

Deep Learning for Crime Forecasting: The Role of Mobility at Fine-grained Spatiotemporal Scales

Ariadna Albors Zumel¹, Michele Tizzoni¹, and Gian Maria Campedelli²

¹*University of Trento*

²*Fondazione Bruno Kessler*

Abstract

Objectives: To develop a deep learning framework to evaluate if and how incorporating micro-level mobility features, alongside historical crime and sociodemographic data, enhances predictive performance in crime forecasting at fine-grained spatial and temporal resolutions.

Methods: We advance the literature on computational methods and crime forecasting by focusing on four U.S. cities (i.e., Baltimore, Chicago, Los Angeles, and Philadelphia). We employ crime incident data obtained from each city's police department, combined with sociodemographic data from the American Community Survey and human mobility data from Advan, collected from 2019 to 2023. This data is aggregated into grids with equally sized cells of 0.077 sq. miles (0.2 sq. kms) and used to train our deep learning forecasting model, a Convolutional Long Short-Term Memory (ConvLSTM) network, which predicts crime occurrences 12 hours ahead using 14-day and 2-day input sequences. We also compare its performance against three baseline models: logistic regression, random forest, and standard LSTM.

Results: Incorporating mobility features improves predictive performance, especially when using shorter input sequences. Noteworthy, however, the best results are obtained when both mobility and sociodemographic features are used together, with our deep learning model achieving the highest recall, precision, and F1 score in all four cities, outperforming alternative methods. With this configuration, longer input sequences enhance predictions for violent crimes, while shorter sequences are more effective for property crimes.

Conclusion: These findings underscore the importance of integrating diverse data sources for spatiotemporal crime forecasting, mobility included. They also highlight the advantages (and limits) of deep learning when dealing with fine-grained spatial and temporal scales.

Keywords: Crime forecasting, Human mobility, Machine Learning, Deep Learning

1 Introduction

The increasing availability of data, coupled with advancements in analytical techniques and the progressive democratization of programming languages, has contributed to facilitate the computational study of criminal phenomena [1]. One of the areas that most benefited from the convergence of these three factors is the spatiotemporal prediction of crime, which is particularly important for theory and policy testing and development, as it can shed light on the behavioral dynamics that explain crime occurrence and provide actionable knowledge that can be leveraged to anticipate and prevent crime [2, 3].

The spatial study of crime has its roots in the first half of the 19th century [4–6] and became extremely popular in the second half of the 20th century. One of the most widely accepted findings to emerge from this area of research is that crime tends to concentrate in small geographical areas. This observation led to the so-called *criminology of place*, first introduced by [7], which focuses on the dynamics of crime in microgeographic units within cities, such as addresses, facilities, street segments, or small clusters of street segments [7, 8]. Since then, the growing interest in these micro-dynamics has led to practical applications in crime prevention, such as hot spots policing [9].

Along with the spatial study of crime, another factor that has long been considered in the empirical and theoretical literature is the temporal dimension of crime [10, 11]. The two dimensions, space and time, are intrinsically interconnected: both components of a crime event are fundamental due to their correlation as time constraints result in space constraints [12]. Given this fundamentally indissoluble link, crime forecasting systems rely on both spatial and temporal information in order to deliver more accurate predictions.

However, beyond the spatial and temporal aspects of crime, many other factors play a role in shaping its concentration and patterned nature. One of the most studied factors is the social context, often examined through dimensions such as socioeconomic inequalities and ethnic diversity. Poverty, for instance, often fosters conditions that increase the likelihood of crime by limiting people's access to opportunities and resources [13–16]. Meanwhile, neighborhoods with significant ethnic diversity sometimes experience crime patterns shaped by the social tensions and discrimination that residents face [17, 18].

Besides these socio-economic dimensions, another central phenomenon impacting the spatiotemporal concentration of crime is human mobility, as suggested by various theories uncovering its link to crime incidents [19, 20]. For example, areas with high foot traffic or frequent visitors often experience more crime due to the increased presence of potential victims and offenders. Conversely, more deserted areas can foster crime due to the absence of individuals who can discourage offenders from committing crimes [19]. Similarly, shifts in population dynamics—such as the influx of new residents or temporary workers—can disrupt established community networks, creating opportunities for criminal behavior [20].

Due to the nonlinear and multidimensional spatiotemporal dynamics that often characterize crime patterns, quantitative criminologists—along with scholars from other fields—

have thus recently become increasingly interested in the promising properties of machine and deep learning methods. Compared to more traditional statistical methods, these algorithms have, in fact, often shown to be more effective in capturing complex patterns, outplaying less flexible methods in prediction and forecasting tasks [21]. However, despite the growing interest in these technological trends and the growing recognition of the importance of mobility in explaining crime, as of the current date, there exist very few studies on crime forecasting employing deep learning models and combining historical crime data with human mobility data. Moreover, such studies use relatively large geographic units of analysis (e.g., census tract or police beats) or have a large temporal granularity (i.e., weeks, months, or years), resulting in predictions that cannot easily yield actionable recommendations. Additionally, these studies often focus on only one city or employ just one year of historical data, leading to models with low generalizability to different contexts [22–25].

To address these gaps, in this research we propose a deep learning framework—based on Convolutional Long Short-Term Memory (ConvLSTM) layers—to perform spatiotemporal crime forecasting in microgeographic units, utilizing historical crime data in conjunction with micro-level mobility and sociodemographic data from 2019 to 2023 across four U.S. cities. More specifically, the main contribution of this work is that, to the best of our knowledge, this is the first study to use mobility data in spatiotemporal crime forecasting with very fine spatial and temporal resolutions (0.077 sq. miles and 12 hours, respectively), covering a period of five years of historical data and including four different cities. This design thus enables a broader comparative analysis, allowing for a more in-depth exploration of the benefits and limitations of predicting crime occurrences at highly detailed geographical and temporal levels. This, in turn, advances our understanding of crime both theoretically and practically, while further investigating whether and how artificial intelligence can serve as an effective ally for criminology. Our short-term forecasting effort, focused on hourly time scales and fine-grained spatial resolution, is motivated by two interrelated goals. By testing the predictive capabilities of deep learning approaches, we aim not only to advance theoretical insights into the mechanisms underlying crime occurrence and distribution, but also to generate practical knowledge that can inform policy decisions and situational interventions. These may include, for example, shift-by-shift allocation of patrol officers, but also extend to non-policing measures, such as adjusting lighting in high-risk areas, modifying urban infrastructure to reduce anonymity or escape routes, deploying community outreach teams during peak risk hours, or temporarily altering the accessibility of specific spaces (e.g., parks, transit hubs) based on dynamic risk assessments.

The remainder of the paper proceeds as follows. In Section 2, we outline the main theories that motivate the use of mobility data for crime forecasting and our contributions to the literature. In Sections 3 and 4, we describe the datasets used, the extensive data processing carried out, and the deep learning algorithm we designed to perform the crime forecasting task. In Section 5, we evaluate the performance of the model and the impact of

using mobility data on forecasting outcomes. Lastly, in Section 6, we discuss the results, highlight the limitations of this study, and address its ethical implications.

2 Background

2.1 Theoretical framework

2.1.1 Crime concentration

The primary framework of this study is the one stemming from the *law of crime concentration*, proposed by Weisburd in his 2014 Sutherland Address to the American Society of Criminology. It states that "for a defined measure of crime at a specific microgeographic unit, the concentration of crime will fall within a narrow bandwidth of percentages for a defined cumulative proportion of crime" [8]. This law was inspired by early descriptive findings on the study of crime that appeared in the 19th century as well as empirical evidence amassed over the last decades of the 20th century and claims that most crime concentrates in small areas regardless of the city [26]. The law of crime concentration, one of the most universally accepted features of crime as a human phenomenon, is among the many applications of the *Pareto principle*, which posits that while numerous factors may contribute to an observed outcome, a small subset of these factors often accounts for the majority of the observations [27]. This principle has been shown to govern many social phenomena. For instance, it appears in education, particularly in the dynamics of classroom disruption [28]; in network theory applied to social influence in online communities [29]; in the field of economics, specifically regarding wealth distribution [30]; and in other areas of criminology, for example, asserting that most crimes in a population are committed by a small number of individuals [31].

We frame the present study starting from the law of crime concentration because, if crime is clustered in space, it follows that there are statistical patterns to be inferred. Learning these patterns, along with their sources of variation, is the key goal of our forecasting framework.

However, while the law of crime concentration represents the fundamental intuition behind this work, there exist complementary theoretical traditions that should be taken into consideration to more comprehensively understand how concentration occurs and, importantly, how to explain and predict concentration in space and time. Modern urban environments are increasingly shaped by factors such as commuting patterns, population flows, gentrification, and urban sprawl [32]: these fluid phenomena are crucial to capturing crime patterns, making it imperative to incorporate such dynamic elements into the study of crime, especially when seeking to forecast crime emergence.

2.1.2 Crime and human mobility

As a social phenomenon, human mobility both influences and is influenced by the environments people engage with. These dynamics also play a critical role in the occurrence of criminal events, which arise from the interplay between offenders, potential targets, and their surroundings. Certain conditions—such as a lack of guardianship, transient populations, or economic disadvantage—can amplify the likelihood of crime occurring. Therefore, to better understand these processes, researchers have proposed theories that examine how human mobility, social structures, and neighborhood characteristics contribute to the spatial and temporal distribution of crime.

For instance, crimes are more likely to occur when offenders and potential targets converge in environments that lack capable guardianship. This concept is central to *Routine Activity Theory* (RAT), introduced by [19]. According to RAT, offenders act when they possess both the motivation and the capacity to commit a crime, often selecting victims either deliberately, based on characteristics like identity or possessions, or opportunistically, based on accessibility or vulnerability. Guardians—such as police officers, family members, or institutions like law enforcement—serve as deterrents but must be able to act effectively upon witnessing criminal behavior. The spatial dimension of RAT is further refined by *crime pattern theory* [33], which incorporates elements from *geometric theory*. Geometric theory explores how the built environment shapes the geographic distribution of crime, leading individuals to form awareness spaces consisting of their main routine activity nodes [34]. Hence, the intersection of offender awareness spaces and potential crime opportunities is where crime hot spots will arise [8, 35]. RAT is therefore important for crime forecasting and for this study specifically, as it provides a conceptual framework for understanding the relationship between human mobility patterns and the dynamics of crime opportunities. Based on these theoretical prepositions, incorporating mobility data can enhance predictive accuracy by allowing us to detect changes in routine activities (e.g., population density, and business hours) that influence the likelihood of crime. The theoretical postulations regarding the relationship between crime incidents and human mobility have been backed by extensive empirical evidence [24, 36–38]. Benefiting from the growing integration between traditional data and more innovative sources (for a review see [39]), these studies usually include mobility features in the form of *footfall*¹, showing that this feature can ameliorate the performance of crime forecasting models when compared to models that only use historical crime features.

Persistent crime patterns, however, are also closely linked to the socioeconomic conditions of neighborhoods. Communities with high levels of poverty, residential instability, and ethnic diversity often experience weakened social networks and limited informal social control. These conditions form the basis of the *Social Disorganization Theory* (SDT) developed by the Chicago School [20]. Social disorganization fosters environments where subcultures of crime and deviance flourish, further contributing to higher crime rates

¹The footfall is defined as the number of people in a given location during a defined time span.

[40]. Despite some challenges and inconsistencies in its empirical research [41, 42], SDT remains highly relevant as it provides a lens through which we can investigate how human mobility, neighborhood dynamics, and social structures intersect to influence crime rates. Hence, using mobility data in combination with sociodemographic data allows us to more comprehensively capture patterns and trends, as well as changes in these patterns and trends, occurring in urban environments. Leveraging both dimensions, we argue, would improve our ability to understand when and where crime will happen.

2.2 Contributions to crime forecasting

Crime forecasting research employing advanced computational methods encompasses a range of scales and methodologies, each designed to address specific contexts and objectives. A review of the past decade's literature reveals five core elements in study design. Here, we outline our contributions and innovations with respect to each of these elements.

Element 1: What urban focus?

Firstly, most studies focus on a single city in the U.S., such as New York City or Chicago [43–46]. A smaller number of studies aim for broader comparability. For example, [23] analyzed crime patterns in four U.S. cities highlighting the benefits of comparative analysis in diverse urban environments. Another notable example is [47], which examined event-level predictions of urban crime across eight U.S. cities. Drawing inspiration from these multi-city approaches, our study investigates four U.S. cities to enhance generalizability of our results, seeking to underscore common patterns and potential differences across diverse urban areas.

Element 2: What geographical scale?

Secondly, the choice of geographic unit, which reflects a fundamental trade-off between spatial detail and computational accuracy, has mostly focused on relatively large areas of 1 km² or larger [23, 24, 46, 48, 48]. Despite this, there are some studies that used very fine geographic units of analysis. For instance, [49] used a highly detailed grid made of square cells measuring 91 × 91 meters in their study of gun violence in Little Rock, AR, and [50] used a grid of cells measuring 183 × 183 meters, approximately 0.122 km² to crime incidents in Portland, OR. However, in such studies, the corresponding temporal granularity was relatively large (i.e., six months and two weeks, respectively). Our analysis adopts a spatial unit of 0.077 sq. miles,² offering a compromise that captures micro-spatial patterns effectively while remaining computationally manageable, also avoiding excessive sparsity, which would have rendered any forecasting exercise theoretically and practically useless.

²0.077 sq. miles are equivalent to 0.2 sq. kilometers in the metric system.

Element 3: What temporal granularity?

As anticipated in the previous paragraph, the temporal granularity of predictions further distinguishes contributions in this field. Many studies focus on daily predictions [23, 24, 43, 48, 51, 52], while some explore larger granularities. For instance, [53] employed a coarser approach, generating yearly forecasts for gun violence. Similarly, [54] examined monthly predictions to highlight long-term trends. Our work focuses on a 12-hour prediction interval, a granularity never explored in the literature in combination with fine spatial resolution. By narrowing the prediction window and using a very fine spatial granularity, we aim to provide more actionable insights for real-time resource allocation and urban safety measures, underscoring not only the potential of this method but also its limits.

Element 4: What temporal focus?

Temporal coverage of the historical data is another critical dimension, influencing the ability to capture trends and variability. Many studies rely just on one year of data, focusing on short-term predictions [23, 55, 56], while few others try to capture long-term patterns by extending their analysis to eight years or more [44, 51]. Other researchers, including [53] and [43], settled on datasets spanning five years, finding a balance between temporal depth and data availability. Our study aligns with this medium-term focus, leveraging five years of data to identify meaningful patterns.

Element 5: How many data sources?

Lastly, recent advancements in crime forecasting have increasingly emphasized the importance of integrating diverse data sources. While traditionally previous studies often relied exclusively on crime data [43, 47, 51, 52], more recent approaches highlight the added value of contextual information as well. In general, the use of mobility data has increased substantially in the past years in criminology [57], and has played a major role in the computational social science revolution [58]. Mobility is measured through various sources, including transportation data [59], GPS tracking [60, 61], and social media or other types of digital traces [37, 62, 63], and it allows researchers to capture individual and collective behaviors and flows at fine-grained geographical and temporal scales. Despite this growing interest and the evident potential of mobility data to enrich crime analysis, their role in short-term crime forecasting remains underexplored and largely unquantified. In particular, we still lack systematic evidence and empirical estimations of the extent to which incorporating mobility features can enhance the accuracy of predicting when and where crimes are likely to occur—an insight that would be highly valuable to both scholars and practitioners. For this reason, our study embraces a multidimensional perspective, integrating crime, mobility, and sociodemographic data to capture the complex interplay of social, spatial, and temporal dynamics in crime occurrence, in the attempt to disentangle the contribution that each dimension provides to the forecasting ability of deep learning.

Therefore, by building on these five elements, our study advances the field by adopting a multi-city framework with exceptionally fine spatial and temporal resolution, utilizing five years of data and integrating three distinct data sources. These design choices align with recent advancements while pushing the boundaries of crime forecasting research.

3 Data

3.1 Sample Selection

In this research, we analyzed four U.S. cities: Baltimore, MD (Bal); Chicago, IL (Chi); Los Angeles, CA (Las); and Philadelphia, PA (Phi). These cities were chosen due to their high crime rates and their good coverage of mobility data (the specific criteria we defined and used are explained in detail in Section 3.3). In addition, their diversity with respect to location, urban structure, and sociodemographics made them particularly appealing for our comparative analysis between cities. For instance, Baltimore is a port city focused on shipping and trade, with a significant African American majority. In contrast, Los Angeles is nearly six times larger, characterized by a sprawling car-dependent structure, and is predominantly home to Latino and Asian communities. A more detailed overview of the four cities can be found in Table 1. Using multiple cities represents a relevant feature of this study since, as previously mentioned, most research on crime forecasting focuses on only one city, making it difficult to argue for the applicability of those models to all contexts.

	Region	Population	Area	Black	Hispanic	Asian	Poverty	Crime rate
Bal	Northeast	565,239	210	61.2%	5.9%	2.6%	19.6%	6,997
Chi	Midwest	2,664,452	591	28.8%	29.0%	7.0%	16.9%	4,381
Las	West	3,820,914	1,216	8.6%	48.1%	11.8%	16.6%	3,332
Phi	Northeast	1,550,542	347	40.1%	15.7%	7.6%	22.7%	4,037

TABLE 1. Comparison of the four U.S. cities we selected. Information obtained from the U.S. Census Bureau QuickFacts for 2023, except for the crime rates, which were obtained from FBI's UCR data and represent the crime rates per 100,000 people in 2019. The area is in sq. kilometers.

In the following subsections, we present the three types of data sources used to create city-specific datasets, as well as the data processing that had to be performed for each of them.

3.2 Crime data

We used city-wide crime datasets containing crime incidents from 2019 to 2023, publicly available through the Open Data portals of the four cities. The final versions of the datasets

contained four variables: the date and time of the crime incident, the category of the crime incident, and the location of the crime given as latitude and longitude coordinates.

Although several crime categories were available depending on the city, we first selected those that were common across all four settings and, among these, the ones more likely to be reported to the police. This latter criterion was important to ensure that the crime data we used for the analysis were not severely impacted by the risk of under-reporting, which would have introduced non-trivial issues to our forecasting exercises [64–66]. Specifically, we selected the following crime categories: burglary, motor vehicle theft (MVT), assault, homicide, and robbery. The first two are property crimes, while the remaining three are violent crimes. Our analysis is thus three-fold: we developed models forecasting all crime together, as well as models focusing on property and violent crimes separately. The percentages of each type of crime and the total number of crime incidents for each city are shown in Table 2. From these data, we observe that assault is the most common crime in all four cities, accounting for around half of the total incidents, while homicide is the least common.

	Property Crimes		Violent Crimes			Total incidents
	Burglary	MVT	Assault	Homicide	Robbery	
Bal	15.14%	18.53%	53.75%	1.22%	11.36%	132,002
Chi	14.81%	29.68%	38.00%	1.29%	16.23%	268,627
Las	15.01%	24.73%	51.19%	0.39%	8.68%	438,905
Phi	11.15%	16.91%	61.36%	0.84%	9.73%	265,097

TABLE 2. Percentage of each crime type over the total amount of crime incidents of those categories between 2019 and 2023 for each of the four U.S. cities we selected.

Furthermore, since our main focus is on 12-hour blocks (half-days) for crime predictions, Figure 1 presents the average number of crime incidents every 12 hours for each city. This visual representation reveals that most types of crime average fewer than 40 incidents per day, except assault, which exceeds this threshold in Los Angeles and Philadelphia. Furthermore, homicide averages are, as expected, much lower than the other crime categories across all four cities, as shown in Table 2, potentially posing a further challenge for our forecasting model due to data sparsity. These findings are consistent across all five years of data.

Using this dataset, we generated a single feature represented as a binary variable: 0 if no crime occurred and 1 if at least one crime occurred within each spatial unit of analysis. This approach was chosen due to the extreme sparsity of the data (clearly illustrated in Appendix A), which made using a binary variable effectively equivalent to examining the actual crime counts. In Baltimore, the percentage of instances—i.e., spatial cells at a given time—with no crime (0s) was, on average, $97.96\% \pm 0.73$, while for Chicago, Los Angeles, and Philadelphia, the corresponding percentages were $99.03\% \pm 0.34$, $99.39\% \pm 0.19$, and $98.38\% \pm 0.62$, respectively. This significant imbalance, with the overwhelming majority

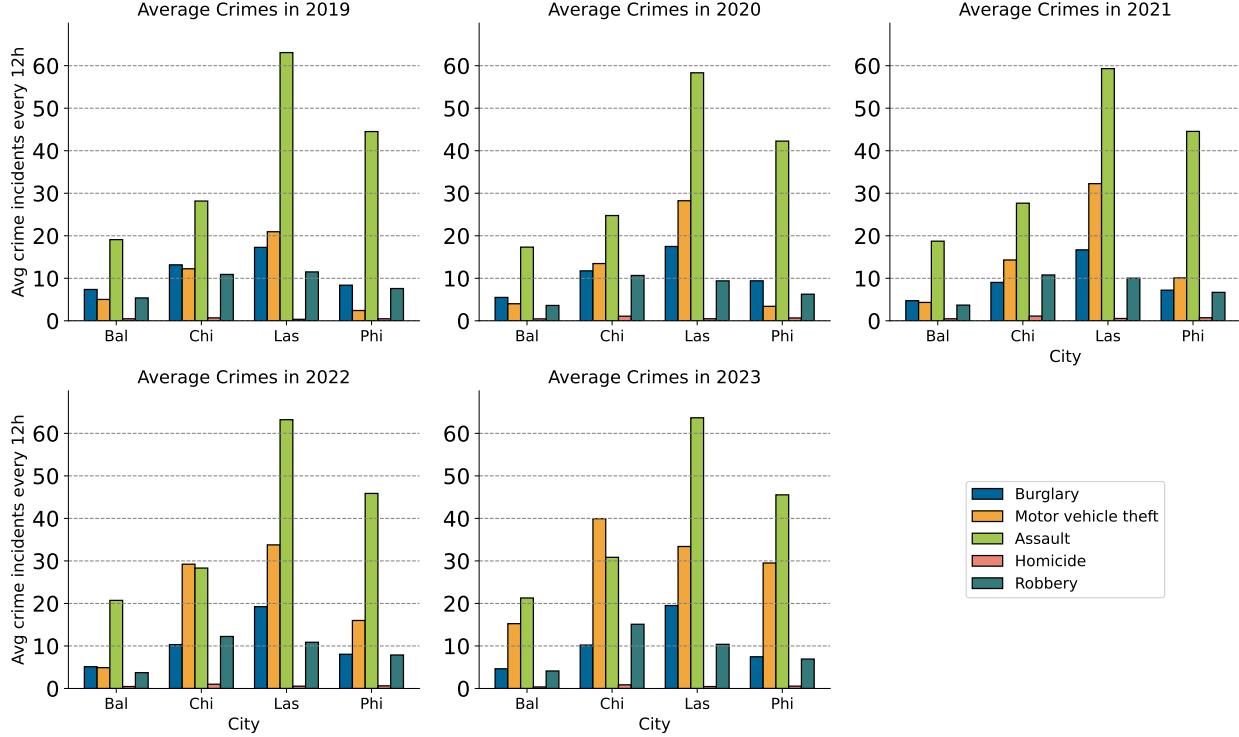


FIGURE 1. Yearly average of crime incidents recorded every 12 hours for each city and crime type.

of cells experiencing no crime (0s), makes the prediction task particularly challenging since the scarcity of positive instances limits the algorithm’s ability to learn effectively from these events, increasing the risk of overfitting or bias toward predicting no crime.

3.3 Mobility data

The source of micro-level mobility data was Advan, a location intelligence company that is mainly focused on providing micro-level data to the financial and real estate industries. They provide anonymized and GDPR³ and CCPA⁴ compliant mobile data from 100 million cellphones, offering a wide set of information about a given location, such as foot traffic and demographics [67]. Advan data starts from 2019 and is updated weekly. Specifically, we exploited their “Weekly Patterns” dataset [68], which contains data from 11,590,987 points of interest (POIs) distributed across the U.S. and Canada. POIs are significant for various reasons, such as their cultural, commercial, or logistical importance. Common examples of POIs include hotels, restaurants, gas stations, stadiums, churches, shopping malls, parks, hospitals, and other places that attract visitors or serve as hubs of social activity in urban contexts.

³General Data Protection Regulation.

⁴California Consumer Privacy Act.

