The Impact of Tectonic Setting on Machine Learning Approaches for Earthquake Prediction

Eldon Taskinen[1] and Zelalem Demissie[1]

[1]Wichita State University

December 27, 2023

Abstract

In prior decades the concept of using mathematical methods to predict earthquakes was considered infeasible. Recent advances in machine learning and predictive modeling offer promising avenues to potentially realize earthquake prediction. In order to test the viability of machine learning methods, experiments were made with earthquake datasets from Kansas and Puerto Rico. The two datasets were chosen for the distinct differences in their tectonic settings. Kansas has few major faults, with a largely inactive subsurface, this produced a smaller dataset with a few large clusters. Puerto Rico is complexly faulted, with an extremely active tectonic setting, this produced a larger dataset with a large number of small clusters. In order to test the effectiveness of these two datasets for machine learning and prediction they were run through three different machine learning algorithms including an LSTM model, Bi-LSTM model, Bi-LSTM model with attention. Not only were the three different machine learning methods compared against each other for accuracy but also the datasets as well. Conclusive findings show that the two different data sets favor different processing methods. The Kansas data performs the best with the Bi-LSTM with attention model, while the Puerto Rico data performs the best with the LSTM model. This is likely due to the tectonic settings of the two regions, since the Kansas dataset has less overall data, and earthquakes are concentrated in a few large clusters, while the Puerto Rico data set has a more even distribution.

The Impact of Tectonic Setting on Machine Learning Approaches for Earthquake Prediction

Enter authors here: E. Taskinen[1], Z. Demissie[2]

[1] Wichita State University.

[2] Wichita State University.

Corresponding author: Eldon Taskinen (eldontaskinen@gmail.com)

†Additional author notes should be indicated with symbols (current addresses, for example).

Key Points:

- Earthquake data distribution is driven by underlying tectonic structures.
- Data distribution can affect how Machine Learning algorithms process data.
- Different Machine Learning algorithms work best for different distributions and tectonic structures.

Abstract

In prior decades the concept of using mathematical methods to predict earthquakes was considered infeasible. Recent advances in machine learning and predictive modeling offer promising avenues to potentially realize earthquake prediction. In order to test the viability of machine learning methods, experiments were made with earthquake datasets from Kansas and Puerto Rico. The two datasets were chosen for the distinct differences in their tectonic settings. Kansas has few major faults, with a largely inactive subsurface, this produced a smaller dataset with a few large clusters. Puerto Rico is complexly faulted, with an extremely active tectonic setting, this produced a larger dataset with a large number of small clusters. In order to test the effectiveness of these two datasets for machine learning and prediction they were run through three different machine learning algorithms including an LSTM model, Bi-LSTM model, Bi-LSTM model with attention. Not only were the three different machine learning methods compared against each other for accuracy but also the datasets as well. Conclusive findings show that the two different data sets favor different processing methods. The Kansas data performs the best with the Bi-LSTM with attention model, while the Puerto Rico data performs the best with the LSTM model. This is likely due to the tectonic settings of the two regions, since the Kansas dataset has less overall data, and earthquakes are concentrated in a few large clusters, while the Puerto Rico data set has a more even distribution.

Plain Language Summary

Advances in machine learning may make earthquake prediction possible. In order to test the viability of machine learning methods, experiments were made with earthquake datasets from Kansas and Puerto Rico. The two datasets were chosen for the distinct differences in their tectonic settings. Three different machine learning algorithms were tested on the two data sets and then compared for accuracy of predictions. Conclusive findings show that the two different data sets favor different processing methods. The Kansas data performs the best with complex models, while the Puerto Rico data performs the best with simplistic models. This is due to the tectonic settings of the two regions, indicating that future exploration into earthquake prediction will need to consider tectonic settings into how data is processed.

1 Introduction

1.1 Problem Statement

Earthquakes are tremendous natural disasters that can cost hundreds or even thousands of lives, often leaving utter destruction in their wake. Current predictions estimate that there will be 7.5∓3.0 catastrophic earthquakes in the 21st Century, and that each of these earthquakes will cause an average of 1.45∓0.58 million casualties (Holzer and Savage, 2013). It is the goal of this research, and all earthquake prediction efforts to prevent as many future casualties as possible, by accurately and precisely predicting earthquakes, so that effective preventative measures can be taken. In the only widely successful instance of earthquake prediction, Chinese authorities were able to warn the town of Haicheng shortly before a magnitude 7.3 earthquake (Wood and Gutenberg, 1935) . Due to the short-term preparations that the people of Haicheng were able to make, the reported death toll was kept under 300 in an area where tens of thousands of fatalities would have been expected (Wood and Gutenberg, 1935). This research aims to implement methods of earthquake prediction that have already shown success in the Horn of Africa. The Horn of Africa experiment was able to predict earthquakes down to a 28.87% error using machine learning algorithms (Abebe et al., 2023). This is a startlingly low margin of error, for the field of earthquake prediction. The goal of this research is to verify these results as well as to see how modifications to the internal data set and hyperparameters change the outcome. If successful, the machine learning techniques used in this experiment can be considered for wider implementation. If the machine learning methods such as the LSTM algorithm used in the Horn of Africa experiment prove effective at predicting earthquakes elsewhere on the globe these advancements could lead to millions of lives being saved.

1.2 Background

Since the 1970's seismologists have been trying to use mathematical methods to predict earthquakes (Bakun and McEvilly, 1984). Early attempts at earthquake prediction relied upon finding patterns of earthquakes and then extrapolating the results over time (Bakun and McEvilly, 1984). A notable early example of this was the Parkfield Experiment in the 1980's, where a 20- to 30- km long section of the San Andreas fault near Parkfield, California had suffered notable earthquakes in 1857, 1881, 1901, 1922, 1934, and 1966 (Bakun and McEvilly, 1984). Considering these repeated earthquakes to be characteristic of a pattern the authors confidently hypothesized that the next notable earthquake in the Parkfield area would happen between 1983 and 1993 (Bakun and McEvilly, 1984). The Parkfield earthquake did not happen until 2004. This was considered a big disappointment for the field of earthquake prediction, and seismologists at the time began to question the viability of statistical approaches as a whole (Jackson and Kagan, 2006). The attitude of most seismologists after the Parkfield Experiment was best summed up in the article "Earthquakes Cannot Be Predicted" where the authors blatantly proposed that earthquakes are inherently unpredictable, citing chaos theory, as well as the unmeasurably fine conditions that surround an earthquake event as reasons why statistical prediction is impossible (Geller et al., 1997). This sentiment has slowed progress inside of the field of seismology, with many funding agencies cautious to support research into predictions that were subject to criticism and showed high degrees of error (Wyss, 2001). However, even then there still remained a few willing to continue working on mathematical methods of

earthquake prediction (Wyss, 1997). That same year an article was published in response titled, "Cannot Earthquakes Be Predicted?" arguing that earthquake prediction is still an emerging field, where early progress was unfolding slowly (Wyss, 1997).

In the 90's and through most of the ensuing time between then and now this is where the discussion of earthquake prediction has remained. Caught in opposition between a majority that believes it to be impossible, and a minority who were still trying to make it happen. The next notion of a change in fortune in the field came in 2017, when a machine learning algorithm successfully predicted earthquakes in a laboratory setting. This was one of the earliest applications of machine learning techniques to the field of earthquake detection (Rouet-Leduc et al., 2017). In the past few years, the development of easy to access software libraries like Theano and TensorFlow has allowed many seismologists to explore the viability of machine learning techniques (Rouet-Leduc et al., 2017). Machine learning allows for many leaps forward in data processing that were previously impossible, especially in terms of the number of data points and variables that can be processed. This all culminated in the Horn of Africa experiment where four of the most recently developed machine learning algorithms are tested against each other for a comparison of accuracy in the Horn of Africa (Abebe et al., 2023).

1.3 Research Objective

In order to conduct this experiment two data sets have been developed, one containing all magnitude 2+ earthquakes whose epicenters were located in the state of Kansas during the years 2018 and 2019, as well as a secondary dataset containing all magnitude 2+ earthquakes whose epicenters were located in Puerto Rico during the years 2018 and 2019. The two regions were chosen as datasets primarily for their structural differences. Of the data sets used in this research Puerto Rico is the more active than the other data set, Kansas. Puerto Rico's tectonic system is located at the eastern edge of the Greater Antilles, and part of the Caribbean tectonic complex (Jansma and Mattioli, 2005). The complexity of the region is due to Puerto Rico, Gonave in the west, and Hispaniola in the center, all occupying their own microplates, which all interact tectonically with one another (Jansma and Mattioli, 2005). Puerto Rico also has large fault zones located on the North and South ends of the island which have characteristic oblique normal faults with a possible right-lateral slip (Jansma and Mattioli, 2005). This makes Puerto Rico a hot spot for earthquake activity. The other dataset used in this experiment is based out of the state of Kansas. Unlike the other location previously discussed, Kansas is a midcontinental region, with only a few subsurface rifts and ridges that date back to the Pennsylvanian or are even older (Berendsen, 1997). This means that with a few regional exceptions that Kansas is rather inactive on the tectonic scale, to complicate things further, recent evidence points towards man-made, or induced earthquakes becoming a risk in Kansas, due to anthropogenic processes (Peterie et al., 2018). By comparing the results of machine-learning predictions in Kansas, and Puerto Rico, it should become apparent how variables such as geography, tectonic setting, and local seismicity play a role in the accuracy of machine-learning predictions.

2 Methods

2.1 Tectonic Setting

In order to achieve the stated goal of this research an understanding must first be established regarding what is known about the tectonic setting of each of the data set locations. In general, the term tectonic setting refers to the principal controlling factors behind a region's

lithology, chemistry, and structure (Veizer and Mackenzie, 2014). This includes dominant faults, tectonic plates, and plate boundaries. The tectonic setting of a region has long played a role in determining its seismicity, that is where and how often earthquakes occur (Bird and Kagan, 2004). That is to say if a region's tectonic setting is dominated by a large subduction zone is more likely to experience earthquakes than a region whose tectonic setting is determined by a continental rift boundary, which in turn is more likely to experience earthquakes than a region that does not have any plate boundaries (Bird and Kagan, 2004). As part of the technical background for this research, two different tectonic settings are described in this section. Firstly, this section will cover the tectonic setting of Puerto Rico. Secondly, this section will cover the tectonic setting of Kansas. The two data sets for this research. Puerto Rico and Kansas were chosen as the data sets for this research due to the prominent differences in their tectonic setting, in order to try and detect differences in machine learning processes later down the line.

2.1.1 Puerto Rico

The island of Puerto Rico is part of the Caribbean archipelago and has an area of roughly 9,104 km2 (Benford et al., 2012). It is also part of the Caribbean plate, which has major boundaries with both North American, South American, Nazca, and Cocos plates.

In the global tectonics system Puerto Rico lies on the Puerto Rico-Virgin Islands microplate, abbreviated PR-VI (Benford et al., 2012). The microplate is part of the larger Caribbean plate, and is a complex geodynamic setting, characterized by oblique collision to the west, oblique subduction at the longitude of Puerto Rico, and frontal subduction to the east (Viltres et al., 2021). One of the most interesting factors of this seismic environment is the trenches located both to the North and South of the island which is caused by double underthrusting by the Caribbean plate. These two trenches along with the three east-west strike-slip zones account for nearly 85% of the relative motion between the Caribbean and North American Plates (Mann et al., 2005). The trenches combine with the Anegada Deep and Mona Deep faults to design the boundary of the PR-VI microplate.

While the microplate is generally considered a single solid block there are two major northwest-southeast faults named the Great Northern and Great Southern faults respectively that can be found further inland. These inland faults meet up with faults in the ocean to create a larger network of faults that has been recorded to be very active during the Quaternary (Jansma and Mattioli, 2005). All of these active faults means that Puerto Rico has a relatively high natural seismicity, making it a good candidate for comparison with the other data set used in this research.

2.1.2 Kansas

The state of Kansas is located in the center of the continental United States and has an area of roughly 213,278 km^2 (U.S. Census Bureau, 2016). It is also part of the North American plate, which has major boundaries with the Pacific, Eurasian, African, South America, and Caribbean plates, and many smaller boundaries with minor plates.

Due to being located in the center of the world's largest continental plate, the tectonic setting of Kansas is remarkably stable. Located on the large, stable craton that composes most of North America, Kansas has had little interaction with any tectonic boundaries since the Precambrian (Merriam, 1963). As such, the state experiences relatively few earthquakes, ranking 45th among 50 of the nation's states in the amount of damage caused by earthquakes on the average year (Buchanan, 2000). The one exception to this is the Humboldt fault zone, part of the Nemaha uplift in the eastern portion of the state (Buchanan, 2000). This fault zone is a series of normal faults that are associated with movement in the North American craton that occurred in the protozoic (Baars et al., 1989).

During the past decade Kansas has experienced a rise in earthquake events. Prominent theories point towards the phenomena known as induced seismicity to be the cause (Peterie et al., 2018). Induced seismicity is a rise in earthquake events that is caused by human activity, usually in association with the production of oil and gas, as well as the disposal of its byproducts. Typically induced seismicity is caused by a single well, which injects the brine byproduct that is produced during oil and gas extraction back into the subsurface, usually in very high quantities. This increases the pore pressure of reservoir rock surrounding the well, which is sometimes large enough to trigger earthquakes a short distance away (Peterie et al., 2018). Recent work done in Kansas and Oklahoma however, suggests that fluid injected into the subsurface is migrating further than expected, and causing earthquakes throughout the state (Peterie et al., 2018). Due to the unique set of circumstances surrounding the earthquake record in Kansas, it will serve as a suitable juxtaposition to the more naturally active tectonic setting of Puerto Rico when the two are compared in the experiment.

2.2 Machine Learning Techniques

Machine learning methods are among the newest and most promising methods for earthquake prediction (Rouet-Leduc et al., 2017). Machine learning uses various different algorithms to develop a prediction model from data that is input into the algorithm. The machine trains itself on the input data, using methods similar to the human brain to sort out noise and identify patterns (Alaskar and Saba, 2021). Then, using what it has learned the machine is able to complete complex tasks, such as predicting future events in a time series, or identify and reproduce text and images. It is specifically the use of machine learning for time series predictions that interest seismologists, since each earthquake happens at a distinct time, the dataset can be represented as a time series, allowing for machine learning analysis (Alaskar and Saba, 2021). The three algorithms explored in this research are all variations on the Long Short-Term Memory model. Long Short-Term Memory (LSTM) models are an advanced form of neural networking, specifically designed for the purpose of time series predictions and have shown successful implementation in the prediction of everything from the weather to the stock market (Karevan and Suykens, 2020; Yadav et al., 2020). Along with the base LSTM model, this

experiment will also be testing the Bidirectional LSTM, and the Bidirectional LSTM with attention layer models, variations on the normal LSTM model, with added layers of complexity in order to optimize performance in some data sets. These three models will be tested against each other to see if the two locations and the two datasets change the relationship between the machine learning methods.

2.2.1 Long Short-Term Memory (LSTM)

The LSTM model is a recurrent neural network that has already seen some success in the field of earthquake prediction (Berhich et al., 2020; Bellamkonda et al., 2021). Recurrent networks like the LSTM function by 'remembering' and 'forgetting' information in a way meant to mimic the human brain (Bengio et al., 1994). The model will store information about past inputs, but only for a set amount of time, that time will depend on the weight given to the data by the algorithm while it is learning. All recurrent networks share 3 core functions, to store information for an arbitrary duration, to be resistant to noise (i.e. fluctuations of the inputs that are random or irrelevant to predicting a correct outcome), and to be trainable (Bengio et al., 1994). This all allows the model to be context sensitive, and show a reasonable degree of flexibility. The reason why the LSTM model has become the preferred recurrent network is because of how it solves an issue that is prevalent among those networks. The range of contextual information that a recurrent network can access is actually quite limited, the problem being that when given influence over a network's learning mechanism most information causes it to either decay or expand exponentially (Graves et al., 2009). This problem, known as vanishing and exploding gradients, is specifically what the LSTM model was introduced to solve (Graves and Schmidhuber, 2005). The LSTM functions by using a series of layers, blocks, and cells. The model contains one input layer, one hidden layer, and one output layer, and it is the hidden layer which contains the memory cells that determines what information the model keeps and what information it discards (Hochreiter and Schmidhuber, 1997). These memory cells are then arranged into blocks where they interact with other connected memory cells, as well as multiplicative units, which are provided by the read, write, and forget gates governed by the algorithm for the entire block (Greff et al., 2017).

