2020 International Conference on Computational Science and Computational Intelligence (CSCI) | 978-1-7281-7624-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/CSCI51800.2020.00279

# *Bilateral Trade Flow Prediction Models Enhanced By Wavelet and Machine Learning Algorithms*

Evdokia Maria Kottou
Farmington High School
Farmington, CT, USA
22kottouev@fpsct.org

Tyler Andrew Grubelich
Farmington High School
Farmington, CT, USA
22grubelichty@fpsct.org

Dr. Xiaodi Wang
Western Connecticut State University
Bethel, CT, USA
wangx@wcsu.edu

***Abstract*—International economy promotes a sense of global interdependence and offers mutual benefits to countries around the world through the system of imports and exports. In this study, we have utilized economic indicators as inputs in an algorithm scheme that is based on Machine Learning methods combined with Wavelet Transforms to predict the bilateral trade flow between pairs of trading countries. Utilizing this methodology will allow countries to strike more successful trade deals, and increase a nation's estimated Gross Domestic Product (GDP) per capita. This will ultimately help reduce poverty as governments reallocate funds to serve struggling populations and target the 2030 United Nations Agenda (UNA) [1].**

***Keywords-Wavelet Transforms; Covolutional Neural Networks; Decision Tree; Gravity Model***

## I. PURPOSE OF RESEARCH

Over 734 million people around the world live on an average income of less than $1.90 a day, a level of extreme poverty that plagues human society in a way that often seems irreparable [2]. With the current recession of the world's economy due to a pandemic, world leaders are looking to utilize trade in order to ignite economic growth in all nations.

Trade deals can be difficult to strike, and often trade representatives bargain with more than one country within a deal in order to achieve an optimal trade agreement. These interactions create global trading relationships and allow for the continuation of international trade for future generations [3]. In developing economies, increases in imports would allow for greater resources while also expanding the competitive market for domestic firms. An increase in exports would allow for a higher GDP, thus a higher GDP per capita, which works to diminish the poverty gap [3] [4].

Using Machine Learning to make better economic predictions may help all countries, the World Bank, and the UNA in making decisions to eliminate global poverty and hence fulfill the very ambitious goals of the 2030 UNA to eradicate global poverty.

## II. RESEARCH DATASETS

The data included in this project is from the Institute of Research on the International Economy (CEPII), World Bank Development Indicator database (World Bank 2020), and CIA World Factbook. Our data set contains both individual countries and country-pairs with an annual time-variant dimension and geographical dimension. To conduct this project, just over 2.7 million data points, between the years 1827 and 2014 were gathered, utilizing data from all possible country pair combinations [5] [6] [7].

## III. PROCESSING DATA

Prior to applying any Machine Learning algorithms, this data must be processed and cleansed, allowing for ease and clarity when implementing regression and other models. We first began by calculating values for the following variables (GDP/Population = GDP per capita, Roadway/Land Area = Roadway Density, Railway/Land Area = Railway Density, Airplane/Land Area = Airplane Density). After establishing the economic features that would be used in our data matrix (listed in Table I) we then reduced the number of observed countries to 112 by eliminating countries that failed to provide enough data, specifically in their trade flow (or response variable). We then assigned each country with a unique number.

TABLE I.

| | | |
|---|---|---|
| Roadway (km) | CIA Factbook | Total length of the roadway network |
| Railway (km) | CIA Factbook | Total length of the railway network |
| Airplane | CIA Factbook | Total number of airports |
| Land Area (km$^2$) | World Bank | Entire land area |
| Year | CEPII TRADHIST | Annual years from 1827 (or from the initiation year of the country) to 2014 |
| Flow (British Sterling Pounds) | CEPII TRADHIST | Net trade flow between country pairs |

978-1-7281-7624-6/20/$31.00 ©2020 IEEE
DOI 10.1109/CSCI51800.2020.00279

| | | |
|---|---|---|
| Nominal GDP (British Sterling Pounds) per capita | CEPII TRADHIST | GDP unadjusted for inflation divided by the annual population |
| Curcol | CEPII TRADHIST | Colonial relationship |
| Common Language | CEPII GeoDist | Whether or not a common language exists |
| Bilateral Distance (km) | CEPII GeoDist | City population-weighted mean of the great-circle distance between each pair of countries |

Once any missing data points were identified, a value of zero was added for any missing cells in flow, railway density, roadway density, and airplane density. A value of zero was added for any missing flow cell of a country as long as the following criteria were fulfilled: the country must have reported trade with more than 30% of all countries included in the dataset for that given year, and the country must be reporting trade with at least 10 different trading partners [5]. A value of zero with a density indicator displays the lack of technological advancements in that country in regard to transportation. Finally, for the time-variant data which was annually recorded, with the exception of flow, there were minimal empty cells. However, for the few missing cells, an average of the two years' data closest to that missing year data was taken and added into that cell. If there was data missing for multiple years, the year closest to these missing years was pasted into the empty cells.

