Predicting crime across cities and regions : a comparative analysis of predictive modelling in three Belgian settings

Reference:
Hardyns Wim, Khalfa Robin.- Predicting crime across cities and regions : a comparative analysis of predictive modelling in three Belgian settings
Applied spatial analysis and policy - ISSN 1874-4621 - Dordrecht, Springer, 16:1(2023), p. 485-508
Full text (Publisher's DOI): https://doi.org/10.1007/S12061-022-09485-9
To cite this reference: https://hdl.handle.net/10067/1920580151162165141

Predicting crime across cities and regions: A comparative analysis of predictive modelling in three Belgian settings

Introduction and background

The importance of the environment in generating or discouraging crime has long been recognised in theoretical and empirical criminology (e.g. Brantingham & Brantingham, 1981, 1991; Taylor & Harrell, 1996; Andresen, 2014). Especially within the subfield of spatial-temporal criminology, the relationship between crime and environmental factors is heavily studied. From this perspective, crime is often explained as the result of an interaction between personal and environmental characteristics (Wikström & Treiber, 2016), in which the spatial-temporal clustering of crime (e.g. Bruinsma & Johnson 2018) represents a situational dimension of criminality. The spatial-temporal clustering of crime thereby reflects the idea that crime is neither randomly distributed nor evenly distributed in a geographical setting and over a stretch of time (Butt et al., 2020). As such, the literature often alludes to so-called 'crime hot spots' and 'burning times', referring to the idea that certain micro-places are more criminogenic than others, also often pertaining to a specific temporal context to which a vast majority of crime relates. This includes the underlying premises of a relatively recent law in the criminological research field, 'the law of crime concentrations at place', for which Weisburd (2015) has stated that "for a defined measure of crime at a specific micro-geographic unit, the concentration of crime will fall within a narrow bandwidth of percentages for a defined cumulative proportion of crime" (p. 138). The criminogenic effects of specific micro-places and time-bounded properties, and their interactions, should thus not be underestimated in efforts to prevent crime from happening (Cohen & Felson, 1979; Levin et al., 2016; Weisburd, 2015). This provides motives for both policymakers and academics to map and, more recently, even predict the spatial-temporal convergence of crime. Last one is often referred to as place-based predictive policing (Ferguson, 2017).

Hence, - especially driven by the emergence of big data - historical crime data, contextual and socio-economic data are increasingly being subjected to methods aimed at predicting the likelihood of where and when crime is more likely to occur. The main objective is to use these predictions within an intelligence-led policing framework to deploy police resources more efficiently and effectively. The spatial dimension is indispensable in that regard, as it provides us with variables that correlate with crime and could therefore act as a proxy to predict new crime events at small spatial levels (Ratcliffe, 2014, 2016; Hardyns & Rummens, 2017; Uchida, 2014). One of the common methods of place-based predictive policing, risk terrain modelling (Kennedy & Caplan, 2010), focuses solely on the presence and proximity of opportunity characteristics such as bars, schools, and parks. It has been proven to be successful in the prediction of among others gun violence, child maltreatment and robberies (e.g. Daley et al., 2016; Drawve et al., 2016; Connealy & Piza, 2019). However, it is equally important to emphasise the temporal dimension when making predictions about criminal opportunities. Neglecting the temporal dimension would entail resorting to one-sided conceptions of crime, in which both the spatial and temporal dimensions of criminal opportunity are merely considered as distinct components (Grubesic & Mack, 2008). For that reason, other predictive methods situated within the realm of crime prediction, such as machine learning methods, do incorporate both spatial and temporal dimensions to leverage the link between environment and crime (Rummens & Hardyns, 2020; Wheeler & Steenbeek, 2021).

Apart from the growing interest in predictive modelling and its deployment for police governance purposes, predictive policing has until now largely been applied in homogenous settings, in particular large metropolitan (US) cities. Mohler et al. (2015), for example, applied predictive policing in Los

Angeles (ca. 4 000 000 inhabitants), while more recently, Ratcliffe et al. (2021) conducted an experimental study in Philadelphia (ca. 1 500 000 inhabitants). As these settings provide the large amount of crime data needed to train complex predictive models, it makes sense that predictive policing systems have been developed and optimised in these settings, like the PredPol system[1] in Los Angeles and the Crime Anticipation System (CAS) in Amsterdam (ca. 870 000 inhabitants), in the Netherlands. Ever since the introduction of these applications, there has been a growing interest in predictive policing programs, both in the US and in Europe. In the US, one more recent application is HunchLab[2] (Ferguson, 2017, 2020) for which, unlike the PredPol system, different data sources can be employed to make predictions about criminal behaviour (Mugari & Obioha, 2021). Thus, not only spatial-temporal crime data can feed the predictive analysis, but also socio-economic crime data and opportunity characteristics can be included, allowing the use of 'maximal' rather than 'minimal' data. In doing so, HunchLab has mainly addressed the accessibility of predictive policing among police administrations by incorporating the predictive system into police patrols via a mobile app (Ferguson, 2017, 2020; Shapiro, 2017). Furthermore, albeit to a lesser extent, emphasis has been put on predicting potential perpetrators and victims (e.g. Hu et al., 2021; Saunders et al., 2016; Singh & Mohapatra, 2021; Ting et al., 2018; Uchida & Swatt, 2013).

In Europe, however, predictive policing applications are scarcer, yet mainly discernible in Germany, the UK and, as already mentioned, the Netherlands. In Germany, the applications are spread across no less than 16 federal states, for which various predictive systems have been rolled out. Applications such as PreMap and PRECOBS are probably the most coveted in that regard (CCI, 2020; Gerstner, 2018)[3]. In the UK, they are rather moving from a commercial system to a government-oriented system. Commercial applications such as PredPol or software such as Palantir were rather deemed not cost-effective, which is why nowadays the UK government prefers to rely on predictive mapping systems provided by the government itself (Couchman, 2019; Jansen, 2018). In the Netherlands, as mentioned, the CAS system has been rolled out, mainly in Amsterdam. Additionally, in 2015, a pilot study was conducted with the aim of testing the potential of the CAS system considering its deployment across different cities and/or areas. Despite the enthusiasm and opportunities conveyed by predictive policing and its application, the report on this pilot study was not splendid. In fact, no significant improvements were observed in terms of crime reduction across the regions where the application was piloted (Mali et al., 2017). Nevertheless, the various initiatives being pioneered all over the world demonstrate that predictive policing applications have not yet reached their limits.

The unprecedented scale on which predictive policing has evolved, reflects the fact that predictive policing applications are no longer restricted to a specific city. The aforementioned initiatives such as the application of the CAS system across Dutch Cities (Mali et al., 2017) and the implementation of Predpol in different American cities (Ferguson, 2017, 2020; Mugari & Obioha, 2021) are just a few examples of the tendency to apply predictive policing applications in several cities or even whole regions at once. However, in that regard, some critical questions need to be asked and addressed. Since most of the applications have originally been designed for a specific setting or city, it should be considered if these applications can be transferred so readily to other cities or whole regions, whose contextual factors might differ significantly from the original setting or city. In other words, can we assume that predictive policing performance will be consistent across cities, or do city-specific contextual factors require a

[1] PredPol, which can be seen as one of the first applications of predictive policing worldwide, has recently changed its name to Geolitica.

[2] The rights of HunchLab have been transferred to Shotspotter in 2018.

