Predicting the Influence of Urban Vacant Lots on Neighborhood Property Values

Muhammad Fazalul Rahman[1], Pradeep Murukannaiah[2], and Naveen Sharma[1]

[1] Rochester Institute of Technology, Rochester NY 14623, USA
{mf3791, nxsvse}@rit.edu
[2] Delft University of Technology, Delft, Netherlands
p.k.murukannaiah@tudelft.nl

Abstract. Vacant lots are municipally-owned land parcels which were acquired post-abandonment or due to tax foreclosures. With time, failure to sell or find alternate uses for vacant lots results in them causing adverse effects on the health and safety of residents, and cost the city both directly and indirectly. Although existing research has tried to define these impacts, cities need quantifiable evidence from within the city to make planning decisions based on these studies. Moreover, trying to understand the impact of vacant lots in an uncontrolled setting makes it difficult to perform A key problem with existing methodologies is that they tend to look at the city as a whole, while ignoring the diverse socio-economic factors at play. Altogether, city planners are left with little or no actionable information to prioritize conversion of vacant lots. In contrast, for our research we try to model the city as blocks, census tracts and neighborhoods while using relevant features to capture key demographic, economic and geographic characteristics. In addition, we build a deep learning model to quantify the impact of vacant lots on changing property values so as to recommend conversions that yields the maximum benefit through property value tax increase. Our results indicate that our model is able to capture the relationship between vacant lots and property values better than conventionally used algorithms and data models. Further, our model specifically caters to small and mid size cities, which are often neglected in the mainstream urban computing research.

Keywords: Urban computing, deep learning, Gaussian processes, spatio-temporal data, computational social science, vacant lots

1 INTRODUCTION

In the past century, cities in the United States have undergone significant changes. While some cities improved, with job opportunities that came with the establishment of new and relocated industries and increased immigration, others suffered from depopulation and job losses [16]. This led to properties in the latter cities getting abandoned or foreclosed due to tax delinquencies [20, 7, 1]. The structures in these properties have to be demolished if found to be in hazardous conditions. Such parcels of city-owned real estate without any known uses are known as *vacant lots*, which comprise about an average of 15% land area across seventy US cities according to a 2000 Brookings Institutions study [17].

While vacant lots seem harmless to cities and residents, they are found to have major impacts on the stability of neighborhoods and lives of neighborhood residents. First, they become dumping grounds for litter and other solid wastes, and eventually, health hazards if left unchecked. Since inspection for city-owned properties are done seasonally, it becomes the responsibility of neighboring residents to call the respective city officials and report such conditions, which leads to code inspections and corrective actions. Further, studies have shown that the presence of abandoned properties and vacant lots can increase crime in the neighboring vicinity [6, 15]. Moreover, vacant parcels can be perceived as a sign of neglect and distress, which can drive down the values of neighboring properties [14]. Property value depreciation further erodes the tax base of the already declining budgets of cities, and incur major losses to these cities over time.

To understand the situation within city offices, we conducted meetings with the city planners, property assessment officers and other city officials, which led do the following observations.

1. Currently, the monetary impact of vacant lot is fixed at USD 6 per lot, which is the amount paid to contractors to perform the seasonal cleaning. Indirect costs like cost of increased police surveillance that result from increased crime, cost of unscheduled cleanups and property tax depreciation are often overlooked.
2. City planners and assessors were unaware of the literature outlining the impact of vacant lots on different neighborhood conditions including health, property values, crime, etc.
3. Although different studies show varying levels of impact of vacant lots on property values, they would not be able to consider these when making policy changes since none of the studies were conducted in Rochester.
4. City policy allowed the leasing of vacant land by city residents for community gardening. The length of the leases are set for 10 months, after which residents can request for renewal.
5. City assessors use real-estate sales values for reassessing the property values every four to five years. Neighborhood conditions are not accounted for directly in the assessments, but real-estate values can be heavily affected by these neighborhood factors.
6. While certain departments employ data scientists, their numbers are few and they are primarily occupied with other responsibilities and would not be able to focus their time and attention for machine learning applications using urban data.

City residents who are themselves interested to lease and convert these lots are hindered by the policies in place since they need to demonstrate that the benefits of conversion outweigh the cost incurred on the city by the lots.

In order to make informed decisions about vacant lot conversion, urban planners and city administrators need data-driven models. However, as we described earlier, there is a dearth of such models for small and medium-sized cities. Models trained for bigger cities would not necessarily work well for smaller cities. In an effort to fill this gap, we develop a data-driven model of vacant lots and their impact on neighborhoods for Rochester NY, a medium-sized, rust-belt city.

