HHS Public Access

Author manuscript

Birth Defects Res. Author manuscript; available in PMC 2021 November 01.

Published in final edited form as:

Birth Defects Res. 2020 November ; 112(18): 1450–1460. doi:10.1002/bdr2.1767.

Exploratory analysis of machine learning approaches for surveillance of Zika-associated birth defects

Richard Lusk[1], **John Zimmerman**[1], **Kelley VanMaldeghem**[1], **Suzanna Kim**[1], **Nicole M. Roth**[2], **James Lavinder**[1], **Anna Fulton**[2], **Meghan Raycraft**[1], **Sascha R. Ellington**[3], **Romeo R. Galang**[4]

[1]Deloitte Consulting, LLP, Atlanta, GA

[2]Eagle Medical Services, LLC, Alpharetta, GA

[3]Division of Reproductive Health, National Center for Chronic Disease Prevention and Health Promotion, US Centers for Disease Control and Prevention, Atlanta, GA

[4]Division of Birth Defects and Infant Disorders, National Center on Birth Defects and Developmental Disabilities, US Centers for Disease Control and Prevention, Atlanta, GA

Abstract

In 2016, Centers for Disease Control and Prevention (CDC) established surveillance of pregnant women with Zika virus infection and their infants in the U.S. states, territories, and freely associated states. To identify cases of Zika-associated birth defects, subject matter experts review data reported from medical records of completed pregnancies to identify findings that meet surveillance case criteria (manual review). The volume of reported data increased over the course of the Zika virus outbreak in the Americas, challenging the resources of the surveillance system to conduct manual review. Machine learning was explored as a possible method for predicting case status. Ensemble models (using machine learning algorithms including support vector machines, logistic regression, random forests, *k*-nearest neighbors, gradient boosted trees, and decision trees) were developed and trained using data collected from January 2016–October 2017. Models were developed separately, on data from the U.S. states, non-Puerto Rico territories, and freely associated states (referred to as the U.S. Zika Pregnancy and Infant Registry [USZPIR]) and data from Puerto Rico (referred to as the Zika Active Pregnancy Surveillance System [ZAPSS]) due to differences in data collection and storage methods. The machine learning models demonstrated high sensitivity for identifying cases while potentially reducing volume of data for manual review (USZPIR: 96% sensitivity, 25% reduction in review volume; ZAPSS: 97% sensitivity, 50% reduction in review volume). Machine learning models show potential for identifying cases of Zika-associated birth defects and for reducing volume of data for manual review, a potential benefit in other public health emergency response settings.

Correspondence John Zimmerman, Deloitte Consulting, LLP, 191 Peachtree Street, Atlanta, GA, 30303, USA. jzimmerman@deloitte.com.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

Keywords

birth defects; machine learning; surveillance; Zika virus

1 | INTRODUCTION

In 2015, Zika virus emerged as a cause of serious defects of the brain and eye and has been associated with neurodevelopmental abnormalities, such as seizures, joint contractures, swallowing difficulties, vision impairments, and hearing loss in offspring of women with Zika virus infection during pregnancy (Honein et al., 2017; Moore et al., 2017; Rasmussen, Jamieson, Honein, & Petersen, 2016; Reynolds et al., 2017; Shapiro-Mendoza et al., 2017; Rice et al., 2018). In the setting of a public health emergency response, Centers for Disease Control and Prevention (CDC) collaborated with state, territorial, and local health departments to establish a national surveillance system of women with laboratory evidence of confirmed or possible Zika virus infection during pregnancy and their infants (Reynolds et al., 2017; Shapiro-Mendoza et al., 2017). Over 7,400 pregnancies completed between December 1, 2015–March 31, 2018 were monitored to better understand the effects of Zika virus infection during pregnancy on women, infants, and children. About 5–10% of infants of women with confirmed Zika virus infection during pregnancy were found to have Zika-associated birth defects (Reynolds et al., 2017; Shapiro-Mendoza et al., 2017).

As part of surveillance, information was abstracted from medical records and reported by U.S. states, territories, and freely associated states. A rule-based process was used to screen data from completed pregnancies for possible abnormal findings. Each completed pregnancy with possible abnormal findings was prioritized for review by subject matter experts at CDC (manual review) to identify those completed pregnancies with findings that met surveillance criteria for Zika-associated birth defects (cases) and to distinguish them from completed pregnancies without these findings (non-cases; Reynolds et al., 2017; Shapiro-Mendoza et al., 2017).

