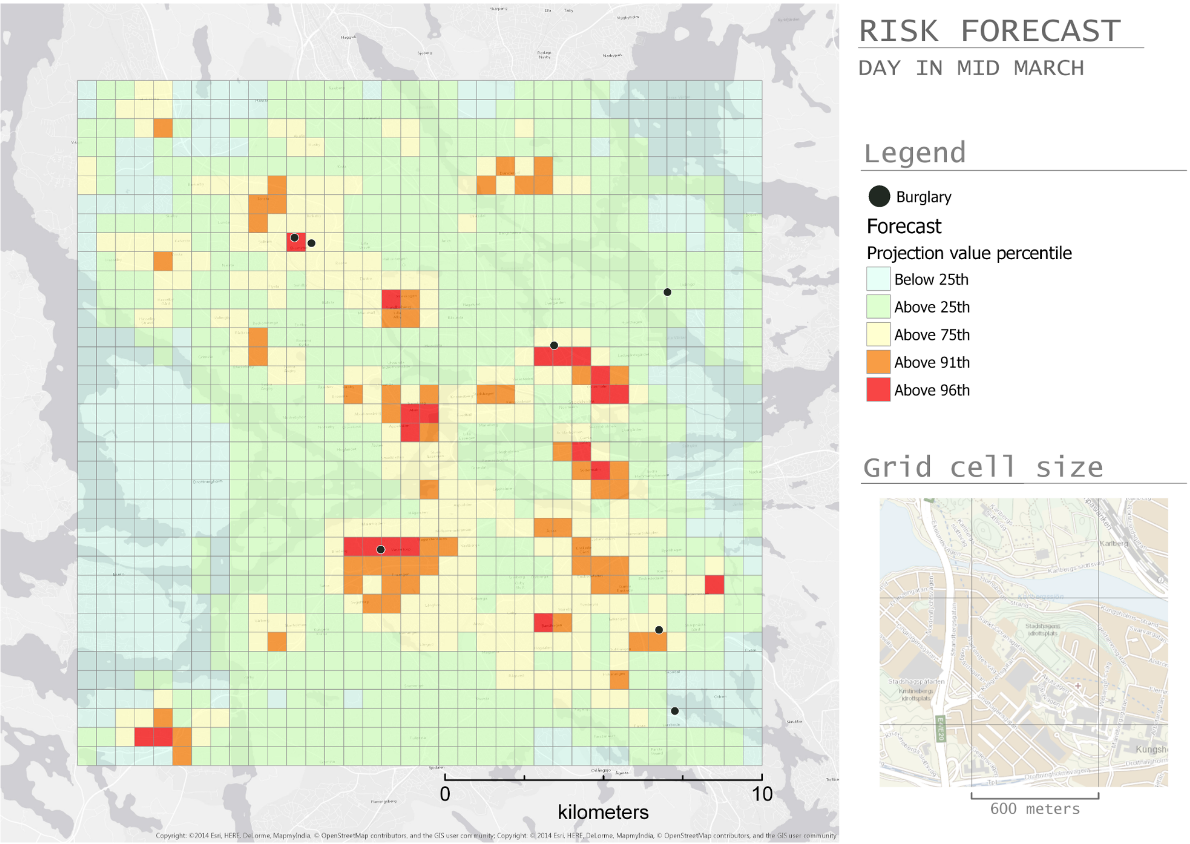

Figure 3.7: Risk forecast for one day in mid-March 2017 (burglary low season) as a continuous surface classified according to percentiles cell of values.