[3] PRECOBS has also been applied in Switzerland.

more case-by-case approach in applying predictive policing? One factor we already know about, is the effect of scale, specifically the effect of high crime density (i.e. large crime counts in a relatively small area). More crime events provide more information for the model to distinguish and learn underlying patterns (Ferguson, 2017; Meijer & Wessels, 2019; Rummens & Hardyns, 2021; Vlahos, 2012). Therefore, one might expect that predictive policing models will perform better in the case of high-frequency crime types such as home burglary, and in urbanized settings, which also tend to have higher crime counts. Yet, as Rummens and Hardyns (2021) have argued, at a certain point, "adding more data will not provide enough new information anymore compared to the data already in the model to justify the increasing efficiency costs in training the model (e.g. the model training process will take considerably longer)" (p. 5). In addition, it is worth noting that earlier studies of predictive policing have largely focused on one setting (i.e. one city or region) at a time. However, a comparative analysis between multiple settings would be very helpful in estimating the effect of contextual factors and to what extent they should be considered when replicating or extending predictive policing to other cities or regions. As such, the aim of this article is to examine the consistency of predictive policing model performance across different urban settings.

To provide the reader with ample transparency, the properties of the respective settings are first and foremost discussed along with the explanatory background of the input variables of the predictive model. A more complex ensemble machine learning model is preferred in this study, where an attempt is made to increase the value and quality of the predictions using multiple input variables. Subsequently, theoretical clarification is provided for each of the indicators according to which the performance of the predictive model is interpreted in this study. We hence provide an integrated overview of both the included predictor variables and the theoretical background to which the predictors pertain. In a next section, the results of this comparative study are presented, primarily by setting. Then, we invert our reasoning by interpreting the results more visually per crime type, to provide a comparative perspective between the settings for each crime type. Finally, the results of this research are interpreted relative to contemporary challenges and future scientific research.

Methodology

Settings and variables used in the analysis

Crime data and supporting data (demographic, socio-economic and proximity variables) were collected from official police and municipal sources for three urban settings (A, B and C)[4] in Belgium, spanning the period from 2012 to 2016. Table 1 presents some basic statistics related to the different settings. Settings A and B are two of the largest cities in Belgium, both consist of one local police department[5]. Setting C is an urban region, including another major city of Belgium and the surrounding urbanised areas. It consists of multiple police departments. The crime data were geocoded and aggregated to a 200x200m grid overlaying the study areas. The grids of settings A and B, respectively, contain 5403 and 4254 grid cells in total. The grid over setting C contains 4284 grid cells in total. Thus, the settings are quite equivalent in terms of grid cells. Some of the supporting data were not available at the grid level and were collected at the statistical sector level instead. This is the smallest unit of analysis for which most statistical data is collected in Belgium, comparable with census tracts in other countries.

Table 1: basic statistics for the three settings

	Setting A	**Setting B**	**Setting C**
Population density (2016)	2529/km²	1655/km²	7378/km²
Area	205 km²	156 km²	161 km²
Number of police departments	1	1	6
Number of municipalities	1	14	19
Number of statistical sectors	300	202	724
Number of grids	5403	4258	4284
Total home burglary crimes (2016)	3031	1697	7512
Total battery crimes (2016)	2063	1151	4589
Total aggressive theft crimes (2016)	1241	466	3543

[4] Upon request of the cities involved, we do not mention the names of the cities for which the analyses are conducted. This does not affect the relevance of the results in any way.

[5] In Belgium, there is a system of an integrated police, structured on two levels. The federal police with federal competences across the entire territory. Its tasks consist mainly of providing specialised functions and support to local police forces where necessary. Secondly, there is the local police organised per local police departments responsible for the basic police tasks. Today, Belgium contains 185 local police departments.

The crime data for home burglary[6], battery[7] and aggressive theft[8] were collected from the local police departments for settings A and B, and from the federal police for setting C[9]. These specific crime types were chosen to represent both property and violent crime. The criminal events were aggregated to the grid cells and temporal distributions using XY-coordinates. Furthermore, some additional variables were included in the predictive model.[10]

First, **crime history variables** were included in the predictive model, which mainly pertain to the temporal dimension for which the crime events are linked to a specific spatial distribution (grid cells). As such, the predictive model is able to discern patterns in the data, varying for different temporal conditions. Instead of only considering the number of criminal incidents in the past month or year, the predictive model also incorporates for example data varying according to when the last criminal incident occurred. As a result, multiple temporal parameters are available concerning the crime events involved, which thus provides more variation and more opportunities for the model to learn definite temporal patterns of the criminal events relative to a specific spatial distribution (McCue, 2014).

Furthermore, demographic and socio-economic variables were collected through the official regional open data repositories and mainly pertain to long-term structural characteristics. The **demographic variables** included in the analysis are *the population rate, the percentage of young people between 15 and 24 years of age, the percentage of immigrants, the percentage of single households and the migration intensity rate*. The **socio-economic variables** included pertain to, amongst others, *unemployment rate, homeownership, number of multi-family homes* and so on. From both a theoretical and an empirical point of view, both groups of predictor variables are often perceived as indicators of vulnerability or mentioned as signs of social (dis)organisation and social cohesion and are thus often associated with an increased or decreased risk of offending and/or victimisation (Sampson, 2009; Sampson et al., 1997; Shaw & McKay, 1969). However, these factors should not be directly associated with criminal behaviour. Rather, they represent risk factors associated with broader structural mechanisms that appear in relation to the environment and correlate to social or situational dimensions of crime. For example, socio-demographic and socio-economic variables are often associated with decreased levels of collective efficacy, involving decreased levels of informal social control and (mutual) social trust, which may thus again lead to increased levels of crime.

Additionally, **environmental** and **proximity variables** were collected via Open Street Map. Both the environmental and the proximity variables commonly refer to opportunity characteristics in committing criminal offences relative to for example the environmental design of the setting (Menting, 2018; Menting et al., 2020; van Sleeuwen et al., 2021). It enhances the predictive model's understanding of situations in which the criminal opportunities are not equally distributed across certain settings. This mainly supports the idea of crime generators and crime attractors. Crime generators are often depicted as venues where people meet, and consequently potential perpetrators and targets converge in time and space, often without specific intent of the perpetrator to engage in criminal behaviour. Crime attractors, in that sense, rather allude to the incitement of a criminal opportunity as such, to which perpetrators deliberately gravitate (Brantingham & Brantingham, 1984, 1995; Kinney et al., 2008; Menting, 2018; Menting et al., 2020). Environmental variables such as *the presence of shops, restaurants and cafés* in

[6] Home burglary is defined as: theft from a residential building while entering illegally, including attempts.

[7] Battery is defined as: the intentional use of force or violence resulting in injuries; intra-familial violence is excluded.

[8] Aggressive theft is defined as: purse snatching or robbery using a weapon or threats, including attempts.

[9] The federal police collect data from the local police departments, making it the most important data source in Belgium for regional data.

[10] A full overview of all the variables included in our model can be found in appendix 1.

a setting sometimes exhibit characteristics of both generators and attractors of criminal opportunities. These encompass places where many people congregate, creating the possibility for potential offenders and targets to encounter each other, as well as the presence of contextual factors that increase the likelihood of a criminal event to occur, such as the presence of alcohol. Likewise, variables such as *the number of areas with green spaces* and *the number of vacant buildings* can be considered as features of criminal opportunity that exhibit their effect via perceptions of (social) disorganisation. Deserted and poorly maintained areas indicate signs of decay, thus suggesting that perpetrators may acquire the perception that the given setting is abandoned. This entails fewer eyes on the street and less ability to identify outsiders (Cozens, 1999; Jacobs, 1993; Quinn, 2019). Moreover, the inclusion of a variable that indicates whether the given setting is in a built-up area or not implies the incorporation of another opportunity variable in the model, as this can provide an indication of the presence of particular targets such as cars (Copes, 1999; Farrell et al., 1995). In addition, a street connectivity score is included, which is not collected via Open Street Map but is based on the walkability tool of the Flemish government (VIGL, n.d.). Better-connected streets can provide enhanced opportunities for offenders to successfully navigate through a setting (Brantingham & Brantingham, 1993; Foster et al., 2014; Frank et al., 2003), which in combination with the 'proximity variables' could result in more opportunities to escape apprehension by law enforcement.