In order to measure the impact, we consider the influence of vacant lots on neighborhood property values. Our discussions with city officials suggest that the property value tax depreciation is a primary source of revenue loss for a city. Accordingly, demonstrating any relationship between vacant lots and property values makes a strong case for policy changes toward vacant lot conversions.

Our approach, first, defines a data model that takes into consideration multiple hierarchies within a city—blocks, census tracts, and neighborhoods. We then extract features relevant to our analysis from each layer in the city hierarchy along with the characteristics of individual property parcels. Finally, we provide the data model thus generated as input to a deep learning framework. Our analysis shows that our model gives much better precision compared to conventional methodologies used for predicting the impact of vacant lots on property values.

The remainder of this paper is organized as follows. In Section 2, we discuss related works. We formally define the problem and describe our data framework in Section 3. The approaches we used are described in Section 4 followed by the results obtained in Section 5. We conclude this paper in Section 7.

2 Related Work

The growing number of vacant lots in cities and their effects on neighborhoods have been well explored in the social science literature, with studies dating back to mid-1950s [4]. These earlier works dealt with the cost of public works that arise due to large areas being left vacant, which in turn increased in cost for installation and maintenance for electric poles, cables, water mains, and so on. However, it was in the late 1900s that population shift started being a more significant issue for smaller cities [16]. This was the time when development in metropolitan areas accelerated much faster than smaller cities, leading to housing abandonment. Burchell and Listokin [5] describe abandonment as both a symptom and disease; one that not only indicated urban decline, but also provided the feedback mechanism to accelerate and perpetuate it. With the increase in housing supplies due to abandonment, rental properties become unable to cover taxes and related costs from the income they produce. The high supply, and consequently declining demand, also make it nearly impossible for landlords to sell them, increasing the pool of abandoned properties even further [9, 15, 21].

Once the problem was well-defined, further studies tried to quantify the exact impact abandoned and vacant properties have on neighborhood dynamics. [15] model the effects of vacant lots on crime. Although the results showed appreciably positive correlation, they were not statistically significant. In a later study by [8], rather than just using data, the authors conducted a randomized control trial by greening a vacant lot cluster to understand changes in crime and safety relating to conversion. Another cluster was used as control group. The results showed an insignificant decrease in violent crime. However, a follow-up study by [3] on 541 randomly sampled vacant lots showed significant improvement in actual and perceived safety in the neighborhoods.

Multiple studies have tried to model the relationship between vacant lots on neighborhood property values. Immergluck and Smith [14] studied the impact of housing foreclosures on property values in Chicago using a hedonic regression model, and found statistically significant relationship between the two. An overall estimated loss of $598 billion in property values was valuated using average property value in the city. However, this model is not adequate enough to make conversion decisions as it has high error rates. A study by [10], in addition to modeling the impact of abandoned properties as a function of distance, used weighted repeated sales model to estimate the impact that duration of abandonment has on property values. The results indicate that both distance from abandoned properties and duration of abandonment have significant impact on property values. However, the use of repeated sales model requires sale prices of the same property during different time periods. Such data would be sparse in historic property value records, and the number of examples available would be too low.

Other studies have focused on the positive impacts of greening vacant land [2]. [12, 11] performed difference-in-difference analysis to understand the impact of converting vacant land into green spaces. Her results suggest that while property values tend to increase all over the city, the properties surrounding converted vacant lots enjoy a greater increase in value. However, the results also indicated that the impact was more pronounced in some parts of the city than others. This shows the need for treating a city not as a whole but as a collection of sub-populations.

To the best of our knowledge, there is no existing work on modeling vacant lots in the computing literature. Although there are computational models of vacant lots in the social science literature, they tend to use simplistic regression models. In line with recent advances in urban computing [25], we seek to use state-of-the-art machine learning techniques for modeling the vacant lot problem.

We address a novel problem. However, our solution borrow ideas from several recent works on urban computing. For example, [22] define a model to understand urban migrant mobility within the new city long with housing price information, geographic information, call behavior and social connections. They then used these features to model the problem of understand migrant churn as a classification task. Similarly, we include multiple features to model vacant lots and develop a baseline classifier.

Huang et.al.[13] use a deep neural network architecture to build a crime prediction framework which can capture the dynamic crime patterns and their inter-dependencies. Their framework models the multidimensional interactions between crime categories and regions over different time periods. [24] also built a deep-learning based model to predict crowd flow within cities. We use a similar deep learning approach to model the characteristics that conventional regression models have not been able to successfully predict. We hypothesize that deep neural networks are more efficient in capturing higher dimensional inter-dependent features than linear regression models.

3 The Vacant Lot Problem

In this section, we outline our research questions and describe the dataset we build to answer those questions.

