

Predicting Disease Outbreaks with Climate Data

Tassallah Abdullahi

Department of Computer Science
University of Cape Town, South Africa
abdtas008@myuct.ac.za

Geoff Nitschke

Department of Computer Science
University of Cape Town, South Africa
gnitschke@cs.uct.ac.za

Abstract—The incidence of most diseases varies greatly with seasons, and global climate change is expected to increase its risk. Predictive models that automatically capture trends between climate and diseases are likely to be beneficial in minimizing disease outbreaks. *Machine learning* (ML) predictive analytic tools have been popularized across many health-care applications, however the optimal task performance of such ML tools largely depends on manual parameter tuning and calibration. Such manual tuning significantly limits the full potential of ML methods, especially for high-dimensional and complex task domains, as typified by real-world health-care application data-sets. Additionally, the inaccessibility of many health-care data-sets compounds innate problems of method comparison, predictive accuracy and the overall advancement of ML based health-care applications. In this study we investigate the impact of *Relevance Estimation and Value Calibration*, an evolutionary parameter optimization method applied to automate parameter tuning for comparative ML methods (*Deep learning* and *Support Vector Machines*) applied to predict daily diarrhoea cases across various geographic regions. *Data-augmentation* is also used to complement real-world noisy, sparse and incomplete data-sets with synthetic data-sets for training, validation and testing. Results support the efficacy of evolutionary parameter optimization and data synthesis to boost predictive accuracy in the given task, indicating a significant prediction accuracy boost for the deep-learning models across all data-sets.

Index Terms—Predictive Machine Learning, Evolutionary Parameter Optimization, Deep-Learning, Support Vector Machines

I. INTRODUCTION

The continuous spread and severity of various infectious diseases has resulted in increased interest in designing applications that aid in reducing their widespread occurrence to avoid outbreaks and epidemics. Disease outbreak and spread is often exacerbated by climate change whose prevalence is said to be one of the major challenges for controlling diseases especially in developing countries [1] [2]. Several observations have indicated that extreme weather events ranging from heat drought or heavy rainfall lead to changes in the ecology of infectious which poses a threat to humans through increased morbidity and mortality [3] [4]. For example, Musengimana *et al.* [5] demonstrated that for every increase in temperature, the rate of some infectious disease hospitalizations increased by eight percent.

Hence, developing computational and formal models able to capture complex relationships and long-term dependencies

between climate factors and diseases may be effective for predicting disease outbreaks, thus offering public health authorities the means for timely detection and control of disease outbreaks across various geographic regions.

Various supervised *Machine Learning* (ML) including *Support Vector Machines* (SVMs) [6] and deep learning techniques such as *Convolutional Neural Networks* (CNNs) [7], *Long Short-Term Memory Networks* (LSTMs) [8] have been applied in medicine and health-care research for developing predictive and diagnostic models for various diseases. For example, CNNs have been effectively applied to Malaria parasite [9] and Tuberculosis [10] related disease detection. Whereas, LSTMs have been used to predict the outbreak of Typhoid, Chicken Pox and Scarlet Fever [11] in sample populations, and SVMs have been used to detect Hepatitis [12]. Despite such advances in health-care applications, overall ML method task performance still largely depends on available training data and method parameter settings which is a significant challenge for most predictive ML methods [13] [14]. Currently, many ML studies still use time consuming and computationally expensive manual parameter tuning techniques, with often sub-optimal predictive performance for complex tasks [13].

Sub-Saharan Africa is a geographic region that experiences significant annual changes in temperature and precipitation, with unpredictability of climate exacerbated by climate change. Such climate factors play a vital role in the long-term trends of diseases [4], [5] [15]. For example, in South Africa, diarrhoea is one major infectious diseases driven by climate change [16]. In Western Cape province of South Africa, the rate of diarrhoea hospitalizations was strongly correlated to increases in minimum and maximum temperatures [5]. Whereas, in the Limpopo province, Ikeda *et al.* [4] found high numbers of diarrhoea cases coincided with below normal precipitation rates. Diarrhoea is a major health concern that accounts for three percent of total annual death records across South Africa [16] and still higher reported rates in other countries [17]. Predictive models would enable governments and healthcare providers to take necessary action and intervene to mitigate risks related to diarrhoea, thus minimizing the costs of delivering related medical care.

The overall aim of this study is to ascertain the suitability of various ML methods, given various climate factors and synthetic (generative) training data, for accurately predicting diarrhoea outbreaks. Specifically, the study aims to elucidate what type of ML method is most suitable when coupled with given types of training and test data. That is, data-sets comprising specific climate variables, data sparseness and noise and a synthetic data compliment, that when coupled with a given ML method enables optimal prediction efficacy.