In order to select the subset of cities to be used for this study, we were aware that not only the number of POIs was relevant, but also the relationship POIs had with population size and the city’s area since population density and POI density allow for a better understanding of the potential human mobility coverage of each POI. For this reason, we defined a variable called the *people-to-POI ratio*, representing the number of people per POI. This ratio was calculated by dividing the population density by the POI density, hence indicating the number of people per POI within 1 km². Therefore, since this is a normalized value, it offers a fairer way to compare the cities, particularly in terms of the richness of the data in capturing as much as possible the underlying mobility dimension of a city. We computed this ratio for the main 735 cities in the U.S. and obtained a list ordered by ascending people-to-POI ratio. From this list, we focused on cities with high crime rates, verified that (a) each did not fall in the upper extreme of the distribution in terms of people-to-POI ratio⁵ and (b) each had complete crime data from 2019 to 2023 publicly available. This resulted in the final selection of four U.S. cities mentioned in the previous subsection (i.e., Baltimore, Chicago, Los Angeles, and Philadelphia). Table 3 presents the people-to-POI ratio for each of the four cities, with a lower ratio indicating better POI coverage. In Appendix B, we visualize the histogram of the distribution of the people-to-POI ratio for all 735 cities.

	Area	Population density	Num. of POIs	People-to-POI ratio
Bal	210	2,930	28,218	21.78
Chi	591	4,581	95,049	28.46
Las	1,216	3,273	102,478	38.80
Phi	347	4,518	55,496	28.25

TABLE 3. Relevant geographical attribute of each of the four cities selected. The area is in km², the population density in people per km², the Num. of POIs is the total number of unique POIs available for that city (not all of them were used during all four years), and the people-to-POI ratio in people per POI per km².

From this dataset, we constructed several mobility features, which we classified into two groups. First, we used the variable “top_category” to assign each POI type to one of 11 categories, following the NAICS system⁶. The 11 categories are: (1) Utilities and construction, (2) Manufacturing, (3) Retail and Wholesale Trade, (4) Transportation and Warehousing, (5) Business and Professional Services, (6) Educational Services, (7) Health care and Social Assistance, (8) Arts, entertainment, and recreation, (9) Accommodation and food services, (10) Public Administration, (11) Other Services. More details are provided in Appendix D. Using the variable “visits_for_each_hour”, we then computed the *footfall* for each POI category. These features allowed us to obtain dynamic quantities reflecting how many

⁵A very high people-to-POI ratio would indicate likely poor coverage and representativity of POIs in a urban context.

⁶NAICS, or the North American Industry Classification System, is a standardized framework used by businesses and governments to classify industries and organize economic data across North America.

people visited each POI category in each cell during each time window, thus representing the underlying mobility characteristics of all our cities, mapping potential time-specific patterns at the daily level (i.e., dynamics changing between day and night), as well as meso-level (i.e., different dynamics on different days of the week) and macro-level (e.g., structural changes due to gentrification) seasonality and trends.

Second, we calculated the POI category diversity using the *Shannon Diversity Index*⁷, defined as:

$$H = - \sum_{i=1}^n p_i \cdot \ln(p_i), \quad (1)$$

where p_i represents the proportion of each POI category i within each spatial unit of analysis, and n is the total number of POI categories (11 in our case). We computed this measure to map the heterogeneity of a given spatial unit in terms of human activities carried out, hypothesizing that this information can be valuable in understanding its potential crime attractors or targets, as well as representing the complexity of its human and social fabric. To calculate the proportion of each POI category within each spatial unit, we first determined the total number of POIs across all categories within each cell by summing the individual counts. The proportion for each category was then computed by dividing the count of that category by the total count within the cell. This process yielded a total of 12 features: 11 representing footfall per POI category and one representing POI diversity. Maps displaying the distribution of POI diversity for each of the four U.S. cities can be found in [Appendix F](#).

3.4 Sociodemographic data

Besides crime and mobility data, we enriched our feature space by including a range of sociodemographic variables. These features allow us to take into consideration the broader social context of each geographic unit beyond the type of activities and premises that are captured by our mobility data.

The data we used were obtained from the American Community Survey. Specifically, we used the ACS 5-year estimates for 2019, 2020, and 2021⁸⁹. These datasets contain more than 20,000 variables related to the social, economic, demographic, and housing characteristics of the U.S. population for different geographic granularities (e.g., region, county, county subdivision, and block group).

Firstly, we chose the block group as our geographic level since it is the smallest granularity available. Afterward, we selected 26 variables related to gender, age, race/ethnicity,

⁷The Shannon Diversity Index (also called the Shannon-Wiener Index) is used to characterize entity diversity in a given community or set of elements. The higher the index value, the greater the entity diversity.

⁸<https://www.census.gov/data/developers/data-sets/acs-5year.2021.html#list-tab-1036221584>

⁹The data for 2022 and 2023 had not been released at the time of this study, so we decided to use the 2021 data for both years, as we do not expect significant variations.

employment, income, education, and marital status—features that have been identified as correlated with crime in the scientific literature [69–71]. The specific variables are listed in Appendix E. Their raw values (in the case of medians) or percentages (in the case of counts) were used as our 26 sociodemographic features.

4 Methodology

4.1 Data preparation

To formalize the microgeographic unit of analysis, we used a grid-based approach, which consisted of defining a grid containing the city borders and then calculating the values of all our features for each spatial cell during every 12-hour period throughout the study’s time span. Therefore the problem definition becomes as follows: Consider that we divide a city into an $N \times N$ grid (with N varying for each city to ensure that each cell has an area of 0.077 sq. miles), hence into a matrix of the form $\mathbf{G} = (g_{i,j})_{0 < i, j \leq N}$. Then for each grid cell $g_{i,j}$, we compute 39 sets of predictive features for each 12-hour block. The first predictive feature is the historical crimes, where we have that the crime feature at the 12-hour block t is represented by $\mathbf{C}_{i,j}^t = \{c_{i,j}^{t-T}, c_{i,j}^{t-T+1}, \dots, c_{i,j}^{t-1}\}$ where T is our look-back period, hence how many blocks of 12 hours before t we use to predict this feature at the 12-hour block t . The second set is the 12 mobility features, which are denoted as $\mathbf{M}_{i,j}^t = \{m_{y_{i,j}}^{t-T}, m_{y_{i,j}}^{t-T+1}, \dots, m_{y_{i,j}}^{t-1}\}$ for each mobility feature y with $y = 1, 2, \dots, 12$ and for each 12-hour block t . Lastly, we have the 26 sociodemographic features, hence for each sociodemographic feature z with $z = 1, 2, \dots, 26$ and each 12-hour block t , we get the features $\mathbf{S}_{z_{i,j}}^t = \{s_{z_{i,j}}^{t-T}, s_{z_{i,j}}^{t-T+1}, \dots, s_{z_{i,j}}^{t-1}\}$ ¹⁰. Consequently, and borrowing this example from computer vision, each frame can be interpreted as an $N \times N$ image with 39 channels, where each channel represents a different feature (see Figure 2). In other words, our forecasting task can be thought of as similar to predicting the next frame in a video: given a series of previous frames, each characterized by a specific set of features, what will the next frame look like?

These frames are then grouped into sequences, with the length determined by the look-back (LB) period T . Each feature within these sequences is then normalized using min-max scaling, ensuring consistent value ranges across the dataset. Also, to maintain uniform grid sizes across all four cities and to augment the number of samples available, we randomly sampled five $M \times M$ subsections for each sequence of frames, keeping only the sequences with at least 2 crimes in that area in order to balance the dataset as much as possible. For this study, we set $M = 16$, which covers approximately 19.31 sq. miles (50 km^2), and used two LB periods of $T = 28$ and $T = 4$ (equivalent to 14 days and 2

¹⁰Since the size of each block group varies significantly, the preprocessing of the sociodemographic features involved identifying the block group(s) that intersected with each cell, and extracting the corresponding sociodemographic data. In cases where multiple block groups overlapped with a cell, a simple average was computed. For cells with no intersection, the value was set to NaN.

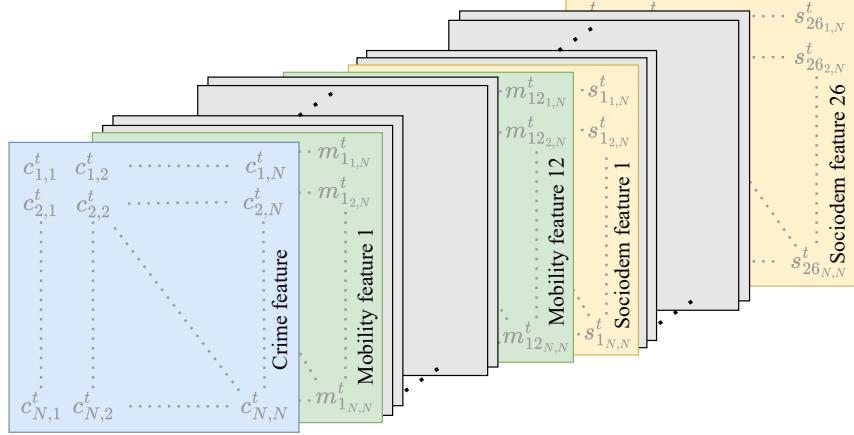


FIGURE 2. Illustration of the structure that we used for our $N \times N$ frames with 39 channels for each 12-hour block t . The first channel corresponds to the crime feature, the following 12 channels are the mobility features, and the last 26 channels are the sociodemographic features.

days respectively) to assess the optimal amount of data required to maximize forecasting outcomes. This dimension of the analysis not only bears practical importance but also theoretical relevance: a shorter LB period would indicate that crime dynamics cluster and change in short time frames, while a longer LB would suggest that crime dynamics are more complex and require long-term memory.

Therefore, the goal is to use T consecutive 12-hour frames to predict the crime occurrence at $T + 1$, where the crime occurrence is defined as a 16×16 matrix with either $y_{i,j} = 0$ (for no crime) or $y_{i,j} = 1$ (when there is at least one crime) for each cell $g_{i,j}$. It should be noted that this binary problem is essentially equivalent to predicting crime counts due to the extreme sparsity of the crime data, as discussed in Section 3.2.

Using this approach, since we had a total of 3,640 half-day frames, using a sliding window of length $T + 1$ resulted in a total of $3,640 - T$ sequences. These sequences were split into training and test sets in a 90/10 ratio, maintaining their chronological order. To prevent data leakage between the last sequences of the training set and the first ones of the test set, the final T sequences from the training set were removed.

To understand how this procedure works, suppose we have 10 half-day frames and a sliding window of length three (i.e., $T = 2$, meaning that we use the first two frames to predict the crime in the third one), as shown in Figure 3.

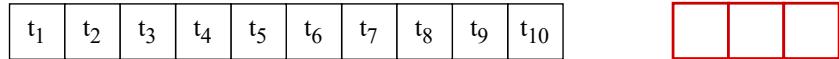


FIGURE 3. Illustration of a toy dataset with 10 half-day frames (left) and a sliding window of size $T + 1 = 3$ (right).

As we slide our window over the data, we obtain eight (i.e., $10 - T$) sequences: (t_1, t_2, t_3) ,

(t_2, t_3, t_4) , (t_3, t_4, t_5) , (t_4, t_5, t_6) , (t_5, t_6, t_7) , (t_6, t_7, t_8) , (t_7, t_8, t_9) , and (t_8, t_9, t_{10}) . Applying a 90/10 chronological split assigns the first seven samples to the train set and the last one to the test set. Next, to avoid data leakage, we remove the last T sequences of the training set, since they contain frames that are part of the test set. This results in a training set with five sequences and a test set with one sequence.

For our actual dataset, when $T = 14$, this procedure resulted in training and test sets containing 3,222 and 362 sequences, respectively, which would increased to 16,110 and 1,810 samples after extracting the five subgrid sections. However, sequences with no crime events were discarded, leading to the loss of some samples. Additionally, to keep the same sample size for all data configurations (i.e., for the four cities and three crime aggregations), we retained the number of samples corresponding to the smallest dataset among those. This yielded a final training set size of 12,546 sequences and a test set size of 1,510 sequences.

Lastly, before using these datasets for our analysis, we applied a spatial mask to exclude all cells that met at least one of the following criteria: (i) they are located outside city borders, (ii) they consist entirely of water, or (iii) they do not intersect with any census block group. A visualization of the included and excluded cells is provided in [Appendix H](#).

4.2 Convolutional LSTMs

Our main model is based on ConvLSTM layers, proposed by [\(author?\) \[72\]](#). This type of neural network consists of two main components: Long Short-Term Memory (LSTM) layers and convolution operations.

LSTMs are powerful neural networks designed to handle temporal patterns. For instance, they have been used to predict daily streamflow through weekly forecast horizons [\[73\]](#), for speech recognition by modeling temporal dependencies in audio signals [\[74\]](#), and for determining the probability that certain social media content will gain popularity [\[75\]](#). They are able to perform such tasks by using an internal structure referred to as *memory cells*. Each cell contains a *cell state* (c), a *hidden state* (h), and three types of gates that determine the flow of information using *sigmoid* and *tanh* activation functions. The types of gates are the *forget gate* f_t , the *input gate* i_t , and the *output gate* o_t . Their corresponding equations for the multivariate version of the LSTM (i.e., the fully connected LSTM or FC-LSTM), where both inputs and outputs are 1-dimensional vectors, are as follows:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{ci} \odot c_{t-1} + b_f) \quad (2)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{cf} \odot c_{t-1} + b_i) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{hc}h_{t-1} + W_{co} \odot c_t + b_o), \quad (4)$$

where h_{t-1} is the previous hidden state, x_t is the current input, W are the weight matrices (where the subscripts indicate the two variables that are connected by this matrix), b_x is the

bias for that gate, and σ represents the sigmoid function [76]. The goal of the forget gate is to determine which information from the previous cell state c_{t-1} should be forgotten. The aim of the input gate is to select which new information should be added to the current cell state c_t . Lastly, the output gate is used to assess what information should be included in the current hidden state h_t .

Each candidate cell is formalized through the equations related to the cell state c_t and the hidden state h_t :

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t). \quad (6)$$

These equations highlight how the candidate cell is used in combination with the forget gate and the input gate in order to obtain the current cell state. Subsequently, the cell state and the output gate are employed to obtain the hidden state, which summarizes all the information learned. Lastly, the hidden state is utilized to make the final prediction y_t by using the appropriate activation function for the task at hand.

Therefore, memory cells allow LSTMs to retain only the information that is most relevant to the forecasting task. The advantage of these internal structures is that they trap the gradient¹¹ in the cell, preventing it from becoming too large or too small and thus avoiding in this way the vanishing/exploding gradients problem that significantly limited its predecessor, the recurrent neural networks [77–79].

However, while LSTMs are effective at capturing temporal information, they are not engineered to account for spatial information. To address this, ConvLSTMs extend traditional LSTMs by replacing the one-dimensional vectors with three-dimensional tensors, where the first dimension represents time and the other two are the spatial dimensions (the number of rows and the number of columns) of the input frame. Therefore, the matrix multiplications in the FC-LSTM equations—used in both the input-to-state and state-to-state transitions—are replaced by convolutional operations, and the weight matrices are replaced by convolutional kernels [72].

To illustrate how the introduction of convolutional operations allows the model to preserve spatial information, consider a simple example where the input x_t is a single-channel 3×3 spatial grid and the output should be of the same shape. Then the operation Wx_t is done as follows. In the case of LSTM, x_t is flattened into a 9×1 vector and this operation becomes a simple matrix multiplication with W being a 9×9 weight matrix. Hence, the operation results in a 9×1 vector, which is then reshaped into a 3×3 matrix to recover its 2-dimensional shape. Therefore, we can see that the fully connected operation disregards the spatial structure of x_t . In contrast, for ConvLSTM, assume W is a 2×2 convolutional kernel. The convolutional operation is then performed by sliding this kernel over the input grid and performing element-wise multiplication of its values with the corresponding elements in the input grid, followed by summing the results. This

¹¹The gradient is a vector that points in the direction of the steepest increase in a multivariate function.

operation is repeated for the rest of the grid, with the kernel sliding across the grid one element at a time in both the horizontal and vertical directions. To ensure the output has the same spatial dimensions as the input, zero-padding¹² is applied. Therefore, this operation results in a 3×3 matrix as output, preserving the spatial dependencies of the input, and maintaining the spatial structure of x_t throughout the computation.

Another advantage of this type of neural network is that the use of three-dimensional tensors considerably reduces the number of parameters to be trained when compared to LSTMs. This ability to incorporate both spatial and temporal information makes ConvLSTMs appealing for capturing the complex spatiotemporal dynamical patterns in crime events, hence offering a suitable algorithmic framework for our task.

4.3 Model architecture

ConvLSTMs were initially defined for precipitation nowcasting, using past radar sequences to predict future radar frames [72]. Since then, they have been applied in various contexts, such as quantifying the contributions of climate change and human activities to vegetation change, and predicting short-term traffic flow [80, 81]. Therefore, this type of neural network is typically used for video frame prediction, as it accounts for both the temporal and spatial components of the data, making it ideal for the data structure we constructed. While this is usually performed with either grayscale images (1 channel) or RGB images (3 channels)¹³, we adjusted the model architecture to train on images with 39 channels (corresponding to our 39 features) and output a 1-channel image.

In order to engineer our model, we started with a basic architecture for video frame prediction using ConvLSTMs and adapted it to our specific problem. Then we fine-tuned several hyperparameters (see [Appendix G](#) for the specific values used) until we obtained the model architecture depicted in Figure 4.

As already mentioned, the goal of this model is to predict crime occurrences at a future point in time (time t) based on past data. Specifically, our model analyzes a sequence of T frames (from $t - T$ to $t - 1$) to make this prediction, with each frame containing 39 channels, one for each feature. The crime occurrence is represented by a 16×16 matrix with zeros and ones, indicating no crime or at least one crime, respectively. Therefore, the input to the model consists of a sequence of T frames, each with a shape of $16 \times 16 \times 39$ (the first value represents the height, the second the width, and the third the number of channels). This sequence is then processed by the first ConvLSTM block, called *ConvLstm1*, which sequentially processes each frame and updates the hidden states. After processing all T frames, the final hidden states of the ConvLSTM layer contain information about both

¹²Zero-padding is the process of adding extra zeros around the border of an input grid before applying the convolution operation. In our example, by adding one layer of zeros around the input, the grid becomes a 5×5 matrix, allowing the kernel to process 9 regions in total (one for each possible position of the kernel). This ensures that the output remains 3×3 , preserving the spatial dimensions of the input.

¹³Grayscale images have only 1 channel, representing intensity information, while RGB images contain separate channels for red, green, and blue.

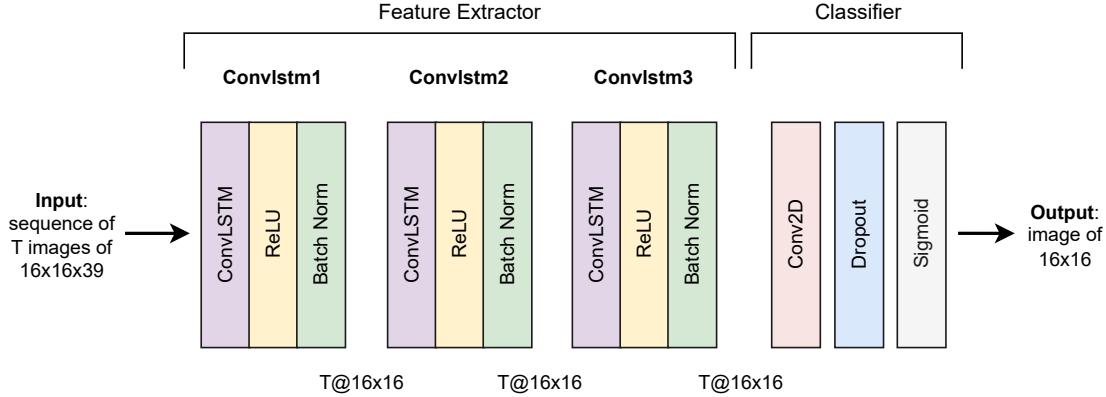


FIGURE 4. Architecture that we designed for our ConvLSTM model, where T is our LB period.

the temporal and spatial context. This process is repeated through two more ConvLSTM blocks, allowing the network to learn more intricate patterns by capturing different levels of temporal and spatial abstraction. This section of the model is therefore used to extract the main features from the input data. Additionally, a *ReLU* activation function and a *batch normalization* layer are applied at the end of each ConvLSTM block to introduce nonlinearity and to improve the learning process by normalizing the inputs to the network, respectively. We then proceed to the classifier section of the model, where the final hidden states of the last ConvLSTM block are processed by a two-dimensional convolutional layer, which produces the final 16×16 output image. This step is followed by a dropout layer, which helps prevent overfitting, and a sigmoid function, which converts the values of each cell into probabilities, resulting in values between 0 and 1 that indicate the likelihood of at least one crime occurring at time t . These probabilities are then turned into a binary classification using a threshold of 0.5. We chose not to optimize the threshold because, due to the extreme sparsity of our data, increasing the threshold resulted in a sharp decrease in recall and a marginal increase in precision (see [Appendix I](#)).

4.4 Model evaluation

Once we obtain the final prediction, we evaluate the model's performance by calculating recall, precision, and F1 score. Specifically, recall indicates how well the model predicts all instances where at least one crime actually occurred, while precision reflects how often the model's prediction of at least one crime is correct. The F1 score then combines precision and recall into a single value, providing a balanced assessment of the model's performance¹⁴. This evaluation is performed over four different random seeds (i.e., 0, 42,

¹⁴Recall, Precision, and F1 score are calculated as follows:

$$\text{Recall} = \frac{tp}{tp + fn}, \quad \text{Precision} = \frac{tp}{tp + fp}, \quad F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

123, and 999) to control for the role of stochasticity in model training and to obtain the average and standard deviation of each model's performance.

Given the unbalanced distribution in our target variable and the differential cost of wrong predictions between false positives and false negatives, we primarily focus on recall to quantify the predictive performance of the model. This is due to the high costs associated with false negatives, as they signify a failure to allocate crime prevention resources to a location in need.

Besides these two metrics, we also calculate a second version of recall and precision that considers false positives with neighboring positive cells¹⁵ as true positives. This choice is motivated by two aspects. First, the presence of crime in a location affects the nearby areas (and the same occurs with police presence) and spillover effects may lead to incidents in locations that fall within adjacent spatial cells. Second, given obfuscated georeferenced crime data, crime incidents that actually occurred in a given cell may be recorded differently and therefore fall in a neighboring one: calculating modified metrics would then allow us to predict them without losing such relevant data points. We calculated these two modified metrics as follows. Let $\mathbf{O} = (o_{i,j})_{1 \leq i,j \leq M}$ denote the ground truth crime occurrence matrix and $\mathbf{P} = (p_{i,j})_{1 \leq i,j \leq M}$ the predicted crime occurrence matrix, where $o_{i,j}, p_{i,j} \in \{0, 1\}$ and M is the size of the matrix as defined in Section 4.1. Next, let $N_{i,j}(\mathbf{O})$ represent the set of indices of the cells with a *Chebyshev distance*¹⁶ of 1 from cell $(i, j) \in \mathbf{O}$ (i.e., its nearest neighbors). Using these three elements, the classification outcomes are defined as follows:

$$\text{Cell Classification: } \begin{cases} \text{True negatives}(tn) : \text{Cells where both } o_{i,j} = 0 \text{ and } p_{i,j} = 0; \\ \text{True positives}(tp) : \text{Cells where both } o_{i,j} = 1 \text{ and } p_{i,j} = 1; \\ \text{False negatives}(fn) : \text{Cells where } o_{i,j} = 1 \text{ and } p_{i,j} = 0; \\ \text{False positives} : \begin{cases} fp_{nn}, & \text{if } \exists (i', j') \in N_{i,j}(\mathbf{O}) \text{ such that} \\ & o_{i',j'} = 1, \text{ (neighboring false positive)} \\ fp, & \text{(otherwise, standard false positive)} \end{cases} \end{cases} \quad (7)$$

Therefore, the modified metrics become:

$$\text{Recall NN} = \frac{tp + fp_{nn}}{tp + fp_{nn} + fn}, \quad \text{Precision NN} = \frac{tp + fp_{nn}}{tp + fp_{nn} + fp}. \quad (8)$$

¹⁵We define as neighboring cells all the cells that are touching the target cell. Hence, there can be between 3 and 8 neighboring cells, depending on the target cells' location in the grid.

¹⁶The Chebyshev distance, also known as the chessboard distance, between two cells (i_1, j_1) and (i_2, j_2) in a grid is defined as:

$$D((i_1, j_1), (i_2, j_2)) = \max(|i_1 - i_2|, |j_1 - j_2|).$$

This metric captures the maximum horizontal, vertical, or diagonal displacement between two cells.

These modified metrics (called NN, an acronym for nearest neighbor) are particularly relevant and applicable in our case since having cells of size 0.077 sq. miles implies that a police officer would take approximately 5 minutes (if the cells share an entire border) or 8 minutes (if they only share a corner) walking (or 2.5 and 4 minutes running, respectively) to go from the center of one cell to the center of a neighboring cell, assuming an average walking speed of 5 km/h. This highlights how small our geographic unit of analysis is and underscores how NN performance can be relevant for first responders.

