medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Machine Learning-Based Prediction of Free IgE Concentration in Allergic Rhinitis Patients Treated with Allergen Immunotherapy and Omalizumab

Kazeem B. Olanrewaju[1] and Laura Marthe Emilie Ngansop Djampou[2]

[1] Prairie View A & M University, Assistant Professor

[2] Prairie View A & M University, Graduate Student

Abstract

Free immunoglobulin E (IgE) concentration is a key biomarker for allergic diseases. Prediction of free IgE concentration can help clinicians diagnose and monitor allergic diseases more effectively. In this study, we used machine learning to predict free IgE concentration in the blood serum of patients with allergic rhinitis who received allergen immunotherapy co-administered with omalizumab. The predictors for free IgE concentration were the number of visits for treatment and baseline checking, and treatment groups (1) omalizumab/ragweed, (2) omalizumab/placebo, (3) placebo/ragweed, and (4) placebo/placebo. Several machine learning algorithms (MLA) were trained with the immunotherapy dataset imported from Immune Tolerance Network (ITN) TrialShare into the Orange data mining platform. The decision tree algorithm model amidst the list of MLAs trained and tested was the best performing model for predicting free IgE concentration, with an R-squared of about 0.6. This study demonstrates that machine learning can be used to predict free IgE concentration with high accuracy. This prediction model could be used to help clinicians diagnose and monitor allergic diseases more effectively.

Keywords: free IgE concentration, machine learning, decision tree, allergen immunotherapy, omalizumab, allergic rhinitis

Background and Significance

Immunoglobulin E (IgE) is an antibody that plays a key role in the development of allergic diseases. When a person with allergies is exposed to an allergen, their body produces IgE antibodies that bind to the allergen[1, 2]. This binding triggers the release of histamine and other inflammatory mediators, which cause the symptoms of an allergic reaction[3]. Free IgE concentration in the blood is a measure of the amount of IgE antibodies that are not bound to allergens. Free IgE concentration is a key biomarker for allergic diseases, and it can be used to diagnose and monitor these diseases[4, 5]. Allergen immunotherapy (AIT) is a treatment that involves exposing a person to gradually increasing doses of an allergen to desensitize them to the allergen. Omalizumab is a monoclonal antibody that targets IgE[6]. It is used to treat severe allergic asthma and other allergic diseases. Prediction of free IgE concentration can help clinicians diagnose and monitor allergic diseases more effectively. For example, if a patient has a high free IgE concentration, it is more likely that they have an allergy. Additionally, monitoring free IgE concentration over time can help clinicians track the effectiveness of AIT and omalizumab treatment[7-9]. Machine learning is a type of artificial intelligence that can be used to learn from data and make predictions. Machine learning algorithms are gradually gaining ground as tools that can be very effective at predicting free IgE concentration and other immunological biomarkers in patients with allergic diseases[10-12]. In this study, we are going to use a machine learning algorithm to predict free IgE in allergic rhinitis patients treated with allergen immunotherapy and omalizumab[5, 6, 9, 13]. The development of machine learningbased prediction models for free IgE concentration in patients with allergic rhinitis treated with AIT and omalizumab would be a significant advance in the field of allergy management. These models could be used to help clinicians diagnose and monitor allergic rhinitis more effectively, and to track the effectiveness of AIT and

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

omalizumab treatment. Potential applications of machine learning-based prediction models for free IgE concentration in patients with allergic rhinitis treated with AIT and omalizumab: ***Diagnosis:*** Machine learning models could be used to develop diagnostic algorithms for allergic rhinitis. These algorithms could be used to identify patients with allergic rhinitis based on their free IgE concentration and other factors[14-17]**.** ***Monitoring:*** Machine learning models could be used to monitor the effectiveness of AIT and omalizumab treatment in patients with allergic rhinitis. This would allow clinicians to adjust the treatment plan as needed to ensure that the patient is receiving the most effective treatment possible[15, 18, 19]. ***Personalized treatment:*** Machine learning models could be used to develop personalized treatment plans for patients with allergic rhinitis. These treatment plans could be based on the patient's free IgE concentration, other factors, and the patient's preferences[20-22]. Overall, the development of machine learning-based prediction models for free IgE concentration in patients with allergic rhinitis treated with AIT and omalizumab has the potential to significantly improve the management of allergic rhinitis.

