SAND2017-10618C

Negative Selection Based Anomaly Detector for Multimodal Health Data

Drew Levin
Ops. Research and Comp. Analysis
Sandia National Laboratories
Albuquerque NM, USA
dlevin@sandia.gov

Melanie Moses
Computer Science
University of New Mexico
Albuquerque NM, USA
melaniem@unm.edu

Tatiana Flanagan
Computer Science
University of New Mexico
Energy Water Sys. Integration
Sandia National Laboratories
Albuquerque NM, USA
tpaz@unm.edu

Stephanie Forrest
Biodesign Inst. Center for Adaptive Computation
Arizona State University
Tempe AZ
Santa Fe Institute
Santa Fe NM, USA

Patrick Finley
Ops. Research and Comp. Analysis
Sandia National Laboratories
Albuquerque NM, USA
pdfinle@sandia.gov

***Abstract*—Early detection of emerging disease outbreaks is crucial to effective containment and response, yet initial outbreak signatures can be difficult to detect with automated methods. Outbreaks may be masked by noisy data, and signs of an outbreak may be hidden across multiple data feeds. Current biosurveillance methods often perform unimodal statistical analyses that are unable to intelligently leverage multiple correlated data of different types while still retaining quantitative sensitivity. In this paper, we propose and implement an anomaly detection system for health data based upon the human immune system. The adaptive immune system operates over a high-dimensional antigen space in a distributed manner, allowing it to efficiently scale without relying on a centralized controller. Our negative selection algorithm based on the immune system provides effective and scalable distributed anomaly detection for biosurveillance. It detects anomalies in the large, complex data from modern health monitoring data feeds with low false positive rates. Our bootstrap aggregation method improves performance on high-dimensional data sets, and we implement a parallelized version of the algorithm to demonstrate the potential to implement it on a scalable distributed architecture. Our negative selection algorithm is able to detect 90% of all outbreaks with a false positive rate of 11.8% in a publicly available multimodal synthetic health record data set. The scalability and performance of the negative selection algorithm demonstrate that immune computation can provide effective approaches for national and global scale biosurveillance.**

This work was supported by Laboratory Directed Research and Development funding from Sandia National Laboratories. MEM acknowledges the partial support of a James S. McDonnell Foundation Complex Systems Scholar Award. SF acknowledges the partial support of NSF (1518878), DARPA (FA8750-15-C-0118), AFRL (FA8750-15-2-0075), the Sandia National Laboratories Academic Alliance, and the Santa Fe Institute. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA0003525.

***Index Terms*—negative selection, anomaly detection, bootstrap aggregation, artificial immune system**

I. Introduction

Early detection of an emergent disease outbreak is crucial for a timely and cost-effective response. Signs of emergent outbreaks can be hidden inside high-dimensional data that is both noisy and incomplete. Biosurveillance detection of these events requires novel data analysis and classification techniques. The design and implementation of such detectors is an ongoing task in the field of electronic biosurveillance [1]–[3]. Current detection mechanisms often derive from established methods of time series analyses from other domains [4]–[6] and may not be best suited to deal with the complexity of modern biosurveillance data.

Current methods applied to health record surveillance rely on standard statistical approaches such as control chart algorithms [7], [8] and Bayesian Belief Networks [9]. While these methods perform well on specific types of data feeds, they often don't handle multivariate data (control charts), continuous data (Bayesian Belief Networks), and frequently do not scale well to the larger data sets available for biosurveillance.

To address these limitations, we turn to a known natural distributed anomaly detector. The adaptive immune system is able to maintain a distributed repertoire of lymphocytes that can recognize and respond to foreign pathogens while avoiding any response to healthy tissue. Applying naturally inspired algorithms in new domains is an established practice. Previous work using immune-inspired classification approaches have been successfully used to detect anomalous UNIX instructions [10], fraudulent ATM transactions [11], unauthorized intru-

