A Comparative Evaluation of Spatio Temporal Deep Learning Techniques for Crime Prediction

Tawanda Matereke, Clement Nyirenda and Mehrdad Ghaziasgar

May 28, 2021

A Comparative Evaluation of Spatio Temporal Deep Learning Techniques for Crime Prediction

Tawanda Matereke
Department of Computer Science
University of Western Cape
Cape Town, South Africa
lloydmatereke23@gmail.com

Clement N. Nyirenda
Department of Computer Science
University of Western Cape
Cape Town, South Africa
cnyirenda@uwc.ac.za

Mehrdad Ghaziasgar
Department of Computer Science
University of Western Cape
Cape Town, South Africa
mghaziasgar@uwc.ac.za

***Abstract*—This paper presents a detailed evaluation of three spatio temporal deep learning architectures for crime prediction. These network architectures are as follows: the Spatio Temporal Residual Network (ST-ResNet), the Deep Multi View Spatio Temporal Network (DMVST-Net), and the Spatio Temporal Dynamic Network (STD-Net). The architectures were trained using Chicago crime data set. The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used as performance metrics to evaluate the model. Results show that the STD-Net achieved the best results with an RMSE of 0.2870, and MAE of 0.2093, while the DMVST-Net achieved an RMSE of 0.4171 and an MAE of 0.3455. The ST-ResNet achieved and an RMSE of 0.4033 and an MAE of 0.3278. Future work will include training these algorithms with crime data which is augmented with external data such climate and socioeconomic data. We also will explore hyperparameter optimization of these algorithms using techniques such as evolutionary computation.**

***Index Terms*—Crime prediction, Spatio Temporal, ST-ResNet, DMVST-Net, STD-Net.**

I. Introduction

Crime is defined as an unlawful act punishable by a state or other authority [1]. According to criminologists, crime can be grouped into five major categories, namely: violent (e.g. homicide, assault) crime; property (e.g. burglary, larceny); white-color (e.g. fraud, Embezzlement); organized crime (e.g. drug trafficking, human trafficking) and consensual (e.g. soliciting, brothel-keeping) [2]. The top three countries in the world with the highest crime rate, as recorded in 2021, are Venezuela, Papua New Guinea, and South Africa [3]. Venezuela's highest crime rate of 84.36 has been attributed to corruption among authorities, a flawed judicial system, and poor gun control. Papua New Guinea has the second-highest crime rate of 80.04, is attributed to violent crimes, primarily fueled by rapid social, economic, and political changes in the country. South Africa has the third highest crime rate of 77.29, and this is ascribed to several factors amongst such as high levels of poverty, inequality, social exclusion, unemployment, and the normalization of violence.

Prevention of crime is one of the key features of a successful society [4]. Predicting where and what crime would be committed, is one of the most famous and prominent tasks the police and other agencies want to achieve [5] and crime prediction is essential because it speeds up the process of solving crimes and reduces crime rates [6]. Police departments spend a lot of resources and time to predict crime events. Accurately predicting crime events enables the police to act proactively and helps them to adequately assign patrols [7]. Machine learning has been transforming the way that governments prevent, detect, and address crime [7]. Some police departments are increasingly relying on predictive software like the Santa Cruz-based PredPol, which uses a machine-learning algorithm to predict crime hotspots before the crimes occur [8]. Traditional machine learning include clustering, classification and regression techniques. These techniques have been mostly used for crime prediction. Random forest, Support Vector Machines (SVM), and Classification and Regression Trees (CART) have shown high accuracy in crime prediction [9]. Deep learning, a subset of machine learning, has shown an improved accuracy on prediction when compared to traditional machine learning techniques in various domains. Deep learning has outperformed some traditional machine learning algorithms such as Support Vector Machines, Random Forest, and K - Nearest Neighbors for crime prediction [10].

The rest of the paper is organized as follows. Section II presents related works and sets the motivational foundation for the work done in this paper. Section III describes the details of the three spatio-temporal deep learning architectures that are compared in this work. Section IV presents the implementation procedures of the algorithms, the training data set, the data prepossessing, and the metrics used in the experiments. Section V presents the results and discusses them. Last, the paper is concluded in Section VI.