Lastly, to determine the effectiveness of our proposed model, we compare it against three baseline models, chosen to provide a comprehensive evaluation by spanning statistical methods, traditional machine learning techniques, and deep learning approaches: (1) Logistic Regression (LR), a commonly used baseline in statistics and machine learning that assumes linear relationships between the features and the target variable; (2) Random Forest (RF), a traditional machine learning ensemble approach well-regarded for its ability to handle nonlinear relationships; and (3) LSTM, a deep learning model capable of capturing temporal dependencies but unable to account for spatial dependencies.¹⁷

5 Results

5.1 General performance

5.1.1 All crimes

Firstly, we assessed the model’s performance against the three chosen baselines—LR, RF, and LSTM—by considering all five crime categories together, to predict whether there will be a crime in the next time window without discriminating by type. On the one hand, this approach has the advantage that it contains more crime incidents, thereby making the dataset relatively more balanced, and stabilizing the learning process. On the other hand, combining crimes with distinct seasonal patterns may introduce further challenges by obscuring individual crime trends.

The results in Figure 5 show that ConvLSTM is the overall best-performing model, which can be clearly seen when comparing the standard and modified F1 scores across the four models. It also achieves the highest standard recall values in all cities, with a notable increase in performance when comparing the results of the modified metrics to the standard metrics. To exemplify, recall for our ConvLSTM model is always higher than 0.65

¹⁷We recognize that self-exciting models (see, for reference, [10] and [82]) were specifically designed to address the challenges of short-term crime forecasting and have demonstrated considerable value in the criminological literature. However, we deliberately chose not to include these models among our baseline comparisons for the present study. This decision is motivated by the fact that a fair and meaningful comparison would require extending the ConvLSTM architecture with attention mechanisms, which better capture the dependencies central to the concept of self-excitation. Such a modification would represent a significant methodological advancement in its own right and is therefore beyond the intended scope of this work.

across all cities, meaning that we are able to correctly forecast crime presence for 65% of the cell-time unit observations in which crime actually occurred. Precision is extremely low, being consistently below 10%, meaning that only one in ten forecasts of crime presence are actually true positives. However, when considering the modified precision metric, performance increases substantially: on average, when we forecast the presence of crime for a given spatial cell, there is approximately a 50% chance that crime will occur in some of the adjacent spatial cells in that same time unit. Given our fine-grained spatial scale and the highly unbalanced nature of our algorithmic task, this represents an important level of accuracy that does not hinder potential timely response in the short aftermath of a crime.

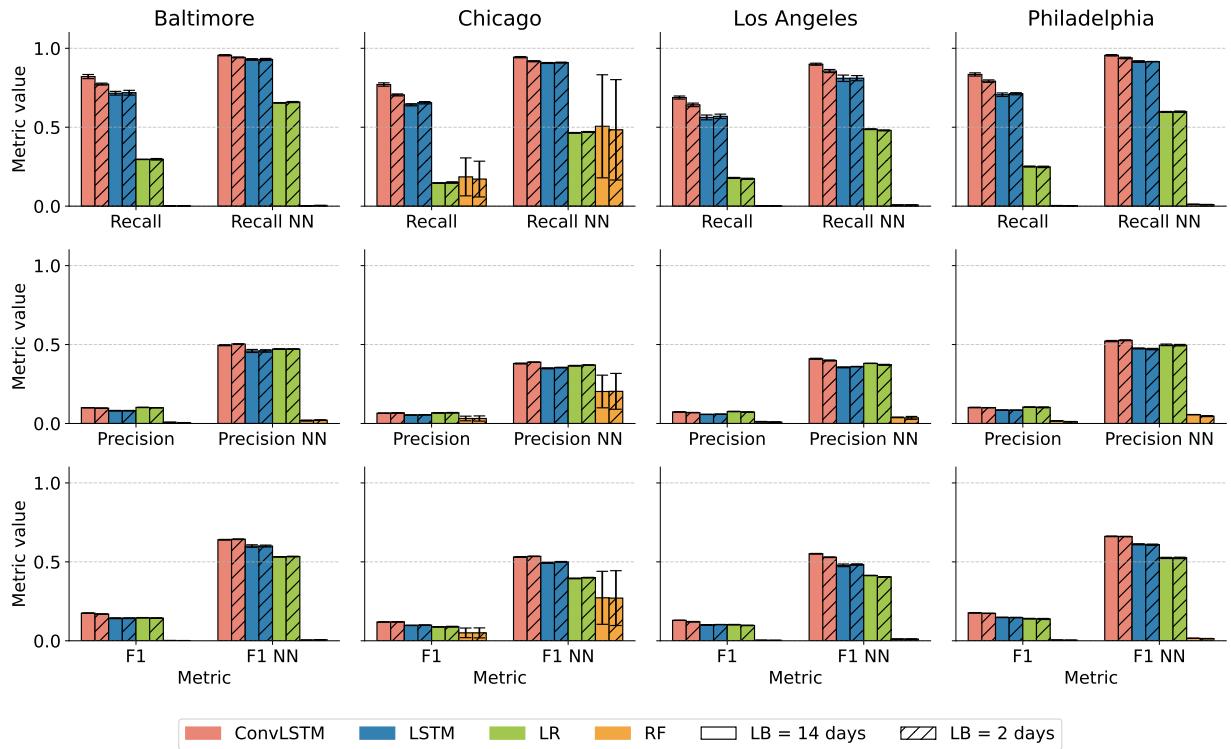


FIGURE 5. Average performance on the test set for our model and three baselines with and without taking into account the NN cells, considering all five crimes together and two different LB periods. Error bars indicate the standard deviation.

Regarding the baseline models, all three exhibit consistently lower standard recall values than ConvLSTM. Specifically, LR and RF remain below 0.30, correctly identifying fewer than 30% of the actual positive instances (cells with at least one crime in a 12-hour block). However, LSTM achieves recall values close to those of ConvLSTM, highlighting how deep learning approaches in general are much more effective compared to less sophisticated (and less computationally costly) models. In terms of precision, LSTM and LR are comparable to ConvLSTM, albeit slightly lower. RF, despite showing similar patterns for Chicago, performs the worst across all metrics, with only marginal improvements for

the modified metrics. Although LSTM is the second-best model for recall, its precision is lower than that of LR, highlighting the importance of the spatial patterns captured by ConvLSTM’s convolutional operations for improving both recall and precision. It is also worth noting that, despite LSTM’s comparable performance, it has approximately ten times more parameters than ConvLSTM, leading to higher computational costs and an increased risk of overfitting, particularly with limited data.

Lastly, the comparison of the LB periods reveals mixed results across models. For ConvLSTM, the 2-day LB period typically shows slightly lower performance compared to the 14-day period, although it achieves higher modified precision in Baltimore, Chicago, and Philadelphia. In contrast, LSTM generally performs better with the 2-day LB period. Meanwhile, LR demonstrates similar performance across both LB periods, while RF seems to benefit from longer LB periods. If we instead focus on the F1 scores, we observe that both LB periods have similar performance, with only Los Angeles showing noticeably higher performance when using the 14-day period for all models except LSTM. A robustness check using 7-day and 1-day LB periods (see [Appendix J](#)) shows that the shorter LB period performs similarly or worse, except for the modified value in Baltimore. These mixed results suggest that the optimal LB period to be used to forecast crime depends both on the chosen algorithm and on the city itself.

5.1.2 Violent crimes

Figure 6 presents the results when considering only violent crimes (i.e., assault, homicide, and robbery). Although this dataset contains fewer crime incidents, it remains relatively balanced since assault accounts for approximately half of the crime incidents in each city, as shown in Table 2.

Consistent with the previous results, the modified metrics achieve higher performance than the standard metrics across all models and cities, although the magnitude of these improvements varies. Notably, ConvLSTM remains the best-performing model for all cities, with generally higher modified recall values compared to when we used all crimes together. This suggests that violent crimes may exhibit stronger spatial dependencies, hence making the use of modified metrics more beneficial. However, precision decreases across all cities. In general, when comparing the F1 scores, we can conclude that the overall performance decreases relative to using all five crimes together.

With respect to the baseline models, LR achieves precision values similar to those of LSTM, slightly exceeding it in all cities for both the standard and modified metrics. Despite this, its recall remains consistently low. In contrast, RF performs poorly across all metrics and remains the worst-performing model, indicating its inability to capture meaningful patterns. Meanwhile, LSTM shows a slight increase in recall, but a decrease in precision compared to its results when considering all crimes combined.

These trends hold true across both LB periods. Nevertheless, we still observe that for ConvLSTM the 2-day LB period performs slightly worse than the 14-day period when

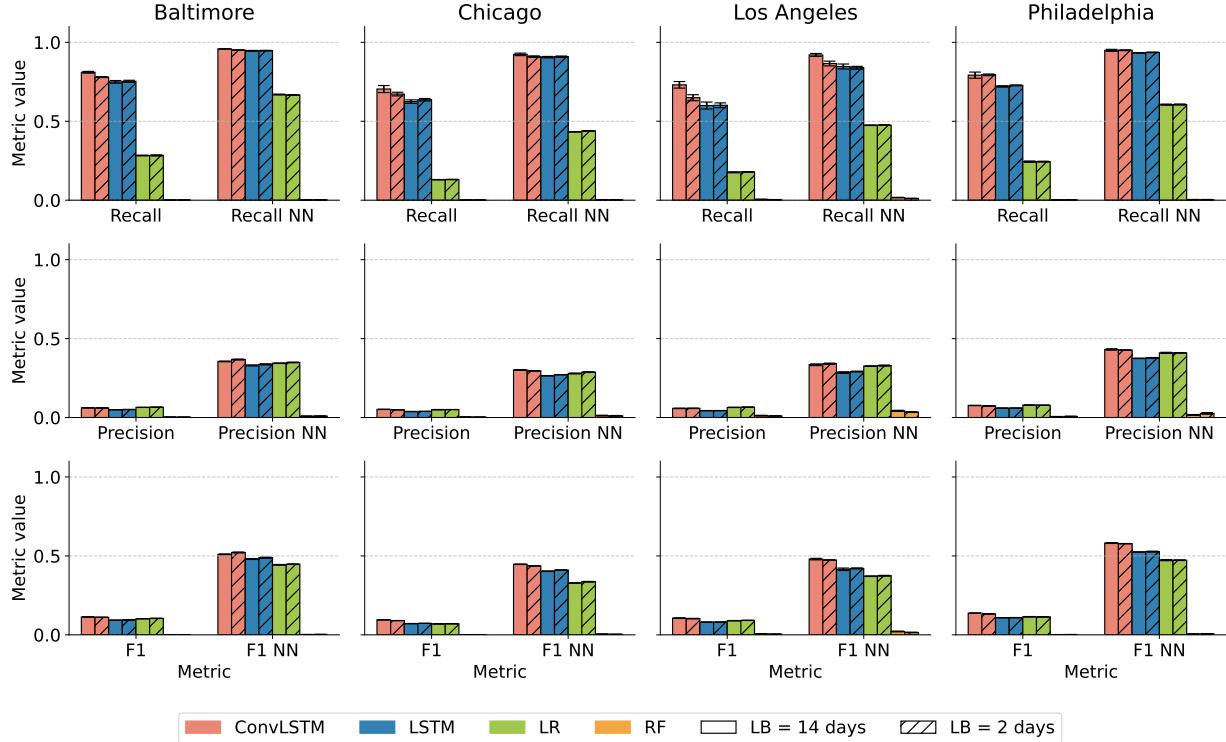


FIGURE 6. Average performance on the test set for our model and three baselines with and without taking into account the NN cells, considering the violent crimes and two different LB periods. Error bars indicate the standard deviation.

using the standard recall and precision, except for the standard recall in Philadelphia. The results are more mixed when we examine the modified metrics instead. However, the F1 score is always slightly higher for the longer LB period, except for the modified value in Baltimore. The same pattern is observed when comparing the 7-day and 1-day LB periods, as shown in Appendix J. These findings suggest that, when using ConvLSTM, violent crimes may often exhibit patterns that benefit from longer timescales.

5.1.3 Property crimes

Lastly, we evaluated the model’s performance on property crimes, specifically burglary and MVT. This configuration had the lowest number of crime incidents, resulting in a highly unbalanced dataset. Nevertheless, property crimes have been shown to be relatively easy to model due to their seasonal trends, dependency on neighborhood characteristics, and correlation with socio-economic factors [83].

The results, presented in Figure 7, indicate a further decline in overall performance compared to violent crimes, which becomes clear when looking at the F1 score values. For the baseline models, we observe that LR consistently achieves again slightly higher precision than LSTM across all cities. However, its standard recall remains very low

(below 0.2). RF continues to be the worst-performing model, generally exhibiting very low recall and precision, although its performance improves in Chicago, Los Angeles, and Philadelphia, compared to its results for violent crimes. LSTM still achieves the highest recall among the baselines but shows slightly worse overall performance (as indicated by the F1 score) compared to its application to all crimes or violent crimes alone. Finally, the ConvLSTM model remains the best-performing model, maintaining strong recall values but showing a decrease in precision—a trend consistent with the baselines.

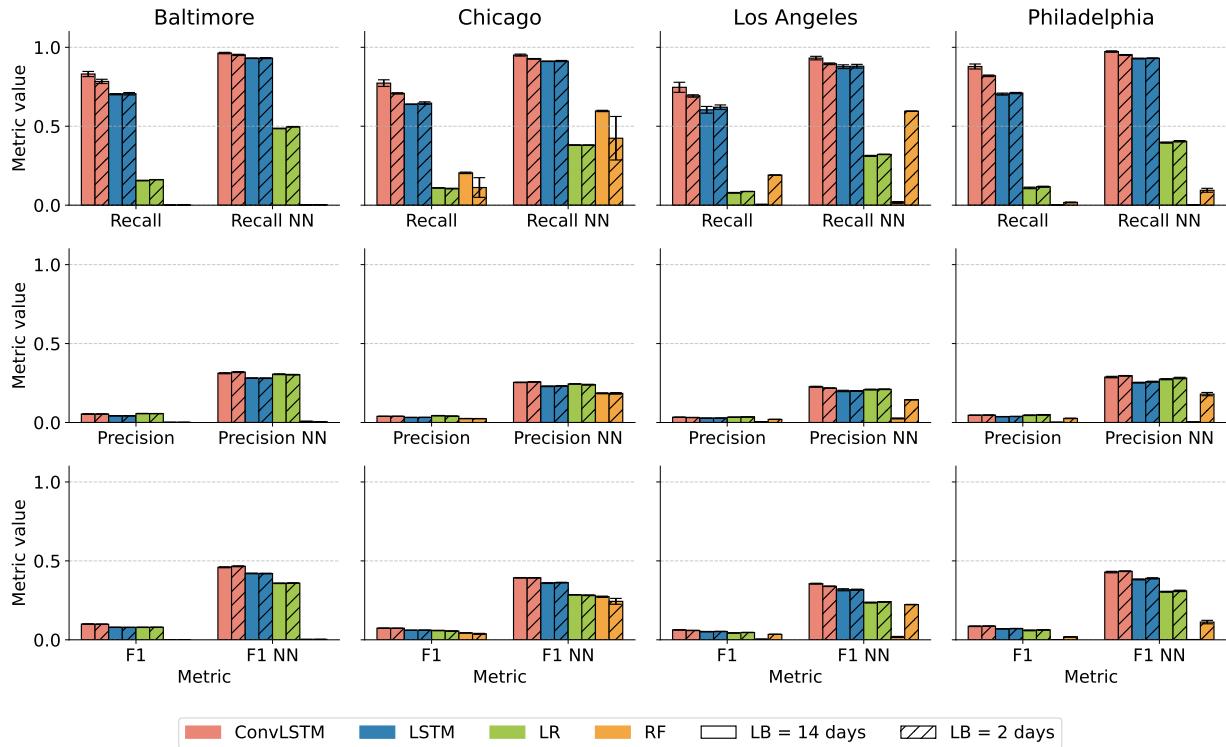


FIGURE 7. Average performance on the test set for our model and three baselines with and without taking into account the NN cells, considering the property crimes and two different LB periods. Error bars indicate the standard deviation.

Therefore, these results indicate that the high data sparsity did not allow the model to learn the patterns in property crime properly when compared to models using all crimes together or only violent crimes.

Lastly, when comparing the values obtained for the two LB periods, we observe a general tendency for the 2-day LB period to achieve slightly better precision than the 14-day period, but lower recall—a pattern also observed for the other two crime aggregations. However, in this case, despite the F1 score indicating that the overall performance is very similar between the two LB periods, the shorter LB performs slightly better when using the modified metric for Baltimore and Philadelphia. This trend is also observed in Chicago, in addition to the two aforementioned cities, when comparing 7-day and 1-day LB periods

(see [Appendix J](#)). These results suggest that the patterns driving property crimes may generally operate over shorter timescales compared to those of violent crimes, although there is some variation depending on the city.

5.2 Effect of mobility features

After evaluating the models' general performance, we sought to assess the impact of including mobility features in our deep learning framework. Thus, we retrained the best-performing model, ConvLSTM, using various feature sets: CMS, CM, CS, and C, where C represents crime features, M represents mobility features, and S represents sociodemographic features. This retraining was conducted using the best configuration alone, hence including all five crime types.

Figure 8 shows the results of the average performance and standard deviation for each of the four cities. If we focus on the results using an LB period of 14 days, the first observation is the considerable variation in standard recall within each city, depending on the feature set used for training. Looking more closely, we see that Chicago and Los Angeles have the lowest performance when focusing on the standard metrics, with the recall always below 0.76 and the precision always below 0.08. The same trend is observed when looking at the modified metrics.

Overall, based on both the standard and modified recall and precision values, the setup incorporating all three feature types consistently performs best, as expected. Specifically, among the 24 configurations seen in Figure 8, the C setup achieves the highest performance in 8.3% of cases, the CM setup in 8.3%, CS in 29.2%, and CMS in 54.2%. Therefore, the results indicate that incorporating additional feature types alongside crime data improves the model's ability to forecast effectively.

Conversely, when analyzing the results with an LB period of 2 days, performance consistently decreases slightly when using the standard metrics. However, Chicago and Los Angeles remain the cities with the lowest performance, and the CMS configuration again emerges as the best overall. In this case, the C setup does not achieve the highest performance in any instance, while the CM setup outperforms others in 12.5% of cases, CS in 29.2%, and CMS in 58.3%. Thus, while shorter LB periods lead to a slight reduction in overall performance, the advantage of incorporating additional feature types alongside crime data becomes more evident.

Afterward, to further understand the impact of mobility features on predictive performance, we calculated the percentage difference between the results of CS and CMS (to assess the effect of incorporating mobility features) and CM and CMS (to evaluate the effect of sociodemographic features) for both LB periods. As shown in Table 4, the results generally indicate a slight decrease in recall when a third feature set is added—both when going from CM to CMS and from CS to CMS—resulting in a negative percentage point difference. In contrast, precision increases in all configurations, with the exception of its modified value for Chicago when using a 14-day LB and Baltimore when using the

2-day LB. Another noteworthy observation is that the LB period of 2 days always shows a greater percentage point increase in overall performance (F1 score) compared to the 14-day period, suggesting that shorter LB periods benefit more from the incorporation of mobility features. Additionally, comparing the F1 scores between CMS-CM and CMS-CS, we can see that sociodemographic features have a greater impact than mobility features in increasing performance; however, both contribute to enhancing the model's ability to forecast crime, with the best performance achieved when they are used together.

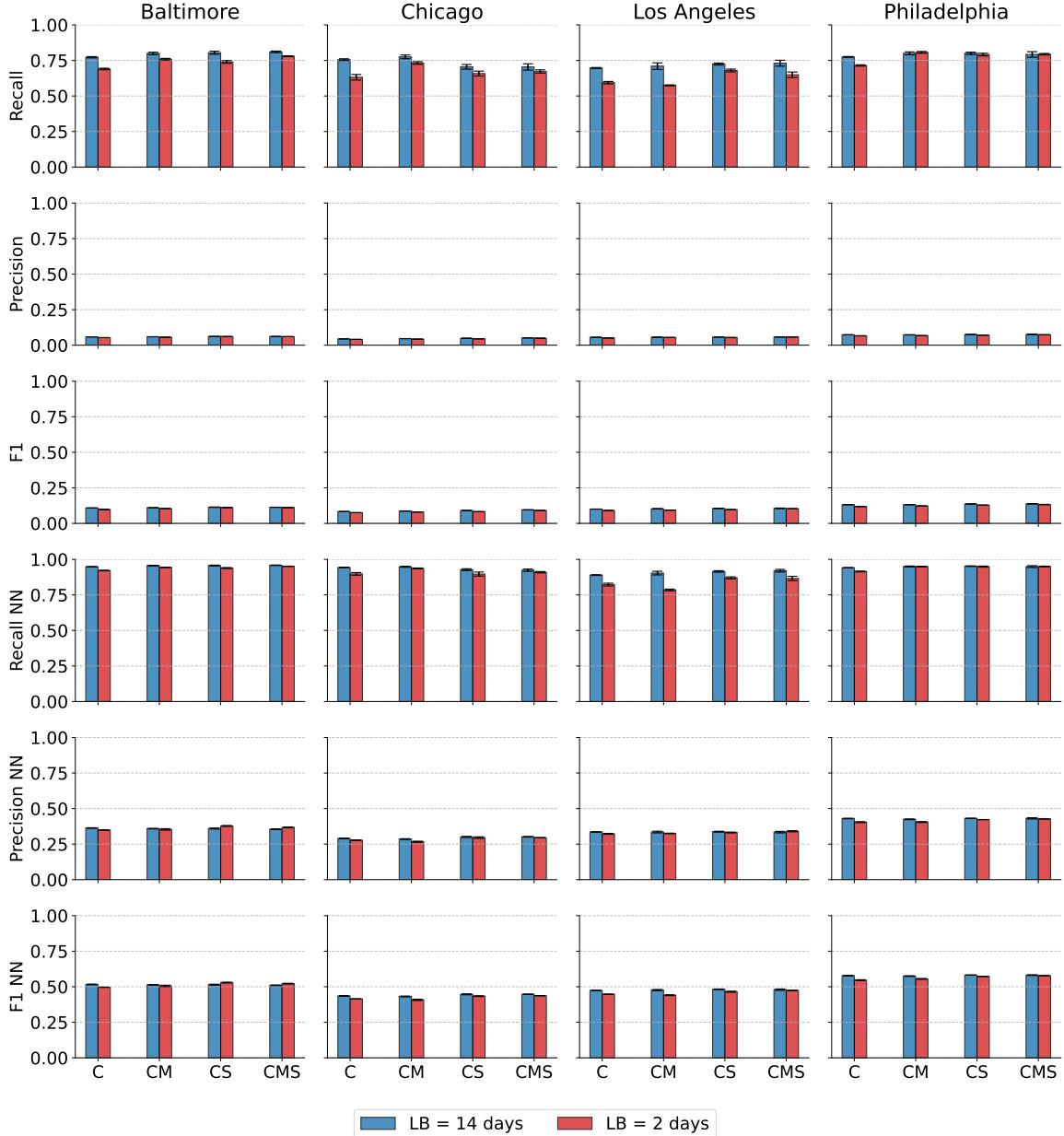


FIGURE 8. Average performance of the ConvLSTM model on the test set, trained using four different feature sets and two different LB periods. Error bars indicate the standard deviation.

LB	Metric	CMS - CM				CMS - CS			
		Bal	Chi	Las	Phi	Bal	Chi	Las	Phi
14 days	Rec	-3.513	-4.743	-0.864	-1.036	-2.157	0.427	-4.920	-0.270
	Prec	8.381	5.606	4.372	4.280	2.689	2.506	3.197	1.630
	F1	6.988	4.372	4.088	3.766	2.077	2.346	2.185	1.453
	Rec NN	-0.986	-1.168	-0.314	-0.206	-0.594	0.068	-1.918	-0.161
	Prec NN	3.266	2.983	1.309	2.076	0.727	-0.102	1.758	0.525
	F1 NN	1.824	1.737	1.065	1.339	0.236	-0.099	0.560	0.283
2 days	Rec	-3.185	-9.972	4.853	-2.698	3.362	-0.923	-2.775	0.123
	Prec	8.640	14.239	5.256	6.390	2.073	6.917	5.612	3.069
	F1	7.164	11.945	5.542	5.175	2.627	6.737	5.323	2.839
	Rec NN	-0.668	-2.875	2.561	-0.606	1.102	0.040	-0.377	0.348
	Prec NN	4.567	7.510	1.929	3.906	-0.243	0.533	1.615	1.353
	F1 NN	2.695	4.482	2.314	2.333	0.404	0.555	1.375	1.105

TABLE 4. Percentage difference in model performance between the CS and CMS configurations (to assess the effect of M), and between the CM and CMS configurations (to assess the effect of S), for each city and using LB periods of 14 and 2 days. Positive values (in green) indicate improved performance, while negative values (in red) indicate a decrease.