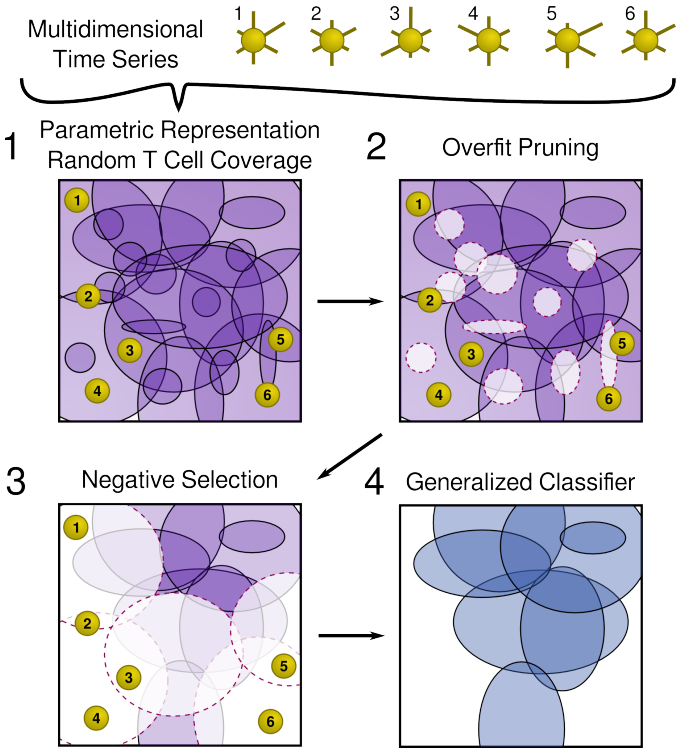

Fig. 1. **Negative Selection Algorithm.** 1) A n-dimensional time series of baseline (normal) data is transformed into its parametric representation in an n-dimensional hyperspace. A large number of n-dimensional polyhedra (detectors) are generated covering the space. 2) Polyhedra with size below a minimum threshold are removed to prevent overfitting. 3) Any polyhedron overlapping a baseline time point is removed. 4) The remaining polyhedra are those that do not overlap any observed baseline points. Future data inside this 'negative space' will be classified as anomalous.

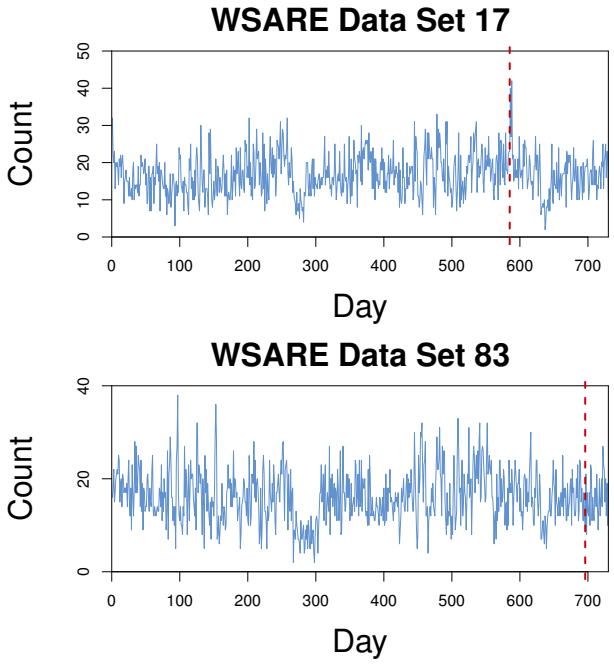

Fig. 2. **Selected WSARE timeseries.** Two randomly selected time series from the WSARE data sets. The blue line represents aggregated daily incident reports while the dashed red line marks the occurrence of an anthrax outbreak. Note that the outbreak in Series 17 represents the global maximum, while the smaller outbreak in Series 83 is not obviously anomalous.

sions [12], anomalous port scans [13], [14], and invalid online media streaming purchases [15].

The biological mechanism that generates and filters the T cell population is known as Negative Selection (NS) [16]. T cells that survive the NS process should not be able to bind to any host molecules; therefore, anything to which they bind can be considered foreign (Fig. 1). The human immune system is able to classify and respond to foreign pathogens quickly and effectively, while avoiding the detection of the host's own cells [17]. Since immune detection and response is a fully distributed, adaptive, robust, and time-sensitive process, computational implementations of immune system processes can serve as effective anomaly detectors for a variety of processes. The NS approach was implemented as computational generalized anomaly detector by Forrest and Pereleson [18]. Since then, the NS algorithm has been applied to detect novelty in time series [19], network intrusions in a Unix environment [20], and industrial tool breakage events [21].

Prior work has discussed limitations in the NS computational algorithm. Ayara et al. [22] demonstrated an exponential runtime increase in the negative selection process as the size of the training data set linearly increased. Stibor et al. [23] claimed that positive (anomalous) training examples are required for the NS algorithm to achieve adequate performance. Stibor et al. [24] showed that NS algorithms operating in high-dimensional space suffer from the ubiquitous 'curse of dimensionality' requiring ever more general detectors to cover an increasingly sparse hyperspace.

Recent advances have alleviated some of these concerns. Textor [25] used combinatorial techniques to improve the efficiency of the traditional NS algorithm such that it no longer suffers from exponential run time. Further, even without these improvements, the claim of exponential run time only holds in the case of uncorrelated data, an invalid assumption in the case of health data. As proof, Textor [26] demonstrated improved and globally competitive NS performance on a range of traditional anomaly detection tasks using his improved techniques. Prasad and Ghosh [27] propose a method to identify key variables to better reduce the dimensionality of the problem space. Yang et al. [28], [29] present an improvement to the NS algorithm that helps detectors target a more efficient low-dimensional subspace. Zhu et al. take advantage of the Map-Reduce algorithm to reduce the the exponential complexity of the NS algorithm to logarithmic [30].