## IV. WAVELET TRANSFORMS

There are many transformations that are used to analyze signals within data, with the most prominent being the Fourier Transform, which decomposes a signal (often a function of time) into its constituent frequencies. The Wavelet Transform has high resolution both in the frequency and time domains which allows for the identification of the time in which certain frequencies are present in a signal [8]. The use of this transformation is tested to determine if it will enable more accurate predictions by machine learning algorithms.

In order to develop a Wavelet Transform, a series of differently scaled waves - called Wavelets - which are time located, are used to extract information from data about its frequency and time. Using an integral called convolution, the Wavelets are shifted through the entire signal in order to find similarities between the different functions [8]. After determining these similarities, different scalars are then applied depending on the information that is necessary to obtain. The lengths of the Wavelets are determined by these scale-factors, where a higher scale factor creates a longer wavelength. Consequently, applying a high scale-factor results in a smaller frequency, thus providing better resolution in the frequency domain [8].

$$\alpha = \begin{bmatrix} 0.3383860978386 \\ 0.53083618701374 \\ 0.72328627674361 \\ 0.23896417190576 \\ 0.04651408217589 \\ -0.14593600755399 \end{bmatrix}$$

$$\beta = \begin{bmatrix} -0.11737701613483 \\ 0.54433105395181 \\ -0.01870574735313 \\ -0.69911956479289 \\ -0.13608276348796 \end{bmatrix}$$

$$\gamma = \begin{bmatrix} 0.40366386892892 \\ -0.62853936105471 \\ 0.46060475252131 \\ -0.40363686892892 \\ 0.07856742013185 \\ 0.24650202866523 \end{bmatrix}$$

**Figure I:** The filter bank for the 3-band Wavelet is shown.

For this research, a discrete 3-band Wavelet was used with a series of filter banks, due to improved decomposition of a signal into its components. Simply, filter banks allow for a signal's decomposition to remain within the corresponding frequency domains, while maintaining its time domain characteristics. The generated set of 3 filter banks (shown above in Figure I) consisted of $\alpha$ (the low pass filter bank), $\beta$ (the first high pass filter bank), and $\gamma$ (the second high pass filter bank) and were ensured to satisfy the conditions listed in the succeeding column.

$$\sum_{i=1}^{6} a_i = \sqrt{3} \tag{1}$$

$$\sum_{i=1}^{6} \beta_i = \sum_{i=1}^{6} \gamma_i = 0 \tag{2}$$

$$\|a\| = \|\beta\| = \|\gamma\| = 1 \tag{3}$$

$$a^*\beta = a^*\gamma = \beta^*\gamma = 0 \tag{4}$$

The application of the Wavelet Transform yields the coordinate of this signal under a wavelet basis that can be divided into 3 coefficient vectors. The first of these vectors is the approximation vector which is a low frequency component of the signal that forms the general trend of the data [2]. The second and third vectors, detail 1 and detail 2, are both the result of the signal passing through high pass filters that provide higher frequency trends of the data, with the latter being the highest frequency trends of the data. Furthermore, the tensor was also created using the approximation, detail 1, and detail 2 vectors. A tensor represents an array of functions of the data set, and is much more general than simply a vector.
When feeding Wavelet Transform data matrices through a variety of models, the natural logarithm of all data was calculated, with the exception of curcol and common language (as these two columns only have values of 0 and 1). The year was also removed when using the wavelet in order to prevent any inaccurate linear relationship being calculated by the model. This allowed the accuracy to greatly increase as the data was more smoothly distributed.

## V. GRAVITY MODEL

Many research efforts in the literature utilize the Gravity Model, seen as

$$T_{ij} = A\frac{Y_i Y_j}{D_{ij}},$$

in order to estimate the bilateral trade flow of two countries. By inputting data for trade flow ($T_{ij}$), GDP per capita ($Y_i$ and $Y_j$), and bilateral distance between countries ($D_{ij}$), the A value can be calculated and a general trend can be made about how the A value changes from year to year. Similar to Wavelet Transforms, the Gravity Model is applied to a machine learning algorithm to improve its prediction accuracy [9].

