GIS Spatial Analysis

1. Site Suitability Analysis of Small Wind Farm using MCA

The purpose of this analysis is to identify two potential wind farms to be built in Western Ontario at the Municipal of Kincardine due to an abundance of suitable space. Wind speed is also suitable as it is greater than or equal to 6 m/s. The suitability of the land is reflected through the already present Wind Farms placed by private, and public companies as well as government-owned Wind Farms. The lands around the proposed Wind Farms are used for agriculture or open green spaces.

A picture containing text, screenshot, map, diagram

Description automatically generated

A picture containing text, map, atlas, font

Description automatically generated

1. Fire Response Coverage Analysis using Network Datasets in the Region of Waterloo

The purpose of this research is to identify and lessen the response time for Fire Calls that are greater than 10 minutes in the Region of Waterloo . To do this, a Network Dataset was first made using the Roads in the Waterloo Region region shapefile as the base for Analysis. The shapefiles for Fire Stations, Fire Calls and Population data are later inputed to the Network Analysis toolkit to produce three types of analysis: Routes, Service Area, and Location-Allocation.

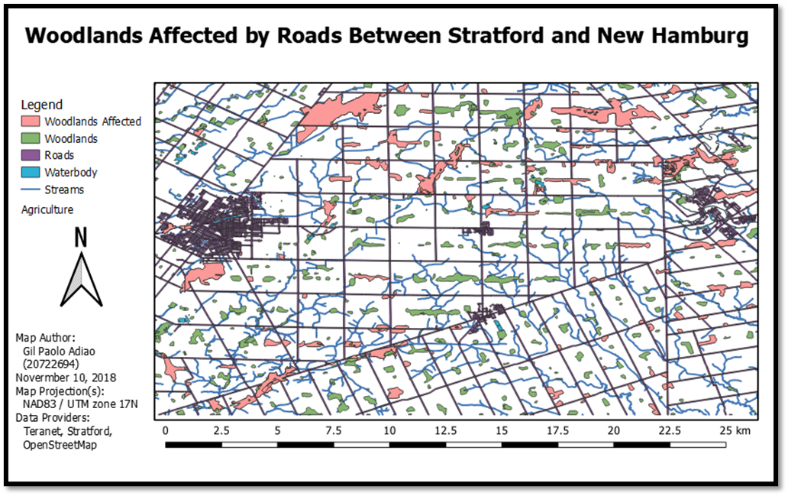
A picture containing text, map, diagram, atlas

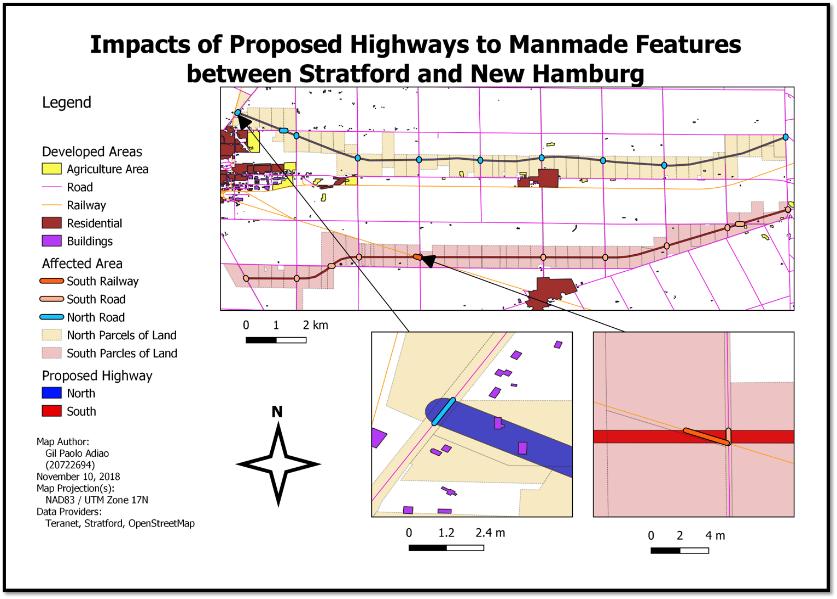
Description automatically generatedA map of fire stations