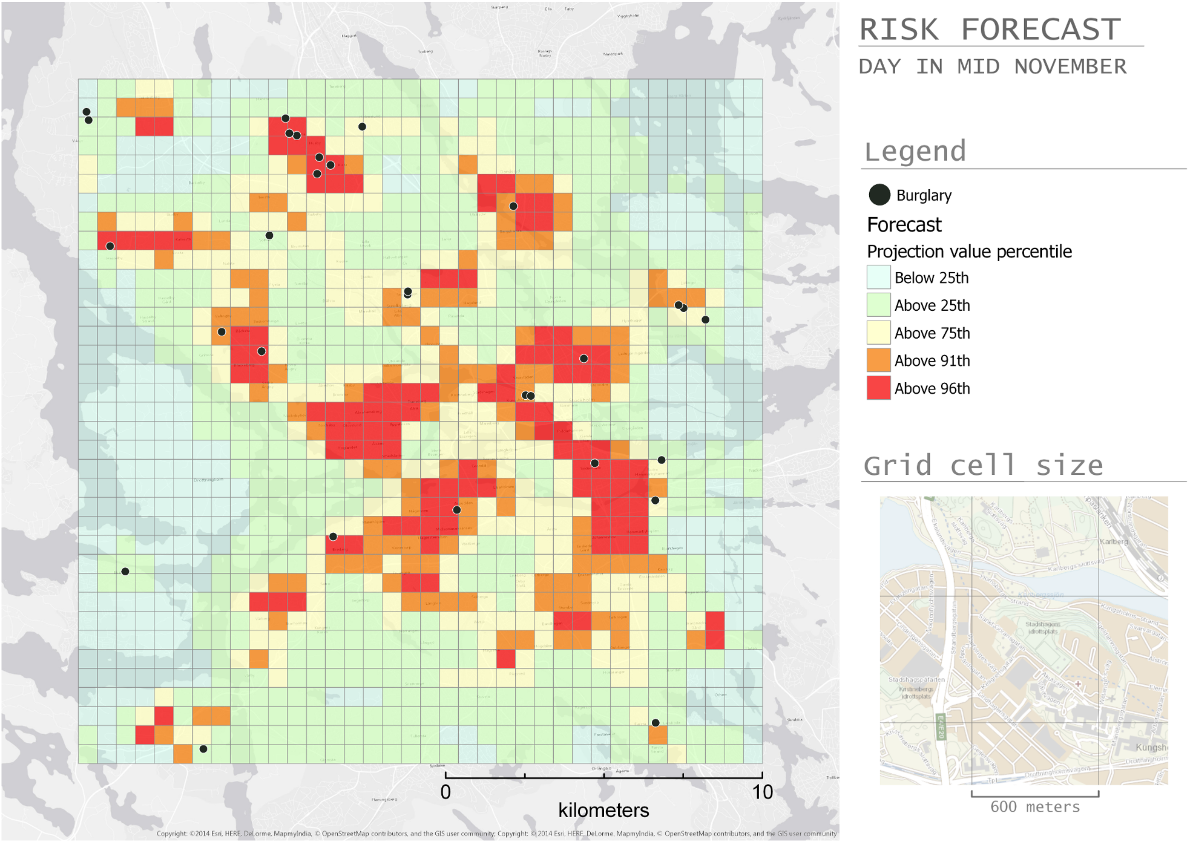

Figure 3.8: Risk forecast for one day in mid-November 2017 (burglary high season) as a continuous surface classified according to percentiles cell of values.

The advantage of alternative one is that it gives a more continuous overview of the risk areas and gives an indication of how the risk between neighboring cells relate to each other. The disadvantage is that the map can be interpreted to indicate more risk areas which can make it harder to pinpoint where to focus resources.