Research Methodology:

Dataset Extraction from Immune Tolerance Network/TrialShare

The Immune Tolerance Network ITN/TrialShare[23] repository contains data from clinical trials of immunotherapy, including both mechanistic and raw clinical datasets. The ITN TrialShare dataset will be used as the primary dataset for this study, as it is expected to be the most comprehensive and up-to-date dataset available. To extract the immunotherapy biomarker dataset[9, 13] from the ITN TrialShare repository, the following steps will be followed:

1. Access the data repository by logging into the TrialShare link on the Immune Tolerance Network (ITN) page.
2. Filter the studies of interest by selecting the therapeutic area (allergy), study type (clinical trial), age group (all adults), phase (all phases), and condition (allergic rhinitis).
3. Select the appropriate study or studies by clicking the "GO TO STUDY" button.
4. On the study page, export the dataset as an Excel file (xlsx).

The dataset will then be uploaded into Orange Data Mining Platform[24] (the Machine Learning Software Platform (python-based) for this study) as CSV files to undergo the various steps leading to the development of the machine learning algorithm model for Free Ige Concentration prediction.

Dataset Preprocessing

The first step in this process is to visualize and preprocess the data to address a range of issues that may affect the learning of the data by the several machine learning algorithms considered before selecting the best performing algorithm for the free IgE concentration predictive model. Preprocessing of the dataset will involve the following steps:

1. Data cleaning: This step involves addressing inconsistency, duplicity, noise, and missing data.
2. Data integration: This step involves combining relevant datasets from the mechanistic and the clinical datasets.
3. Data reduction: This step involves reducing data attributes and dimensionality to features with significant impact (high information gain) on the target variable (s).
4. Data transformation: This step involves creating a function that can map old values into a new set of values through smoothing functions (Fourier transform), feature construction, aggregation, and normalization such that each old value is identified with a new value.
5. Data discretization: This step involves applying interval or conceptual labels to datasets such as age groups and immunotherapy patients' weekly injection duration.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Once the data has been preprocessed, it will be ready for use in the machine learning algorithm model development process.

Machine Learning Algorithm Training and Testing

The preprocessed dataset will advance to the next stage within the Orange data mining platform, which focuses on machine learning and evaluation, specifically the prediction phase. During this step, the identification and categorization of the dataset attributes into predictors and target variables will be done. Free IgE concentration will be designated as the target variable while the treatment group ((1) omalizumab/ragweed, (2) omalizumab/placebo, (3) placebo/ragweed, and (4) placebo/placebo), V_visit_common (weekly visit no treatment administered), and V_Visit_num (treatment IgE assessment weekly visit with treatment administered). Other attributes were not considered in the predictors list because they are attributes meant for participant identification. These attributes are treated as meta-data or simply ignored in the machine learning algorithm procedures. Typically, machine learning employs a single variable as the target, but occasionally, multiple variables can be designated as targets simultaneously. This machine learning approach, where there is a well-defined target variable, is known as supervised machine learning. Among the supervised machine learning algorithms under consideration are k-nearest neighbors (kNN), decision trees, linear regression, random forests, artificial neural networks (ANN), gradient boosting, and support vector machines.

The training of the machine learning algorithm will consist of using 70-80% of the dataset for training and reserving 20-30% for testing. During this process, the algorithm was tuned (sensitivity analysis) by employing techniques such as cross-validation with various fold numbers, bootstrapping for random forest algorithms, and adjusting algorithm parameters, such as the number of nearest neighbors in kNN or modifying neurons and activation methods in ANN. After the training phase is completed, the testing of the machine learning algorithm on the dataset immediately follows. This machine learning process within the Orange Data Mining platform will involve a sequence of widgets encompassing data manipulation, transformation, visualization, modeling, and evaluation. These widgets are Python-based tools provided by the Orange Data Mining platform.

Initial datasets have about 5.2 % missing values which are completely from the free IgE concentration. The missing data were addressed during the preprocessing stage in three ways. These include filling the missing free IgE concentration values with the average/most frequent values, replacing them with random values, removing rows with missing values, and replacing them with predicted values from the bestperforming MLA model. Thus, the machine learning algorithm with the best-performing metrics, including coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) will be chosen as the Machine Learning Algorithm Model for predicting the Free IgE concentration. Amidst the list of MLA performance metrics, the coefficient of determination (R^2) will be the top determinant metric in making the best selection.

Results and Discussion

Different scenarios for addressing missing values and the performance metric results

<u>Scenario 1: imputing missing values with the average/most frequent value</u>

The missing values in the datasets were addressed by imputing missing free IgE values with the average/most frequent with cross-validation of 10 folds and 80 % sampled dataset for MLA training and 20% dataset to test the MLA model for free IgE concentration prediction. In addition, the parameter for each MLA considered was tuned to achieve the optimum performance metric. The resultant performance metric for this approach can be seen in the Test and Score widget window below. Figure 1 below shows the metric performance measured for MLA trained and tested in this scenario. Thus, the decision tree, the best

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

performing MLA model, has its optimum parameter at 29 minimum number of instances in leaves with split subsets not smaller than 80. The worst-performing MLA model, linear regression, has its optimum parameter at alpha 30 for the ridge regression as depicted in Figure 2.