Finally, the predictive model includes **(6) a so-called 'seasonal indicator',** which is a mere 'non-traditional data source'. The latter takes into account the effects of seasonal conditions on criminal behaviour, since both normal conditions such as rain or snow and more drastic conditions such as storms, might equally affect individuals' routine activities and their access to potential targets (McCue, 2014).

Despite the potential benefits of adding predictor variables, it remains important to consider the potential side effects of this choice. Including more variables in the predictive model may also lead to more noise in the trained data set (McCue, 2014). Therefore, it is important to consider both issues of data quality and include an ethical appraisal in the training process.

Statistical model and performance measures

Using the crime and supporting data, a predictive policing model is trained to make predictions for each of the three settings. For this purpose, we employ a machine learning method by using an ensemble of neural networks (ensemble modelling). In general, neural networks can be considered as prominent methods used to handle the complexity of big data sets since the statistical power emanating from such networks is of very large venue for making predictions. However, ensemble-modelling is a more useful technique when one is envisaging the analysis of more complex problems/phenomena, such as criminal events (Rummens & Hardyns, 2020). This method implies combining multiple neural networks in order to maintain sufficient statistical power and to generate better results in relation to the complexity of the problem, whereby the strengths of the respective models outweigh one another's weaknesses (Rummens et al., 2017; McCue, 2014). As such, ensembling modelling provides us with a machine learning method in which the prediction performance of the model ought to be better than when using the distinct types of neural networks independently (Rummens & Hardyns, 2020; Zhou, 2012). More specifically, we used an averaged neural network, which means that the model scores are averaged before being transformed into prediction classes. Model averaging is a type of ensemble modelling approach for reducing the variance error in the final neural network model. When leveraging a more 'conventional' neural network model, the overall model would produce different predictions every time the same model configuration is trained on the same dataset. Model averaging attempts to solve this challenge by merging the findings of both classification and logistic regression machine learning algorithms in a single forecast (Brownlee, 2020; Ellman et al., 2019; Naftaly et al., 1997). The function we employed has one hidden layer, and the variance in the dataset is reduced via bootstrap aggregation, i.e. bagging.

In the present study, predictions are made for three test months: January, May and September 2016.[11] This was done to capture variations between different months and to account for possible seasonal differences. The comparative analysis is done using the retrospective analysis procedure: only the data leading up to each month were used to train the models and optimize them, after which the data of the test months were used to test and evaluate the prediction performance. The number of predictions made for each month was determined in the same way for each setting: a fixed number of predicted risk locations is determined based on the expected frequency of the crime type in question. The resulting predictions for each of the settings are compared against each other by calculating their prediction performance based on the actual locations of the crime events for each test month. To evaluate the performance of the predictive model, four measures are used: **(1) the direct hit rate, (2) the near hit rate, (3) a precision indicator and (4) the F1-score**. Using several measures allows to have a more complete picture of the prediction performance, mitigating disadvantages of using single performance measures.

The **direct hit rate or recall** shows the proportion of correctly predicted crime events relative to the total number of actual crime events. Consequently, **the near hit rate** is reported, which is a less strict alternative for the direct hit rate, where crime events adjacent to the predicted risk location are also included into the calculation as correct predictions. These are also often referred to as measures of sensitivity (Rummens et al., 2017; Rummens & Hardyns, 2020, 2021). **Precision** is the proportion of correct predictions relative to the total number of predictions. In other words, precision is a measure of efficiency, reflecting how many 'attempts' were needed to obtain a certain number of correct predictions. A model would perform good to very good when all indicators display high scores. Obviously, this would represent an ideal situation. However, a balance needs to be struck as these first three indicators are dependent on the number of risk locations predicted, in the sense that direct hit and near hit rates will generally be higher the more risk locations are predicted, while the precisions will be lower and vice versa for a lower number of predicted risk locations. **The F1 score** is often used to express the balance between hit rates and precision, as it is the harmonic average of the direct hit rate and precision. The F1-score is a number between 0 and 1; the higher, the better the balance between the hit rate and the precision scores.[12]

An inherent difficulty is associated with these indicators, as no universal threshold exists with regard to defining what are good or worse values for the indicators. The threshold often bears a relation to the phenomenon for which predictions are made. Some allowance must thus be made considering the context and realm to which the predictions relate. As such, setting specific thresholds concerns a process that is often entrusted to the researchers involved, for whom the intrinsic motivation is often related to the content of the predictive model and the dependent variable. This implies that imposing thresholds to the values of the performance indicators of a predictive model will differ when forecasting criminal activity than when forecasting for example the probability of an individual to be ill or not (McCue, 2014). Yet, irrespective of the performance of a prediction model, there will always be a general concern of producing both false positives and/or false negatives (Kuhn & Johnson, 2013). This problem pertains to a mere modern statistical issue, for which it has been noted that researchers, and analysts in general, should take into account the purposes of the outcome variables when setting thresholds with regard to

[11] For the modelling process, the caret package was used in R.

[12] Higher F1 scores will be obtained if both the precision and recall are high. Lower F1 scores will be obtained if both measurements are low. It is true that the F1 score is very sensitive towards disparate values for both precision and recall, but this is always a balanced and harmonic mean, not a conventional mean.

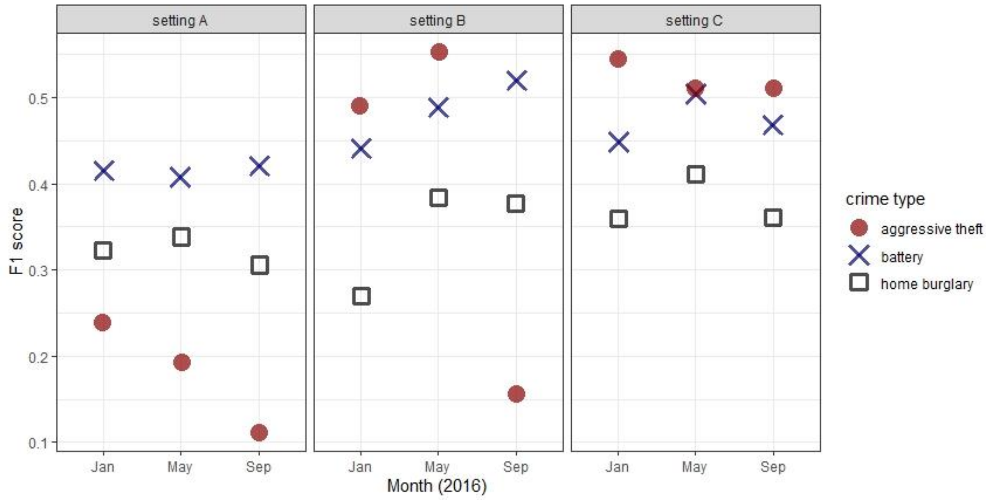
the accuracy and performance of the predictive model, which generally involves a trade-off between the costs of false negatives versus the costs of false positives (Saunders et al., 2016).

Results

General prediction performance comparison between settings

First, we will look at the results obtained for the F1-scores, as this provides the most general and balanced indication of prediction performance. Figure 1 summarises the F1 scores for each setting according to crime type for each of the three months. Low F1 scores indicate a poor balance between hit rates and precision scores, while conversely, high F1 scores suggest a better balance between hit rates and precision scores, thus indicating a better prediction performance of the model in that regard (Manning et al., 2018; Rummens & Hardyns, 2020).