Here, we extend the NS algorithm to detect anomalies in health records consisting of multimodal data types. Our implementation simultaneously handles both continuous and categorical data. Further, the distributed nature of the NS algorithm allows us to parallelize our implementation for significantly improved performance across multiple CPUs. Finally, we implement a form of bootstrap aggregation to help

improve scalability. Our results show that the NS algorithm is able to detect anomalies in health reports with few false positives.

II. Data

Due to privacy concerns it is difficult to compare different anomaly detection algorithms with a common real-world dataset. Therefore, we trained and evaluated our NS implementation on the publicly available WSARE synthetic data set. The WSARE data set is hosted by the Auton Lab at https://www.autonlab.org/datasets. The data consists of 100 unique time series representing two years of data each with a simulated anthrax outbreak [31].

The data contains lists of health care events as described in Table I. An event is defined as either a clinical visit, a purchase of medication, or an irregular absence from school or work. Each data set contains approximately 25,000 individual records, a subset of which are caused by a simulated anthrax infection event. Records contain a coarse spatial component consisting of a single location on a 3×3 grid along with general symptom information. Because each record is an independent event, daily counts were obtained by aggregating individual records (Fig. 2). For the purposes of our experiments, we transformed the individual datasets by aggregating individual records into daily frequency counts for none, rash, respiratory, and nausea events. This transformation resulted in the removal of categories that were not consistent across the aggregation step. Therefore we removed columns 1 (XY), 2 (age), 3 (gender), 8 (action), and 10 (drug), and combined 11 and 12 into a single day of the year column. This resulted in a nine-dimensional data set of the following columns: flu prevalence, day of the week, weather, season, day of the year, aggregate none, aggregate nausea, aggregate rash, and aggregate respiratory.

The WSARE data are public, allowing researchers to compare outbreak detectors on the same data. The data contain multivariate sources including patient demographics, seasonal data, and a variety of reported symptoms. Each time series contains single labeled outbreak events of various sizes with which to test detection algorithms. While the data do not contain missing records, the dataset attempts to reflect the random and noisy nature of real biosurveillance time series.

ID	Label	Description
1	XY	Spatial region of record (3×3 grid)
2	age	Patient age: child, working, or senior
3	gender	Patient gender
4	flu	Flu prevalence: none, low, high, or decline
5	day_of_week	Saturday, Sunday, or a weekday
6	weather	Hot or cold
7	season	Winter, Spring, Summer, or Fall
8	action	Record type: purchase, evisit, or absent
9	reported_symptom	None, respiratory, nausea, or rash
10	drug	Drugs: none, nyquil, apririn, or anti-vomit
11	date	From Jan-01-2002 to Dec-31-2003
12	daynum	Date converted to a single integer index

TABLE I

Fields included in a single WSARE record. 100 unique data sets exist. A single data set spans two years and contains multiple records per day. Each data set contains one simulated anthrax outbreak.

III. Implementation

A. Negative Selection Algorithm

We implement a mechanistic version of the biological NS algorithm for use as a biosurveillance detector (Fig. 1). First, detectors are generated at random to cover the full space of possible data points.

Once generated, each detector is tested against the selected training set of baseline data known not to contain anomalies. A detector reacts with a single data point if that point lies within the detector's boundaries in every dimension. If a detector reacts to any baseline point, the detector is removed from the population (Fig. 1, Lower Left). This process mimics negative selection in natural immune systems and ensures that the remaining detector population does not react with any data previously seen.

Once training is complete, test data overlapping any of the surviving detectors is marked as anomalous. To achieve a numeric metric, each data point of the test set is scored according to how many detectors react to it. Because the remaining detectors survived both overfit pruning and negative selection, we assert they can be considered both appropriately specific and general.

B. Detecting Anomalies Across Multiple Types of Data

The artificial NS classifier identifies anomalies within multidimensional time series data. NS-based anomaly detectors provide an independent range for each data dimension defined in the data and will recognize a data point if the point lies within the range defined for each dimension. Each input dimension contains one of three possible data types (Table II):

- **Quantitative**: Quantitative data are numerical values where the quantity is of interest. An example would be the number of patient visits to a local clinic each day. Quantitative detectors maintain a single threshold value and react with a quantitative value if the value is above the that threshold.
- **Identifier:** Identifier data refer to sequential non-quantitative numerical values. An example of an identifier datum would be the day of the month. Because the quantity itself is not of interest detectors maintain a two-sided range of values. A detector will react with a specific point if the value is inside the chosen range.
- **Category:** Categorical data are anything non-numeric. Categorical data are defined by the set of all possible items in the category. An example of categorical data would be a list of possible disease symptoms, such as rash, nausea, diarrhea, and congestion. A detector stores a fixed subset of the possible categories and reacts with a categorical item if that item is contained in the detector's subset.