Manual review is considered a gold standard method for distinguishing cases from non-cases (case status). This process involves consideration of many data points for each completed pregnancy, including findings from infant physical examination, neuroimaging, hearing screening, and eye examination. Manual review of each completed pregnancy with a possible abnormality maximizes sensitivity for detecting cases at the expense of both time and human resources for conducting the review. As the Zika virus outbreak in the Americas progressed, the volume of data reported to CDC increased, challenging the resources of the surveillance system for conducting manual review.

Automation of systematic surveillance processes by supervised machine learning (ML) and natural language processing (NLP) methods might present an opportunity to conduct more efficient manual review. ML/NLP methods create classification decisions based on patterns derived from existing data, and these classification decisions can be automated and applied to future data. ML has been proposed and explored in other clinical scenarios for predicting therapeutic outcomes, identifying abnormalities in radiology, and identifying risk in chronic disease (Kang, Schwartz, Flickinger, & Beriwal, 2015; Wall, Kosmicki, Deluca, Harstad, &

Fusaro, 2012; Wei, Yang, Nishikawa, & Jiang, 2005). For surveillance, results of previous manual reviews might be used to guide development of ML/NLP methods to predict case status algorithmically, thereby augmenting the processes of screening and manual review, reducing overall review volume, and potentially improving overall surveillance system performance in a public health emergency setting (Lee et al., 2018; Xiong et al., 2018).

Using the results of manual review by subject matter experts as the gold standard, we explored the use of ML/NLP processing and variable feature engineering methods for predicting case status. We estimated the volume of non-cases predicted by ML/NLP methods and the potential reduction in the volume of completed pregnancies undergoing manual review.

2 | METHODS

Figure 1 presents a basic flow chart of our methods approach to exploring a ML application for the Zika Case Review Process

2.1 | Data source

Pregnancies with laboratory evidence of confirmed or possible Zika virus infection were followed to pregnancy completion and into early childhood through active surveillance methods. Data on pregnancy, birth outcome, and findings from clinical evaluations of the infant were abstracted using standardized methods from prenatal, birth hospitalization, pediatrician, and specialty care medical records and reported to CDC. Data were reviewed to identify cases meeting the CDC surveillance criteria of Zika-associated birth defects. A trained reviewer scrutinized each completed pregnancy with data indicating a possible abnormality (from rules-based process), and review results were discussed with a panel of subject matter experts for final case classification.

Data collected from Puerto Rico was stored and analyzed separately due to differences in the amount and format of data collected. Data collected from the U.S. states, non-Puerto Rico territories, and freely associated states are referred to as the U.S. Zika Pregnancy and Infant Registry (USZPIR) in this analysis, and data collected from Puerto Rico are referred to as the Zika Active Pregnancy Surveillance System (ZAPSS).

Abstracted medical record data were included for 7,155 completed pregnancies (3,212 from the USZPIR; 3,943 from ZAPSS) which were reported to CDC from January 1, 2016 to October 31, 2017. Data from manual reviews completed by the time of this analysis were included for model development (324 from USZPIR and 523 for ZAPPS); data for remaining completed pregnancies were used for model validation exploration. Abstracted data included both non-text variables, such as dates of events (e.g., date of maternal symptoms, diagnosis, birth, and clinical evaluations), numerical values (e.g., head circumference measurements), and categorical information for outcomes (e.g., pregnancy outcome and birth defects), and open-text reporting fields (e.g., verbatim descriptions of results from neuroimaging studies).

All independent variables from USZPIR and ZAPSS were initially considered for model development and testing. Variables were excluded from model development if they were not used for case status determination. For example, infant date of birth, estimated date of delivery, city of birth, and name of birth hospital did not provide useful information for determining if a Zika-associated birth defect is present or absent in an individual case. Included variables were standardized into binary and categorical variables. From USZPIR, a total of 198 non-text variables and 19 open-text reporting fields were included for analysis. From ZAPSS, 800 non-text variables and 142 open-text reporting fields were included for analysis.

Case status was designated using the 2017 CDC surveillance criteria for Zika-associated birth defects, which at the time included structural abnormalities of the brain and eye with and without microcephaly, neural tube defects and other early brain malformations, and consequences of central nervous system dysfunction (Honein et al., 2017; see Box 1).

2.2 | Natural language processing

NLP techniques were used to derive quantitative data (variable features) from qualitative data contained in open-text variables to be used for ML model development and training.

First, open-text data were standardized using (a) text cleaning by converting to all lower-case characters; (b) removal of punctuation, numbers, whitespace, and common English words; and (c) root-word reduction by stemming technique. Standardized open-text data were then processed into variable features using term frequency-inverse document frequency (TF-IDF), a technique which estimates the importance of a word by taking into account the number of times the word appears in open-text, the total number of words in open-text, and the number of the cases in which the word appears. These variable features were then amended to data for each case to create an ML training data set (matrix). For each open-text variable, a TF-IDF metric was created and appended to the matrix of non-text variables. Bi-grams and tri-grams did not improve model performance and were removed in the matrices.