II. Related Works and Motivation for this Work

This section will review the related works as well as present the motivation behind the work presented in this paper.

Li et al. [11] carried out a spatio-temporal analysis and prediction of urban crime in China using an Auto-regressive moving average (ARIMA) model. Three other models of the Global Model (GM), Pooled Model (PM), and Hierarchical Model (HM) were used for comparative experiments. Error analysis of the prediction results showed that the prediction error rate of the ARIMA model was at least 5% lower than the other three models. Thus, it was considered that the prediction results of ARIMA are more accurate and met the expected requirements.

Kumar et al. [12] carried out forecasting the annual crime rate in India using Time Series Models such as Auto-Regressive Integrated Moving Average (ARIMA) and Exponential Smoothing. The data used for the experiments was obtained from the National Crime Record Bureau of India and was dated from 1953 to 2013. The ARIMA model was compared to the Holt Linear model using the Mean Absolute Error, Mean Absolute Squared Error, Mean Absolute Percentage Error techniques. The ARIMA model exhibited the best results.

Almaw and Kadam [13] proposed a system for crime data analysis and prediction using ensemble learning. The crime data set used was obtained from Denver Open Data Catalog, which is an open data portal for accessing government data. They did a performance comparison of Naive Bayes, Decision tree, Random forest, and Ensemble (Combination of the Decision tree and Random Forest) techniques. The performance of each technique was evaluated using accuracy measure, true positive rate, false negative rate, precision, recall and F-measure. Random Forest classifier achieved the highest accuracy of 82.02%.

Yadav et al. [14] proposed a system for city crime mapping for San Fransisco using machine learning techniques. The crime data set was acquired from the San Francisco Police Department Crime Incident Reporting System, and contained records from 2003 to 2015. Because the crime data set had both numeric and nominal data, a preprocessing step was taken to assign numeric value to nominal data, so that classification techniques can consume it. They explored four classification models along with one clustering model; Gaussian Naive Bayes, Multilayer Perceptron, K-Nearest Neighbors, Classification and Regression tree and k-means. Accuracy and Area Under Curve (AUC) of the Receiver Operator Curve (ROC) metrics were used to evaluate the techniques performance. The Classification and Regression tree (CART) out performed the other techniques in predicting the resolution of crime in San Francisco, and achieved an accuracy of 80.852% and AUC of 0.81.

Bappee, Soares, and Matwin [15] proposed a system for predicting crime using geospatial features for different categories of crime. The model attempted to predict the relationship between criminal activity and geographical regions. Nova Scotia (NS) crime dataset was used as the target of study. They did a comparison of Linear Regression (LR), Support Vector Machine (SVM), and Random Forest (RF), and an Ensemble with all the previous classifiers. Evaluation of the classifiers performance was done by means of accuracy measurement and Area Under the Curve (AUC) of the Receiver Operator Characteristic (ROC) analysis. 10-fold cross validation was used to estimate the classifiers performance on unseen data. Their experiment results showed that the Support Vector Machines and Ensemble learning techniques out performed the other techniques. The Ensemble achieved the highest accuracy and AUC results of 81.8% and 0.69 respectively for the data with engineered features.

Zhang, Zheng, and Qi [16] applied the ST-ResNet to Citywide crowd flow prediction. They used the Beijing and and New York City bike trajectory data. They used min-max normalization to scale the data, and one-hot coding to transform the categirical data into a binary vector. Root Mean Square Error (RMSE) was used for model evaluation. They compared the performance of the ST-ResNet, Historical Average (HA), Auto-regressive Moving Average (ARIMA), Seasonal Auto-regressive Moving Average (SARIMA), Vector Auto-regression (VAR), Spatio Temporal Artificial Neural Network (ST-ANN), and Deep Spatio Temporal Neural Network (Deep ST). The ST-ResNet exhibited the best performance.

Wang et al. [17] carried out real-time crime forecasting on an hourly timescale by applying an ST-ResNet model. They considered all types of crime in Los Angeles (LA) over the last six months of 2015. In total there were 104,957 crimes. Due to the low regularity of the crime data in both space and time, both spatial and temporal regularization of the data was performed. They compare two similar DNN structures except that one has CNN layers (the ST-ResNet). RMSE was used to evaluate the accuracy of the models. The ST-ResNet achieved the best results with an average accuracy of of 84.78% in the top 25 predictions.