The vacant lot problem is a well-defined and well explored problem in the realm of urban sciences. However based on existing results, it is not possible to drive data-driven decision making in cities. We therefore need to build a city-specific model that can capture the direct impact of vacant lots on nearby properties, rather than use the same approximate impact percentage for every property. Such a model that can explain effects in the micro-level can help urban planners make informed decisions, and can help improve the conditions in neighborhoods. This can also help prioritize neighborhoods within the city that require immediate attention so as to allocate limited budgets more efficiently. Our objective through this research is to answer two key research questions:

RQ$_1$ What impacts do vacant lots in a neighborhood have on the property values of that neighborhood?

RQ$_2$ How can we choose vacant lots to convert so as to maximize the benefits of conversion (minimize the negative impacts on property values if the lots are not converted)?

By answering RQ$_1$, we hope to provide a monetary assessment that can help city officials in understanding the "silent impact" vacant lots have on the city, and in particular, on individual properties within the city. Once we have a model trained using data from a city, we can identify vacant lots within the city that will have the most impact on property values if converted. For our model, this would be based on the depreciation in property value the vacant lots cause. Then, we can use this information to answer RQ$_2$, i.e., to prioritize vacant lots to convert based on budget constraints.

3.1 Data Collection

Data from larger cities are being made publicly available more often now, while smaller cities neither have the resources nor the manpower to do the same. When it comes to vacant lots, this has led to the lack of research about how to manage vacant land within tight budgets. The common solution provided for the problems caused by vacant lot is to convert them to green spaces, which is not always feasible for the rust-belt cities. For our research, we therefore chose to use one such city for analysis. Since the city of Rochester fits the profile of a declining city, and because their data was available for analysis, we chose to center our research around Rochester, NY. Table 1 shows the number of vacant lots in the city.

Table 1: The number of Vacant lots in Rochester, NY

Population		208046
Number of property parcels		65622
Number of vacant lots	Total	5037
	Residential	4198
	Commercial	839
Number of neighborhoods		48
Number of street blocks		2644

The data about property parcels, 311 calls, crime and demographics for Rochester are publicly available. While some of the data was in GIS format, others varied depending on the software used in the respective offices from which the data has to be collected. All the data was made GIS-compatible by geo-coding addresses and using unique IDs assigned by the city.

Although we focus our research on Rochester, NY, our models and methodologies are generic. We conjecture that our methods work for other similar sized cities, too, with appropriate tuning.

Fig. 1: The hierarchical structure for vacant lot model

3.2 The Vacant Lot Hierarchy

In order to approach the vacant lot problem as a data problem, it is necessary to capture the complex relationships that occur in an urban settings that can contribute towards the different outcomes that can be observed in cities. Therefore,

http://www.cityofrochester.gov/innovation/

approaching a complex setting naively might lead to loss of critical information about the underlying hierarchical structures and the inter- and intra-hierarchical relationships between them. To this end, we model a city as three layers as shown in Figure 1 to capture the essential characteristics of each level.

Level 1: City Block. The lowest level in the vacant lot model hierarchy, the city block is the smallest area that is surrounded by streets. Each city block contains one or more property parcels, and is used to obtain neighborhood characteristics which are mostly distance-sensitive.

Level 2: Neighborhood. Neighborhoods form larger geographic boundaries within cities, and are sometimes given official or semi-official statuses through resident associations or watch groups. While neighborhood data fails to capture distance-sensitive features, trends and policies that are usually similar for each neighborhood can be acquired in this level.

Level 3: City. At the city level most diverse characteristics of residents are lost; however, aggregated city data can help differentiate between cities and can help adapt models based on city-specific characteristics. City level aggregated data also helps understand where neighborhood and city blocks stand in terms of features like property values and crime.

3.3 Feature Modeling

Based on the hierarchies defined above, we collected features for each of the three levels of the hierarchy. The features can be subdivided into three categories:

Spatio-Temporal Featuresinclude crime incidents, 311 calls, code violations and property values. The examples in these datasets, with the exclusion of property values, occur at a location mostly only once; therefore, they are aggregated for each year for each level in the hierarchy. Property values (per square feet) data is available for multiple years, although the ranges of available data vary depending on the city. For example, for the city of Philadelphia, property value data is available for every year from 2012 to 2017, whereas for Rochester, data is available quadrennially from 1990 to 2017.

Spatial Features include property parcel information and locations of parks, schools, libraries and city facilities. Distances and densities of these features can help model the block or neighborhood characteristics. Distance to the nearest vacant lot and density of vacant lots in the city block are also used to incorporate any impacts they might have on the models.