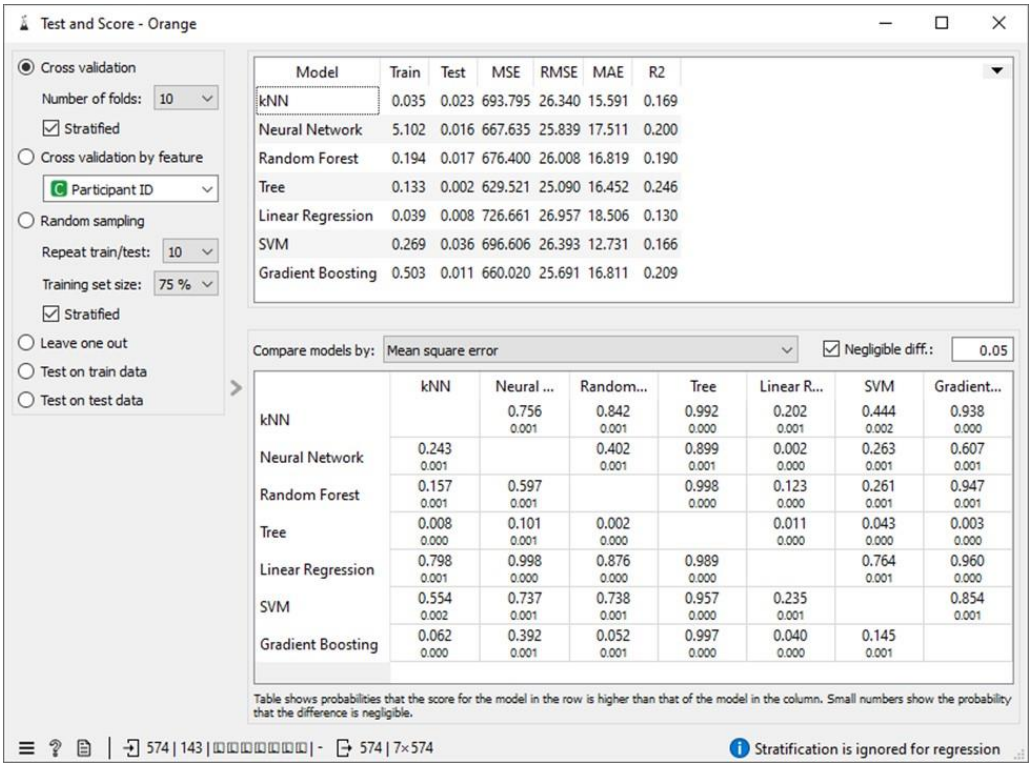


Figure 1: Test and Score showing the model evaluation metric for the MLAs under scenario 1

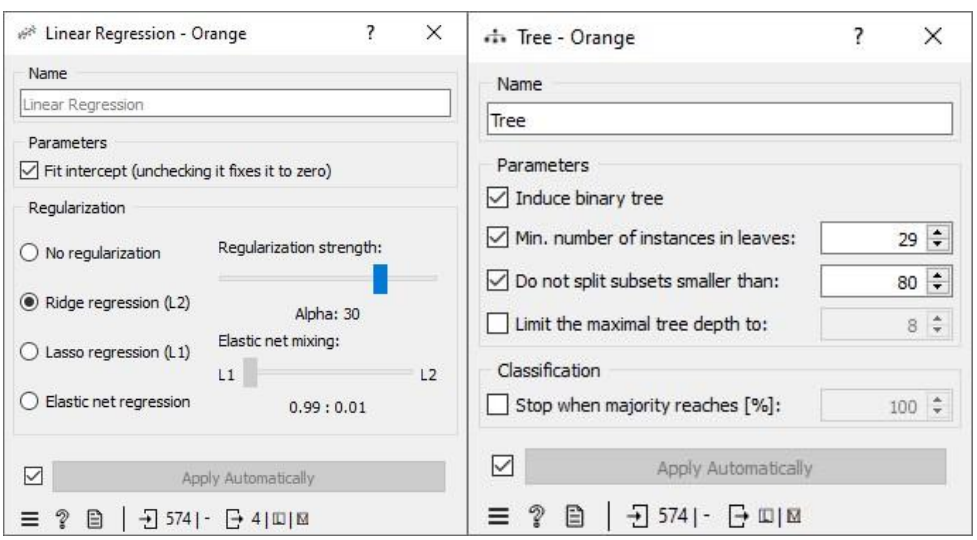


Figure 2: Linear regression and Decision Tree parameter windows under scenario 1

<u>Scenario 2: Replacing missing values with random value</u>

Missing values in the dataset were replaced with random values. Then, the dataset was split into 10 folds, and 80% of the data was used to train the machine learning algorithms. The remaining 20% of the data was

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

used to test the algorithms and to tune their parameters. The decision tree algorithm was the bestperforming algorithm, with the lowest errors measured relative to other algorithms. Its optimum parameter was 29 minimum number of instances in leaves with split subsets not smaller than 120. The worst performing algorithm was support vector machine (SVM), with its optimum parameters at Cost (c) of 24.00 and regression loss epsilon (ε) equals 4.10 with RBF kernel.