Variable features which summarize large numbers of open-text data were created using Latent Dirichlet Allocation (LDA) and document word embeddings (Doc2Vec; Li et al., 2018). LDA is a form of topic modeling that attempts to identify general topics from a body of text data. Word embeddings are a general approach to quantify mathematically the meaning of a word based upon adjacent words. Combining the two techniques has been previously demonstrated to substantially improve model performance (Pratapa et al., 2018).

Custom variable features were created to group data from similar text features (e.g., all text for the hearing exam, all text related to ultrasound exams). Due to the large number of variable features from TF-IDF with a value of zero (i.e., no data), a truncated singular value decomposition technique was applied to reduce both the overall sparsity (measure of number of cells in a matrix with a value of zero relative to number of individual data points) and size of the matrix.

In ML, a feature vector is a series of individual data points about an object which can be represented spatially; feature vectors are the equivalent of vectors of explanatory variables

that are used in statistical procedures such as linear regression. In this analysis, final features vectors were normalized by L2 Euclidean normalization in order to have all features uniformly scaled. Not all text features were included in the model training phase. Each model was structured to reduce dimensionality (i.e., obtaining a set of principal variables to reduce the number of random variables under consideration) and keep only features that significantly contribute to explaining variation based upon a predetermined threshold.

2.3 | ML methods

A binary classification of case versus non-case was used as the dependent variable for ML model development. Models were trained using data sets of completed pregnancies that had been prioritized for manual review and received a designation of case versus non-case by the time of this analysis (USZPIR: 324 of 3,212 cases [10.1%]); ZAPSS: 523 of 3,943 cases [13.2%]).

First, exploratory of ML methods focused on reducing the number of non-cases that would undergo manual review. The ML models were then calibrated to maximize the number of cases of completed pregnancies with a Zika-associated birth defect identified by models. Because ascertainment of cases was prioritized, models that missed cases identified by manual review were considered invalid for further applications.

A fivefold cross-validation approach was used to train and test models including voting ensemble models (Hastie, Tibshirani, Friedman, & Franklin, 2005). Model sensitivity was calculated as the percentage of completed pregnancies identified as cases by manual review that were predicted as cases by the model. Review depth was defined as the percentage of completed pregnancies that were prioritized by the ML model to undergo manual review process to achieve a chosen model sensitivity threshold. Review depth was the predetermined factor used for efficiency gains, and the corresponding model sensitivity was used to evaluate model results, with the objective of achieving 100% sensitivity for predicting a case also identified by manual review while minimizing review depth. The potential percent reduction in review volume at a given model threshold was defined as the percentage of completed pregnancies identified as non-cases by manual review that were not prioritized by the ML model to undergo manual review.

ML algorithms (support vector machines [SVM], logistic regression, random forests [RF], *k*-nearest neighbors [KNN], gradient boosted trees [GBT], and decision trees) were assessed individually and then developed into a voting ensemble model for each data set. A voting ensemble model was explored to not rely on one algorithm as algorithm performance may vary across use cases of ML as data features change. Ensemble model weighting schemes were tested iteratively with weights of .1 to .9 given to each model and iterated to maximize sensitivity. Probability thresholds for results were then explored to improve ensemble model sensitivity and review depth. Potential percent reduction in review volume was calculated for voting ensemble models at sensitivities ≥94% (see Appendix for individual model parameters). Sensitivity and review depth were compared between ensemble and individual models to identify gains or losses in improvement (i.e., did the ensemble do better or worse than individual models).

2.4 | Ensemble model validation

Model validation was performed by applying the ensemble model with 100% sensitivity to USZPIR and ZAPSS data that were not prioritized for manual review. The model was expected to predict a low number of cases from these data, which had been screened away from prioritized manual review as previously described because they either had only normal findings among reported data, data uninformative for case status classification, or missing data. Completed pregnancies predicted as cases by the ensemble model (possible cases) were then manually reviewed by subject matter experts to assign case status.

3 | RESULTS

3.1 | U.S. Zika Pregnancy and Infant Registry

At a model prediction threshold of .95 (wherein model-identified cases have a probability of ≥.95 of being a case identified by manual review), individual model sensitivity using fivefold cross-validation was greatest for logistic regression (sensitivity 32%; review depth 17%), followed by GBT (sensitivity 29%, review depth 15%), RF (sensitivity 8%, review depth 4%), and KNN (sensitivity 5%, review depth 4%; Table 1). Individual review depth to achieve maximum sensitivity varied from 87% (RF) to 100% (SVM). The KNN model did not reach 100% sensitivity and was not included in the voting ensemble model. SVM, RF, logistic regression, and GBT were included in the voting ensemble model and were given equal weighting.