Hedonic Features are the broken down constituent parts of a component like real estate or consumer electronics that can be used to predict a dependent variable [19]. For property parcels, the hedonic features include the area of the lot and any units in the lot, number of rooms, stories, bathrooms, etc. This can particularly help when we try to fit regression models so that the cost identified with these hedonic features can be accounted for, while being able to identify the costs associated with vacant lot related features.

Demographic Features are collected from various census datasets and incorporated to model the social dynamics. This includes population, average income, education attainment and market demand data. In addition, a diversity

Table 2: Features used for the hierarchical model

Block	Census Tract	Property
Population growth 2000-2010	Total population in 2000	Number of units
Total violent crime in 2017	Total population in 2010	Number of stories
Total property crime in 2017	Total population in 2017	Number of rooms
Ratio of crime in 2018 to 2014	Total households in 2000	Number of bedrooms
Number of properties	Total households in 2010	Number of bathrooms
Distance to the nearest school	Total households in 2017	Vacant lots within 50 feet
Distance to the nearest library	Percentage of bachelor's degree	Vacant lots within 100 feet
Distance to the nearest park	Percentage of graduate degree	Vacant lots within 150 feet
Distance to the nearest vacant lot	Education base	Vacant lots within 200 feet
Area of the property parcel	Diversity index	
Total area of parcels	Average income in 2017	
Total area of all vacant lots		
Total area of residential vacant lots		

index is included which shows the probability that two people chosen at random belong to the same ethnic/racial group. The complete list of features extracted from the hierarchical model has been listed in Table 2

3.4 Data Modeling

Once the required data has been identified and collected as mentioned in the previous sections, they have to be modeled to preserve neighborhood and block characteristics, while avoiding the need to have multiple models within the city. We therefore collect data about individual properties (parcel area, distance to facilities, property prices, etc.) and join them with block level data (crime count, demographics, property count, etc.) before scaling them among neighborhoods. To this end, we consider a set of neighborhoods within the City of Rochester (i.e., N = $(N_1, ..., N_I)$). Each neighborhood contains multiple census tracts $(C_1, ..., C_j) \in N_i$ and street blocks $(B_1, ..., B_k) \in C_j$. Within these blocks, there are multiple non-vacant property parcels (P_x) and vacant property parcels (V_y). We seek to model the effects of vacant lots $(V_y \in B_k)$ on the properties within the same block $(P_x \in B_k)$. We first construct a neighborhood matrix (N_{S*F}), where each sample $\vec{S}$ contains F features (as mentioned in the previous section) for every property in the neighborhood collected from its corresponding layers N_i, C_j and B_j as well as the individual parcel's characteristics $(\vec{P_i})$ and vacant lot characteristics $(\vec{V})$. That is,

$$\vec{S_i} = \vec{B}_k \frown \vec{C}_j \frown \vec{P}_x \frown \vec{V}_y$$

Once we have the neighborhood matrix, we standardize the features among each neighborhood as opposed to standardizing the data for the entire city. This can be represented as follows:

$$F' = \frac{F - \overline{\mathrm{F}}}{\sigma}$$

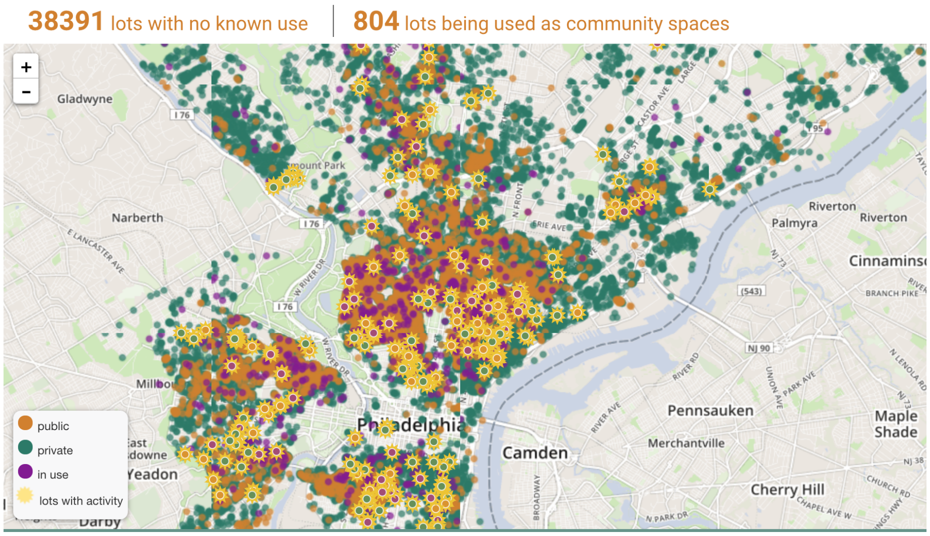

Fig. 2: Converted vacant lots in Philadelphia, PA

where F is the original feature vector, $\overline{\mathrm{F}}$ is the mean of the feature vector and σ is its standard deviation within the neighborhood.

