Leveraging AI to Assess Inequities in Pavement Maintenance and Rehabilitation Strategies

A Thesis presented to the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Civil Engineering

by

CHARLES TOWNSEND

Dr. Yaw Adu-Gyamfi, Thesis Supervisor

December 2022

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled.

Leveraging AI to Assess Inequities in Pavement Maintenance and Rehabilitation Strategies

Presented by Charles Townsend,

A candidate for the degree of Master of Science,

And hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Yaw Adu-Gyamfi

Dr. Bill Buttlar

Dr. Timothy Matisziw

ACKNOWLEDGEMENTS.

Firstly, I would like to thank my advisor, Dr. Yaw Adu-Gyamfi, who has always made himself available anytime I have had difficulties with either my research or my course work. He has always found time in his busy schedule to provide advice and direction for anything I have worked on and has supported me thoroughly along my academic journey. I consider myself fortunate to have had the opportunity to collaborate with him.

Besides my advisor, I would like to thank all the professors who have educated me and guided me along my academic journey, Dr. Buttlar, Dr. Sun, and Dr. Matisziw.

I would like to thank my parents, who have provided emotional and financial support for me throughout this academic journey and throughout life. Words alone cannot describe how grateful I am for the support and love they have given me.

Lastly, I would like to thank my friends and colleagues, from whom I have learned much from and whose friendship has helped me through these projects. I would like to highlight my girlfriend, Brianna Davis, whose love and patience has helped me at every stage in this project.

TABLE OF CONTENTS

LIST OF FIGURES:

LIST OF TABLES:

ABSTRACT

Pavement maintenance and repair (M&R) is an important social need that should be accessible to all regardless of social or economic circumstances. In this project, pavement condition is analyzed through the use of machine learning algorithms and then compared to the socioeconomic factors of the surrounding communities. This analysis is crucial for assessing inequities that disadvantaged communities often face, especially pertinent toward pavement M&R policies. In addition to the goal of determining equity, a secondary goal of this project was to develop a methodology that can be repeated across any city. Utilizing an image-set provided by Google Street-View's API and census data of Kansas City, Missouri, a comparative analysis was conducted to determine whether equity between advantaged and disadvantaged groups was present. To aid in pavement distress identification, YOLOv5 (You only look once), a popular deep learning algorithm, was used to identify seven unique pavement distresses across Kansas City Road segments. The resulting methods were able to demonstrate a comparison between pavement conditions and socioeconomic metrics, demonstrating a trend indicating road segments in disadvantaged communities show slightly worse conditions. The strongest correlative factor borne out of analysis shows that median household income demonstrates the greatest gap between advantaged and disadvantaged census blocks.

CHAPTER 1: INTRODUCTION

In modern society, road infrastructure has essential and active roles in the advancement of cities and communities (Amir Shtayat et al, 2022). Proper maintenance and rehabilitation (M&R) of road infrastructure facilities is an essential factor for developing the socioeconomic development of communities. However, according to the United States Government Accountability Office (2022) even when controlling for factors such as climate type and traffic density, pavements are less likely to be in good condition on roads with census tracts with higher percentages of disadvantaged persons. Ensuring that all communities receive an equitable M&R standard is crucial, as mobility is a key factor for reducing poverty and social exclusion (Wachs, M., 2010). Thus, pavement management policy should consider the people affected by M&R activities and should ensure that disadvantaged communities receive an equitable standard comparable to all else in the network.

To make certain that communities receive proper M&R in road facilities, the use and implementation of artificial intelligence and machine learning should be considered. In recent years, applications of AI algorithms and machine learning have proven popular for assessing faults in transportation infrastructure and systems (Vasudevan, M. et al. 2020). The application of this technology has improved the efficiency of modern agencies and has found greater application in adjacent subjects such as autonomous vehicles and traffic detection. Applications of machine learning can be used for more than simply assessing traffic flow or pavement conditions and can be used to ensure that all streets within a transportation system are treated equally regardless of where they are located.

Many authors have used this technology to perform their own surveys of street conditions considering many pavement condition systems (PCS) and including budgetary or scheduling considerations for developing M&R strategies. Equity is considered on occasion, but usually in terms of assessing what streets are in worse condition within a network and optimizing maintenance scheduling. Rarely is the development of a pavement management system (PMS) concerned with emphasizing disadvantaged communities and the socioeconomic factors of those who utilize roadway facilities. Addressing these factors in the development of a PMS is important for maximizing equity within a network.

As for the subject of this project, Kansas City, Missouri was selected to conduct this analysis. The selection of Kansas City was predicated on a few reasons, first being that it is a major city with well documented lines where disadvantaged communities have historically been. A chief example is that of Troost Avenue which represented a historic divide between African American and white citizens (Euston and Reidy, 2020). A second reason is accessibility of data, as Kansas City provides a helpful database of geographic data that was utilized for analysis. These reasons aid in the analysis as presented in this project.

1.1 Problem Statement

This project seeks to address the topic of equity in transportation infrastructure by addressing pavement distresses against socioeconomic factors such as race, income, and education. This will address aspects of horizontal equity, in terms of assessing gaps in coverage across all roads, and vertical equity, in terms of socioeconomic considerations.

The goal is to provide a framework that is able to address these equity goals and to provide a means to visualize the relation between pavement conditions and disadvantaged groups.

1.2 Objectives of Study

The objectives of this study can be summarized as follows:

- The main objective of this thesis is to develop a framework that uses machine learning to assess distresses in pavements and then compares those distresses to socioeconomic factors such as race, income, education, and employment.
- To create a system capable of visualizing pavement conditions against socioeconomic factors.
- To develop a scoring system for assessing the condition of pavement using PASER as a foundation.
- Make use of open-source and accessible data for cost-effective analysis.

CHAPTER 2. LITERATURE REVIEW

To better break down and understand the subject of this project, an extensive literature review was conducted to investigate relevant topics. The key focus of this literature review was to investigate how the quality of transportation infrastructure affects social groups and how pavement conditions are typically assessed. Other topics of concern included policy considerations in transportation management systems and pavement distress indexing/identification methods. Below is the summary of this investigation.

2.1 Equity in Transportation

Within the literature, there are many papers that discuss the topic of equity as it relates to transportation or infrastructure in general. Equity refers to the distribution of impacts (benefits and costs) and whether that distribution is considered fair and appropriate (Litman, 2017). Two distinctions of equity, horizontal and vertical equity, describe specifically what equity means with the former assuming that people with similar needs and abilities should be treated similarly while the latter assumes disadvantaged people should receive favorable treatment. As such, equity assumes that so long as all else is equal, similar treatment should be applied, however if a group is disadvantaged then greater consideration or focus should be given to that group.

Litman (2017) provides an extensive breakdown on the subject of equity in transportation by breaking down the distinction between horizontal and vertical equity. Horizontal equity addressing fairness with regards to resource allocation and external costs imposed by travel activities and vertical equity addressing inclusivity with how the system serves its population, affordability with understanding how the system impacts lower-income people, and social justice which addresses structural injustices such as racism and sexism. In addition, the author discusses two types of equity strategies called programmatic strategies and structural strategies. Programmatic strategies provide special benefits to designated people such as universal design, special mobility services, and senior/student transit fare. Structure strategies include reform planning practices to create more inclusive, affordable, and resource-efficient transportation systems such as multimodal planning, pricing reforms, and smart growth development policies.

Returning to an article mentioned in the introduction of this paper, the findings of the United States Government Accountability Office (2022) do point to inequity in pavement conditions for disadvantaged communities. This article references data taken from Federal Highway Administration (FHWA), which assessed whether states are making progress towards state-wide pavement condition targets, as well as other agencies to determine how pavement conditions compare against key categories in census metrics. The article found that even when accounting for factors such as traffic density and climate type, pavement is less likely to be in good condition on roads with census tracts that have higher percentages of underserved racial and ethnic groups and higher poverty rates. The article concludes with recommendations calling for executive action to analyze and address roads with respect to the local community or other characteristics (such as race, ethnicity, and poverty) and that pavement management strategies should include these factors when determining maintenance selections.

Although establishing that disadvantaged groups need greater attention is important to recognize, Wachs, M. (2010) outlines how poverty and poor access to transportation are intrinsically related. The article discusses how social and physical mobility is severely hampered within disadvantaged communities and how this is worsened when considering the intersection of gender, race, physically disabled persons, etc. In a similar vein, Lucas et al. (2016) discusses that the poorest groups tend to be less mobile and often suffer from a lack of both private and public transport services. This is emphasized by how poor urban areas located at the edges of cities with low amenity value lack local employment opportunities and basic facilities. This coupled with lack of access to transportation options creates what the authors call a 'poverty trap' which limits access to jobs, education and health facilities, etc.

In the article written by Manaugh & El-Geneidy (2012), the effects of proposed transit infrastructure in Montreal, Canada are analyzed through the lens of social equity. The stated purpose of the paper is to examine the extent to which proposed transit infrastructure projects in the city transportation plan benefit disadvantaged populations with many metrics measuring travel time and accessibility to commercial centers as core concerns. The article found that from a regional standpoint the introduction of new transit infrastructure would improve access to employment centers as well as reducing travel time for regular commutes. This shows that proper investment towards disadvantaged communities can result in tangible benefits even if only on a microscale. This perspective is also seen in Lewis (2011), which covers the negative economic impact of poor transportation and how better investment can better improve worker productivity.

In the article written by France-Mensha et al (2019), a series of optimization models that consider environmental and socially equitable factors for determining roadway maintenance policy. The paper addresses that several highway agencies lack specific evaluation standards to facilitate the comparison of alternative highway M&R schemes to assess social equality performance in budget-constrained programs. Thus, these models sought to measure social equity by comparing economically disadvantaged communities (SED) to pavement assets in the rest of their network. Of the tested policies, Policy C, which supports the unequal allocation of resources to reduce the gap in performance, showed the best overall performance at reducing the gap between advantaged and disadvantaged groups and reducing inequity.

Naseri et al (2020) addresses the topic of developing an equitable maintenance and rehabilitation scheduling system on a network scale. The author identifies that agencies tend to focus on improving pavement conditions while minimizing maintenance expenditures at the expense of equity. The author writes that equity is considered ideal when all segments are in identical condition as road users will compare the roads, they drive on to other parts of the road network in their city or country. The article utilizes water cycle and genetic algorithms with an equity index that directly considers the International Roughness Index to develop an optimized plan for pavement management.

2.2 Maintenance and Rehabilitation Strategies

To facilitate the proper maintenance and rehabilitation (M&R) of pavements, it is important for agencies to have strategies in place to manage their road networks. A common approach for managing networks is the development of pavement management systems (PMS) which can more efficiently address M&R of road sections. PMS often include an inventory of pavement segments in a network, pavement distress ratings and traffic information, scheduling maintenance, and budget allocation to determine what roads need maintenance and when (Abaza et al. 2004). Generally, a better designed PMS should result in better coverage of road systems as resources can be better distributed for M&R strategies.

An important note is that many solutions for developing M&R strategies involve the implementation of algorithms, API, and models to distribute resources for projects or aid in the decision as to what pavements receive maintenance. Sundin & Braban-Ledoux (2001) wrote about the implementation of AI based decision support technologies and how they can be used for pavement management. Their article serves as a summary of articles covering AI use in PMS with decision tools such as artificial neural networks (ANNs). A more recent example covering a similar topic, Xu and Zhang (2022) covers the topic AI algorithm usage for pavement management including distress detection and classification.

In general, many articles have approached the subject of developing more efficient M&R strategies by using optimization models to aid in road segment selection, budget allocation, and scheduling. Abaza et al. (2004) outlines the development of an integrated pavement management system (IPMS) using a Markovian Prediction Model that considers deterioration rates and improvement rates resulting from M&R actions. The purpose of this model was to aid in the selection process for determining what roads receive maintenance. Concerning budget allocation optimization, France-Mensah and O'Brien (2018) conducted a case study comparing multiple methods of budget allocation factoring pavement condition scores against available budget and network coverage. Similarly, an article written by Boyles et al (2010) created two algorithms to address both long-term and short-term budget allocation and planning to aid M&R scheduling.