To address this objective, we compared the predictive task-performance (number of expected daily diarrhoea cases) of three ML methods (CNNs, LSTMs and SVMs) across nine South African provinces. Given the sparse and noisy nature of the data-sets used for method training and testing, we necessarily *augmented* [18] the available data with synthetic data generated using *Generative Adversarial Networks* (GANs) [19]. Also, given deficient related work to guide ML method parameter tuning and calibration for disease outbreak prediction, we used *Relevance Estimation and Value Calibration* (REVAC) [20], an evolutionary algorithm design for meta-heuristic parameter tuning [21], and applied to optimise ML method parameters used in this study.

This study’s key contribution was its comprehensive study and application of pertinent ML methods to real-world health-care data sourced from various South African medical institutions. This enabled us to devise an effective predictive ML methodology for sub-Saharan Africa, which is (globally) one of the regions most adversely affected by diarrhoea outbreaks [17]). The study’s second contribution was its demonstration of the efficacy of automated (evolutionary) parameter tuning [20], [21], combined with data-augmentation (data synthesis to complement deficient training data [14]), for the given predictive ML task and data-sets.

II. METHODS

This section presents an overview of ML methods, parameter tuning, and data-sets used for training and testing.

A. Data-sets

Data-sets¹ used focused on nine South African provinces: *Western Cape, Eastern Cape, Northern Cape, North West, Free State, Limpopo, KwaZulu Natal, Gauteng, and Mpumalanga*. For each province, data-sets consisted of nine features: Diarrhoea cases and eight climate features². Data-set sources were classified as *real-world* (1) and *synthetic* (2) data-sets.

1) *Real-world Data-sets*: For each province, a ten-year period (2008–2018) of daily sales records of *Loperamide*, an anti-diarrhoea compound evaluated in the treatment of patients with chronic non-specific diarrhoea in South

Africa was obtained from *Clicks pharmaceuticals*. These data were used as proxy for diarrhoea cases in the region. The number of diarrhoea cases per day for a specific province was computed as the number of Loperamide sales per day in the given province. Six-hourly data on eight climate factors between the period of (2008–2019) were obtained from the National Centers for *Atmospheric Research and Environmental Prediction*³. These data-sets were re-gridded to daily average format to match the daily diarrhoea case data format (making a total of 3,763 time-series samples).

2) *Synthetic Data-sets: Generative Adversarial Networks* (GANs) [14] were used to generate 20,000 synthetic time-series samples with 24 time step for the diarrhoea data and eight climate data in each country province. GANs were selected as they have been previously demonstrated as effective for generating different types of realistic data [22] [23]. The aim of generating synthetic was to have sufficient data for making predictions, where synthetic data was augmented with the real-world data-sets in two ways: *upward augmentation* and *downward augmentation*. When the data-sets were augmented upwards, the training set included a combination of the real-world and synthetic samples, but the test set included only the synthetic data-sets and when the data-sets were augmented downwards, the training set included mainly the synthetic data-sets and the test set included the real-world data-set. Based on previous studies [14], augmentation with considerable amounts of synthetic data enhance performance while too much synthetic data slows down performance. Thus, to effectively utilize the augmented data, for all experiments, we used the following mixture (percentage) ratios when combining our synthetic and real-world data-sets (respectively): 90:10, 80:20, 70:30, 60:40, and 50:50.

B. Configuration of ML Algorithms

1) *CNN Method*: Our CNN method was designed with 1D convolutions to match the sequential nature of our input data. *CNNs* are a class of feed-forward, deep neural networks that consist of multiple convolutional and activation layers, pooling layers, and a fully connected layer. These layers are designed to perform specific tasks such as feature extraction from the input data. After several iterations of convolutions, activation and pooling the final output is computed in the fully connected layer of the network [7].

2) *LSTM Method*: *LSTMs* are example of *Recurrent Neural Networks* (RNNs) [7] that address the issue of exploding and vanishing gradients. They contain memory cells that maintain its state overtime and is managed by gating units that control how it memorize, erase, and expose information. These gating units are called the input gate, forget gate and output gate respectively. The ability of the gating units enables the network to handle sequential data effectively [7].

¹<https://github.com/ProjectRepo2021/predicting-outbreaks>

²Climate variables: *Maximum, Minimum and Air temperature, Humidity, Evaporation and Precipitation rate, Surface pressure, Wind velocity*.