Lastly, as a robustness check and to investigate further how long-term versus short-term patterns influence crime prediction, we repeated the analysis in this subsection using two additional LB periods, 7 days and 1 day. The results, which can be found in Appendix K, also show lower overall performance for the shorter LB period and an improvement (or very small decrease) in performance when incorporating a third feature set, regardless of whether it is the mobility or the sociodemographic feature. We also observe that the shorter LB period still has a greater percentage point increase in overall performance (F1 score). However, the relative advantage between 1 and 2 days, as well as between 7 and 14 days, is not clear from this analysis.

6 Discussion and conclusions

Extensive research has explored the spatiotemporal concentration of crime and how resulting patterns can be exploited for crime forecasting [22, 84]. In recent years, thanks also to the vast availability of data, studies in this field have started to explore more advanced computational methods (e.g., machine and deep learning), to investigate a range of questions [1]. Moreover, motivated by several theories highlighting the correlation between human mobility and the dynamics of crime, some studies started incorporating mobility data in combination with more traditional data [23, 37, 38, 85]. Despite this, many questions remain unanswered as current scholarship is characterized by relatively coarse spatial and temporal granularities, very few years of historical data, and/or focus on a single city. This paper aimed to advance the literature on crime forecasting and spatial criminology by offering four main contributions.

Firstly, we used five years of human mobility data, in combination with crime and sociodemographic data from four U.S. cities, to assess the extent to which these features could enhance the prediction of the spatiotemporal occurrence of crime. Secondly, we defined a small geographic unit of analysis of 0.077 sq. miles and a fine temporal granularity of 12 hours in order to assess whether this fine-grained spatiotemporal unit of analysis allows us to meaningfully forecast crime in four different urban contexts in the U.S. To evaluate this, we designed a deep learning model using ConvLSTM layers and compared its performance to three baseline models to determine whether the performance improvement of using an advanced computational method compensated for the complexity involved in building it. Lastly, we compared the results from two different LB periods to assess the impact of longer versus shorter time windows on predictive performance.

The results led to several key findings. We first demonstrated that ConvLSTM consistently outperformed the three baseline models, particularly in recall, but struggled with precision, especially when using standard metrics. LSTM showed the closest performance in both recall and precision, though it remained slightly below ConvLSTM. LR, while competitive in precision, suffered from low recall, and RF failed to capture meaningful patterns across all data configurations. Therefore, on the one hand, the effort involved in building such an elaborate model was worthwhile, as the other models either struggled with high-dimensional data or, as in the case of LSTM, had to be highly complex to achieve comparable performance. Moreover, it should be noted that ConvLSTM's high recall values suggest it can predict most instances where at least one crime is likely to occur, minimizing false negatives as intended. However, the low precision results highlight how the extremely unbalanced nature of our forecasting tasks—a byproduct of the fine-grained spatial and temporal scales of the study—makes this approach highly imperfect in terms of reducing false positives.

Concerning the crime types, we found that our model performed best when using all crime types together, rather than separating them into violent and property crimes. This outcome was expected since this configuration results in a more balanced dataset, hence leading to more stable training. Additionally, the comparison across four different LB periods indicated that violent crimes tend to yield higher overall predictive performance with longer LB periods, whereas property crimes appear to benefit from shorter ones.

When evaluating our model on different sets of features, we were able to determine that the overall best performance was achieved when using all three feature sets together (i.e., crime, mobility, and sociodemographic data), while the worst performance occurred when only historical crime data were used. This suggests that including additional data sources, in combination with crime data, allows the algorithm to learn more effectively the patterns leading to crime occurrence. Moreover, the percentage difference between CS and CMS indicates that mobility data improve performance, especially when using shorter LB periods, stressing how mobility captures aspects of an urban context that transcend its static socio-economic ecology.

These findings not only contribute to those strands of criminological literature con-

cerned with spatiotemporal crime forecasting and the investigation of the link between mobility and crime, but are also relevant to policymakers. In fact, they contribute to the ongoing debate on the challenges associated with the use of computational approaches by law enforcement agencies [3, 86–89]. Although offering an exhaustive, systematic analysis of the limits of deep learning for crime forecasting is beyond the scope of this work, our results underscore that not even extremely sophisticated algorithms can reach high performance in both recall and precision when focusing on fine-grained spatial and temporal scales. This finding—detected for all cities and crime types—bears important implications. Relying on the predictions of our models, for instance, would certainly allow first responders and law enforcement to predict practically all actual offenses, but at the cost of wrongfully targeting certain micro-areas due to high rates of false positives, possibly leading to unintended consequences. While we show that this issue is largely diminished when considering the modified nearest neighbor metrics, the low baseline precision scores speak to the limits that computational models still face when dealing with such a sparse, unbalanced, and volatile phenomenon as crime.

6.1 Limitations

This work, despite its promising results, is not without limitations. First, the use of a grid-based approach, while being very common in the literature (see, among others, [82, 90, 91]) imposes an arbitrary division on our space, and it has been shown that network-based models can outperform grid-based models for crime prediction [92]. Furthermore, an arbitrary division of urban space may also pose a problem, for instance, in terms of administrative or police precinct boundaries, since two points in the same spatial cell might fall under the jurisdiction of two different police precincts. However, a grid-based approach was mandatory given our design focused on the adaptation of computer vision models to crime forecasting.

A second potential limitation is that the crime data we used suffered from location obfuscation due to privacy concerns, which limits how finely we can define our spatial unit of analysis. However, our cell size is still considerably smaller than the typical block size in American cities (see [93] for some estimates), including those analyzed in this research. Additionally, given the low precision obtained with our current grid cell size, we are concerned that further reducing the cell size would exacerbate the problem of false positives. Therefore, we argue that obfuscation does not constitute a major issue in our current design, as it is not as micro in scale.

Finally, arguably the main limitation of this study is that it does not rely on individual trajectories to capture mobility dynamics in the cities under analysis. More nuanced mobility features, such as mobility flows, have been shown to further improve forecasting performance [94] and would certainly have offered a richer set of opportunities to study how crime concentrates and varies over time along with variation in human movements across different parts of the same urban context. Unfortunately, we were unable to obtain

such information due to the limited availability of mobility data for crime research.

6.2 Ethical Implications

This study adheres to ethical research practices, particularly regarding the use of mobility and crime data. The data used in this study do not contain individual-level information; instead, they represent aggregate mobility information at the premise level, further aggregated at the spatial cell level. This fully ensures the privacy and anonymity of individuals while allowing us to study broader patterns of human activity and their relationship with crime.

We also acknowledge the inherent limitations in the reporting of crime data. Certain crimes may be underreported in specific areas of a city due to socioeconomic, cultural, or institutional factors, potentially introducing biases into the analysis. Nonetheless, we have sought to design our framework focusing on those crimes that are less likely to suffer from this issue, as well as crimes that are less prone to be the subject of specific policing decision-making processes, such as those concerning drug-related offenses.

Additionally, our model's statistical results, particularly its low precision, raise significant concerns that could have ethical implications if deployed in real-world scenarios. Relying on these results to allocate police resources risks unjustified overpolicing certain areas, as false positives are pervasive across cities and crime types, possibly exacerbating existing inequalities or disproportionately impacting specific communities.

We therefore emphasize that the results of this study are intended to advance the theoretical understanding of crime and to contribute to the development of methodological approaches in spatiotemporal crime forecasting, rather than to serve as direct recommendations for enforcement actions. Future research should focus on addressing these ethical concerns, improving model precision, and incorporating diverse stakeholder perspectives to ensure fair and equitable algorithmic design.

Acknowledgments

Anonymized for submission. We thank Laia Albors Zumel for comments and suggestions regarding earlier versions of this manuscript and Giuseppe Veltri for partial financial support in the first phase of this research.

Author Contributions

This work is the result of the joint efforts by all authors, A.A.Z., M.T., and G.M.C. A.A.Z., M.T., and G.M.C. jointly contributed to design the study and to its methodological setup, A.A.Z. performed the coding and modeling part, A.A.Z. and G.M.C. both contributed to the analysis of the results, A.A.Z. wrote the first draft of the paper and M.T. and G.M.C.

reviewed and refined it. M.T. contributed funding to acquire the mobility data. G.M.C. supervised the project.

Competing Interests

The authors declare that they have no financial or non-financial interests, either directly or indirectly, related to the content of this work.

Data and Code Availability

The code and links to the raw data used in this study have been deposited in Zenodo and are accessible via DOI: 10.5281/zenodo.16042193.

References

- [1] G. M. Campedelli, *Machine Learning for Criminology and Crime Research: At the Crossroads*. Routledge Advances in Criminology, Taylor & Francis, 2022.
- [2] W. L. Perry, B. McInnis, C. C. Price, S. Smith, and J. S. Hollywood, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Santa Monica, CA: RAND Corporation, 2013.
- [3] R. Berk, "Artificial intelligence, predictive policing, and risk assessment for law enforcement," *Annual Review of Criminology*, vol. 4, pp. 209–237, 2021. First published as a Review in Advance on November 13, 2020.
- [4] F. Cullen and P. Wilcox, "Quetelet, adolphe: Explaining crime through statistical and cartographic techniques," 2010.
- [5] L. Turnbull, E. Hendrix, and B. Dent, *Atlas of Crime: Mapping the Criminal Landscape*. Bloomsbury Academic, 2000.
- [6] S. Chamarad, *History of Crime Mapping and Its Use by American Police Departments*, vol. 23. Alaska Justice Forum, 2006.
- [7] L. W. Sherman, P. R. Gartin, and M. Buerger, "Hot spots of predatory crime: Routine activities and the criminology of place," *Criminology*, vol. 27, pp. 27–56, 1989.
- [8] D. Weisburd, "The law of crime concentration and the criminology of place*," *Criminology*, vol. 53, no. 2, pp. 133–157, 2015.

- [9] L. W. Sherman and D. Weisburd, "General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial," *Justice Quarterly*, vol. 12, no. 4, pp. 625–648, 1995.
- [10] G. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 100–108, 2011.
- [11] S. van Sleeuwen, W. Steenbeek, and S. Ruiter, "When do offenders commit crime? an analysis of temporal consistency in individual offending patterns," *Journal of Quantitative Criminology*, vol. 37, 12 2021.
- [12] M. Andresen and N. Malleson, "Intra-week spatial-temporal patterns of crime," *Crime Science*, vol. 4, 2015.
- [13] P. Sharkey, M. Besbris, and M. Friedson, "623Poverty and Crime," in *The Oxford Handbook of the Social Science of Poverty*, Oxford University Press, 05 2016.
- [14] R. Jarjoura, R. Triplett, and G. Brinker, "Growing up poor: Examining the link between persistent childhood poverty and delinquency," *Journal of Quantitative Criminology*, vol. 18, pp. 159–187, 06 2002.
- [15] C. Graif and S. Matthews, "The long arm of poverty: Extended and relational geographies of child victimization and neighborhood violence exposures," *Justice Quarterly (online Jan 2017)*, vol. 34, 01 2017.
- [16] R. Berk, K. Lenihan, and P. Rossi, "Crime and poverty: Some experimental evidence from ex-offenders," *American Sociological Review*, vol. 45, p. 766, 10 1980.
- [17] J. Legewie and M. Schaeffer, "Contested boundaries: Explaining where ethnoracial diversity provokes neighborhood conflict," *American Journal of Sociology*, vol. 122, no. 1, pp. 125–161, 2016.
- [18] S. Olzak, *The Dynamics of Ethnic Competition and Conflict*. Stanford, CA: Stanford University Press, 1992.
- [19] L. E. Cohen and M. Felson, "Social Change and Crime Rate Trends: A Routine Activity Approach," *American Sociological Review*, vol. 44, no. 4, pp. 588–608, 1979.
- [20] R. J. Sampson and W. B. Groves, "Community Structure and Crime: Testing Social-Disorganization Theory," *American Journal of Sociology*, vol. 94, no. 4, pp. 774–802, 1989.
- [21] Y. Du and N. Ding, "A Systematic Review of Multi-Scale Spatio-Temporal Crime Prediction Methods," *ISPRS International Journal of Geo-Information*, vol. 12, no. 6, 2023.

- [22] O. Kounadi, A. Ristea, A. de Araujo, and M. Leitner, "A systematic review on spatial crime forecasting," *Crime Science*, vol. 9, 05 2020.
- [23] J. Wu, S. M. Abrar, N. Awasthi, E. Frias-Martinez, and V. Frías-Martínez, "Enhancing short-term crime prediction with human mobility flows and deep learning architectures," *EPJ Data Science*, vol. 11, p. 53, 11 2022.
- [24] A. Stec and D. Klabjan, "Forecasting crime with deep learning," 2018.
- [25] V. Mandalapu, L. Elluri, P. Vyas, and N. Roy, "Crime prediction using machine learning and deep learning: A systematic review and future directions," *IEEE Access*, vol. 11, p. 60153–60170, 2023.
- [26] A. Braga, M. Andresen, and B. Lawton, "The law of crime concentration at places: Editors' introduction," *Journal of Quantitative Criminology*, vol. 33, 03 2017.
- [27] P. Bauchet and V. Pareto, "Cours d'économie politique," *Revue économique*, vol. 16, p. 811, 09 1965.
- [28] R. J. Skiba and M. K. Rausch, "Zero tolerance, suspension, and expulsion: Questions of equity and effectiveness," in *Handbook of Classroom Management: Research, Practice, and Contemporary Issues* (C. M. Evertson and C. S. Weinstein, eds.), pp. 1063–1089, New York: Routledge, 2006.
- [29] D. J. Watts and P. S. Dodds, "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, vol. 34, pp. 441–458, 05 2007.
- [30] T. Piketty and A. Goldhammer, *Capital in the Twenty-First Century*. Harvard University Press, 2014.
- [31] T. Thornberry and M. Krohn, "Taking stock of delinquency, an overview of findings from contemporary longitudinal studies," 01 2003.
- [32] A. Antipova, *City Structure and Spatial Patterns*, pp. 153–204. Cham: Springer International Publishing, 2018.
- [33] P. Brantingham and P. Brantingham, *Environment, Routine, and Situation: Toward a Pattern Theory of Crime: Advances in Criminological Theory*, vol. 5, pp. 259–294. Routine Activity and Rational Choice, 10 1993.
- [34] M. Andresen, *Environmental Criminology: Evolution, Theory, and Practice*. 03 2014.
- [35] A. Piquero, *The Handbook of Criminological Theory*. Wiley Handbooks in Criminology and Criminal Justice, Wiley, 2015.

- [36] A. Bogomolov, B. Lepri, J. Staiano, E. Letouzé, N. Oliver, F. Pianesi, and A. Pentland, "Moves on the street: Classifying crime hotspots using aggregated anonymized data on people dynamics," *Big Data*, vol. 3, no. 3, pp. 148–158, 2015.
- [37] C. Kadar and I. Pletikosa, "Mining large-scale human mobility data for long-term crime prediction," *EPJ Data Science*, vol. 7, 12 2018.
- [38] M. De Nadai, Y. Xu, E. Letouzé, M. C. Gonzalez, and B. Lepri, "Socio-economic, built environment, and mobility conditions associated with crime: A study of multiple cities," *Scientific Reports*, vol. 1, 4 2020.
- [39] M. Luca, G. M. Campedelli, S. Centellegher, M. Tizzoni, and B. Lepri, "Crime, inequality and public health: a survey of emerging trends in urban data science," *Frontiers in Big Data*, vol. 6, p. 1124526, 2023.
- [40] N. Smelser and P. Baltes, *International Encyclopedia of Social & Behavioral Sciences*. International Encyclopedia of the Social & Behavioral Sciences, Elsevier Science, 2001.
- [41] P. Bellair, "Social disorganization theory," 07 2017.
- [42] R. Wickes and M. Sydes, "Social disorganization theory," 11 2017.
- [43] L. Duan, T. Hu, E. Cheng, J. Zhu, and C. Gao, "Deep Convolutional Neural Networks for Spatiotemporal Crime Prediction," p. 61–67, 2017.
- [44] E. Cesario, P. Lindia, and A. Vinci, "Multi-density crime predictor: an approach to forecast criminal activities in multi-density crime hotspots," *Journal of Big Data*, vol. 11, 05 2024.
- [45] X. Zhao, W. Fan, H. Liu, and J. Tang, "Multi-type urban crime prediction," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 4388–4396, 06 2022.
- [46] A. Wheeler and W. Steenbeek, "Mapping the risk terrain for crime using machine learning," *Journal of Quantitative Criminology*, vol. 37, 06 2021.
- [47] V. Rotaru, Y. Huang, T. Li, J. Evans, and I. Chattopadhyay, "Event-level prediction of urban crime reveals a signature of enforcement bias in us cities," *Nature Human Behaviour*, vol. 6, pp. 1–13, 08 2022.
- [48] H.-W. Kang and H.-B. Kang, "Prediction of crime occurrence from multi-modal data using deep learning," *PLOS ONE*, vol. 12, p. e0176244, 04 2017.
- [49] G. Drawve, S. C. Moak, and E. R. Berthelot, "Predictability of gun crimes: A comparison of hot spot and risk terrain modelling techniques," *Policing & Society*, vol. 26, no. 3, pp. 312–331, 2016.

- [50] Y. Zhuang, M. Almeida, M. Morabito, and W. Ding, "Crime hot spot forecasting: A recurrent model with spatial and temporal information," in *IEEE International Conference on Big Knowledge (ICBK)* (X. D. Wu, T. Ozu, J. Hendler, and R. Lu, eds.), pp. 143–150, 2017.
- [51] S. Wang and K. Yuan, "Spatiotemporal analysis and prediction of crime events in atlanta using deep learning," pp. 346–350, 07 2019.
- [52] L. Xia, C. Huang, Y. Xu, P. Dai, L. Bo, X. Zhang, and T. Chen, "Spatial-temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning," pp. 1631–1637, 08 2021.
- [53] M.-M. Magnusson, "Predicting gun violence in stockholm, sweden, using sociodemographics, crime and drug market locations," *European Journal on Criminal Policy and Research*, pp. 1–22, 08 2023.
- [54] Y.-L. Lin, M.-F. Yen, and L.-C. Yu, "Grid-based crime prediction using geographical features," *ISPRS International Journal of Geo-Information*, vol. 7, p. 298, 07 2018.
- [55] B. Chen and Y. Liao, "Spatio-temporal deep fusion graph convolutional networks for crime prediction," in *Proceedings of the 2023 7th International Conference on Machine Learning and Soft Computing, ICMLSC '23*, (New York, NY, USA), p. 75–81, Association for Computing Machinery, 2023.
- [56] W. Liang, Y. Wang, H. Tao, and J. Cao, "Towards hour-level crime prediction: A neural attentive framework with spatial-temporal-categorical fusion," *Neurocomputing*, vol. 486, 11 2021.
- [57] C. R. Browning, N. P. Pinchak, and C. A. Calder, "Human Mobility and Crime: Theoretical Approaches and Novel Data Collection Strategies," *Annual Review of Criminology*, vol. 4, pp. 99–123, Jan. 2021.
- [58] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, "Computational Social Science," *Science*, vol. 323, pp. 721–723, Feb. 2009.
- [59] R. Rosés, C. Kadar, C. Gerritsen, and O. C. Rouly, "Simulating Offender Mobility: Modeling Activity Nodes from Large-Scale Human Activity Data," *Journal of Artificial Intelligence Research*, vol. 68, pp. 541–570, July 2020.
- [60] H. Wang, D. Kifer, C. Graif, and Z. Li, "Crime Rate Inference with Big Data," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, (New York, NY, USA), pp. 635–644, Association for Computing Machinery, Aug. 2016.

- [61] C. Huang, J. Zhang, Y. Zheng, and N. V. Chawla, "DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, (New York, NY, USA), pp. 1423–1432, Association for Computing Machinery, Oct. 2018.
- [62] N. Malleson and M. A. Andresen, "Spatio-temporal crime hotspots and the ambient population," *Crime Science*, vol. 4, p. 10, May 2015.
- [63] J. R. Hipp, B. Christopher, L. Moshe, and P. Smyth, "Using Social Media to Measure Temporal Ambient Population: Does it Help Explain Local Crime Rates?," *Justice Quarterly*, vol. 36, pp. 718–748, June 2019. Publisher: Routledge _eprint: <https://doi.org/10.1080/07418825.2018.1445276>.
- [64] A. D. Biderman and A. J. Reiss, "On exploring the "dark figure" of crime," *The ANNALS of the American Academy of Political and Social Science*, vol. 374, no. 1, pp. 1–15, 1967.
- [65] N.-J. Akpinar, M. De-Arteaga, and A. Chouldechova, "The effect of differential victim crime reporting on predictive policing systems," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, (New York, NY, USA), p. 838–849, Association for Computing Machinery, 2021.
- [66] J. Wu, E. Frias-Martinez, and V. Frias-Martinez, "Addressing under-reporting to enhance fairness and accuracy in mobility-based crime prediction," in *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '20, (New York, NY, USA), p. 325–336, Association for Computing Machinery, 2020.
- [67] Advan, "Location Data For Finance & Commercial Real Estate." <https://advanresearch.com/>, 2023. Accessed: 25-05-2023.
- [68] Advan Research, "Foot traffic / weekly patterns [dataset]," 2022.
- [69] P. Buonanno, "The socioeconomic determinants of crime. a review of the literature," no. 63, 2003.
- [70] S. N. Z. Hamzah and E. Lau, "The role of social factors in explaining crime," *Theoretical and Applied Economics*, vol. 0, pp. 99–118, 6 2013.
- [71] R. Flowers, *Demographics and Criminality: The Characteristics of Crime in America*. Contributions in criminology and penology, Greenwood Press, 1989.
- [72] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. kin Wong, and W. chun Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, 2015.

- [73] G. Nearing, D. Cohen, V. Dube, M. Gauch, O. Gilon, S. Harrigan, A. Hassidim, D. Klotz, F. Kratzert, A. Metzger, S. Nevo, F. Pappenberger, C. Prudhomme, G. Shalev, S. Shenzis, T. Tekalign, D. Weitzner, and Y. Matias, "Global prediction of extreme floods in ungauged watersheds," *Nature*, vol. 627, pp. 559–563, 03 2024.
- [74] A. Graves, A. rahman Mohamed, and G. E. Hinton, "Speech recognition with deep recurrent neural networks," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6645–6649, 2013.
- [75] J. G. Lee, S. Moon, and K. Salamatian, "Modeling and predicting the popularity of online contents with cox proportional hazard regression model," *Neurocomputing*, vol. 76, no. 1, pp. 134–145, 2012. Seventh International Symposium on Neural Networks (ISNN 2010) Advances in Web Intelligence.
- [76] J. Zhao, F. Deng, Y. Cai, and J. Chen, "Long short-term memory - fully connected (lstm-fc) neural network for pm2.5 concentration prediction," *Chemosphere*, vol. 220, pp. 486–492, 2019.
- [77] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [78] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020.
- [79] B. Cortez, B. Carrera, Y.-J. Kim, and J.-Y. Jung, "An architecture for emergency event prediction using LSTM recurrent neural networks," *Expert Systems with Applications*, vol. 97, pp. 315–324, 2018.
- [80] Z. Liang, R. Sun, and Q. Duan, "Attribution of vegetation dynamics in the yellow river water conservation area based on the deep convlstm model," *Remote Sensing*, vol. 16, p. 3875, 10 2024.
- [81] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A hybrid deep learning model with attention-based conv-lstm networks for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, pp. 1–11, 06 2020.
- [82] G. Mohler, "Marked point process hotspot maps for homicide and gun crime prediction in chicago," *International Journal of Forecasting*, vol. 30, pp. 491–497, July–September 2014.
- [83] S. J. Linning, M. A. Andresen, and P. J. Brantingham, "Crime seasonality: Examining the temporal fluctuations of property crime in cities with varying climates," *International journal of offender therapy and comparative criminology*, vol. 61, no. 16, pp. 1866–1891, 2017.