France-Mensah and O'Brien (2018) addresses the subject of equity as the fair distribution of M&R funds to in need pavement sections and uses equity as a metric for evaluating their model serving as an integration of horizontal equity in their M&R strategy. Equity, assessed as "equity in outcome" in this article, is a core component of the scoring system laid out by the author and serves as a core consideration for the model's outcomes. The equity consideration would measure the extent each proposed policy would widen or narrow the gap between the highest and lowest pavement conditions scores in the network. Of the three policies tested, cost-benefit analysis and needs based models outperformed their integer-linear programming model.

Kothari et al. (2022) writes on the development of a budget allocation model that accounts for trade-offs of economics, environment, and social equity aspects of sustainable planning. The goal of the article was to develop a pavement management plan that draws a comparison between economic factors, such as network condition and pavement condition, environmental factors, such as GHG Emissions from materials and construction, and social equity factors, minimizing the disparagement between disadvantaged groups and the rest of the network. The article presents a series of policies that either maximize one of these three aspects or optimize for all three.

2.3 Pavement Distress Identification

The process of identifying, classifying, and ranking distresses in pavements is important to understand the constraints related to pavement management. A number of articles have covered the topic of pavement distress identification for the purpose of shortening the time needed to inspect and maintain roadways while also providing more objective measures of their severity.

A major problem when it comes to assessing pavement severity comes from collecting the street data to begin with. Attoh-Okine & Adarkwa (2013) outline the subject of pavement condition surveys and their goals of recording data collection, condition rating, and quality management. In addition, data collection methods such as automatic and manual collection are discussed including the benefits and drawbacks of both approaches. Generally, manual data collection, although inexpensive relative to automatic methods, tends to require longer collection times, risk to personnel, and are subjective in the determination of severity. Automatic methods, despite the potential costs, are more objective and can cover larger datasets in less time.

Returning to a previously mentioned article, Xu and Zhang (2022) provides an overview of AI applications for pavement management. The article reviews work of many authors and their use of various AI algorithms such as artificial neural networks, deep neural networks, and tree-based algorithms greatly assist the assessment of pavement distresses. Although these findings have found that the use of AI algorithms have assisted in M&R strategies, they tend to have trouble assessing distress detection, classification, and quantification all at once.

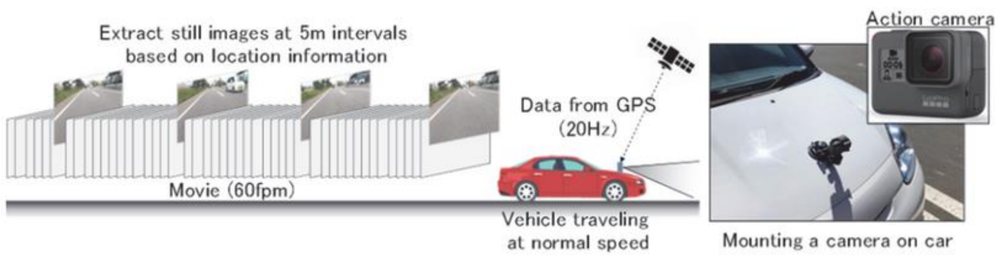

Figure 1. Road monitoring system based on using action cameras. (Asada et al., M.2020)

Reviewing examples of automatic image gathering method applications in the literature, Cafiso et al. (2006) collects images using the high-speed digital acquisition system of a mobile laboratory and then identifies the distresses captured on those images. In a similar vein Asada et al. (2020) utilized car mounted action cameras and U-Net deep learning model to perform automatic pavement distress detection. The mounted camera captures images of roadways as the car drives over, where after the convolution neural network identifies pavement distresses in 5-meter intervals. In both articles, software is utilized for the purpose of image detection and can detect distresses such as potholes, block cracks and alligator cracks. Gopalakrishnan et al. (2017) approaches distress detection of pavements via a pre-trained convolutional neural network to perform detection across more complex surfaces. Their approach utilized large data samples for more accurate distress detection and found greater success in performing detections on more complicated surfaces.

The Article written by Majidifard et al. (2020) details the development of a new pavement condition index utilizing several deep learning models. The approach of this article follows a similar approach of using Google Street-View API to build a pavement image dataset (PID) for distress detection as well as YOLOv5 for the crack classification model but, performs an in-depth classification of the detected distresses using a U-Net density model. The built pavement condition index utilizes the Pavement surface evaluation and rating (PASER) system and is designed for general application across any road system.

CHAPTER 3: METHODS

3.0.0 Design Approach:

This project sought to establish a comparative analysis between street segments and socioeconomic factors while utilizing existing APIs and programming. To provide an overview of this project, several datasets such as street segment data and census data for Kansas City, Missouri were collected for analysis. An image-set of street segments was then generated using Google's Street-View API. The image-set was used to derive pavement conditions along those segments using YOLOv5 to provide a dataset of individual distresses across the pavement segments. Next, census data provided relevant geographic data which gave a means to overlay the distress data over census blocks. Socioeconomic data and pavement distress data were then compared to see if any correlative trends can be gleaned from the comparison. Figure 2 illustrates the steps taken to conduct this analysis.

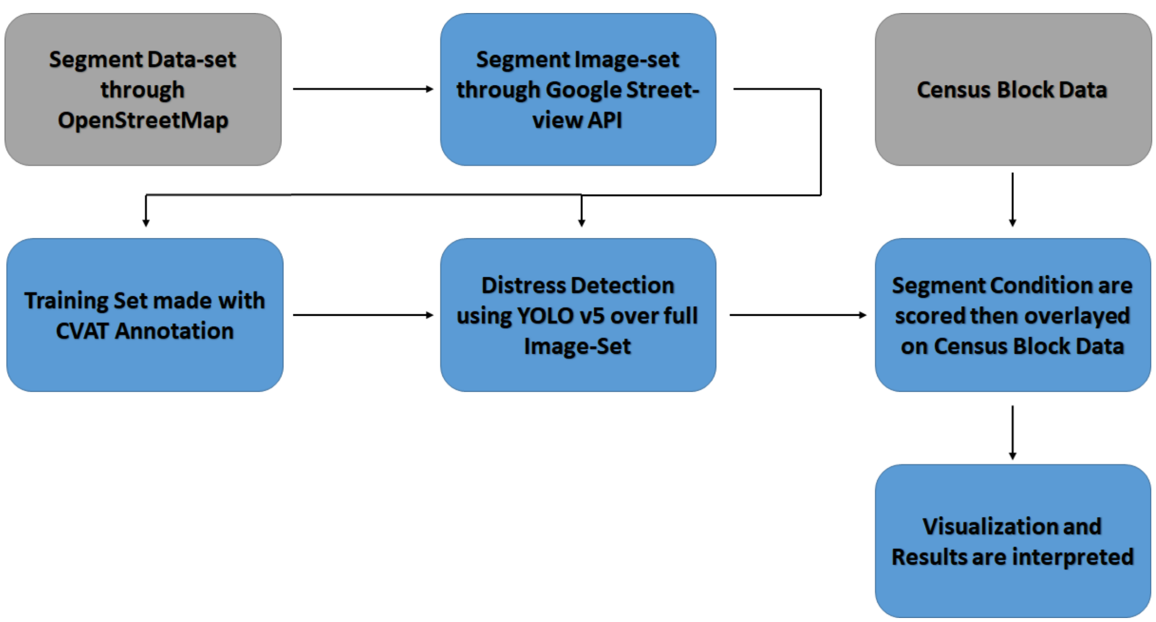

Figure 2. Outline of Project

3.0.1 Determining Equity

To perform the analysis, it is important to establish what equity means in the context of this project. As mentioned in previous sections, equity can be defined in several ways, in terms of vertical and horizontal equity. The analysis of this project is focused on addressing the condition of road segments and equity can be determined by visualizing the gaps between the upper and lower ends of the pavement conditions. In essence, a narrow gap in the quality of roads would suggest equity while a wide gap would suggest otherwise. This, however, is not the only metric this project will consider, as factors such as economic status, race, and employment are considered for establishing groups that are considered disadvantaged. The purpose of establishing these groups is to provide an additional contrast as to how equity can be measured; now if a wide gap exists and it is shown that poorer pavement conditions correlate with disadvantaged communities the argument of inequity is strengthened.

As for how disadvantaged groups are determined, specifically in the context of this project, Littman (2017) outlines groups to include nicely in their section on social justice. Disadvantaged groups as defined in their paper include "racial and sexual minorities, women, immigrants, lower-income groups, etc." including those who lack access to basic services and activities such as education, jobs, and healthcare. The data for this project is derived from census data which contains information specific to the above-mentioned categories which will aid in the equity analysis. For this, the categories of race, income, employment, and education status will be used for determining disadvantaged groups in the data.

Returning to the determination of equity, the selected categories used for determining disadvantaged communities will be selected and directly compared to the condition of pavements located within their census blocks. Once the pavement sections are arranged from best to worst, the disadvantaged groups categories are directly compared to the scores given to those pavements and any trends or correlations that may exist will be determined. Special consideration will be made towards blocks that express multiple or all of these characteristics and weighed more heavily in the analysis. In addition, location and the number of segments in a census block will be considered due to some census blocks having less road segments than others.

3.0.2 Condition Rating:

Determining a good system for rating the condition of pavements is a chief consideration for this project. Attoh-Okine, N., & Adarkwa, O. (2013) discusses various pavement condition systems (PCS) such as the Present Serviceability Index (PSI), Condition Rating Survey (CRS), and Pavement Surface Evaluation and Rating System (PASER). Each PCS has their own rationales and advantages and are adopted by different states to fulfill their agency's specific needs or preferences. Not all PCS use the same metrics nor use the same scales; as an example, the Pavement Condition Index (PCI) has values associated with distress types and severity and determines a condition score are based on a 0-100 scale, while the Present Serviceability Rating (PSR) scores based on ride quality and are based on a 0 to 5 scale.

For this project, PASER is used as the pavement condition system that influences the condition scores in the later analysis. The rationale for using PASER to influence the scoring system is due to ratings being estimated based on visual evaluation, which is the primary means that this project uses for analysis. Another advantage is that the types of distresses as well as the severity often have thresholds that determine their score. As an example, new pavements with no distress are considered a rating of 10-9, first signs of reflection cracking and transverse cracking indicates a score of 8 at max, and the appearance of alligator cracking implies a maximum score of 3. Figure 3, provided by the University of Wisconsin-Madison (2002) in their PASER Manual, shows in greater detail how PASER's 0-10 based scoring system can be broken down.

Rating system

Surface rating	*Visible distress**	*General condition/ treatment measures*
10 Excellent	None.	New construction.
9 Excellent	None.	Recent overlay. Like new.
8 Very Good	No longitudinal cracks except reflection of paving joints. Occasional transverse cracks, widely spaced (40' or greater). All cracks sealed or tight (open less than 1/4").	Recent sealcoat or new cold mix. Little or no maintenance required.
7 Good	Very slight or no raveling, surface shows some traffic wear. Longitudinal cracks (open 1/4") due to reflection or paving joints. Transverse cracks (open 1/4") spaced 10' or more apart, little or slight crack raveling. No patching or very few patches in excellent condition.	First signs of aging. Maintain with routine crack filling.
6 Good	Slight raveling (loss of fines) and traffic wear. Longitudinal cracks (open 1/4"–1/2"), some spaced less than 10'. First sign of block cracking. Sight to moderate flushing or polishing. Occasional patching in good condition.	Shows signs of aging. Sound structural condition. Could extend life with sealcoat.
5 Fair	Moderate to severe raveling (loss of fine and coarse aggregate). Longitudinal and transverse cracks (open 1/2") show first signs of slight raveling and secondary cracks. First signs of longitudinal cracks near pavement edge. Block cracking up to 50% of surface. Extensive to severe flushing or polishing. Some patching or edge wedging in good condition.	Surface aging. Sound structural condition. Needs sealcoat or thin non-structural overlay (less than 2")
4 Fair	Severe surface raveling. Multiple longitudinal and transverse cracking with slight raveling. Longitudinal cracking in wheel path. Block cracking (over 50% of surface). Patching in fair condition. Slight rutting or distortions (1/2" deep or less).	Significant aging and first signs of need for strengthening. Would benefit from a structural overlay (2" or more).
3 Poor	Closely spaced longitudinal and transverse cracks often showing raveling and crack erosion. Severe block cracking. Some alligator cracking (less than 25% of surface). Patches in fair to poor condition. Moderate rutting or distortion (1" or 2" deep). Occasional potholes.	Needs patching and repair prior to major overlay. Milling and removal of deterioration extends the life of overlay.
2 Very Poor	Alligator cracking (over 25% of surface). Severe distortions (over 2" deep) Extensive patching in poor condition. Potholes.	Severe deterioration. Needs reconstruction with extensive base repair. Pulverization of old pavement is effective.
1 Failed	Severe distress with extensive loss of surface integrity.	Failed. Needs total reconstruction.