³<https://psl.noaa.gov/>

3) *SVM Method*: SVMs are mathematical models whose main function is to find hyper-planes capable of creating margins that separates data points in a high dimensional feature space with the smallest structural risk using kernel functions. SVMs are widely accepted for their ability to solve nonlinear regression estimation problems [6]. In addition, their non-parametric nature enables them to represent complex and non-linear functions easily. We used an SVM with a *Radial Basis Function* (RBF) Kernel [6] for all our predictions.

Deep learning methods (CNN, LSTM) were implemented with the *Keras and TensorFlow*⁴ library while the Python *Scikit-Learn*⁵ library was used to develop our SVM. All methods were configured for reproducible experiments, we thus used fixed random seeds for all experiments⁶. All method parameters (table II) were tuned using either *Grid-search* (section II-D2) or REVAC (section II-D1). For all experiments, our DL models used *Min-Max normalization* since it largely adopted for most neural network regression models [24]. The *standard scaling* technique [25] was adopted for SVM since SVMs assume that the data given as input is within a standard range. We used the *Scikit-Learn package*⁵ to implement all normalization. Data-sets used for testing and training were divided into a ratio of 70:30. Data-sets with earlier dates were used for training and data-sets with later dates were used to test and verify method predictive accuracy.

C. Performance Evaluation Criteria

To compare and evaluate the performance of our ML methods, the *Root Mean Square error* (RMSE) was used given its demonstrated efficacy in many ML prediction studies [3], [11]. Specifically, the RMSE metric is superior at disclosing differences in method performance and is recommended given evaluations based on understanding predictions [26]. RMSE is the square root of the mean of the squared differences between actual outcomes and the method predictions. RMSE is calculated as presented in equation (1).

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

In equation (1), x_i is the actual value while y_i is the predicted value and n the total number of observations. The method with the smallest RMSE error is considered to be the best performing in terms of prediction accuracy.

D. Determining Optimal ML Parameters

1) *Relevance Estimation and Value Calibration*: REVAC is an evolutionary algorithm formally designed to automate meta-heuristic parameter tuning. Given an objective, a population of parameter vectors, N iterations and a task environment, REVAC explores, selects, and evaluates

favourable parameter set values. Using artificial evolution operators such as mutation, recombination, selection and replacement, it improves and updates the distribution of the parameter vectors such that after each iteration, there is an increasing chance of obtaining optimal performance given that *tuned* parameters values are adopted for training an ML method. In this study, REVAC was implemented as a meta-layer to search for optimal parameter values for each of our ML methods (SVM, LSTM, CNN) when applied to the task of predicting diarrhoea cases. The REVAC implementation used for this study was based on the methodology used by Nannen & Eiben [20]. The list of REVAC tuned parameters and their resultant values is presented in table II.

2) *Grid-Search Parameter Tuning*: Grid-search is an automatic parameter tuning technique that trains an ML method with a combination of possible parameters on the training set and evaluates and outputs the best parameters based on a given task performance metric [13]. Grid-search parameter tuning was implemented for each ML method using the Python *Scikit-Learn Grid Search CV package*⁵.

Both REVAC and Grid search selected combinations of possible parameter values from ranges of user-specified values. Table II presents the list and range of parameters values specified for each ML method. Deep learning parameters not specified in table II, used the default values of the *Keras* package. The objective of the both techniques was to minimize the RMSE output of each ML method's prediction, for a given input data-set. For each ML method and given province data, the best *tuned* parameters (leading to the lowest RMSE outputs during REVAC and Grid-search parameter tuning), were the parameters used for final method predictions.

III. EXPERIMENTS

Experiments⁷ (table I) were carried out to predict the number of daily diarrhoea cases across geographic regions (nine South African provinces in this case study).

We considered past observations (*lags*) in our method predictions since patterns of the past are likely to be repeated in the future. Thus, we tested the predictions of the three ML methods with respect to four different lag periods from all input features (climate features and previous diarrhoea cases). The lag periods we considered include lag of *one day, five days, two weeks* and *three weeks*. For example, a one day lag meant that the predictions made by a method for January 6th, 2018 was made with input variables (for all features) for January 5th, 2018, while a five day lag meant predictions for January 6th, 2018 was made with input variables (for all features) for January 1st, 2018. These lag periods were chosen since preliminary experiments showed such lags produced more accurate predictions for all methods.