4 Approaches

In this section we outline the approaches we explore using the data models described in the previous section.

4.1 Gaussian Process Regression

Gaussian Processes are supervised non-parametric learning approaches in which we consider the predictions to be probabilistic [18]. This is with the underlying assumption that the probability distribution of a set of arbitrary points from the dataset is jointly Gaussian with some mean and covariance. Just like in any supervised learning models where we assume that the target variables are similar for similar predictor variables, Gaussian Processes also follow the same assumption and use a covariance matrix $\Sigma(x)$ to define the similarities. A characteristic length scale (σ_l) is used to define the maximum distance between input values, beyond which the target values become uncorrelated.

For our analysis, we provided the set of features $\vec{S_i}$ as independent variables and the change in property value value ΔP as the dependent variable. In addition, linear regression is also used as a baseline to compare with more complicated analyses, since ordinary least square and linear regression is generally used in social science literature to identify correlations.

4.2 Artificial Neural Network

While conventional or Gaussian process regression works for most cases, they are not always able to capture key relationships within the data. Especially when it comes to a hierarchical framework like our model where there are multiple parameters affecting each other, regression might not be able to provide us with the best possible fit. To overcome this issue, we experiment with the application of neural networks. We build a neural network that optimizes on different hyper-parameters like the optimizer, activation functions, batch sizes and epochs. Since there is no one size fits all, we believe that this is essential to find the best fit, given the unknown underlying relationships within and across the different layers in our city hierarchy.

The input layer to the neural network architecture consists of the features from different hierarchies that we had discussed earlier ($\vec{S_i}$). We define the expected output to be the change in property values over the years (ΔP). We define the first hidden layer to contain the same number of neurons as the input layer, while the number of neurons in the layers after the first is chosen dynamically to optimize the loss function. We use mean squared error (MSE) as the loss function for our experiment. We also perform parameter tuning using these different activation functions, optimizers, and data models and try to find the right combination that optimizes our result.

4.3 Conversion Prioritization

One we have a model that captures the cost associated with the vacant features for each property, we can modify the data in a way which would reflect the changes in the neighborhood if all the vacant lots are converted. This gives us the pre- and post-conversion property values, helping us determine which vacant lots are causing the most impact in the neighborhood. As mentioned earlier, conversion of all the lots in the city is infeasible; therefore, choosing lots to convert is based on budget constraints.

This then becomes similar to a bin packing problem [23], where the number of bins would be the number of vacant lots that can be converted based on the city's constraints and selecting the vacant lots from the entire pool so as to optimize profit is the goal. Having to check every possible combination of vacant lots to convert is an NP-hard problem. However, it is possible to sort the vacant lots in the decreasing order of property value impact and then the first fit for the budget can be found.

5 Results and Discussion

In this section we perform experiments to evaluate how well our data and machine learning models perform on real-world datasets collected from the City of Rochester. Our aim is to identify how well our models perform compared to currently used techniques used for the purpose of property value impact prediction.

We begin by visualizing mean changes in property values with respect to the number of vacant lots in each block. That is, we group the properties based on the number of vacant lots that are present in its block and then average the property values within the groups. As it can be seen from Figure 3, the average

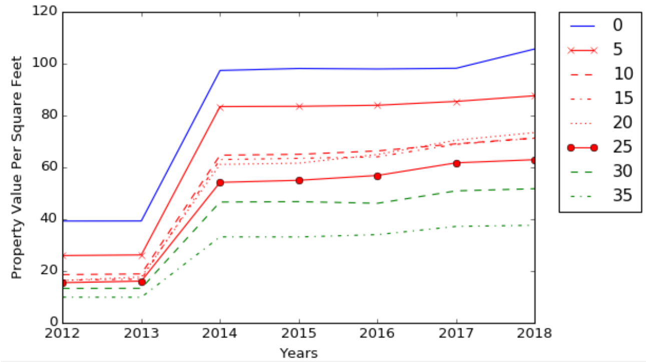

Fig. 3: Change in average property values over the years. Each line represents the average property value on blocks with varying number of vacant lots

property values tend to be showing intuitive results, with properties without any vacant lots nearby having the highest property values, and as the number of vacant lots increase, the average property values seem to be lower. However, the variance was very high for these averages, and therefore it is not possible to directly understand the changes in property values using averages.