## VI. GAUSSIAN PROCESS REGRESSION

Gaussian Process Regression (GPR) utilizes Prior, Predictive, and Posterior distributions to find a probable function that "fits" the trend of a dataset. This method allows for predictions of data points using the Predictive Posterior Distribution:

$$f^* = \mu^* + K(X^*, X)[K(X,X) + \sigma_n^2 I]^{-1}(y\text{-}\mu)$$

Within GPR, we utilized two different Gaussian kernel-based subsets to train our dataset: The Exponential Kernel and Matern 5/2 Kernel. Both subsets calculated the covariance of the datasets. The Exponential Kernel used the equation:

$$k(x_i, x_j|\theta) = \sigma_f^2 \exp\left(-\frac{r}{\sigma_l}\right) \text{ where } r^2 = (x_i\text{-}x_j)^T(x_i\text{-}x_j)$$

and the Matern 5/2 Kernel used the equation:

$$k(x_i, x_j) = \sigma_f^2\left(1 + \frac{[\sqrt{5}r]}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right)e^{-\sqrt{5}r/\sigma_l}$$

to calculate covariance [10].

## VII. DECISION TREES

Decision Tree is a Machine Learning tool that approximates a sine curve from a dataset, where a deeper tree has greater accuracy. This tool breaks down a dataset into subsets, while simultaneously forming a regression model in the shape of a tree [11].

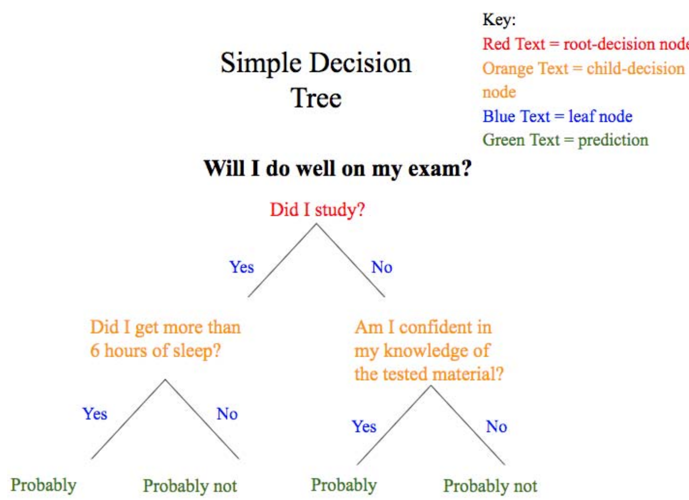


**Figure II:** A simple Decision Tree example is shown

After the Decision Tree model is trained by the dataset, it is split along the "best" feature. Figure II depicts a basic example of a Decision Tree with the "Did I study?" decision node. To determine the best feature, the machine simply assesses the target variable that is necessary to obtain, and determines which feature it is split along, based on a variety of methods. The method for choosing the best features within a numerical dataset is the Reduction in Variance method, using the equation: Variance = (Xi - )2Nwhere $X_i$ is the sample, $N$ is the number of samples, and μ is the mean. During each split, the variance of each child-decision node is calculated and the average of the child-decision node variances is the variance for the root-decision node. The variable with the most homogeneous node will be considered most effective, while the feature with the variance closest to 0 is the most homogeneous [11]. We utilized Fine Tree, which is a subset of Decision Tree and is a specific Machine Learning tool that emphasizes the importance of having many leaves within a tree, where it has a maximum of 100 leaves. We also used the Medium Tree, which has a maximum of 20 leaves, making it less likely to overfit a dataset. The final two subsets of Decision Tree which were tested were two Ensembles - Bagged Tree and Boosted Tree. Being an Ensemble simply means that these two models incorporate several Decision Trees in hopes of producing better results and accuracies.

## VIII. NUERAL NETWORKS

The multi-layered algorithmic set of Neural Networks was first modeled loosely after the human brain and allows programming systems to learn from observational data. This multi-layered algorithm first begins with the processing nodes or neurons which have been organized into layers, thus allowing information which has been entered as an input to be processed into this network [12].