⁴<https://keras.io/>

⁵<https://scikit-learn.org/>

⁶<https://www.tensorflow.org/>

⁷<https://github.com/ProjectRepo2021/predicting-outbreaks>

TABLE I
EXPERIMENTS OVERVIEW

Experiment Description	Parameter Tuning Technique	Datasets Used	Research Objective
(1) Predictions with real-world data	Grid Search	Real-world data	Determine best predicting ML model given real-world data
(2) Predictions with real-world and synthetic data	Grid Search	Upward and downward augmented data	Determine the effect of augmented data on predicting performance of ML models.
(3) Predictions with augmented data and REVAC tuned parameters	REVAC tuning	Upward and downward augmented data	Determine the impact of REVAC tuning on predicting performance of ML models

TABLE II
EXPERIMENT PARAMETERS AND THEIR RANGES

SVM	LSTM	CNN	REVAC
C: [1, 100]	dropout rate: [0.1,0.2,1.0]	pool size: [1,2]	generation size: [100]
Gamma: [0.001, 0.1]	LSTM layers: [1,2,3]	convolutional layers: [1,2,3]	population size: [80]
	neurons: [6,12,16,18,24,28,32,50,64,100]	kernel size: [6,12,16,18,24,28,32,64]	crossover parent size: [2]
	batch size: [4,16,18,32,64]	batch size: [4,16,18,32,64]	child size: [1]
	learning rate: [0.001, 0.01]	learning rate: [0.001, 0.01]	replacement size: [1]
	epochs size: [40,50,60,70,100,120,150,200]	epochs size: [40,50,60,70,100,120,150,200]	

Thereafter, we determined the best performing ML method via comparing the RMSE from the predictions made by each method (with respect to the four lag periods). This was done in three experiments, where for each ML method predictions were repeated three times for each lag, across each province data-set and the average RMSE computed. Experiment 1 was implemented with the real-world data, with the objective of determining which ML method performs best given the amount of data instances contained in the given data-sets.

We then measured the degree of impact of each climate variable on the best performing diarrhoea prediction methods for each given data-set (country province), via conducting a sensitivity analysis [27]. We adopted the *Backward step-wise* method [27] to measure the effect of one variable at a time while keeping the other variables fixed. Sensitivity was then measured by observing changes in the RMSE error of each given method based on the omission of a specific variable. This sensitivity analysis demonstrated the larger the increase in RMSE, the higher the importance of the omitted variable.

Experiment 2 determined the impact of data-augmentation (data synthesis) applied to training and testing data on the predictive task performance of the three ML methods. The input data-sets used for predictions were the upward and downward augmented data-sets (combinations of the synthetic and real-world data-sets, section II). Predictions were made with each input data-set separately for each province. Each input data was divided in the ratio 70:30 for training and testing, respectively. Parameters selected by Grid-search tuning in experiment 1 were maintained for each ML method and their respective province data-sets.

Experiment 3 was performed to determine the impact of REVAC parameter tuning on the prediction task performance of the three ML methods given data-augmentation (partially synthetic data). The key difference of experiment 3 (compared to experiments 1 and 2), is its use of REVAC to tune method parameters given partially synthetic data-sets (section II). The number of REVAC generations was set to 100 with an initial population size of 80 for each parameter (based on demonstrated efficacy in previous work [20], [21]). All REVAC algorithm parameters are presented in table II. Parameters tuned by REVAC for each method and a given province data-set, were then used for final method predictions.

IV. RESULTS AND DISCUSSION

Results overall, indicate the CNN, followed by the LSTM method out-performed SVM predictive task-performance (average RMSE calculated over all data-sets and lag-periods, section III). However, for all data-sets and data-augmentations, there was no single method that consistently yielded a significantly higher (Wilcoxon test [28], $p > 0.05$) task performance (lower average RMSE), compared to other methods. Prediction results made with real-world data-sets indicate that the CNN yielded the best predictive task performance (lowest average RMSE), followed by the LSTM, and the SVM yielding the poorest performance (highest RMSE). For each country province data-set, the CNN, given well established universal function approximation and pertinent feature detection [7], yielded a comparatively higher task performance (despite limited training set size).