Description automatically generated

1. Impact Analysis of Proposed Highways in Ontario

The purpose of this analysis is to identify whether the proposed North or South Highway between Stratford and New Hamburg is suitable through a highway expansion study and environmental assessment. The North Highway is longer and intersects more woodlands. The South Highway does not impact nearby residential areas as well as having an obstructed highway route of less than 30% therefore is the more suitable route.

A picture containing text, diagram, screenshot, font

Description automatically generated

Remote Sensing

1. Airborne LiDAR Data Processing of EV3 Building of UWaterloo

Airborne LiDAR Data Processing portrayed to show the granular scale of detail in accuracy, resolution, and modelling of real-world objects. The 3D model is stitched together using Aerial and ground-based LiDAR data. The main drawback of a full 3D coverage of a study area that includes buildings is the inability for ground LiDAR devices to capture the vertical surfaces of buildings. The device cannot penetrate further than the area the device is designed to image. The Aerial LiDAR data was able to capture accurate vertical measurements but not granular enough to image the building from above.

*Figure 5: Picture of the EV3 buiilding using aerial lidar data in conjunction with one ground lidar data to show where the building is.*

A picture containing tree, building, plant, drawing

Description automatically generated

*Figure 6: All 7 .las ground datasets stitched together to reveal the outputted building layer. Note the missing roof layers.*

A picture containing building, tree, house

Description automatically generated

1. Man-made and Natural Feature Image Classifications

A comparison of ISOCLUS unsupervised pixel-based classification, MDM (Minimum Distance to Mean) pixel-based classification, and OBIA (Object-based Image Analysis) classification. ISOCLUS classification is observed to work on a massive scale to identify features but versatility in classification comes from user specified aggregation. MDM classification accuracy comes from the accuracy of the given training data. However, the most aesthetic can be seen at OBIA but can overgeneralize features.

|  |  |  |  |
| --- | --- | --- | --- |
|  | ISOCLUS Classification | Corresponding MDM  classification subset | Corresponding OBIA  classification subset |
| Water | A blue and green background  Description automatically generated | A blue and green water  Description automatically generated | A colorful map of a city  Description automatically generated |
| Forest | A pixelated image of a green and black background  Description automatically generated | A green and yellow background  Description automatically generated | A green screen with a white cross  Description automatically generated |
| Fields | A pixelated image of a green and black pattern  Description automatically generated | A green yellow and white pixelated pattern  Description automatically generated | A screenshot of a video game  Description automatically generated |
| Urban/built up | A colorful pattern with different colors  Description automatically generated | A colorful pattern with black and yellow and green  Description automatically generated | A screenshot of a video game  Description automatically generated |
| Legend | A screenshot of a list  Description automatically generated | A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

1. Amazon Rainforest Deforestation through RADAR Data Processing

The process of filtering is used to greatly enhance the extent of damage deforestation has done to the Amazon Rainforest. This is first done by taking SAR satellite data and then using VV (Vertical Vertical Polarization) and VH (Vertical Horizontal Polarization) images from the subset of 2017 and 2018 images. For the RGB images on the bottom, Red represents the division of the year 2017 and 2018 for the relevant band, Green represents the year 2017 for the band, and Blue represents the year 2018 for the band. By doing this, the green and blue pixels together fill in the non-relevant features that do not cover forest cover change while the red pixels will show which areas that have little to no forest cover as its values are flipped as seen on the table above where bright areas for the divided bands are flipped making it the ideal band to show tree cover. The polarization for VV is the most useful for identifying forest cover as it covers a higher count of red pixels where areas of no forests are seen.