2.2.2 Bidirectional LSTM (BiLSTM)

A common alteration that is made to the LSTM model is to run it as a bidirectional model. The basic idea of the bidirectional LSTM is that it can run its training sequence both forwards and backwards, then each learning method is connected to the same output layer (Graves and Schmidhuber, 2005). This allows the algorithm to identify patterns that occur in either direction, not only does this generally increase algorithm performance, but it is also essential to some pattern recognitions, like voice recognition.

2.2.3 BiLSTM with an Attention Layer

Another common method for improving the LSTM model is to implement an attention layer. The attention layer is capable of storing inputs and outputs from the hidden layer in order to help direct weights given to stored data (Bahdanau et al., 2014). This helps to increase accuracy even further, especially in larger data sets. One of the issues with the LSTM and all recurrent networks is that it still must process all of the information sequentially. This is not a limitation shared by the attention layer, which processes all input information before determining weights. By using the weights determined by the attention layer the LSTM has less error when

determining the weight of information while processing. In smaller data sets this improvement can be negligible, however the importance of attention for large data sets can not be understated (Bahdanau et al., 2014).

2.3 Model Hyperparameters

All the machine learning algorithms share nearly identical hyperparameters. This includes the LSTM model, BiLSTM model, and BiLSTM model with an attention layer. Each of the models was compiled using the mean squared error as the loss, and the adam optimizer. The model parameters are set to 30 epochs and a batch size of 32. The only exception to these parameters is the Bi-LSTM with an Attention layer model, which had its number of epochs reduced to 25 in order to save time while processing. All hyperparameters for the models are provided on the table below.

LSTM Hyperparameter	**Value**
Loss	'mean_squred_error'
optimizer	'adam'
epochs	30
Batch size	32
Epochs for Bi-LSTM with an attention layer	25

Table 1. The hyperparameters for the LSTM models.

3 Data

3.1 Data Set & Preprocessing

In order to create the data sets used in this experiment, earthquake data was collected from the US Geological Survey (USGS) earthquake database and the Kansas Geological Survey (KGS) earthquake database. During preprocessing the data from the two databases was combined, all duplicate instances were removed, and then data was split into two sets based on location. This resulted with a Puerto Rico data set containing 1436 earthquakes, and a Kansas data set containing 908 earthquakes. Both data sets contain only earthquakes with a magnitude of 2+ and span from January 2018 to December 2019.

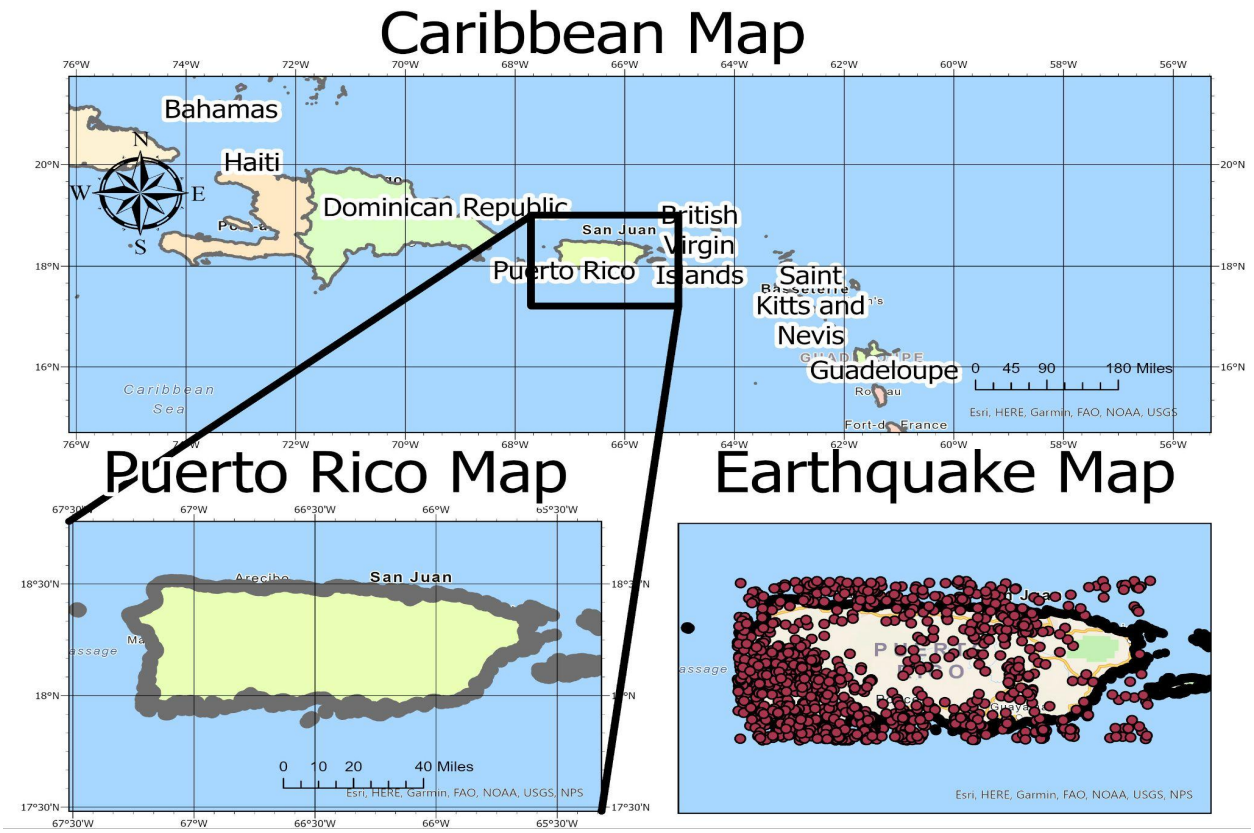

Figure 1. Distribution of earthquake epicenters (red circles) within the Puerto Rico data set.

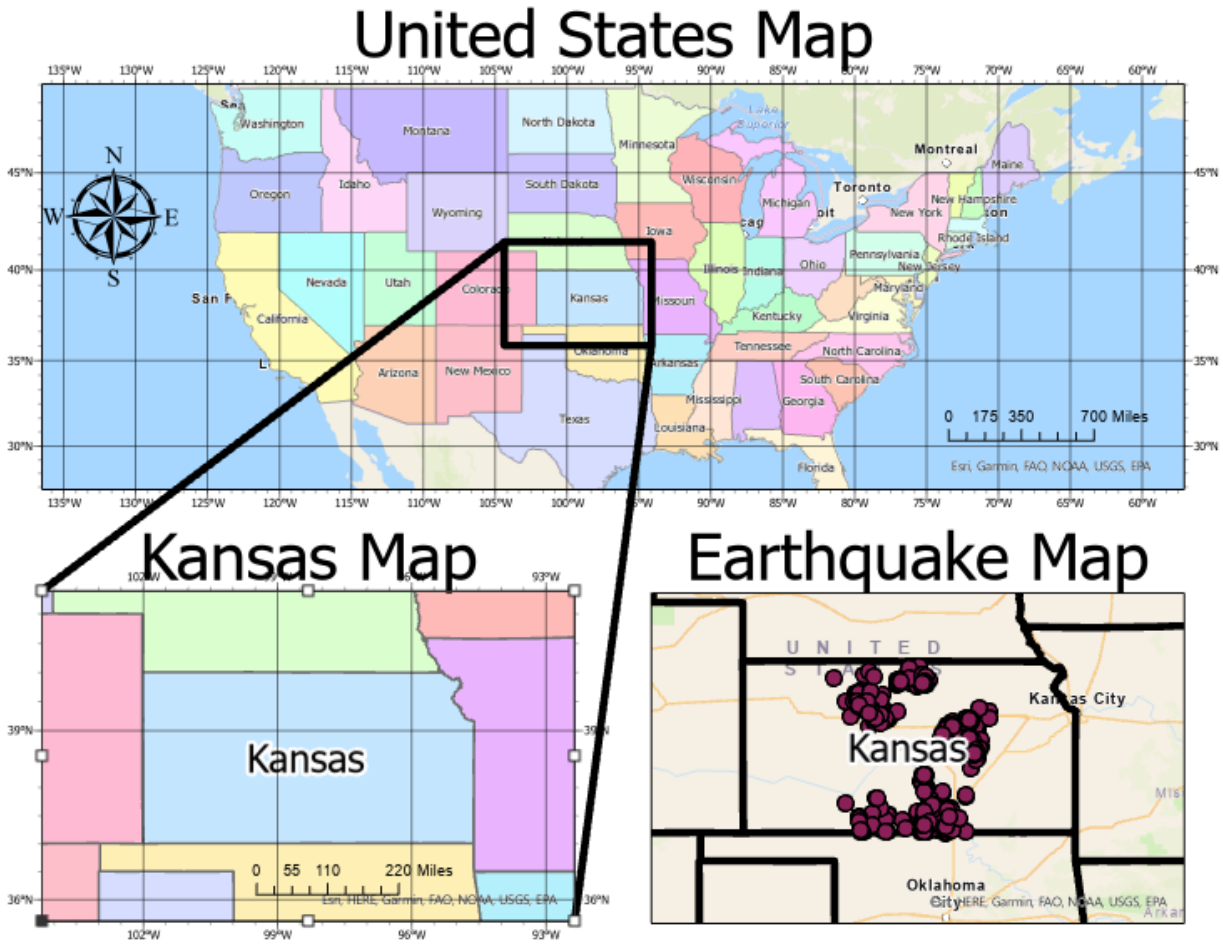

Figure 2. Distribution of earthquake epicenters (red circles) within the Kansas data set.

The resulting data sets include the following attributes associated with each earthquake: time, latitude, longitude, and magnitude. These four variables are also what the machine learning algorithms will be using to train their models, and what they will be ultimately attempting to predict. Data preprocessing included data merging, sampling, determining if there were any missing values, making a statistical summary, and data exploration.

	time	lat	long	mag
count	1436	1436	1436	1436
mean	3/2/2019 15:54:35	18.12247987	-66.72783294	2.348461003
std	228.9885707	0.2189891398	0.4076399862	0.3260447147
min	1/3/2018 2:36:40	17.8805	-67.2743	2
25%	9/30/2018 5:46:25	17.97575715	-67.00095173	2.130011044
50%	2/17/2019 6:40:38	18.0774	-66.8362	2.26
75%	8/3/2019 2:02:46	18.2692026	-66.45471415	2.566910962
max	12/31/2019 23:54:01	18.595	-65.5701	5

Table 2. The statistical Summary of the cleaned Puerto Rico data set.

	time	lat	long	mag
count	908	908	908	908
mean	12/30/2018 8:32:48	38.18453439	-98.04513143	2.365969151
std	208.3535793	0.9261152324	0.7349208297	0.4256889572
min	1/1/2018 4:27:24	37.00058	-100.00507	2
25%	8/12/2018 18:13:16	37.56403719	-98.53752839	2.08075755
50%	12/17/2018 17:52:27	38.01521	-97.967775	2.2
75%	5/18/2019 22:52:20	38.8050316	-97.55273448	2.651180752
max	12/30/2019 20:42:30	39.91831	-96.5312	4.8000002

Table 3. The statistical Summary of the cleaned Kansas data set.

3.2 Data Exploration

The Python programming language was used for all data processing and analysis. Numpy, Pandas, and Matplot libraries were applied for data cleaning and exploration. The Spyder development environment was used for all programming tasks, including data cleaning, visualization, and processing. Both graphs below were produced using Google Sheets, in order to visualize the distributions.

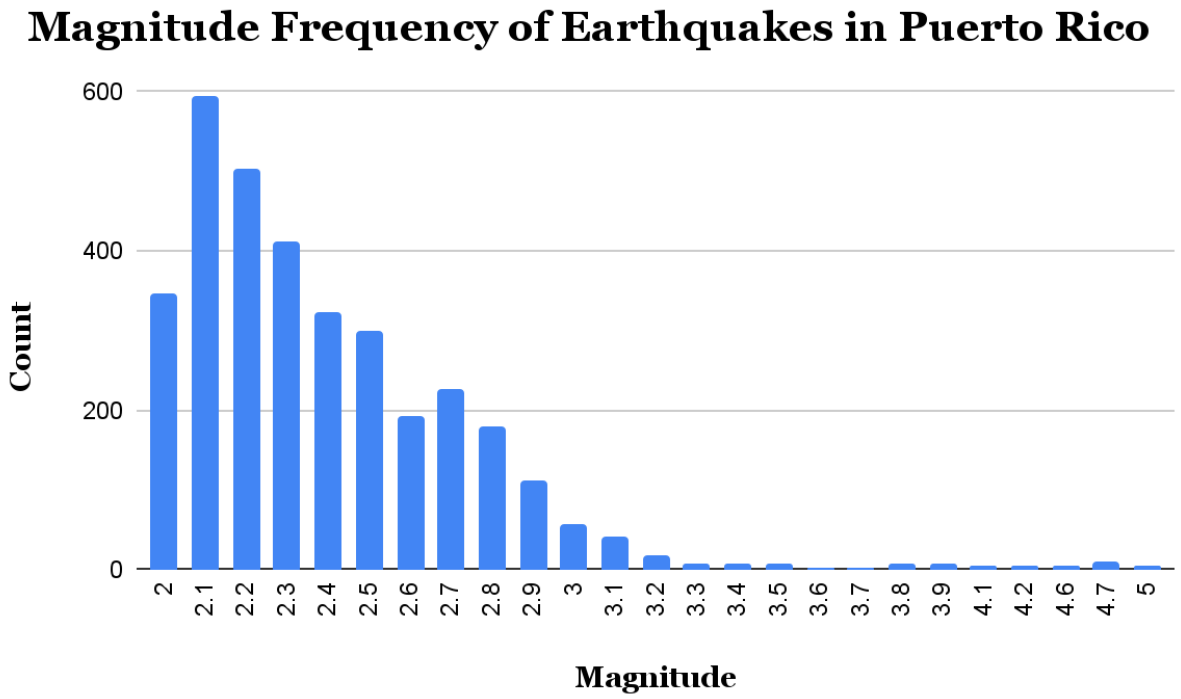

Figure 3. Frequency Plot of earthquake magnitude for Puerto Rico. For the years 2018-2019, only recording earthquakes with magnitude 2+.

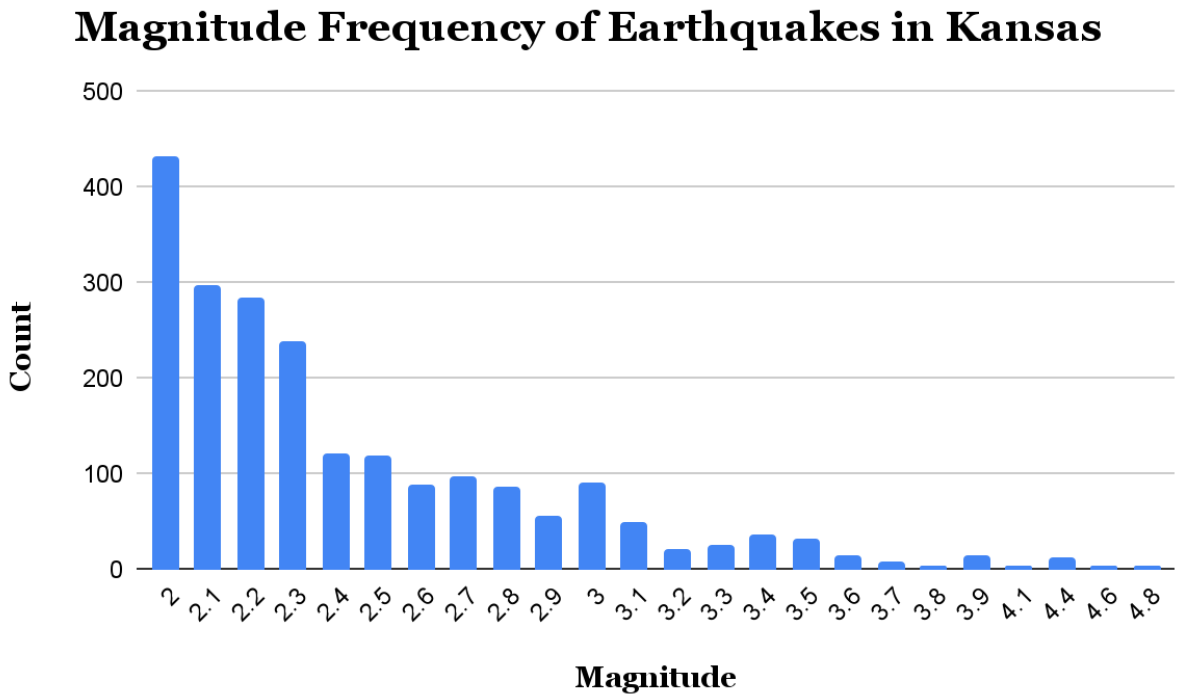

Figure 4. Frequency Plot of earthquake magnitude for Kansas. For the years 2018-2019, only recording earthquakes with magnitude 2+.