Figure 1. Predictive model performance according to the F1 score for each setting



For setting A, we see a clear separation between the three different crime types, with battery scoring the highest, followed by home burglary and aggressive theft scoring the lowest. Thus, for setting A, a better balance between the hit rate and the precision indicator is obtained for battery, followed by home burglary and aggressive theft. For aggressive theft, we clearly observe decreasing F1 scores every month. In the other two settings, however, aggressive theft generally scores higher (September in setting B being the exception), despite aggressive theft displaying the lowest rates in each setting. Setting A scores lower in general for all three crime types, even though it is the setting with the second highest crime counts. Both setting A and B have a very low score for aggressive theft in September, but this is not the case for setting C, where aggressive theft scores the highest of all three crime types in September. The F1 scores for both home burglary and battery are more consistent across settings and months, except for home burglary in January in setting B, even though January tends to be a month with a high number of home burglaries. While setting A (for aggressive theft only) and especially B (January for home burglary and September for aggressive theft) show some clear differences between months, the scores for setting C, the setting with the highest crime counts, are more consistent for all three crime types. In the following sections, the prediction performance of the predictive model is discussed in more detail for each setting. The resulting prediction parameter scores for each setting are shown in tables 3, 4 and 5.

Setting A prediction performance

Table 1: Prediction performance scores for setting A

	Home burglary	Battery	Aggressive theft
January 2016			
Direct hit rate	34,02%	51,10%	25,00%
Near hit rate	63,23%	68,13%	36,54%
Precision	29,20%	32,00%	22,00%
F1-score	0,32	0,42	0,24
May 2016			
Direct hit rate	38,85%	47,87%	23,26%
Near hit rate	65,38%	71,09%	30,23%
Precision	27,20%	34,50%	16,00%
F1-score	0,33	0,41	0,20
September 2016			
Direct hit rate	33,48%	50,26%	13,79%
Near hit rate	70,59%	70,26%	48,28%
Precision	25,60%	33,00%	8,00%
F1-score	0,30	0,42	0,11

In general, the predictive model performs relatively well for both home burglary and battery in setting A (see Table 3), with consistent scores for all three months. The direct hit rates exceed 30% for both home burglary and battery in every month, while the near hit rates pertain around 65-70% for the same types of crime. The prediction performance scores for aggressive theft on the other hand, are much lower, especially for September. However, the near hit rate for aggressive theft in September is the highest of all three months; 48,28% compared to 36,54% and 30,23%, indicating that the model did not predict the exact locations very well, but was able to predict the more general neighbourhood of the actual events. The enhancement in the performance of the model for both home burglary and battery relative to aggressive theft might be explained by more clustered crime patterns being more stable over time. This is equally presented in the precision scores, which are generally lower for aggressive theft. In general, between 8% and 34,5% of the predictions made by the model for setting A were correct.

Setting B prediction performance

Table 2: Prediction performance scores for setting B

	Home burglary	**Battery**	**Aggressive theft**
January 2016			
Direct hit rate	36,70%	75,74%	77,78%
Near hit rate	60,55%	87,77%	77,78%
Precision	16,57%	15,45%	20,00%
F1-score	0,27	0,45	0,49
May 2016			
Direct hit rate	58,75%	84,27%	54,32%
Near hit rate	70,00%	94,38%	62,53%
Precision	17,14%	14,09%	10,00%
F1-score	0,38	0,49	0,55
September 2016			
Direct hit rate	60,94%	84,62%	20,00%
Near hit rate	68,75%	96,15%	40,00%
Precision	13,14%	17,73%	10,00%
F1-score	0,37	0,51	0,15

Overall, setting B (see Table 4) scores good hit rates, with relatively worse precision scores. The precision scores are never higher than 20%, which might indicate that the model is overpredicting, i.e. predicting a much larger number of risk zones relative to the actual number of crime events (Kuhn & Johnson, 2013). The direct hit rates are much higher compared to setting A and setting C, except for the prediction of home burglary in the month of January. Especially for battery, the direct and near hit rates are very good, which is indicative for a larger proportion of correct predictions compared to the total number of actual criminal offences in the same month, thus also considering the number of correct predictions in adjacent prediction locations. Equally, the F1-scores demonstrate that the predictive model performs comparatively well in setting B, not hindered by it being the smallest-scale setting of the three, with the lowest crime counts. The model performs bad for home burglary in January and for aggressive theft in September. Especially the lower score for aggressive theft in September is striking, as in the other two months, the prediction performance was best for aggressive theft compared to the other two crime types.

Setting C prediction performance

Table 3: Prediction performance scores for setting C

	Home burglary	Battery	Aggressive theft
January 2016			
Direct hit rate	41,63%	52,97%	68,89%
Near hit rate	83,46%	78,31%	88,19%
Precision	28,62%	35,11%	40,00%
F1-score	0,35	0,44	0,54
May 2016			
Direct hit rate	47,79%	57,67%	60,61%
Near hit rate	86,19%	82,21%	84,42%
Precision	34,83%	44,44%	41,73%
F1-score	0,41	0,51	0,51
September 2016			
Direct hit rate	41,92%	55,70%	65,57%
Near hit rate	80,94%	79,39%	85,86%
Precision	30,51%	37,78%	37,31%
F1-score	0,36	0,47	0,51

Finally, the predictive performance for setting C (see Table 5) shows the most consistency across the three months and the different performance parameters. The direct hit rates tend to be higher for aggressive theft, with medium direct hit rates for battery and the lowest percentages of direct hit rates displayed for home burglary. Despite the differences in performance between the crime types, there are no outlying values, as was the case for setting A and B. This is also reflected by the F1 scores, which are equally distributed across crime types. There are no signs of overpredicting. In general, prediction performance is considered best for setting C.

Comparing the results between settings per crime type

In conjunction with the presented results for each setting, we aim to visually present the inter-setting differences regarding the performance indicators of the predictive model for each of the crime types. Hence, so-called 'radar charts' were constructed per crime type, visualising the values of the performance indicators for each setting according to the considered prediction month. Higher values on the indicators are reflected by a closer positioning of the lines at the outside of the chart. Thus, the more central the values are, the lower the values are for the corresponding performance indicator. The indicators are presented in chronological order of the respective months (January, May and September).

In general, as illustrated by figure 2, the performance pattern for home burglary can be considered stable for every setting. However, the performance indicators for the predictions of home burglary are relatively better for setting C compared to setting A and B. Especially the near hit rates are higher for setting C, which is, as stated, indicative for a higher number of correct predictions relative to the total number of criminal incidents, for which incidents in the adjacent areas are also included as correct predictions. Except for the direct hit rates in September, all indicators' values for setting C are higher or equal to the values of the indicators in the other settings. In addition, for setting B, a combination of relatively high hit rates and low precision is observed, which is again suggestive for the fact that the model is overpredicting. Setting A scores average, often better than setting B, yet often worse than setting C.