** Individual pavements will not have all of the types of distress listed for any particular rating. They may have only one or two types.*

Figure 3. PASER Rating System Outlined (Walker et al. 2001)

For this project, pavement distresses will be categorized individually and given a score to represent their severity. For each distress category a numeric score is assigned; these scores are defined by the rationale outlined by PASER. The distresses considered for this project are alligator cracking, block cracking, transverse cracks, longitudinal cracks, sealed cracks, patching, and manholes (only for detection purposes as they will not impact segment's condition score). Table 1 shows the scores assigned to these distresses.

Table 1. Distresses and the Assigned Scores

Distress	Score
Alligator Cracking	2
Block Cracking	4
Transverse Crack	8
Longitudinal Crack	7
Sealed Cracks	9
Patching	9
Manhole	No Score

As for how the scores are assigned to pavement segments, pavement segments are assigned ratings based upon all detections seen in the segment. Every detection for each distress is recorded along each street segment, then the totals are multiplied by their assigned score, and finally are divided by the total detections across the segment. Following the condition scoring, that data will then be overlaid within the census block data and the mean across those segments will represent the overall condition score for the census block. To mathematically represent this, the following equations were derived:

$$C_s = \frac{n_1(s_1) + n_2(s_2) + \ldots + n_i(s_i)}{n_1 + n_2 + \ldots + n_i}$$

$$if\ C_s \in S, B_i = \sum_{c \in S} \frac{C_s}{|S|}$$

Table 2. Variables used for pavement condition calculations

Value	Description
C_s	Condition score for segment
s_i	PASER value for distress
n_i	Number of detections made across pavement segment
S	Set of condition scores within a census block
B_i	Condition score for census block

3.1.0 Data Acquisition

With the framework of the project established, it is important to discuss the data required to conduct this analysis. A series of datasets were gathered and generated to formulate the final dataset used in the comparative analysis. These datasets included a general road segment dataset generated from OpenStreetMap, segment image-set generated from Google's Street-View API, and census block data sourced from OpenDataKC. The selection of these datasets was predicated on their accessibility as obtaining these datasets is relatively simple and can logically be implemented for any city. The sections following elaborate further on their contents.

3.1.1 Pavement Segment Data

One of the first challenges to overcome in this project is finding data to help produce an image-set to train and test the model. This necessitated the collection of geographic coordinates that correspond to the location of the streets that are located in Kansas City, Missouri, the subject city of this project. There are many ways to obtain data of street coordinates for most major cities, often available in GIS databases or public open-source databases. For this project, street coordinates were obtained in the form of street segments from the OpenStreetMap database. OpenStreetMap is a community driven open-source database of map data utilized for use in websites, apps, and hardware devices. The main benefit of using OpenStreetMap data is due to its open data policy and the API integration for coding purposes.

The data used for this project included the street segment data selecting for drivable roads in Kansas City, Missouri. The segments of Kansas City selected are those that comprise the city proper within the state of Missouri, including the streets located within the counties of Jackson, Clay, Platte, and Cass. Data included in this dataset include street name, highway (road type), segment length, street coordinates, etc. Segments represented in this dataset are measured at one kilometer or less in length. Later in this project the data will be aggregated to select for roadways around residential areas to better focus the analysis of this paper towards streets where residents and businesses reside. A proper breakdown of the dataset and how it is modified for this project is touched upon later in this paper. Visualization of the street segments before and after data aggregation are present in fig 4.

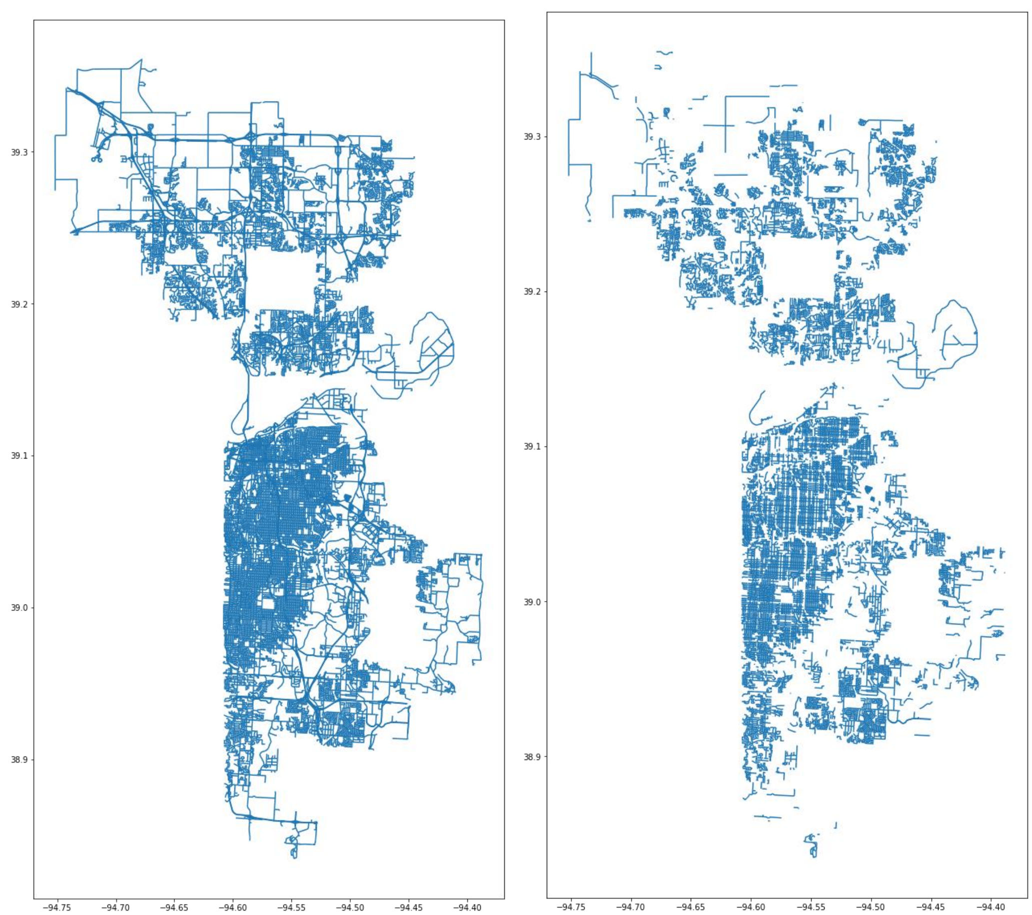

Figure 4. The full data set of street segments (left) and the narrowed data set of street segments (right).

3.1.2 Google Street-View Image-Set Generation

With the acquisition of street segment data an image set was able to be generated across all road segments. The image set was needed to train and test the YOLO model so that a proper dataset of pavement condition could be generated. There are many ways to collect image-sets for this type of analysis, but for this project Google's Street-View API was utilized to generate images using the street segment data. The Street-View API can take the geographic coordinates from the segment data to generate images that correspond with those street segments. The generated images are static snapshots taken from a panoramic image that, with specified conditions, can be taken in any direction within 3-Dimensional space.

The images generated by the Street-View API are sourced from a library of images and videos stored online via Google's servers. These images were originally gathered using panoramic cameras gathered by third party sources or through google themselves. Gathered images are used by Google for their maps and street-view API to be used for various apps like Google Maps, Google Earth, and Google Street-view. Images can be extracted from the API for data processing or research purposes given the proper inputs; as it relates to this project, images of streets are the primary interest.

Images sourced from this API are collected regularly and updated frequently with Google providing a yearly schedule of when and where new image-set collection is occurring across the world. This said, factors such as weather, road closures, and available workers can cause image segments of cities to be outdated leading to their database being inconsistent based on when images were taken. This has the consequence that the image set used for this project has images that we collected across several years. This means that when condition is assessed across these segments, it is important to understand that the images gathered are not necessarily the most current.

3.1.3 Socioeconomic Data

With segment image data fully gathered, the last information needed for analysis was socioeconomic and demographic data. For this project, relevant socioeconomic and demographic data was collected from the 2010 Census Block Group Data provided by the OpenDataKC website and sourced by the U.S. Census Bureau (Lebofsky, 2014). Census block data was selected for this analysis due to its accessibility, as well as its ability to be mapped visually using geographic coordinates. By default, this data does not contain geographic polygons to represent the census blocks visually, so a file containing census block coordinates was obtained from the Mid-America Regional Council's (MARC) website. Information on how these datasets is modified is discussed later in this paper.

As is implied by the dataset's title, data contained in this dataset was collected through censuses and surveys from the citizens of Kansas City, Missouri in 2010. The data covers many columns of data touching on many aspects of demographic information such as race, income, and employment segmented across census blocks. For this project, racial demographics, median income, employment and unemployment, and educational status was extracted from the dataset including geographic polygons representing the census blocks of Kansas City.

To better describe how the data contained in the census block data will be used, these categorizations will be used to create contrasts between advantaged and disadvantaged groups based upon the logic presented in the previous sections. This will be done by recognizing the baseline or average assumptions to draw a comparison against. For race, according to the 2010 census blocks data, the majority racial population category for Kansas City, Missouri is roughly 69.8% as compared to minority populations of roughly 26.95%. Median income is measured as $49,916 in 2010 with the percent of population below the poverty line measured as 15.9%. Unemployment is measured at 9.74% for the overall labor force and education shows 12.55% had less than a high school diploma and 26.27% had a bachelor's degree or greater. The purpose of establishing these baselines will be used for comparison when interpreting results. For example, census blocks that feature a greater minority percentage than the white population will be considered a disadvantage census block and census blocks that feature a poverty percentage greater than the average for Kansas City or the national average will also be considered disadvantaged. Table 3 features the above-mentioned census block information in addition to 2020 census data as reported by Census Reporter, an open source hub for compiling census data.

To note, census data from 2020 was not used for this analysis due to a lack of census block specificity in that dataset. The 2010 Census Block dataset provides information with deference to each census block, making the information more useful for broader analysis. It is still important however, to consider the information collected in the 2020 census data for analysis in the later parts of the paper.

Table 3. Census Data Breakdown of Analyzed Categories.

Category	Sub-Category	2010 (census block data)	2020 (census reporter)
Race	White%	~69.8	~56%
	Minority%	~26.95	~32%
Median Income	Median Income	$49,916	$63,396
	Poverty%	15.9%	~13.4%
Employment	Labor force Unemployed%	9.74%	NA
Education	Less than HS diploma%	12.55%	~7.6%
	HS diploma & less than BA degree%	60.57%	~53.8%
	BA degree or greater%	26.27%	~38.7%

3.2.0 Data Processing

The section will outline the methodology considered for the training and testing of the YOLO model used for this project. In brief, a small segment of images was obtained from the image-set to train the model. Once the model is trained, the remaining image-set is fed into the model in batches to detect the distress across all segments. These detected distresses are then compiled into a single dataset and then utilized for the final dataset.