4 Results

4.1 Performance Measures

After completing data preprocessing, and modeling each of the four machine learning algorithms was run on the two data sets. After processing the Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) were calculated to evaluate the accuracy of each model.

4.1.1 Mean Absolute Error (MAE)

Mean Absolute Error is the measure of error between the calculated and expected results. It is commonly used to measure machine learning model performance, because its values are absolute, thus returning the same results independent of whether or not the model over or underestimates its predictions (Schneider and Xhafa, 2022). In general, MAE can be found using the following equation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - x\right|$$

Where:

n = the number of data points.

x_i = predicted value

x = true value

4.1.2 Mean Squared Error (MSE)

Mean Squared Error is another way to measure the error of a model. It assesses the average squared difference between the observed and predicted values. Squaring the differences serves several purposes, it eliminates negative values and ensures that the MSE is always greater than or equal to zero. Additionally, it increases the impact of larger errors, causing the MSE to rise substantially if there are large errors in the tested model (Frost, 2023). In general, MSE can be found using the following equation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(x - x_i\right)^2$$

Where:

n = the number of data points.

x_i = predicted value

x = true value

4.1.3 Root Mean Squared Error (RMSE)

Root Mean Squared Error is another way to measure the error of a model. It is the square root of the model's MSE, and is generally representative of the scatter of the observed values around the predicted values (Frost, 2023). RMSE is a commonly accepted statistic for model performance, and is simple enough to calculate that it is used by many with both little experience, and extensive experience with machine learning. However, there is a known weakness to using RMSE in that it is scale sensitive, this means that RMSE values can vary widely depending on the scale and number of values in a machine learning model. This means that unlike the other performance measures used in this experiment the RMSE may be a less reliable way to compare this model against others. In general, RMSE can be found using the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x - x_i\right)^2}$$

Where:

n = the number of data points.

x_i = predicted value

x = true value

4.1.4 Mean Absolute Percentage Error (MAPE)

Mean Average Percentage Error is another way to measure the error of a model. It measures accuracy as a percentage, and can be calculated as the average absolute percent error for each instance. MAPE is the most common measure used in forecasting analysis, likely because its units are scaled to a percentage, which makes it very easy to understand (Glen, 2023). MAPE works its best when there are no extremes or outliers in the data set, since large fluctuations can skew the statistic (Glen, 2023). In general, MAPE can be found using the following equation:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x-x_i}{x}\right|$$

Where:

n = the number of data points.

x_i = predicted value

x = true value

4.2 Model Performance

As each model was run the graph of its training loss and validation loss were recorded. The training loss indicates how well the model is fitting the training data, while the validation loss indicates how well the model fits new data (Abebe et al., 2023). When the training loss is greater than the validation loss the model is overfitting data, meaning that it is picking up too much noise to make accurate predictions. However, when the validation loss is greater than the training loss the model is underfitting data, meaning that it does not have enough information to make accurate predictions. When running a machine learning model the training loss should be as close to the validation loss as possible in order to represent an optimized model (Abebe et al., 2023). Each machine learning model was run 10 times, and then the average MAE, MSE, RMSE, and MAPE scores were recorded, and have been provided alongside the graph of the training and validation loss.

4.2.1 LSTM Performance

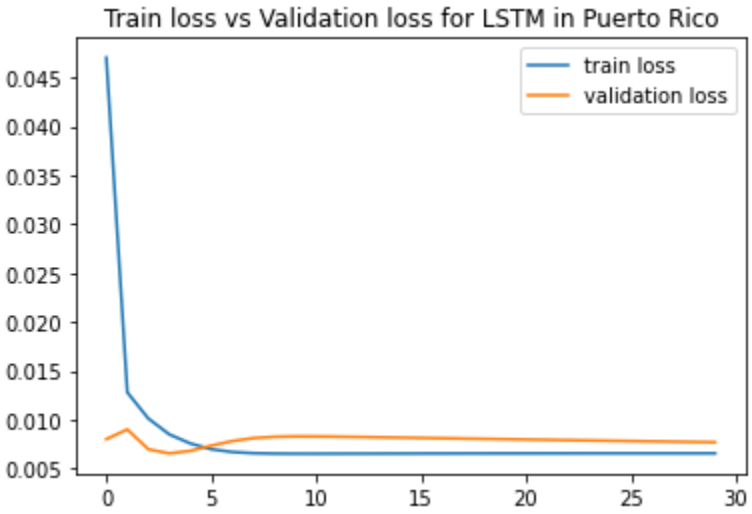

Figure 5. The training loss vs validation loss graph for the LSTM model when used on the Puerto Rico data set.

The graph of training loss versus validation loss for Puerto Rico shows that the LSTM model fits the data set well. There is an initial amount of overfitting, however once excess noise has been eliminated by the fifth epoch, the training loss and the validation loss stabilize within very close proximity to one another. After fitting the model displayed a MAE of 0.066, MSE of 0.008, RMSE of 0.088, and MAPE of 17.03%.

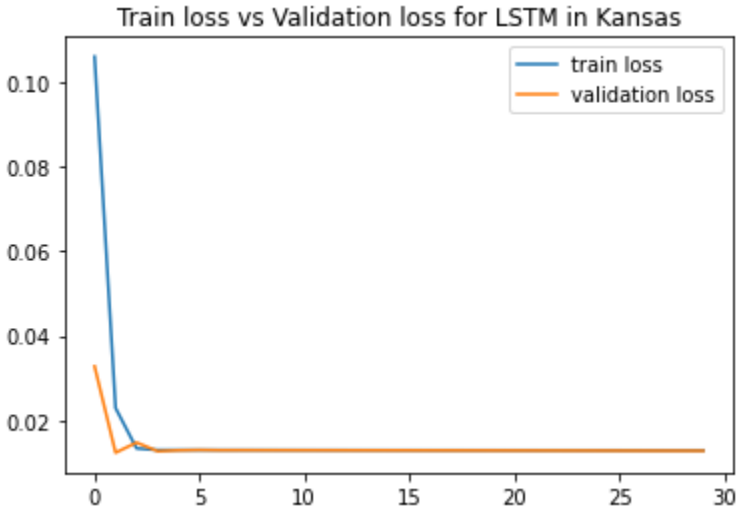

Figure 6. The training loss vs validation loss graph for the LSTM model when used on the Kansas data set.

The graph of training loss versus validation loss for Kansas shows that the LSTM model fits the data set well. There is an initial amount of overfitting, however once excess noise has

been eliminated by the fifth epoch, the training loss and the validation loss stabilize within very close proximity to one another. After fitting the model displayed a MAE of 0.096, MSE of 0.013, RMSE of 0.114, and MAPE of 29.19%.

4.2.2 BiLSTM Performance

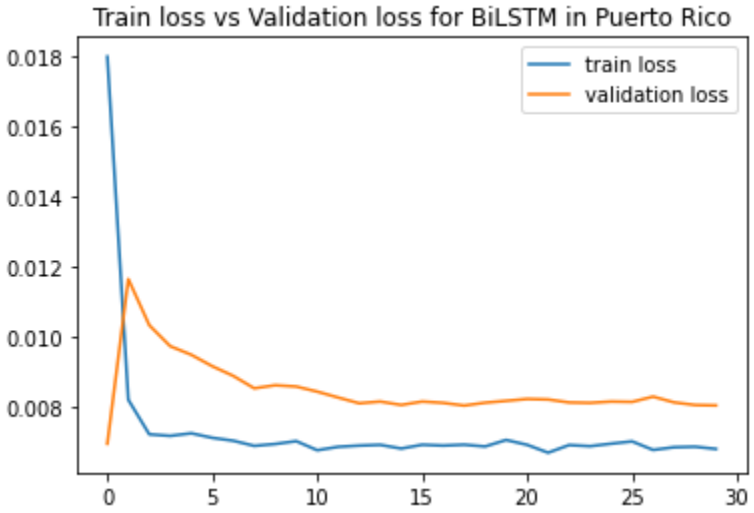

Figure 7. The training loss vs validation loss graph for the BiLSTM model when used on the Puerto Rico data set.

The graph of training loss versus validation loss for Puerto Rico shows that the BiLSTM model fits the data set well. There is an initial amount of overfitting, however once excess noise has been eliminated by the second epoch, the training loss and the validation loss stabilize within close proximity to one another, with a slight amount of underfitting. After fitting the model displayed a MAE of 0.067, MSE of 0.008, RMSE of 0.090, and MAPE of 17.40%.

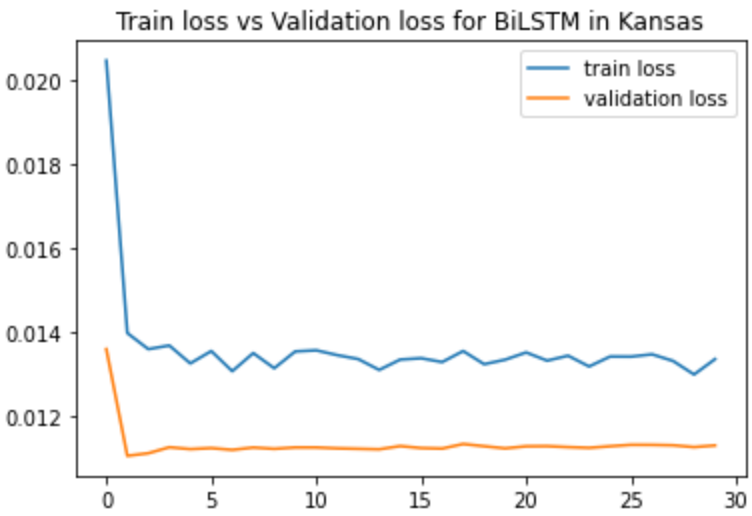

Figure 8. The training loss vs validation loss graph for the BiLSTM model when used on the Kansas data set.

The graph of training loss versus validation loss for Kansas shows that the BiLSTM model fits the data set well. There is an initial amount of overfitting, however once excess noise has been eliminated by the second epoch, the training loss and the validation loss stabilize within close proximity to one another, with a slight amount of overfitting. After fitting the model displayed a MAE of 0.081, MSE of 0.011, RMSE of 0.106, and MAPE of 23.39%.

4.2.3 BiLSTM with an attention layer Performance

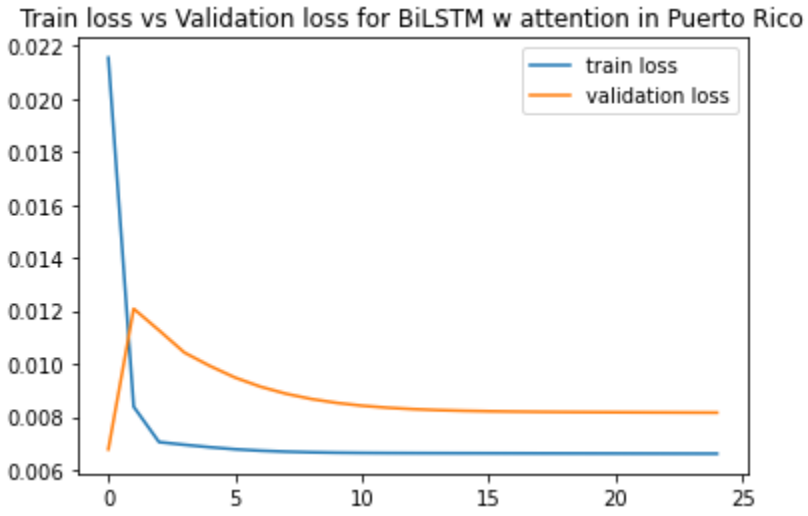

Figure 9. The training loss vs validation loss graph for the BiLSTM with an attention layer model when used on the Puerto Rico data set.

The graph of training loss versus validation loss for Puerto Rico shows that the BiLSTM with an attention layer model fits the data set well. There is an initial amount of overfitting, however once excess noise has been eliminated by the third epoch, the training loss and the validation loss stabilize within close proximity to one another, with a slight amount of underfitting. After fitting the model displayed a MAE of 0.068, MSE of 0.008, RMSE of 0.090, and MAPE of 17.40%.

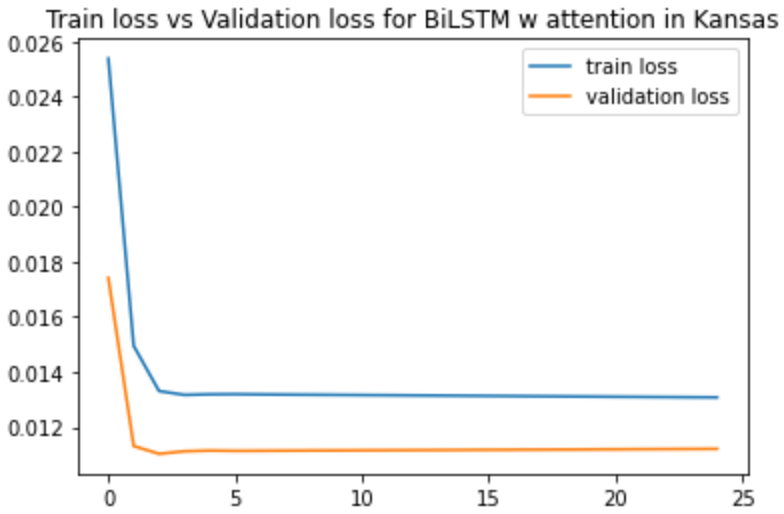

Figure 10. The training loss vs validation loss graph for the BiLSTM with an attention layer model when used on the Kansas data set.

The graph of training loss versus validation loss for Kansas shows that the BiLSTM with an attention layer model fits the data set well. There is an initial amount of overfitting, however once excess noise has been eliminated by the third epoch, the training loss and the validation loss stabilize within close proximity to one another, with a slight amount of overfitting. After fitting the model displayed a MAE of 0.080, MSE of 0.011, RMSE of 0.106, and MAPE of 22.98%.

4.3 Discussion

The results provided by this experiment help to prove that while machine learning is making earthquake prediction feasible, there is still much about the process that needs further understanding. This experiment was only able to cover two data sets and three processing models, so the room for further investigation into the realm of machine learning for earthquake prediction remains vast. Once the data sets were fully processed their performance measures were collected and displayed on the table and figures below.