Wang et al. [18] continued their experiments on real-time crime forecasting in LA using the ST-ResNet model and proposed some improvements in terms of the model's accuracy and performance on mobile devices. They used the LA crime dataset which has records since 2015 and they also used external weather data. RMSE was used for model evaluation. The ST-Resnet was compared with HA, KNN, ARIMA models. The ST-ResNet produced the best results with a low error in crime density of 0.659.

Stalidis, Semertzidis, and Daras [19] demonstrated that deep learning-based methods outperform the existing best-performing traditional methods on crime classification and prediction. They carried out an evaluation of the effectiveness of different parameters in the deep learning architectures and gave insights for configuring them in order to achieve improved performance in crime classification and finally crime prediction. They used five different datasets. These 5 datasets include incident reports from Seattle, Minneapolis, Philadelphia, San Francisco, and Metropolitan DC police departments. Ten algorithms were compared namely; CCRBoost, ST-Resnet, Decision Trees, Naive Bayes, LogitBoost, Random Forests, Support Vector Machines (SVM), K Nearest Neighbors (KNN), Multi Layer Preceprtron (MLP). F1 score, Area Under Reciever Operater Chracteritic (AU ROC), and PAI (Prediction Accuracy Index) were used to evaluate the model's performance. From the experiment, the ST-ResNet was found dominating other methods.

Stec and Klabjan. [10] used a joint RCNN for the purpose of predicting crime. Chicago and Portland crime datasets were used for the experiments. They combined crime data with additional weather, public transportation, and census data. They conducted experiments to determine the best network structures between the Feed Forward network, CNN, RNN, and RCNN. Mean Absolute Scaled Error (MASE) was used for model evaluation. The RCNN exhibited the best accuracy results on both Chicago (with an accuracy of 75.6%) and Portland (with an accuracy of 65.3%) data sets.

Yao et al. [20] proposed a DMVST-Net framework to model both spatial and temporal relations. The model consisted of three views: temporal view (modeling correlations between future demand values with near time points via LSTM), spatial view (modeling local spatial correlation via local CNN), and semantic view (modeling correlations among regions sharing similar temporal patterns). They used a large-scale online taxi request dataset collected from Didi Chuxing, which is one of the largest online car-hailing companies in China. Mean Average Percentage Error (MAPE) and Root Mean Square Error (RMSE) for model evaluation. The proposed model was compared to HA, ARIMA, Linear regression, MLP, XGBoost, and ST-ResNet. The DMVST-Net achieved the best results with MAPE - 0.1616 and RMSE - 9.642.

Ali et al. [21] proposed a deep hybrid neural network composed of recurrent and convolutional networks to predict citywide traffic crowd flows by leveraging Spatio-temporal patterns. They used the TaxiBj and BikeNYC datasets which were normalized using the Min-Max normalization technique. Root Mean Square Error (RMSE) and Mean Average Percentage Error (MAPE) techniques were used to evaluate the model performances. They compared the performance of their proposed model to that of HA, ARIMA, LinUOTD, XGBoost, MLP, ConvLSTM, STD-Net, and ST-ResNet.

Zhang et al. [22] proposed a model for attention-based supply demand for autonomous vehicles. The dataset they used was obtained from an online car-hailing company in China. The Mean Average Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) techniques were used for evaluating model performance. They compared the performance of their model to that of ARIMA, LSTM, ConvLSTM, Reduced-ConvLSTM, ST-ResNet, DMVST-Net, STD-Net, and Reduced-STD-Net, baseline models. Amongst the baseline models, the STD-Net achieved the best results with MAPE - 21.08%, RMSE - 0.1634, and MAE- 0.1348. The STD-Net achieved the best results as compared to other baseline models that were tested.

Whilst the ST-ResNet has proven to be state-of-the-art deep learning technique for crime prediction [17], [19], [22], [24], there are other recent spatio temporal deep learning techniques, such as the DMVST-Net and STD-Net, have shown better results than the ST-ResNet in other spatio-temporal domains. Therefore, in this paper, we compare the performance of the ST-ResNet, DMVST-Net and STD-Net architectures for crime prediction using the Chicago dataset.