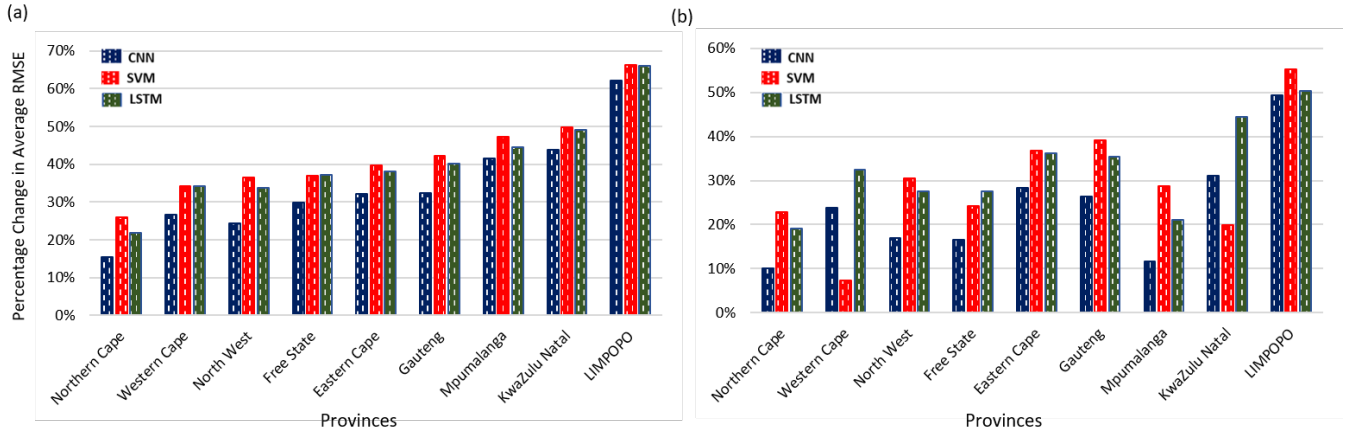


Fig. 1. Percentage change in predictive performance (average RMSE) of each method, for each country province data-set, given (a) *upward* augmented data-set and (b) *downward* augmented data-set (section II).

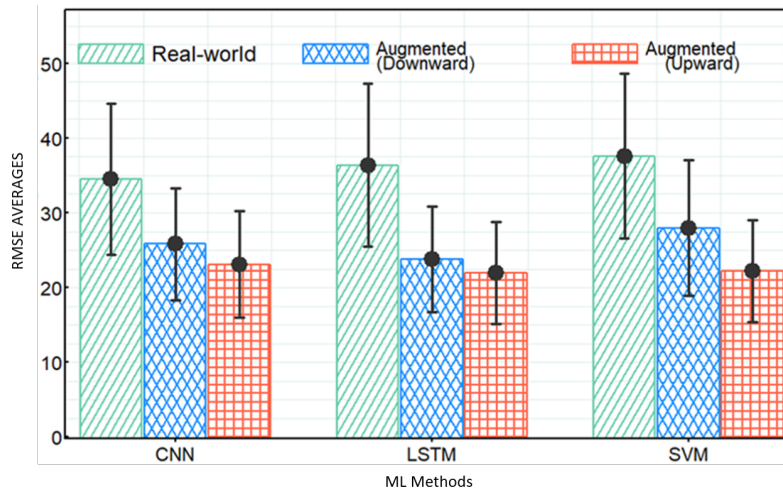


Fig. 2. A comparison of method prediction accuracy, averaged for all country provinces data-sets. Lower RMSE average indicates better performance accuracy. CNN: *Convolutional Neural Network*, LSTM: *Long-Short Term Memory Neural Networks*, SVM: *Support Vector Machine*.

This result is supported by related work similarly demonstrating increased prediction accuracy of CNNs compared to other supervised-learning methods applied to various infectious disease prediction tasks [9], [10]. Figure 1 indicates that data augmentation (section II) significantly improved (Wilcoxon test, $p < 0.05$) the task performance for all methods and data-sets. Figure 2 indicates the CNN method's overall average (over all data-sets) RMSE for predictions using *real-world*, *upward augmented*, and *downward augmented* data-sets were 31.55, 23.11 and 25.80 respectively, whereas the SVM method's overall RMSE was 33.89, 22.17 and 27.97 respectively, and LSTM method RMSE results were 32.91, 21.93 and 23.78 respectively.

Thus, significant task performance increases (lower average RMSE) were observed for all methods using augmented data-sets compared to predictions by the same methods using only real-world data-sets. This indicates that all tested methods benefit from larger training data-sets and also supports the

efficacy of data-augmentation [29] for boosting predictive task performance. However, LSTM out-performed (though without significant difference, Wilcoxon test, $p > 0.05$) other methods given any augmented data-sets. We surmise that the LSTM method yields an increased task performance as training data-set size increases, where the relatively poor LSTM task performance in experiment 1 was the result of sparse and noisy real-world data-set with small size (compared to augmented data-set size, section II).

For example, Yang *et al.* [30] demonstrated LSTM task performance benefits given large training set size, and Jia *et al.* similarly reported that [11] the capability of LSTMs to learn long-term dependencies and patterns in sequential data (regardless of noise) given sufficient training set size. However, the task and methodological designs resulting in the relatively poor LSTM task performance observed in experiment 1 remains a focus of ongoing research.