Figure 2. Comparison of the performance indicators between settings for home burglary

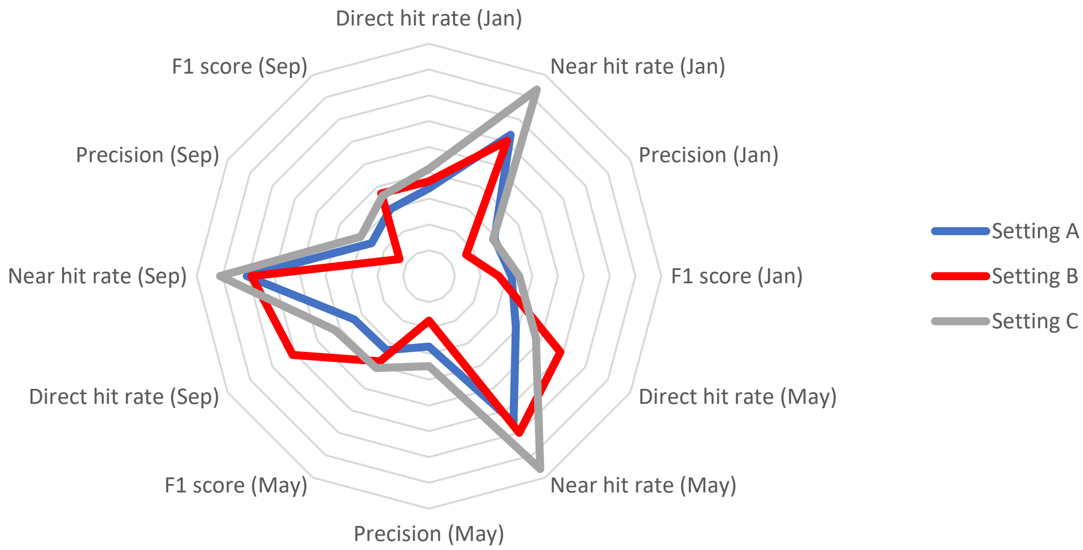
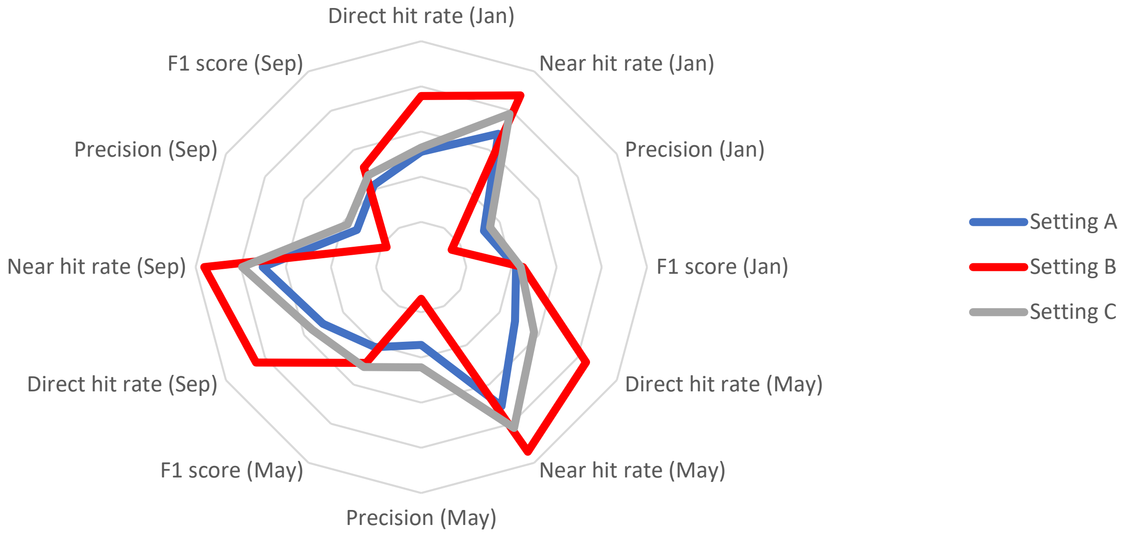


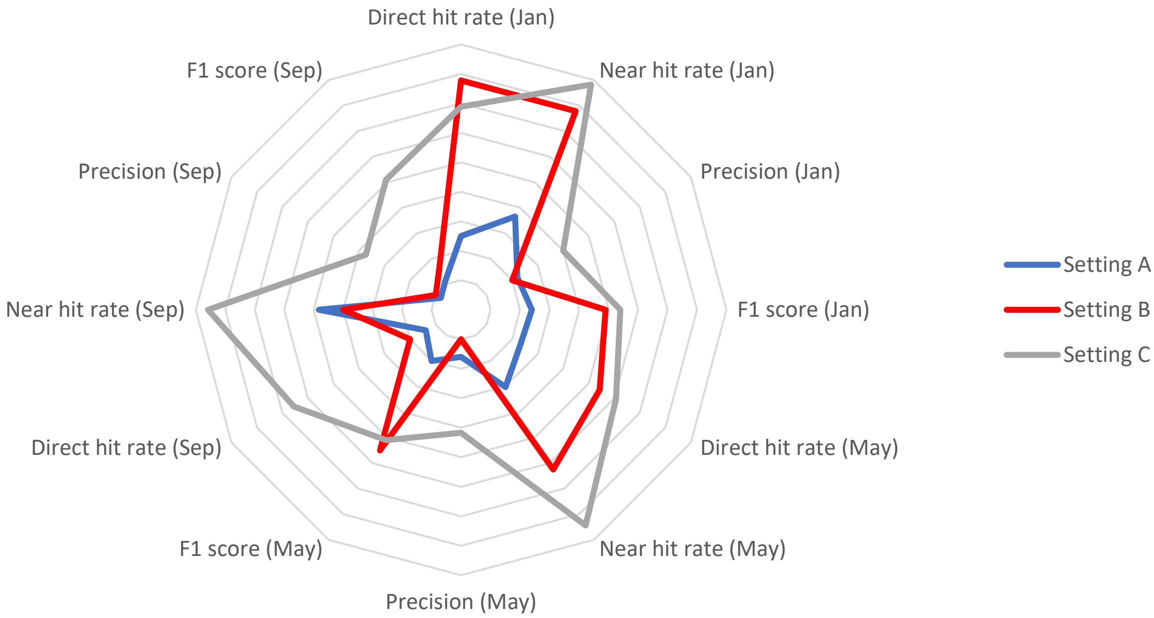
Figure 3 shows a yet more 'erratic' performance pattern for battery but which is still relatively stable for each of the three settings. The performance indicators for the predictions of battery take relatively higher values and are thus considered to indicate a better prediction performance of the model for setting B (red line) compared to setting A and setting C. Nevertheless, it is noticeable that the precision indicator is lower for setting B, in every month, and that the F1 score is similar between the different settings. Again, this reflects the idea that the combination of good hit scores and low precision values could be indicative for overpredicted results for setting B.

Figure 3. Comparison of the performance indicators between settings for battery



Finally, figure 4 presents the performance indicators for aggressive theft for the three settings. The pattern differs considerably across the three settings. As for home burglary, higher values are found for the performance indicators for setting C compared to settings A and B. In particular, the near hit rates are high in all three months for setting C, as well as for setting B in January and May. In September, the model shows low hit rates for the predictions in setting B, indicating less correct predictions relative to the number of criminal incidents compared to the other months. However, once again, relatively higher hit rates for setting B are accompanied by low precision scores, which in turn may indicate that the model is overpredicting for this specific setting. Noteworthy are the results for setting A, which indicate relatively poor performance of the predictive model for aggressive theft. The values for all indicators are very low.

Figure 4. Comparison of the performance indicators between settings for aggressive theft



Discussion and conclusions

As stated in the introductory part, scale is an important factor to consider when applying predictive policing across various settings, as more events equal more information for the predictive model (Rummens et al., 2017; Rummens & Hardyns, 2021; McCue, 2014). This is in line with the findings of this study, which suggest that predictive modelling performance will be more consistent in settings with higher crime counts relative to their area, as prediction performance will exhibit fewer extreme fluctuations. However, it is not always true that settings with lower crime counts cannot benefit from predictive policing. If the frequency of a specific crime type is too low, it can be compensated for by aggregating the data to a higher spatial or temporal level, or by grouping related crime types in a larger group (Rummens & Hardyns, 2021). Not only may the higher scale enhance prediction performance, it may also aid in acquisition of cross-border crimes which do not take into account the borders between different police departments. However, despite these considerations, the findings show that scale does not provide the complete picture and that significant differences between settings can exist that go beyond scale and frequency, that are associated to the crime type and/or a combination of contextual factors. This is a convincing argument for adapting the predictive policing approach to the specific context (Gerstner, 2018), which may be achieved by undertaking a comprehensive pre-implementation contextual analysis and tailoring the predictive model to the findings.