To create a detector set, detectors are generated randomly such that they contain unique boundaries in each dimension. To generate random detectors, first a training data set with no

Type	Format	Boundary	Example
Quantitative	`numeric`	threshold	clinic visits
Identifier	`numeric`	range	day of month
Category	`text`	subset	symptom list

TABLE II
NS DATA TYPES THE NEGATIVE SELECTION ALGORITHM CAN SIMULTANEOUSLY OPERATE ON MULTIPLE TYPES OF DATA.

known disease outbreaks is examined. Detector values are chosen uniformly from the upper and lower ranges of numerical data, and sampled without replacement from categorical data. Identifier specificity is constrained to be within a min and max numerical range of the generated center value and categorical specificity is set as the size of the Detector's subset of possible items. The minimum size constraint on ranges ensures that generated detectors are appropriately general (Fig. 1, Top Right).

C. Distributed Implementation

The NS Algorithm consists of two distributed stages. The first involves independently training a large number of detectors through the negative selection process. The second involves testing new data by independently evaluating the data versus the previously generated detector set. Training and evaluation implementations can be distributed across the individual detectors to improve throughput. We implemented a parallel version of the NS algorithm in Python 3.5.2, using the `multiprocessing` library to take advantage of multiple computer cores.

During the training phase a specified number of new processes are created such that each processing core independently generates a large number of detector candidates and evaluates each one against the training data set which is maintained in shared memory. Once the appropriate number of valid detectors have been generated, a signal is sent to each process to halt.

Parallel processing during the evaluation phase is implemented similarly. Because the total number of data points to be evaluated is known *a priori*, we split the test data into subsets of equal size and evaluate each subset with a unique processor. In this instance, we place the detector repertoire in shared memory as it is used statically by each process. Once each process completes the evaluation of its subset, the disjoint subsets are concatenated back in order for final use.

Hyperparameter	Setting
Detector Count	10,000
Minimum Range (Numeric)	10%
Maximum Range (Numeric)	75%
Minimum Size (Set)	1
Maximum Size (Set)	$\|S\| - 1$
Bagging Dimensions	4 of 9
Distance Norm	$\mathcal{L}^\infty$ (Rectangular)
Processor Count	1, 2, 4, 8, 16

TABLE III
NS HYPERPARAMETERS. CONFIGURATION SETTINGS USED IN OUR IMPLEMENTATION OF THE NS ALGORITHM

D. Bootstrap Aggregation

Bootstrap aggregation, or bagging, is the process of using multiple weak classifiers on bootstrapped subsamples of data to reduce classification variance of a volatile data set [32]. The native NS algorithm can be thought of a an *ensemble* method: it combines multiple weak classifiers (detectors) to achieve a strong classification over the problem space.

Applying the bagging approach helps alleviate the curse of dimensionality that plagues spatial NS implementations in high-dimensional spaces. For example, if a randomly generated detector spans half of the possible range of each dimension in the problem space, a single detector will cover one quarter of the area of a two-dimensional space, but only one-tenth of one percent of the hypervolume of a ten dimensional space. Thus, as the NS algorithm is extended into higher dimensional problem sets, detectors must either drastically increase in size, or the total number of detectors must be increased exponentially.

Rather than having each detector operate over every possible dimension, individual detectors randomly select a small subset of dimensions (with replacement) and only evaluate data in terms of those dimensions. Continuing the previous example, a two-dimensional detector operating on a ten-dimensional problem space would still cover one quarter of the total hypervolume (assuming the detector covered half of the range of each of its two sampled dimensions). It is assumed detectors react with the full range across any dimension they do not explicitly cover. Allowing each individual detector to select its own random subset of dimensions helps avoid potential blind spots arising from the dimensionality flattening process. Bagging allows detectors to cover a high-dimensional space without requiring exponential computational complexity or extreme detector generalization.