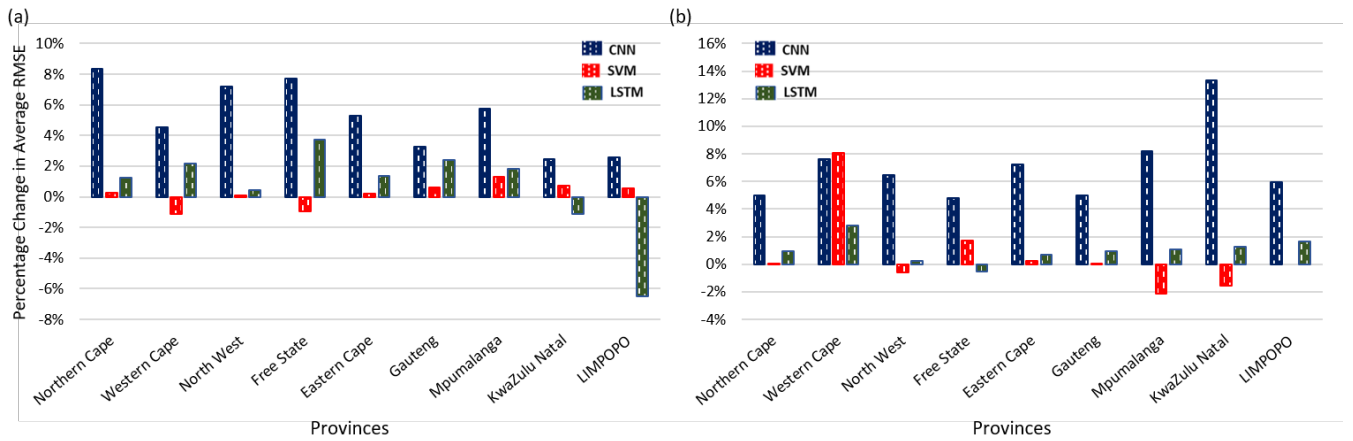


Fig. 3. Percentage change in performance of each method over each province given predictions made with REVAC tuned parameters. (a) *upward* augmented data and (b) *downward* augmented data-set. Higher percentage RMSE indicates higher task performance improvement.

TABLE III
ML METHOD'S OVERALL RMSE AVERAGES FOR PREDICTIONS WITH BOTH REVAC AND GRID SEARCH METHOD PARAMETERS

ML Method	REVAC tuning		Grid search tuning	
	Upward Augmented data	Downward Augmented data	Upward Augmented data	Downward Augmented data
CNN	22.07	23.86	23.11	25.80
LSTM	21.60	23.61	21.93	23.78
SVM	22.17	27.30	22.17	27.97

In terms of elucidating the impact of parameter tuning in company with data augmentation on method task performance, figure 1 indicates that using Grid-search tuned parameters, the average percentage increase in task performance of the CNN method was the least when compared to the other methods in each individual country province. Additionally, table III, indicates that the SVM average RMSE (22.17) out-performed the CNN's RMSE average (23.11) when Grid-search parameters were used with the upward augmented data-sets (see experiment 2, table I).

The instances where SVM out-performed the CNN was hypothesized to be the result of CNN parameter settings used during training, as similarly reported in related deep-learning based predictive applications [9], [10], [11]. To test this hypothesis we thus setup an experiment where all methods had parameters tuned with REVAC evolutionary optimization (experiment 3). Figure 3 presents the percentage change in RMSE results, for all methods, when method parameters were tuned with REVAC instead of Grid-search. These results indicate that CNN prediction results improved for all province data-sets, where the highest percentage increase recorded for the CNN was approximately 12% and the least recorded increase was approximately 2.5%.

However, table III indicates that LSTM method prediction task performance (given REVAC tuning) still out-performed

other methods, further supporting suppositions that LSTM task performance benefits from augmented data-sets. Figure 3, also indicates that SVM yields the least average percentage increase and the highest average percentage decrease across all country provinces. These results (figure 3), thus indicate that REVAC parameter tuning is more suitable for deep learning methods and less suitable for SVM methods. We theorize this to be a result of the low dimensional search space constituted by SVM's few method parameters. That is, an SVM (using an RBF kernel) method parameters are gamma and C only, meaning that parameter tuning will likely have limited impact on SVM task performance.

Finally, the results from our sensitivity analysis presented in figure 4, indicated that the prediction of diarrhoea outbreak by all methods is influenced by specific climate factors. The most prominent factors are *precipitation*, *humidity*, *evaporation* and *temperature*, where levels of influence differ across provinces. These findings are supported by related work [1], [5] that similarly observed that diarrhoea cases increases for every one degree centigrade increase in temperature. Other related work [4], [3], has similarly demonstrated that precipitation rate and humidity are strongly related to increases in diarrhoea related disease and hospitalizations.