The computation

$$Z = f(\text{weight, input, bias})$$

allows for a representation of the function of each node, where $f$ is called the activation function of this network. Once this data has been entered into the network, it begins to train the network by randomly assigning weights and threshold values - denoted by the variables W and T. As the data is fed into the input layer, weights and biases are used (weights are multiplied and bias values are added) as the data travels through the sequential layers before finally reaching the output layer. Then, the back-propagation gradient descent algorithm is applied to the network to find the optimal W according to the threshold values.

A subset of Feed-Forward Neural Networks known as Convolutional Neural Networks (CNN) were tested as a model in this research. CNNs work to avoid cases of overfitting by assembling complex patterns through small and simple patterns in the data. These complex patterns are automatically created as CNNs assume that all inputs are multi-dimensional tensors [12].

Additionally, these CNNs contain layers of neurons in all three dimensions: width, height and depth. In our case, the height is 11407, the width is 13, and since there is only 1 channel of the row vectors of the data the depth is 1 [12].

The CNN designed in this research consists of four two-dimensional convolutional layers. In the first layer of the CNN, 50 filters which have a kernel size of 3 are used while a Rectified Linear Unit (ReLU) activation function is engaged in this process as well. This activation function is a non-linear operation where any negative value is replaced with zero. Between the first convolution layer and the ReLU layer, a batch normalization layer is used to normalize the inputted data, ultimately speeding up the network's training and reducing its sensitivity. Next, the network proceeds to the succeeding convolutional layers which further reduce the dimensions of the output shape. The second convolutional layer uses 100 filters with a size of 3, following it are two identical convolutional layers with 200 filters and size of 3.

The pooling function was also used in our network in order to reduce the dimensionality of each convolved feature while still maintaining the most influential information by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. More specifically, two average pooling layers were implemented throughout the network. Finally, a Dropout layer was implemented in the training phase to randomly drop 20% of the data. This ultimately allows all neurons to evenly contribute to the model and eliminates any one neuron from dominating the system [12].

These CNNs are a subset of the Feed-Forward Network as output values are not entered back into the model, thus only interacting with other nodes one time. If entering these output values back into the model is desired, a Recurrent Neural Network (RNN) is used rather than the Feed-Forward Neural Network. Simply, a RNN cycles the information previously received through a loop, and uses this information along with its current input when making decisions.

## IX. LONG SHORT-TERM MEMORY

A subset of the RNN, the Long Short-Term Memory (LSTM) network is often used in deep learning to find long-term dependencies within a dataset. Within the LSTM network, there exists 3 gates: the Input Gate, the Output Gate, and the Forget Gate. These gates are used to keep and discard information [13]. The Input Gate is the location where new information or data is stored into the cell. Once inside of the cell state, the Forget Gate decides which information can be kept for training purposes, and which information is unnecessary and can be forgotten by the machine. This information is classified as important or unimportant based on a predetermined activation function in order to optimize the performance of the LSTM network. The Output Gate is the final gate used in the LSTM network, where it selects the most important information from the cell state in order to produce an output. Although the LSTM algorithm is commonly used for time-dependent data where the time cycles are within a matter of seconds, we nevertheless decided to test our annual trade data using this model.

## X. EXPERIMENTAL RESULTS

TABLE II. ($R^2$ values)

| MODEL | Original Data | Original Data + Gravity Model | Approximation Wavelet Transform | Approximation Wavelet Transform + Gravity Model | Detail 1 Wavelet Transform | Tensor Wavelet Transform |
|---|---|---|---|---|---|---|
| Fine Tree | 0.97 | 0.20 | 0.88 | 0.00 | 0.85 | 0.85 |
| Medium Tree | 0.88 | 0.33 | 0.84 | 0.01 | 0.81 | 0.81 |
| Matern 5/2 Gaussian | 0.95 | 0.39 | 0.93 | 0.02 | 0.40 | 0.47 |
| Exponential Gaussian | 0.92 | 0.38 | 0.93 | 0.018 | 0.40 | 0.56 |
| Boosted Tree 30 learners | 0.91 | 0.29 | 0.72 | 0.0117 | 0.71 | 0.56 |
| Boosted Tree 38 learners | 0.93 | 0.35 | 0.79 | 0.008 | 0.78 | 0.62 |
| Bagged Tree | 0.92 | 0.35 | 0.88 | 0.0033 | 0.85 | 0.87 |
| Long Short - Term Neural Network | 0.0017 | 0.0014 | 0.000678 | 0.00376 | 0.002670 | 0.0296 |
| Convolutional Neural Network | 0.8148 | 0.7936 | 0.8120 | 0.7903 | 0.753 | 0.8028 |

## XI. ANALYSIS OF RESULTS

We are able to draw various conclusions from the results of Decision Trees, Gaussian Processes, LSTM network, and CNNs using the original data (with and without the Gravity model), tensor matrix, approximation matrix (both with and without the Gravity model), and the detail 1 matrix.