Model	LSTM	LSTM	BiLSTM	BiLSTM	BiLSTM w attention	BiLSTM w attention	Best Model and Data Set
Data Set	Puerto Rico	Kansas	Puerto Rico	Kansas	Puerto Rico	Kansas	
MAE	0.066	0.096	0.067	0.081	0.068	0.080	**LSTM (Puerto Rico)**
MSE	0.008	0.013	0.008	0.011	0.008	0.011	**LSTM (Puerto Rico)**
RMSE	0.088	0.114	0.090	0.106	0.090	0.106	**LSTM (Puerto Rico)**
MAPE	17.03%	29.19%	17.40%	23.39%	17.40%	22.98%	**LSTM (Puerto Rico)**

Table 4. Comparison of performance measures across machine learning models and data sets

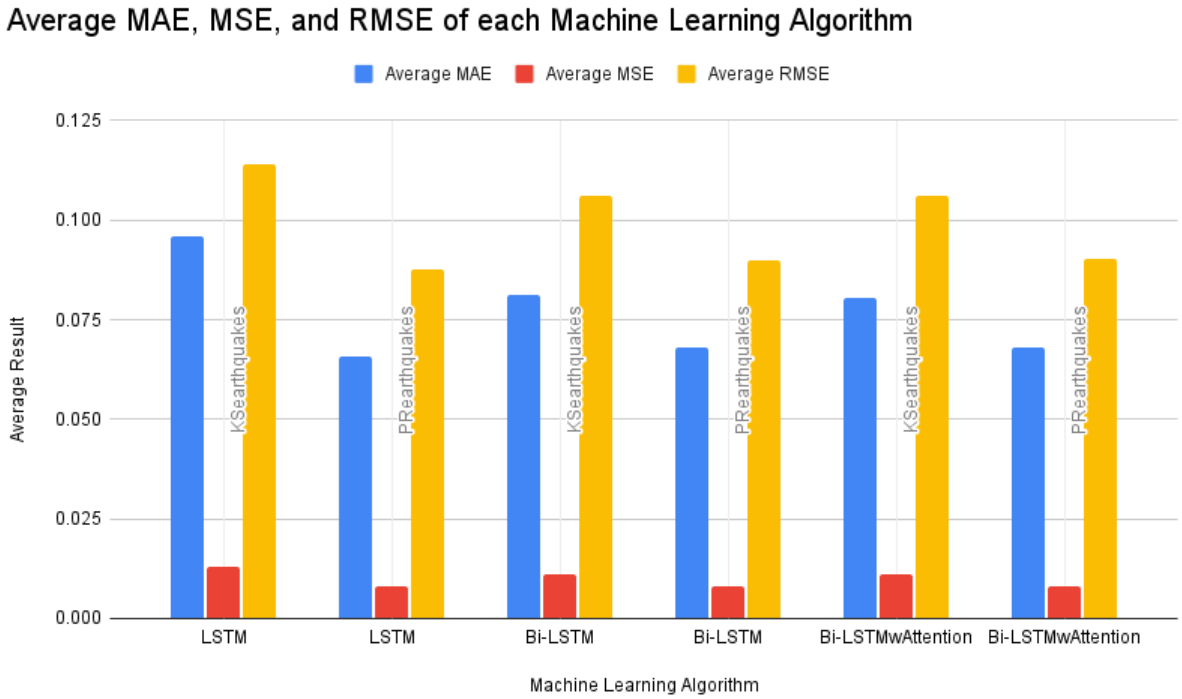

Figure 11. The average MAE, MSE, and RMSE between both data sets and all four machine learning models.

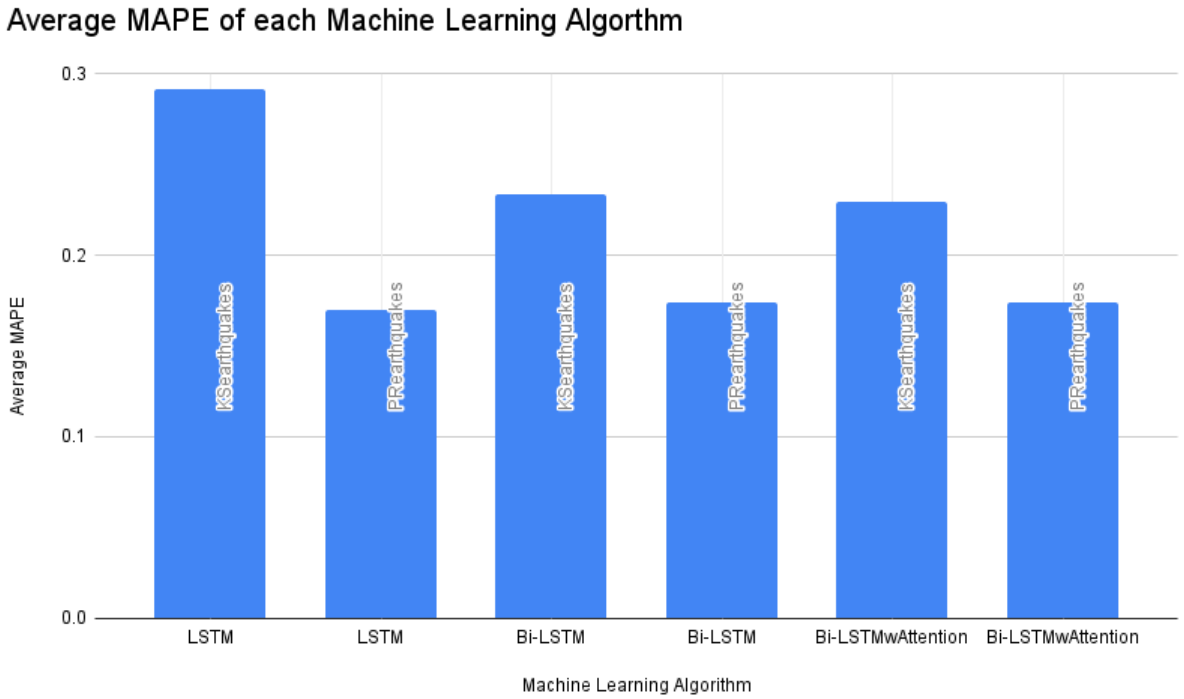

Figure 12. The average MAPE between both data sets and all four machine learning models.

As displayed on table 3 and on figures 11 and 12 the Puerto Rico data set combined with the normal LSTM model produced the lowest error, and thus the best results, with a MAE of 0.066, MSE of 0.008, RMSE of 0.088, and MAPE of 17.03%. The Kansas data set performed the best with the BiLSTM with an attention layer model, resulting in an MAPE of 17.40%. This is likely due to a difference in the data sets. Meaning, this experiment does succeed in establishing some connection between location, tectonic setting, and machine learning performance on earthquake data sets. When comparing the Puerto Rico data set with the Kansas data set some interesting differences emerge. Firstly, not only does the Puerto Rico data set perform the best of the two, but it performs at its best when used along with the least complex of the machine learning models, the LSTM. The data set performs similarly, but marginally worse on the other two LSTM based models. Each new model introduces another level of complexity to the traditional LSTM model, by either making it a bidirectional model, or introducing an attention layer. This could also be related to the fact that the Puerto Rico data set continually underfit its predictions, since the model determined too much of the data set as noise, it was unable to make use of the tools provided by the more complex models. Looking back at Figure 1. The map of Puerto Rico, and its associated earthquake data set, it can be seen how widely spread the earthquakes are with only minor amounts of clustering where major faults are the most active. This distribution of earthquakes is a direct result of the island's tectonic setting, where complex fault networks are regularly stressed by nearby plate boundaries. In this way the tectonic setting of Puerto Rico could be the reason as to why simple machine learning models work better there than complex ones. Conversely, the Kansas data set showed the most success with the BiLSTM with an attention layer model. Unlike Puerto Rico the Kansas data set showed a decrease in error

each time that the LSTM model increased in complexity. Another difference between data sets is that the Kansas data set regularly overfit its estimations, rather than underfit them, meaning that it was learning too well and began to consider some of the noise in the data set as part of the pattern. Many of the newest and more complex varieties of the LSTM, like the BiLSTM or the attention layer are specifically designed to combat overfitting errors (Graves and Schmidhuber, 2005). The reason why the Kansas data set is overfit rather than underfit is likely due to the distribution of events, and tectonic setting. Since it is located in the center of a continental plate Kansas has a far less active tectonic setting, and only one major fault. This means that almost all of the events, as they are displayed on Figure 2. are grouped together in one of three to four clusters around the state. This makes the task of predicting earthquakes easier on the machine, since it can identify the clusters and can predict earthquakes within the clusters with reasonable assuredness. The Puerto Rico data set ultimately still performed better, however that is more likely due to the simple fact that it had the largest data set.

5 Conclusions

With the use of machine learning to predict earthquakes a growing possibility, it is valuable to take a step back and look at the methods behind the predictions, and how scientists in the future can cater the machine learning models that they use towards the region that they survey. The results of this experiment show that in regions where earthquake clustering is not present, or only weakly present, such as in complexly faulted locations, or locations near multiple plate boundaries that simpler machine learning models such as the LSTM work best. Whereas, in locations with earthquake clustering is present, such as in locations with no plate boundaries or little tectonic activity complex machine learning models work better, like the BiLSTM with an attention layer model. This is due to how the distribution of earthquakes affects machine learning, thus indicating a correlation between the tectonic setting and the performance of machine learning models in a region. Given the error results produced by this experiment it is possible that earthquakes will be predicted by machine learning programs in the near future, however even the most accurate of results still show a reasonable degree of error, meaning that there are further improvements that can still be made in the field.

Acknowledgments

This study is supported by the Disaster Resilience Analytics Center of Wichita State University.

Eldon Taskinen and Zelalem Demissie are supported by the Geology Department at Wichita

State University. The authors thank the editor and two anonymous reviewers for their

constructive reviews that helped improve the manuscript.

Data Availability Statement

Datasets for this research are available upon request.

References

Abebe, E., Kebede, H., Kevin, M., Demissie, Z., (2023), Earthquakes magnitude prediction using deep learning for the Horn of Africa, *Soil Dynamics and Earthquake Engineering*, 170, 1-44. doi:10.1016/j.soildyn.2023.107913

Alaskar, H., and Saba, T., (2021), Machine Learning and Deep Learning : A Comparative Review Machine Learning and Deep Learning : A Comparative Review doi:10.1007/978-981-33-6307-6.

Baars, D. L., Watney, W. L., Steeples, D. W., Brostuen, E. A., (1989), Petroleum: a primer for Kansas, *Kansas Geological Survey*, Education, URL:HTTP://www.kgs.ku.edu/Publications/Oil/index.html

Bahdanau, D., Cho, K., Bengio, Y., (2014), Neural Machine Translation by Jointly Learning to Align and Translate, *ICLR Computation and Language*. doi:10.48550/arXiv.1409.0473

Bakun, W. H., McEvilly, T. V., (1984), Recurrence models and Parkfield, California, earthquakes, *Solid Earth*, 89(B5), 3051-3058. doi:10.1029/JB089iB05p03051

Bellamkonda, S., Settipalli, L., Vedantham, R., and Vemula, M. K., (2021), An Enhanced Earthquake Prediction Model Using Long Short-Term Memory, *Turkish Journal of Computer and Mathematics Education*, 12, 2397–2403.

Benford, B., DeMets, C., Calais, E., (2012), GPS estimates of microplate motions, northern Caribbean: evidence for a Hispaniola microplate and implications for earthquake hazard, *Geophysical Journal International*, 191(2), 481-490. doi:10.1111/j.1365-246X.2012.05662.x

Bengio, Y., Simard, P., Frasconi, P., (1994), Learning long-term dependencies with gradient descent is difficult, *IEEE Transactions on Neural Networks*, 5(2), 157-166. doi:10.1109/72.279181

Berendsen, P., (1997), Tectonic evolution of the Midcontinent Rift System in Kansas, Middle Proterozoic to Cambrian Rifting, Central North America, *Geologic Society of America*, 235-241. doi:10.1130/0-8137-2312-4.235

Berhich, A., Belouadha, F., and Kabbaj, M. I., (2020), LSTM-based Models for Earthquake Prediction, *Association for Computing Machinery*, doi:10.1145/3386723.3387865.

Bird, P., Kagan, Y.Y., (2004), Plate-Tectonic Analysis of Shallow Seismicity: Apparent Boundary Width, Beta, Corner Magnitude, Coupled Lithosphere Thickness, and Coupling in Seven Tectonic Settings, *Bulletin of the Seismological Society of America*, 94(6), 2380-2399, doi:10.1785/0120030107

Brosius, L., Steeples, D.W., (2014), Earthquakes, *Kansas Geological Survey*, Publications, URL:https://www.kgs.ku.edu/Publications/pic3/pic3_1.html

Buchanan, R., (2000), Earthquake Risk Low, but Real in Some Parts of State, Kansas Geological Survey, News Release, URL:http://www.kgs.ku.edu/General/News/2000/earthquake.html

Frost, J., Mean Squared Error (MSE), Statistics by Jim, URL:https://statisticsbyjim.com/regression/mean-squared-error-mse/ , Accessed: 9/25/2023

Frost, J., Root Mean Squared Error (RMSE), Statistics by Jim, URL:https://statisticsbyjim.com/regression/root-mean-square-error-rmse/ , Accessed: 9/25/2023

Geller, R. J., Jackson, D. D., Kagan, Y. Y., Mulargia, F., (1997), Earthquakes Cannot Be Predicted, *Science*, 275(5306) 1616. doi:10.1126/science.275.5306.1616

Geography, US Census Bureau. "State Area Measurements and Internal Point Coordinates". Archived from the original on March 16, 2018. Retrieved May 31, 2016.

Glen, S., Mean Absolute Percentage Error (MAPE) StatisticsHowTo: Elementary Statistics for the rest of us! Accessed: 9/25/2023

URL:https://www.statisticshowto.com/mean-absolute-percentage-error-mape/

Graves, A., Schmidhuber, J., (2005), Framewise phoneme classification with bidirectional LSTM networks, *IEEE International Joint Conference on Neural Networks*. doi:10.1109/IJCNN.2005.1556215

Graves, A., Liwicki, M., Fernandez, S., Bertolami, R., Bunke, H., Schmidhuber, J., (2009), A Novel Connectionist System for Unconstrained Handwriting Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5), 855-868. doi:10.1109/TPAMI.2008.137

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., Schmidhuber, J., (2017), LSTM: A Search Space Odyssey, *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222-2232. doi:10.1109/TNNLS.2016.2582924

Holzer, T. L., Savage, J. C., (2013), Global Earthquake Fatalities and Population, *Earthquake Specra*, 29(1), 155-175. doi:101193/1.4000106

Jackson, D. D., Kagan, Y. Y., (2006), The 2004 Parkfield Earthquake, the 1985 Prediction, and Characteristic Earthquakes: Lessons for the Future, *Bulletin of the Seismological Society of America*, 96(4B), S397-S409. doi:10.1785/0120050821

Jansma, P. E., Mattioli, G. S., (2005), GPS Results from Puerto Rico and the Virgin Island: Constraints on Tectonic Setting and Rates of Active Faulting, Active Tectonics and Seismic Hazards of Puerto Rico, the Virgin Islands, and Offshore Areas, *Geological Society of America*, 13-25. doi:10.1130/0-8137-2385-X.13

Karevan, Z., Suykens, J. A., (2020), Transductive LSTM for time-series prediction: An application to weather forecasting, *Neural Networks*, 125, 1-9. doi:10.1016/j.neunet.2019.12.030

Mann, P., Prentice, C. S., Hippolyte, J. C., Grindlay, N. R., Abrams, L. J., Lao-Davila, D., (2005), Reconnaissance study of Late Quaternary faulting along Cerro Goden fault zone, western Puerto Rico, *GSA Special Papers: Active Tectonics and Seismic Hazards of Puerto Rico, the Virgin Islands, and Offshore Areas*, 385, 115-138. doi:10.1130/0-8137-2385-X.115

Merriam, D. F., (1963), The Geologic History of Kansas, Kansas Geological Survey Bulletin 162, URL:http://www.kgs.ku.edu/Publications/Bulletins/162/index.html

Peterie, S. L., Miller, R. D., Intfen, J. W., Gonzales, J. B., (2018), Earthquakes in Kansas Induced by Extremely Far-Field Pressure Diffusion. *Geophysical Research Letters*, 45(3), 1395-1401. doi:10.1002/2017GL076334

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., Johnson, P. A., (2017), Machine Learning Predicts Laboratory Earthquakes, *Geophysical Research Letters,* 44(18), 9276-9282. doi:10.1002/2017GL074677

Schneider, P., Xhafa, F., (2022), Anomaly Detection and Complex Event Processing over IoT Data Streams with Application to eHealth and Patient Data Monitoring, *Chapter 3- Anomaly Detection: Concepts and methods*, 49-66. ISBN:9780128238189

Veizer, J., Mackenzie, F. T., (2014), 9.15 - Evolution of Sedimentary Rocks, Treatise on Geochemistry, *Reference Module in Earth Systems and Environmental Science*, 9, 399-435. doi: 10.1016/B0-08-043751-6/07103-6

Viltres, R., Nobile, A., Vasyura-Bathke, H., Trippanera, D., Xu, W., Jonsson, S., (2021), Transtensional Rupture within a Diffuse Plate Boundary Zone during the 2020 Mw 6.4 Puerto Rico Earthquake, *Seismology Research Letters*, 93, 567-583. doi:10.1785/0220210261

Wood, H. O., Gutenberg, B., (1935), Earthquake Prediction, *Science*, 82(2123), 219-220. doi:10.1126/science.82.2123.219

Wyss, M., (1997), Cannot Earthquakes Be Predicted?, *Science*, 278(5337) 487-490. doi:10.1126/science.278.5337.487

Wyss, M., (2001), Why is earthquake prediction research not progressing faster?, *Tectonophysics*, 338(3-4) 217-223. doi:10.1016/S0040-1951(01)00077-4

Yadav, A., Jha, C. K., Sharan, A., (2020), Optimizing LSTM for time series prediction in Indian stock market, *Procedia Computer Science,* 167, 2091-2100 doi:10.1016/j.procs.2020.03.257

The Impact of Tectonic Setting on Machine Learning Approaches for Earthquake Prediction

Enter authors here: E. Taskinen[1], Z. Demissie[2]

[1] Wichita State University.