Fig. 4. Sensitivity analysis results for the CNN applied to each country province data-set. X-axis indicates method prediction accuracy once a given variable (Y-axis) is omitted. The longer the bar, the larger the loss in CNN accuracy and the more important the variable. This sensitivity analysis indicates: *precipitation*, *humidity*, *evaporation*, and *temperature*, climate variables as having the greatest impact on predictive task-performance.

Overall, this study demonstrated the benefits (predictive task performance increases), given the application of evolutionary parameter tuning (REVAC in this case study) combined with data-augmentation (generative data synthesis), for all evaluated ML methods (CNN, LSTM, and SVM). However, the impact of data-sets of varying complexity (size, noise and sparseness) as well as varying data-set compositions (synthetic versus real-world) on predictive task performance for a broader range of ML methods and predictive task domains remains the topic of current research.

V. CONCLUSIONS

This study evaluated the application of *Support Vector Machines* (SVMs), *Long-Short Term Memory Neural Networks* (LSTM) and *Convolutional Neural Networks* (CNNs), on the task of predicting daily diarrhoea given a set of disparate training and test data-sets (indicative of nine South African provinces). Prediction task performance was with respect to climate variable sets and varying

combinations of real-world and synthetic (augmented) data-sets training and testing data. Results indicated that overall (for all real-world data-sets), the CNN yielded the highest accuracy predictions, and a sensitivity analysis revealed the *precipitation*, *humidity*, *evaporation*, and *temperature* climate variables as having the greatest impact on predictive CNN task-performance. Also, for most data-sets, synthetic data-augmentation and evolutionary parameter tuning by the *Relevance Estimation and Value Calibration* (REVAC) method resulted in a significantly higher CNN predictive task performance, compared to the LSTM and SVM. The study's key contribution was its methodological guide (empirical data for predictive machine learning algorithm design, parameter tuning and data-set calibration), for future research using climate variable oriented data-sets to predict disease outbreak across varying geographic regions. The study thus contributes to ongoing efforts to develop automated early-warning systems for broad-spectrum disease outbreak prediction across various geographic areas and climate zones.