III. Spatio Temporal Deep learning Algorithms

In this section we give a detailed description of the spatio temporal techniques used in this study.

A. Spatio Temporal Residual Network

The ST-ResNet is an extension of the ResNet [23] for prediction in the spatio-temporal domain. We adopt the ST-ResNet architecture proposed by Zhang, Zheng and Qi [24]. Fig. 1 show the overview of the architecture. The architecture

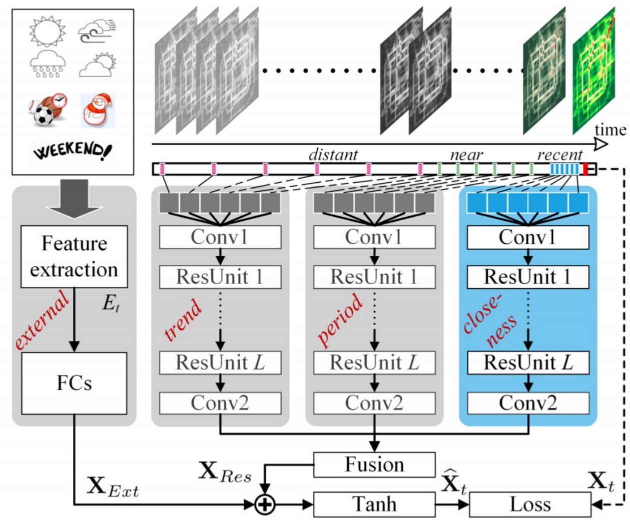


Fig. 1: ST-ResNet Architecture, adopted from Zhang, Zheng and Qi [24].

has four main components that model the temporal *closeness, period, trend* and *external* influence. The *trend, period* and *closeness* components each have layers of convolutional neural networks followed by residual units sequences. These three components structures capture the spatial dependency between close and remote regions. The *external* component was meant to used to extract features from external datasets, however in our study we did not use any external data, hence we did not include this component. the outputs of *trend, period* and *closeness* components were combined using parameter matrices to give X_{Res}. Parameter matrices allocate unique weights to the results of each component in each region. Each component is be represented by

$$X_c^{(1)} = f\left(W_c^{(1)} * X_c^{(0)} + b_c^{(1)}\right) \tag{1}$$

where * represents the convolution; f is a activation function; $W_c^{(1)}$, $b_c^{(1)}$ are the trainable parameters in the first layer. The ResUnits were stacked on the convolution layers as follows:

$$X_c^{(l+1)} = X_c^{(1)} + F(X_c^{(l)}; \theta_c^{(l)}), l = 1, ..., L \tag{2}$$

where F represents a residual function i.e. a combination of ReLU and convolution.

B. Deep Multi View Spatio Temporal Network

The DMVST-Net framework was adopted from Yao, Wu, and Ke [20]. The DMVST-Net is a multi-view model that jointly considers the spatial, temporal and semantic relations. In the framework, convolutional networks are used to learn the local characteristics of regions in relation to their neighbors. The spatial view is represented by a local CNN. The local CNN performs transformation on each k convolutional layer as follows:

$$Y_t^{i,k} = f(Y_t^{i,k-1} * W_t^k + b_t^k) \tag{3}$$

where * denotes the convolutional operation and $f(.)$ is an activation function i.e., $f(z) = max(0, z)$; W_t^k and b_t^k are two sets of parameters in the k^{th} convolution layer.

The Long Short-Term Memory (LSTM) network is used as the temporal view component. The architecture of the LSTM is formulated as follows:

$$\begin{aligned} i_t^i &= \sigma\big(W_i g_t^i + U_i h_{t-1}^i + b_i\big), \\ f_t^i &= \sigma\big(W_f g_t^i + U_i h_{t-1}^i + b_f\big), \\ o_t^i &= \sigma\big(W_o g_t^i + U_i h_{t-1}^i + b_o\big), \\ \theta_t^i &= tanh\big(W_g g_t^i + U_i h_{t-1}^i + b_g\big), \\ c_t^i &= f_t^i \cdot c_{t-1}^i + i_t^i \cdot \theta_t^i, \\ h_t^i &= o_t^i \cdot tanh(c_t^i) \end{aligned} \tag{4}$$

where $(\cdot)$ represents the Hadamard product and *tanh* is the hyperbolic tanget function. Both of these function are element-wise. W_a, U_a, b_a ($a \in \{i, f, o, g\}$) are trainable parameters. The semantic view of the network relates the locations with similar functionality may have similar demands. The semantic view is represented with a graph of locations with functional similarity among regions. Fig. 2a shows the overview of the DMVST-Net.