3.2.1 Training Image Set

Before the model can be used for detecting pavement distresses from images, it is important to create a training dataset to teach the model to perform that detection. To facilitate this, a small sub section of the generated images was selected from the broader image set to be annotated for the training set. This necessitates the Computer Vision Annotation Tool (CVAT) which is an open-source web-based image and video annotation tool. With CVAT, a sub-selection of 403 images from the image-set was selected and various distresses were annotated manually for those images. The specific distresses identified in these images include alligator or fatigue cracking, block cracking, transverse cracking, longitudinal cracking, sealed cracked, and patching. Manholes were also used for detection, but this was only done to differentiate them from patch work. Although there are many more distress categorizations outside of those outlined in this paper, these distresses were selected for analysis due to their commonality across all pavements and ability to be seen clearly on the pavement's surface. Table? shows examples of the specified distresses as well as the scores they were assigned earlier in the paper.

Table 4. Distress Types and IDs

Distress Type	Distress ID	Score	Image Example
Alligator Cracking	cls0	2	
Block Cracking	cls1	4	
Transverse Cracking	cls2	8	
Longitudinal Cracking	cls3	7	
Sealed Cracks	cls4	9	
Patching	cls5	9	
Manhole	cls6	n/a	

3.2.2 YOLOv5 Distress Detection

With a training image-set made, the next step was to use that image-set to train the You Only Look Once (YOLO) version 5 algorithm to detect the pavement distresses across the rest of the image-set. To describe YOLOv5 in greater detail, YOLOv5 is an object detection algorithm that was designed to create features from input images which then uses those features through a prediction system to draw boxes around objects and predict their classes (Solawetz, 2020). The YOLOv5 networks consist of three components, the backbone which is convolutional neural network that aggregates and forms image features at different granularities, the neck which is a series of layers to mix and combine image features to pass for prediction, and the head which consumes features from the neck and takes box and class prediction steps (fig 5, Solawetz 2020).

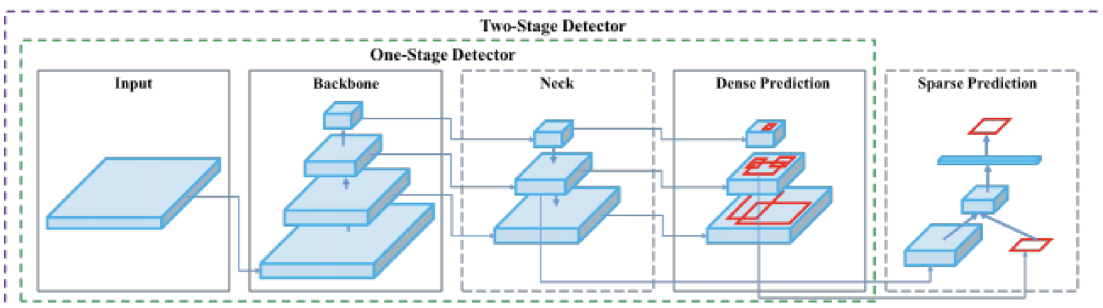

Figure 5. YOLO Architecture (Solawetz, 2020)

The training image set was used to train the model to detect the pavement distresses. Once trained, the model was given batches of images from the broader image-set. The resulting files were then compiled together to form a complete dataset of all distresses collected across the all street segments. Model performance and accuracy are elaborated on later in this paper.

3.3.0 Condition Scoring for Census Data

After reviewing the annotations made by the model and making small adjustments as needed, the next step is to take the resulting file and score the street segments. The scoring was done following the rationale expressed during the condition scoring segment of this paper by assigning values to each segment based upon the total number of distresses detected and their assigned scores. Once the street segments were scored, the geographic coordinates of the road segments were used to conduct a spatial join with the census data, taking every street that fell within the boundaries of that census block. From here all scores that fell within the census block were averaged creating a single unique score for each block. All census blocks that possessed no pavement segment data were removed from the dataset. Figure 6 shows the census block data after the previous steps were taken.

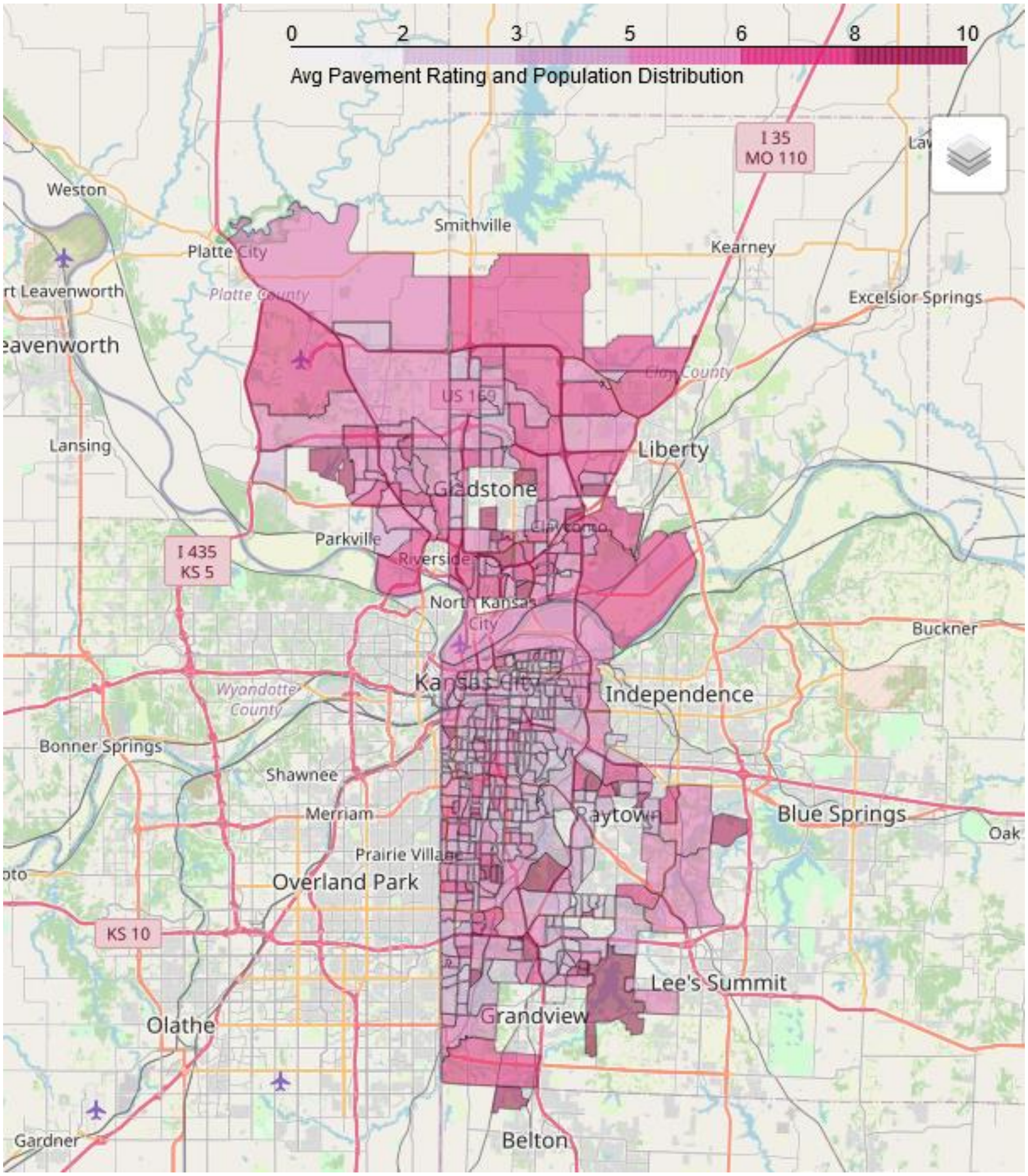

Figure 6. Visualization of Census Blocks Overlaid with Pavement Condition Scores

CHAPTER 4: DATA ANALYSIS AND RESULTS

4.1.0 Data Analysis

This section will analyze the structure and contents of the various datasets utilized for this project. Street segment data, segment image set, and census block data served as the main sources of information needed to conduct a proper comparison. All utilized datasets provide much needed data to compose the final dataset for the comparative analysis.

Discussing the street segment data collected from OpenStreetMap, the OSMnx python library was used to develop a shapefile of the city streets segments within Kansas City, Missouri proper. The generated shapefile is composed of 32,458 rows and 19 columns of data with each row representing a single pavement segment across 3,689 streets. Relevant data columns utilized for this project include street name, highway (road categorization), and geometry. Of note, the geometry column is composed of coordinate pairs that represent the geographic beginning and end of street segments. The columns were extracted from the data set and another column was added based upon calculating the compass bearing for each street segment. The compass bearing calculation was performed so that generated images would be facing down the roadway segment; this will be elaborated on further when discussing the Google Street-View API. Geographic heading is represented as a compass bearing and is calculated with the following formulas.

$$\beta = atan2(X, Y),$$

$$X = cos\theta b * sin\Delta L, Y = cos\theta a * sin\theta b - sin\theta a * cos\theta b * cos\Delta L$$

Table 5. variables for compass bearing calculations

Value	Description
a	Coordinates for beginning point
b	Coordinates for ending point
β	Compass bearing

The last change of note made to this dataset was the narrowing down to segments that were considered either residential or a living street in classification. The narrowing of road segments is to better focus the analysis on road segments where residents live and work rather than highway systems where there is little to no residency. That said, highway segments that were considered a part of residential areas were kept in the dataset. The narrowed dataset is composed of 22,384 rows of data represented over 3,425 streets. An example of the resulting data frame is shown in fig 7.

street address	coordinates	headings
'1257, West 72nd Terrace, Ward Parkway, Kansas...	38.99745625165482,-94.6064905682812	272.9959498067638
'1248, West Gregory Boulevard, Ward Parkway, K...	39.00015317984346,-94.60562754559737	272.00352397526615
'1249, West 72nd Street, Ward Parkway, Kansas ...	38.99832395,-94.60570584999999	273.8148185048187
"McDonald's, West 79th Street, Kansas City, Ja...	38.98562648358928,-94.60764000012138	270.5562835612352
'1573, West 76th Street, Ward Parkway, Kansas ...	38.9908985,-94.60743725	269.9304918177278

Figure 7. Snapshot of Data Frame

With the street segment data frame properly formatted, the information was then input into the Google Street-View API. For the API to produce an image of a road segment, the API requires a number of parameters such as geographic coordinates, camera heading/pitch (in terms of 3-dimensional space), the camera's field of view (FOV) and the image’s resolution. With segment data, the centroid and geographic headings of the segments are calculated in order to augment the parameters of the API and a pitch of -90 is used to point the camera towards the ground. Following this, the image resolution is set to 640x640 and the FOV is set to 120 to provide the greatest overall viewability of the generated image. From there, the name of the street as well as the centroid coordinates are appended to a separate column to be used for the image's name. Figure 8 shows samples of generated images.

Figure 8. Examples of Generated Images

Of note, the produced images often would have post-processing corrections applied to them done by Google, creating a smudging effect in the center of many images. This would still make any relevant distresses visible, but the clarity of distresses would be poor. This coupled with the relatively small, capped image size of 640x640 means that some distress can be very difficult to make out in some images. This is only mentioned to account for some possible errors that may arise in distress detection and may account for difficulties with regard to model accuracy.

Finally, the last dataset of importance to this project is that of the census block data. This data consists of two separate data frames, the first being the census block data (OpenDataKC) containing all relevant socioeconomic data and the second providing the geographic context for the socioeconomic data (MARC). The OpenDataKC dataset has a shape of 115 columns and 820 rows containing many forms of census data including those needed for this project's comparison. The 820 rows represent every census block contained in the 4 counties that make up Kansas City, Missouri which are Clay, Platte, Jackson, and Cass. Most columns contain numerical and percentile data related to racial populations, employment percentages, median income, median house costs, education status, etc. As for the MARC dataset, the data has a shape of 135 columns and 1672 rows covering population and geographic data in the Kansas City metropolitan area from 2010 and 2000. Similar to the OpenDataKC dataset, this includes population breakdowns by race but does not include other socioeconomic data.