[2] Wichita State University.

Corresponding author: Eldon Taskinen (eldontaskinen@gmail.com)

†Additional author notes should be indicated with symbols (current addresses, for example).

Key Points:

- Earthquake data distribution is driven by underlying tectonic structures.
- Data distribution can affect how Machine Learning algorithms process data.
- Different Machine Learning algorithms work best for different distributions and tectonic structures.

Abstract

In prior decades the concept of using mathematical methods to predict earthquakes was considered infeasible. Recent advances in machine learning and predictive modeling offer promising avenues to potentially realize earthquake prediction. In order to test the viability of machine learning methods, experiments were made with earthquake datasets from Kansas and Puerto Rico. The two datasets were chosen for the distinct differences in their tectonic settings. Kansas has few major faults, with a largely inactive subsurface, this produced a smaller dataset with a few large clusters. Puerto Rico is complexly faulted, with an extremely active tectonic setting, this produced a larger dataset with a large number of small clusters. In order to test the effectiveness of these two datasets for machine learning and prediction they were run through three different machine learning algorithms including an LSTM model, Bi-LSTM model, Bi-LSTM model with attention. Not only were the three different machine learning methods compared against each other for accuracy but also the datasets as well. Conclusive findings show that the two different data sets favor different processing methods. The Kansas data performs the best with the Bi-LSTM with attention model, while the Puerto Rico data performs the best with the LSTM model. This is likely due to the tectonic settings of the two regions, since the Kansas dataset has less overall data, and earthquakes are concentrated in a few large clusters, while the Puerto Rico data set has a more even distribution.

Plain Language Summary

Advances in machine learning may make earthquake prediction possible. In order to test the viability of machine learning methods, experiments were made with earthquake datasets from Kansas and Puerto Rico. The two datasets were chosen for the distinct differences in their tectonic settings. Three different machine learning algorithms were tested on the two data sets and then compared for accuracy of predictions. Conclusive findings show that the two different data sets favor different processing methods. The Kansas data performs the best with complex models, while the Puerto Rico data performs the best with simplistic models. This is due to the tectonic settings of the two regions, indicating that future exploration into earthquake prediction will need to consider tectonic settings into how data is processed.

1 Introduction

1.1 Problem Statement

Earthquakes are tremendous natural disasters that can cost hundreds or even thousands of lives, often leaving utter destruction in their wake. Current predictions estimate that there will be 7.5∓3.0 catastrophic earthquakes in the 21st Century, and that each of these earthquakes will cause an average of 1.45∓0.58 million casualties (Holzer and Savage, 2013). It is the goal of this research, and all earthquake prediction efforts to prevent as many future casualties as possible, by accurately and precisely predicting earthquakes, so that effective preventative measures can be taken. In the only widely successful instance of earthquake prediction, Chinese authorities were able to warn the town of Haicheng shortly before a magnitude 7.3 earthquake (Wood and Gutenberg, 1935) . Due to the short-term preparations that the people of Haicheng were able to make, the reported death toll was kept under 300 in an area where tens of thousands of fatalities would have been expected (Wood and Gutenberg, 1935). This research aims to implement methods of earthquake prediction that have already shown success in the Horn of Africa. The Horn of Africa experiment was able to predict earthquakes down to a 28.87% error using machine learning algorithms (Abebe et al., 2023). This is a startlingly low margin of error, for the field of earthquake prediction. The goal of this research is to verify these results as well as to see how modifications to the internal data set and hyperparameters change the outcome. If successful, the machine learning techniques used in this experiment can be considered for wider implementation. If the machine learning methods such as the LSTM algorithm used in the Horn of Africa experiment prove effective at predicting earthquakes elsewhere on the globe these advancements could lead to millions of lives being saved.

1.2 Background

Since the 1970's seismologists have been trying to use mathematical methods to predict earthquakes (Bakun and McEvilly, 1984). Early attempts at earthquake prediction relied upon finding patterns of earthquakes and then extrapolating the results over time (Bakun and McEvilly, 1984). A notable early example of this was the Parkfield Experiment in the 1980's, where a 20- to 30- km long section of the San Andreas fault near Parkfield, California had suffered notable earthquakes in 1857, 1881, 1901, 1922, 1934, and 1966 (Bakun and McEvilly, 1984). Considering these repeated earthquakes to be characteristic of a pattern the authors confidently hypothesized that the next notable earthquake in the Parkfield area would happen between 1983 and 1993 (Bakun and McEvilly, 1984). The Parkfield earthquake did not happen until 2004. This was considered a big disappointment for the field of earthquake prediction, and seismologists at the time began to question the viability of statistical approaches as a whole (Jackson and Kagan, 2006). The attitude of most seismologists after the Parkfield Experiment was best summed up in the article "Earthquakes Cannot Be Predicted" where the authors blatantly proposed that earthquakes are inherently unpredictable, citing chaos theory, as well as the unmeasurably fine conditions that surround an earthquake event as reasons why statistical prediction is impossible (Geller et al., 1997). This sentiment has slowed progress inside of the field of seismology, with many funding agencies cautious to support research into predictions that were subject to criticism and showed high degrees of error (Wyss, 2001). However, even then there still remained a few willing to continue working on mathematical methods of

earthquake prediction (Wyss, 1997). That same year an article was published in response titled, "Cannot Earthquakes Be Predicted?" arguing that earthquake prediction is still an emerging field, where early progress was unfolding slowly (Wyss, 1997).

In the 90's and through most of the ensuing time between then and now this is where the discussion of earthquake prediction has remained. Caught in opposition between a majority that believes it to be impossible, and a minority who were still trying to make it happen. The next notion of a change in fortune in the field came in 2017, when a machine learning algorithm successfully predicted earthquakes in a laboratory setting. This was one of the earliest applications of machine learning techniques to the field of earthquake detection (Rouet-Leduc et al., 2017). In the past few years, the development of easy to access software libraries like Theano and TensorFlow has allowed many seismologists to explore the viability of machine learning techniques (Rouet-Leduc et al., 2017). Machine learning allows for many leaps forward in data processing that were previously impossible, especially in terms of the number of data points and variables that can be processed. This all culminated in the Horn of Africa experiment where four of the most recently developed machine learning algorithms are tested against each other for a comparison of accuracy in the Horn of Africa (Abebe et al., 2023).

1.3 Research Objective

In order to conduct this experiment two data sets have been developed, one containing all magnitude 2+ earthquakes whose epicenters were located in the state of Kansas during the years 2018 and 2019, as well as a secondary dataset containing all magnitude 2+ earthquakes whose epicenters were located in Puerto Rico during the years 2018 and 2019. The two regions were chosen as datasets primarily for their structural differences. Of the data sets used in this research Puerto Rico is the more active than the other data set, Kansas. Puerto Rico's tectonic system is located at the eastern edge of the Greater Antilles, and part of the Caribbean tectonic complex (Jansma and Mattioli, 2005). The complexity of the region is due to Puerto Rico, Gonave in the west, and Hispaniola in the center, all occupying their own microplates, which all interact tectonically with one another (Jansma and Mattioli, 2005). Puerto Rico also has large fault zones located on the North and South ends of the island which have characteristic oblique normal faults with a possible right-lateral slip (Jansma and Mattioli, 2005). This makes Puerto Rico a hot spot for earthquake activity. The other dataset used in this experiment is based out of the state of Kansas. Unlike the other location previously discussed, Kansas is a midcontinental region, with only a few subsurface rifts and ridges that date back to the Pennsylvanian or are even older (Berendsen, 1997). This means that with a few regional exceptions that Kansas is rather inactive on the tectonic scale, to complicate things further, recent evidence points towards man-made, or induced earthquakes becoming a risk in Kansas, due to anthropogenic processes (Peterie et al., 2018). By comparing the results of machine-learning predictions in Kansas, and Puerto Rico, it should become apparent how variables such as geography, tectonic setting, and local seismicity play a role in the accuracy of machine-learning predictions.

2 Methods

2.1 Tectonic Setting

In order to achieve the stated goal of this research an understanding must first be established regarding what is known about the tectonic setting of each of the data set locations. In general, the term tectonic setting refers to the principal controlling factors behind a region's

lithology, chemistry, and structure (Veizer and Mackenzie, 2014). This includes dominant faults, tectonic plates, and plate boundaries. The tectonic setting of a region has long played a role in determining its seismicity, that is where and how often earthquakes occur (Bird and Kagan, 2004). That is to say if a region's tectonic setting is dominated by a large subduction zone is more likely to experience earthquakes than a region whose tectonic setting is determined by a continental rift boundary, which in turn is more likely to experience earthquakes than a region that does not have any plate boundaries (Bird and Kagan, 2004). As part of the technical background for this research, two different tectonic settings are described in this section. Firstly, this section will cover the tectonic setting of Puerto Rico. Secondly, this section will cover the tectonic setting of Kansas. The two data sets for this research. Puerto Rico and Kansas were chosen as the data sets for this research due to the prominent differences in their tectonic setting, in order to try and detect differences in machine learning processes later down the line.

2.1.1 Puerto Rico

The island of Puerto Rico is part of the Caribbean archipelago and has an area of roughly 9,104 km2 (Benford et al., 2012). It is also part of the Caribbean plate, which has major boundaries with both North American, South American, Nazca, and Cocos plates.

In the global tectonics system Puerto Rico lies on the Puerto Rico-Virgin Islands microplate, abbreviated PR-VI (Benford et al., 2012). The microplate is part of the larger Caribbean plate, and is a complex geodynamic setting, characterized by oblique collision to the west, oblique subduction at the longitude of Puerto Rico, and frontal subduction to the east (Viltres et al., 2021). One of the most interesting factors of this seismic environment is the trenches located both to the North and South of the island which is caused by double underthrusting by the Caribbean plate. These two trenches along with the three east-west strike-slip zones account for nearly 85% of the relative motion between the Caribbean and North American Plates (Mann et al., 2005). The trenches combine with the Anegada Deep and Mona Deep faults to design the boundary of the PR-VI microplate.

While the microplate is generally considered a single solid block there are two major northwest-southeast faults named the Great Northern and Great Southern faults respectively that can be found further inland. These inland faults meet up with faults in the ocean to create a larger network of faults that has been recorded to be very active during the Quaternary (Jansma and Mattioli, 2005). All of these active faults means that Puerto Rico has a relatively high natural seismicity, making it a good candidate for comparison with the other data set used in this research.

2.1.2 Kansas

The state of Kansas is located in the center of the continental United States and has an area of roughly 213,278 km^2 (U.S. Census Bureau, 2016). It is also part of the North American plate, which has major boundaries with the Pacific, Eurasian, African, South America, and Caribbean plates, and many smaller boundaries with minor plates.

Due to being located in the center of the world's largest continental plate, the tectonic setting of Kansas is remarkably stable. Located on the large, stable craton that composes most of North America, Kansas has had little interaction with any tectonic boundaries since the Precambrian (Merriam, 1963). As such, the state experiences relatively few earthquakes, ranking 45th among 50 of the nation's states in the amount of damage caused by earthquakes on the average year (Buchanan, 2000). The one exception to this is the Humboldt fault zone, part of the Nemaha uplift in the eastern portion of the state (Buchanan, 2000). This fault zone is a series of normal faults that are associated with movement in the North American craton that occurred in the protozoic (Baars et al., 1989).

During the past decade Kansas has experienced a rise in earthquake events. Prominent theories point towards the phenomena known as induced seismicity to be the cause (Peterie et al., 2018). Induced seismicity is a rise in earthquake events that is caused by human activity, usually in association with the production of oil and gas, as well as the disposal of its byproducts. Typically induced seismicity is caused by a single well, which injects the brine byproduct that is produced during oil and gas extraction back into the subsurface, usually in very high quantities. This increases the pore pressure of reservoir rock surrounding the well, which is sometimes large enough to trigger earthquakes a short distance away (Peterie et al., 2018). Recent work done in Kansas and Oklahoma however, suggests that fluid injected into the subsurface is migrating further than expected, and causing earthquakes throughout the state (Peterie et al., 2018). Due to the unique set of circumstances surrounding the earthquake record in Kansas, it will serve as a suitable juxtaposition to the more naturally active tectonic setting of Puerto Rico when the two are compared in the experiment.

2.2 Machine Learning Techniques

Machine learning methods are among the newest and most promising methods for earthquake prediction (Rouet-Leduc et al., 2017). Machine learning uses various different algorithms to develop a prediction model from data that is input into the algorithm. The machine trains itself on the input data, using methods similar to the human brain to sort out noise and identify patterns (Alaskar and Saba, 2021). Then, using what it has learned the machine is able to complete complex tasks, such as predicting future events in a time series, or identify and reproduce text and images. It is specifically the use of machine learning for time series predictions that interest seismologists, since each earthquake happens at a distinct time, the dataset can be represented as a time series, allowing for machine learning analysis (Alaskar and Saba, 2021). The three algorithms explored in this research are all variations on the Long Short-Term Memory model. Long Short-Term Memory (LSTM) models are an advanced form of neural networking, specifically designed for the purpose of time series predictions and have shown successful implementation in the prediction of everything from the weather to the stock market (Karevan and Suykens, 2020; Yadav et al., 2020). Along with the base LSTM model, this

experiment will also be testing the Bidirectional LSTM, and the Bidirectional LSTM with attention layer models, variations on the normal LSTM model, with added layers of complexity in order to optimize performance in some data sets. These three models will be tested against each other to see if the two locations and the two datasets change the relationship between the machine learning methods.

2.2.1 Long Short-Term Memory (LSTM)

The LSTM model is a recurrent neural network that has already seen some success in the field of earthquake prediction (Berhich et al., 2020; Bellamkonda et al., 2021). Recurrent networks like the LSTM function by 'remembering' and 'forgetting' information in a way meant to mimic the human brain (Bengio et al., 1994). The model will store information about past inputs, but only for a set amount of time, that time will depend on the weight given to the data by the algorithm while it is learning. All recurrent networks share 3 core functions, to store information for an arbitrary duration, to be resistant to noise (i.e. fluctuations of the inputs that are random or irrelevant to predicting a correct outcome), and to be trainable (Bengio et al., 1994). This all allows the model to be context sensitive, and show a reasonable degree of flexibility. The reason why the LSTM model has become the preferred recurrent network is because of how it solves an issue that is prevalent among those networks. The range of contextual information that a recurrent network can access is actually quite limited, the problem being that when given influence over a network's learning mechanism most information causes it to either decay or expand exponentially (Graves et al., 2009). This problem, known as vanishing and exploding gradients, is specifically what the LSTM model was introduced to solve (Graves and Schmidhuber, 2005). The LSTM functions by using a series of layers, blocks, and cells. The model contains one input layer, one hidden layer, and one output layer, and it is the hidden layer which contains the memory cells that determines what information the model keeps and what information it discards (Hochreiter and Schmidhuber, 1997). These memory cells are then arranged into blocks where they interact with other connected memory cells, as well as multiplicative units, which are provided by the read, write, and forget gates governed by the algorithm for the entire block (Greff et al., 2017).

2.2.2 Bidirectional LSTM (BiLSTM)

A common alteration that is made to the LSTM model is to run it as a bidirectional model. The basic idea of the bidirectional LSTM is that it can run its training sequence both forwards and backwards, then each learning method is connected to the same output layer (Graves and Schmidhuber, 2005). This allows the algorithm to identify patterns that occur in either direction, not only does this generally increase algorithm performance, but it is also essential to some pattern recognitions, like voice recognition.

