

A Machine Learning Algorithm to Predict Hyperglycemic Cases Induced by PD-1/PD-L1 Inhibitors in the Real World

Jincheng Yang (✉ jincool.1129@163.com)

National Cancer Center

Weilong Lin

National Cancer Center

Liming Shi

National Cancer Center

Ming Deng

National Cancer Center

Wenjing Yang

National Cancer Center

Research

Keywords: machine learning, hyperglycemia, prediction, PD-1/PD-L1, ADR, SVM

Posted Date: December 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-135595/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Diabetes mellitus and cancer are amongst the leading causes of deaths worldwide; hyperglycemia plays a major contributory role in neoplastic transformation risk. From reported adverse events of PD-1 or PD-L1 (programmed death 1 or ligand 1) inhibitors in post-marketing monitoring, we aimed to construct an effective machine learning algorithm to predict the probability of hyperglycemic adverse reaction from PD-1/PD-L1 inhibitors treated patients efficiently and rapidly.

Methods: Raw data was downloaded from US Food and Drug Administration Adverse Event Reporting System (FDA FAERS). Signal of relationship between drug and adverse reaction based on disproportionality analysis and Bayesian analysis. A multivariate pattern classification of Support Vector Machine (SVM) was used to construct classifier to separate adverse hyperglycemic reaction patients. A 10-fold-3-time cross validation for model setup within training data (80% data) output best parameter values in SVM within R software. The model was validated in each testing data (20% data) and two total drug data, with exactly predictor parameter variables: *gamma* and *nu*.

Results: Total 95918 case files were downloaded from 7 relevant drugs (cemiplimab, avelumab, durvalumab, atezolizumab, pembrolizumab, ipilimumab, nivolumab). The number-type/number-optimization method was selected to optimize model. Both *gamma* and *nu* values correlated with case number showed high adjusted r^2 in curve regressions (both $r^2 > 0.95$). Indexes of accuracy, F1 score, kappa and sensitivity were greatly improved from the prediction model in training data and two total drug data.

Conclusions: The SVM prediction model established here can non-invasively and precisely predict occurrence of hyperglycemic adverse drug reaction (ADR) in PD-1/PD-L1 inhibitors treated patients. Such information is vital to overcome ADR and to improve outcomes by distinguish high hyperglycemia-risk patients, and this machine learning algorithm can eventually add value onto clinical decision making.

1. Background

Diabetes mellitus and cancer are amongst the leading causes of deaths worldwide. As a major contributory role in neoplastic transformation risk, hyperglycemia (high blood glucose) is also influenced by cancer treatment. Higher blood glucose level, dealing with or without insulin, correlated with cancer risk, progression, and mortality in a higher degree.^{1,2}

Recently, immunotherapy is a vast improvement over previous anti-cancer therapies.³ In 2018 and 2019, two reviews of JAMA reported treatment-related adverse events from programmed death 1 (PD-1) or programmed death ligand 1 (PD-L1) inhibitors, or immune checkpoint inhibitor regimens in clinical trials. Among the endocrine dysfunctions, hyperglycemia was on the top three following hypothyroidism and hyperthyroidism in all-grade immune-related adverse events (irAE), and on the top of grade 3 or higher irAEs were hyperglycemia.^{4,5} Prolonged exposure to hyperglycemia can epigenetically modify gene expression profiles in human cells and this effect is still sustained even after hyperglycemic control is therapeutically achieved. Cancer cells exposed to hyperglycemic situation would develop permanent aggressive growth, even after returned to euglycemic conditions. This phenomenon is called hyperglycemic memory. This metabolic memory effect contributes substantially to the pathology of various diabetic complications.^{6,7}

World-wide public database on adverse events could provide many drug usage information^{8,9}, becoming a new information source in drug post marketing phase. The Food and Drug Administration Adverse Event Reporting System (FAERS) is a database that contains adverse event reports, medication error reports and product quality complaints resulting in adverse events. The database is designed to support the FDA's post-marketing safety surveillance program for drug and therapeutic biologic products. The informatic structure of the FAERS database adheres to the international safety reporting guidance issued by the International Conference on Harmonisation.

10

Machine learning is the scientific discipline that focuses on how computers learn from data, with its emphasis on efficient computing algorithms.¹¹ Machine Learning techniques have increased to be a hot spot in data mining with health record data, and modeling this Big Data information requires managing overfitting, model interpretability and computational cost.¹² Support Vector Machine (SVM) is a type of supervised learning method which analyzes data and recognizes patterns, mainly used for statistical classification and regression, especially in nonlinear classification.^{13,14} SVM based approach doing well in managing sparse data in high dimensions. It is a multivariate pattern classification dividing samples into worthy bifurcations (binary variables) through a line or plane in multidimensional feature, which is called Maximum Margin Hyperplane.¹⁵ SVM overcomes the other state-of-the-art competitors providing the best compromise between predictive performance and computation time.¹² The SVM has been applied to seizure prediction, detection and classification.¹⁶ One advantage of SVM is its classification in small number of training samples; another is solving linear and nonlinear regression problems.¹⁴ In 2019, Bernardini *et al.* successfully predicted Type 2 Diabetes in electronic health records by establishing a cross-validation sparse balanced SVM model, proving that the capability of SVM model to predict the clinical outcomes.¹²

However, to date, neither the association of hyperglycemic incidence and personal-related features, nor which special features could impact hyperglycemia have been reported. Predicting the hyperglycemic incidence could provide better guidance for clinical decision-making for avoiding terrible adverse drug reaction (ADR). In this study, we aimed to construct an effective SVM-based machine learning algorithm model to predict hyperglycemic reaction within PD-1/PD-L1 treating patients.

2. Methods

2.1 Subjects (Data sources: raw data):

Raw data was downloaded from FDA FAERS (Food and Drug Administration, Adverse Event Reporting System) Public Dashboard (<https://www.fda.gov/>, Date as of March 31, 2020).

2.1.1 Inclusive criteria:

☒. Search by post-marketing product: Cemiplimab (cem, PD-1), Avelumab (avel, PD-L1), Durvalumab (dur, PD-L1), Atezolizumab (atez, PD-L1), Pembrolizumab (pem, PD-1), Ipilimumab (ipi, CTLA-4), Nivolumab (nivo, PD-1). (As two two PD-1 combination therapies have been approved by the FDA: nivo with ipi, and pem with chemotherapy¹⁷, ipi is also included by product search for better predicting hyperglycemic ADR in real world drug usage.)

☒. Search by reaction term: Diabetes Mellitus, Hyperglycemia (Hyperglycaemia), Type 1 Diabetes Mellitus, Type 2 Diabetes Mellitus, and Blood Glucose Increased.

2.1.2 Exclusive criteria:

☒. Other products.

☒. No information on blood glucose in reactions.

☒. Case with missing items.

2.2 Experimental Design

The procedure of SVM machine learning is: raw data download from FAERS, algorithm selection, model setup (parameters optimization and determination), parameter curve regressions, and data prediction (Fig. 1).

Step 1: Data form selection

Raw data were fixed into two parts: original-data (with missing values) and complete-data (without missing values, rows containing missing value(s) were deleted).