- [84] S. D. Johnson, "A brief history of the analysis of crime concentration," *European Journal of Applied Mathematics*, vol. 21, no. 4-5, p. 349–370, 2010.
- [85] C. Graif, B. N. Freelin, Y.-H. Kuo, H. Wang, Z. Li, and D. Kifer, "Network spillovers and neighborhood crime: A computational statistics analysis of employment-based networks of neighborhoods," *Justice Quarterly*, vol. 38, no. 2, pp. 344–374, 2021. Received 13 Feb 2018, Accepted 27 Mar 2019, Published online: 26 Apr 2019.
- [86] K. Lum and W. Isaac, "To predict and serve?," *Significance*, vol. 13, no. 5, pp. 14–19, 2016. First published: 07 October 2016.
- [87] S. Brayne, "Big data surveillance: The case of policing," *American Sociological Review*, vol. 82, no. 5, pp. 977–1008, 2017. First published online August 29, 2017.
- [88] P. J. Brantingham, M. Valasik, and G. O. Mohler, "Does predictive policing lead to biased arrests? results from a randomized controlled trial," *Statistics and Public Policy*, vol. 5, no. 1, pp. 1–6, 2018. Published online: 09 Apr 2018, Received 01 May 2017, Accepted 01 Feb 2018.
- [89] A. Meijer and M. Wessels, "Predictive policing: Review of benefits and drawbacks," *International Journal of Public Administration*, vol. 42, no. 12, pp. 1031–1039, 2019. Published online: 12 Feb 2019.
- [90] A. Rummens, W. Hardyns, and L. Pauwels, "The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context," *Applied Geography*, vol. 86, pp. 255–261, September 2017.
- [91] X. Zhao and J. Tang, "Modeling temporal-spatial correlations for crime prediction," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM'17)*, (New York, NY, USA), pp. 497–506, ACM, 2017.
- [92] G. Rosser, T. Davies, K. Bowers, S. Johnson, and T. Cheng, "Predictive crime mapping: Arbitrary grids or street networks?," *Journal of Quantitative Criminology*, vol. 33, pp. 1–26, 09 2017.
- [93] M. D. Major, "The invention of a new scale – the paradox of size and configuration in american cities," *Journal of Space Syntax*, vol. 6, pp. 170–191, October 2015.
- [94] C. Kadar, S. Feuerriegel, A. Noulas, and C. Mascolo, "Leveraging mobility flows from location technology platforms to test crime pattern theory in large cities," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, pp. 339–350, 05 2020.

Appendix A Crime distribution

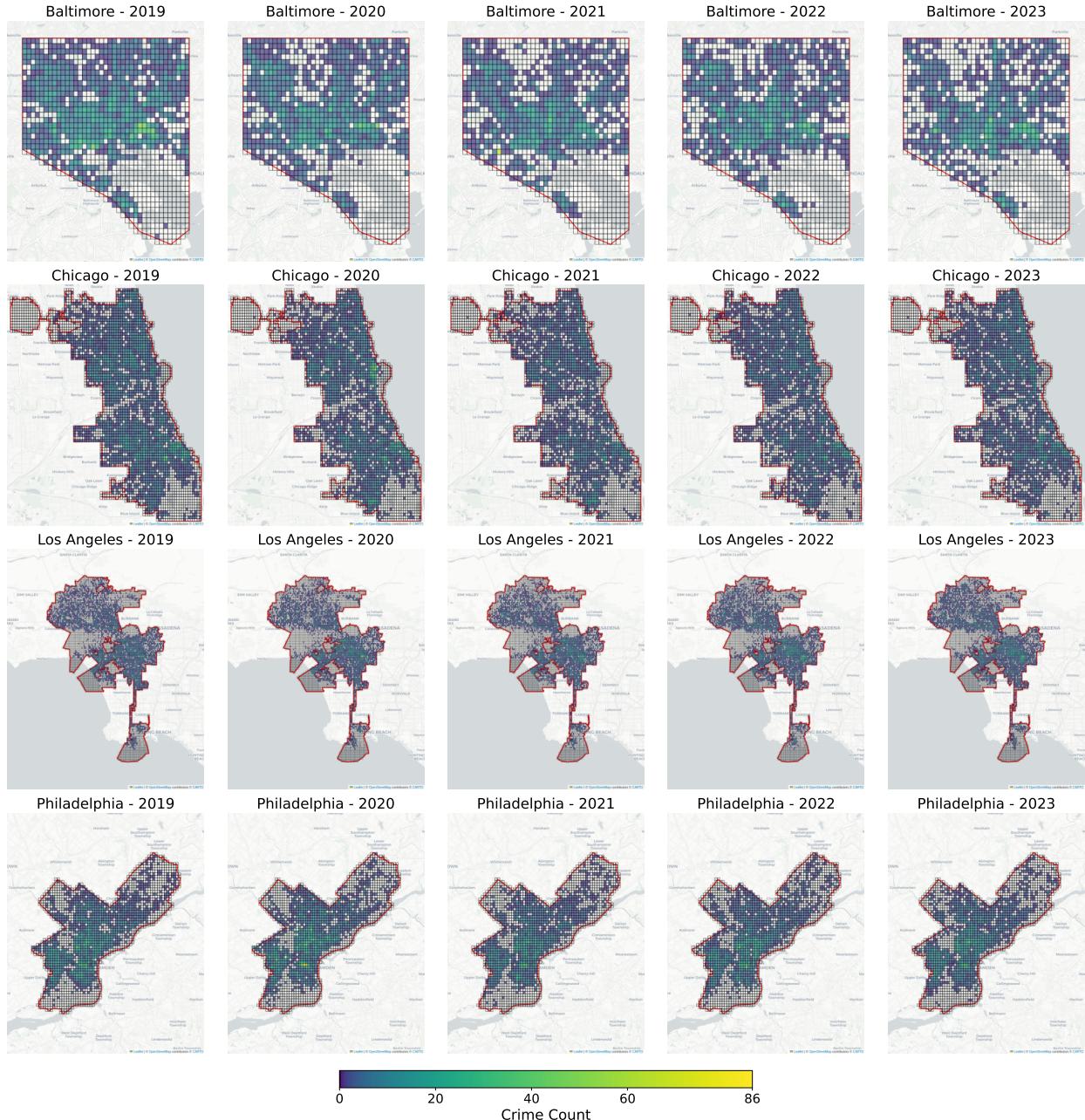


FIGURE 9. Map of the crime distribution for each of the four U.S. cities for each of the years between 2019 and 2023.

Appendix B People-to-POI ratio

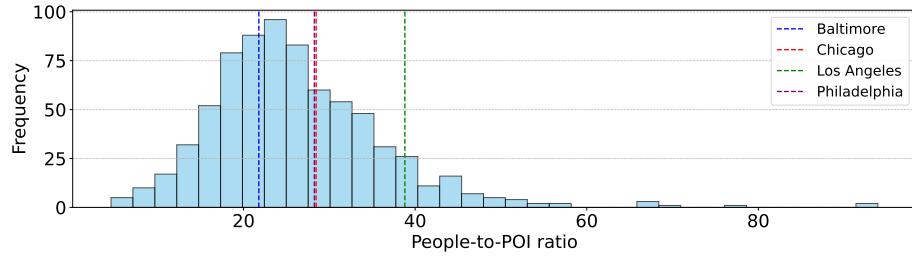


FIGURE 10. Histogram of the distribution of the people-to-POI ratio for the 735 cities in the U.S., with vertical lines indicating the position of our four cities within this distribution.

Appendix C POI distribution

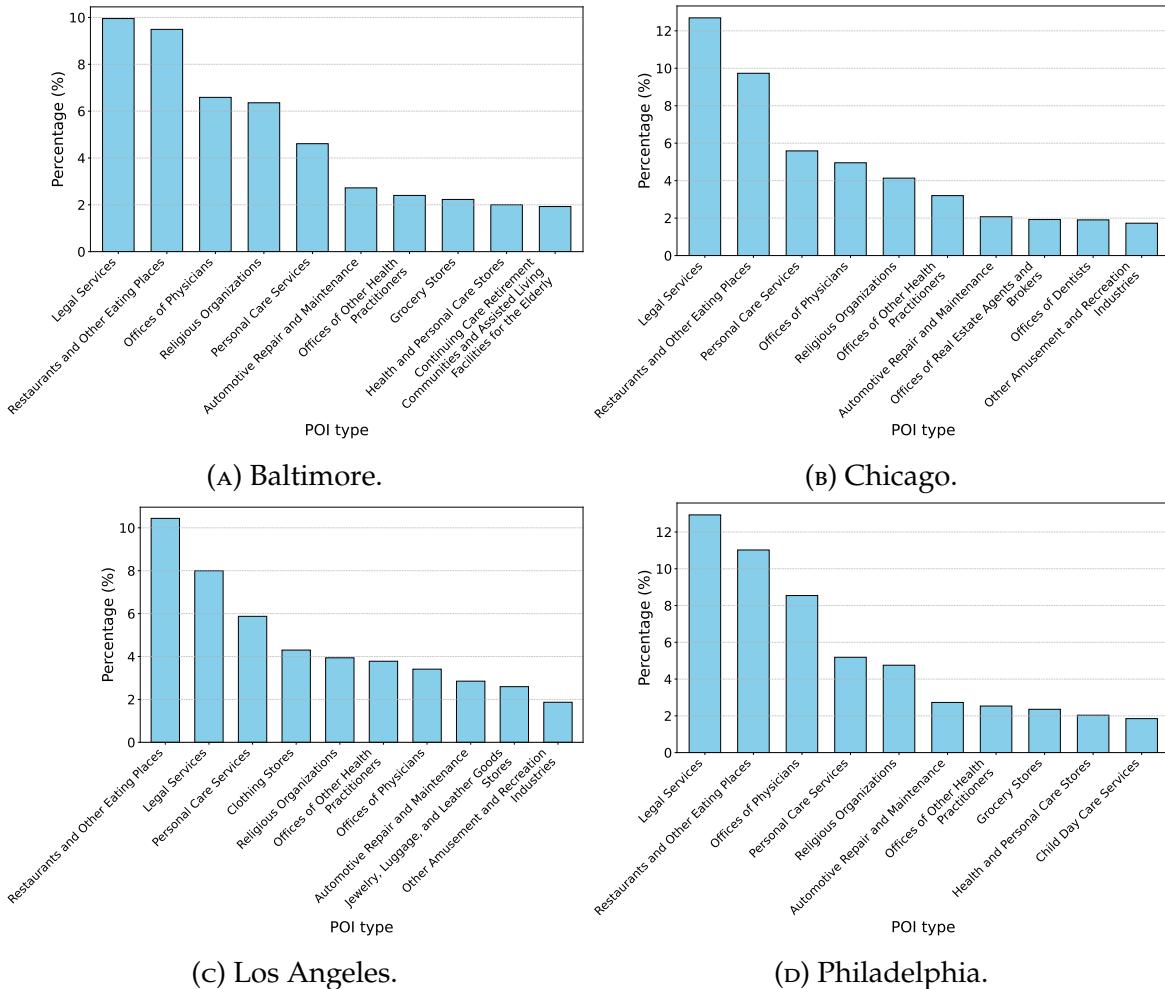


FIGURE 11. Top 10 POI types by percentage for each of the four U.S. cities between 2019 and 2023.

Appendix D POI Categories

1. **Utilities and Construction:** "Building Material and Supplies Dealers", "Foundation, Structure, and Building Exterior Contractors", "Building Equipment Contractors", "Building Finishing Contractors", "Other Specialty Trade Contractors", "Utility System Construction", and "Electric Power Generation, Transmission and Distribution".
2. **Manufacturing:** "Bakeries and Tortilla Manufacturing", "Beverage Manufacturing", "Other Miscellaneous Manufacturing", "Glass and Glass Product Manufacturing", "Other Wood Product Manufacturing", "Metalworking Machinery Manufacturing", "Steel Product Manufacturing from Purchased Steel", "Greenhouse, Nursery, and Floriculture Production", and "Printing and Related Support Activities".
3. **Retail and Wholesale trade:** "Shoe Stores", "Grocery Stores", "Clothing Stores", "Electronics and Appliance Stores", "Health and Personal Care Stores", "Gasoline Stations", "Used Merchandise Stores", "Automotive Parts, Accessories, and Tire Stores", "Jewelry, Luggage, and Leather Goods Stores", "Beer, Wine, and Liquor Stores", "Specialty Food Stores", "Home Furnishings Stores", "Other Motor Vehicle Dealers", "Office Supplies, Stationery, and Gift Stores", "General Merchandise Stores, including Warehouse Clubs and Supercenters", "Department Stores", "Book Stores and News Dealers", "Automobile Dealers", "Furniture Stores", "Lawn and Garden Equipment and Supplies Stores", "Hardware, and Plumbing and Heating Equipment and Supplies Merchant Wholesalers", "Machinery, Equipment, and Supplies Merchant Wholesalers", "Drugs and Druggists' Sundries Merchant Wholesalers", "Chemical and Allied Products Merchant Wholesalers", "Petroleum and Petroleum Products Merchant Wholesalers", "Motor Vehicle and Motor Vehicle Parts and Supplies Merchant Wholesalers", "Miscellaneous Durable Goods Merchant Wholesalers", "Grocery and Related Product Merchant Wholesalers", "Household Appliances and Electrical and Electronic Goods Merchant Wholesalers", "Lumber and Other Construction Materials Merchant Wholesalers", "Direct Selling Establishments", "Other Miscellaneous Store Retailers", "Professional and Commercial Equipment and Supplies Merchant Wholesalers", and "Sporting Goods, Hobby, and Musical Instrument Stores".
4. **Transportation and warehousing:** "Support Activities for Air Transportation", "Specialized Freight Trucking", "Rail Transportation", "Taxi and Limousine Service", "Other Transit and Ground Passenger Transportation", "Transit and Ground Passenger Transportation", "Scenic and Sightseeing Transportation", "Support Activities for Road Transportation", "Freight Transportation Arrangement", "Support Activities for Water Transportation", "Warehousing and Storage", "Automotive Equipment Rental and Leasing", "Commercial and Industrial Machinery and Equipment Rental and Leasing", "Interurban and Rural Bus Transportation", and

“Postal Service”.

5. **Business and Professional Services:** “Investigation and Security Services”, “Activities Related to Credit Intermediation”, “Offices of Real Estate Agents and Brokers”, “Other Professional, Scientific, and Technical Services”, “Accounting, Tax Preparation, Bookkeeping, and Payroll Services”, “Management, Scientific, and Technical Consulting Services”, “Advertising, Public Relations, and Related Services”, “Legal Services”, “Data Processing, Hosting, and Related Services”, “Architectural, Engineering, and Related Services”, “Specialized Design Services”, “Business Support Services”, “Employment Services”, “Travel Arrangement and Reservation Services”, “Freight Transportation Arrangement”, “Management of Companies and Enterprises”, “Activities Related to Real Estate”, “Agencies, Brokerages, and Other Insurance Related Activities”, “Cable and Other Subscription Programming”, “Consumer Goods Rental”, “Depository Credit Intermediation”, “General Rental Centers”, “Insurance Carriers”, “Lessors of Real Estate”, “Motion Picture and Video Industries”, “Nondepository Credit Intermediation”, “Other Financial Investment Activities”, “Other Information Services”, “Radio and Television Broadcasting”, “Sound Recording Industries”, and “Wired and Wireless Telecommunications Carriers”.
6. **Educational services:** “Elementary and Secondary Schools”, “Colleges, Universities, and Professional Schools”, “Junior Colleges”, “Technical and Trade Schools”, “Other Schools and Instruction”, and “Educational Support Services”.
7. **Health care and social assistance:** “Offices of Physicians”, “Offices of Other Health Practitioners”, “Specialty (except Psychiatric and Substance Abuse) Hospitals”, “General Medical and Surgical Hospitals”, “Psychiatric and Substance Abuse Hospitals”, “Outpatient Care Centers”, “Nursing Care Facilities (Skilled Nursing Facilities)”, “Continuing Care Retirement Communities and Assisted Living Facilities for the Elderly”, “Home Health Care Services”, “Individual and Family Services”, “Community Food and Housing, and Emergency and Other Relief Services”, “Residential Intellectual and Developmental Disability, Mental Health, and Substance Abuse Facilities”, “Child Day Care Services”, “Medical and Diagnostic Laboratories”, “Nursing and Residential Care Facilities”, “Offices of Dentists”, and “Other Ambulatory Health Care Services”.
8. **Arts, entertainment, and recreation:** “Museums, Historical Sites, and Similar Institutions”, “Amusement Parks and Arcades”, “Spectator Sports”, “Other Amusement and Recreation Industries”, “Performing Arts Companies”, “Promoters of Performing Arts, Sports, and Similar Events”, “Social Advocacy Organizations”, “Civic and Social Organizations”, and “Gambling Industries”.

9. **Accommodation and food services:** "Traveler Accommodation", "Special Food Services", "Restaurants and Other Eating Places", "Drinking Places (Alcoholic Beverages)", and "RV (Recreational Vehicle) Parks and Recreational Camps".
10. **Public administration:** "Justice, Public Order, and Safety Activities", "Administration of Economic Programs", "Administration of Human Resource Programs", and "National Security and International Affairs".
11. **Other services:** "Florists", "Other Personal Services", "Religious Organizations", "Personal and Household Goods Repair and Maintenance", "Drycleaning and Laundry Services", "Death Care Services", "Personal Care Services", "Social Assistance", "Grantmaking and Giving Services", "Couriers and Express Delivery Services", "Waste Management and Remediation Services", "Remediation and Other Waste Management Services", "Services to Buildings and Dwellings", "Waste Treatment and Disposal", "Waste Collection", "Automotive Repair and Maintenance", and "Electronic and Precision Equipment Repair and Maintenance".

Appendix E Sociodemographic Variables

Gender:

1. Percentage of females.

Age:

2. Median age of males.
3. Median age of females.

Race/ethnicity:

4. Percentage of white alone.
5. Percentage of black or African American.
6. Percentage of American Indian and Alaskan native.
7. Percentage of Asian alone.
8. Percentage of native Hawaiian and other.
9. Percentage of some other race alone.

Employment (population over 16 years old):

10. Percentage employed.
11. Percentage unemployed.

12. Percentage in the armed forces.
13. Percentage not in the labor force.

Income:

14. Median household income in the past 12 months.

Education (population over 25 years old):

15. Percentage with no schooling.
16. Percentage with a regular high school diploma.
17. Percentage with a Bachelor's degree.
18. Percentage with a professional school degree.

Marital status:

19. Percentage of males never married.
20. Percentage of males currently married.
21. Percentage of males widowed.
22. Percentage of males divorced.
23. Percentage of females never married.
24. Percentage of females currently married.
25. Percentage of females widowed.
26. Percentage of females divorced.

Appendix F POI diversity

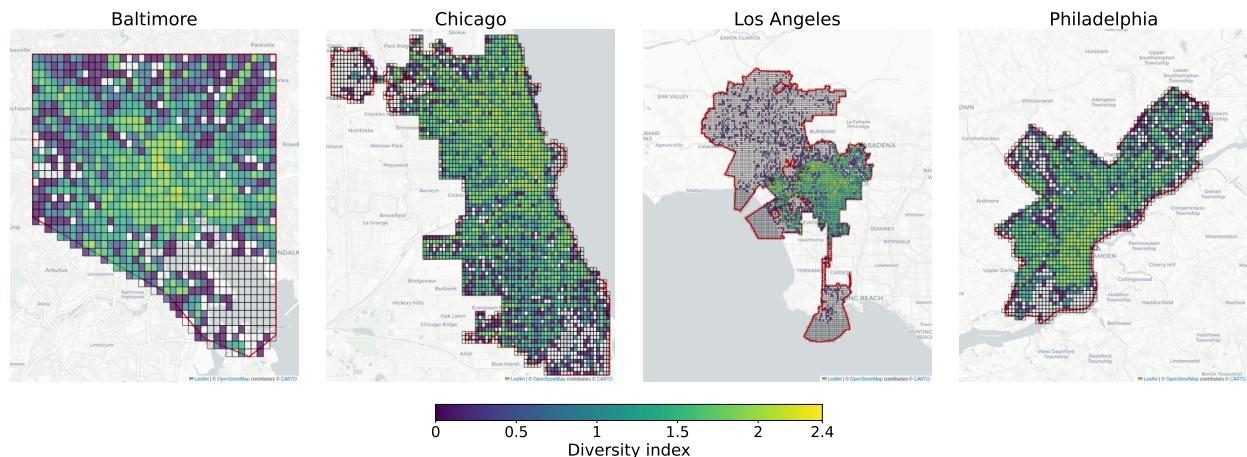


FIGURE 12. Map of the POI diversity distribution for each of the four U.S. cities between 2019 and 2023.

Appendix G Models hyperparameters

Model	Hyperparameter	Value	Description
ConvLSTM	Num. of layers	1,2,3*,4,5	Total ConvLSTM layers in the neural network.
	Num. of neurons	28	Number of neurons in each hidden layer.
	Kernel size	(3,3)*,(5,5),(7,7)	Shape of filter applied in convolutional layers.
	Optimizer	Adam	The optimization algorithm used for training.
	Learning rate	0.001,0.0001,0.00001*	The step size for the optimizer.
	Dropout rate	0.8	Probability of dropping a unit during training.
	Batch size	55	Number of samples per gradient update.
	Epochs	200	Number of full passes through the train set.
	Loss function	Binary cross-entropy	Difference between predictions and targets.
	Activation function	ReLU*,tanh	Activation function used in the hidden layers.
LR	Multiclass	ovr	One-vs-Rest strategy.
	Class weight	Balanced	The class weights are inversely proportional to their frequency.
	Max iter	400	Maximum number of iterations.
RF	Num. of trees	100	Number of decision trees in the ensemble.
	Class weight	Balanced	The class weights are inversely proportional to their frequency.
	Criterion	Entropy*,Gini	Metric used to evaluate splits.
LSTM	Num. of layers	3	Total LSTM layers in the neural network.
	Num. of neurons	28	Number of neurons in each hidden layer.
	Optimizer	Adam	The optimization algorithm used for training.
	Learning rate	0.0001	The step size for the optimizer.
	Loss function	Binary cross-entropy	Difference between predictions and targets.
	Activation function	ReLU	Activation function used in the hidden layers.

TABLE 5. Hyperparameter configurations for each model. In the case of hyperparameter tuning, the best-performing values are marked with an asterisk (*).

Appendix H Cells included in the analysis

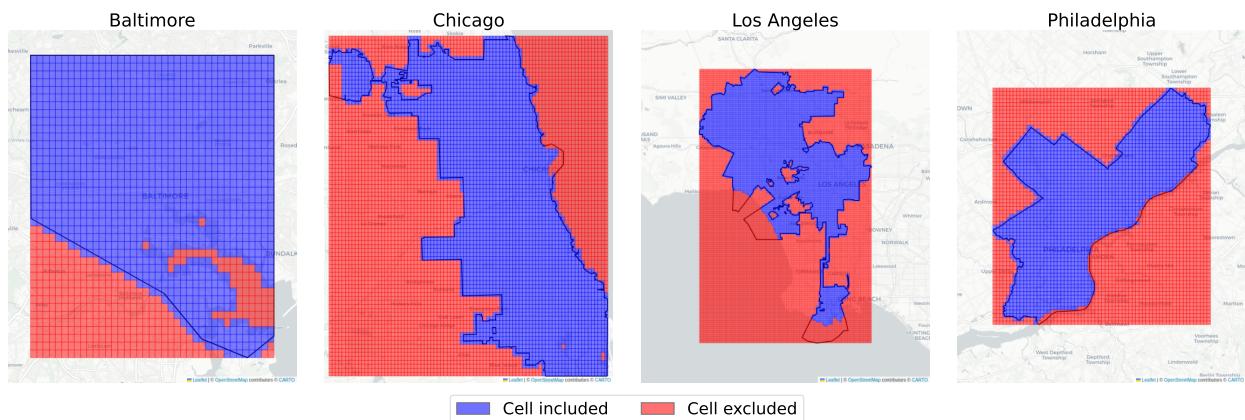


FIGURE 13. Map of the cells included and excluded from the analysis.