Compared to a scenario in which all completed pregnancies (cases and non-cases) are manually reviewed to achieve maximum sensitivity (sensitivity 100%; review depth 100%), the USZPIR voting ensemble model achieved 100% sensitivity with a review depth of 91% (Table 2). In this scenario, all of cases and 82% of non-cases would be manually reviewed—a potential 18% reduction in volume of non-cases for manual review. The USZPIR voting ensemble model achieved 96% sensitivity with a review depth of 75%. In this scenario, 96% of cases and 50% of non-cases would be manually reviewed—a potential reduction of 50% in non-cases for manual review. Compared to individual model performances, the voting ensemble review depth at 100% sensitivity was greater than GBT individually.

3.2 | Zika Active Pregnancy Surveillance System

At a model prediction of threshold .95, individual model sensitivity using fivefold cross-validation was greatest for decision trees (sensitivity 52%, review depth 14%), followed by logistic regression (sensitivity 28%, review depth 8%), GBT (sensitivity 10%, review depth 3%), SVM (sensitivity 7%, review depth 2%), and RF (sensitivity 2%, review depth 0%; Table 3). Individual review depth to achieve maximum sensitivity varied from 94% (RF) to 99% (GBT and SVM models). SVM, RF, GBT, logistic regression, and decision trees were included in the voting ensemble model. GBT and RF models were given twice the weight of the other three models (logistic regression, decision trees, and SVM).

Using the voting ensemble model, to achieve 100% sensitivity, 94% ZAPSS of completed pregnancies would require manual review (Table 2). In this scenario, 100% of cases and 92% of non-cases would be manually reviewed—a potential 8% reduction in volume of non-

cases for manual review. To achieve 97% sensitivity, 50% of all reported completed pregnancies would undergo manual review. In this scenario, 97% of cases and 34% of non-cases would be manually reviewed—a potential 66% reduction in volume of non-cases for manual review. Compared to the RF model individually, the best performing individual model at 100% sensitivity, the voting ensemble review depth at 100% sensitivity was equal to that of RF individually. Voting ensemble review depth was lower than that of RF at model sensitivity of 90–99%.

3.3 | Ensemble model validation

Ensemble models with 100% sensitivity identified an additional 204 possible cases from USZPIR and an additional 101 possible cases from ZAPSS for manual review. These possible cases represented 13.0% of unreviewed USZPIR cases (n = 1,566) and 5.8% of unreviewed ZAPPS cases (n = 1,736), respectively. Manual review of these possible cases did not yield additional cases of Zika-associated birth defects.

4 | DISCUSSION

ML methods have potential to reduce modestly the number of non-cases that undergo manual review (6–9% reduction in overall review volume) without sacrificing sensitivity for ascertaining cases of Zika-associated birth defects. Moreover, ensemble models that achieved 96–97% sensitivity may better reflect what is achievable in a real-world scenario, while reducing the number of non-cases undergoing review by 50–66% (reduction in overall review volume by 25–50%). Resources to manually review additional cases would need to be allocated to approach 100% sensitivity. Additionally, larger reductions in overall review volume were observed for ZAPSS compared to USZPIR, which might be attributable to the larger number of non-text variables and open-text variables fields in ZAPSS available for developing models.

Reductions in overall burden on resources to conduct manual review might be appealing; however, there are important considerations for implementation. Because knowledge of the range of outcomes of novel exposures may change over time, there are no reasonable estimates of how long it will take to develop a final model for case status prediction. Case status designations by ML models would benefit from validation against manual review to ensure similar sensitivity and specificity for case identification. With buy-in from stakeholders, ML models with desired levels of sensitivity and specificity might be incorporated into routine surveillance procedures. Possible cases identified by ML models as having the highest likelihood of being true cases might be prioritized for manual review or, if acceptable, bypass manual review completely. Alternatively, as done for model validation in this analysis, ML models might be used routinely to scan batches of data for non-cases in order to identify missed possible cases.

ML methods cannot replace manual review completely. For instance, manual review methods were a necessary step in creating a data set for model training and in scrutinizing data for additional possible cases, especially where identification of possible cases differed between voting ensemble models and current methods. This is especially important when

evaluating initial ML models that achieve predictive performances that have an undesirable or unbalanced number of missed cases or number of non-cases undergoing review.