Next, we plot similar graphs with the number of 311 calls (non-informational) made, number of code violations, and vacant lots, and found similar results, but they too suffered the same problem with high variance. However, when we tried plotting the relationship between the average number of crime incidents in the block and number of vacant lots, the graphs did not show any correlation.

5.1 Gaussian Process Regression

This section discusses the results from the experiments described in Section 4.2. We used the hierarchical data model to run linear and Gaussian Process Regression (GPR) algorithms to understand whether the independent variables are able to predict the change in property prices. We tried using different kernels for GPR to find the best fit for our data. We used data from the city of Rochester, with the target variable being the ratio of property values in the year 2018 to property values in 2014. These two years were chosen in particular as the housing market prices, after having suffered a sudden drop during the financial crisis, has shown signs of improvement in recent years. As it can be observed in Figure 3, the mean property prices have been improving significantly between 2013 and 2014, and then becomes steadier after that.

The results obtained, as shown in Table 3, show that Gaussian process regression models are able to provide slightly better results compared to basic linear regression. However, this improvement is not sufficient to correctly predict property value changes as on an average, the Matern 5/2 kernel would give an error of approximately 11%. This shows that even with a non-parametric probabilistic model, the relationships in a social setting is difficult to establish, showing the need for more complex multi-layer models.

Table 3: Results for different regression models

Model	Kernel	MSE
Linear Regression	-	0.01406
Gaussian Process Regression	Rational Quadratic	0.01320
	Exponential	0.01297
	Squared Exponential	0.01344
	Matern 5/2	**0.01294**

5.2 Artificial Neural Networks

Similar to GPR, we used the same hierarchical model as input to our neural networks. We gave this as input to a network with a single hidden layer having 34 neurons (equal to the number of features). More hidden layers were incrementally added and number of neurons reduced to optimize the errors. Upon tuning these two parameters, the optimal results were obtained with 4 hidden layers- three with 30 neurons each and the last layer with a single neuron.

We then trained the model by using different combinations of the optimizers and activation functions to find the model with the least mean squared error (MSE). As shown in Figure 4, the results for sigmoid and relu activation functions were close to each other, while the use of tanh as activation function gives much higher error comparatively. The best results were obtained using sigmoid as the activation function and Adagrad as the optimizer. The error seems to flatten out as the number of epochs approach 1000, with the MSE for this combination being 0.00493, which translates to an average error of 6-7%. This model outperforms other regression-based models commonly used in social science literature, and provides better estimates about the effects of vacant lots on nearby property values. This confirms our intuition that the use of deep learning for predicting property value prices in the neighborhood shows much more promising results than linear and Gaussian process regressions.

These observations lead us to answer our research questions.

RQ$_1$: *What impacts do vacant lots in a neighborhood have on the property values of that neighborhood?* The process of modeling social relationships is complex, and it is likely that even with all the available data, key neighborhood characteristics are lost in the modeling process. However, by exploiting existing literature, we were able to include some of the most relevant feature. We used neighborhood aggregation, which was not used in prior literature, to compare properties within a neighborhood rather than comparing with an entire city. While different studies have shown varying results, the results we obtained using conventional regression models were inadequate to make property value predictions based on vacant lot features.

With the use of neural networks, the performance seems to improve significantly, allowing for better predictions. Based on experimental results, we conclude that a hierarchical data model combined with a deep neural network architecture can be used to capture neighborhood characteristics and perform property value predictions with low error margins. We used the model thus

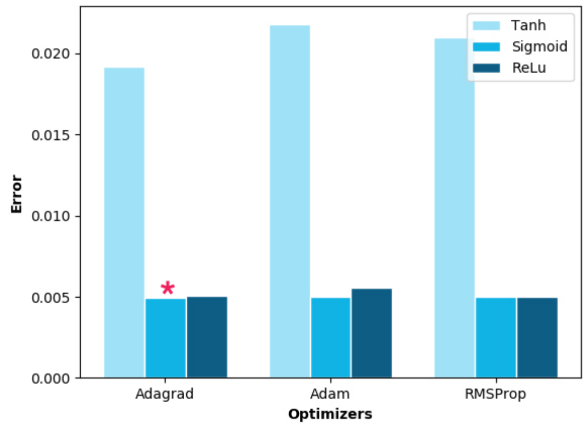

Fig. 4: Hyperparameter tuning for deep learning

trained and changed the data to reflect the conditions that would occur if all the vacant lots in the city are converted. That is, almost all of the features related to vacant lots would become zero. We use this as the input to our model for prediction. Based on the results obtained, by converting every vacant lot in the city, the total property value increase in Rochester is approximately 1.54 million.