C. Spatio Temporal Dynamic Network

The STD-Net architecture has been adopted from Ali et al. [21]. In this architecture, the CNN is used to extract spatial features and LSTM is used to extract the temporal features. The architecture has four main components; *closeness, period, trend* and *external*. Fig. 2b shows the overview of the STD-Net.

The mathematical formulation of STD-Net is given by:

$$\begin{aligned} i_t &= \sigma\big(W_{yi} * Y_{r,t} + W_{di} * Y_{r,t} + W_{ri} \cdot C_{t-1} + b_i\big), \\ f_t &= \sigma\big(W_{yf} * Y_{r,t} + W_{df} * Y_{c,t} + W_{ri} \cdot C_{t-1} + b_f\big), \\ C_t &= f_t \cdot C_{t-1} + i_t \cdot tanh\big(W_{yr} * Y_{r,t} + W_{dr} * Y_{r,t-1} + b_c\big), \\ o_t &= \sigma\big(W_{yo} * Y_{r,t} + W_{ho} * Y_{c,t-1} + W_{ro} \cdot C_t + b_o\big), \\ Y_{r,t} &= o_t \cdot tanh(C_t), \end{aligned} \tag{5}$$

where * denotes the convolutional operation and $\cdot$ represents the Hadamard product, and, (W_{yi}, W_{di}, W_{yf}, W_{df}, W_{yr}, W_{dr}, W_{yo}, W_{ho}, W_{ro}, b_i, b_f, b_c, b_o are training parameters. The variables i_t, f_t, C_t and o_t represent the input and output gates, respectively. The convolution layers used ReLU as the activation and used a kernel of size 3×3. For this implementation we did not include the *external* component because we did not include any external data to the crime dataset, therefore we also did not use the fusion.

IV. Implementation of Algorithms

The codes for ST-ResNet [23], DMVST-Net [20], and STD-Net [21] from the publishers GitHub repositories. The experiments were performed on a hyper threaded linux virtual machine on Azure Cloud [25], with 8 virtual CPU's, 128GB RAM, and a processor speed of 2.3 GHz

A. Procedures

In our experiments, the hyper parameters were set as follows: The number of epochs used were 10. The learning rate is chosen to be 0.001. Batch normalization was used. We adopted the ADAM optimizer to optimize the loss function. The length of the temporal sliding window was fixed at a value of at 30.

B. Data set Used in this study

We obtained our crime dataset from the Chicago Data Portal. The dataset includes crime incident reports dating back to 2001, with 7.29 million records. Each report includes location information (in latitude and longitude), a time and a type of crime. There are 32 distinct crime types in the dataset. We selected records in the 3 year period from 20017 to 2020 for training our algorithms.

We adopted the spatial grid resolution that is used in a related work in [36]. These ranges are 16×16 cells. We aim to predict the number of times a certain crime type will happen in the a given area by evaluating past incidents in the current month. Thus, the past crime incidents will be grouped in incident maps I of timespan t of 1 day, and for a period T of 30 days. Therefore, 30 daily incident maps are used as input to forecast a certain crime type for the next period. We used a daily timespan for incidents so that enough temporal detail can be extracted while the time series are adequately populated.