|  |  |
| --- | --- |
| Corrected Unfiltered: VV\_2017 | Lee filtered: VV\_2017 |
| A close-up of a grey surface  Description automatically generated | A close-up of a grey surface  Description automatically generated |
| Corrected Unfiltered: VV\_2018 | Lee Filtered: VV\_2018 |
| A close-up of a grey surface  Description automatically generated | A close-up of a grey surface  Description automatically generated |
| Corrected Unfiltered: VH\_2017 | Lee Filtered: VH\_2017 |
| A close-up of a grey surface  Description automatically generated | A close-up of a grey surface  Description automatically generated |
| Corrected Unfiltered: VH\_2018 | Lee Filtered: VH\_2018 |
| A close-up of a stone  Description automatically generated | A close-up of a stone surface  Description automatically generated |
| Band Division of VV\_2017 and VV\_2018 | Band Division of VH\_2017 and VH\_2018 |
| A close-up of a black and white background  Description automatically generated | A close-up of a black and white background  Description automatically generated |
| RGB Composite:  R: VV\_2017/VV\_2018  G: VV\_2017  B: VV\_2018 | RGB Composite:  R: VH\_2017/VH\_2018  G: VH\_2017  B: VH\_2018 |
| A blue and orange background  Description automatically generated | A blue and orange rock  Description automatically generated |

Data Analysis

1. Predicting Canadian Crop Yield using Neural Networks

Do deep neural networks such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) return more precise results and operate more efficiently in comparison to a traditional statistical model for crop yield forecasting? LSTM (Long Short Term Memory) is a part of neural network that is adept with time-series data. A Convolution model makes predictions based on a fixed-width history and is able to see how things are changing over time by using convolutions and filters to output a specified value. Mean Absolute Percentage Error is a statistic used to measure the quality of a model’s predictions and measures the average difference between the actual and the predicted values. Both the LSTM and CNN has performed quite well compared to the CCYF, though more studies will be required because of the obstacles encountered in implementations. Some evidence of overfitting has been found in all the models, especially the canola crop yield prediction. The machine learning models have potential to replace the traditional model in crop yield prediction.

Research Question:

Do deep neural networks such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) return more precise results and operate more efficiently in comparison to a traditional statistical model for crop yield forecasting?

Research Goal:

Investigate the potential for a deep neural network machine learning algorithm (LSTM or CNN) to be an effective replacement or alternative for the Canadian Crop Yield Forecaster’s currently employed statistical model.

Mean Absolute Percentage Error (MAPE)

* Mean Absolute Percentage Error is a statistic used to measure the quality of a model’s predictions
* Measures the average difference between the actual and the predicted values

LSTM

* LSTM (Long Short Term Memory) is a part of neural network that is adept with time-series data. (Schwalbert et al., 2020)
* For every time step, input variables are fed into LSTM cell, and that cell produces a prediction. That prediction is fed into another LSTM cell on the next time step; the next LSTM cell will use that input in addition to the original input variables to make another prediction. This process is repeated until the end of time interval. (Jiang et al., 2019)

A map of the united states

Description automatically generated with low confidenceA map of the united states

Description automatically generated with low confidence

CNN (Convolution

* A Convolution model makes predictions based on a fixed-width history and is able to see how things are changing over time by using convolutions and filters to output a specified value.
* Uses temporal convolution which is a 1-D Convolution specialized for time-series data repeated over each variable on a given time-step for the whole dataset over a single epoch. (Wan et al, 2019)

A map of the united states

Description automatically generated with medium confidenceA map of the united states

Description automatically generated with low confidence

Limitations

1. Source of data (AAFC) was not clean:
   * Missing historical data points for some CARDs for specific crop types
   * Null values/Obsure values (negative numbers, 999999, string)
   * Incorrect data type for specific entries
2. Variability in size of CARUID, something determined by AAFC
3. Some models required more rows of data than available
   * Long format vs Wide Format
4. Overfitting

Conclusion

* Both the LSTM and CNN has performed quite well compared to the CCYF, though more studies will be required because of the obstacles encountered in implementations.
* Some evidence of overfitting has been found in all the models, especially the canola crop yield prediction.
* The machine learning has potential to replace the traditional model in crop yield prediction.