2.2.3 BiLSTM with an Attention Layer

Another common method for improving the LSTM model is to implement an attention layer. The attention layer is capable of storing inputs and outputs from the hidden layer in order to help direct weights given to stored data (Bahdanau et al., 2014). This helps to increase accuracy even further, especially in larger data sets. One of the issues with the LSTM and all recurrent networks is that it still must process all of the information sequentially. This is not a limitation shared by the attention layer, which processes all input information before determining weights. By using the weights determined by the attention layer the LSTM has less error when

determining the weight of information while processing. In smaller data sets this improvement can be negligible, however the importance of attention for large data sets can not be understated (Bahdanau et al., 2014).

2.3 Model Hyperparameters

All the machine learning algorithms share nearly identical hyperparameters. This includes the LSTM model, BiLSTM model, and BiLSTM model with an attention layer. Each of the models was compiled using the mean squared error as the loss, and the adam optimizer. The model parameters are set to 30 epochs and a batch size of 32. The only exception to these parameters is the Bi-LSTM with an Attention layer model, which had its number of epochs reduced to 25 in order to save time while processing. All hyperparameters for the models are provided on the table below.

LSTM Hyperparameter	**Value**
Loss	'mean_squred_error'
optimizer	'adam'
epochs	30
Batch size	32
Epochs for Bi-LSTM with an attention layer	25

Table 1. The hyperparameters for the LSTM models.

3 Data

3.1 Data Set & Preprocessing

In order to create the data sets used in this experiment, earthquake data was collected from the US Geological Survey (USGS) earthquake database and the Kansas Geological Survey (KGS) earthquake database. During preprocessing the data from the two databases was combined, all duplicate instances were removed, and then data was split into two sets based on location. This resulted with a Puerto Rico data set containing 1436 earthquakes, and a Kansas data set containing 908 earthquakes. Both data sets contain only earthquakes with a magnitude of 2+ and span from January 2018 to December 2019.

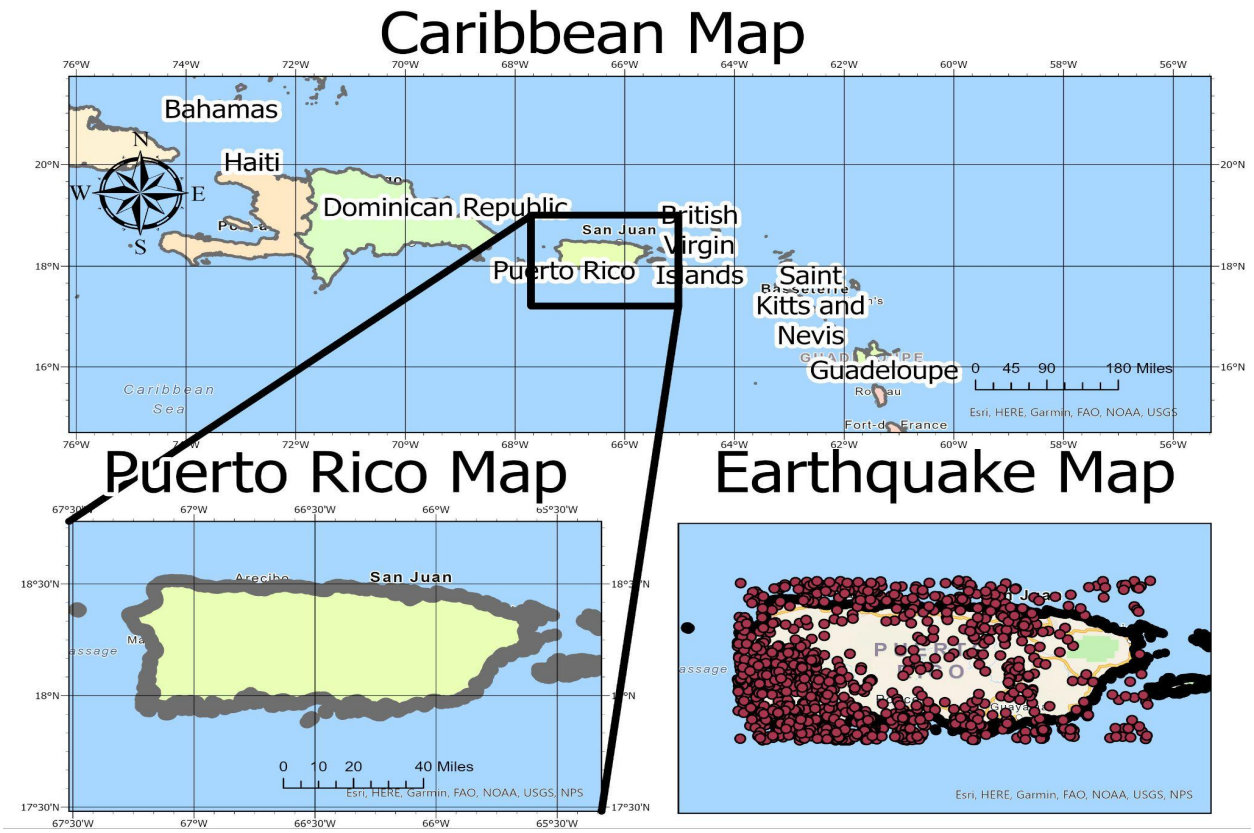

Figure 1. Distribution of earthquake epicenters (red circles) within the Puerto Rico data set.

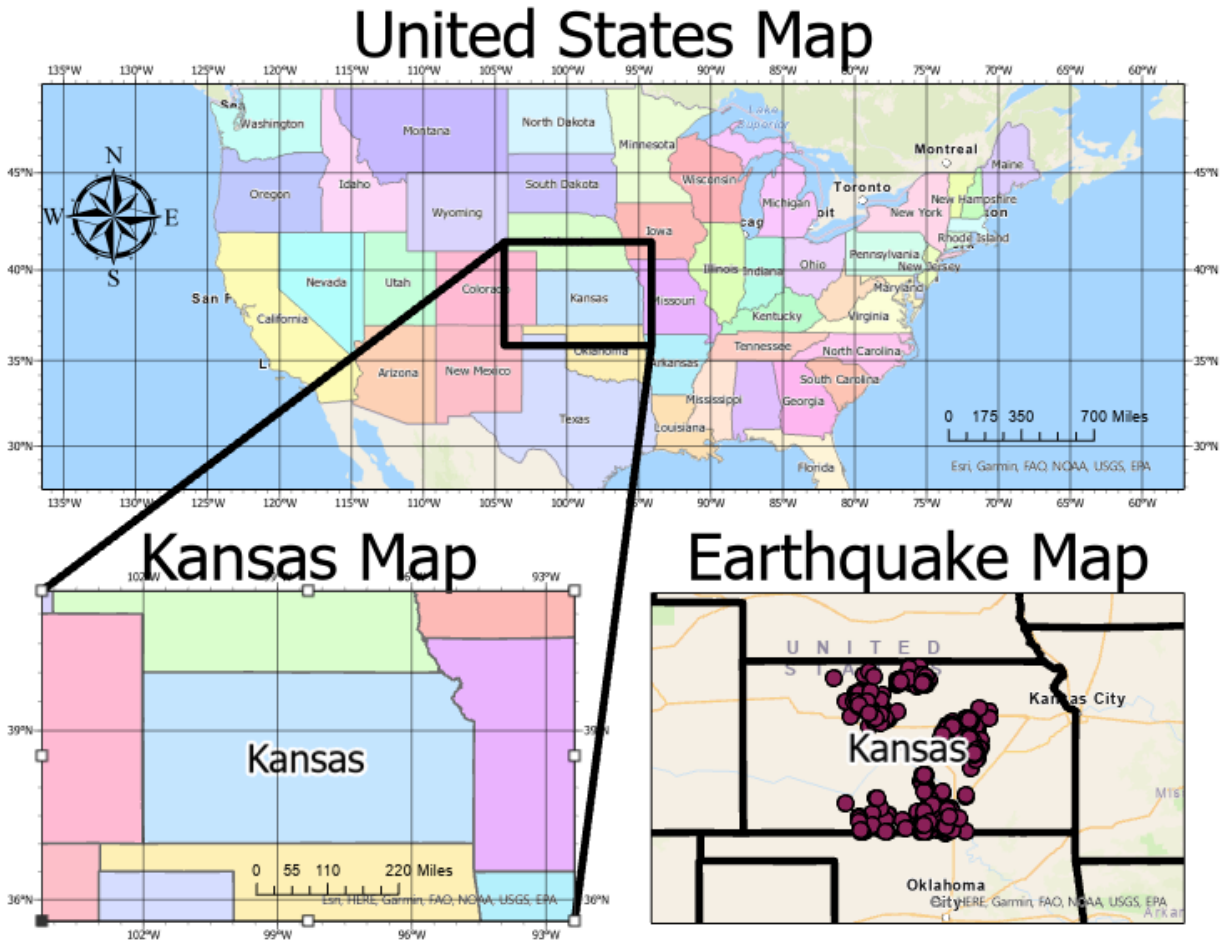

Figure 2. Distribution of earthquake epicenters (red circles) within the Kansas data set.

The resulting data sets include the following attributes associated with each earthquake: time, latitude, longitude, and magnitude. These four variables are also what the machine learning algorithms will be using to train their models, and what they will be ultimately attempting to predict. Data preprocessing included data merging, sampling, determining if there were any missing values, making a statistical summary, and data exploration.

	time	lat	long	mag
count	1436	1436	1436	1436
mean	3/2/2019 15:54:35	18.12247987	-66.72783294	2.348461003
std	228.9885707	0.2189891398	0.4076399862	0.3260447147
min	1/3/2018 2:36:40	17.8805	-67.2743	2
25%	9/30/2018 5:46:25	17.97575715	-67.00095173	2.130011044
50%	2/17/2019 6:40:38	18.0774	-66.8362	2.26
75%	8/3/2019 2:02:46	18.2692026	-66.45471415	2.566910962
max	12/31/2019 23:54:01	18.595	-65.5701	5

Table 2. The statistical Summary of the cleaned Puerto Rico data set.

	time	lat	long	mag
count	908	908	908	908
mean	12/30/2018 8:32:48	38.18453439	-98.04513143	2.365969151
std	208.3535793	0.9261152324	0.7349208297	0.4256889572
min	1/1/2018 4:27:24	37.00058	-100.00507	2
25%	8/12/2018 18:13:16	37.56403719	-98.53752839	2.08075755
50%	12/17/2018 17:52:27	38.01521	-97.967775	2.2
75%	5/18/2019 22:52:20	38.8050316	-97.55273448	2.651180752
max	12/30/2019 20:42:30	39.91831	-96.5312	4.8000002

Table 3. The statistical Summary of the cleaned Kansas data set.

3.2 Data Exploration

The Python programming language was used for all data processing and analysis. Numpy, Pandas, and Matplot libraries were applied for data cleaning and exploration. The Spyder development environment was used for all programming tasks, including data cleaning, visualization, and processing. Both graphs below were produced using Google Sheets, in order to visualize the distributions.

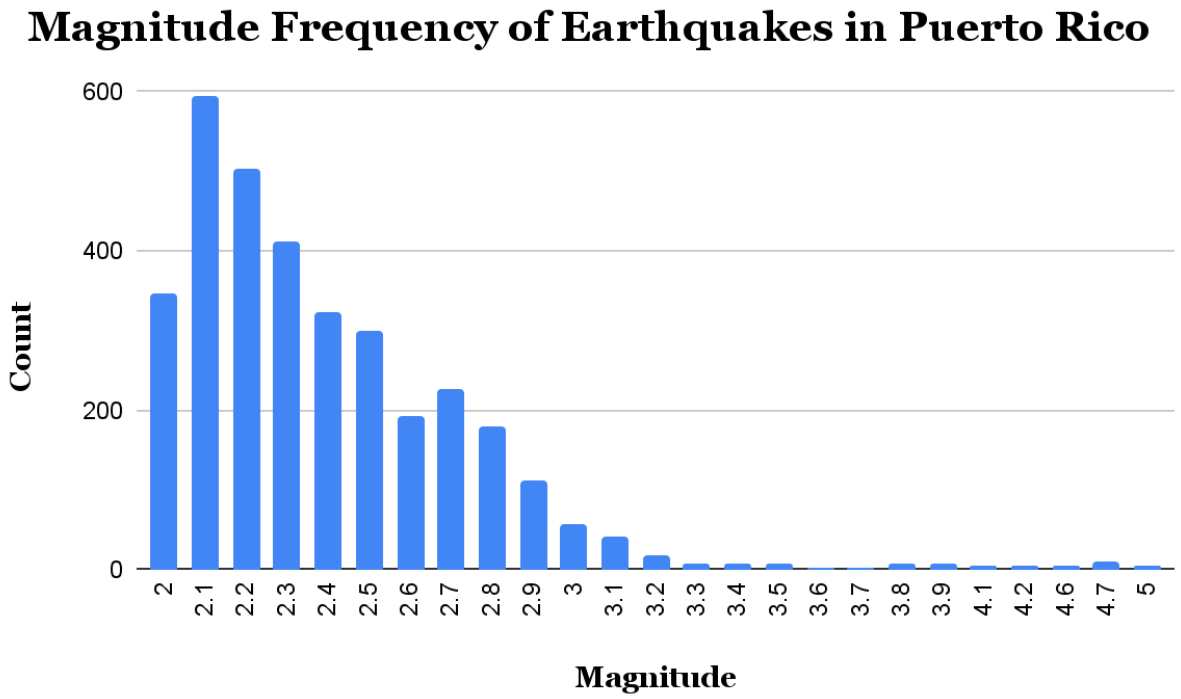

Figure 3. Frequency Plot of earthquake magnitude for Puerto Rico. For the years 2018-2019, only recording earthquakes with magnitude 2+.

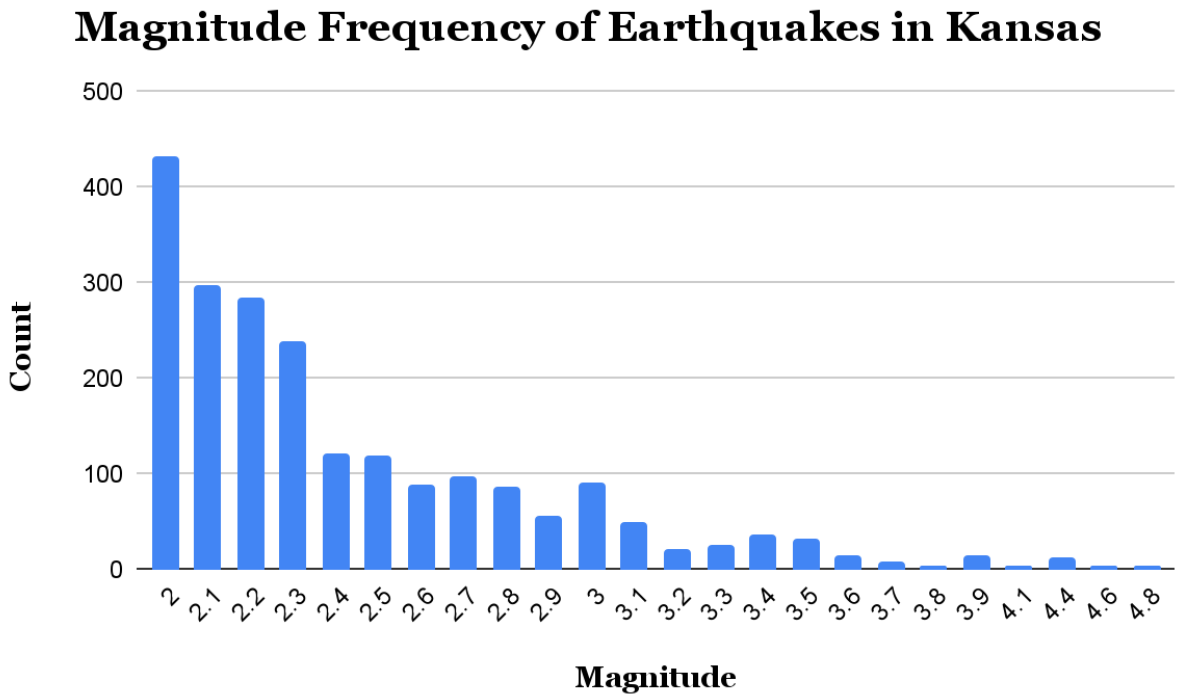

Figure 4. Frequency Plot of earthquake magnitude for Kansas. For the years 2018-2019, only recording earthquakes with magnitude 2+.