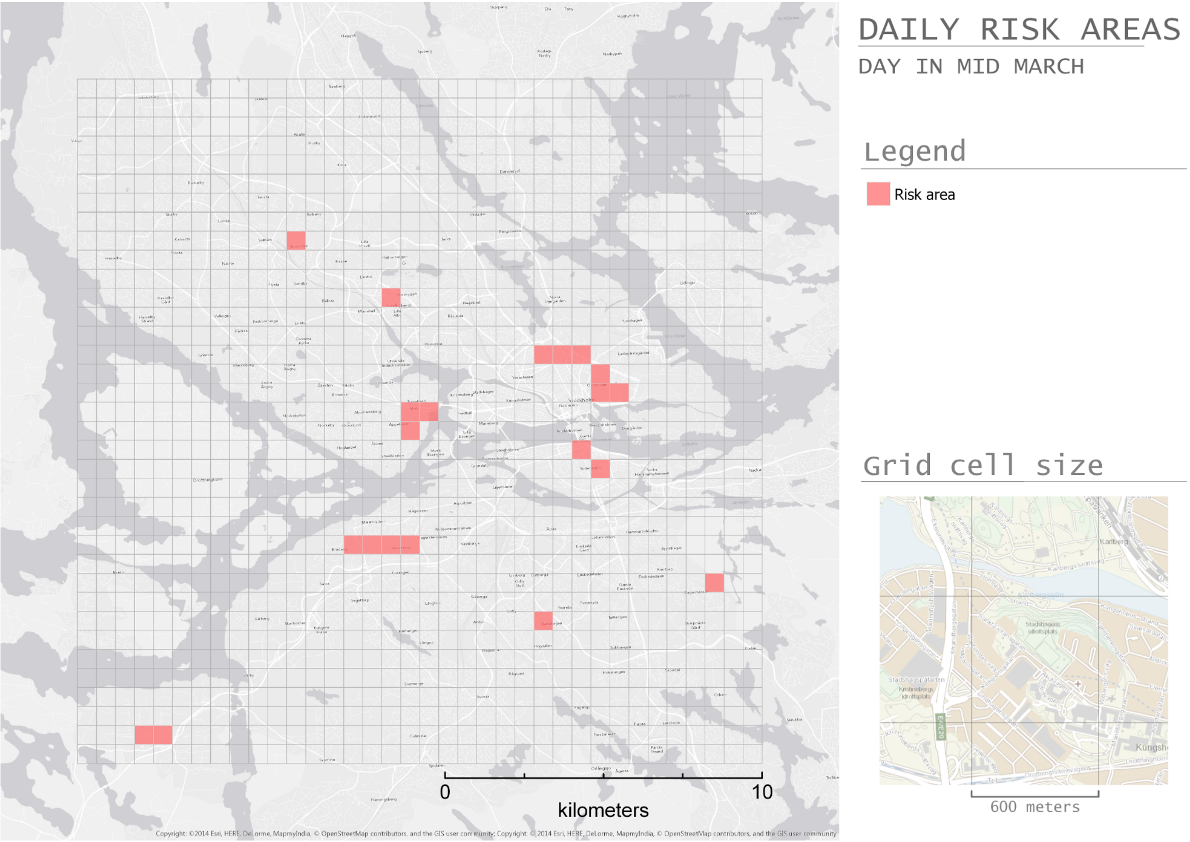

Figure 3.9: Risk areas outline by values exceeding the $96th$ percentile threshold for one day in mid-March 2017 (burglary low season).