During a public health emergency response, sensitivity of methods for identifying possible cases for manual review is prioritized over specificity, especially when the full range of outcomes of a new and emerging exposure are not yet known. In this analysis, ensemble model validation identified additional completed pregnancies as possible cases, although none were found to be true cases after manual review of their data. Close inspection of data showed similarities between these possible cases and cases confirmed by manual review, mainly characteristics which are known to increase the suspicion for the presence of a Zika-associated birth defect. For example, presence of symptoms consistent with Zika virus disease in the mother were found in both the additional possible cases and cases confirmed via manual review.

There are several limitations to ML methods as applied in this exploratory analysis. First, ML models were not developed to classify individual types of Zika-associated birth defect. Manual review is still required to confirm the presence and type of Zika-associated birth defect for public reporting. Additionally, the definition of Zika-associated birth defects used for manual review at the time of this analysis has since been revised to exclude neural tube defects. Finally, ML models developed in this analysis are specific for the methods of data collection, review and classification used by USZPIR and ZAPSS surveillance systems.

Despite these limitations, the results of this ML exploration demonstrate potential for ML to help reduce the number of cases undergoing manual review or at least to prioritize reviews. Surveillance for prenatal exposures and infant outcomes might benefit from integration of ML methods into manual review procedures, especially when data collection is expected to be ongoing, when the burden of manual review is anticipated to reach resource capacity, and where manual review methods for case classification are relatively stable. While ML models may need to be retrained if data collection methods are changed or if the definition of an outcome of interest is refined, establishing use of ML methods concurrently with manual review might improve the ability to integrate ML methods early in an emergency response and relieve burden of manual review. This may be especially important in scenarios with potential for exponential increases in the volume of data for review and for the need of rapid reporting of surveillance data to be used for public health action.

ACKNOWLEDGEMENTS

This work was funded by the US Centers for Disease Control and Prevention. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

DATA AVAILABILITY STATEMENT

Authors elect to not share data due to privacy concerns and confidentiality agreements, but code and supporting materials including data dictionaries can be made available

APPENDIX A.

U.S. Zika Pregnancy and Infant Registry	
Model	**Parameters**
Support vector machine	Unbalanced class weights {0:40%, 1:60%} Kernel = “radial basis function” Penalty parameter = 1
Random forest	Number of trees: 100 Number of features: 150
k-nearest neighbors	Number of neighbors = 5
Logistic regression	Penalty = “12” Cost strength = 1
Gradient boosted machine	Learning rate = 0.1 Minimum sample split = 2 Number of boosting stages = 100 Max features = none Max depth = 3
Zika Active Pregnancy Surveillance System	
Model	**Parameters**
Support vector machine	Unbalanced class weights: N/A—SVM model for ZAPSS was attempted in initial model training stages; however, was abandoned as it provided no meaningful change in model performance Kernel = “radial basis function” Penalty parameter = 1
Random forest	Number of trees = 100 Number of features = 250
Decision trees	Imbalanced class weight {0:20%, 1:80%} Number of features: all Minimum split: 2
Logistic regression	Penalty = “12” Cost strength = 1
Gradient boosted machine	Learning rate = 0.1 Minimum sample split = 2 Number of boosting stages = 100 Max features = 250 Max depth = 3

REFERENCES

Hastie T, Tibshirani R, Friedman J, & Franklin J (2005). The elements of statistical learning: Data mining, inference and prediction. Mathematical Intelligencer, 27(2), 83–85.

Honein MA, Dawson AL, Petersen EE, Jones AM, Lee EH, Yazdy MM, … Jamieson DJ (2017). Birth defects among fetuses and infants of US women with evidence of possible Zika virus infection during pregnancy. JAMA, 317(1), 59–68. 10.1001/jama.2016.19006 [PubMed: 27960197]

Kang J, Schwartz R, Flickinger J, & Beriwal S (2015). Machine learning approaches for predicting radiation therapy outcomes: A clinician's perspective. International Journal of Radiation Oncology, Biology, Physics, 93(5), 1127–1135. 10.1016/j.ijrobp.2015.07.2286

Lee Y, Ragguett RM, Mansur RB, Boutilier JJ, Rosenblat JD, Trevizol A, … McIntyre RS (2018). Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review. Journal of Affective Disorders, 2, e100. 10.1038/tp.2012.10

Li C, Lu Y, Wu J, Zhang Y, Xia Z, Wang T, … Guo J (2018). LDA Meets Word2Vec: A novel model for academic abstract clustering. In Companion proceedings of the the web conference 2018 (pp. 1699–1706). Lyon, France: International World Wide Web Conferences Steering Committee.