The latter was equally illustrated by the so-called ‘radar charts’ in the preceding section, demonstrating that differences between the various settings are evident as regards the performance of the predictive model for the different crime types. In general, different performance patterns were observed for each crime type when compared across settings. For home burglary, performance patterns were relatively stable, however for battery, performance patterns were yet more ‘erratic’ and eventually for aggressive theft, the performance patterns were fluctuating and varying a lot across the three settings. This could be suggestive for the fact that different performance patterns are obtained when predictive modelling is applied for different crime types. Additionally, better performance was reached for the relative ‘larger’ urban setting (C) compared to a slight smaller urban setting (A) (setting B not included because of the signs of over-prediction of the results). This could equally indicate that the setting size of the urban region/city itself also has a bearing on the prediction performance of the model. The degree of urbanisation might thus also be a relevant factor contributing to different levels of prediction performance across settings. Moreover, population density might also be a relevant factor that contributes to the prediction performance of the model. Our model seems to perform better in setting C, which has a higher overall population density. This would be in line with the findings of the study of Kadar et al. (2019), which is an important prior to this study. Comparing prediction performance across different ML techniques and population densities, they found that a hyper-ensemble Machine Learning model predicts crime more accurately for cities with a higher population density. In that regard, it was concluded that areas with a higher population density generally have higher crime rates, implying that the algorithm can be fed with more positive cases and because such areas typically “coincide with a wider distribution of the features, which allows the algorithm to discover more discriminative patterns” (Kadar et al., 2019, p. 115). However, when the number of crimes per crime type is divided by the population density (crime rates), it can be observed that the crime rates of setting B and C are almost equal for battery and home burglary, while setting A generally has higher crime rates (appendix 2). As such, we do not assume that population density nor the density of crimes sufficiently explain the performance of the prediction model. Prediction performance was also generally better for setting C, with lower crime rates, than for setting A. The fluctuation and distribution of crime incidences could also have led to differential class-imbalance problems across the three settings in each prediction month, which in turn may have affected the performance of the model, i.e. the model's ability to recognise 'discriminative patterns' in the test dataset, based on the patterns discovered from the training dataset. We could thus also assume that the distribution of historical crime features differs significantly from setting to setting, which would also be in line with the assumptions and findings of the study of Kadar et al. (2019), who argued that the importance of historical crime features increases when population density increases. Nonetheless, based on the F1 scores, which is generally a better measure to present when facing class-imbalance problems, we can still conclude that the predictive performance of the model varies a lot from month to month, taking into account the differential effects across the three settings. One could therefore also argue that the spatial stability of crime varies a lot across the three settings, which makes it again difficult for the model to recognise these patterns. After all, if a specific proportion of micro-places is responsible for a large proportion of crime for several years, this does not immediately imply that the same proportion of micro-places will be responsible for a large proportion of crime the following years (Andresen et al., 2017). This could thus also suggest that more structural factors – other than the variation in crime frequencies or crime densities – may partially explain the monthly variation in F1 scores across the three settings. One might for example contend that the positive results in setting C are due to differences or changes in the geographical composition of the city, which is again associated with the crime patterns discovered by the model in the dataset. For example, as setting C is the largest of the three settings, it could be assumed that the positive results in setting C can be partially explained by a higher concentration of opportunity characteristics at the grid-level, which again may vary per crime type. Yet, these conclusions cannot be made so easily and one-on-one, as machine learning models have a so-called 'black box structure', making it difficult to reveal the throughput of the neural network, i.e. linking the input with the output of the model (Rummens et al., 2017).

Despite the contributions of this study to the limited field of research pertaining to predictive policing and its application across settings, some limitations should be taken into account in future studies. As has been clearly stated, this study was limited to the urban context so that mere suburban or rural areas were neglected to a certain extent. However, this mainly bears to the fact that only data at the urban level were disposable. In addition, this study was mainly focusing on the Belgian context, so that in the future, studies should not only aspire to mere cross-regional applications but should equally study and apply predictive policing from more cross-national and international perspectives. Moreover, it could be seen as a limitation that we only devoted attention to 2016 as prediction year and that the predictions only pertained to three specific non-consecutive months in that year. Hence, research should pay attention to predict for mere consecutive months and for a longer period. In addition, research should also aim to predict crime across different settings, taking into account more granular temporal resolutions such as weeks or days, to see how the prediction performance varies across contexts. The approach that was adopted in this study significantly varies from the more commercial systems such as PredPol or HunchLab, which aim to make spatiotemporal predictions of crime at a time granularity of days or even hours. However, as Rummens and Hardyns (2021) have shown, in European police zones, monthly predictions of crime can be more reliable, although this should be balanced against the spatial resolution. Moreover, as we only employed one specific method group (ensemble modelling), future research should explore different forecasting techniques in predicting crime across different contexts such as hyper-ensemble models, which can be used to account for imbalance distributions of crime events or other features. Furthermore, research should also employ new developing forms of big data, such as data from new technologies like CCTV and track and trace data from police patrols. These can be incorporated in a predictive policing model. Studies should also elaborate upon the problem of the dark figure of crime in crime prediction and how other non-police data sources, such as self-reported data or data from private security companies, might supplement and improve the quality of official police statistics. Although we did not particularly include an ethical assessment in the development of our prediction model, this should be a goal in future studies. For example, research might employ particular frameworks to examine all potential errors and biases that may arise during the data collection or training process.

Nevertheless, future research should mainly envisage the extent to which predictive policing is applicable in more isolated rural settings compared to highly urban areas. The spatiotemporal distribution of crime patterns may appear quite different, with distinct spatiotemporal dynamics at work or exerting effects in a different way. In line with Weisburd and Telep (2014) and Santos and Santos (2015), it is not entirely implausible to assert that urban areas are more likely to be exposed to long-term clusters of crime, while more suburban or rural areas are more likely to be exposed to short-term clusters of crime experiencing crime 'flare ups' rather than stable crime patterns. A such, the distribution of crime might be significantly different in rural areas, which in turn might impact the performance of a model predicting spatiotemporal patterns of criminality and its application in a rural context. Dynamic hot spots of crime, compared to more static hot spots, may also require different police strategies, or at least the need to 'cool' the hot spots in a somewhat different way. For example, as Sherman (1990) argued, police patrols in so-called short-term hot spots should be short, rotated by target and rely on unpredictability. Of course, scale issues arise once again, as the hot spots policing literature provides little evidence for rural areas being equally exposed to high crime concentrations at small geographical levels relative to urban areas. Moreover, research is indicative for finding less 'hot' hot spots in small cities and rural areas compared to large urban cities (Santos & Santos, 2015; Weisburd & Telep, 2014). This could point to the fact that smaller (rural) areas with a lower population density are less or differently exposed to crime. It is thus inadvisable to take the transitivity of predictive policing models for granted. Carefully planning and investigating its application might initially prove to be time-consuming and costly, in the long run it will prove to be much more beneficial in terms of potential benefits. Future research should thus focus on more diverse settings and aggregate the obtained results (e.g. through a meta-analysis) to gain more insight into how the predictive modelling of crime can be replicated and extended to various settings. In addition to research that should focus on applying predictive crime models across different contexts, it is equally important that future research gives due consideration to the complexity of the statistical modelling procedure, as we repeatedly encountered the impression that the model was over-predicting for one setting. After all, measuring a model's efficacy entails more than simply determining its overall accuracy. It is hence necessary to ascertain the nature of the deficiencies when predicting crime across settings and evaluate which deficiencies are acceptable and which are not (McCue, 2014).