4 Results

4.1 Performance Measures

After completing data preprocessing, and modeling each of the four machine learning algorithms was run on the two data sets. After processing the Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) were calculated to evaluate the accuracy of each model.

4.1.1 Mean Absolute Error (MAE)

Mean Absolute Error is the measure of error between the calculated and expected results. It is commonly used to measure machine learning model performance, because its values are absolute, thus returning the same results independent of whether or not the model over or underestimates its predictions (Schneider and Xhafa, 2022). In general, MAE can be found using the following equation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - x\right|$$

Where:

n = the number of data points.

x_i = predicted value

x = true value

4.1.2 Mean Squared Error (MSE)

Mean Squared Error is another way to measure the error of a model. It assesses the average squared difference between the observed and predicted values. Squaring the differences serves several purposes, it eliminates negative values and ensures that the MSE is always greater than or equal to zero. Additionally, it increases the impact of larger errors, causing the MSE to rise substantially if there are large errors in the tested model (Frost, 2023). In general, MSE can be found using the following equation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(x - x_i\right)^2$$

Where:

n = the number of data points.

x_i = predicted value

x = true value

4.1.3 Root Mean Squared Error (RMSE)

Root Mean Squared Error is another way to measure the error of a model. It is the square root of the model's MSE, and is generally representative of the scatter of the observed values around the predicted values (Frost, 2023). RMSE is a commonly accepted statistic for model performance, and is simple enough to calculate that it is used by many with both little experience, and extensive experience with machine learning. However, there is a known weakness to using RMSE in that it is scale sensitive, this means that RMSE values can vary widely depending on the scale and number of values in a machine learning model. This means that unlike the other performance measures used in this experiment the RMSE may be a less reliable way to compare this model against others. In general, RMSE can be found using the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x - x_i\right)^2}$$

Where:

n = the number of data points.

x_i = predicted value

x = true value

4.1.4 Mean Absolute Percentage Error (MAPE)

Mean Average Percentage Error is another way to measure the error of a model. It measures accuracy as a percentage, and can be calculated as the average absolute percent error for each instance. MAPE is the most common measure used in forecasting analysis, likely because its units are scaled to a percentage, which makes it very easy to understand (Glen, 2023). MAPE works its best when there are no extremes or outliers in the data set, since large fluctuations can skew the statistic (Glen, 2023). In general, MAPE can be found using the following equation:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x-x_i}{x}\right|$$

Where:

n = the number of data points.

x_i = predicted value

x = true value

4.2 Model Performance

As each model was run the graph of its training loss and validation loss were recorded. The training loss indicates how well the model is fitting the training data, while the validation loss indicates how well the model fits new data (Abebe et al., 2023). When the training loss is greater than the validation loss the model is overfitting data, meaning that it is picking up too much noise to make accurate predictions. However, when the validation loss is greater than the training loss the model is underfitting data, meaning that it does not have enough information to make accurate predictions. When running a machine learning model the training loss should be as close to the validation loss as possible in order to represent an optimized model (Abebe et al., 2023). Each machine learning model was run 10 times, and then the average MAE, MSE, RMSE, and MAPE scores were recorded, and have been provided alongside the graph of the training and validation loss.

4.2.1 LSTM Performance

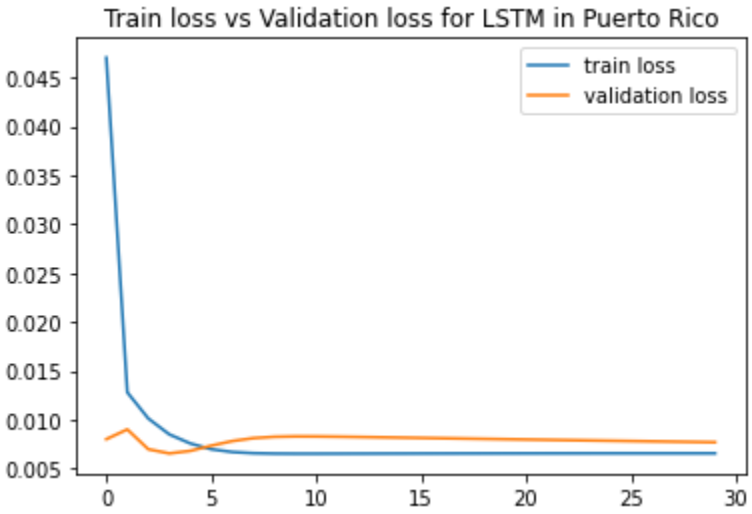

Figure 5. The training loss vs validation loss graph for the LSTM model when used on the Puerto Rico data set.

The graph of training loss versus validation loss for Puerto Rico shows that the LSTM model fits the data set well. There is an initial amount of overfitting, however once excess noise has been eliminated by the fifth epoch, the training loss and the validation loss stabilize within very close proximity to one another. After fitting the model displayed a MAE of 0.066, MSE of 0.008, RMSE of 0.088, and MAPE of 17.03%.

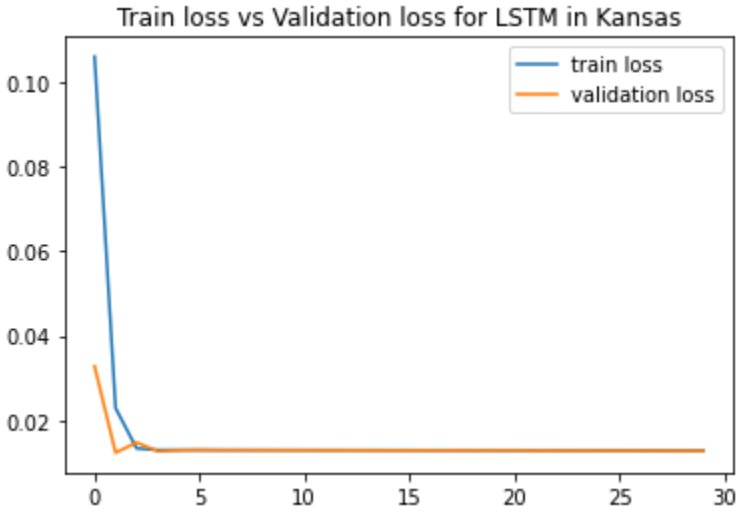

Figure 6. The training loss vs validation loss graph for the LSTM model when used on the Kansas data set.

The graph of training loss versus validation loss for Kansas shows that the LSTM model fits the data set well. There is an initial amount of overfitting, however once excess noise has

been eliminated by the fifth epoch, the training loss and the validation loss stabilize within very close proximity to one another. After fitting the model displayed a MAE of 0.096, MSE of 0.013, RMSE of 0.114, and MAPE of 29.19%.

4.2.2 BiLSTM Performance

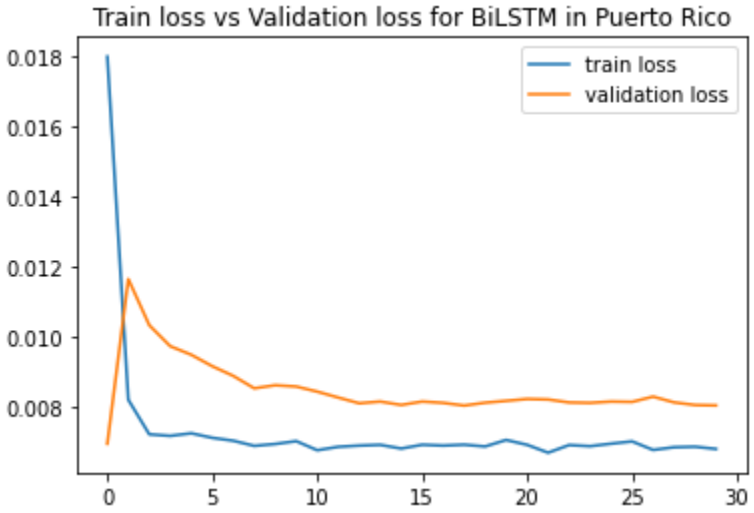

Figure 7. The training loss vs validation loss graph for the BiLSTM model when used on the Puerto Rico data set.

The graph of training loss versus validation loss for Puerto Rico shows that the BiLSTM model fits the data set well. There is an initial amount of overfitting, however once excess noise has been eliminated by the second epoch, the training loss and the validation loss stabilize within close proximity to one another, with a slight amount of underfitting. After fitting the model displayed a MAE of 0.067, MSE of 0.008, RMSE of 0.090, and MAPE of 17.40%.

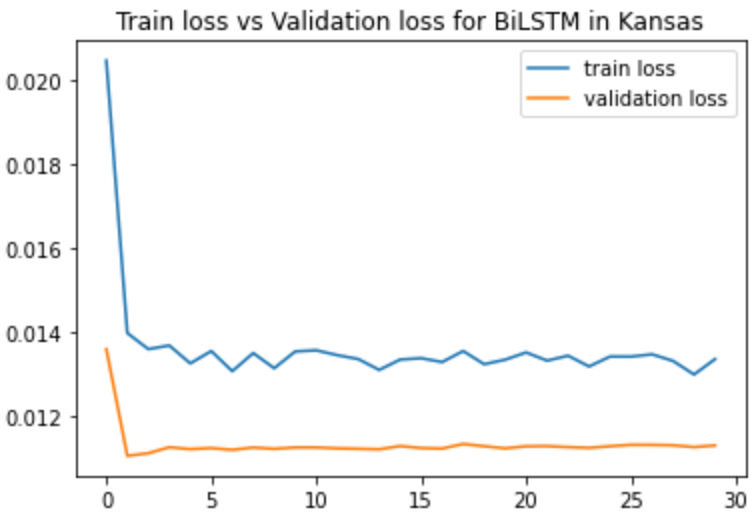

Figure 8. The training loss vs validation loss graph for the BiLSTM model when used on the Kansas data set.

The graph of training loss versus validation loss for Kansas shows that the BiLSTM model fits the data set well. There is an initial amount of overfitting, however once excess noise has been eliminated by the second epoch, the training loss and the validation loss stabilize within close proximity to one another, with a slight amount of overfitting. After fitting the model displayed a MAE of 0.081, MSE of 0.011, RMSE of 0.106, and MAPE of 23.39%.

4.2.3 BiLSTM with an attention layer Performance

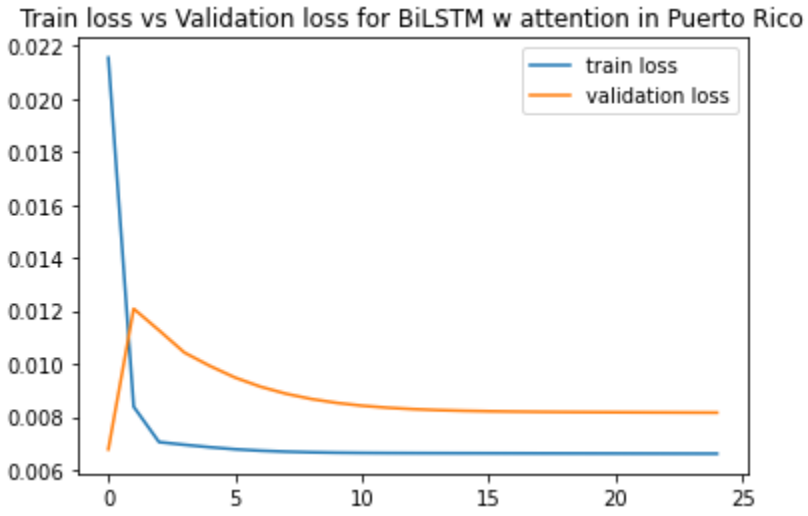

Figure 9. The training loss vs validation loss graph for the BiLSTM with an attention layer model when used on the Puerto Rico data set.

The graph of training loss versus validation loss for Puerto Rico shows that the BiLSTM with an attention layer model fits the data set well. There is an initial amount of overfitting, however once excess noise has been eliminated by the third epoch, the training loss and the validation loss stabilize within close proximity to one another, with a slight amount of underfitting. After fitting the model displayed a MAE of 0.068, MSE of 0.008, RMSE of 0.090, and MAPE of 17.40%.

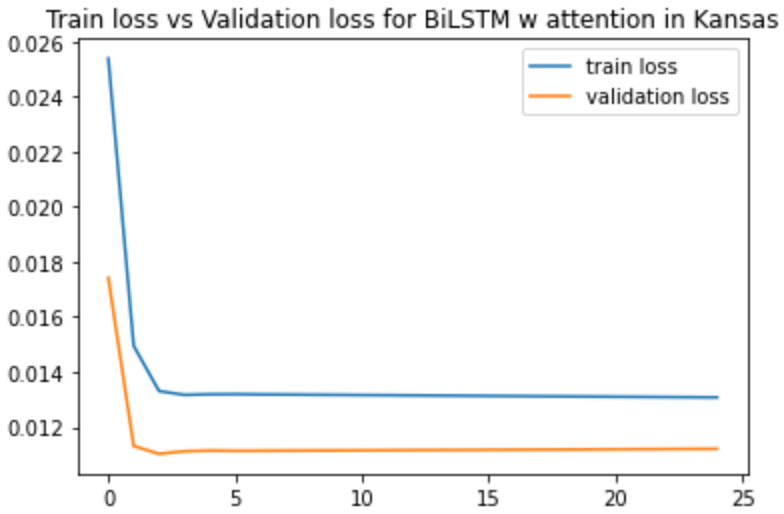

Figure 10. The training loss vs validation loss graph for the BiLSTM with an attention layer model when used on the Kansas data set.

The graph of training loss versus validation loss for Kansas shows that the BiLSTM with an attention layer model fits the data set well. There is an initial amount of overfitting, however once excess noise has been eliminated by the third epoch, the training loss and the validation loss stabilize within close proximity to one another, with a slight amount of overfitting. After fitting the model displayed a MAE of 0.080, MSE of 0.011, RMSE of 0.106, and MAPE of 22.98%.

4.3 Discussion

The results provided by this experiment help to prove that while machine learning is making earthquake prediction feasible, there is still much about the process that needs further understanding. This experiment was only able to cover two data sets and three processing models, so the room for further investigation into the realm of machine learning for earthquake prediction remains vast. Once the data sets were fully processed their performance measures were collected and displayed on the table and figures below.

Model	LSTM	LSTM	BiLSTM	BiLSTM	BiLSTM w attention	BiLSTM w attention	Best Model and Data Set
Data Set	Puerto Rico	Kansas	Puerto Rico	Kansas	Puerto Rico	Kansas	
MAE	0.066	0.096	0.067	0.081	0.068	0.080	**LSTM (Puerto Rico)**
MSE	0.008	0.013	0.008	0.011	0.008	0.011	**LSTM (Puerto Rico)**
RMSE	0.088	0.114	0.090	0.106	0.090	0.106	**LSTM (Puerto Rico)**
MAPE	17.03%	29.19%	17.40%	23.39%	17.40%	22.98%	**LSTM (Puerto Rico)**

Table 4. Comparison of performance measures across machine learning models and data sets

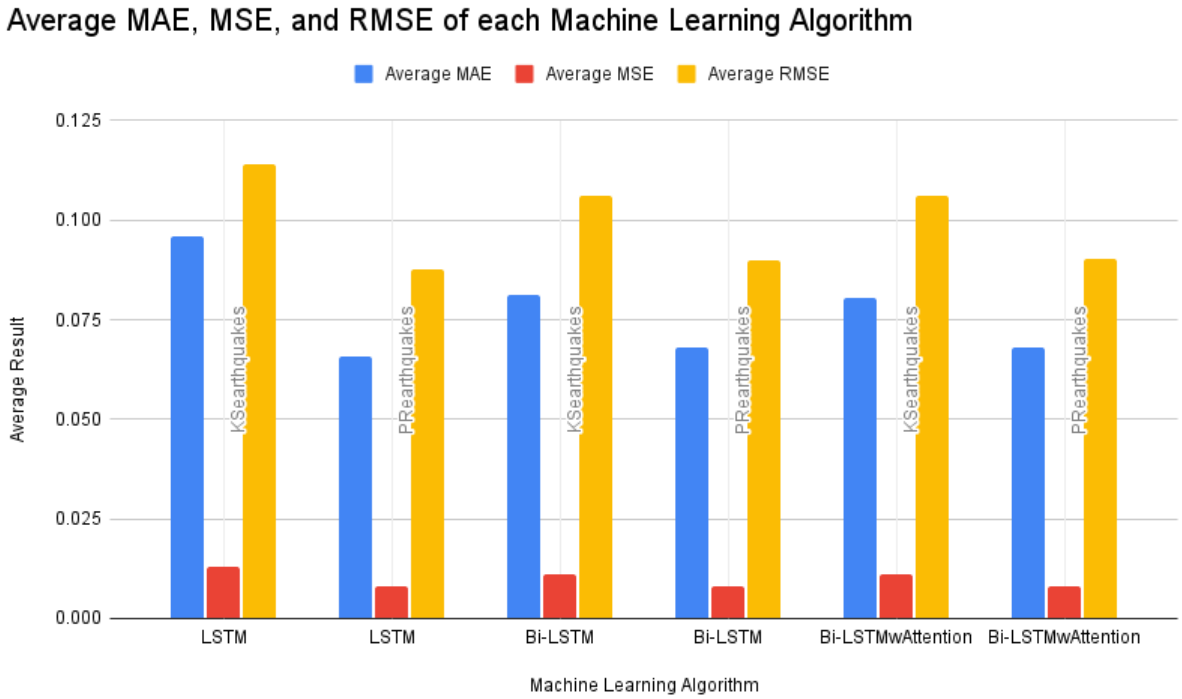

Figure 11. The average MAE, MSE, and RMSE between both data sets and all four machine learning models.