Appendix I Average performance at different thresholds

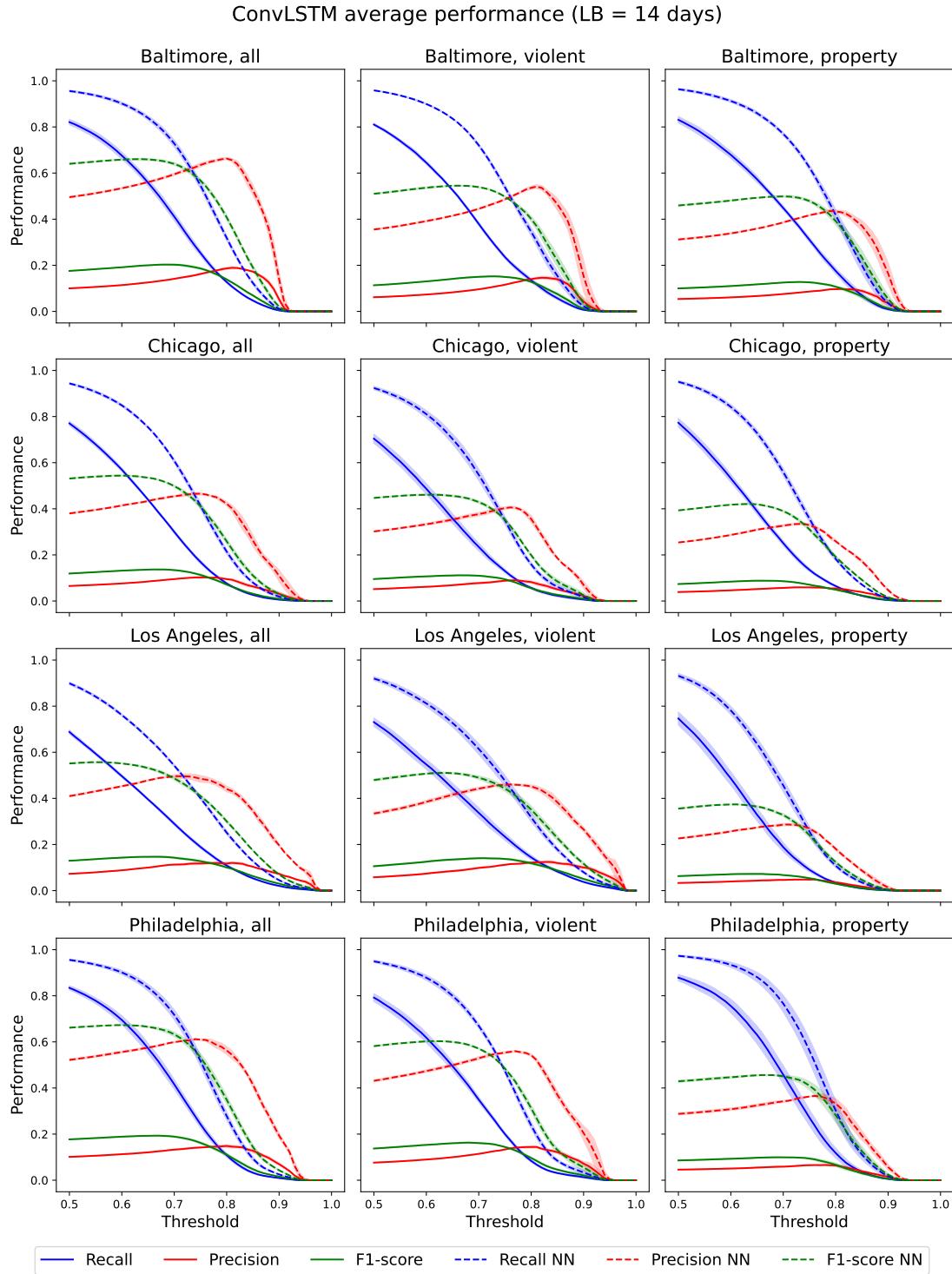


FIGURE 14. Average performance on the test set using ConvLSTM and an LB period of 14 days. The shaded area indicates the standard deviation over the four different random seeds.

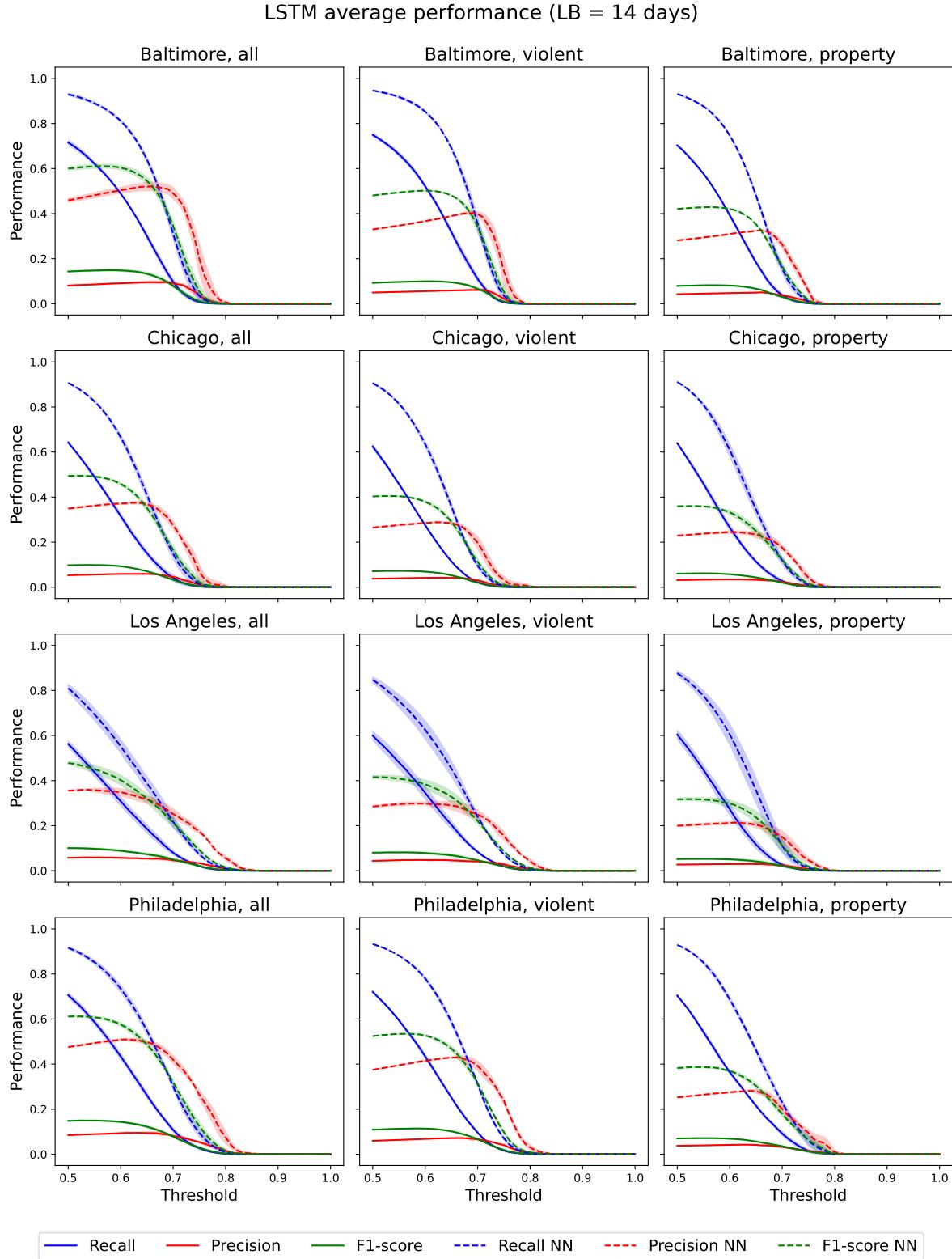


FIGURE 15. Average performance on the test set using LSTM and an LB period of 14 days. The shaded area indicates the standard deviation over the four different random seeds.

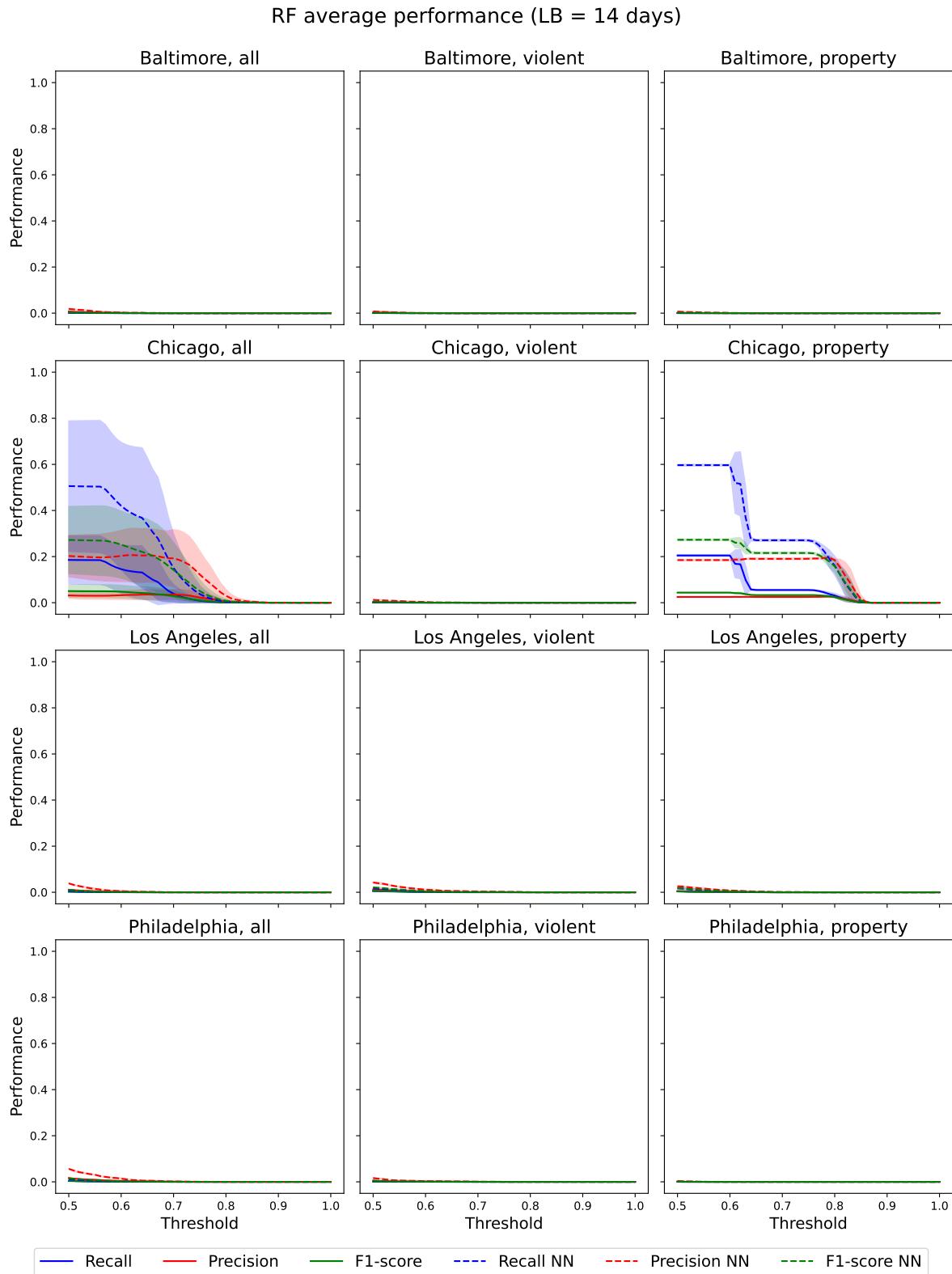


FIGURE 16. Average performance on the test set using RF and an LB period of 14 days. The shaded area indicates the standard deviation over the four different random seeds.

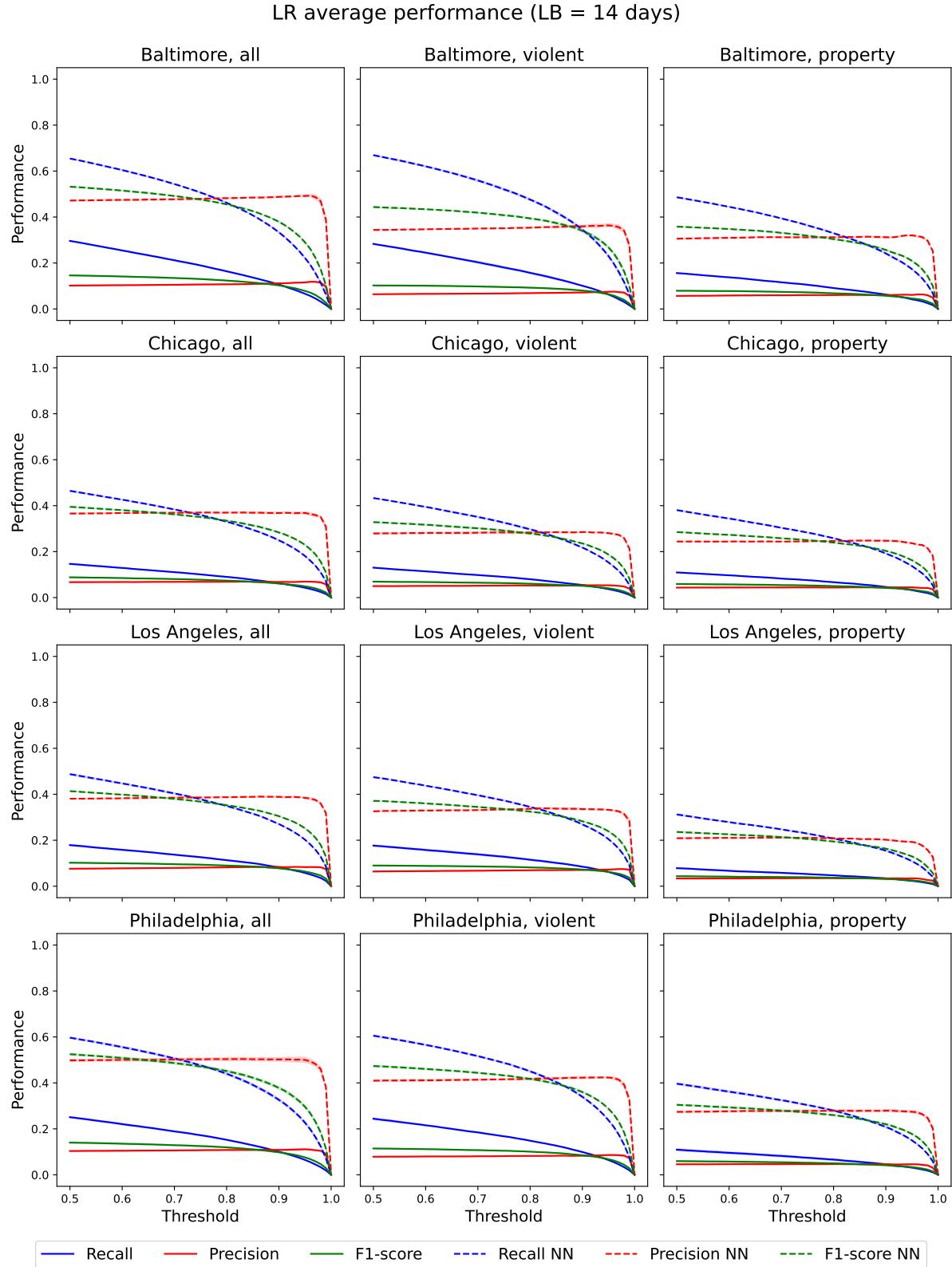


FIGURE 17. Average performance on the test set using LR and an LB period of 14 days. The shaded area indicates the standard deviation over the four different random seeds.

ConvLSTM average performance (LB = 7 days)

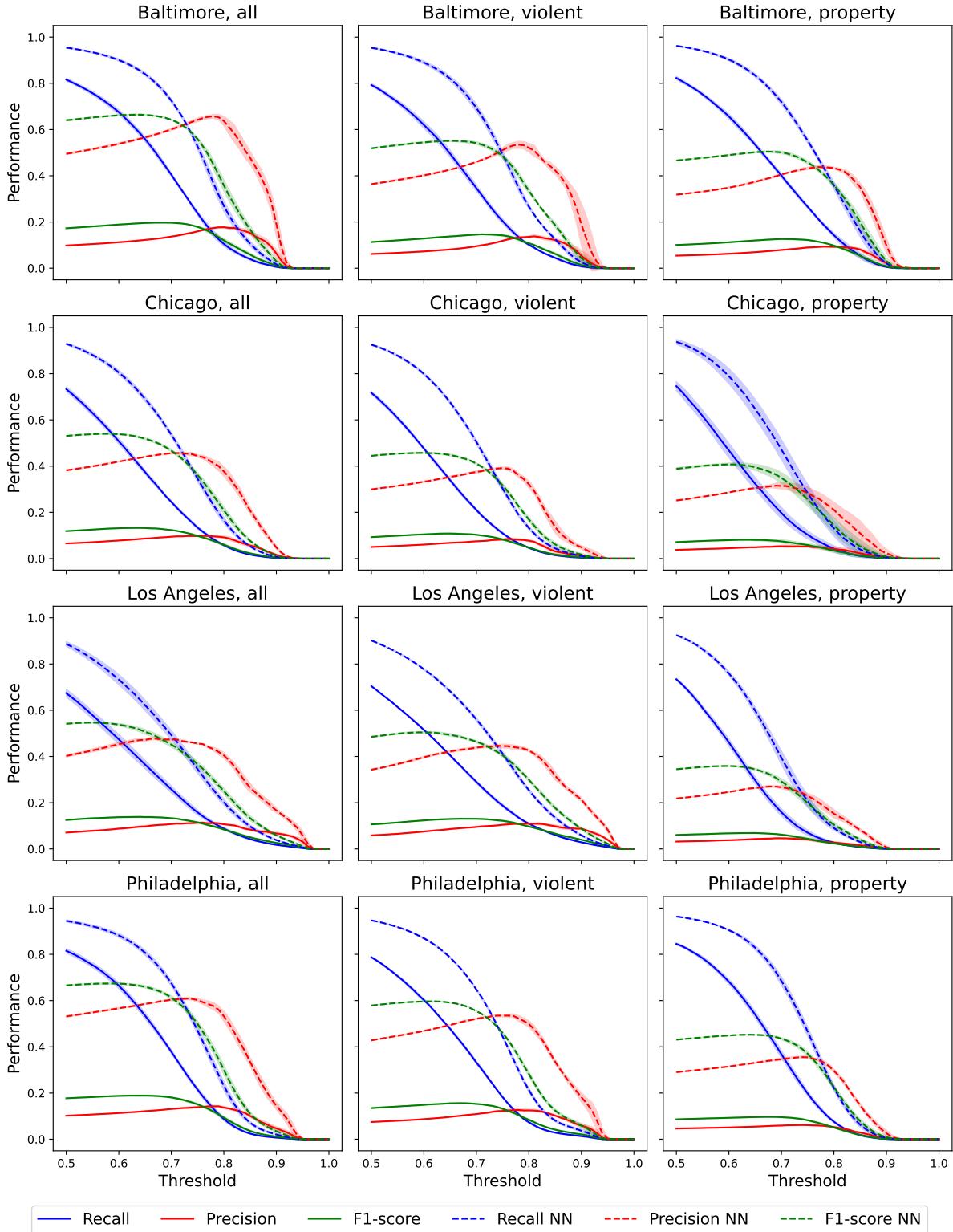


FIGURE 18. Average performance on the test set using ConvLSTM and an LB period of 7 days. The shaded area indicates the standard deviation over the four different random seeds.

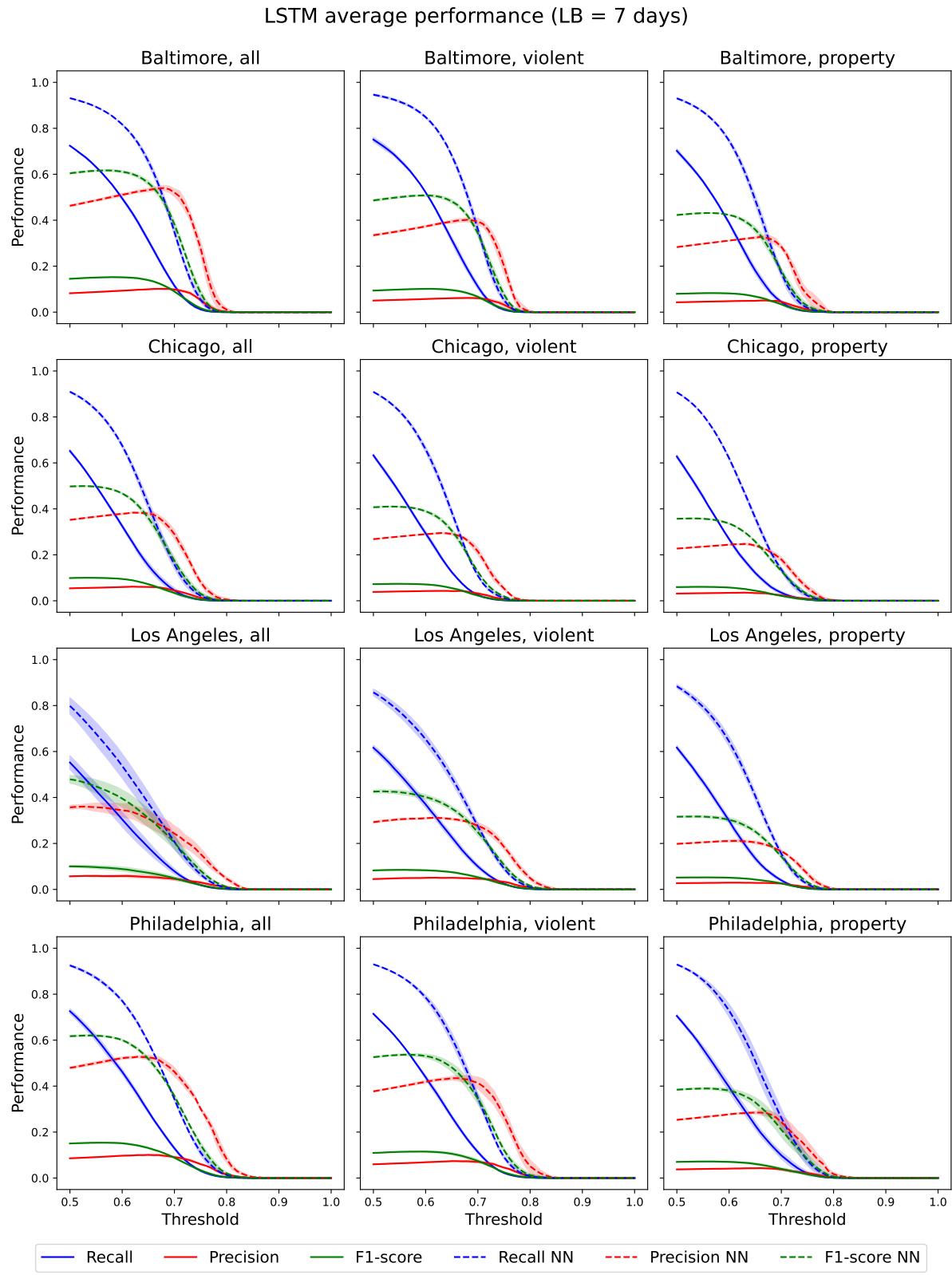


FIGURE 19. Average performance on the test set using LSTM and an LB period of 7 days. The shaded area indicates the standard deviation over the four different random seeds.

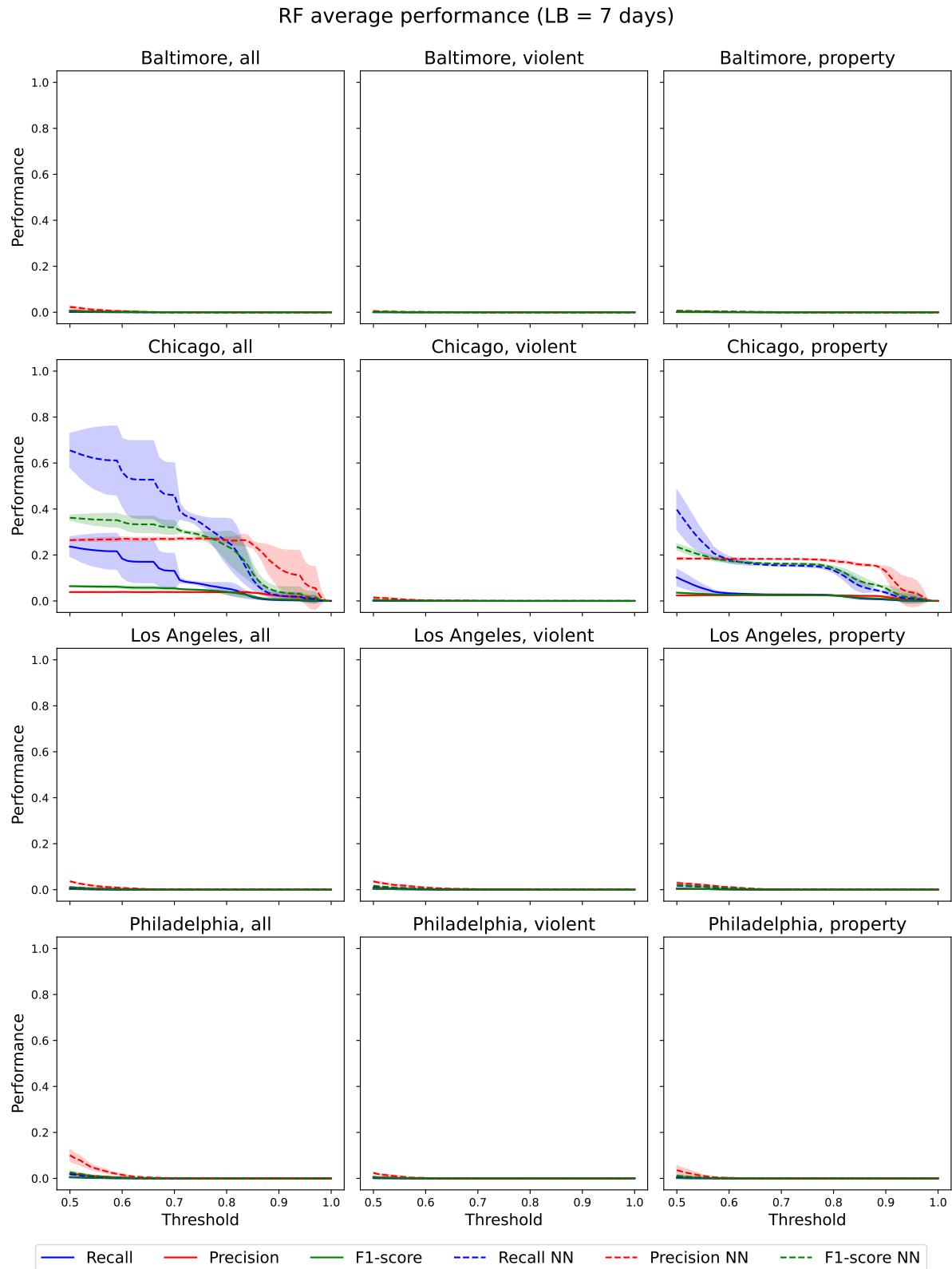


FIGURE 20. Average performance on the test set using RF and an LB period of 7 days. The shaded area indicates the standard deviation over the four different random seeds.