Geographic data for census blocks can be obtained through a number of means, but the selection of the MARC data was selected for a few reasons. These reasons include a similar dataset shape of 820 columns covering all the same census blocks, block ids in similar format between both data frames, and that both datasets cover similar data including population and racial data. These datasets were joined with one another to take the geographic data from the second dataset and join it with the first. Relevant columns taken from the OpenDataKC dataset include block id, population total, median age, racial categories (Whites, African Americans, Native Americans, etc.), employment percentages, education (less than high school diploma, equal to high school diploma, and bachelor's degree or greater), household income, and home value. From the MARC dataset only the block id, block geometry was extracted for visualization purposes. Fig 9 shows a sample of the resulting data frame.

index	geometry	GEOID10	Population	MedianAge	Whites	AfricanAmericans	NativeAmericans	Asians	PacificIslanders	PersonsofOtherRaces	PersonsofMorethanOneRace
0	POLYGON ((-5810233.734 11808378.074, -5809059....	2.903706e+11	967.0	51.0	901.0	30.0	0.0	12.0	1.0	15.0	8.0
1	POLYGON ((-5814852.000 11803613.257, -5814903....	2.903706e+11	1597.0	37.5	1396.0	98.0	5.0	6.0	0.0	29.0	63.0
2	POLYGON ((-5814905.304 11800940.170, -5814565....	2.903706e+11	1624.0	41.7	1465.0	50.0	4.0	10.0	15.0	30.0	50.0
3	POLYGON ((-5816679.710 11797146.765, -5817425....	2.903706e+11	1264.0	38.1	1111.0	52.0	6.0	5.0	2.0	56.0	32.0
4	POLYGON ((-5819776.063 11796197.080, -5819741....	2.903706e+11	2090.0	33.4	1762.0	165.0	12.0	43.0	5.0	42.0	61.0

Figure 9. Snapshot of final census block data frame

Of note, due to this dataset covering all census blocks from 4 counties, a number of census blocks are disregarded for the final analysis. The reason they are disregarded is because they are not a part of Kansas City proper and instead a part of the smaller cities and towns that surround Kansas City or are simply rural blocks that are not a part of the city. As such, this reduces the number of census blocks represented in the final dataset once the condition scores are applied, reducing the total rows from 820 to 445.

4.1.1 Model Performance

This section will discuss the model performance of the YOLOv5 algorithm in both the training and testing phase of this project. The model training was conducted with batch sizes of 16 and 200 epochs. The model performed adequately, showing low training and value loss, indicating that the model is able to detect distresses reliably and returned a value of 0.65 in terms of precision. From here, the training data was used to perform the automatic annotation detection on the test image set. This was done in batches of 3000 images at a time with a confidence of 0.4, intersection over union (IOU) of 0.999 to allow for more precise overlap of bounding boxes, and test time augmentation to improve results.

In assessing the overall performance of the model, the Weights and Biases web platform was used to track the performance of the YOLOv5 model. Both precision and recall showed adequate performance with precision measured at a range of 0.6 to 0.8 and 0.6 to 0.7 respectively. The confusion matrix showed that the model performed well in detecting alligator cracking, block cracking, patching, and manholes while having some difficulties noticing longitudinal crack. Regardless of the difficulties with some detections, most distresses were reliably detected at an acceptable threshold for the purpose of analysis. Table 6 shows the number of detections made across the outlined detections.

Table 6. Number of Distress Detections

Distress	Number of Detections
Alligator Cracking	**2,689**
Block Cracking	**8,970**
Transverse Cracking	**3,274**
Longitudinal Cracking	**2,388**
Sealed Cracks	**3,288**
Patching	**2,056**
Manholes	**2,524**
Total Detections	**25,239**

Overall, the model's performance could be improved with more training data but for the purpose of this project, these metrics were acceptable for detections. Fig10 shows the performance metrics including precision, recall, and loss. Fig 11 and Fig 12 display street segments where the model detected distresses well and where the model failed to detect some distress.

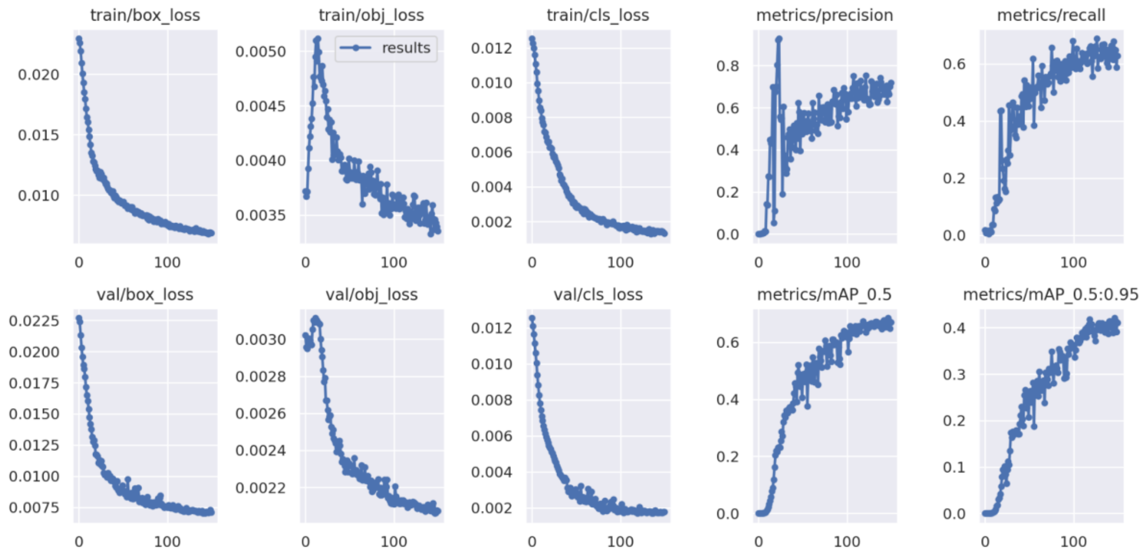

Figure 10. Training Metrics for YOLOv5 Model

Figure 11. Segments where distress detection performed well

Figure 12. Segments where the model failed to detect some distresses

4.1.2 Final Dataset Analysis

Once the pavement distresses are identified by the model, the batches of distress detections are merged to produce a single dataset which will be used for the condition data. As laid out in the methods section of this paper, each distress classification is assigned a value that is congruent with PASER classification. The dataset was then modified to represent each street segment with a single value indicating a PASER score for that segment. Those segment scores were then overlaid upon the census block data, treating segments as points that fall within the geographic polygons stored within the census blocks. These segments then had their scores averaged based upon the census block they fell within, producing the final condition score for the entire census block.

This final join and aggregation of data left many census blocks without any condition scores. There are two reasons as to why this is the case, the first being that those census blocks are outside of the street segment data set used for this paper. Many of those census blocks represent cities that make up the Kansas City metropolitan area or are rural blocks that are not considered a part of the city proper. As such, these blocks will not be considered for broader analysis in this paper. The second reason is that the model does not return information about segments in which no distresses were detected. This means that technically these are segments that should return a high rating but instead were simply not reflected for the dataset. To correct this, each road segment where no distresses were detected were assigned a score of 10.

To address the first objective of this project, it is important to see how wide the gap in road condition scores is in the final dataset. Taking the entire dataset from top to bottom shows that the highest score out of any census block is a condition score of 10 while the bottom condition score is rated 2. This alone only shows that a significant gap in the highest and lowest scores exist but taking a look at the broader dataset shows that a majority of the data falls between 4 and 8. This means that scores of 10 and 2 (the upper and lower threshold of the scoring system) could be interpreted as outliers. To account for this, all census blocks where 10 or less total detections are present were dropped from the final dataset to generate better analysis.

4.2.0 Results

Outlined in this section are the findings and results of the comparative analysis made between pavement condition and various socioeconomic factors. First the dataset was analyzed to determine whether or not a significant gap exists in the scoring of pavement sections. The following sections were divided along the four metrics of race, income, employment, and education selected to compare against the pavement conditions followed by a joint comparison considering the previously mentioned metrics. Each metric was graphed on a map for visual purposes and the top and bottom scores with respect to these metrics were also graphed to show a more direct comparison. All mapping projections in this section are displayed as choropleth maps where the color gradient represents the condition scores where the color purple indicates better condition scores and red represents worse condition scores. Following the generation of this choropleth map projections, the data was displayed graphically using scatter plots to display the data in a more visual form and statistical analysis was applied after. Following this section will be a discussion of the results where trends and correlations will be discussed.

Visualizing the top and bottom 25 census blocks based upon their condition scores shows some interesting details. For the top 25 condition scores, the values seem to be more spread out across all areas of the city while the bottom 25 seem more concentrated towards the downtown and the denser urban parts of the city. This makes sense considering that downtown features a much higher density of roads which are likely traveled on more regularly. These trends become far less visible when increasing the top and bottom from 25 census blocks to 100 census blocks with the top end showing more coverage across the city and the bottom representing scores of up to 7 which represents more average or decent quality road segments. Figure 13 shows the top 25 and bottom 25 census blocks highlighted and shown on a map visualized by Folium.

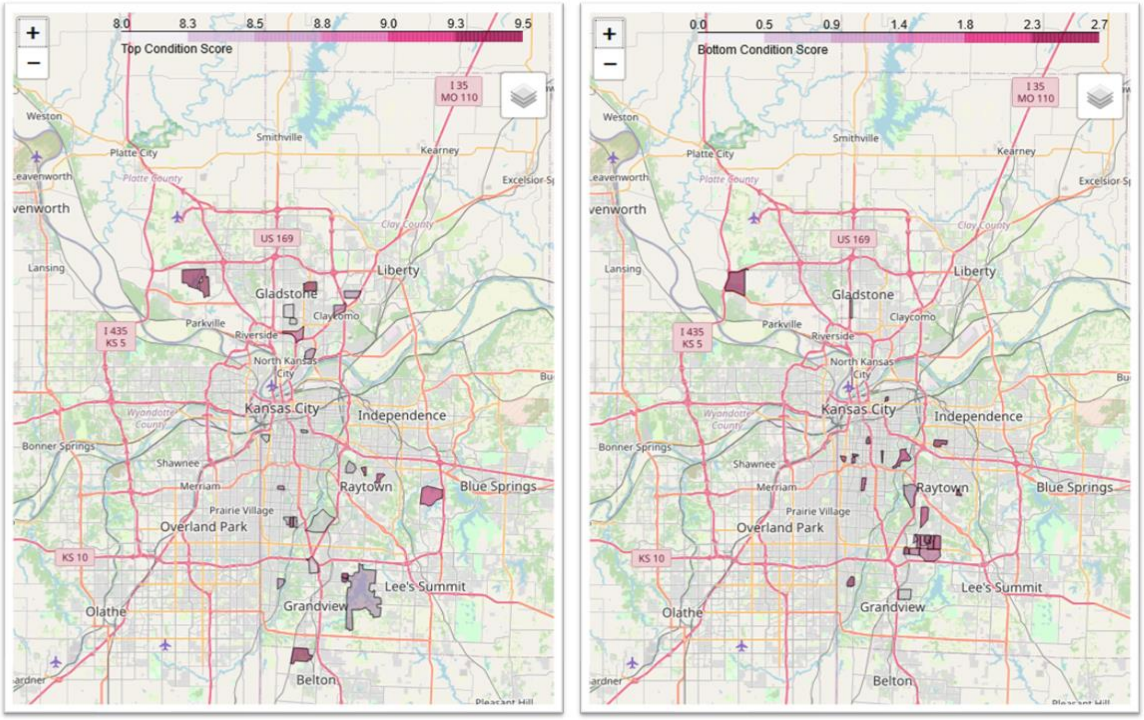

Figure 13. Top and Bottom 25 census blocks based on condition score

4.2.1 Socio Economic Considerations

With the previous analysis established, it is now important to look at this data with the socioeconomic data overlaid to see if any trends exist. Starting with race, Kansas City's racial breakdown features primarily white comprising nearly 70% of the total population. The census data does provide a metric for the percentage of minority persons which represents any non-white person within that census block. This column was used to isolate any census blocks that feature minority persons as the majority of persons with that block. The results show that blocks which feature more minority persons are almost all located in downtown Kansas City while majority white census blocks feature almost all other census blocks (Figure 14).