C. Metrics

Finally, we evaluate our proposed method by using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

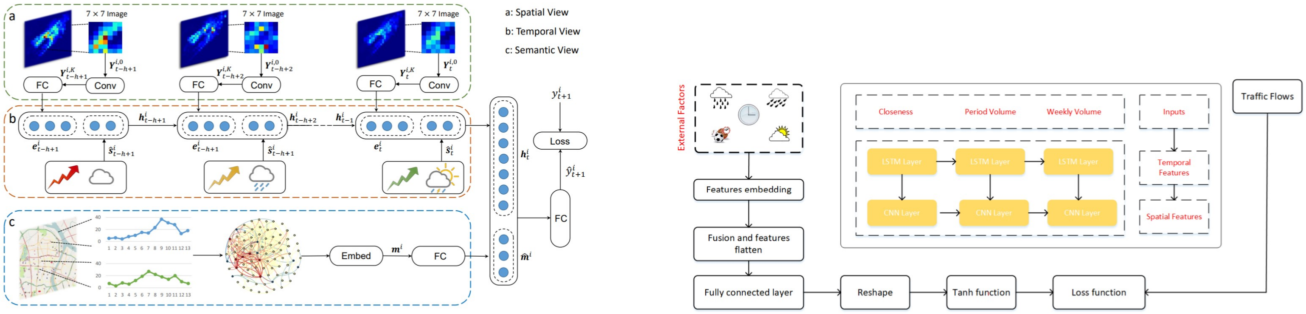
where $\hat{x}_i$ is the predicted value, x_i is the true value and N is the total number of data points.

V. Results

In this section we discuss the results of the performance of the ST-ResNet, the DMVST-Net and thin e STD-Net on the classification of the Chicago crime dataset. We also discuss the loss during the training of the models.

We plotted graphs of losses of each model during training time, over 10 epochs as shown in Fig 3, 4, and 5. From the graphs, it is evident that loss during training drops in each training iteration, hence the ST-ResNet, the DMVST-Net and the STD-Net algorithms are able to interpret the data points well during training.

We trained the models over 10 epochs. We also used early stopping call backs to stop the training when the training loss in the next training step is greater than the previous step. Fig 6



(a) DMVST-Net Architecture [20]. (b) STD-Net Architecture [21].

Fig. 2: DMVST-Net and STD-Net architectures.

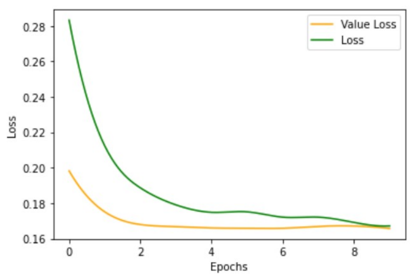


Fig. 3: ST-ResNet Training Loss

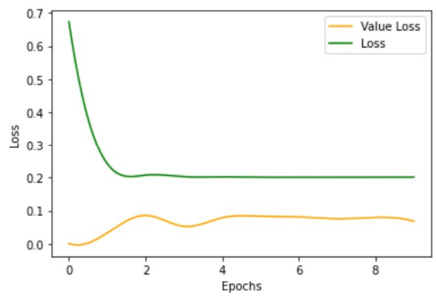


Fig. 4: DMVST-Net Training Loss

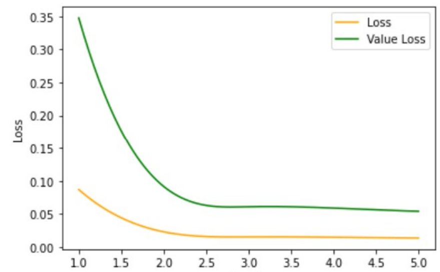
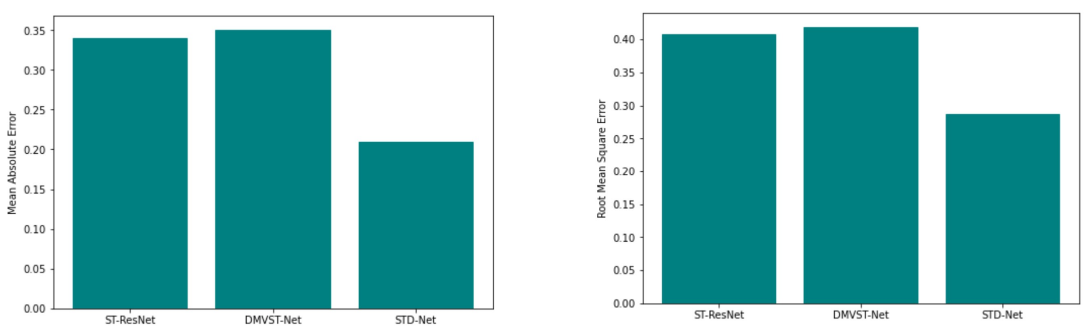


Fig. 5: STD-Net Training Loss



(a) MAE (b) RMSE

Fig. 6: Results for ST-ResNet, DMVST-Net, and STD-Net for crime prediction using Chicago crime dataset..

shows the the MAE and the RMSE results of exhibited by the three algorithms.