E. Hyperparameter Choices and Configuration

Our implementation of the NS algorithm requires a number of hyperparameters, inputs that determine the size, granularity, and complexity of an analysis run. Hyperparameters include the number of valid detectors to generate, the minimum and maximum sizes of the detectors in each dimension, the number of sampled dimensions for the bagging process, and which spatial distance norm to use. For the purposes of this paper, we used values that performed well in experimentation (Table III). Specifically, we limited spatial detector size to be between 10% and 75% of a dimension's full range and set size to be between size one and one less than the total number of categories. Detectors sampled from four of the possible nine dimensions included in our transformed WSARE data set (with replacement). Finally, detectors were represented as $\mathcal{L}^\infty$ hyper-rectangles (as opposed to $\mathcal{L}^2$ Euclidean hyper-ellipses or $\mathcal{L}^1$ Manhattan diamond configurations).

IV. RESULTS

A. Performance

We evaluated our NS implementation across the full WSARE data set. For each for the 100 example datasets we

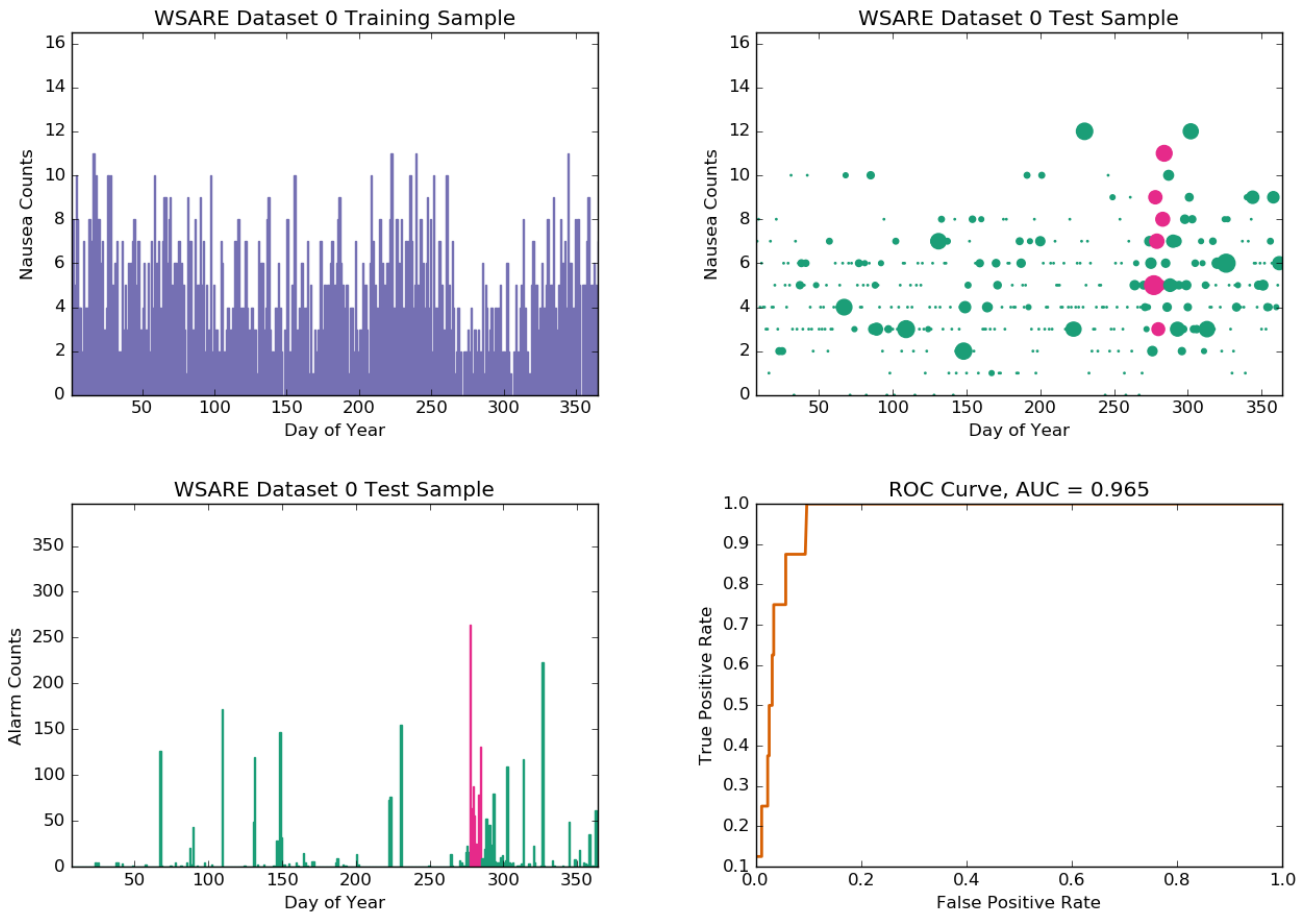

Fig. 3. **Application of negative selection for anomaly detection on a sample dataset.** Upper Left: Nausea counts of the WSARE dataset as an example dimension. Upper Right: The generated detectors were applied to the second year of the dataset. A true anomalous outbreak occurs at day 277 and is shaded pink. The size of each data point represents the number of detectors that overlap the point. Lower Left: The alarm rate for each day of the test set, analogous to the size of the data points in the upper right plot. Right: The ROC curve for the generated alarm rate as compared to the true outbreak.