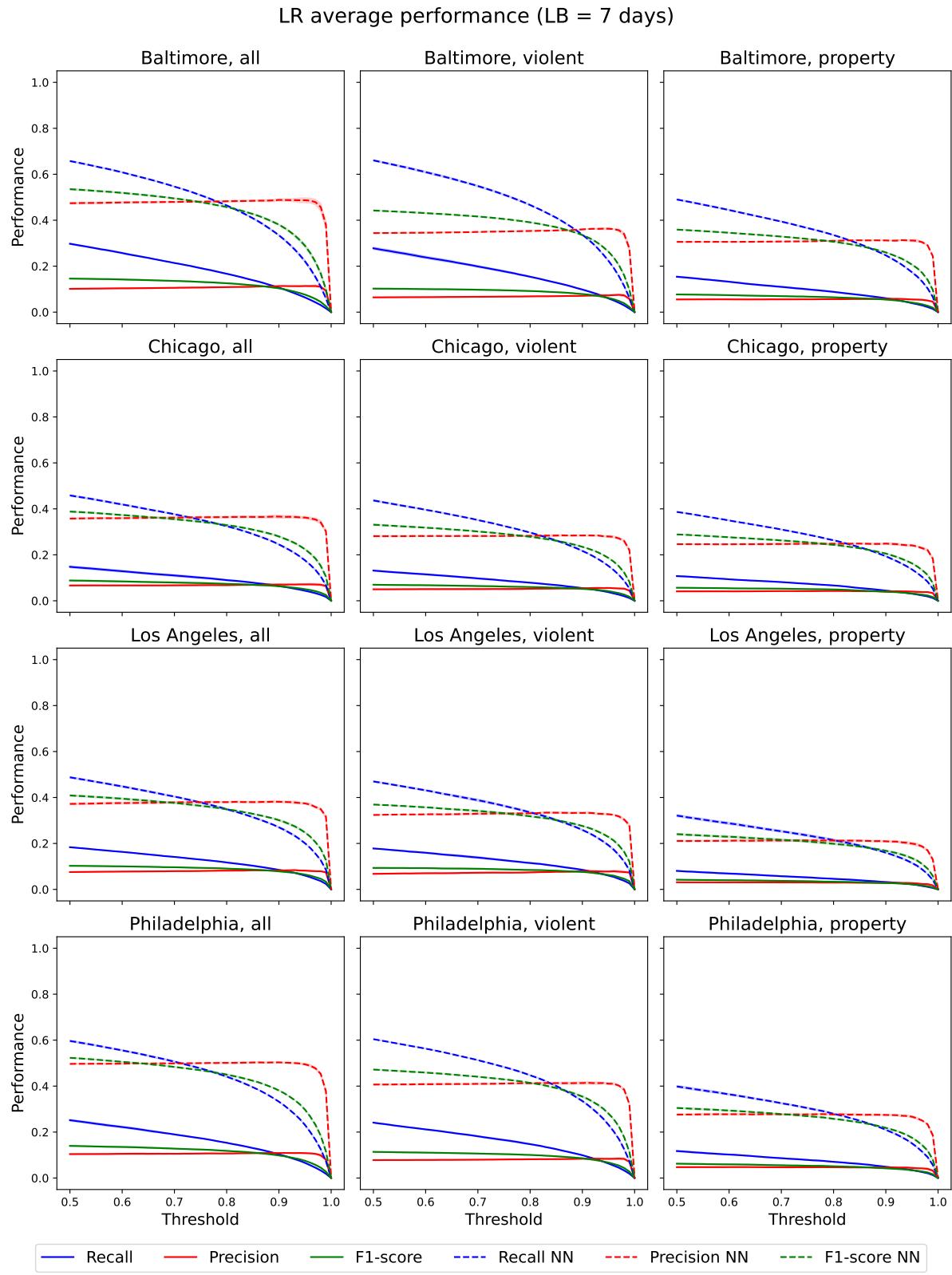


FIGURE 21. Average performance on the test set using LR and an LB period of 7 days. The shaded area indicates the standard deviation over the four different random seeds.

ConvLSTM average performance (LB = 2 days)

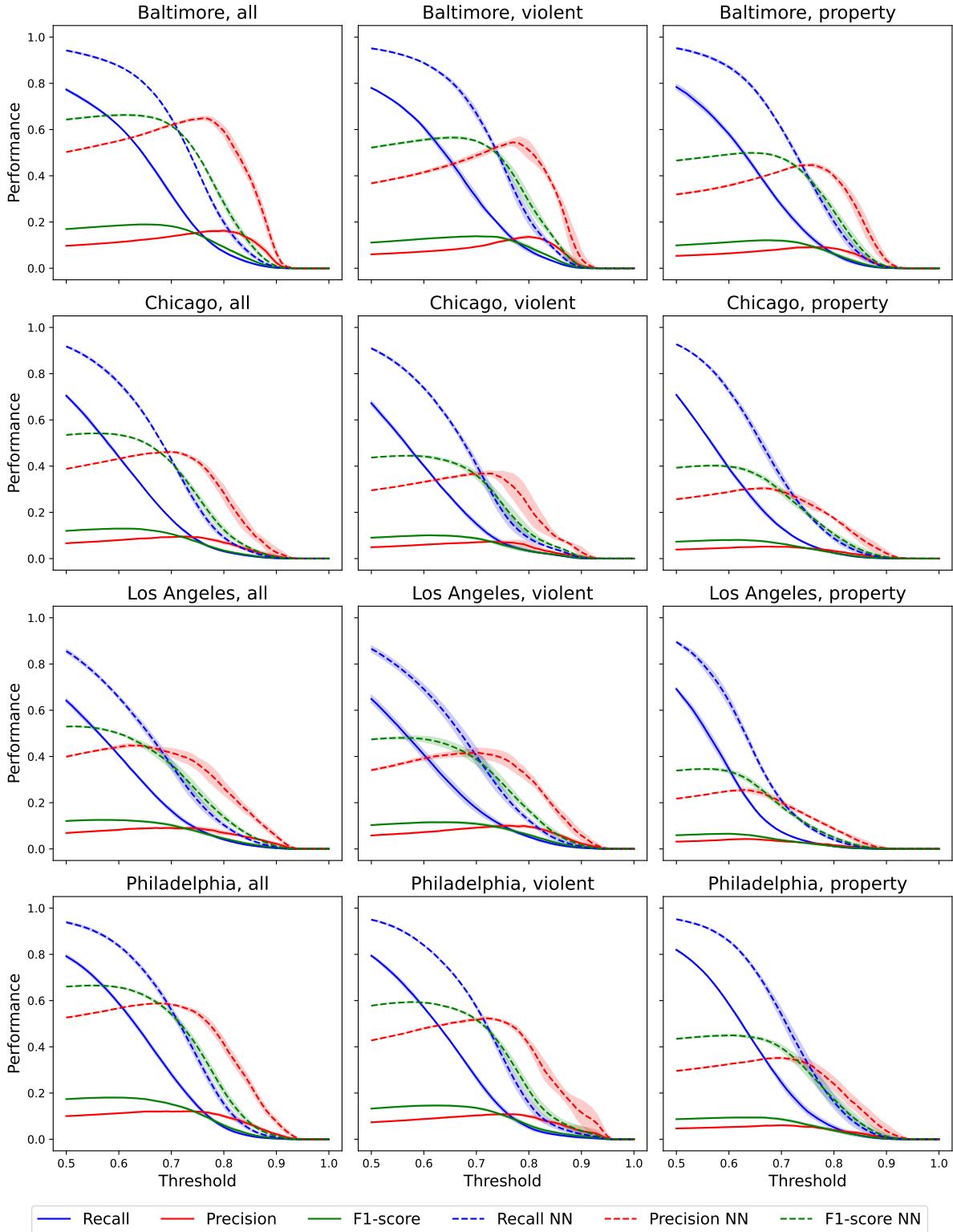


FIGURE 22. Average performance on the test set using ConvLSTM and an LB period of 2 days. The shaded area indicates the standard deviation over the four different random seeds.

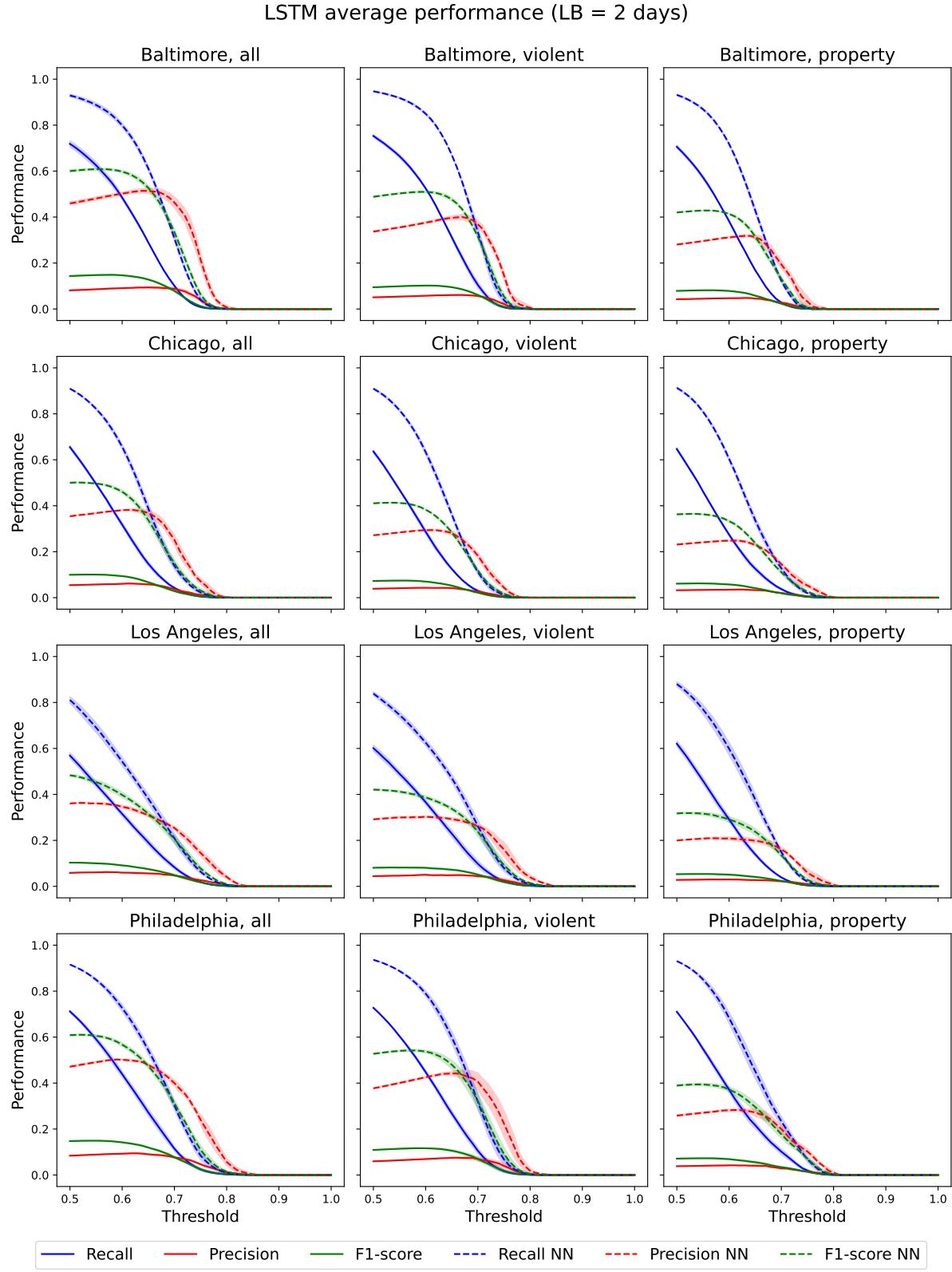


FIGURE 23. Average performance on the test set using LSTM and an LB period of 2 days. The shaded area indicates the standard deviation over the four different random seeds.

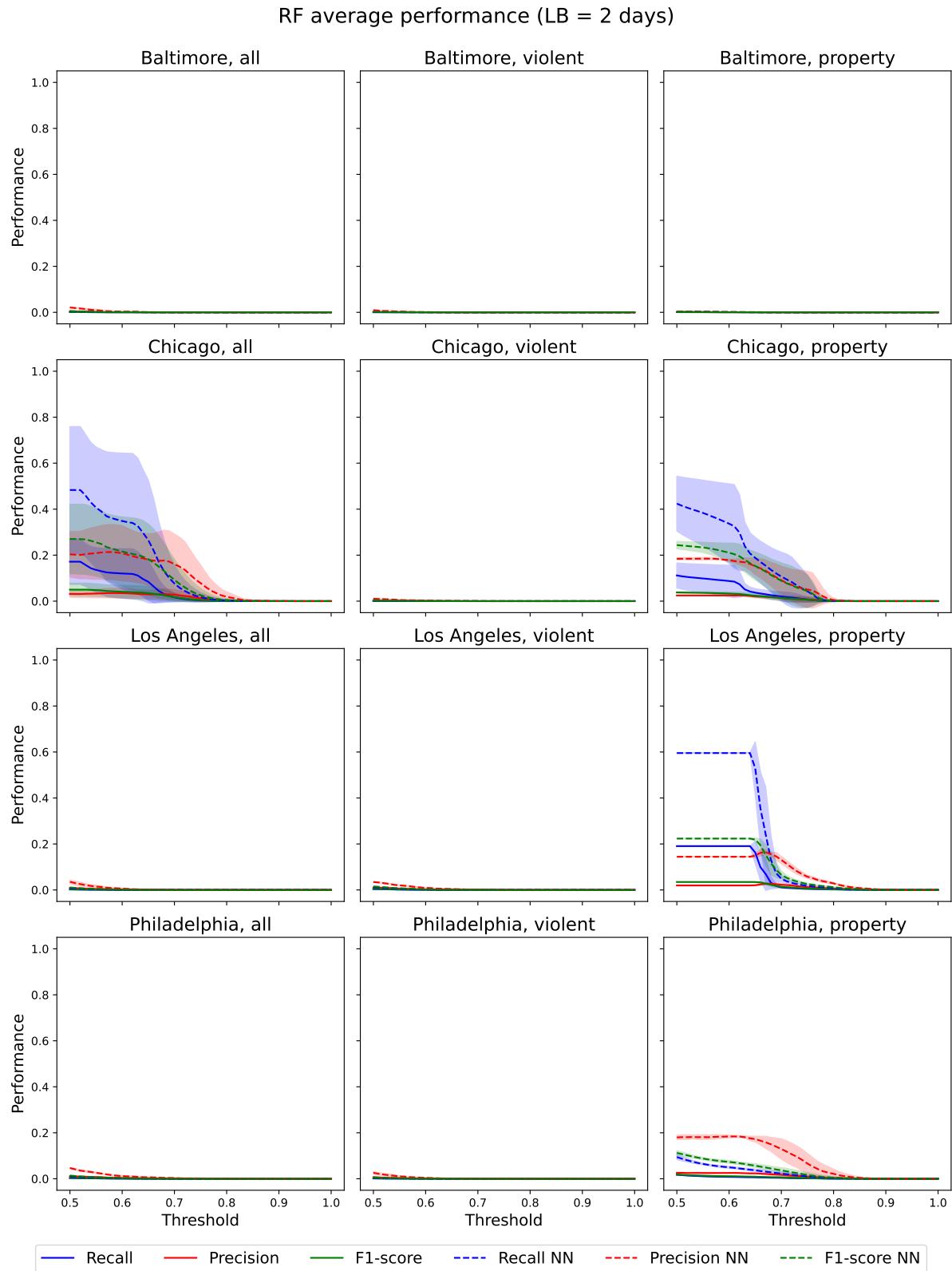


FIGURE 24. Average performance on the test set using RF and an LB period of 2 days. The shaded area indicates the standard deviation over the four different random seeds.

LR average performance (LB = 2 days)

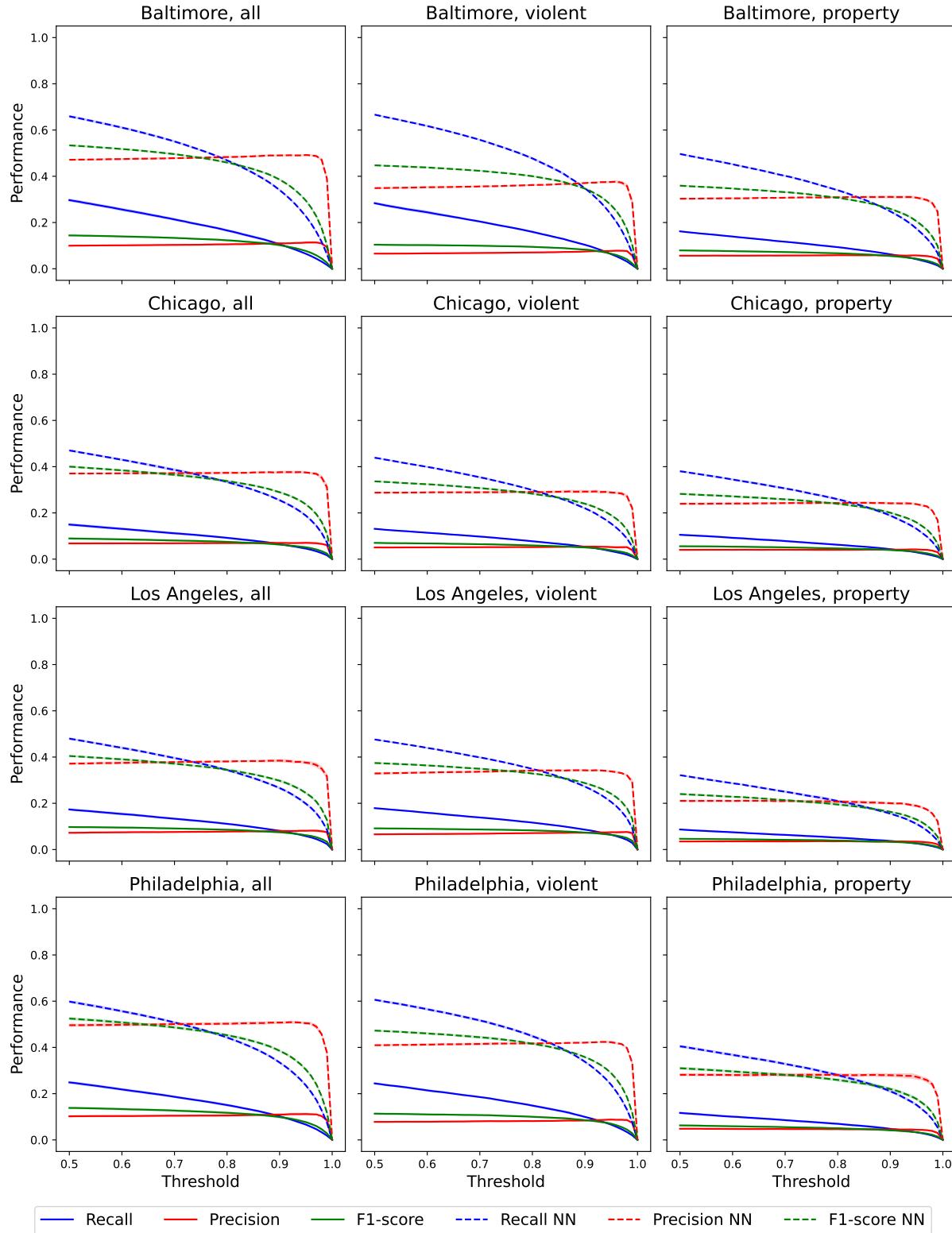


FIGURE 25. Average performance on the test set using LR and an LB period of 2 days. The shaded area indicates the standard deviation over the four different random seeds.

ConvLSTM average performance (LB = 1 days)

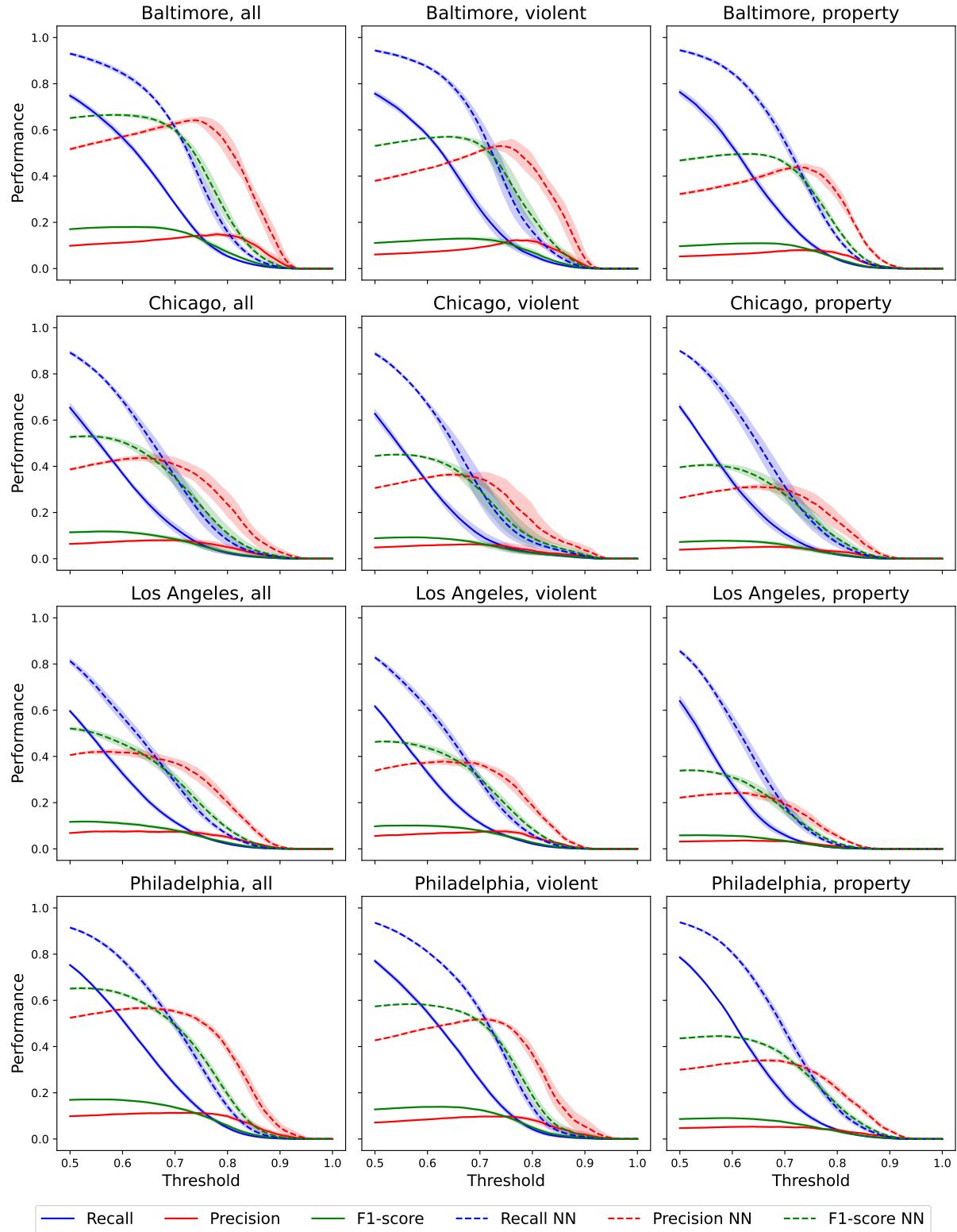


FIGURE 26. Average performance on the test set using ConvLSTM and an LB period of 1 day. The shaded area indicates the standard deviation over the four different random seeds.