Step 1.1. Data extraction (original data):

Items of *Case ID, Reactions, Serious, Sex, Event Date, Case Priority, Patient Age, Patient Weight, Reporter Type, Report Source, Country where Event occurred* were obtained from the downloaded excel files (raw data).

Excluded items were: *Suspect Product Names, Suspect Product Active Ingredients, Reason for Use, Outcomes, Latest FDA Received Date, Sender, Concomitant Product Names, Latest Manufacturer Received Date, Initial FDA Received Date, Reported to Manufacturer, Manufacturer Control Number, Literature Reference, Compounded Flag*.

Step 1.2. Data mining (complete data):

Within original data, algorithms on signal of relationship between a drug and a special adverse reaction usually based on disproportionality analysis and Bayesian analysis, including four statistical procedure: proportional reporting ratio (PRR), reporting odds ratio (ROR), information component (IC), and empirical Bayes geometric mean (EBGM).^{18,19} The four algorithms' computation and criteria are according to reference.²⁰

Step 2: SVM optimization methods

Bicategorically, target reaction could be changed into factor-type ("Yes" and "No"), or treated as number-type ("1" and "0") in SVM model setup. The parameter optimization could also be accomplished via two methods: number-optimization and R-function-optimization.

In number-optimization, the three parameters were tested separately to select a range covering best value. After best proportion were defined by combination of three parameters, the best values were set to predict. In R-function-optimization, the parameter range was also input to a built-in function (`tune.svm`), and then best values were output. If the best values closed to range boundary, the adjusted new range would be input for optimization again. The factorized data were both optimized via number and function methods.

Step 3: SVM model

After missing values deleted from original data to generate complete data, disproportionate and Bayesian analysis were adopted to quantify the signal, the association between the reported features and ADR. The general modelling set was from stratifying random-split cross-validation into training data (80% data) and testing data (20% data) in each drug, containing proportional positive and negative cases respectively. Once the algorithm was optimized by training data, no more changes were made and it was evaluated on testing data checks.

Step 3.1. Model variable selection

The key to construct a SVM model that can accurately screen the active markers is to select the appropriate variable indexes.¹⁴ Variable selection was according to two methods: Near Zero Variance Method (R function: NearZeroVar) and Model Assessment (R function: varImp).

Step 3.2. Model setup

Parameter algorithm selection: SVM-Type (*C-classification*, "C", *one-classification*, "one", *eps-regression*, "eps", *nu-regression*, "nu") and SVM-Kernel (*linear*, "l", *poly*, "p", *rbf*, "r", *sigmoid*, "s").

Parameters (eg. *gamma*, *nu*, *cost*, *degree*, *coef0*) optimization separately: their value ranges were determined by the best outputs.

Parameters (eg. *gamma*, *nu*, *cost*) optimization mutually: the complete-data was divided into training data and testing data according to random seed; best parameter values were set by training data through 10-fold-3-time cross validation.

Parameters (eg. *gamma*, *nu*) determination: Accuracy (total precise rate), F1 score, sensitivity (positive precise rate) and kappa (consistence) values were chose as evaluation indicators. Confusion matrix was calculated according to appendix-Table 1.

Step 4: Curve regression

The variables (eg. *gamma*, *nu*) and relevant case number from SVM model setup were selected and tested in curve regression. The computing formula were given beside the curve.

Step 5: Data prediction

The prediction model was performed by the testing data and included the sixth drug (ipi) as well as the exceptional drug (dur). Four indexes (accuracy, F1 score, sensitivity and kappa) were checked from this model.

2.3 Statistical analysis

Descriptive analysis was used to summarize patient demographic characteristics, with mean values and range values for continuous variables, ratios for categorical variables.

To explain the impacting factors to hyperglycemia, Reaction is defined as dependent variable, others as response variables. For classification algorithm, we used SVM with statistical software R, version 4.0.0 for windows. T test

was performed for comparing normal distributions and defining 95% confidence intervals, and Wilcoxon rank test for comparing other unknown distributions.

3. Results

From FAERS database, total 95918 case files were downloaded from 7 relevant drugs (cem:376, avel:1116, dur:2980, atez:7061, pem:20507, ipi:20862, nivo:43016). Hyperglycemic signal detection for those drugs were listed in Table 1, where Bayesian confidence propagation neural network (BCPNN) indicated all 6 PD-1/PD-L1 inhibitors and ROR indicated 4 of them would induce hyperglycemia (Table 1).

Table 1
Signal Detection for Hyperglycemia in PD-1/PD-L1 Drugs

Drugs	N	ROR (95% two-sided CI)	PRR (χ^2)	BCPNN (95% two-sided CI)	MGPS (95% two-sided CI)
cem	2	0.41(0.10–1.63)	0.41(1.72)	51.80(12.90-207.70)*	0.41(0.10–1.65)
avel	18	1.25(0.78–1.99)*	1.25(0.87)	56.50(35.50–90.00)*	1.24(0.78–1.98)
dur	40	1.04(0.76–1.41)*	1.04(0.05)	59.10(43.20–80.70)*	1.04(0.76–1.41)
atez	90	0.98(0.80–1.21)	0.98(0.029)	61.50(49.90–75.70)*	0.98(0.80–1.21)
pem	437	1.66(1.51–1.82)*	1.66(111.40)*	65.30(59.40–71.80)*	1.64(1.49–1.81)
nivo	1050	1.91(1.79–2.03)*	1.90(440.26)*	67.60(63.60–71.90)*	1.88(1.77-2.00)
Criteria: ROR: 95% CI > 1, N ≥ 2; PRR: PRR ≥ 2, χ^2 ≥ 4, N ≥ 3; BCPNN: IC025 > 0; MGPS: EBG05 > 2, N > 0.					
N: hyperglycemic reaction cases in raw data.					
*Statistical significance.					

As no hyperglycemic ADR reported in cem as well as too many missing items in its original-data (only 2 cases in complete-data), cem is not included in following analysis yet. Cases with unavailable items were removed and the rest demographic summary was listed in Appendix-table 2; and the following complete-data were based on the listed cases.

Model for original-data was setup and optimized via number-type and number-optimization, while complete-data was via bilateral methods (Table 2). Accuracy and sensitivity results showed number-optimization is more precise than R-function-optimization (avel, pem, nivo); number-type was a little higher than factor-type (avel, dur, atez) (Fig. 2). The following SVM model setup chose number-type and number-optimization method within complete-data.