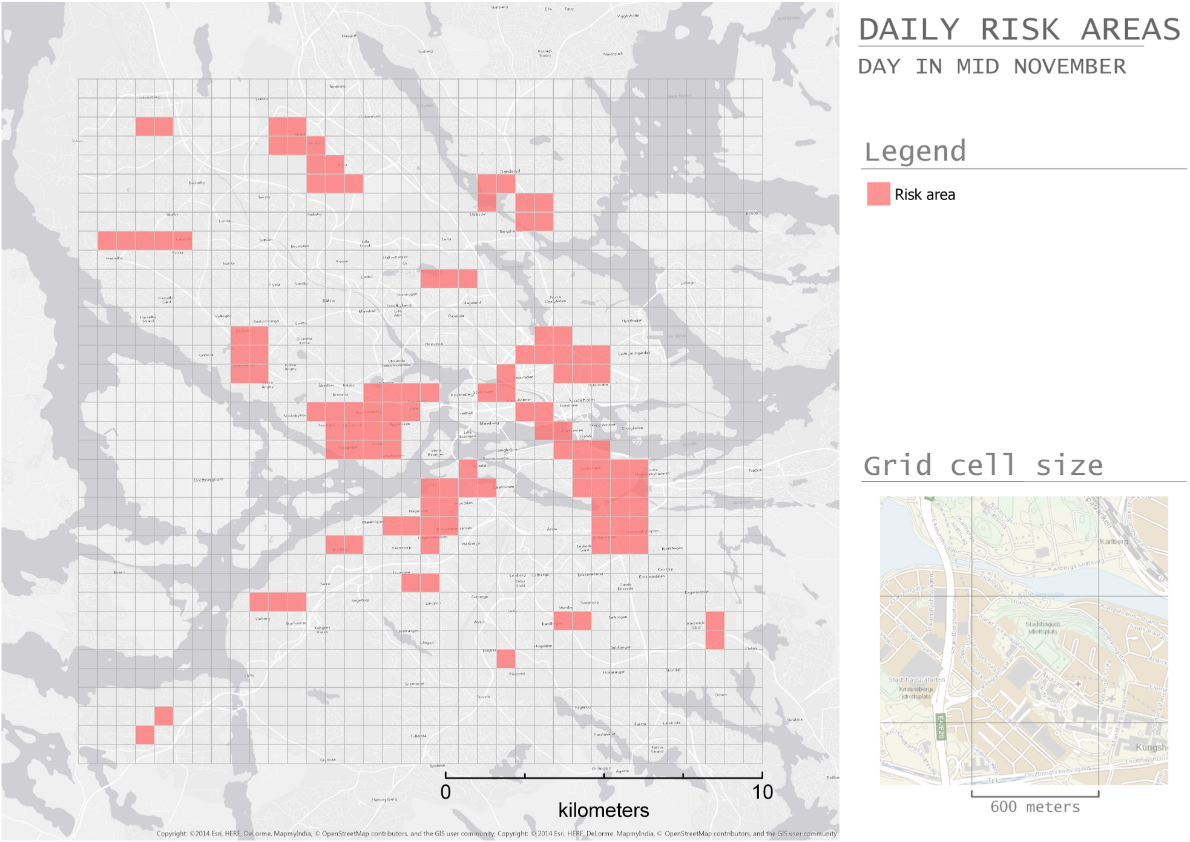

Figure 3.10: Risk areas outline by values exceeding the $96th$ percentile threshold for one day in mid-November 2017 (burglary high season).

The advantage of alternative two is that the binary classification makes it easy to identify risk areas and direct resources towards these. The disadvantage is that cells that fall just below the $96th$ percentile are classified equally as a cell in for instance the $25th$ percentile which can be misleading when studying the broader picture.

4 Discussion

The aim of this thesis was to investigate the potential to predict residential burglary in the near future. The results show that our neural network trained on ten years worth of burglary data does succeed in identifying burglary risk areas on a daily basis. The constructed model builds on a deep convolutional neural network which uses LSTM and max pooling to learn spatio-temporal patterns in a grid covering Stockholm which all reported burglaries were aggregated to. Our results and accuracy indicate that this is a functioning method which could be useful for law enforcement in order to reduce the number of burglaries committed. Further comments on the methodology and limitations of the data can be found in Section 4.1 and comments regarding the implementation of the neural network in Section 4.2.

As mentioned in Section 1.7 most of the similar research aiming at creating prediction models with machine learning also used deep convolutional neural networks. One differentiation between our study and many of the other studies is their use of deeper and more complex architectures which are considerably more time consuming. Due to lack of suitable hardware to train more complex architectures in a reasonable time the depth and complexity of our ANN was limited. This study also implements the convolutional-LSTM layers invented by Shi et al. (2015) which is an architecture that none of the other studies have tried out and it proved to be successful for crime prediction.

Regardless of the differences, there are similar findings that machine learning can be a tool for predicting crime. The findings are in agreement with the conclusion of Duan et al. (2017) that even though there are still many uncertainties in how the deep and complex algorithms in the neural network work, it is a field worth researching for preventing crime in the near future.

4.1 Data quality and filtering

The quality of crime data used in the analysis is considered good for several reasons. The report rate of occurred burglaries is high and the number of unrecorded cases is so low that it is negligible which makes the amount of burglaries in the data true to reality. The spatial quality of the reported burglaries is also considered precise due to all burglaries being reported at an address which is geocoded. Additionally, the methodology of aggregation to a grid eliminates smaller errors in the reported location. The least correct information in the burglary data is the start and end time of the incident report. This is a result of both uncertainty of the actual time of crime due to people being out of home, but also because of irregularities when the crime is reported in the system. The methodology of filtering crimes with a start- and end time difference of more than 24 hours eliminates the crimes that are reported with a large time uncertainty. However, there are also many crimes reported within 24 hours that are probably as uncertain. Some crimes have the same start and end time which seem to have been arbitrarily reported, for instance 00:00. These issues could not be managed due to lack of information and risk of lowering the amount of data.