REFERENCES

- [1] M. Dhimal, K. Karki, K. Aryal, S. Shrestha, and B. Pradhan, *Final Report on Assessment of Effects of Climate Factors on Diarrheal Diseases at National and Sub-national Levels in Nepal*. World Health Organization Country Office, Kathmandu, Nepal: Nepal Health Research Council Report., 2016.
- [2] WHO, *Protecting Health from Climate Change: Connecting Science, Policy and People*. WHO, Geneva, Switzerland: WHO, 2009.
- [3] Y. Wang, J. Li, J. Gu, Z. Zhou, and Z. Wang, "Artificial Neural Networks for Infectious Diarrhea Prediction using Meteorological Factors in Shanghai," *Applied Soft Computing*, vol. 35, pp. 280–290, 2015.
- [4] T. Ikeda, T. Kapwata, S. Behera, N. Minakawa, M. Hashizume, N. Sweijd, A. Mathee, and C. Wright, "Climatic Factors in Relation to Diarrhoea Hospital Admissions in Rural Limpopo, South Africa," *Atmosphere*, vol. 10, no. 9, p. 522, 2019.
- [5] G. Musengimana, F. Mukinda, R. Machezano, and H. Mahomed, "Temperature Variability and Occurrence of Diarrhoea in Children under Five-years-old in Cape Town Metropolitan Sub-districts," *Journal of Environmental Research and Public Health*, vol. 13, no. 9, p. 859, 2016.
- [6] N. Sapankevych and R. Sankar, "Time Series Prediction using Support Vector Machines: A Survey," *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 24–38, 2009.
- [7] Y. Le Cun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 1, pp. 436–444, 2015.
- [8] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Predicting Healthcare Trajectories from Medical Records: A Deep Learning Approach," *Journal of Biomedical Informatics*, vol. 69, pp. 218–229, 2017.
- [9] K. Fuhad, J. Tuba, M. Sarker, R. Ali, S. Momen, N. Mohammed, and T. Rahman, "Deep Learning Based Automatic Malaria Parasite Detection from Blood Smear and its Smartphone Based Application," *Diagnostics*, vol. 10, no. 5, p. 329, 2020.
- [10] Z. Abideen, M. Ghafoor, K. Munir, M. Saqib, A. Ullah, T. Zia, S. Tariq, G. Ahmed, and A. Zahra, "Uncertainty Assisted Robust Tuberculosis Identification With Bayesian Convolutional Neural Networks," *IEEE Access*, vol. 8, pp. 2812–2825, 2020.
- [11] W. Jia, Y. Wan, Y. Li, K. Tan, W. Lei, Y. Hu, Z. Ma, X. Li, and G. Xie, "Integrating Multiple Data Sources and Learning Models to Predict Infectious Diseases in China," *AMIA Summits on Translational Science Proceedings*, p. 680, 2019.
- [12] Y. Guo, Y. Feng, F. Qu, L. Zhang, B. Yan, and J. Lv, "Prediction of Hepatitis E using Machine Learning Models," *PLOS ONE*, vol. 15, no. 9, p. e0237750, 2020.
- [13] J. Wu, X. Chen, H. Zhang, L. Xiong, H. Lei, and S. Deng, "Hyperparameter Optimization for Machine Learning Models based on Bayesian Optimization," *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, 2019.
- [14] Z. Che and et al., "Boosting Deep Learning Risk Prediction with Generative Adversarial Networks for Electronic Health Records," in *Proceedings of International Conference on Data Mining*. IEEE, 2017, pp. 787–792.
- [15] C. Davis and K. Vincent, "Climate Risk and Vulnerability: A Handbook for Southern Africa," Pretoria, South Africa, 2017.
- [16] T. Kapwata, A. Mathee, W. Le Roux, and C. Wright, "Diarrhoeal Disease in Relation to Possible Household Risk Factors in South African Villages," *Journal of Environmental Research and Public Health*, vol. 15, no. 8, p. 1665, 2018.
- [17] C. Troeger and et al., "Estimates of the Global, Regional, and National Morbidity, Mortality, and Aetiologies of Diarrhoea in 195 countries: A Systematic Analysis for the Global Burden of Disease Study 2016," *The Lancet Infectious Diseases*, vol. 18, no. 11, pp. 1211–1228, 2018.
- [18] S. Acton, S. Abramowitz, L. Toledo, and G. Nitschke., "Efficiently Coevolving Deep Neural Networks and Data Augmentations," in *Proceedings of the IEEE Symposium Series on Computational Intelligence*. Canberra, Australia: IEEE, 2020, p. DOI: 10.1109/SSCI47803.2020.9308151.
- [19] P. Kumar and P. Jayagopal, "Generative Adversarial Networks: A Survey on Applications and Challenges," *International Journal of Multimedia Information Retrieval*, 2020.
- [20] V. Nannen and A. Eiben, "Efficient Relevance Estimation and Value Calibration of Evolutionary Algorithm Parameters," in *IEEE Congress on Evolutionary computation*. IEEE, 2007, pp. 103–110.
- [21] ———, "Relevance Estimation and Value Calibration of Evolutionary Algorithm Parameters," in *Proceedings of the International Joint Conference on Artificial Intelligence*. Hyderabad, India: Morgan Kaufmann, 2007, pp. 975–980.
- [22] Q. Wen, L. Sun, X. Song, J. Gao, X. Wang, and H. Xu, "Time Series Data Augmentation for Deep Learning: A Survey," *arXiv preprint arXiv:2002.12478*, 2020.
- [23] C. Esteban, S. Hyland, and G. Rätsch, "Real-valued (Medical) Time Series Generation with Recurrent Conditional GANS," *arXiv preprint arXiv:1706.02633*, 2017.
- [24] S. Jorge and J. Sevilla, "Importance of Input Data Normalization for the Application of Neural Networks to Complex Industrial Problems," *IEEE Transactions on Nuclear Science*, vol. 44, no. 3, pp. 1464–1468, 1997.
- [25] S. Borer and A. Graf, "Normalization in Support Vector Machines," in *Joint Pattern Recognition Symposium*. Munich, Germany: Springer, 2001, pp. 277–282.
- [26] R. Pelanek, "Metrics for Evaluation of Student Models," *Journal of Educational Data Mining*, vol. 7, no. 2, pp. 1–19, 2015.
- [27] M. Gevrey, I. Dimopoulos, and S. Lek, "Review and Comparison of Methods to Study the Contribution of Variables in Artificial Neural Network Models," *Ecological Modelling*, vol. 160, no. 3, pp. 249–264, 2003.
- [28] B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes*. Cambridge, UK: Cambridge University Press, 1986.
- [29] L. Taylor and G. Nitschke, "Improving Deep Learning with Generic Data Augmentation," in *Proceedings of the IEEE Symposium Series on Computational Intelligence*. IEEE Press, 2018, pp. 1542–1547.
- [30] S. Yang, X. Yu, and Y. Zhou, "LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example," in *Proceedings of the International Workshop on Electronic Communication and Artificial Intelligence*. Qingdao, China: IEEE, 2020, pp. 98–101.