Table 2
Accuracy and Sensitivity of Two Algorithm-Optimization Methods

Drugs	Original-data		Complete-data (without missing values)					
	(with missing values)		NT + NUO		FT + RFO		FT + NUO	
	A	S	A	S	A	S	A	S
avel	98.3%	85.7%	100.0%	100.0%	98.8%	46.2%	99.8%	92.3%
dur	97.8%	34.5%	100.0%	100.0%	99.9%	90.0%	99.9%	90.0%
atez	98.7%	62.9%	100.0%	100.0%	99.9%	89.7%	99.9%	89.7%
pem	96.6%	27.4%	99.5%	87.1%	98.5%	44.4%	99.6%	84.7%
nivo	96.4%	24.8%	99.8%	94.0%	97.6%	9.9%	99.8%	93.5%
S = sensitivity, A = accuracy, NT = number-type, FT = factor-type, NUO = number-optimization, RFO = R-function-optimization								

The two methods (Near-Zero-Variance-Method and Overall-Value-of-Model-Assessment) showed one common point: variables of "Reporter" (means Consumer or Health professional etc), "Serious" and "Source" were not important in model (Table 3). The three variables ("Reporter", "Serious" and "Source") and "ID" were not included in following algorithm. Clinically, "Source" and "Reporter" introduced where the ADR came from, without more information in disease control. Other 8 variables ("Reactions", "Reason", "Country", "Weight", "Year", "Age", "Priority" and "Sex") were used to model setup.

Table 3
Variable Selection from nzv and overall Methods

Drugs		avel		dur		atez		pem		nivo	
No.	Variable	nzv	overall	nzv	overall	nzv	overall	nzv	overall	nzv	overall
1	reactions	F	1	F	1	F	1	F	1	F	1
2	reason	F	1	F	0.37141	F	0.69288	F	0.34268	F	0.21317
3	country	F	0.05873	F	0.00494	F	0.01639	F	0.00251	F	0.01057
4	age	F	0.01599	F	0.00321	F	0	F	0	F	0
5	weight	F	0.01533	F	0.00495	F	0.00529	F	0	F	0.00110
6	year	F	0	F	0	F	0.00304	F	0.00137	F	0
7.	sex	F	0	F	0	F	0	F	0	F	0
8	priority	T	0	F	0	T	0	F	0.00254	F	0.00061
9	reporter	T	0	T	0	T	0	T	0	F	0
10	serious	T	0	T	0	T	0	T	0	F	0
11	source	T	0	T	0	T	0	T	0	F	0
nzv: near zero variance. Overall: overall value of model assessment.											
T: TRUE. F: FALSE (T means this variable should be deleted)											

In SVM algorithms, SVM-kernel includes "l", "p", "r", "s" and SVM-type includes "C", "one", "eps", "nu". The five PD-1/L1 drugs were test by the 4 × 4 cross combinations. The 16 combinations showed "rbf" and "nu-regression" (r-nu) displaying good prediction, especially at F1 score and kappa (Fig. 3).

Parameters in SVM mainly included *degree*, *cost*, *gamma*, *nu*, *coef0*. In former method section, all the results showed the best *degree* and *coef0* value is "1" (the function default number), so the following model setup section took another three parameters (*nu*, *cost* and *gamma*) into account. Based on the best model and to avoid overfitting, *cost* value was defined as "1" (the function default number); and *nu* and *gamma* value was defined as the mean value in the best prediction range. Based on adjusting *gamma* and *nu* value from training data, the model performance was checked in testing data, four indexes (accuracy, F1 score, sensitivity, and kappa values) were showed in Fig. 4.

As the four-index trends from dur were exceptional (excessively high at Sensitivity in Fig. 4), the curve regression did not include parameters from dur (ie, curve regression based on avel, atez, pem and nivo). From model optimization, we found out the *gamma* values were quadratic to case number and *nu* values were exponential to minus case number (Fig. 5), which are as follows:

$$y = 3.78 \times 10^{-4}x + 6.47 \times 10^{-8}x^2 \quad (y = \text{gamma}, x = \text{case number})$$

[Formula-I]

$$y = 1.251 \times e^{(-x/2650)} \quad (y = \text{nu}, x = \text{case number}, 0 < \text{nu} \leq 1)$$

[Formula-II]

New predictions from optimized model (Type: *nu*-regression; Kernel: *rbf*, parameter: *gamma* and *nu*) were applied on four testing data and two total data (dur and ipi). Four indexes (accuracy, F1 score, sensitivity and Kappa) over curve model were greatly improved than initial prediction (Fig. 6, appendix-table 3).

Receiver operating characteristic (ROC) analysis is a tool used to describe the discrimination accuracy of a diagnostic test or prediction model.²¹ The diagnostic values from this model prediction and single items (“Reactions”, “Reason”, “Country”, “Weight”, “Year”, “Age”, “Priority” and “Sex”) on the test-parts and two drugs were evaluated by ROC curves, whereas the predictive values were much higher from model than single FAERS items in Fig. 7.

4. Discussion

In this study, we have developed a machine learning algorithm that predicted hyperglycemic cases in PD-1/PD-L1 used patients based on US FDA adverse events reporting system (FAERS) within 5 respective drugs. To the best of our knowledge, this is the first time to predict hyperglycemic incidence from real-world clinical practice via machine learning.

Hyperglycemia is an important ADR in cancer treatment²², as it influences the outcome of cancer therapy via various mechanisms such as chemoresistance enhancement²³, metabolic reprogramming and molecular alterations²⁴, neoplastics deactivation²⁵, affecting pharmacokinetics, pharmacodynamics and ADRs²⁶, and inflammation sponsoring¹, immune destruction.^{27,28}

One emphasis of pharmacovigilance is ADR from drug post-marketing phase. On one hand, it is challengeable to find out the effective and fast means from clinical interheterogeneity and intraheterogeneity results. Interestingly, and on the other hand, the prediction is important in that occurrence is usually unknown in clinical treatment, as real time monitoring is inconvenient and expensive for discovering ADRs. This algorithm from clinical available features at the time of presentation was proved robust and generalizable in later testing sequence. This non-invasive and precise prediction could greatly help clinical practitioners to distinguish high-risk patients. Therefore, this study provided an orientation to predict occurrence of hyperglycemic ADR of these drugs.

FDA is responsible for protecting the public health by ensuring the safety, efficacy, and security of human and veterinary drugs, biological products, and medical devices. The reports in FAERS are evaluated by clinical reviewers, in the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER), to monitor the safety of products after they are approved by FDA.¹⁰ SVM is a kind of structural dependence model to find Maximum Margin Hyperplane with ADR reaction and reported features. To train our algorithm in prediction section, new cases are projected into the same situation to test which side of the hyperplane they located into.¹⁶ Each drug database was divided into training and testing parts proportionally

and randomly, in order to set up and check model on real data. As the adverse events may have occurred in a small fraction of patients, we adjusted, for class-imbalance, data divide within R software function (createDataPartition).

As omission values and duplicated cases, variables should be selected in machine learning. Variable selection was defined by two methods, all of which was performed by package of caret in R software. In Near Zero Variance Method, variables displayed as "TRUE" should be deleted. In Overall value of Model Assessment, variables displayed as "0" should be deleted. Compared the two results and clinical analysis, variables of "reactions", "reason", "country", "age", "weight" and "year" are included in model setup. "pem" and "nivo" indicated the meanings of "priority" in variable selection in "overall" method, and "dur" in "nzv" method; so, "priority" is included. Though there were no meanings of "sex" in "overall" method, all the five drugs indicated its meanings at the "nzv" method; so, "sex" is included. For the clinical and algorithmic reasons, "Serious", "Source" and "Reporter" were not included in model setup (Table 3).