The results that are yielded from Decision Tree indicate that all of its subsets are closely related in making predictions using individual countries. The most accurate of Decision Tree subsets, Fine Tree, generates an average $R^2$ value of 0.97 (Figure III). The least accurate of the 5 subsets, the Medium Tree model, is only slightly less accurate in its predictions with a $R^2$ value of 0.88 using the Original Individual Countries data.

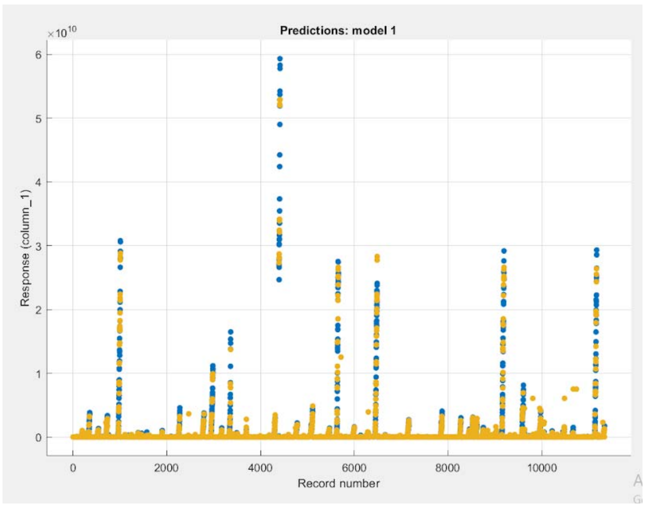


**Figure III:** A response plot of the individual country 61 is shown in the Fine Tree model, with an $R^2$ value of 0.97

When assessing the results of GPRs, the Matern 5/2 Gaussian and the Exponential Kernel Gaussian also generate predictions with similar $R^2$ outputs. Throughout all of the testing, the $R^2$ values of the Matern 5/2 Gaussian and the Exponential Kernel Gaussian are consistently within 0.09 of each other with all input datasets. Overall, the $R^2$ value of GPRs is very high, with the highest of these predictions being 0.95 with the Matern 5/2 Gaussian using the Original Individual Countries data (see Figure IV).

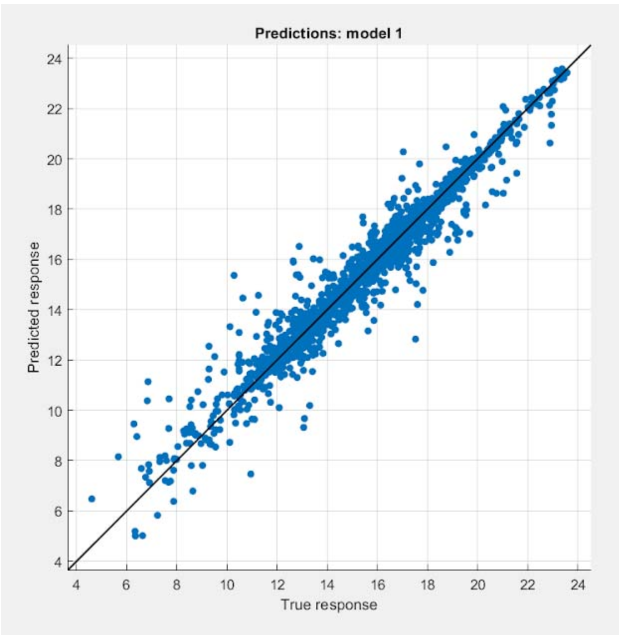


**Figure IV:** An $R^2$ value of 0.93 was reached when testing the Approximation Wavelet Transform of country 64 with the Matern 5/2 Kernel.

When testing the LSTM network, the accuracy of our predictions are significantly lower than Decision Trees and GPRs. Out of all of our tests, the most accurate prediction using the LSTM network is generated using the tensor of the individual countries dataset, but the $R^2$ value is only a mere 0.0296. The LSTM network does not generate accurate predictions with our time-dependent data, as our time changes by year, while the LSTM network favors data that cycles by a matter of seconds.