LSTM average performance (LB = 1 days)

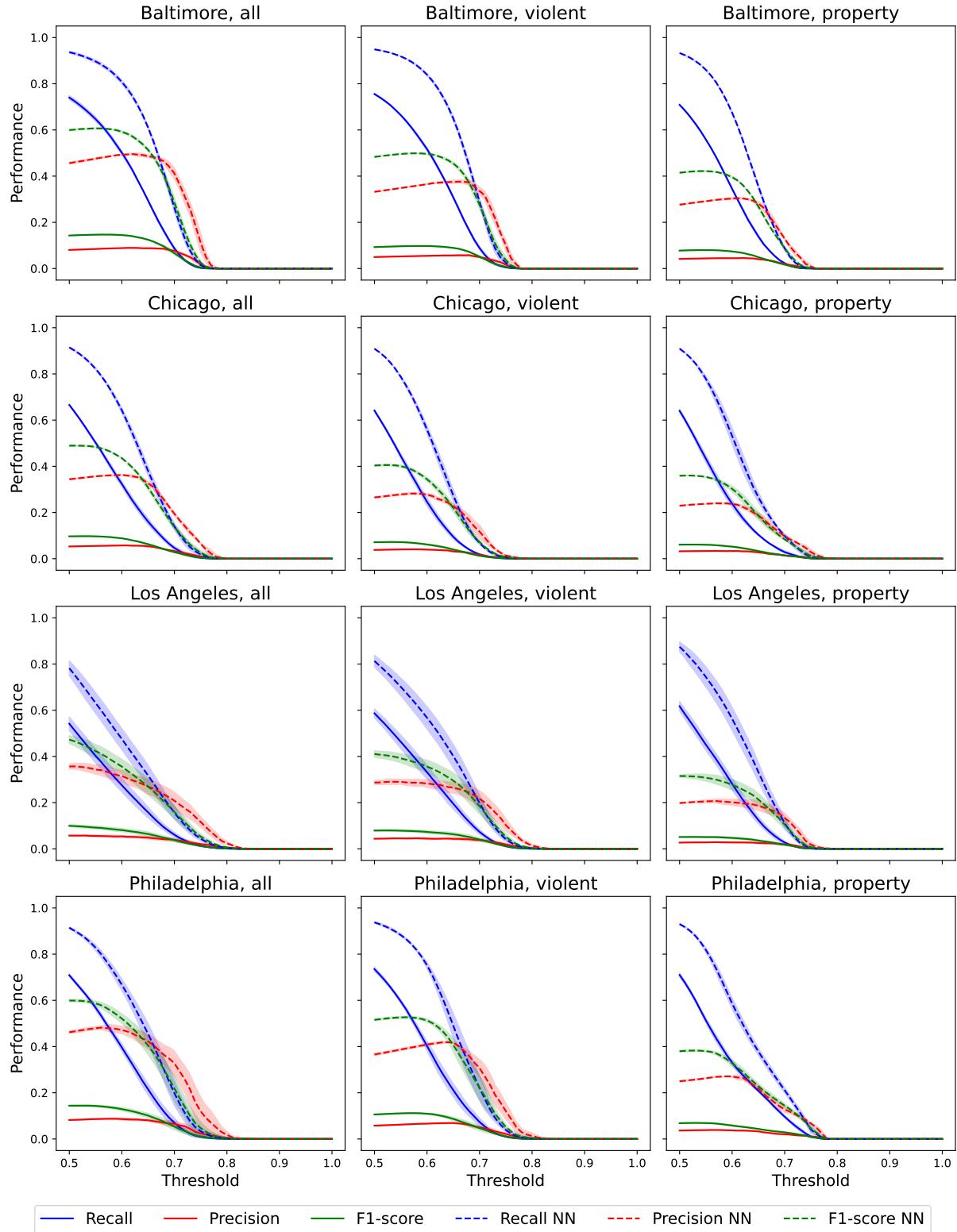


FIGURE 27. Average performance on the test set using LSTM and an LB period of 1 day. The shaded area indicates the standard deviation over the four different random seeds.

RF average performance (LB = 1 days)

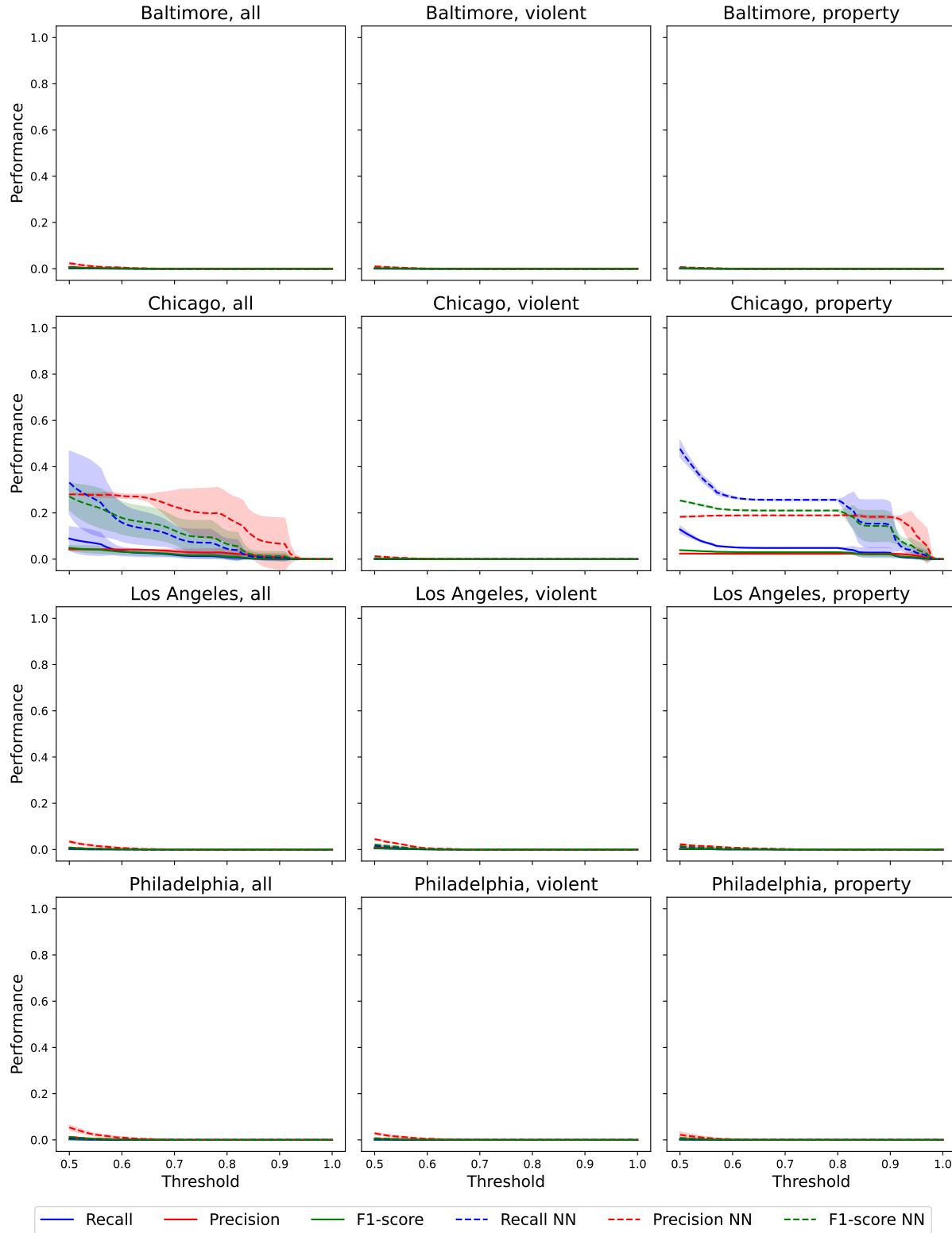


FIGURE 28. Average performance on the test set using RF and an LB period of 1 day. The shaded area indicates the standard deviation over the four different random seeds.

LR average performance (LB = 1 days)

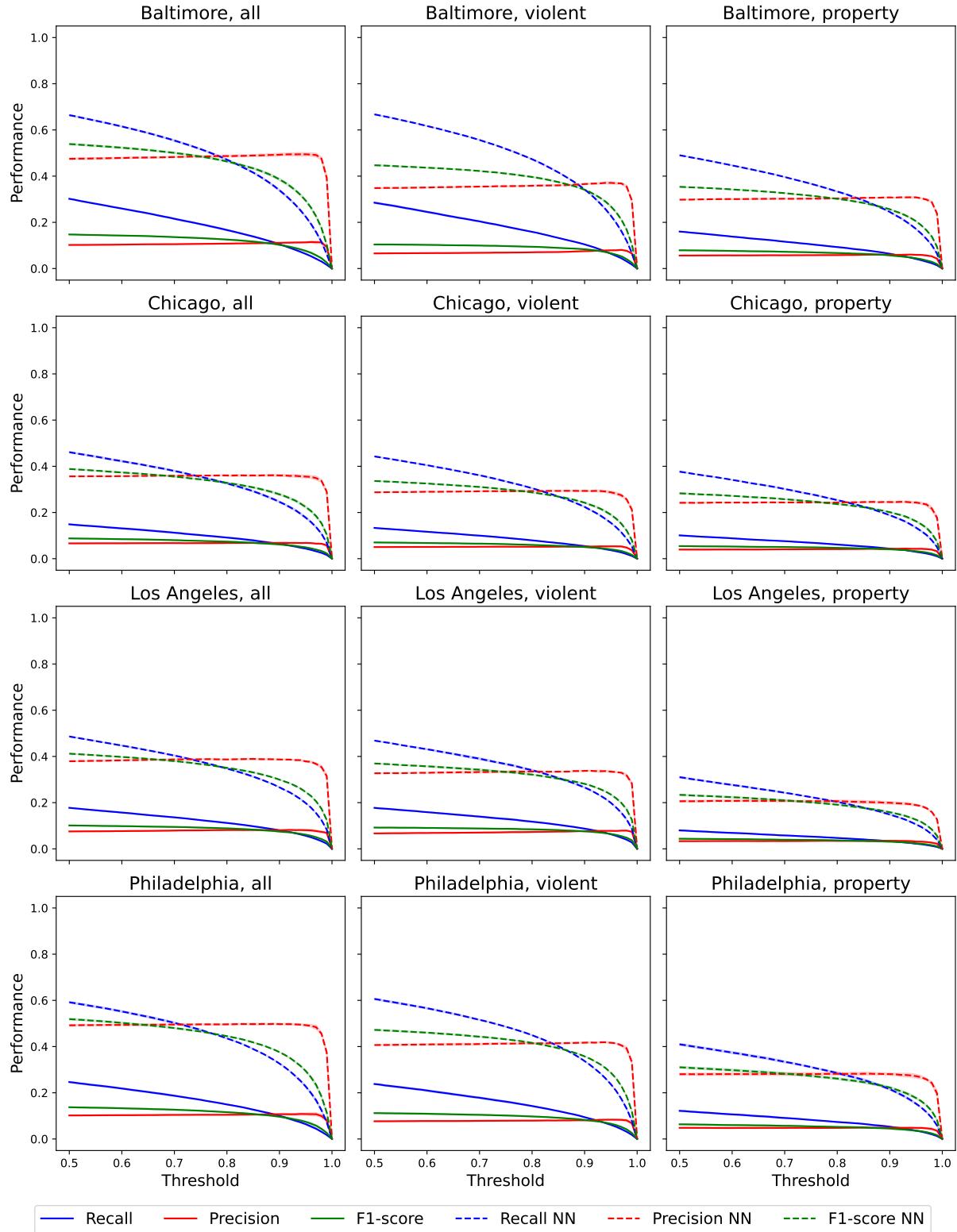


FIGURE 29. Average performance on the test set using LR and an LB period of 1 day. The shaded area indicates the standard deviation over the four different random seeds.

Appendix J General performance for 7- and 1-day LBs

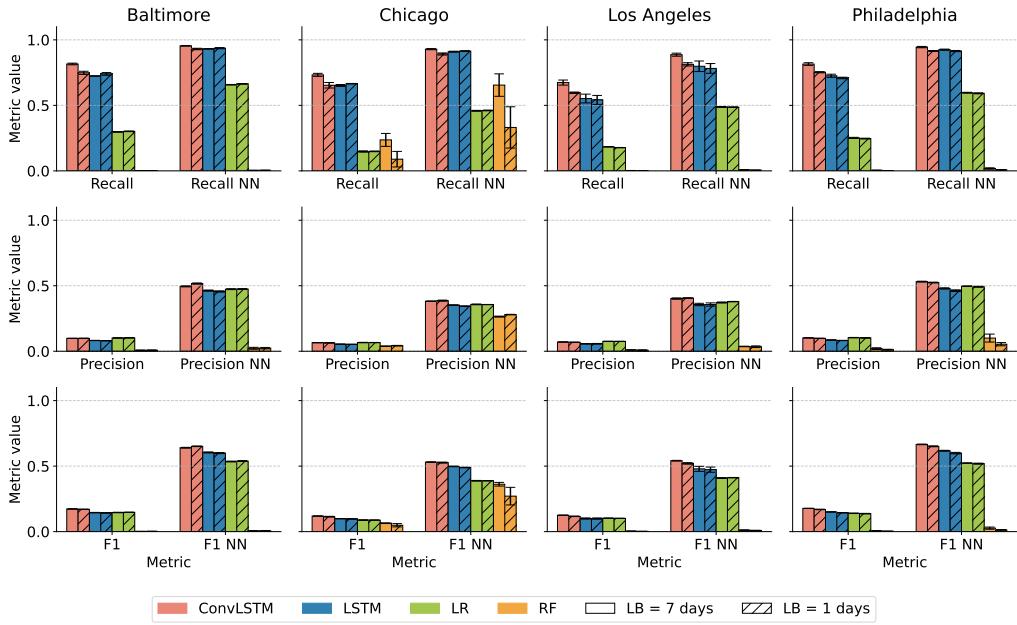


FIGURE 30. Average performance on the test set for our model and three baselines with and without taking into account the NN cells, considering all five crimes together and using LB periods of 7 and 1 day. Error bars indicate the standard deviation.

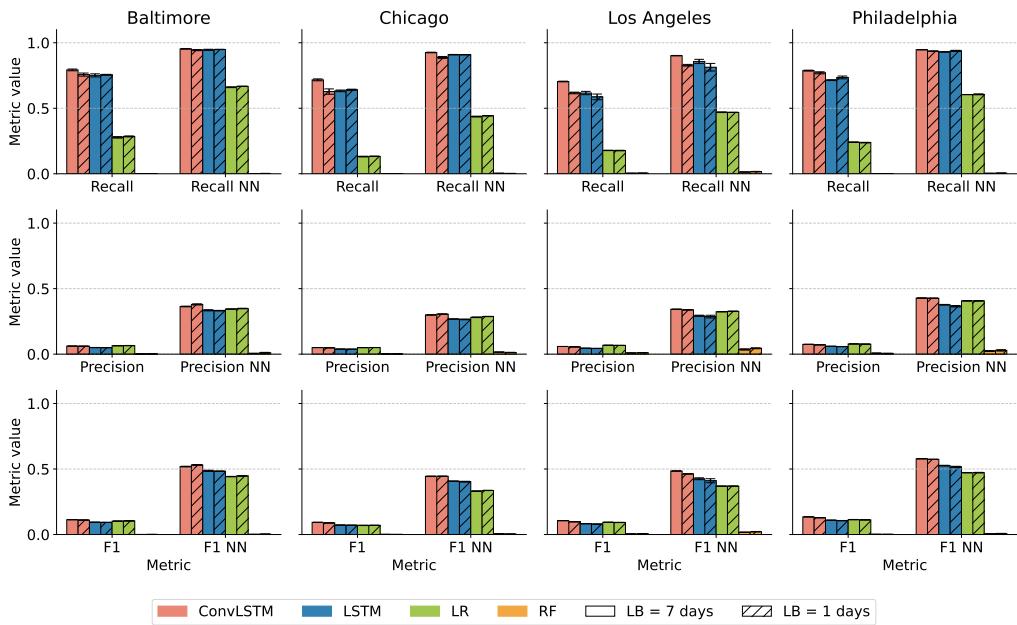


FIGURE 31. Average performance on the test set for our model and three baselines with and without taking into account the NN cells, considering only violent crimes and using LB periods of 7 and 1 day. Error bars indicate the standard deviation.

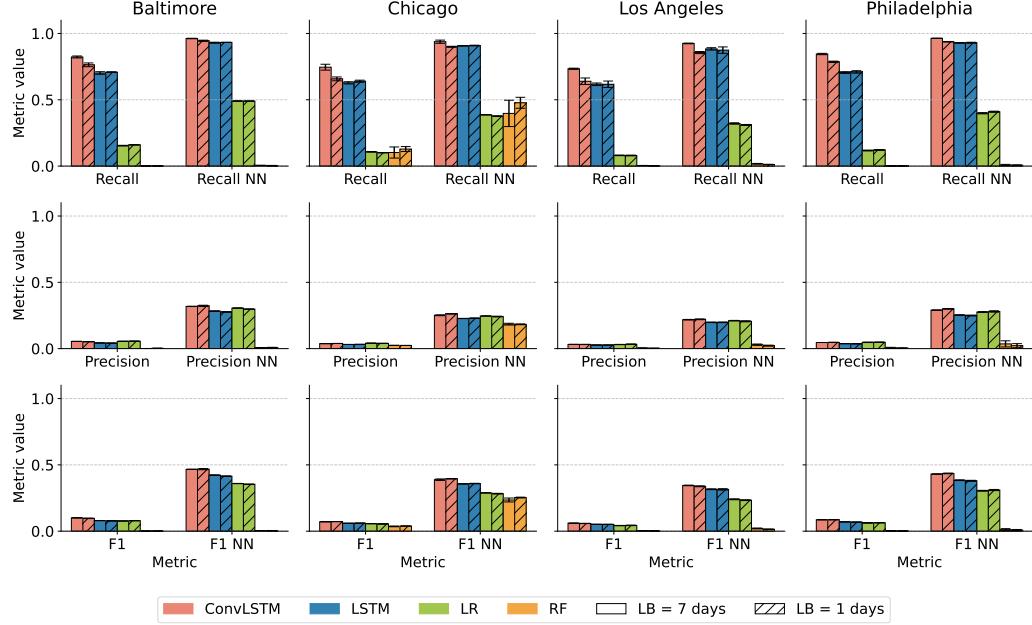


FIGURE 32. Average performance on the test set for our model and three baselines with and without taking into account the NN cells, considering only property crimes and using LB periods of 7 and 1 day. Error bars indicate the standard deviation.

Appendix K Effect of mobility features for 7- and 1-day LBs

LB	Metric	CMS - CM				CMS - CS			
		Bal	Chi	Las	Phi	Bal	Chi	Las	Phi
7 days	Rec	-2.025	-8.628	-1.039	-2.083	0.196	-1.976	-0.612	-1.545
	Prec	8.264	7.814	5.247	4.107	1.193	4.854	1.509	1.957
	F1	7.171	6.274	4.632	3.507	1.077	4.304	1.528	1.421
	Rec NN	-0.488	-2.552	-0.873	-0.597	-0.016	-0.638	0.116	-0.681
	Prec NN	2.949	4.411	1.937	3.156	-0.359	0.182	-0.203	0.866
	F1 NN	1.826	2.310	1.169	1.776	-0.222	-0.006	-0.053	0.249
1 day	Rec	3.597	-5.895	28.822	1.878	4.945	-0.352	-5.450	-3.413
	Prec	3.859	10.652	1.875	4.244	3.662	6.491	3.913	4.925
	F1	3.719	9.882	10.795	4.028	3.624	6.613	2.240	3.649
	Rec NN	1.147	-1.084	21.804	0.599	1.201	0.335	-2.406	-1.532
	Prec NN	1.924	6.311	5.507	1.952	0.864	2.999	2.010	2.964
	F1 NN	1.533	4.587	13.724	1.549	0.973	2.511	0.375	1.397

TABLE 6. Percentage difference in model performance between the CS and CMS configurations (to assess the effect of M), and between the CM and CMS configurations (to assess the effect of S), for each city and using LB periods of 7 and 1 day. Positive values indicate improved performance, while negative values indicate a decrease.

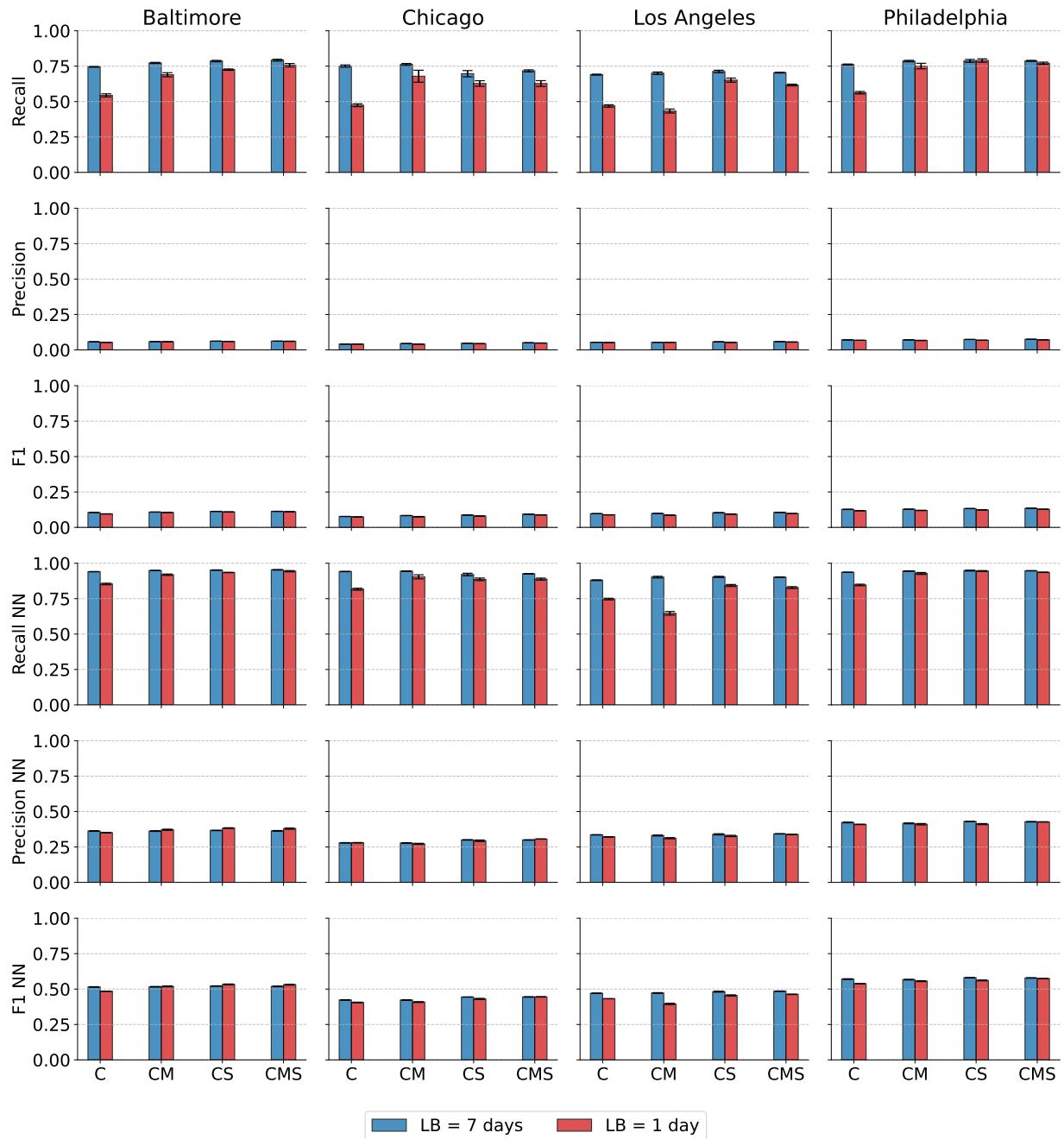


FIGURE 33. Average performance of the ConvLSTM model on the test set, trained using four different feature sets and LB periods of 7 and 1 day. Error bars indicate the standard deviation.