SVM model algorithm is defined by its kernel and type. The "*r-nu*" from 4×4 crossing combination showed upper value than average in accuracy and sensitivity, and top values in F1 score and kappa (Fig. 3). Hence, this combination was selected as the main algorithm in SVM model.

Five parameters (*degree*, *cost*, *gamma*, *nu* and *coef0*) were tested in parameter selection section. As no any change of result occurred in regulating the parameter of *degree* and *coef0*, the two values were set as "1". *Cost* was used to prevent overfitting level. Higher *cost* value means more overfitting probability. SVM model indicated *cost* value could be set as "1", where adjusting the other two parameters (*gamma* and *nu*) could also get the best results in 3×10 (10 folds 3 times) cross validation within training data. Then the other parameters were determined by the mean value of cross validation. Using the two mean values, the testing data were predicted with such parameter values. In Fig. 3, dur sensitivity result (highest) did not match other four results (lowest) and its *gamma* value (*gamma* = 5) was excessive at the *gamma* parameter group; so, the curve regression did not include the parameters provided from dur drug.

In the exponential curves (Formula-II), the constant value (2650) was determined by the adjusted r^2 in liner regression between *nu* value and $e^{(-\text{casenumber}/\text{constant})}$. We tried many constant values, during which all the constant (2630, 2640, 2650 and 2660) indicated the same highest r^2 value (0.9567); so, we selected the middle one (2650) as the formula constant.

Though SVM machine learning have many advantages, it does not solve all the problems based on spontaneous reporting systems. FAERS data itself have limitations: no certainty that the reported event was due to the product, insufficient detail, incomplete reports and duplicate reports.¹⁰ In this study, only 5 ADR items (at inclusive criteria ☒) could be selected as the restriction of larger volume data download from website. Other drug factors (eg. doses, frequency) and biological factors (eg. genomic data) were not included in the development of the present algorithm, preventing accuracy as well as F1 score up to 100%.

The graphical ROC curve is produced by plotting sensitivity (true positive rate) on the y-axis against 1 – specificity (false positive rate) on the x-axis for the various values tabulated.²⁹ Areas under the ROC curves (AUC) from eight items were far less than that from model prediction (the uppermost red line in Fig. 7). It means the

prediction of total combination from our model is much more powerful, since single factors do nothing meaningful in prediction.

Although these limitations described above, the algorithm may have provided information that supported either identifying signals between the hyperglycemic ADR and PD-1/PD-L1 regimens, or adjusting care goals for these patients, or providing signs for further well-organized clinical studies.

5. Conclusion

In summary, the SVM model established here can non-invasively and precisely predict occurrence of hyperglycemic ADR in PD-1/PD-L1 inhibitors treated patients from given personal-related variables and case number. The SVM model showed good prediction performance within testing data, whose result proved this model is robust and generalizable. We also believe that the availability of drug regime and dosage will promote this model for facilitating the prediction of hyperglycemic ADR occurrence. Such information is vital to improve to overcome ADR and patient outcomes. The prediction algorithm can eventually add value to clinical decision making.

Abbreviations

PD-1/PD-L1 programmed death 1/programmed death ligand 1

FDA FAERS Food and Drug Administration Adverse Event Reporting System

SVM Support Vector Machine

ADR adverse drug reaction

irAE immune-related adverse events

PRR proportional reporting ratio

ROR reporting odds ratio

IC information component

EBGM empirical Bayes geometric mean

BCPNN Bayesian confidence propagation neural network

ROC receiver operating characteristic

Declarations

Acknowledgements

Not applicable

Ethical Approval and Consent to participate

All procedures performed in studies involving human participants were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. As this was an observational and retrospective study, and the database is open to the public, therefore, ethical review board approval was not required.

Availability of supporting data

The datasets and algorithms used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

All the authors declare that there is no competing interest.

Funding

The work was supported by National Natural Science Foundation of China (grant number: 81603218) and Clinical Pharmacy Research 2019, Beijing Pharmacy Association.

Authors' contributions

Conception/design: Wenjing Yang, Jincheng Yang. Collection and/or assembly of data: Weilong Lin, Liming Shi. Data analysis and interpretation: Jincheng Yang, Liming Shi. Manuscript writing: Jincheng Yang, Ming Deng. Final approval of final manuscript: All authors

Consent for publication

Not applicable

References

1. Ramteke P, Deb A, Shepal V, Bhat MK. Hyperglycemia Associated Metabolic and Molecular Alterations in Cancer Risk, Progression, Treatment, and Mortality. *Cancers*. 2019;11(9):23.
2. Yang J, Jia B, Qiao Y, Chen W, Qi X. Variations of blood glucose in cancer patients during chemotherapy. *Nigerian Journal of Clinical Practice*. 2016;19(6):704–8.
3. PdS I. S L, C Q, et al. Site-specific response patterns, pseudoprogression, and acquired resistance in patients with melanoma treated with ipilimumab combined with anti-PD-1 therapy. *Cancer*. 2020;126(1):86–97.
4. Y W, S Z, F Y, et al. Treatment-Related Adverse Events of PD-1 and PD-L1 Inhibitors in Clinical Trials: A Systematic Review and Meta-analysis. *JAMA oncology* 2019; 5(7): 1008–19.
5. R B-S WTB, AC G-C, et al. Incidence of Endocrine Dysfunction Following the Use of Different Immune Checkpoint Inhibitor Regimens: A Systematic Review and Meta-analysis. *JAMA oncology*. 2018;4(2):173–82.
6. Lee C, An D, Park J. Hyperglycemic memory in metabolism and cancer. *Horm Mol Biol Clin Investig*. 2016;26(2):77–85.