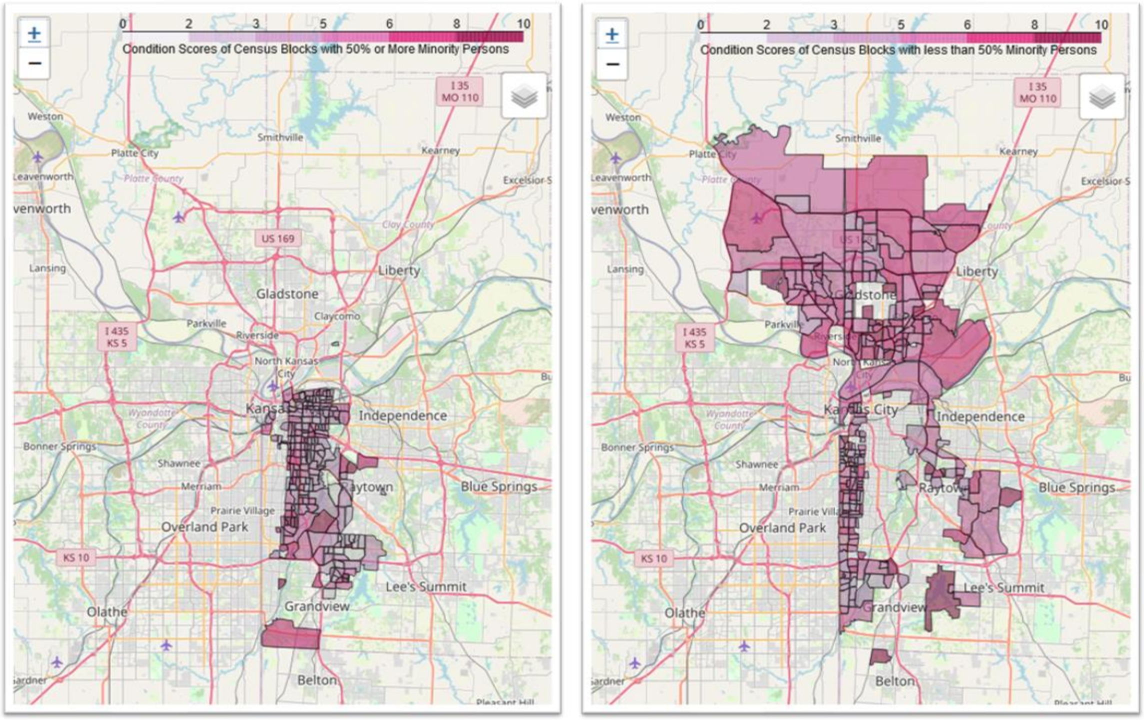

Figure 14. Census blocks that feature 50% or greater minority persons (left) and census blocks that feature majority white persons (right)

With regards to income, the data shows the median income across census blocks and poverty the percentage of persons below the poverty level. This provides two means of visualizing the data, with one showing the census blocks conditions scores with regard towards median income above and below the $49,916 median income across the city and the other visualizing the data with condition scores reflecting census blocks with a poverty rate higher that the 15.9% average. Interestingly, median household income below the average represented 304 out of the 445 census blocks covering the entire range of scores from 2 to 10 with a similar range present for the census blocks that feature a median household income above the average. Coincidentally, there is a fair amount of overlap between lower median household income and minority persons as the maps cover many of the same areas. As for poverty level, the resulting data showed very similar projections to that of median household income and similar condition score ranges. Even when selecting census blocks with a percentage of persons below the poverty level of 50% or greater, the overall condition scores reflect the full range of scores. Figure 15 shows the median household income projections and fig? shows the poverty level projections.

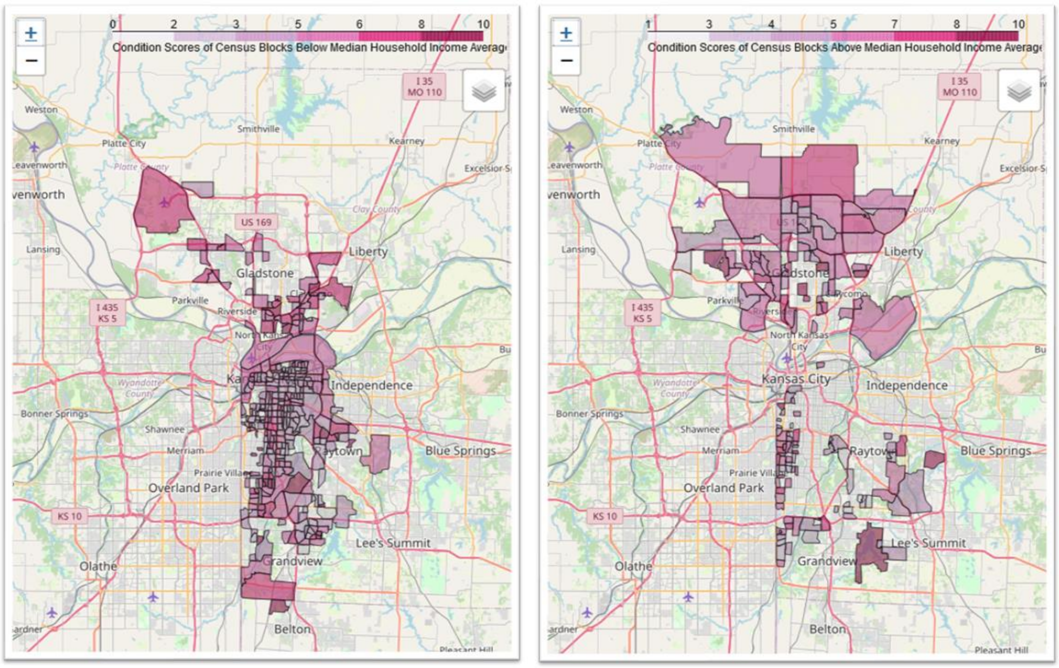

Figure 15. Census Blocks that feature a median household income less than the average (left) and greater than the average (right)

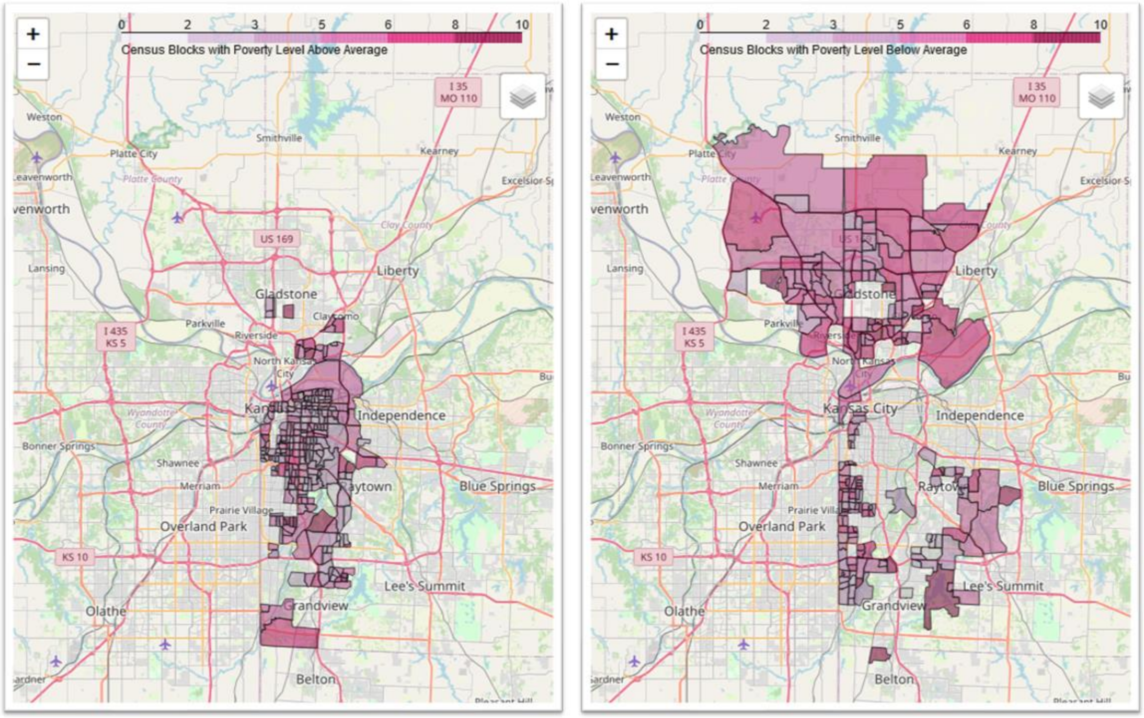

Figure 16. Census Blocks that feature a poverty level greater than the average (left) and less than the average (right)

Moving along to analyze unemployment, the dataset features data that includes the percentage of the labor force unemployed. The context of this statistic included every workforce eligible individual aged 16 and older who were at time of the statistical record unemployed. As such, census blocks that featured an unemployment percentage greater than the average, being 9.74%, were selected for analysis. Looking at how unemployment compares to condition scoring, it seems that high unemployment is mostly present around downtown Kansas City, with scores representing the full range scores. Figure 17 shows the projection created by comparing above average poverty rates to condition scores.

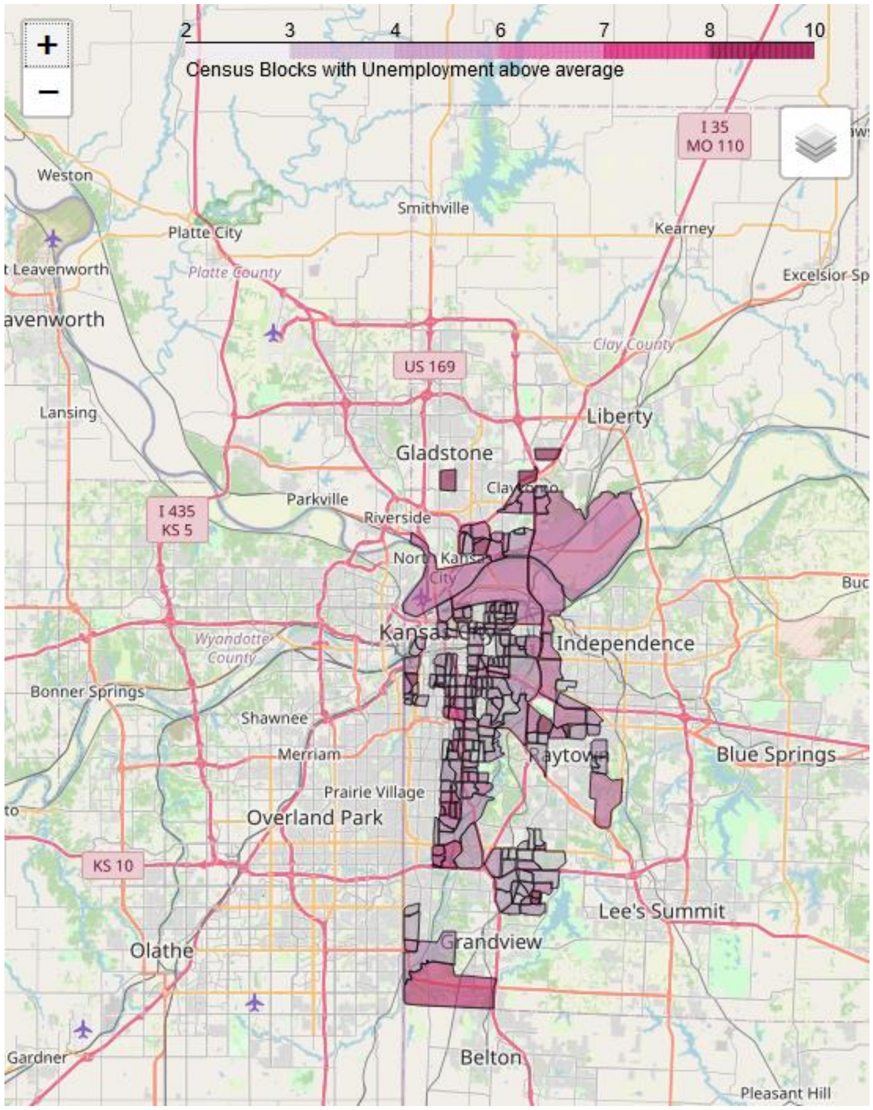

Figure 17. Census blocks that feature unemployment greater than the city average

Lastly, education status was analyzed across all census blocks in Kansas City. The dataset features three columns separating the educational classifications as less than a high school diploma, high school diploma with less than a bachelor’s degree, and bachelor’s degree or greater. For analysis, census blocks were selected that featured a greater percentage of persons that have less than a high school diploma and greater percentage of persons with a bachelor’s degree or greater. Results show that several census blocks in the northern part of downtown show highest percentages of persons with less than a high school diploma featuring a condition score range of 2.7 to 7.8. Conversely, census blocks consisted mostly of persons with a bachelor's degree or greater feature more broadly across Kansas City with a condition score range of 2 to 10. Figure 18 shows these projections.