4.2 Neural network implementation

Building the model based on recurrent convolutional neural networks requires the conversion from point vector-data with perfect spatial resolution to a raster format. The advantage of this is that there are several well developed machine learning frameworks which one can use that supports convolution based learning on raster data. The disadvantage is that the raster format restricts the spatial precision to one grid cell. Considering that a prediction model like this probably cannot achieve a spatial precision much higher than what the raster resolution in this case allows, the advantage of using well established frameworks outweighs the cons of moving from vector to raster. However one of the problematic parts of aggregating point data to a raster is so called edge effects which means that some of the points will be right at the edge of one cell hence being almost as close to its neighboring cell but still being aggregated only to one cell. This problem is hard to avoid but the effects of this can be regulated by changing the cell size where a smaller cell causes more spatially accurate aggregation. Another possible approach to reduce the edge effects, would be to use two overlapping shifted versions of the grid in both the training and prediction phase resulting in two models and two predictions. The output predictions could then be combined and averaged which would smooth the results.

Because of the aggregation of points to areal units the modifiable areal unit problem (MAUP) is a potential source of bias in our implementation. The potential sources of bias are different cell sizes and the placement of the grid. However, because of the use of a uniform grid instead of administrative regions the potential effects of MAUP is lower. Larger cell sizes would create less cells with no burglaries, hence reduce the sparsity of the training data and probably increase the accuracy, but it would also result in larger risk areas with lower spatial precision.

Using a deep neural network architecture involves considerable amounts of work trying to identify the optimal configuration and parameters to use in order to achieve the best possible result. A great deal of effort was put into optimizing the model. However, due to the large amount of parameters, number of layers etc. that are possible to vary we are convinced that even though we have identified a well performing model there are better configurations to be found.

Apart from parameter tweaking one factor that could contribute to greatly alter the result would be to modify the loss function. The cross entropy loss function that we use only considers one to one matches between individual cells when calculating the loss. This means that the loss of classifying one cell as a cell with elevated risk, and then having a burglary occur in its neighboring cell is considered equally bad as not having anything happen at all. This is something that definitely could alter the behavior of the model and maybe improve its performance. The use of an asymmetric loss that is biased towards false positives is one thing that seems to be crucial when working with data as sparse as burglary data.

The model was trained using shuffled training data. One could argue that a disadvantage with this could be that the model fails to catch seasonal patterns in burglary rates but as shown in Section 3.1.3 the model is still able the pick up on these kind of patterns. Therefore the usage of shuffled training data is something that focuses the model on learning only the patterns given from one sequence of data. It is clear that a sequence containing an elevated amount of incidents do increase the models amount of predictions for the frame following that sequence. In Section 3.1.3 it was mentioned that even though the model adapts to seasonally varying rates of occurring burglaries a slight lag in this adaption can be noticed due to "old" data remaining in it's sequence. This lag could be reduced by using shorter input sequences, hence reducing the time span which influences the models prediction. This would require changing the model architecture since the maxpooling is designed for a input sequence of 16 frames. An interesting method to tackle this issue could be to build and train two separate models, one that trains and performs predictions using a long sequence and one that trains and performs predictions using a short sequence. The output predictions of these two models could then be combined as for instance a weighted average thus allowing the user to bias towards the time span that is most relevant.

The choice to classify burglaries that occur in a cell next to a cell with a classified elevated risk as correct could be seen as a bit controversial. Our logic and motivation behind this decision is that an increased presence or awareness in one particular cell would definitely have an effect on the neighboring cells which is why it is reasonable that this would have a preventive effect in those cells as well.

5 Conclusion and future work

5.1 Conclusion

As shown in the review of related works several attempts have and are being made to predict crime in order to make crime prevention more efficient. In some areas, the prediction methods have also been implemented in the daily police work with successful results. Although the accuracy of the prediction model we have developed has room for further improvement this work is a first step into a potential future where the Swedish Police authority utilizes modern and innovative solutions to help reduce crime rates. It is also a lookout to new technologies where machine learning is becoming more understood, sophisticated and used to greater extent. With more refined data processing, storage and analysis methods comes new possibilities to utilize the data that the police has gathered for years. This is a natural step in today's digitalization efforts.

One potential usage of the predicted risk areas is to send police to patrol the area if there is nothing more important to attend. However, limited resources will probably make this a neglected task. Instead, this could be used in an ecosystem of crime preventive parties where municipalities, neighborhood watches and individual citizens cooperate with the police to reduce crime rates in their areas.