These results show that winning architecture is the STD-Net with an MAE of 0.2093 and RMSE of 0.2870.

VI. Conclusion

This paper has presented a performance comparison of three spatio temporal deep learning algorithms for crime prediction. DMVST-Net and STD-Net, which have achieved stellar results in other application domain, have been compared against the state-of-the-art ST-ResNet for crime prediction using the Chicago crime data set. Results show that STD-Net achieves the best MAE and RMSE.

In the future, we will explore if adding external data to these frameworks, similar to worldly semantics, socioeconomic, climate, road guides and focal points in the region can assist these models with learning better insights. We also want to explore optimization of hyper parameters by using techniques such as evolutionary computation.

References

[1] Education and Crime. (n.d.). A Political Ecology of Youth and Crime. doi:10.1057/9781137291486.0008

[2] Social Problems: Continuity and Change. (n.d.). Retrieved October 16, 2020, from https://saylordotorg.github.io/text_social-problems-continuity-and-change/s11-02-types-of-crime.html

[3] Crime rate by country 2020. (n.d.). Retrieved October 16, 2020, from https://worldpopulationreview.com/country-rankings/crime-rate-by-country

[4] Harrendorf, S., Heiskanen, M., & Malby, S. (n.d.). INTERNATIONAL STATISTICS on CRIME AND JUSTICE. Retrieved October 11,

2020, from https://www.unodc.org/documents/data-and-analysis/Crime-statistics/International_Statistics_on_Crime_and_Justice.pdf

[5] Baraniuk, C. (2018, November 26). Exclusive: UK police wants AI to stop violent crime before it happens. Retrieved September 14, 2020

[6] Varshitha D N, Vidyashree K P, Aishwarya P, Janya T S, KR Dhananjay Gupta and Sahana R. 2017. Paper on Different Approaches for Crime Prediction system, InternationalJournal of Engineering Research and Technology (IJERT).

[7] Liv, N. (2019). Crime Prediction Technologies. The International Institute of Counter-Terrorism.

[8] Law Enforcement: PredPol Law Enforcement Intelligence Led Policing Software: PredPol Law Enforcement Intelligence Led Policing Software. (2020, September 30). Retrieved October 18, 2020, from https://www.predpol.com/law-enforcement/

[9] N.A. Zaidi, A. Mustapha, S.A. Mostafa, and M.N. Razali, "A classification approach for crime prediction. Communications in Computer and Information Science," 68-78. doi:10.1007/978-3-030-38752-5_6, 2020.

[10] A. Stec, and D. Klabjan, "Forecasting Crime with Deep Learning," arXiv e-prints, 2018.

[11] Z. Li et al., "Spatio-Temporal Pattern Analysis and Prediction for Urban Crime," 177-182. 10.1109/CBD.2018.00040, 2018.

[12] M. Kumar et al. "Forecasting of Annual Crime Rate in India: A case Study," In proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2018.

[13] A. Almaw and K. Kadam, "Crime Data Analysis and Prediction Using Ensemble Learning," In proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 1918-1923, doi: 10.1109/ICCONS.2018.8663186, 2018.

[14] N. Yadav, A. Kumar, R. Bhatnagar, and V.K. Verma, "City Crime Mapping Using Machine Learning Techniques," In proceedings of the International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019), Advances in Intelligent Systems and Computing, vol 921. Springer, Cham. https://doi.org/10.1007/978-3-030-14118-9_65, 2019.

[15] F.K. Bappee, A. Soares Júnior, and S. Matwin, "Predicting Crime Using Spatial Features. In Bagheri E., Cheung J. (eds) Advances in Artificial Intelligence. Canadian AI 2018. Lecture Notes in Computer Science, vol 10832. Springer, Cham. https://doi.org/10.1007/978-3-319-89656-4_42, 2018.