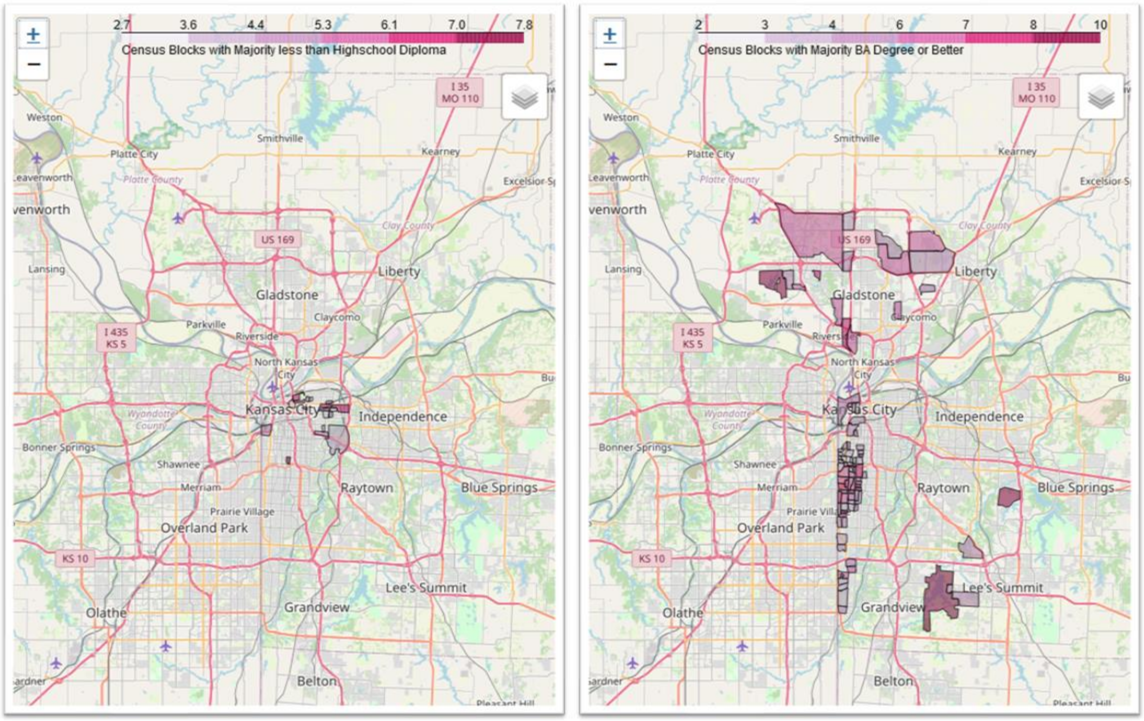

Figure 18. Census Blocks that feature majority persons with less than a high school diploma (left) and bachelor's degree or greater (right)

As for how these various factors intersect, it is important to emphasize the census blocks that express multiple aspects of being disadvantaged as to see if any trends exist. As such, for each of the previous categorizations a placeholder value was assigned to each and then totaled to represent a level of disadvantage. Essentially, the greater the placeholder value, the more disadvantaged the census block for the purpose of analysis. The resulting map shows trends typical of the previous maps, where census blocks with higher scores (more disadvantaged) typically appeared in the downtown area of Kansas City, East of Troost Ave as demonstrated by fig? Looking at how the disadvantaged scoring compares to condition scores; the more disadvantaged census blocks appear to be concentrated towards the downtown area South of the Missouri River and East of Troost Ave. This falls in line with the general trend seen in the previous maps and visually confirms that these census blocks are the most disadvantaged by these metrics.

4.2.2 Graphical Results

Viewing the data graphically a better picture of any trends can be seen. To facilitate a graphical depiction of the socioeconomic factors used for this paper, a combination of scatter plots showing trends via a line of best fit. The x access of each graph is one of the aforementioned socioeconomic factors while the y access represents the condition scores. Each graph also displays a line of best fit which serves as the indicator for the trends that exist in the dataset.

First, looking at the graphs of the 5 categories featuring persons of racial minorities, median income, poverty, unemployment, and less than high school education with respect towards pavement conditions a slight but clear trend is shown across all metrics (Figure 19). For all graphs, according to the line of best fit, the condition score difference between advantaged and disadvantaged blocks is between 0.25 and 0.5 points. This is a relatively low difference, but it does show that by all metrics there is a slight trend showing that disadvantaged blocks reflect a trend of worse pavement scores.

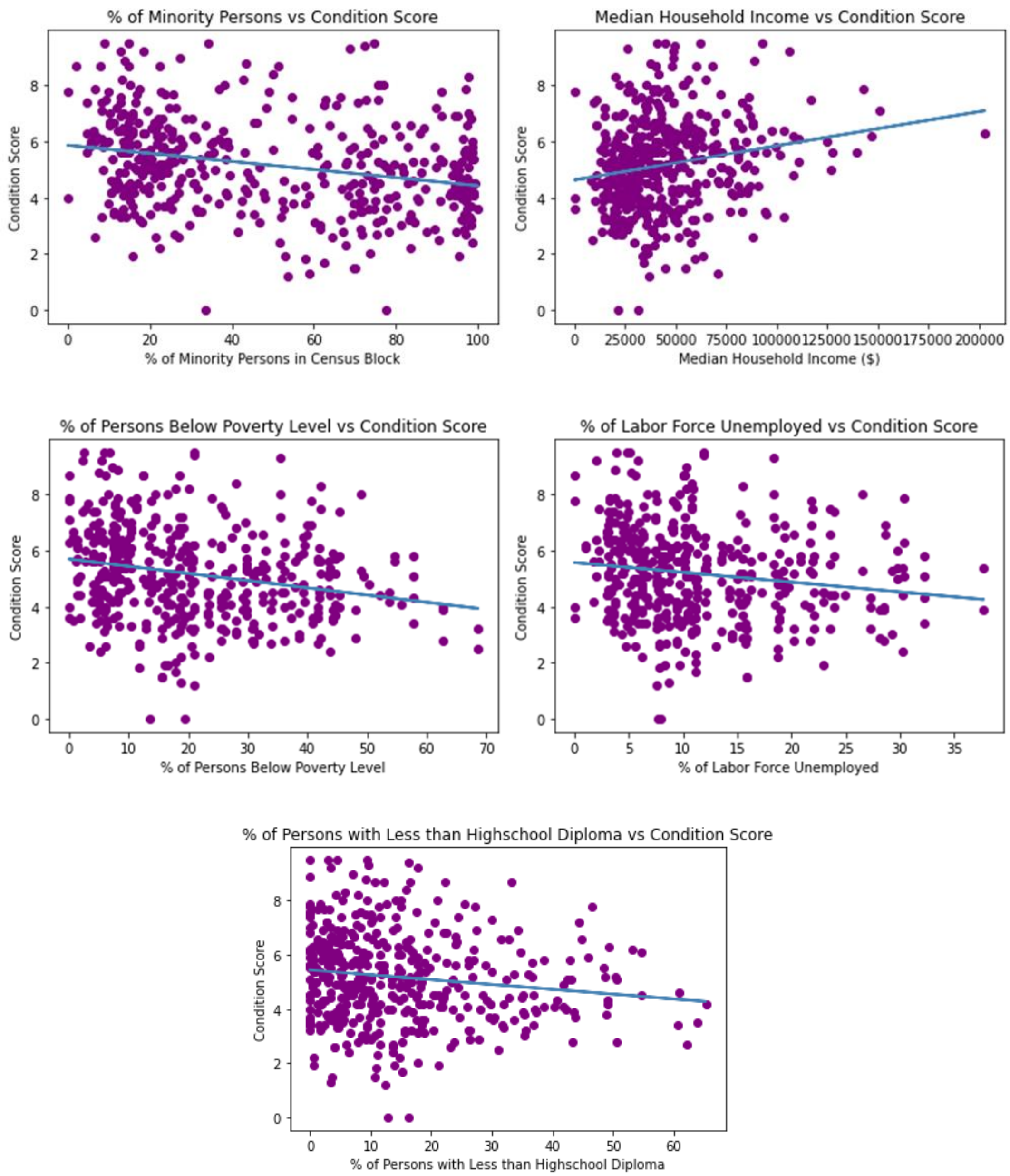

Figure 19. Scatter Plots of Socioeconomic Factors vs. Condition Scores

As for the last plot which measures all metrics of socioeconomic disadvantage in the form of a disadvantage score vs condition of pavement, the same trend as the previous graphs is present (Figure 20). The line of best fit shows only a difference between a disadvantage score of 0 (least disadvantaged) and 5 (most disadvantaged) as less than 0.5.

This does still point to a trend that favors advantaged census blocks, but the trend is very slight.

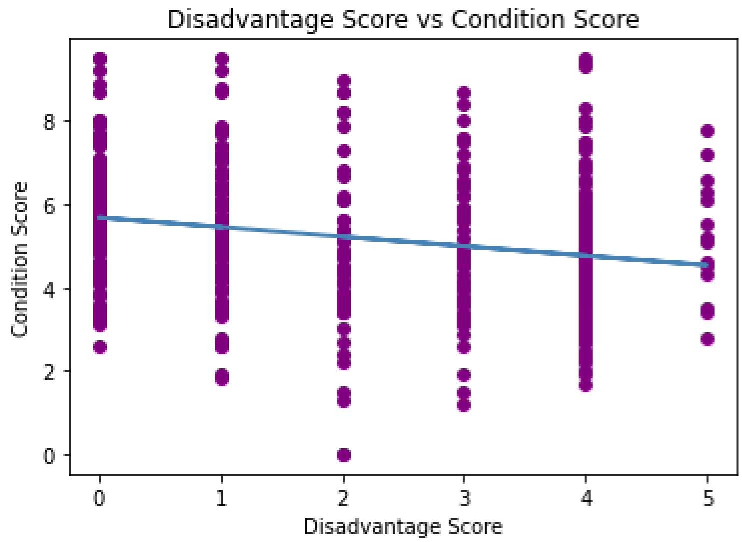

Figure 20. Disadvantaged Scores vs. Condition Scores

Using statistical analysis to aid in the interpretation of these results, the following table (table 7) was derived. First the R-Squared metric is measured across all plots to help determine the extent any variations are determined by outside factors. Following this the correlation coefficient r was calculated to determine the strength of the correlative factors as they relate to the condition scores. Lastly, a statistical test of significance was conducted across the plots to determine to what extent the hypothesis of disadvantaged census blocks demonstrating worse conditions is true. Calculations for the test of significance will utilize an α value of 0.05 and the null hypothesis will be the assumption that pavement condition scores across all communities is equal.