Furthermore, several other reasons should be considered too when one is advocating to tailor the predictive policing approach to the specific context. First and foremost, some contexts may contain interesting data that is not obtainable throughout all settings, which especially relates to the availability of data. The higher the level of hierarchy, the more likely the data will be heterogeneous. A general model, on the other hand, would be unable to use this data because it would be lacking for some settings. In addition, predictive policing applications generally require 'big data', as their premises rely on mere complex statistical models and procedures as demonstrated in this study. However, as Ferguson (2017) puts it: "…some small towns do not have adequate data to build a strong model". It might thus be reasonable to assume that police departments in smaller geographical settings bear less possibilities to both extract big data and employ big data techniques and to ensure data quality, relative to police departments in larger geographical settings. In addition, if myriad settings are combined, the data for some variables may be collected or calculated differently, introducing bias into the predictions, something that should be avoided. Data quality is hence another reason to assert that big data and the application of techniques in function of big data are not always straightforward when applied across different settings.

A tailored approach therefore allows for the predictive model to be adjusted to the specific aspirations of the police department in question (e.g. local priorities). While priorities may overlap across zones, it may be beneficial in some cases to tailor the predictive policing approach to address key challenges within a police department. It may also be beneficial to investigate alternative methods and techniques, as this may be related to context specifics. For example, while machine learning and hotspot analysis performed better than risk terrain modelling when compared (Rummens & Hardyns, 2020), Ohyama & Amemiya (2018) discovered the opposite in a study conducted in Tokyo, Japan: Risk Terrain Modelling outperformed machine learning and hot spot analysis. A more hybrid approach combining both the general multi-setting and the more specific single setting approaches could be a viable solution. While more rural areas may benefit from a more general approach led by the federal police, major cities may benefit from a(n) (additional) tailor-made predictive model. In doing so, it is equally pertinent to bear in mind policing practice. The principles on which predictive policing draws, lend themselves extensively to the realm of operational policing tactics applicable within the field, proactive foot and car patrols in particular (Ratcliffe, 2016). Nevertheless, given the differences between larger and smaller settings, both in terms of spatial-temporal crime dynamics and contextual factors, it is not incongruous to expect differences in the application of certain policing strategies in smaller areas compared to larger areas when supported by the insights of predictive analytics. More evaluation research is thus also needed, applying different predictive models across different settings and different policing strategies, pertaining to both urban and rural settings.

References

Andresen, M. A. (2014). *Environmental criminology: Evolution, theory, and practice*. Routledge.

Andresen, M. A., Linning, S. J., & Malleson, N. (2017). Crime at places and spatial concentrations: Exploring the spatial stability of property crime in Vancouver BC, 2003–2013. *Journal of quantitative criminology*, *33*(2), 255-275.

Brantingham, P. J., & Brantingham, P. L. (Eds.). (1981). *Environmental criminology*. Sage Publications.

Brantingham, P. J., & Brantingham, P. L. (1984). *Patterns in crime*. Macmillan.

Brantingham, P. J., and Brantingham, P.L. (1991). 'Introduction: The Dimensions of Crime'. In P. Brantingham and P. Brantingham (eds), *Environmental Criminology* (2nd ed.). (pp. 7-26). Waveland Press

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3-28.

Brantingham, P. L., & Brantingham, P. J. (1995). Criminality of place. *European journal on criminal policy and research*, *3*(3), 5-26.

Brownlee, J. (2020, August 28). *How to Develop an Ensemble of Deep Learning Models in Keras*. Machine Learning Mastery. Retrieved 21 June 2022, from https://machinelearningmastery.com/model-averaging-ensemble-for-deep-learning-neural-networks/

Bruinsma, G. J. N., & Johnson, S. D. (2018). Environmental criminology: Scope, history, and state of the art. In G. Bruinsma, & S. D. Johnson (Eds.), *The Oxford Handbook of Environmental Criminology* (pp. 1-31). Oxford University Press.

Butt, U. M., Letchmunan, S., Hassan, F. H., Ali, M., Baqir, A., & Sherazi, H. H. R. (2020). Spatio-temporal crime HotSpot detection and prediction: a systematic literature review. *IEEE Access*, *8*, 166553-166574.

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American sociological review*, 588-608.

Connealy, N. T., & Piza, E. L. (2019). Risk factor and high-risk place variations across different robbery targets in Denver, Colorado. *Journal of criminal justice*, 60, 47-56.

Copes, H. (1999). Routine activities and motor vehicle theft: A crime specific approach. *Journal of Crime and Justice*, *22*(2), 125-146.

Cozens, P., & Love, T. (2015). A review and current status of crime prevention through environmental design (CPTED). *Journal of Planning Literature*, *30*(4), 393-412.

Couchman, H. (2019). Policing by machine: predictive policing and the threat to our rights. *Liberty.*

Cutting Crime Impact (CCI). 2020. *Fact Sheet: Predictive Policing.* CCI

Daley, D., Bachmann, M., Bachmann, B. A., Pedigo, C., Bui, M. T., & Coffman, J. (2016). Risk terrain modeling predicts child maltreatment. *Child abuse & neglect*, *62*, 29-38.

Drawve, G., Moak, S. C., & Berthelot, E. R. (2016). Predictability of gun crimes: A comparison of hot spot and risk terrain modelling techniques. *Policing and Society*, *26*(3), 312-331.

Ellmann, S., Seyler, L., Evers, J., Heinen, H., Bozec, A., Prante, O., ... & Baeuerle, T. (2019). Prediction of early metastatic disease in experimental breast cancer bone metastasis by combining PET/CT and MRI parameters to a Model-Averaged Neural Network. *Bone*, *120*, 254-261.

Farrell, G., Phillips, C., & Pease, K. (1995). Like taking candy-why does repeat victimization occur. *British Journal for Criminology*, *35*, 384.

Ferguson, A. G. (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York University Press.

Ferguson, A. G. (2020). Predictive Policing Theory. In T. R. Lave, & E. J. Miller (eds.), *The Cambridge Handbook of Policing in the United States.* Cambridge University Press.

Foster, S., Knuiman, M., Villanueva, K., Wood, L., Christian, H., & Giles-Corti, B. (2014). Does walkable neighbourhood design influence the association between objective crime and walking? *International Journal of Behavioral Nutrition and Physical Activity*, *11*(1), 1-9.

Frank, L., Engelke, P., & Schmid, T. (2003). *Health and community design: The impact of the built environment on physical activity*. Island Press.

Gerstner, D. (2018). Predictive policing in the context of residential burglary: An empirical illustration on the basis of a pilot project in Baden-Württemberg, Germany. *European Journal for Security Research*, *3*(2), 115-138.

Grubesic, T. H., & Mack, E. A. (2008). Spatio-temporal interaction of urban crime. *Journal of Quantitative Criminology*, *24*(3), 285-306.

Hardyns, W., & Rummens, A. (2017). Predictive policing as a new tool for law enforcement? Recent developments and challenges. *European Journal on Criminal Policy and Research*, 24(3), 201-218. 10.1007/s10610-017-9361-2

Hu, X., Zhang, X., & Lovrich, N. P. (2021). Forecasting identity theft victims: Analyzing characteristics and preventive actions through machine learning approaches. *Victims & Offenders*, *16*(4), 465-494.

Jacobs, J. (1993). *The Death and Life of Great American Cities.* Random House, Inc.

Jansen, F. (2018). Data driven policing in the context of Europe. *Data Justice Lab*.

Kadar, C., Maculan, R., & Feuerriegel, S. (2019). Public decision support for low population density areas: An imbalance-aware hyper-ensemble for spatio-temporal crime prediction. *Decision Support Systems*, *119*, 107-117.

Kennedy, L. & Caplan, J. (2010). *Risk Terrain Modelling Manual: Theoretical framework and technical steps of spatial risk assessment for crime analysis*. Rutgers.