References:

Jiang, H., Hu, H., Zhong, R., Xu, J., Xu, J., Huang, J., Wang, S., Ying, Y., & Lin, T. (2019). A deep learning approach to conflating heterogeneous  
 geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. *Global Change Biology*, *26*(3), 1754–1766.  
 https://doi.org/10.1111/gcb.14885

Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P. V. V., & Ciampitti, I. A. (2020). Satellite-based soybean yield forecast: Integrating  
 machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology*, *284*,  
 107886. https://doi.org/10.1016/j.agrformet.2019.107886

1. Hot Spot Analysis of Home Depots in Ontario

The Hot Spot Analysis done completely differs from a standard hotspot analysis featuring circular blobs of colors as Figure 2 and 3 looks more like a choropleth map with a gradient using z-score and p-value. In terms of aesthetics, it differs from a traditional hotspot analysis because it the hotspot calculations were grouped together using a fishnet from a given shapefile and in this case it is the OSD files for the province of Ontario. To have better look, OLS and GWR calculations were used. Multivariate OLS and GWR have different representations in showing explanatory variables to help predict where future stores will be located. With both a high R-squared obtained from the OLS and GWR, we can prove that Home Depot stores are strategically placed in high populated with well-endowed individuals who can afford the tools and equipment in the store. More than likely, more stores will be built alongside the development of the suburbs of expanding cities like Toronto and Ottawa.

A picture containing text, map, atlas, diagram

Description automatically generatedA picture containing text, map, diagram, atlas

Description automatically generated

*OLS Summary Statistics*

A picture containing text, screenshot, font, receipt

Description automatically generated

*Histogram of Standardized Residuals*

A picture containing text, diagram, plot, line

Description automatically generated

**Test OLS Residuals for spatial Structure**

*Figure 9 – Spatial Autocorelation Report*

A screenshot of a graph

Description automatically generated with low confidence

**Geographically Weighted Regression (GWR)**

*Figure 10 –* GWR Results

A picture containing text, screenshot, number, font

Description automatically generated

*Figure 11 – GWR using TC0005*

A map of a city

Description automatically generated with low confidence

*Figure 11 – GWR using TC1272*

A map of a city

Description automatically generated with low confidence

1. PCA of Lake ice cover break-up in Canada

Figure 3: Plot of original data and projected data using only PC1 to PC6 (the red line shows original data)

Chart, line chart

Description automatically generated

Using only six principal components, the projected data not only kept up with the shape of the original data, but came relatively close to predicting the accurate amount of ice break up with relatively good examples between the 8-10 year period and the especially the years 15 to 20.

Scatterplot Matrix of Latitude, Longitude, PC1, PC2, and PC3

A picture containing table

Description automatically generated

Looking at Figure 2, PC1 and latitude is seen to have a slight positive correlation with significant uniformity in the distribution of points. PC2 has a very slight negative correlation with longitude as seen from its significant uniformity while PC3 has a slight positive correlation with somewhat significant uniformity with longitude.

Screeplot of eigenvalues and principal components

Chart, histogram

Description automatically generated

As seen from the Figure 1 screeplot and supported from Table 1, the PC1 to PC6 as these 6

principal components covers about 65% of the variance in all 30 principal components.

Using only six principal components, the projected data not only kept up with the shape of the original data, but came relatively close to predicting the accurate amount of ice break up with relatively good examples between the 8-10 year period and the especially the years 15 to 20. Looking at Figure 2, PC1 and latitude is seen to have a slight positive correlation with significant uniformity in the distribution of points. PC2 has a very slight negative correlation with longitude as seen from its significant uniformity while PC3 has a slight positive correlation with somewhat significant uniformity with longitude.