[16] J. Jin, Y. Lin, Z. Wu, and H. Wan, "Spatio-Temporal Recurrent Convolutional Networks for Citywide Short-term Crowd Flows Prediction," In Proceedings of the 2nd International Conference on Compute and Data Analysis (ICCDA 2018), Association for Computing Machinery, New York, NY, USA, 28–35. DOI:https://doi.org/10.1145/3193077.3193082, 2018.

[17] B. Wang et al., "Deep Learning for Real Time Crime Forecasting", arXiv e-prints, 2017.

[18] B. Wang, P. Yin, A.L. Bertozzi, P. J. Brantingham, S.J. Osher, and J. Xin, "Deep Learning for Real-Time Crime Forecasting and Its Ternarization," Chinese Annals of Mathematics, Series B, 40(6), 949-966. doi:10.1007/s11401-019-0168-y, 2019.

[19] P. Stalidis, T. Semertzidis, and P. Daras, "Examining Deep Learning Architectures for Crime Classification and Prediction," arXiv:1812.00602, 2018.

[20] H. Yao et al, "Deep Multi-view Spatial-Temporal Network for Taxi Demand Prediction," in Proc. Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, pp. 2588-2595, Feb. 2018.

[21] A. Ali, Y. Zhu, Q. Chen, J. Yu and H. Cai, "Leveraging Spatio-Temporal Patterns for Predicting Citywide Traffic Crowd Flows Using Deep Hybrid Neural Networks," In proceedings of the IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS), Tianjin, China, pp. 125-132, doi: 10.1109/ICPADS47876.2019.00025, 2019.

[22] Z. Zhang, H. Dong, Y. Li, Y. You, and F. Zhao, "Attention-Based Supply-Demand Prediction for Autonomous Vehicles," In the 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT). doi:10.1109/pdcat46702.2019.00085, 2019.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770–778, 2016

[24] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," AAAI, 2017.

[25] M. Copeland, J. Soh, A. Puca, M. Manning, and D. Gollob, "Microsoft Azure and Cloud Computing," In Microsoft Azure. Apress, Berkeley, CA, 2015.

[26] The Commission on Crime Prevention and Criminal Justice. United Nations: Office on Drugs and Crime, //www.unodc.org/unodc/en/commissions/CCPCJ/index.html. Accessed 16 Mar. 2021.

[27] Y. LeCun, Y. Bengion, and G. Hinton, "Deep learning," Nature, 521:436–444, 2015.

[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput, 9:1735–1780, 1997.

[29] H. Lee and H. Kwon, "Going Deeper With Contextual CNN for Hyperspectral Image Classification," in IEEE Transactions on Image Processing, vol. 26, no. 10, pp. 4843-4855, doi: 10.1109/TIP.2017.2725580, Oct. 2017.

[30] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-rnn: Deep learning on spatio-temporal graphs," CVPR, 2016.

[31] K. Hara, H. Kataoka, and Y. Satoh, "Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition," IEEE International Conference on Computer Vision Workshops (ICCVW). doi:10.1109/iccvw.2017.373, 2017.

[32] H. Zunair, A. Rahman, N. Mohammed, and J.P. Cohen, "Uniformizing techniques to Process Ct scans with 3d CNNs for tuberculosis prediction" Predictive Intelligence in Medicine, 156-168. doi:10.1007/978-3-030-59354-4, 2019.

[33] Francois Chollet et al. Keras. https://github.com/kerasteam/keras, 2015.

[34] S. Ji, W. Xu, M. Yang and K. Yu, "3D Convolutional Neural Networks for Human Action Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221-231, doi: 10.1109/TPAMI.2012.59, Jan. 2013.

[35] M. Abadi et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from tensorflow.org. 2015. URL: https : / / www.tensorflow.org/. https://doi.org/10.1007/978-1-4842-1043-7_1

[36] C-H. Yu et al. "Hierarchical Spatio-Temporal Pattern Discovery and Predictive Modeling". In: IEEE Transactions on Knowledge and Data Engineering 28(4), pp. 979–993,Jan. 2013,