trained the NS algorithm on the first full year of data and evaluated the performance on the second full year (Fig. 3). Once detectors are generated on the training set, each point in the test set is scored against the detector repertoire based on how many detectors overlap the point. Each example contains one simulated outbreak in the second year, thus we judge the performance of our algorithm by its ability to detect the outbreak within one week of the initial occurrence. Ideally, the algorithm would detect the outbreak with no false positives. A threshold can be placed over the alarm rates to convert the numeric scores into a binary signal. The sensitivity and specificity of the possible threshold values was evaluated by a ROC curve for illustrative purposes.

Because each data set contains one guaranteed outbreak we evaluate the performance of our implementation in terms of the number of false alarms required in order to detect the true positive at least once within the first seven days. The sensitivity of the algorithm can be controlled by setting the level of the signaling threshold. Due to the variability of the individual WSARE data sets, some outbreaks can be easily detected by our implementation, while a small number are not detected at all (Fig. 4A). At an aggregate level, we evaluate our performance by calculating the required beta (false positive) error rate necessary to obtain a desired alpha (false negative) error rate (Fig. 4B). Our results show that to achieve a 10% alpha error rate (missing 10 outbreaks out of 100) requires a 11.8% beta error rate.

B. Parallel Processing

We evaluated the parallelization of our implementation on a Xeon E7 v3 processor with 12 hyperthreaded CPU cores (resulting in 24 logical cores). We define throughput as the number of valid detectors generated per second. Throughput was evaluated by generating 10,000 valid detectors over a single WSARE data set as well as the entire data set. The size of the data set appears to have no significant impact on the throughput. Our results show a distinct performance improvement as the NS algorithm is applied to an increasing number of processing cores (Fig. 5). Note that Fig. 5 is logarithmically

Fig. 4. **Aggregate Performance on WSARE.** Left: Each individual WSARE dataset is evaluated based upon the number of false positives required in order to detect the simulated outbreak. The data sets are displayed in sorted order. Middle: The trade-off between the allowable alpha error (false negatives) and the resulting required beta error (false negatives). For example, an alpha error of 10% results in a required false positive rate of 11.8% (dashed lines). Right: The entire 100 WSARE datasets displayed in terms of the alarm rate for each day.

scaled in both dimensions, so the linear trend with near 0.83 slope suggests approximate $O(n^{5/6})$ performance scaling.

We hypothesize that the sublinear scaling is due in large part to the inability of the Python implementation to make concurrent use of shared read-only memory. Because we use the `multiprocessing` library's internal shared memory manager, we are subject to process exclusivity constraints when accessing shared memory. Further, due to Python's Global Interpreter Lock, we are not able to take advantage of a more light-weight threaded implementation. Implementation using a language that supports concurrent memory access using threading could potentially improve performance scaling. Our results also show a decreasing trend as the processor count increases to 16, which we suspect is due to the inability of the 12 hyperthreaded cores to perform full computations in parallel.

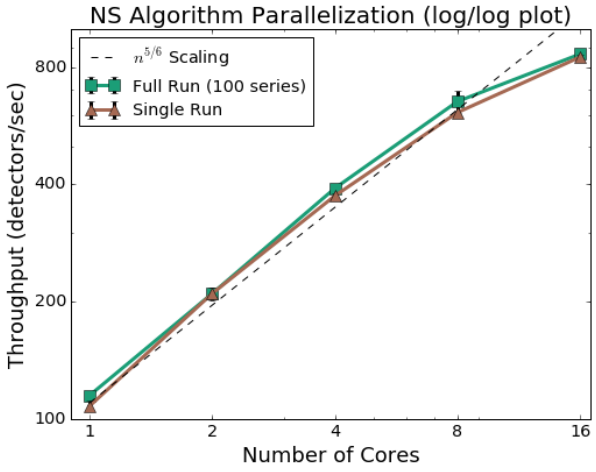

Fig. 5. **Parallelization.** 10,000 valid detectors were generated covering the space defined by the full WSARE dataset and a single WSARE time series. Each sample was replicated five times. Error bars show plus and minus two standard errors from the mean. Evaluation was performed on a 12-core Xeon E7 v3 processor (24 logical hyperthreaded cores).

V. Conclusions

Anomaly detection within multimodal health records requires methods that can operate over a wide class of input types. Due to the expansive size and anticipated growth rate of modern biosurveillance data feeds, any potential approach must lend itself well to distributed computation. The human immune system is a naturally occurring distributed anomaly detector and a natural inspiration for computational application to the health record domain.