5.2 Future work

This is a first prototype of a prediction model which could be implemented to predict and furthermore prevent crime. However, some things are left unfinished or unperformed due to lack of time and scope during this thesis.

In the design of the architecture for the prediction model there are a wide range of parameters that needs to be configured. These parameters have been identified and selected based on related works studies, empirical tests and best practice information. Further work into the design of the architecture and configuration of the model's parameters may improve the accuracy of the predictions.

By using the uniform grid instead of administrative regions the data is aggregated to areas disregarding natural dividers such as water, larger roads or forests. For the convolutional operations to work the grid structure with eight neighbors per cell is needed. However, there could be potential in using administrative regions and defining each region's neighbors using a graph structure with the condition of eight neighbors. However creating a division like this compatible with convolutional operations is not an easy task.

The burglary data was filtered to disregard all burglaries with start and end time difference of over 24 hours. A potential alternative method would be to include all reported burglaries and evenly distribute them over the days they reportedly occurred. A burglary with start and end time spanning four days would then have a weight of 0.25 for each day. A different normalization method rather than the binary normalization used in our method should then be used.

Our work only utilized reported residential burglaries. By including different type of data the accuracy and reliability of the predictions could potentially increase. By allowing the neural network learn using more factors that potentially affect the levels of residential burglaries it will receive more information to base the predictions on. Such data could be weather, socio-economic factors, public building locations or housing types. Additionally, specifics about space around housing such as if it borders to a forest could be a desirable extension. However, such specific data is unlikely available or existing and would require a large effort to gather.

A possible further extension of including more data is to refocus a different neural network learning to identify which factors that burglars like and dislike in their targets. By identifying these risk factors, such as having a large hedge or a lawn beside a forest the preventive actions against residential burglaries could be improved further.

To implement the model in daily police work a seamless implementation needs to be configured. We propose a web map service which trains during the night and delivers new risk areas for the current day in the morning. Automatic scripts for collecting, processing, training and visualization will need to be implemented to deliver an easily interpreted risk area map for normal police officers and managers. Studies on the density requirements of areas need to be performed, as rural areas with very few amount of burglaries will be difficult to predict.

References

Bishop, Christopher M. (1995). *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc. ISBN: 0198538642.

Bowers, Kate J., Shane D. Johnson & Ken Pease (2004). "Prospective Hot-Spotting: The Future of Crime Mapping?" In: *The British Journal of Criminology* 44.5, pp. 641–658.

Brå (2006). *Konsten att läsa statistik om brottslighet*. ISBN: 91–38–32253–6.

– (2018). *Bostadsinbrott*. URL: `https://www.bra.se/brott-och-statistik/statistik-utifran-brottstyper/bostadsinbrott.html` (visited on 04/17/2018).

Carlstedt, Malena (2001). *Upprepad utsatthet för brott*. Brottsförebyggande rådet, Information och förlag. ISBN: 91-38-31772-9.

Chainey, Spencer & Jerry Ratcliffe (2005). *GIS and Crime Mapping*. John Wiley & Sons, Inc., pp. 145–256. ISBN: 978-11-1868-518-1.

Cohen, Lawrence E. & Marcus Felson (1979). "Social Change and Crime Rate Trends: A Routine Activity Approach". In: *American Sociological Review* 44.4, pp. 588–608. ISSN: 00031224.

Duan, L., T. Hu, E. Cheng, J. Zhu & C. Gao (2017). "Deep Convolutional Neural Networks for Spatiotemporal Crime Prediction". In: *2017 International Conference on Information and Knowledge Engineering (IKE)*, pp. 61–67. ISBN: 1-60132-463-4.

Flores, Paulina, Mario Vergara, Pablo Fuentes, Francisco Jaramillo, David Acuna, Aramis Perez & Marcos Orchard (2015). "Modeling and Prediction of Criminal Activity Based on Spatio-Temporal Probabilistic Risk Functions". In: *Annual Conference of the Prognostics and Health Management Society 2015*. Vol. 6.

Graham, Benjamin (2014). "Spatially-sparse convolutional neural networks". In: *Computing Research Repository (CoRR)*.