Moore CA, Staples JE, Dobyns WB, Pessoa A, Ventura CV, Da Fonseca EB, … Rasmussen SA (2017). Characterizing the pattern of anomalies in congenital Zika syndrome for pediatric clinicians. JAMA Pediatrics, 171(3), 288–295. 10.1001/jamapediatrics.2016.3982 [PubMed: 27812690]

Olson SM, Delaney A, Jones AM, Carr CP, Liberman RF, Forestieri NE, … Cragan JD (2019). Updated baseline prevalence of birth defects potentially relatedto Zika virus infection. Birth Defects Research, 111(13), 938–940. 10.1002/bdr2.1546 [PubMed: 31264801]

Pratapa A, Choudhury M, & Sitaram S (2018). Word embeddings for code-mixed language processing. In Proceedings of the 2018 conference on empirical methods in natural language processing (pp. 3067–3072). Brussels, Belgium: Association for Computational Linguistics.

Rasmussen SA, Jamieson DJ, Honein MA, & Petersen LR (2016). Zika virus and birth defects—Reviewing the evidence for causality. New England Journal of Medicine, 374(20), 1981–1987. 10.1056/NEJMsr1604338

Reynolds MR, Jones AM, Petersen EE, Lee EH, Rice ME, Bingham A, … Honein MA (2017). Vital signs: Update on Zika virus-associated birth defects and evaluation of all US infants with congenital Zika virus exposure—US Zika pregnancy registry, 2016. Morbidity and Mortality Weekly Report, 66(13), 366–373. 10.15585/mmwr.mm6613e1 [PubMed: 28384133]

Rice ME, Galang RR, Roth NM, Ellington SR, Moore CA, Valencia-Prado M, … Honein MA (2018). Vital signs: Zika-associated birth defects and neurodevelopmental abnormalities possibly associated with congenital Zika virus infection—US territories and freely associated states, 2018. Morbidity and Mortality Weekly Report, 67(31), 858–867. 10.15585/mmwr.mm6731e1 [PubMed: 30091967]

Shapiro-Mendoza CK, Rice ME, Galang RR, Fulton AC, VanMaldeghem K, Prado MV, … Meaney-Delman D (2017). Pregnancy outcomes after maternal Zika virus infection during pregnancy—US territories, January 1, 2016–April 25, 2017. Morbidity and Mortality Weekly Report, 66(23), 615. 10.15585/mmwr.mm6623e1 [PubMed: 28617773]

Wall DP, Kosmicki J, Deluca TF, Harstad E, & Fusaro VA (2012). Use of machine learning to shorten observation-based screening and diagnosis of autism. Translational Psychiatry, 2 (4), e100. [PubMed: 22832900]

Wei L, Yang Y, Nishikawa RM, & Jiang Y (2005). A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. IEEE Transactions on Medical Imaging, 24(3), 371–380. [PubMed: 15754987]

Xiong Z, Liu T, Tse G, Gong M, Gladding PA, Smaill BH, … Zhao J (2018). A machine learning aided systematic review and meta-analysis of the relative risk of atrial fibrillation in patients with diabetes mellitus. Frontiers in Physiology, 9, 835. 10.3389/fphys.2018.00835 [PubMed: 30018571]

BOX 1

U.S. Zika Pregnancy and Infant Registry criteria for Zika-associated birth defects[a,b,c]

Brain abnormalities with and without microcephaly

Confirmed or possible congenital microcephaly[c]

Intracranial calcifications

Cerebral atrophy

Abnormal cortical formation (e.g., polymicrogyria, lissencephaly, pachygyria, schizencephaly, gray matter heterotopia)

Corpus callosum abnormalities

Cerebellar abnormalities

Porencephaly

Hydranencephaly

Ventriculomegaly/hydrocephaly (excluding “mild” ventriculomegaly without other brain abnormalities)

Fetal brain disruption sequence (collapsed skull, overlapping sutures, prominent occipital bone, scalp rugae)

Other major brain abnormalities

Neural tube defects and other early brain malformations

Neural tube defects including anencephaly, acrania, encephalocele, spina bifida

Holoprosencephaly (arhinencephaly)

Eye abnormalities

Microphthalmia/anophthalmia

Coloboma

Cataract

Intraocular calcifications

Chorioretinal anomalies involving the macula (e.g., chorioretinal atrophy and scarring, macular pallor, gross pigmentary mottling and retinal hemorrhage; excluding retinopathy of prematurity)

Optic nerve atrophy, pallor, and other optic nerve abnormalities

Consequences of central nervous system dysfunction

Congenital contractures (arthrogryposis, club foot with associated brain abnormalities, congenital hip dislocation or developmental dysplasia of the hip with associated brain abnormalities)