We present a novel implementation of the negative selection algorithm and demonstrate its utility by testing it against a realistic data set consisting of a wide variety of data types. Our implementation is able to operate over these disparate data types simultaneously in a 9-dimensional space. Performance on higher dimensional data is aided by the use of bootstrap aggregation, allowing individual detectors to operate over lower dimensional regions thus avoiding the so-called curse of dimensionality. Finally, we demonstrate the benefits of the distributed nature of the negative selection process by allowing our algorithm to make simultaneous use of multiple computer cores, which results in a near linear increase in computational performance. The end result is a robust anomaly detection algorithm able to operate efficiently and accurately over complex data sets inaccessible to traditional statistical methods.

The negative selection based anomaly detection methods demonstrated in this paper hold great potential to improve accuracy and timeliness of national-scale electronic biosurveillance. Currently, electronic biosurveillance for the US operates through CDCs National Syndromic Surveillance Program (NSSP) [33]. Hospital emergency department admission data

is funneled to CDC in near-real-time to provide early notice of disease outbreak, but outbreak determination is often reliant on single-variable, non-adaptive statistical algorithms. Recently published review of immediate needs for large scale biosurveillance [34] lists "Methods and systems to support the fusion of various types of data" and "Enhanced and adaptive detection algorithms" as being the most important research priorities for the coming decade. The negative selection methods described in this paper address both of these identified gaps. Additionally, the documented performance gains possible from the novel application of bagging and parallel processing to NS anomaly detection directly addresses prior concerns that multi-dimensional NS approaches cannot scale sufficiently to tackle large scale problems.

REFERENCES

[1] G. Shmueli and S. E. Fienberg, "Current and potential statistical methods for monitoring multiple data streams for biosurveillance," in *Statistical Methods in Counterterrorism*. Springer, 2006, pp. 109–140.

[2] S. Unkel, P. Farrington, P. Garthwaite, C. Robertson, and N. Andrews, "Statistical methods for the prospective detection of infectious disease outbreaks: a review," *Journal of Royal Statistical Society: Series A*, vol. 175, no. 1, pp. 49–82., 2012.

[3] K. N. Gajewski, A. E. Peterson, R. A. Chitale, J. A. Pavlin, K. L. Russell, and J. P. Chretien, "A review of evaluations of electronic event-based biosurveillance systems," *PLoS ONE*, vol. 9, no. 10, pp. 7–10, 2014.

[4] A. Goldenberg, G. Shmueli, R. A. Caruana, and S. E. Fienberg, "Early Statistical Detection of Anthrax Outbreaks by Tracking Over-the-Counter Medication Sales on JSTOR," pp. 5237–5240, 2002.

[5] K. E. Cheng, D. J. Crary, J. Ray, and C. Safta, "Structural models used in real-time biosurveillance outbreak detection and outbreak curve isolation from noisy background morbidity levels," *Journal of the American Medical Informatics Association*, pp. 435–440, 2012.

[6] G. Shmueli, "Wavelet-Based Monitoring for Biosurveillance," *Axioms*, vol. 2, no. 3, pp. 345–370, 2013.

[7] A. P. Morton, M. Whitby, M.-L. McLaws, A. Dobson, S. McElwain, D. Looke, J. Stackelroth, and A. Sartor, "The application of statistical process control charts to the detection and monitoring of hospital-acquired infections," *Journal of Quality In Clinical Practice*, vol. 21, no. 4, pp. 112–117, dec 2001.

[8] W. H. Woodall, M. A. Mohammed, J. M. Lucas, and R. Watkins, "The Use of Control Charts in Health-Care and Public-Health Surveillance," *Journal of Quality Technology*, vol. 38, no. 2, p. 89, 2006.

[9] H. S. Burkom, L. Ramac-Thomas, S. Babin, R. Holtry, Z. Mnatsakanyan, and C. Yund, "An integrated approach for fusion of environmental and human health data for disease surveillance." *Statistics in medicine*, vol. 30, no. 5, pp. 470–9, feb 2011.

[10] S. Forrest, S. Hofmeyr, A. Somayaji, and T. Longstaff, "A sense of self for Unix processes," in *Proceedings 1996 IEEE Symposium on Security and Privacy*. IEEE Comput. Soc. Press, 1996, pp. 120–128.

[11] M. Ayara, J. Timmis, R. de Lemos, and S. Forrest, "Immunising Automated Teller Machines," in *Artificial Immune Systems*, 2005, vol. 3627, pp. 404–417.

[12] J. Greensmith, J. Twycross, and U. Aickelin, "Dendritic Cells for Anomaly Detection," in *2006 IEEE International Conference on Evolutionary Computation*. IEEE, 2006, pp. 664–671.