7. A V-D-S RM, dQ B, dCR ART. WB D. Hyperglycemia and aberrant O-GlcNAcylation: contributions to tumor progression. *J Bioenerg Biomembr.* 2018;50(3):175–87.
8. J Y, Y W, K L, W Y, J Z. Risk Factors for Doxorubicin-Induced Serious Hyperglycaemia-Related Adverse Drug Reactions. *Diabetes therapy: research, treatment and education of diabetes and related disorders* 2019; 10(5): 1949–57.
9. Yang J, Yang J. Hyperglycemic ADR Distribution of Doxorubicin From VigiBase. *Am J Ther.* 2019;26(3):e428-e30.
10. US.FOOD&DRUG Administration. <https://www.fda.gov/> (accessed 2020/3/31 2020).
11. RC D. Machine Learning in Medicine. *Circulation.* 2015;132(20):1920–30.
12. L MB. R, P M, E F. Discovering the Type 2 Diabetes in Electronic Health Records Using the Sparse Balanced Support Vector Machine. *IEEE journal of biomedical health informatics.* 2020;24(1):235–46.
13. WJ HZWLYX, SY WLLL. Y. Rapid and accurate assessment of seizure liability of drugs by using an optimal support vector machine method. *Toxicology in vitro: an international journal published in association with BIBRA.* 2011;25(8):1848–54.
14. S L, L W, Z D, et al. Identification of the lipid-lowering component of triterpenes from *Alismatis rhizoma* based on the MRM-based characteristic chemical profiles and support vector machine model. *Analytical and bioanalytical chemistry* 2019; 411(15): 3257–68.
15. Z YL. C, Q L, et al. Support vector machine-based multivariate pattern classification of methamphetamine dependence using arterial spin labeling. *Addiction biology.* 2019;24(6):1254–62.
16. JH Z, HW XH. Z, et al. Personalized prediction model for seizure-free epilepsy with levetiracetam therapy: a retrospective data analysis using support vector machine. *Br J Clin Pharmacol.* 2018;84(11):2615–24.
17. Tang JYJ, Hubbard-Lucey VM, Neftelinov ST, Hodge JP, Lin Y. Trial watch: The clinical trial landscape for PD1/PDL1 immune checkpoint inhibitors. *Nat Rev Drug Discov.* 2018;17(12):854–5.
18. A TS, K K T. Y O. Data mining of the public version of the FDA Adverse Event Reporting System. *International journal of medical sciences.* 2013;10(7):796–803.
19. PS A, Z C, CL C, BC T. Data mining spontaneous adverse drug event reports for safety signals in Singapore - a comparison of three different disproportionality measures. *Exp Opin Drug Saf.* 2016;15(5):583–90.
20. J S, J Y, B Z. A Survey of the FDA's Adverse Event Reporting System Database Concerning Urogenital Tract Infections and Sodium Glucose Cotransporter-2 Inhibitor Use. *Diabetes therapy: research, treatment and education of diabetes and related disorders* 2019; 10(3): 1043–50.
21. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol.* 2018;63(7):07TR1.
22. Yang J, Jia B, Yan J, He J. Glycaemic adverse drug reactions from anti-neoplastics used in treating pancreatic cancer. *Niger J Clin Pract.* 2017;20(11):1422–7.
23. IP Y, ZF M, CW H, et al. High blood sugar levels but not diabetes mellitus significantly enhance oxaliplatin chemoresistance in patients with stage III colorectal cancer receiving adjuvant FOLFOX6 chemotherapy. *Therapeutic Advances in Medical Oncology.* 2019;11:1–13.
24. Q D, C HL. G, et al. High glucose promotes pancreatic cancer cells to escape from immune surveillance via AMPK-Bmi1-GATA2-MICA/B pathway. *Journal of experimental clinical cancer research: CR.* 2019;38(1):192.

25. MC G, vdV DL, JW B, et al. Impact of hyperglycemia on the efficacy of chemotherapy-A systematic review of preclinical studies. *Critical reviews in oncology/hematology* 2017; **113**: 235 – 41.
26. H M-S AHM. H N, A S. The effect of diabetes mellitus on pharmacokinetics, pharmacodynamics and adverse drug reactions of anticancer drugs. *Journal of cellular physiology*. 2019;234(11):19339–51.
27. C LLYC. C, S F, Q X, J Z. Glucose negatively affects Nrf2/SKN-1-mediated innate immunity in *C. elegans*. *Aging*. 2018;10(11):3089–103.
28. PB P. H S, K S, SK S, KV R. Aldose Reductase Mediates NLRP3 Inflammasome-Initiated Innate Immune Response in Hyperglycemia-Induced Thp1 Monocytes and Male Mice. *Endocrinology*. 2017;158(10):3661–75.
29. Hoo ZH, Candlish J, Teare D. What is an ROC curve? *Emerg Med J*. 2017;34(6):357–9.