Kinney, J. B., Brantingham, P. L., Wuschke, K., Kirk, M. G., & Brantingham, P. J. (2008). Crime attractors, generators and detractors: Land use and urban crime opportunities. *Built environment*, *34*(1), 62-74.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society*, *1*(1), 1-12.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26, p. 13). Springer.

Levin, A., Rosenfeld, R., & Deckard, M. (2017). The law of crime concentration: An application and recommendations for future research. *Journal of Quantitative Criminology*, *33*(3), 635-647.

Mali, B., Bronkhorst-Giesen, C., & den Hengst, M. (2017). *Predictive policing: lessen voor de toekomst.* Politieacademie Apeldoorn.

Manning, C., Raghavan, P., & Schütze, H. (2008). Evaluation in information retrieval. In: Introduction to Information Retrieval, 139–161. Cambridge University Press.

McCue, C. (2014). *Data mining and predictive analysis: Intelligence gathering and crime analysis*. Butterworth-Heinemann.

Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration*, *42*(12), 1031-1039.

Menting, B. (2018). Awareness × opportunity: Testing interactions between activity nodes and criminal opportunity in predicting crime location choice. *The British Journal of Criminology*, *58*(5), 1171-1192.

Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2020). The influence of activity space and visiting frequency on crime location choice: findings from an online self-report survey. *The British Journal of Criminology*, *60*(2), 303-322.

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing. *Journal of the American statistical association*, *110*(512), 1399-1411.

Mugari, I., & Obioha, E. E. (2021). Predictive Policing and Crime Control in The United States of America and Europe: Trends in a Decade of Research and the Future of Predictive Policing. *Social Sciences*, *10*(6), 234.

Naftaly, U., Intrator, N., & Horn, D. (1997). Optimal ensemble averaging of neural networks. *Network: Computation in Neural Systems*, *8*(3), 283.

Ohyama, Tomoya & Amemiya, Mamoru. (2018). Applying Crime Prediction Techniques to Japan: A Comparison Between Risk Terrain Modeling and Other Methods. *European Journal on Criminal Policy and Research,* 24. 10.1007/s10610-018-9378-1.

Quinn, A. (2019). Vehicle Crime, CPTED, and Offending under the Influence: A Qualitative Investigation of Offender Perceptions. *Social Sciences*, *8*(3), 88.

Ratcliffe, J. (2014). What is the future of predictive policing? *Translational Criminology, 6*, 4–5.

Ratcliffe, J. H. (2016). *Intelligence-led policing*. Routledge.

Ratcliffe, J. H., Taylor, R. B., Askey, A. P., Thomas, K., Grasso, J., Bethel, K. J., Fisher, R. & Koehnlein, J. (2021). The Philadelphia predictive policing experiment. *Journal of Experimental Criminology*, *17*, 15-41.

Rummens, A., & Hardyns, W. (2020). Comparison of near-Repeat, Machine Learning and Risk Terrain Modelling for Making Spatiotemporal Predictions of Crime. *Applied Spatial Analysis and Policy, 13*(4), 1035-1053.

Rummens, A., & Hardyns, W. (2021). The effect of spatiotemporal resolution on predictive policing model performance. *International Journal of Forecasting,* 37(1), 125-133.

Rummens, A., Hardyns, W., & Pauwels, L. (2017). A scoping review of predictive analysis techniques for predicting criminal events. In G. Vermeulen & E. Lievens (Eds.), *Data Protection and Privacy under Pressure* (pp. 253–292). Maklu.

Rummens, A., Hardyns, W., & Pauwels, L. (2017). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. *Applied geography*, *86*, 255-261.

Sampson, R. J. (2009). Disparity and diversity in the contemporary city: social (dis) order revisited 1. *The British journal of sociology*, *60*(1), 1-31.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *science*, *277*(5328), 918-924.

Santos, R. G., & Santos, R. B. (2015). Practice-based research: Ex post facto evaluation of evidence-based police practices implemented in residential burglary micro-time hot spots. *Evaluation Review*, *39*(5), 451-479.

Saunders, J., Hunt, P., & Hollywood, J. S. (2016). Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology*, *12*(3), 347-371.

Shapiro, A. (2017). Reform predictive policing. *Nature*, *541*(7638), 458-460.

Shaw, C. R., & McKay, H. D. (1969). *Juvenile Delinquency and Urban Areas: A Study of Rates of Delinquency in Relation to Differential Characteristics of Local Communities in American Cities (1942)*. University of Chicago Press.

Sherman, L. W. (1990). Police crackdowns: Initial and residual deterrence. *Crime and justice*, *12*, 1-48.

Singh, A., & Mohapatra, S. (2021). Development of risk assessment framework for first time offenders using ensemble learning. *IEEE Access*, *9*, 135024-135033.

Taylor, R. B., & Harrell, A. (1996). *Physical environment and crime*. US Department of Justice, Office of Justice Programs, National Institute of Justice.

Ting, M. H., Chu, C. M., Zeng, G., Li, D., & Chng, G. S. (2018). Predicting recidivism among youth offenders: Augmenting professional judgement with machine learning algorithms. *Journal of Social Work*, *18*(6), 631-649.

Uchida, C. (2014). Predictive policing. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 3871–3880). Springer.

Uchida, C. D., & Swatt, M. L. (2013). Operation LASER and the effectiveness of hotspot patrol: A panel analysis. *Police Quarterly, 16*(3), 287-304

van Sleeuwen, S. E., Steenbeek, W., & Ruiter, S. (2021). When do offenders commit crime? An analysis of temporal consistency in individual offending patterns. *Journal of quantitative criminology*, *37*(4), 863-889.

Vlahos, J. (2012). The department of pre-crime. *Scientific American*, *306*(1), 62-67.

Vlaams Instituut Gezond Leven. (z.d.). *Walkabilityscore-tool | Gezond Leven*. Gezondleven.be. Geraadpleegd op 17 februari 2022, van https://www.gezondleven.be/settings/gezonde-gemeente/gezonde-publieke-ruimte/walkability-tool

Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology*, *53*(2), 133-157.

Weisburd, D., & Telep, C. W. (2014). Hot spots policing: What we know and what we need to know. *Journal of Contemporary Criminal Justice*, *30*(2), 200-220.

Wheeler, A. P., & Steenbeek, W. (2021). Mapping the risk terrain for crime using machine learning. *Journal of Quantitative Criminology*, *37*(2), 445-480.

Wikström, P. O. H., & Treiber, K. (2016). Situational theory: The importance of interactions and action mechanisms in the explanation of crime. *The handbook of criminological theory*, *1*, 415-444.

Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms.* Machine learning and pattern recognition series. CRC Press.

Appendix 1: List of included variables wihtin the predictive model

*Only available at the statistical sector level instead of grid level

(1) Crime history variables
Crime events in the previous month
Time since last crime event (in months)
Crime events in the past year
Crime events in the previous period in the neighbourhood*
Crime events in the same month last year
(2) Demographic variables
Population
Youth (15- to 24-year-olds) (percentage)*
Immigrants (percentage)*
Single households (percentage)*
Migration intensity rate*
(3) Socio-economic variables
Unemployment rate*
Homeownership / renting rate*
Housing stock*
Number of single-family homes*
Number of multi-family homes*
Mean surface area of houses*
(4) Environmental variables
Number of shops
Number of bars/cafés
Number of restaurants
Green space (area)
Number of vacant buildings*
Built-up area *
Street connectivity score
(5) Proximity variables
Distance to nearest train stop (m)
Distance to nearest highway (m)
Distance to nearest bus stop (m)
(6) Indicator variables
Seasonal indicator (winter, summer)

Appendix 2: Crime rates (per population density)

	Setting A	**Setting B**	**Setting C**
Home burglary	1.20	1.03	1.02
Battery	0.82	0.70	0.62
Aggressive theft	0.49	0.28	0.48