The results from the CNN are very consistent amongst all forms of data application and are quite accurate in its predictions overall. The range of $R^2$ values is less than 0.06 between every dataset, where the highest $R^2$ value, from the Original Country Pairs data, is 0.81. These high $R^2$ values can also be shown in regard to the RMSE decreasing, as seen below in Figure V.

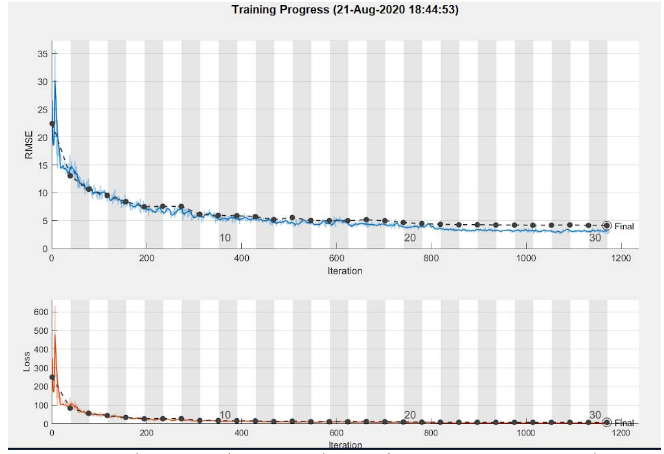


**Figure V:** The RMSE decreases when testing country 152 Tensor data matrix. A $R^2$ value of 0.80 was achieved.

However, the prediction accuracies made by the CNN are most likely lower compared to the predictions of Decision Tree algorithms because any subset of the Feed-Forward Neural Network typically thrives using unstructured data such as audio files or images. With the structure of our data being in a data table, CNN does not fully succeed in predicting the bilateral trade flow values using our structured dataset.

When applying the Wavelet Transforms to our data, the accuracy of the predictions ends up being lower than the original data without Wavelets. Although the decrease is minimal, the $R^2$ value of these predictions typically is lowered between 0.04 and 0.19 from using the Original Individual Countries dataset and the natural logarithm of the approximation matrix from the Wavelet Transforms. This decrease can be attributed to the fact that Wavelet Transforms are primarily used for more dynamic data - data that does not cyclically change - and since our data cyclically changes each year, the Wavelet Transforms do not provide greater accuracy.

When applying the Gravity model to the data, it can be seen that the machine performs significantly worse than without the use of it. The $R^2$ values decrease significantly when using the Gravity model with the Individual Country data as well as with the Approximation matrices of these countries. The only exception to this is when applying the Gravity model using the CNN, where the decrease in $R^2$ values from the Original Individual Countries data is only 0.02, with a shift of $R^2$ values from 0.81 to 0.79. This great disparity can largely be attributed to the lack of features that can be used within Tinbergen's Gravity model equation, where we had to omit the implementation of the features that model infrastructure density.

## XII. CONCLUDING REMARKS

The results achieved from this research allow us to conclude that testing multiple algorithms such as Decision Trees, Gaussian Process Regression, Long-Short Term Neural Network, and CNN lead to optimal bilateral trade flow predictions. With our applications of different Wavelet Transforms and the Gravity model, we were able to add on to previous literature work which had not experimented with these tools. Although the highest accuracies were received without the Wavelet Transform being applied, our research now provides more insight into which applications can contribute to the most accurate trade flow predictions (Wavelet Transforms should not be applied). By contributing better results than those presented in other literature, such as the $R^2$ value of 0.97 reached using the Fine Tree of the Original Data Matrix, we expect to aid the future of trade predictions with this research in hopes of ending global poverty.

As poverty continues to accelerate throughout the globe at an alarming rate, such measures of international trade between both developed and developing countries are necessary in aiding future generations. Today's strategic and determined world leaders who strive for growth in their countries could utilize the evolving technological field which allows for accurate measurements and precise predictions, in order to develop opportunities and connections for their countries which ultimately align with the 2030 UNA. Entering the annually recorded data of their country into our model would not only allow trade representatives to determine the trading partner which would allow for optimal growth, but the optimal amount of trade would also be generated. For instance, developing countries including Rwanda and Egypt can enter their data regarding roadway density, railway density, airplane density, nominal GDP per capita, year, colonial relationships, distance, past net trade values, and common languages and receive immediate trade flow predictions for their country and other countries around the world. Although this model is very successful in predicting trade flow values, it is not entirely complete. The model will be fully complete and more successful if all countries around the world input their own historical data pertaining to the features discussed above (roadway density, railway density, airplane density, nominal GDP per capita, year, colonial relationships, distance, past net trade values, and common languages). With this, all countries will receive a precise prediction of bilateral trade flow with every country that they trade with [1].