Figures

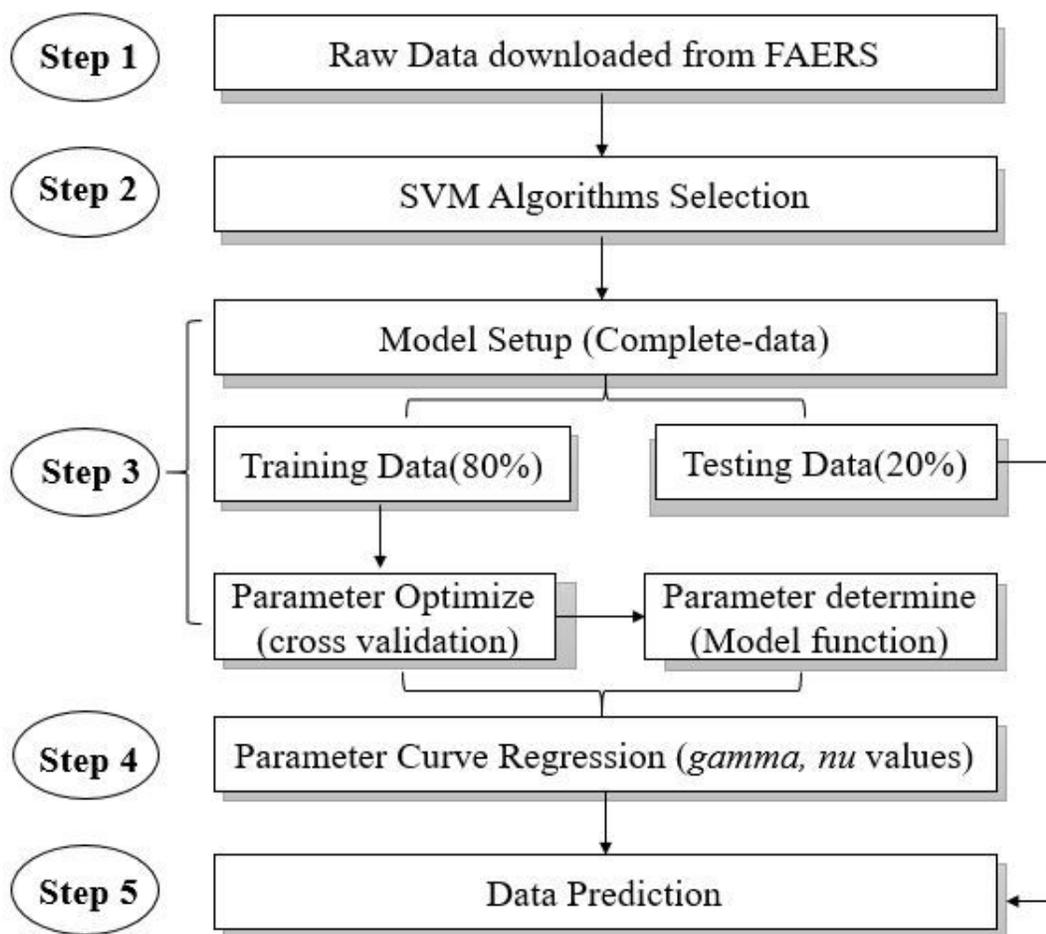


Figure 1

Flow Diagram of the Study Protocol

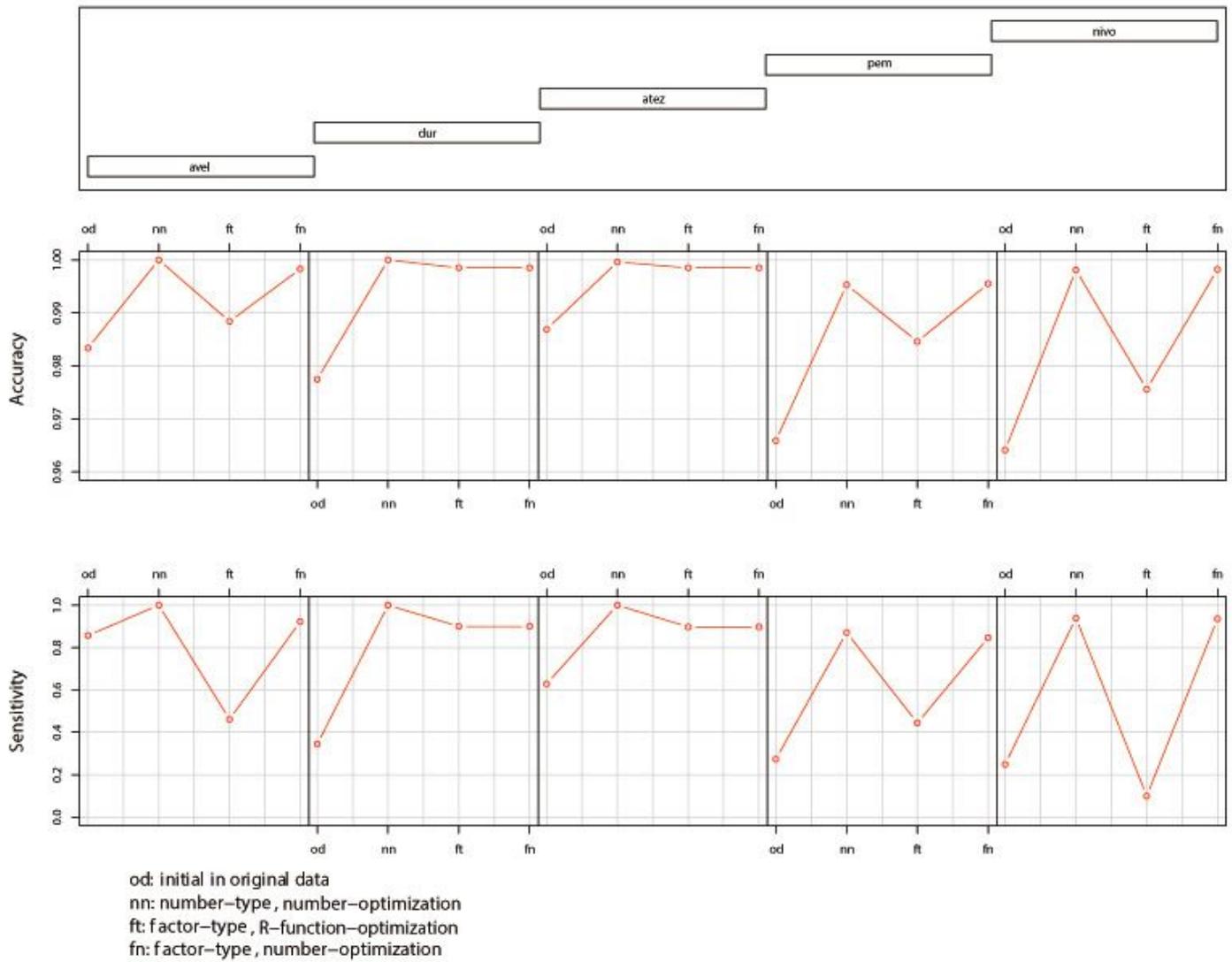


Figure 2

Three SVM Algorithms and Optimization Methods

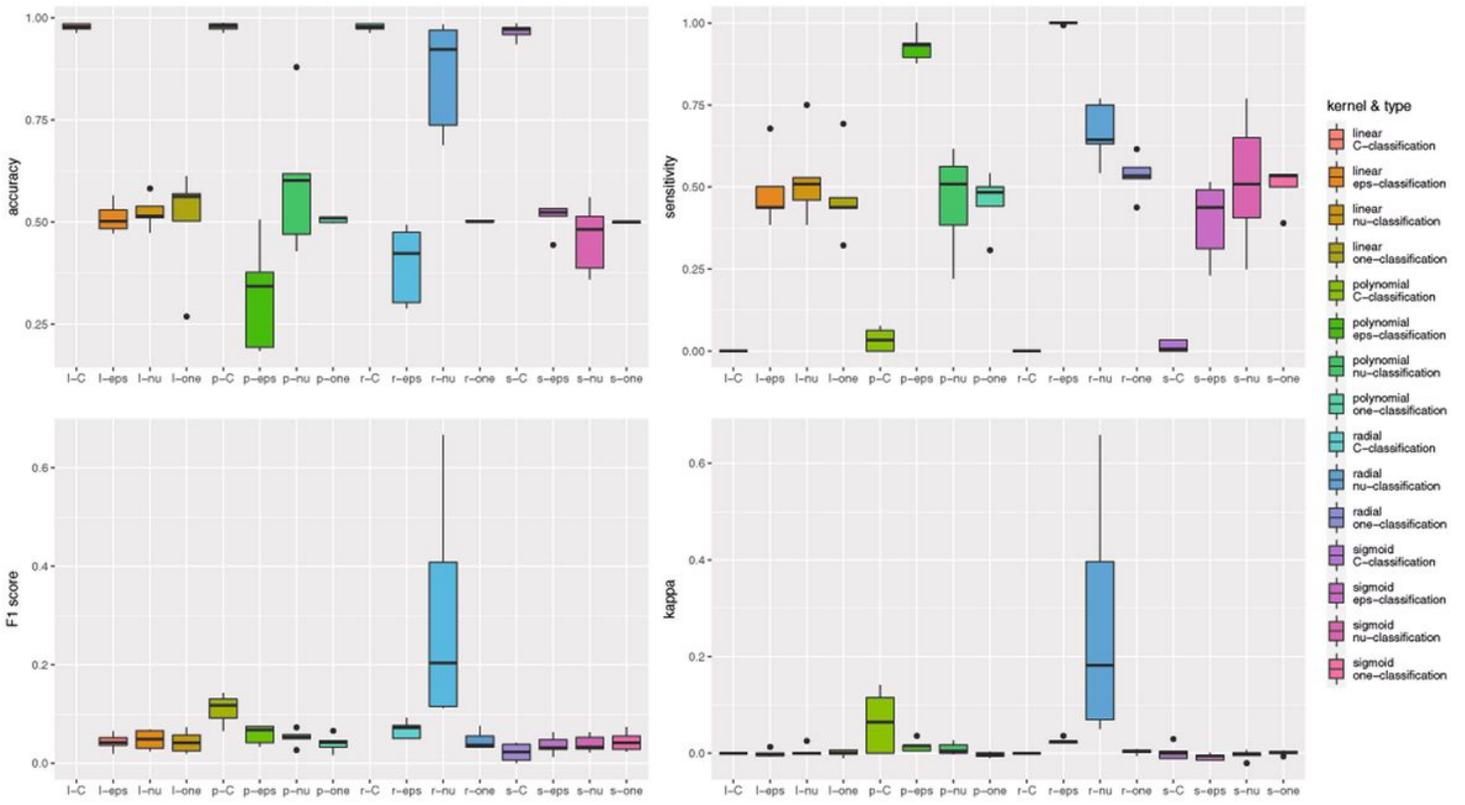


Figure 3

Parameter Algorithm Selection