[13] J. Greensmith and U. Aickelin, "The deterministic dendritic cell algorithm," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5132 LNCS, pp. 291–302, 2008.

[14] J. Greensmith, U. Aickelin, and G. Tedesco, "Information fusion for anomaly detection with the dendritic cell algorithm," *Information Fusion*, vol. 11, no. 1, pp. 21–34, jan 2010.

[15] R. Huang, H. Tawfik, and A. Nagar, "On the use of innate and adaptive parts of artificial immune systems for online fraud detection," in *2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA)*. IEEE, sep 2010, pp. 1669–1676.

[16] G. J. V. Nossal, "Negative selection of lymphocytes," pp. 229–239, jan 1994.

[17] C. A. Janeway, "How the immune system works to protect the host from infection: a personal view." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 13, pp. 7461–7468, 2001.

[18] S. Forrest, A. Perelson, L. Allen, and R. Cherukuri, "Self-nonself discrimination in a computer," *Proceedings of 1994 IEEE Computer Society Symposium on Research in Security and Privacy*, pp. 202–212, 1994.

[19] D. Dasgupta and S. Forrest, "Novelty Detection in Time Series Data using Ideas from Immunology," in *Proceedings of The International Conference on Intelligent Systems*, 1995.

[20] S. A. Hofmeyr and S. Forrest, *Architecture for an Artificial Immune System*, 2000, vol. 8, no. 4.

[21] D. Dasgupta and S. Forrest, "Artificial immune systems in industrial applications," in *Proceedings of the Second International Conference on Intelligent Processing and Manufacturing of Materials. IPMM'99 (Cat. No.99EX296)*. IEEE, 1999, pp. 257–267 vol.1.

[22] M. Ayara, J. Timmis, R. de Lemos, L. N. de Castro, and R. Duncan, "Negative selection: How to generate detectors," *Proceedings of the 1st International Conference on Artificial Immune Systems (ICARIS)*, vol. 1, pp. 89–98, 2002.

[23] T. Stibor, P. Mohr, J. Timmis, and C. Eckert, "Is negative selection appropriate for anomaly detection?" *Proceedings of the 2005 conference on Genetic and evolutionary computation*, pp. 321–328, 2005.

[24] T. Stibor, J. Timmis, and C. Eckert, "On the use of hyperspheres in artificial immune systems as antibody recognition regions," in *ICARIS*, 2006, pp. 215–228.

[25] J. Textor, "Efficient negative selection algorithms by sampling and approximate counting," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7491 LNCS, no. PART 1. Springer Berlin Heidelberg, 2012, pp. 32–41.

[26] ——, "A comparative study of negative selection based anomaly detection in sequence data," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7597 LNCS. Springer Berlin Heidelberg, 2012, pp. 28–41.

[27] J. V. Prasad and K. Ghosh, "Negative selection algorithm for monitoring processes with large number of variables," in *Control Applications (CCA), 2014 IEEE Conference on*. IEEE, 2014, pp. 778–783.

[28] T. Yang, W. Chen, and T. Li, "A real negative selection algorithm with evolutionary preference for anomaly detection," *Open Physics*, vol. 15, no. 1, pp. 121–134, 2017.

[29] ——, "An antigen space density based real-value negative selection algorithm," *Applied Soft Computing*, 2017.

[30] F. Zhu, W. Chen, H. Yang, T. Li, T. Yang, and F. Zhang, "A quick negative selection algorithm for one-class classification in big data era," *Mathematical Problems in Engineering*, vol. 2017, 2017.

[31] W.-k. Wong, A. Moore, G. Cooper, and M. M. Wagner, "Bayesian Network Anomaly Pattern Detection for Disease Outbreaks," in *Proceedings of the Twentieth International Conference on Machine Learning*, T. Fawcett and N. Mishra, Eds. Menlo Park, California: AAAI Press, 2003, pp. 808–815. [Online]. Available: http://www.autonlab.org/autonweb/14642.html

[32] L. Breiman, "Bagging Predictors," *Machine Learning*, vol. 24, no. 421, pp. 123–140, 1996.

[33] "US Centers for Disease Control: National Syndromic Surveillance Program (NSSP)," 2017. [Online]. Available: https://www.cdc.gov/nssp/index.html

[34] R. S. Hopkins, C. C. Tong, H. S. Burkom, J. E. Akkina, J. Berezowski, M. Shigematsu, P. D. Finley, I. Painter, V. J. Del Rio Vilas, and L. C. Streichert, "A Practitioner-Driven Research Agenda for Syndromic Surveillance," *Public Health Reports*, vol. 132, no. 1_suppl, pp. 116S–126S, jul 2017.