It is very necessary for trade representatives to consider our model when coming to a consensus and term of trade with their partner. The accurate predictions made by our model can aid trade representatives in determining an optimal amount of trade between the pair. This will ultimately benefit both nations by granting them the opportunity to distribute new resources fairly amongst their citizens, thus targeting the 2030 UNA by eradicating global poverty.

Allowing undeveloped countries an opportunity to grow by providing them an easily accessible and usable model of optimal trading partners and trade amounts, government officials could use this new income and increase in GDP to increase the GDP per capita, thus providing all citizens with a humane lifestyle. Our research conducted in this project is solely the commencement of the move to eradicate global poverty and provide all humans with a basis of living and pursuing a high standard of life.

## XIII. ACKNOWLEDGMENTS

Both partners, Evdokia Kottou (EK) and Tyler Grubelich (TG), contributed to this paper. Firstly, EK contributed the idea for this research, including its economic applications. TG explored a variety of Machine Learning techniques and observed those which would provide the best results through literature review. After, EK wrote codes such as the Wavelet Transform code, Convolution Neural Network, and Long Short-Term Network code using the MATLAB programming language. Following these codes, EK tested numerous data sets on different models. Both TG and EK contributed to the writing of this research paper. TG and EK are grateful that Western Connecticut State University provided them with an opportunity to take summer research classes.We would like to thank Dr. Xiaodi Wang for helping us throughout our research process.

## XIV. REFERENCES

[1] "Transforming Our World: the 2030 Agenda for Sustainable Development .:. Sustainable Development Knowledge Platform." *United Nations*, United Nations, 2020, sustainabledevelopment.un.org/post2015/transformingourworld/.

[2] "Overview." *World Bank*, 2020, www.worldbank.org/en/topic/poverty/overview.

[3] Reinsdorf, Marshall B. "Terms of Trade Effects: Theory and Measurement ." *Bureau of Economics*, Oct. 2009, www.bea.gov/system/files/papers/WP2009-1.pdf.

[4] "How Trade Works." *OECD*, 2020, www.oecd.org/trade/understanding-the-global-trading-system/how-trade-works.

[5] Fouquin, Michel, and Jules Hugot. "Two Centuries of Bilateral Trade and Gravity Data: 1827-2014." *CEPII*, 14 May 2016, www.cepii.fr//en/bdd_modele/presentation.asp?id=32.

[6]"World Development Indicators." *DataBank*, 2020, databank.worldbank.org/reports.aspx?source=2.

[7] "COUNTRY COMPARISON :: RAILWAYS." *Central Intelligence Agency*, Central Intelligence Agency, 2020,

[8] Dumitrescu, Catalin & Costea, Ilona & Nemtanu, Florin & Stan, Valentin & Gheorghiu, Andrei. (2014). Application of the Wavelet transform in machine-learning. University Politehnica of Bucharest Scientific Bulletin-Series a-Applied Mathematics and Physics. 76. 167-178.

[9] Tsamboulas, Dimitrios, et al. "The Gravity Model Specification for Modeling International Trade Flows and Free Trade Agreement Effects: A 10-Year Review of Empirical Studies ." *The Open Economics Journal*, Mar. 2010, benthamopen.com/contents/pdf/TOECONSJ/TOECONSJ-3-1.pdf.

[10] Perez-Cruz, Fernando, et al. Gaussian Processes for Nonlinear Signal Processing. July 2013, arxiv.org/pdf/1303.2823.pdf.

[11] Freund, Yoav, and Llew Mason. *The Alternating Decision Tree Algorithm*. cseweb.ucsd.edu/~yfreund/papers/atrees.pdf.

[12] Hardesty, Larry. "Explained: Neural Networks." *MIT News*, 14 Apr. 2017, news.mit.edu/2017/explained-neural-networks-deep-learning-0414

[13] Hochreiter, Sepp. *LONG SHORT-TERM MEMORY*. 1997, www.bioinf.jku.at/publications/older/2604.pdf