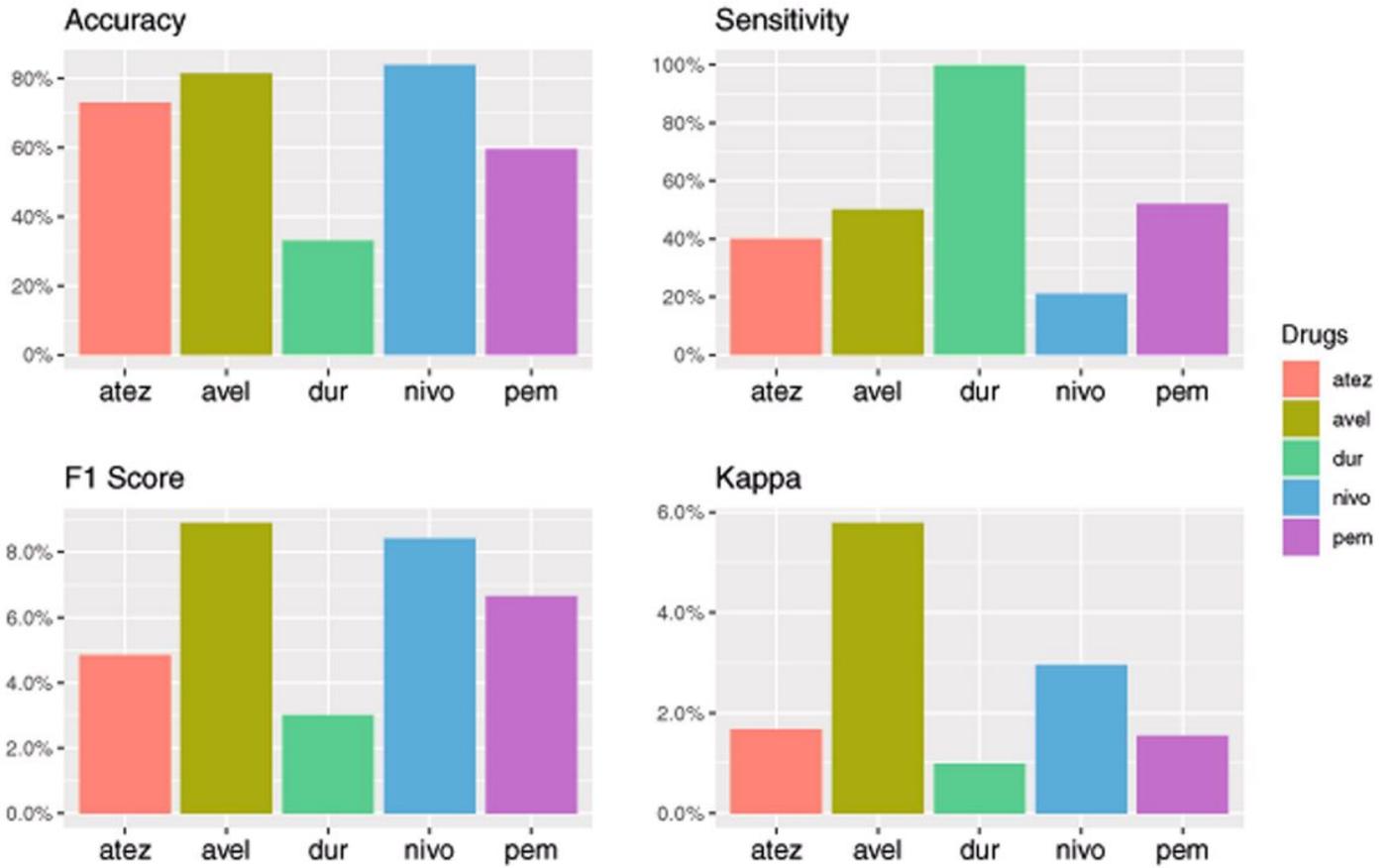


Figure 4

SVM Prediction Model Setup

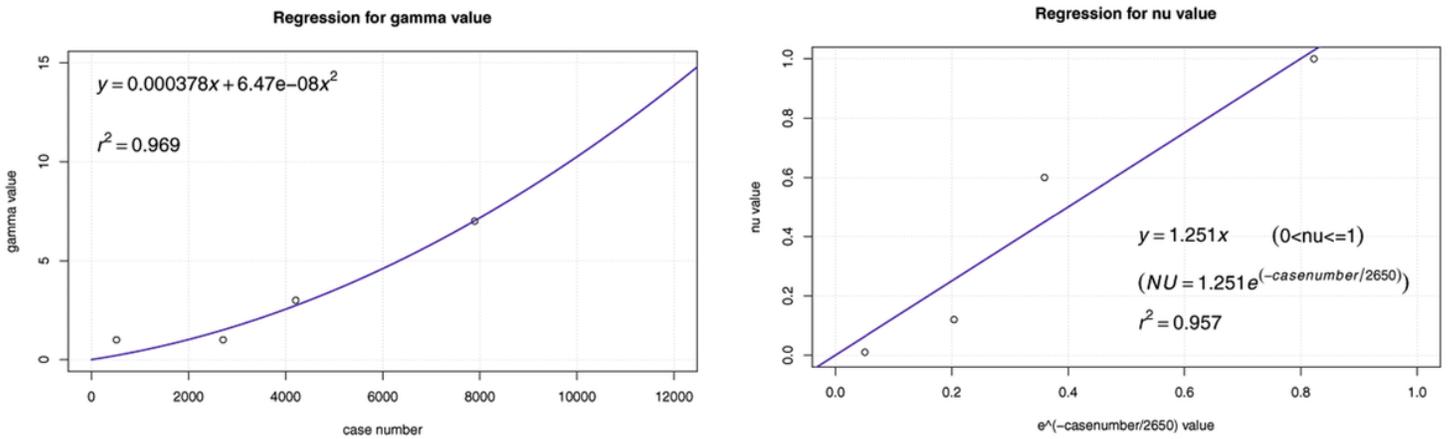


Figure 5

Curve Regressions of gamma and nu Values with Case Number

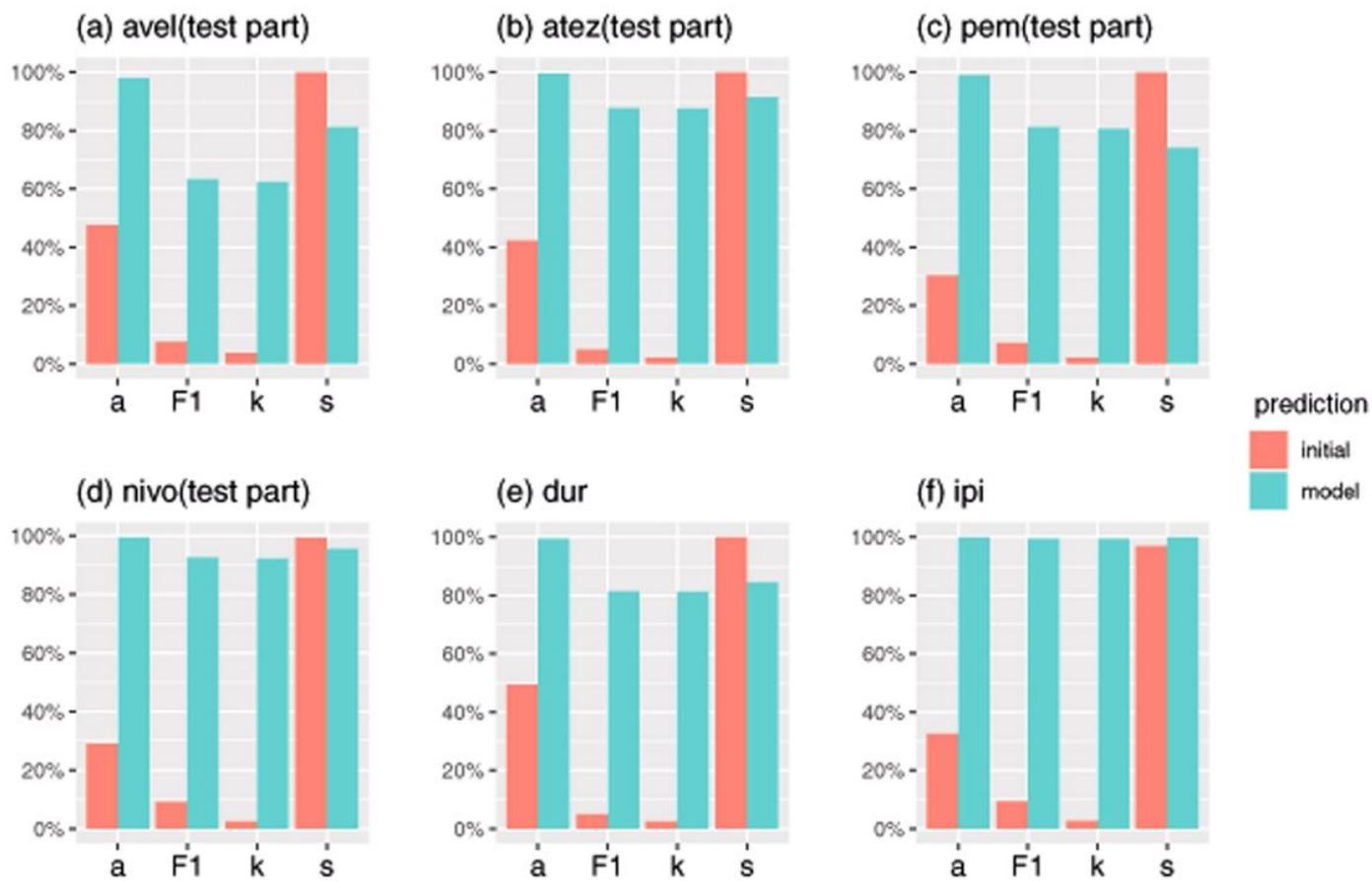


Figure 6