Congenital deafness documented by postnatal audiological testing

[a]Adapted from "Birth defects among fetuses and infants of U.S. women with evidence of possible Zika virus infection during pregnancy" by Honein et al., 2017. [b]Since the completion of this analysis, the criteria for Zika-associated birth defects has been updated to remove those with NTDs and other early brain malformations or CNS dysfunction with no other qualifying defects (Olson et al., 2019). [c]Live births: measured head circumference (adjusted for gestational age and sex) less than the third percentile at birth or, if not measured at birth, within first 2 weeks of life. Pregnancy loss: prenatal head circumference more than 3 *SDs* below the mean based on ultrasound or postnatal head circumference less than the third percentile. Birth measurements are evaluated using the Intergrowth-21st standards (http://intergrowth21.ndog.ox.ac.uk/) based on measurements within 24 hr of birth.

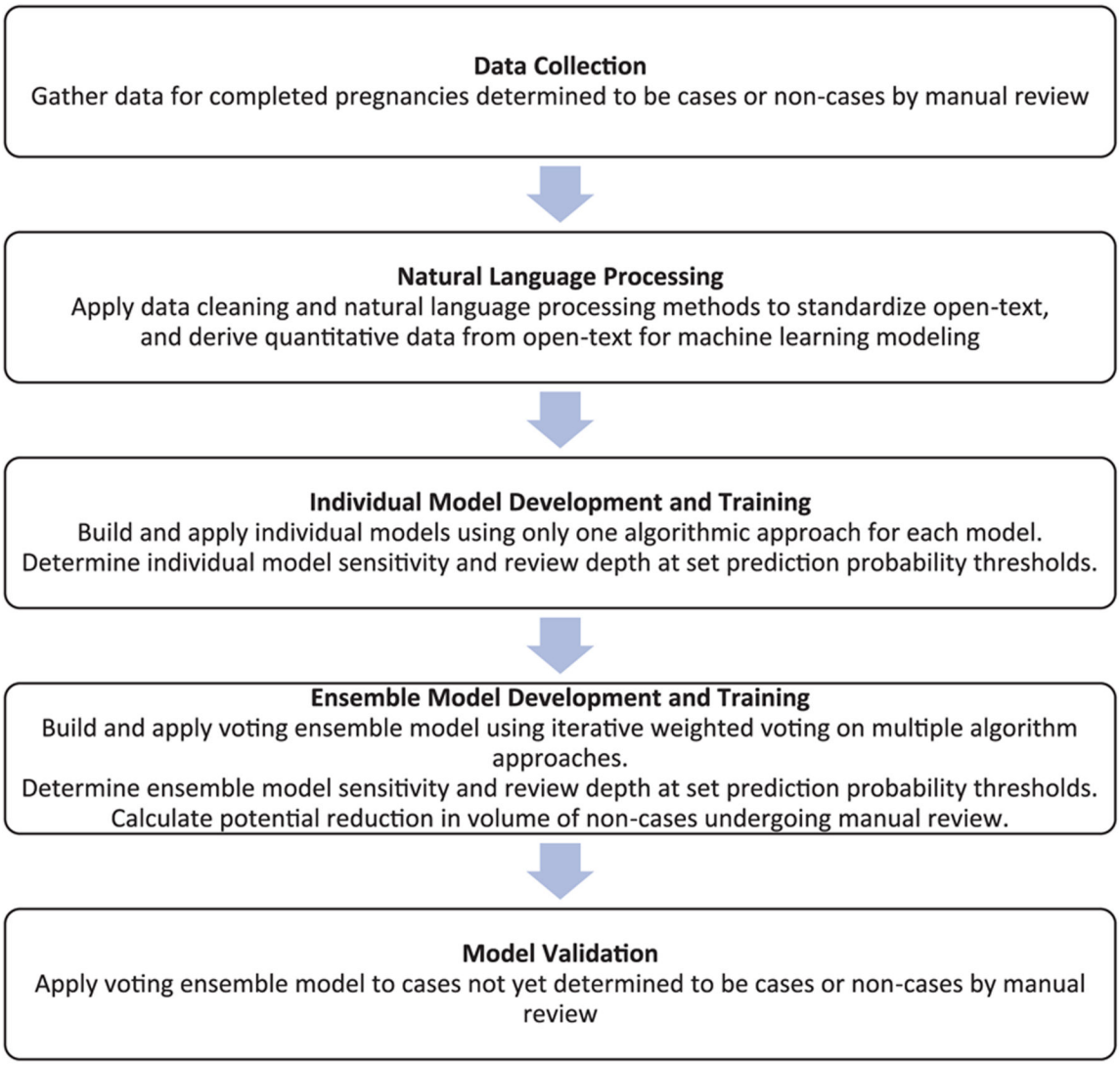

FIGURE 1.
Flow chart of methods approach to exploring ML application in the Zika case review process