Table 7. Statistic Regression Analysis of Results

Plots vs. Condition Scores	R^2	R	t-statistic	p-value
% Minority Persons	0.074123	-0.272255	28.51674	4.69974…e-09
Median Income*	0.036659	0.191465	36.04567	4.52251…e-05
% Poverty Level	0.051665	-0.227299	21.53125	1.16590…e-06
% Unemployment	0.024626	-0.156927	17.09855	0.00085
% Less than Highschool Diploma	0.020772	-0.144126	14.70301	0.00223
Disadvantaged Score	0.051835	-0.227672	-26.75521	1.11852…e-06

Interpreting the results of the statistical analysis, a number of interesting trends can be seen. First, R^2or the coefficient of determination shows that by every factor reflects a value less than 10%, this means that the predictability of the data is relatively low. R or the correlation factor reflects correlation percentages between 14% and 27%, generally correlation percentages between 0% and 20% indicate very weak correlation or no association while 20% and 40% indicate weak correlation which means that every factor indicates either very weak or weak correlation with respect to the accepted hypothesis. To determine the accepted hypothesis, the calculation of the T-test and p-value was conducted. For all factors, the t-statistic's absolute value is high which suggests the results are reliable and the p-value is well below the significance factor (α) which indicates that the null hypothesis should be rejected.

4.3.0 Discussion of Results

From the tests conducted and the results as displayed it is clear to see that the data reflects trends in support of the notion that pavement distresses tend to be worse in disadvantaged communities. This said, the trends are not strong, showing a weak correlation measured statistically via the correlation factor measured in the previous step. In terms of which factor demonstrated the greatest correlation, Race appears to be the most significant factor to support this hypothesis while the percentage of persons with less than a high school diploma is the least significant. When comparing all factors and examining census blocks that display one or more of these disadvantaged factors, the trend reflects the individual metrics. What is important to emphasize is that every factor measured reflects trends indicating that disadvantaged socioeconomic factors indicate worse conditions, with no factor indicating otherwise.

Reflecting on why the trends are weaker than they could be a possible explanation could be the quantity of data used for this project. A collection of 22,384 segments over 3,245 streets may seem like a significant data set, but in truth may simply not be enough samples to properly provide a complete picture of the condition of these streets. Aspects of street condition can change dramatically depending on what part of the road is being inspected, as such a greater volume of pavement snapshots, especially along roads that contain very few data samples, could provide more conclusive results.

In addition, the performance of the model could in part explain some discrepancies. Although the YOLOv5 model performed well in detecting most distresses, it did struggle with some distresses such as block cracking. The most likely explanation would be that the quality of images is quite poor along several segments which may explain the model's difficulties when detecting distresses. For future trials conducted using the methodology derived from this project, increasing the number of images used for training would likely improve the performance of the model. In addition, training the model to detect a variety of additional distresses such as raveling, scaling, potholes, etc. could also improve the results by providing a more complete picture of the road's conditions.

With the above considerations, it is important to understand that these suggestions are conjecture about ways to improve the methodology and may not reflect dramatic changes in the results. Although these suggested improvements would undoubtedly improve the accuracy of the results, the results recorded in this paper are acceptable for the purpose of establishing a framework for this concept. The process of detecting distresses, rating the condition of roadways, and displaying the data graphically and visually has proven successful in supporting the hypothesis of gap in pavement conditions with respect to socioeconomic factors. Whether or not the trends are extremely significant, the demonstration of a trend of any kind is significant enough to warrant further investigation.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

This project was successful for the purpose of developing a framework to implement AI applications to detect pavement conditions and simultaneously provide a means of conducting sociological analysis. In this study, the use of various open-source tools, APIs, and AI algorithms proved useful in developing a cost-effective, customizable, and broadly applicable system to conduct this form of analysis. The geographic census data provided context for the determination of inequities that may exist within a roadway network while also providing a visual means of assessing the data used. The added socioeconomic context can provide another means for agencies to address aspects of a transportation network to better aid disadvantaged communities.

Due to the nature of this project, there is plenty of room to incorporate more data and tools to further improve the accuracy of this methodology. A greater volume of road segments along each roadway gained from the API could provide added context and more accurate reading of pavement conditions in any city. Provided that the YOLOv5 model is trained to detect additional distresses, a more complete rating could be established for the pavement sections analyzed for this project. Both aspects can be heavily modified and adjusted to agency preferences and can provide additional modes of analysis for developing Pavement Management Systems.

Speaking to the results of this project, it is important to stress that the results only glimpse at the reality that is reflected in the data. The results clearly show where disadvantaged communities are within the city and glimpse pavement conditions estimated from a limited dataset. This does not mean that the results are inaccurate or do not reflect some truths regarding the state of transportation infrastructure within Kansas City, but rather that the methods utilized for this project suggest a trend of inequity. At most the results should only be utilized to influence future investigations of infrastructural assessment in disadvantaged communities and the methodology considered for PMS development.

Addressing some key limitations of this research methodology, it is important to understand the need for expanding the scope of this project. Kansas City is only of many cities that demonstrate historical racial segregation boundaries and to strengthen the arguments made in this project more cities should be analyzed. In addition, other pavement condition systems should be considered as due to the simplicity of PASER, and how it is implemented into this project, only analyzes individual distresses and not necessarily their severity nor overall coverage. Although if PASER is still used, the formulations used in this project should be adjusted to consider severity of distresses in the condition score calculations to better reflect PASER documentation. Lastly, although the use of Google Street-View to provide an image set for this project is time/cost-effective and convenient it is at the expense of accuracy as one cannot be certain as to whether the images pulled for analysis are reflective of the roadway's current condition. As such, alternative means of collecting image sets for analysis should be considered to maximize the accuracy of the results.

Overall, the performance of the models and the implementation of various open-source technologies was satisfying. The ability to collect a large dataset of roadway data and develop a model to determine inadequacies within pavement sections is important for effectively managing roadway assets. This coupled with sociological analysis can help highlight where coverage is lacking with respect to disadvantaged communities and can hold agencies accountable to ensure that resources are being utilized equitably.

CHAPTER 6: REFERENCES

Shtayat, A., Moridpour, S., Best, B., & Rumi, S. (2022). An Overview of Pavement Degradation Prediction Models. *Journal of Advanced Transportation*, vol. 2022, Article ID 7783588, 15 pages, 2022. https://doi.org/10.1155/2022/7783588

United States Government Accountability Office (2022). *Analysis of Available Data Could Better Ensure Equitable Pavement Condition,* (GAO-22-104578). https://www.gao.gov/products/gao-22-104578

Vasudevan, M., Townsend, H., Schweikert, E., Wunderlich, K. E., Burnier, C., Hammit, B. E., ... & Ozbay, K. (2020). *Identifying real-world transportation applications using artificial intelligence (AI)-real-world AI scenarios in transportation for possible deployment* (No. FHWA-JPO-20-810). United States. Department of Transportation. Intelligent Transportation Systems Joint Program Office.

Euston, D. and Reidy, T. (2020). *Dissecting the Troost Divide and racial segregation in Kansas City*. Martin City Telegraph https://martincitytelegraph.com/2020/06/30/dissecting-the-troost-divide-and-racial-segregation-in-kansas-city/

Litman, T. (2017). *Evaluating transportation equity*. Victoria Transport Policy Institute. https://www.researchgate.net/profile/Todd-Litman-2/publication/284050013_Evaluating_transportation_equity/links/5c4f42bba6fdccd6b5d00a9d/Evaluating-transportation-equity.pdf

Wachs, M. (2010). Transportation policy, poverty, and sustainability: history and future. *Transportation research record*, *2163*(1), 5-12.

Lucas, K., Mattioli, G., Verlinghieri, E., & Guzman, A. (2016, December). Transport poverty and its adverse social consequences. In *Proceedings of the institution of civil engineers-transport* (Vol. 169, No. 6, pp. 353-365). Thomas Telford Ltd.

Manaugh, K., & El-Geneidy, A. (2012). Who benefits from new transportation infrastructure? Using accessibility measures to evaluate social equity in public transport provision. In *Accessibility analysis and transport planning* (pp. 211-227). Edward Elgar Publishing.

Lewis, D. (2011). Economic perspectives on transport and equality. International Transport Forum Discussion Paper.

France-Mensah, J., Kothari, C., O'Brien, W. J., & Jiao, J. (2019). Integrating social equity in highway maintenance and rehabilitation programming: A quantitative approach. *Sustainable Cities and Society*, *48*, 101526.

Naseri, H., Fani, A., & Golroo, A. (2022). Toward equity in large-scale network-level pavement maintenance and rehabilitation scheduling using water cycle and genetic algorithms. *International Journal of Pavement Engineering*, *23*(4), 1095-1107.

Abaza, K., Khatib, I., & Ashur, S. (2004). Integrated pavement management system with a Markovian prediction model.

Sundin, S., & Braban-Ledoux, C. (2001). Artificial intelligence–based decision support technologies in pavement management. *Computer-Aided Civil and Infrastructure Engineering*, *16*(2), 143-157.

Xu, Y., & Zhang, Z. (2022). Review of Applications of Artificial Intelligence Algorithms in Pavement Management. *Journal of Transportation Engineering, Part B: Pavements*, *148*(3), 03122001.

France-Mensah, J., & O'Brien, W. J. (2018). Budget allocation models for pavement maintenance and rehabilitation: Comparative case study. *Journal of Management in Engineering*, *34*(2), 05018002.

Boyles, S. D., Zhang, Z., & Waller, S. T. (2010). Optimal maintenance and repair policies under nonlinear preferences. *Journal of Infrastructure Systems*, *16*(1), 11.

Kothari, C., France-Mensah, J., & O'Brien, W. J. (2022). Developing a Sustainable Pavement Management Plan: Economics, Environment, and Social Equity. *Journal of Infrastructure Systems*, *28*(2), 04022009.

Attoh-Okine, N., & Adarkwa, O. (2013). Pavement condition surveys–overview of current practices. *Delaware Center for Transportation, University of Delaware: Newark, DE, USA*.

Cafiso, S., Di Graziano, A., & Battiato, S. (2006, October). Evaluation of pavement surface distress using digital image collection and analysis. In *Seventh International congress on advances in civil Engineering* (pp. 1-10).

Asada, T., Kachi, K., & Arimura, M. (2020, October). Exhaustive Pavement Inspection on Residential Roads Using Car-Mounted Cameras and U-Net. In *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)* (pp. 698-699). IEEE.

Gopalakrishnan, K., Khaitan, S. K., Choudhary, A., & Agrawal, A. (2017). Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction and building materials*, *157*, 322-330.

Majidifard, H., Adu-Gyamfi, Y., & Buttlar, W. G. (2020). Deep machine learning approach to develop a new asphalt pavement condition index. *Construction and building materials*, *247*, 118513.

Miller, J. S., & Bellinger, W. Y. (2003). Distress identification manual for the long-term pavement performance program (No. FHWA-RD-03-031). United States. Federal Highway Administration. Office of Infrastructure Research and Development.

Lebofsky, Steve. (2014). 2010 Census/ACS Basic Block Group Data. OpenDataKC. https://data.kcmo.org/Census/2010-Census-ACS-Basic-Block-Group-Data/xq6k-hd5i

Walker, D., Entine, L., and Kummer, S. (2013). Pavement Surface Evaluation and Rating - PASER Manual: Asphalt Roads. University of Wisconsin-Madison.

Solawetz, Jacob. (2020). YOLOv5 New Version - Improvements and Evaluation. Roboflow. https://blog.roboflow.com/yolov5-improvements-and-evaluation/#data-augmentation-in-yolov5