RQ$_2$: *How can we choose vacant lots to convert so as to maximize the benefits of conversion (minimize the negative impacts on property values if the lots are not converted)?* While it is not feasible for most cities to re-purpose every vacant lot, based on the data obtained from the model, it is possible to sort out the vacant lots that have the highest impact on nearby property values. However, it is not necessary to convert every single vacant lot near a property to observe improvement in property values. That is, the effects of vacant lots are observed when there is a cluster of such lots near a property. Converting even a couple of these lots can bring about changes to the property values.

To optimally chose vacant lots to convert, it is necessary to iteratively change vacant lot density data for each property to reflect conversion and test the change in property value with those parameters. If the budget allotted by the city for vacant lot conversion is x, it is necessary to try every possible combination of vacant lots that can be converted, and the corresponding total change it would bring to property values. This can then help order the vacant lots in the descending order of impact and the top x vacant lots can be selected for intervention.

5.3 Social Implications

As it was demonstrated in this study, we were able to use data that was publicly available to build models that can predict the impact of vacant lots on neighborhood property values. However, the key motivation for understanding this impact was to show that it is possible to gather evidence from within the city to drive changes in city policy. While this study was restricted to one use case and one city, similar implementations can help both city officials and residents derive evidence for other urban problems as well.

With the prioritization of vacant lots based on impact, it becomes possible for city officials to find locations where interventions can bring about the most impact. These interventions can be in the form of incentives to residents for fostering conversions of these lots, or through investments or subsidization by the city that might make these lots more desirable for purchase. While different studies have shown the impact of conversions of these lots, the data about these impacts is difficult to acquire. With such data, it would also become possible to recommend actions that yield the best outcome.

However, unlike conventional machine learning applications, the use of data for urban planning decision making can have implications on the lives of cities' residents. As demonstrated in this paper, it is possible to apply optimization techniques on social problems and minimize for errors, but without thoroughly understanding the reasoning behind why a model has given a particular result or recommendation, it would be risky to deploy it for decision making.

With the vacant lot impact assessment tool, the same problem arises. While the model was able to provide better performance compared to Gaussian process or linear regression, it isn't apparent what led the model to made these conclusions. Since the key set of features included demographics, it is possible that the model might have learned with inherent biases. Another problem that might arise could be gentrification. In an ideal scenario where all vacant lots get converted or sold, it is likely that the real estate demand would go up in an area. This can further lead to increase in property value assessments and subsequently, higher property value taxes, leading to gentrification. Although this is speculative, since it affects the lives of citizens, it is always better to err on the side of caution. We therefore believe that it is necessary to improve the model to (1) provide better explainability before deployment, and (2) conduct a longitudinal impact study about the impact of conversions on different neighborhood factors.

6 Limitations

Firstly, with the large number of unknown variables that might occur in a social setting, it becomes almost impossible to do a intervention-control study to understand the causal relationships behind the impact of vacant land on neighborhood property values. Such causal studies have been performed in social sciences in various cities, and therefore we make the assumption that the same causal relationships exist in Rochester as well.

Secondly, while we have done our best to ensure that the features collected and used are as accurate as possible, there still exists the possibility that the changes in property values were also tied to some unknown variable. Based on the discussion with the city assessor, it was evident that vacant lots do not directly impact property assessments, but rather have a more indirect impact through real estate values. In fact, it was mentioned during the meeting that no neighborhood factors (crime, blight, demographics, etc.) are taken into consideration when assessing properties.

7 Conclusion

Urban data is often under-utilized yet highly valuable in making informed policy and urban planning decisions. One domain where the exploitation of such data can bring about positive change is the vacant lot problem. While the effects of property abandonment and subsequent generation of vacant lots have been extensively discussed, the models built using these methodologies are inaccurate to be used to understand the effects vacant lots have on a smaller scale. Moreover, no tools and models for using this knowledge to make informed decisions exist. With our research, we propose a novel way of representing data so as to capture key neighborhood characteristics, and also understand the impact of vacant lots on neighborhood property values. We also propose a deep learning framework that can predict changes in property values with respect to a set of vacant-lot-related features. We then show experimental evidence that our model shows better results compared to baseline methods. Unlike other models, our model caters to small and mid-sized cities,making it easier to make informed policy decisions while taking budget constraints into consideration.

Notwithstanding the improvement and accuracy obtained, some directions exist for future work. Firstly, this framework could further be improved by using recurrent neural networks and using time series crime and 311 call data. Secondly, only limited data was made available to us about property values. With data spanning longer periods of time along with observable conversion of vacant lots for residential or public use, it would be possible to generate better estimates about the impact of conversion depending on what the lot is being re-purposed into. Lastly, with data from multiple cities, a model learned from one city can be transferred to another without the need for re-training using transfer learning.