Predictions from Optimized Parameters Compared with Initial SVM Results Note: a = accuracy, F1 = F1 score, k = kappa, s = sensitivity

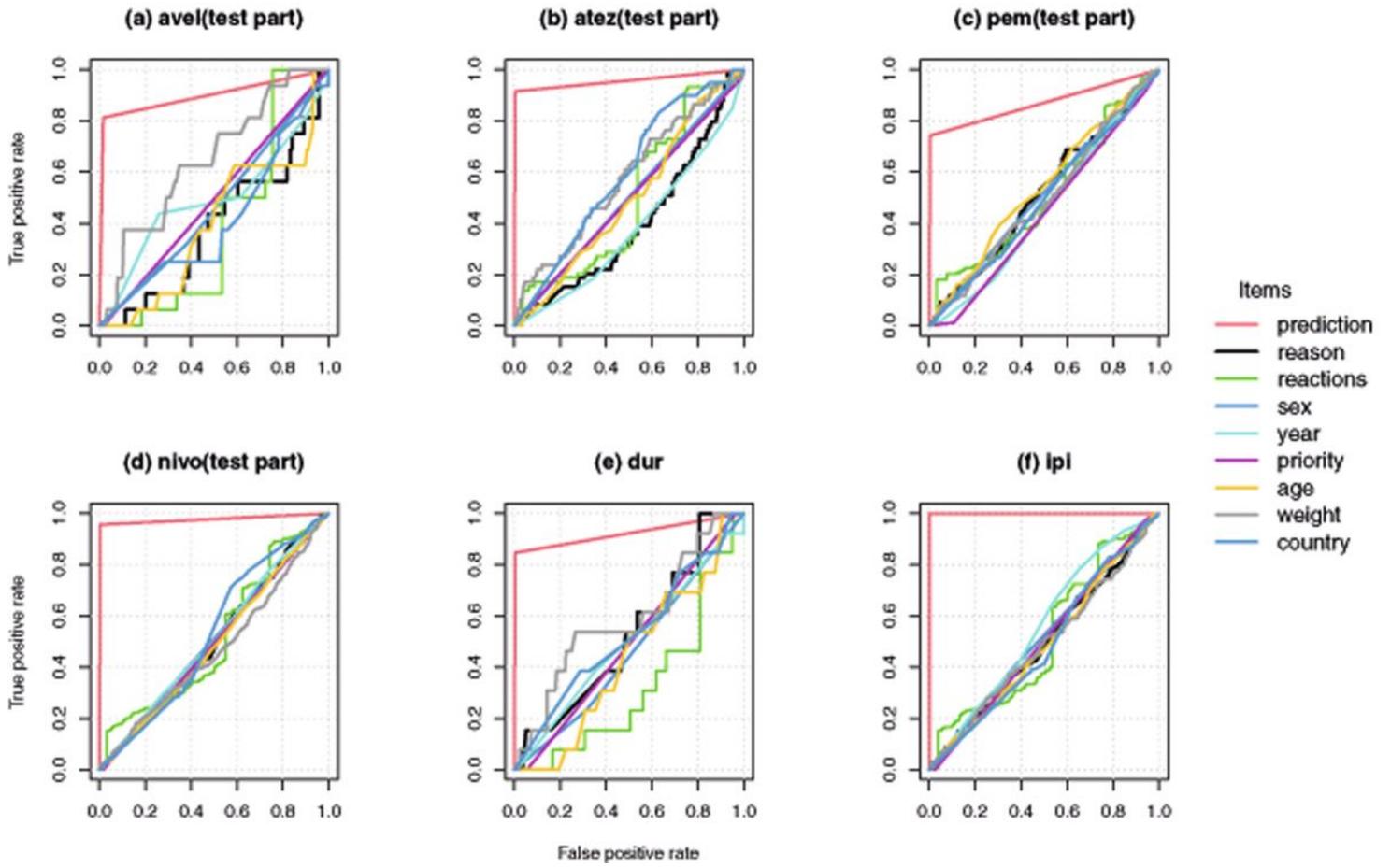


Figure 7

Predictive Evaluations on ROC Curves

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [tables13appendix.docx](#)