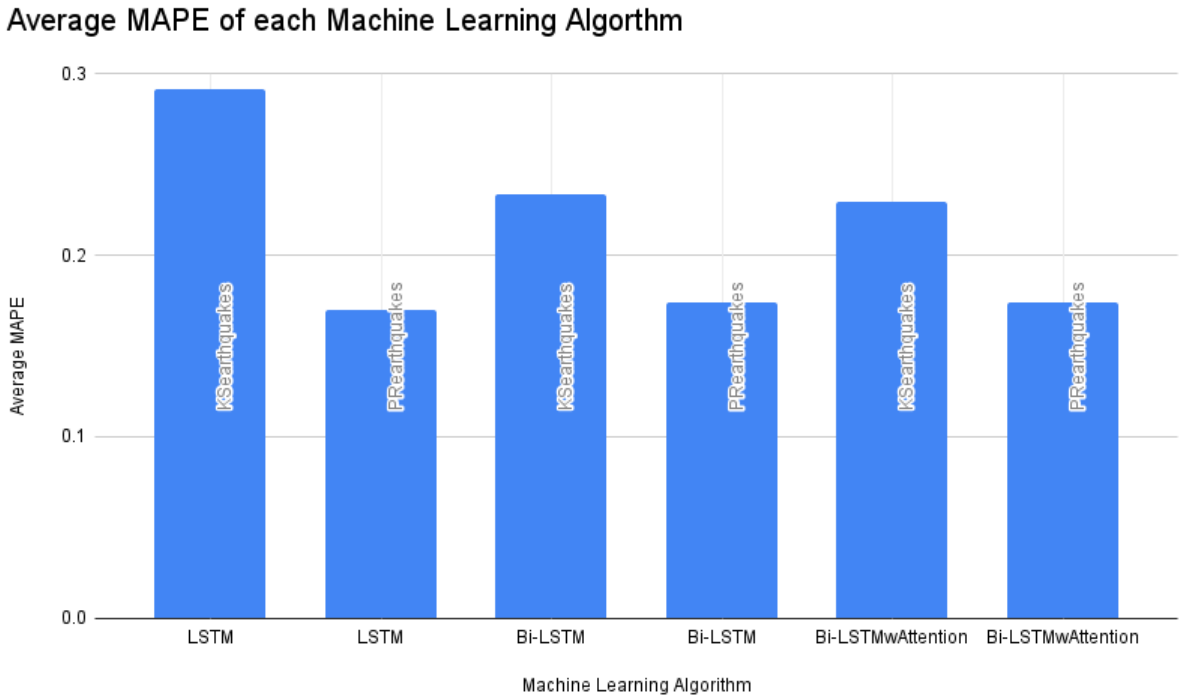

Figure 12. The average MAPE between both data sets and all four machine learning models.

As displayed on table 3 and on figures 11 and 12 the Puerto Rico data set combined with the normal LSTM model produced the lowest error, and thus the best results, with a MAE of 0.066, MSE of 0.008, RMSE of 0.088, and MAPE of 17.03%. The Kansas data set performed the best with the BiLSTM with an attention layer model, resulting in an MAPE of 17.40%. This is likely due to a difference in the data sets. Meaning, this experiment does succeed in establishing some connection between location, tectonic setting, and machine learning performance on earthquake data sets. When comparing the Puerto Rico data set with the Kansas data set some interesting differences emerge. Firstly, not only does the Puerto Rico data set perform the best of the two, but it performs at its best when used along with the least complex of the machine learning models, the LSTM. The data set performs similarly, but marginally worse on the other two LSTM based models. Each new model introduces another level of complexity to the traditional LSTM model, by either making it a bidirectional model, or introducing an attention layer. This could also be related to the fact that the Puerto Rico data set continually underfit its predictions, since the model determined too much of the data set as noise, it was unable to make use of the tools provided by the more complex models. Looking back at Figure 1. The map of Puerto Rico, and its associated earthquake data set, it can be seen how widely spread the earthquakes are with only minor amounts of clustering where major faults are the most active. This distribution of earthquakes is a direct result of the island's tectonic setting, where complex fault networks are regularly stressed by nearby plate boundaries. In this way the tectonic setting of Puerto Rico could be the reason as to why simple machine learning models work better there than complex ones. Conversely, the Kansas data set showed the most success with the BiLSTM with an attention layer model. Unlike Puerto Rico the Kansas data set showed a decrease in error

each time that the LSTM model increased in complexity. Another difference between data sets is that the Kansas data set regularly overfit its estimations, rather than underfit them, meaning that it was learning too well and began to consider some of the noise in the data set as part of the pattern. Many of the newest and more complex varieties of the LSTM, like the BiLSTM or the attention layer are specifically designed to combat overfitting errors (Graves and Schmidhuber, 2005). The reason why the Kansas data set is overfit rather than underfit is likely due to the distribution of events, and tectonic setting. Since it is located in the center of a continental plate Kansas has a far less active tectonic setting, and only one major fault. This means that almost all of the events, as they are displayed on Figure 2. are grouped together in one of three to four clusters around the state. This makes the task of predicting earthquakes easier on the machine, since it can identify the clusters and can predict earthquakes within the clusters with reasonable assuredness. The Puerto Rico data set ultimately still performed better, however that is more likely due to the simple fact that it had the largest data set.

5 Conclusions

With the use of machine learning to predict earthquakes a growing possibility, it is valuable to take a step back and look at the methods behind the predictions, and how scientists in the future can cater the machine learning models that they use towards the region that they survey. The results of this experiment show that in regions where earthquake clustering is not present, or only weakly present, such as in complexly faulted locations, or locations near multiple plate boundaries that simpler machine learning models such as the LSTM work best. Whereas, in locations with earthquake clustering is present, such as in locations with no plate boundaries or little tectonic activity complex machine learning models work better, like the BiLSTM with an attention layer model. This is due to how the distribution of earthquakes affects machine learning, thus indicating a correlation between the tectonic setting and the performance of machine learning models in a region. Given the error results produced by this experiment it is possible that earthquakes will be predicted by machine learning programs in the near future, however even the most accurate of results still show a reasonable degree of error, meaning that there are further improvements that can still be made in the field.

Acknowledgments

This study is supported by the Disaster Resilience Analytics Center of Wichita State University.

Eldon Taskinen and Zelalem Demissie are supported by the Geology Department at Wichita

State University. The authors thank the editor and two anonymous reviewers for their

constructive reviews that helped improve the manuscript.

Data Availability Statement

Datasets for this research are available upon request.

References

Abebe, E., Kebede, H., Kevin, M., Demissie, Z., (2023), Earthquakes magnitude prediction using deep learning for the Horn of Africa, *Soil Dynamics and Earthquake Engineering*, 170, 1-44. doi:10.1016/j.soildyn.2023.107913

Alaskar, H., and Saba, T., (2021), Machine Learning and Deep Learning : A Comparative Review Machine Learning and Deep Learning : A Comparative Review doi:10.1007/978-981-33-6307-6.

Baars, D. L., Watney, W. L., Steeples, D. W., Brostuen, E. A., (1989), Petroleum: a primer for Kansas, *Kansas Geological Survey*, Education, URL:HTTP://www.kgs.ku.edu/Publications/Oil/index.html

Bahdanau, D., Cho, K., Bengio, Y., (2014), Neural Machine Translation by Jointly Learning to Align and Translate, *ICLR Computation and Language*. doi:10.48550/arXiv.1409.0473

Bakun, W. H., McEvilly, T. V., (1984), Recurrence models and Parkfield, California, earthquakes, *Solid Earth*, 89(B5), 3051-3058. doi:10.1029/JB089iB05p03051

Bellamkonda, S., Settipalli, L., Vedantham, R., and Vemula, M. K., (2021), An Enhanced Earthquake Prediction Model Using Long Short-Term Memory, *Turkish Journal of Computer and Mathematics Education*, 12, 2397–2403.

Benford, B., DeMets, C., Calais, E., (2012), GPS estimates of microplate motions, northern Caribbean: evidence for a Hispaniola microplate and implications for earthquake hazard, *Geophysical Journal International*, 191(2), 481-490. doi:10.1111/j.1365-246X.2012.05662.x

Bengio, Y., Simard, P., Frasconi, P., (1994), Learning long-term dependencies with gradient descent is difficult, *IEEE Transactions on Neural Networks*, 5(2), 157-166. doi:10.1109/72.279181

Berendsen, P., (1997), Tectonic evolution of the Midcontinent Rift System in Kansas, Middle Proterozoic to Cambrian Rifting, Central North America, *Geologic Society of America*, 235-241. doi:10.1130/0-8137-2312-4.235

Berhich, A., Belouadha, F., and Kabbaj, M. I., (2020), LSTM-based Models for Earthquake Prediction, *Association for Computing Machinery*, doi:10.1145/3386723.3387865.

Bird, P., Kagan, Y.Y., (2004), Plate-Tectonic Analysis of Shallow Seismicity: Apparent Boundary Width, Beta, Corner Magnitude, Coupled Lithosphere Thickness, and Coupling in Seven Tectonic Settings, *Bulletin of the Seismological Society of America*, 94(6), 2380-2399, doi:10.1785/0120030107

Brosius, L., Steeples, D.W., (2014), Earthquakes, *Kansas Geological Survey*, Publications, URL:https://www.kgs.ku.edu/Publications/pic3/pic3_1.html

Buchanan, R., (2000), Earthquake Risk Low, but Real in Some Parts of State, Kansas Geological Survey, News Release, URL:http://www.kgs.ku.edu/General/News/2000/earthquake.html

Frost, J., Mean Squared Error (MSE), Statistics by Jim, URL:https://statisticsbyjim.com/regression/mean-squared-error-mse/ , Accessed: 9/25/2023

Frost, J., Root Mean Squared Error (RMSE), Statistics by Jim, URL:https://statisticsbyjim.com/regression/root-mean-square-error-rmse/ , Accessed: 9/25/2023

Geller, R. J., Jackson, D. D., Kagan, Y. Y., Mulargia, F., (1997), Earthquakes Cannot Be Predicted, *Science*, 275(5306) 1616. doi:10.1126/science.275.5306.1616

Geography, US Census Bureau. "State Area Measurements and Internal Point Coordinates". Archived from the original on March 16, 2018. Retrieved May 31, 2016.

Glen, S., Mean Absolute Percentage Error (MAPE) StatisticsHowTo: Elementary Statistics for the rest of us! Accessed: 9/25/2023

URL:https://www.statisticshowto.com/mean-absolute-percentage-error-mape/

Graves, A., Schmidhuber, J., (2005), Framewise phoneme classification with bidirectional LSTM networks, *IEEE International Joint Conference on Neural Networks*. doi:10.1109/IJCNN.2005.1556215

Graves, A., Liwicki, M., Fernandez, S., Bertolami, R., Bunke, H., Schmidhuber, J., (2009), A Novel Connectionist System for Unconstrained Handwriting Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5), 855-868. doi:10.1109/TPAMI.2008.137

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., Schmidhuber, J., (2017), LSTM: A Search Space Odyssey, *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222-2232. doi:10.1109/TNNLS.2016.2582924

Holzer, T. L., Savage, J. C., (2013), Global Earthquake Fatalities and Population, *Earthquake Specra*, 29(1), 155-175. doi:101193/1.4000106

Jackson, D. D., Kagan, Y. Y., (2006), The 2004 Parkfield Earthquake, the 1985 Prediction, and Characteristic Earthquakes: Lessons for the Future, *Bulletin of the Seismological Society of America*, 96(4B), S397-S409. doi:10.1785/0120050821

Jansma, P. E., Mattioli, G. S., (2005), GPS Results from Puerto Rico and the Virgin Island: Constraints on Tectonic Setting and Rates of Active Faulting, Active Tectonics and Seismic Hazards of Puerto Rico, the Virgin Islands, and Offshore Areas, *Geological Society of America*, 13-25. doi:10.1130/0-8137-2385-X.13

Karevan, Z., Suykens, J. A., (2020), Transductive LSTM for time-series prediction: An application to weather forecasting, *Neural Networks*, 125, 1-9. doi:10.1016/j.neunet.2019.12.030

Mann, P., Prentice, C. S., Hippolyte, J. C., Grindlay, N. R., Abrams, L. J., Lao-Davila, D., (2005), Reconnaissance study of Late Quaternary faulting along Cerro Goden fault zone, western Puerto Rico, *GSA Special Papers: Active Tectonics and Seismic Hazards of Puerto Rico, the Virgin Islands, and Offshore Areas*, 385, 115-138. doi:10.1130/0-8137-2385-X.115

Merriam, D. F., (1963), The Geologic History of Kansas, Kansas Geological Survey Bulletin 162, URL:http://www.kgs.ku.edu/Publications/Bulletins/162/index.html

Peterie, S. L., Miller, R. D., Intfen, J. W., Gonzales, J. B., (2018), Earthquakes in Kansas Induced by Extremely Far-Field Pressure Diffusion. *Geophysical Research Letters*, 45(3), 1395-1401. doi:10.1002/2017GL076334

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., Johnson, P. A., (2017), Machine Learning Predicts Laboratory Earthquakes, *Geophysical Research Letters,* 44(18), 9276-9282. doi:10.1002/2017GL074677

Schneider, P., Xhafa, F., (2022), Anomaly Detection and Complex Event Processing over IoT Data Streams with Application to eHealth and Patient Data Monitoring, *Chapter 3- Anomaly Detection: Concepts and methods*, 49-66. ISBN:9780128238189

Veizer, J., Mackenzie, F. T., (2014), 9.15 - Evolution of Sedimentary Rocks, Treatise on Geochemistry, *Reference Module in Earth Systems and Environmental Science*, 9, 399-435. doi: 10.1016/B0-08-043751-6/07103-6

Viltres, R., Nobile, A., Vasyura-Bathke, H., Trippanera, D., Xu, W., Jonsson, S., (2021), Transtensional Rupture within a Diffuse Plate Boundary Zone during the 2020 Mw 6.4 Puerto Rico Earthquake, *Seismology Research Letters*, 93, 567-583. doi:10.1785/0220210261

Wood, H. O., Gutenberg, B., (1935), Earthquake Prediction, *Science*, 82(2123), 219-220. doi:10.1126/science.82.2123.219

Wyss, M., (1997), Cannot Earthquakes Be Predicted?, *Science*, 278(5337) 487-490. doi:10.1126/science.278.5337.487

Wyss, M., (2001), Why is earthquake prediction research not progressing faster?, *Tectonophysics*, 338(3-4) 217-223. doi:10.1016/S0040-1951(01)00077-4

Yadav, A., Jha, C. K., Sharan, A., (2020), Optimizing LSTM for time series prediction in Indian stock market, *Procedia Computer Science,* 167, 2091-2100 doi:10.1016/j.procs.2020.03.257