References

1. Accordino, J., Johnson, G.T.: Addressing the vacant and abandoned property problem. Journal of Urban Affairs **22**(3), 301–315 (2000)
2. Branas, C., Cheney, R., M MacDonald, J., Tam, V., Jackson, T., R Ten Have, T.: A difference-in-differences analysis of health, safety, and greening vacant urban space. American journal of epidemiology **174**, 1296–306 (11 2011)
3. Branas, C.C., South, E., Kondo, M.C., Hohl, B.C., Bourgois, P., Wiebe, D.J., MacDonald, J.M.: Citywide cluster randomized trial to restore blighted vacant land and its effects on violence, crime, and fear. Proceedings of the National Academy of Sciences (2018)
4. Brown, E.R.: The vacant lot problem in american cities. American Journal of Economics and Sociology **17**(1), 41–42 (1957)
5. Burchell, R., Listokin, D.: Property abandonment in the united states. In: The Adaptive Reuse Handbook: Procedures to Inventory, Control, Manage, and Reemploy Surplus Municipal Properties. Rutgers University, Center for Urban Policy Research (1981)
6. Cui, L., Walsh, R.: Foreclosure, vacancy and crime. Journal of Urban Economics **87**, 72 – 84 (2015)

7. Garvin, E., Branas, C., Keddem, S., Sellman, J., Cannuscio, C.: More than just an eyesore: local insights and solutions on vacant land and urban health. Journal of Urban Health **90**(3), 412–426 (2013)
8. Garvin, E.C., Cannuscio, C.C., Branas, C.C.: Greening vacant lots to reduce violent crime: a randomised controlled trial. Injury Prevention **19**(3), 198–203 (2013)
9. Goldstein, J., Jensen, M., Reiskin, E.: Urban vacant land redevelopment: Challenges and progress (2001)
10. Han, H.S.: The impact of abandoned properties on nearby property values. Housing Policy Debate **24**(2), 311–334 (2014)
11. Heckert, M.: Access and equity in greenspace provision: A comparison of methods to assess the impacts of greening vacant land. Transactions in GIS **17**(6), 808–827 (2012)
12. Heckert, M.: A spatial difference-in-differences approach to studying the effect of greening vacant land on property values. Cityscape **17**(1), 51 (2015)
13. Huang, C., Zhang, J., Zheng, Y., Chawla, N.V.: Deepcrime: Attentive hierarchical recurrent networks for crime prediction. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. pp. 1423–1432. CIKM '18, ACM, New York, NY, USA (2018)
14. Immergluck, D., Smith, G.: The external costs of foreclosure: The impact of single-family mortgage foreclosures on property values. Housing Policy Debate **17**(1), 57–79 (2006)
15. Immergluck, D., Smith, G.: The impact of single-family mortgage foreclosures on neighborhood crime. Housing Studies **21**(6), 851–866 (2006)
16. Inman, R.P.: Making cities work: Prospects and policies for urban America. Princeton University Press (2009)
17. Pagano, M.A., Bowman, A.O.: Vacant land in cities: An urban resource. Brookings Institution, Center on Urban and Metropolitan Policy Washington, DC (2000)
18. Rasmussen, C.E., Nickisch, H.: Gaussian processes for machine learning (gpml) toolbox. Journal of machine learning research **11**(Nov), 3011–3015 (2010)
19. Rosen, S.: Hedonic prices and implicit markets: Product differentiation in pure competition. Journal of Political Economy **82**(1), 34–55 (1974)
20. Schilling, J., Logan, J.: Greening the rust belt: A green infrastructure model for right sizing america's shrinking cities. Journal of the American Planning Association **74**(4), 451–466 (2008)
21. Sternlieb, G., Burchell, R.W., Hughes, J.W., James, F.J.: Housing abandonment in the urban core. Journal of the American Institute of Planners **40**(5), 321–332 (1974)
22. Yang, Y., Liu, Z., Tan, C., Wu, F., Zhuang, Y., Li, Y.: To stay or to leave: Churn prediction for urban migrants in the initial period. CoRR **abs/1802.09734** (2018)
23. Yao, A.C.C.: New algorithms for bin packing. J. ACM **27**(2), 207–227 (Apr 1980)
24. Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction. CoRR **abs/1610.00081** (2016)
25. Zheng, Y., Capra, L., Wolfson, O., Yang, H.: Urban computing: Concepts, methodologies, and applications. ACM Transaction on Intelligent Systems and Technology (October 2014)