TABLE 1

Individual model sensitivity and depth by model prediction threshold—US Zika Pregnancy and Infant Registry

	Gradient boosted trees		Logistic regression		Random forests		*k*-nearest neighbors		Support vector machine	
Model prediction threshold	Model sensitivity (%)	Review depth (%)	Model sensitivity (%)	Review depth (%)	Model sensitivity (%)	Review depth (%)	Model sensitivity (%)	Review depth (%)	Model sensitivity (%)	Review depth (%)
.01	100	100	100	100	100	100	98	92	100	100
.03	100	99	100	100	100	99	98	92	100	100
.05	100	88	100	96	100	96	98	92	100	100
.10	96	74	98	83	100	87	97	91	100	100
.15	93	67	95	72	98	80	94	89	100	100
.20	92	63	93	65	97	73	89	82	100	100
.25	90	60	91	61	94	68	86	79	100	100
.30	90	58	89	58	91	62	82	75	100	100
.35	88	55	87	55	88	58	78	71	100	100
.40	86	54	87	53	86	55	71	64	95	95
.45	84	52	83	50	81	51	63	57	90	88
.50	81	50	81	49	78	47	58	52	51	46
.60	76	46	75	44	70	41	43	38	3	3
.70	70	41	66	37	60	35	28	25	0	0
.75	64	37	62	34	52	30	21	20	0	0
.80	60	34	56	31	45	25	15	14	0	0
.85	54	30	51	27	35	20	10	9	0	0
.90	43	23	44	23	19	10	7	6	0	0
.95	29	15	32	17	8	4	5	4	0	0

TABLE 2

Voting ensemble model sensitivity, depth, and percent reduction in number of non-cases for manual review

U.S. Zika Pregnancy and Infant Registry voting ensemble model		
Model sensitivity (%)	**Review depth (%)**	**Potential % reduction in non-cases for manual review**
100	100	0
100	91	18
96	75	50
94	67	67
83	50	N/A
75	44	N/A
49	27	N/A
29	16	N/A
8	5	N/A
Zika Active Pregnancy Surveillance System voting ensemble method		
Model sensitivity (%)	**Review depth (%)**	**Potential % reduction in non-cases for manual review**
100	100	0
100	94	8
97	50	66
94	40	79
86	30	N/A
73	20	N/A
39	10	N/A
20	5	N/A
5	1	N/A

Note: Percent reduction in number of non-cases undergoing manual review was calculated for voting ensemble models at sensitivities ≥94%. Review depth was defined as the number of completed pregnancies the ML model classified to undergo the manual review process (numerator) over the total number of available completed pregnancies (denominator) to achieve a chosen sensitivity threshold.

TABLE 3

Individual model sensitivity and depth by model prediction threshold—Zika Active Pregnancy Surveillance System

	Support vector machine		Random forests		Decision trees		Logistic regression		Gradient boosted trees	
Model prediction threshold	Model sensitivity (%)	Review depth (%)	Model sensitivity (%)	Review depth (%)	Model sensitivity (%)	Review depth (%)	Model sensitivity (%)	Review depth (%)	Model sensitivity (%)	Review depth (%)
.01	100	100	100	94	99	98	99	86	100	100
.03	100	100	99	83	99	98	95	62	100	99
.05	100	99	98	71	99	98	92	50	97	52
.10	98	82	97	53	94	77	87	38	92	36
.15	88	43	93	43	84	38	84	33	89	30
.20	79	30	89	35	83	37	82	30	86	27
.25	72	24	86	31	83	37	79	27	83	25
.30	67	21	81	27	83	37	77	25	81	24
.35	59	18	79	24	81	33	75	24	80	23
.40	51	15	76	22	81	33	72	22	77	22
.45	48	14	72	20	79	29	70	21	74	21
.50	44	13	66	18	77	28	68	20	73	20
.60	42	12	58	15	76	26	64	19	69	19
.70	39	11	42	11	71	22	58	16	65	17
.75	37	11	34	9	67	20	54	15	60	16
.80	35	10	25	6	64	19	48	14	52	14
.85	31	9	17	4	61	18	44	12	40	10
.90	22	6	10	2	57	16	37	10	26	7
.95	7	2	2	0	52	14	28	8	10	3

Birth Defects Res. Author manuscript; available in PMC 2021 November 01.