Gül, Serdar K. (2009). "An Evaluation of the Rational Choice Theory in Criminology". In: *GAU Journal of Social and Applied Sciences* 4.8, pp. 36–44.

Hochreiter, Sepp & Jürgen Schmidhuber (1997). "Long Short-term Memory". In: *Neural computation* 9, pp. 1735–1780.

Johnson, Shane D. & Kate J. Bowers (2004). "The Burglary as Clue to the Future: The Beginnings of Prospective Hot-Spotting". In: *European Journal of Criminology* 1.2, pp. 237–255.

Koehn, Josh (2012). *Algorithmic Crimefighting*. URL: http://www.sanjose.com/2012/02/22/sheriffs_office_fights_property_crimes_with_predictive_policing/ (visited on 02/06/2018).

Kulldorff, Martin & Ulf Hjalmars (1999). "The Knox Method and Other Tests for Space-Time Interaction". In: *Biometrics* 55.2, pp. 544–552. ISSN: 0006341X, 15410420. URL: http://www.jstor.org/stable/2533804.

Lecun, Yann, Léon Bottou, Yoshua Bengio & Patrick Haffner (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE*, pp. 2278–2324.

Mohler, G. O., M. B. Short, P. J. Brantingham, F. P. Schoenberg & G. E. Tita (2011). "Self-Exciting Point Process Modeling of Crime". In: *Journal of the American Statistical Association* 106.493, pp. 100–108.

Newman, G.R., R.V.G. Clarke & S.G. Shoham (1997). *Rational Choice and Situational Crime Prevention: Theoretical Foundations*. Commonwealth Parliamentary Association Series. Dartmouth. ISBN: 978-18-5521-947-2.

Perry, Walter L., Brian McInnis, Carter C. Price, Susan Smith & John S. Hollywood (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation. ISBN: 978-0-8330-8148-3.

Scott, Michael S. (2003). *The benefits and consequences of police crackdowns*. Office of Community Oriented Policing Services. ISBN: 1-932582-24-X.

Sehlin, Ann (2018). *Sökes – pigga pensionärer för att förhindra inbrott*. URL: `https://www.svt.se/nyheter/lokalt/stockholm/sokes-pigga-pensionarer-for-att-forhindra-inbrott` (visited on 04/17/2018).

Sherman, Lawrence W. (1990). "Police Crackdowns: Initial and Residual Deterrence". In: *Crime and Justice* 12, pp. 1–48. ISSN: 0192-3234.

Shi, Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong & Wang-chun Woo (2015). "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting". In: *Computing Research Repository (CoRR)*.

Solna Stad (2018). *Grannstödsbil i Solna stad*. URL: `https://www.solna.se/sv/familj-omsorg/nyheter-familj-omsorg/grannstodsbil-i-solna-stad/` (visited on 04/17/2018).

Tran, D., L. Bourdev, R. Fergus, L. Torresani & M. Paluri (2015). "Learning Spatiotemporal Features with 3D Convolutional Networks". In: *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 4489–4497.

Wang, Bao, Penghang Yin, Andrea L. Bertozzi, P. Jeffrey Brantingham, Stanley J. Osher & Jack Xin (2017). "Deep Learning for Real-Time Crime Forecasting and its Ternarization". In: *Computing Research Repository (CoRR)*.

Weisburd, David et al. (2016). *Place Matters: Criminology for the Twenty-First Century*. Cambridge University Press.

Yu, Chung-Hsien, Wei Ding, Ping Chen & Melissa Morabito (2014). "Crime Forecasting Using Spatio-temporal Pattern with Ensemble Learning". In: *Advances in Knowledge Discovery and Data Mining*. Springer International Publishing, pp. 174–185. ISBN: 978-3-319-06605-9.

Zhang, Junbo, Yu Zheng & Dekang Qi (2016). "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction". In: *Computing Research Repository (CoRR)*.

Zhuang, Y., M. Almeida, M. Morabito & W. Ding (2017). "Crime Hot Spot Forecasting: A Recurrent Model with Spatial and Temporal Information". In: *2017 IEEE International Conference on Big Knowledge (ICBK)*. Vol. 00, pp. 143–150.

TRITA TRITA-ABE-MBT-18268