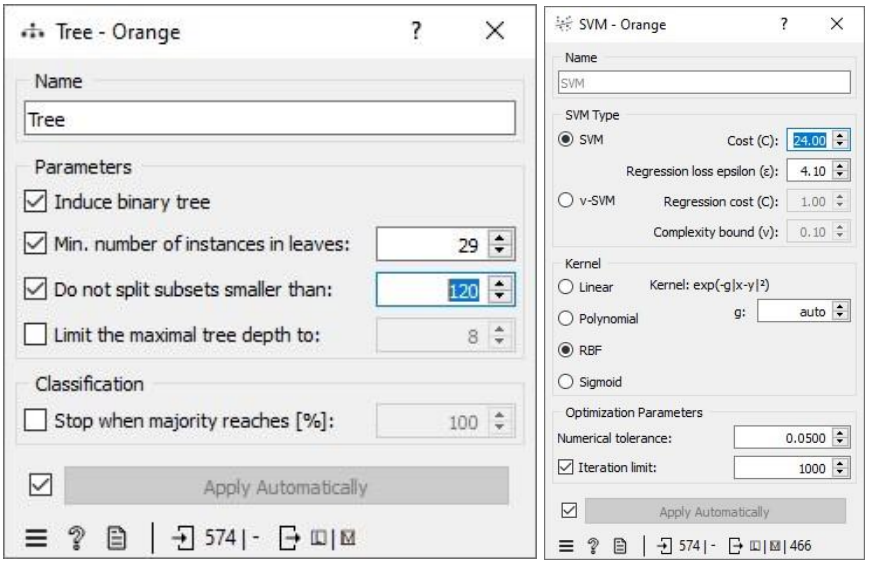


Figure 3. Decision Tree and Support Vector Machine parameter windows under scenario 2

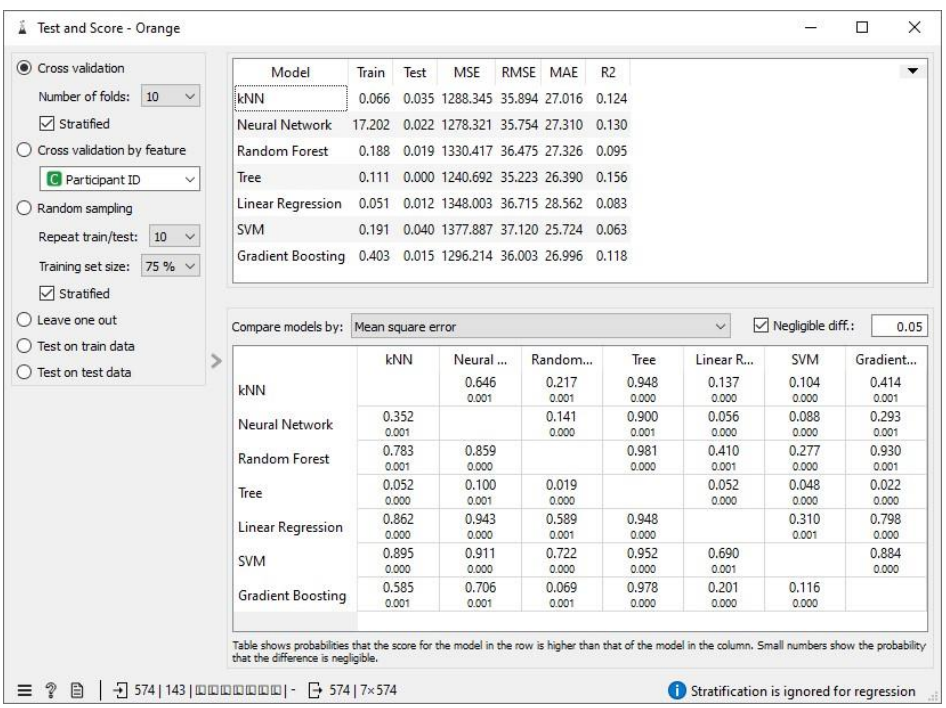
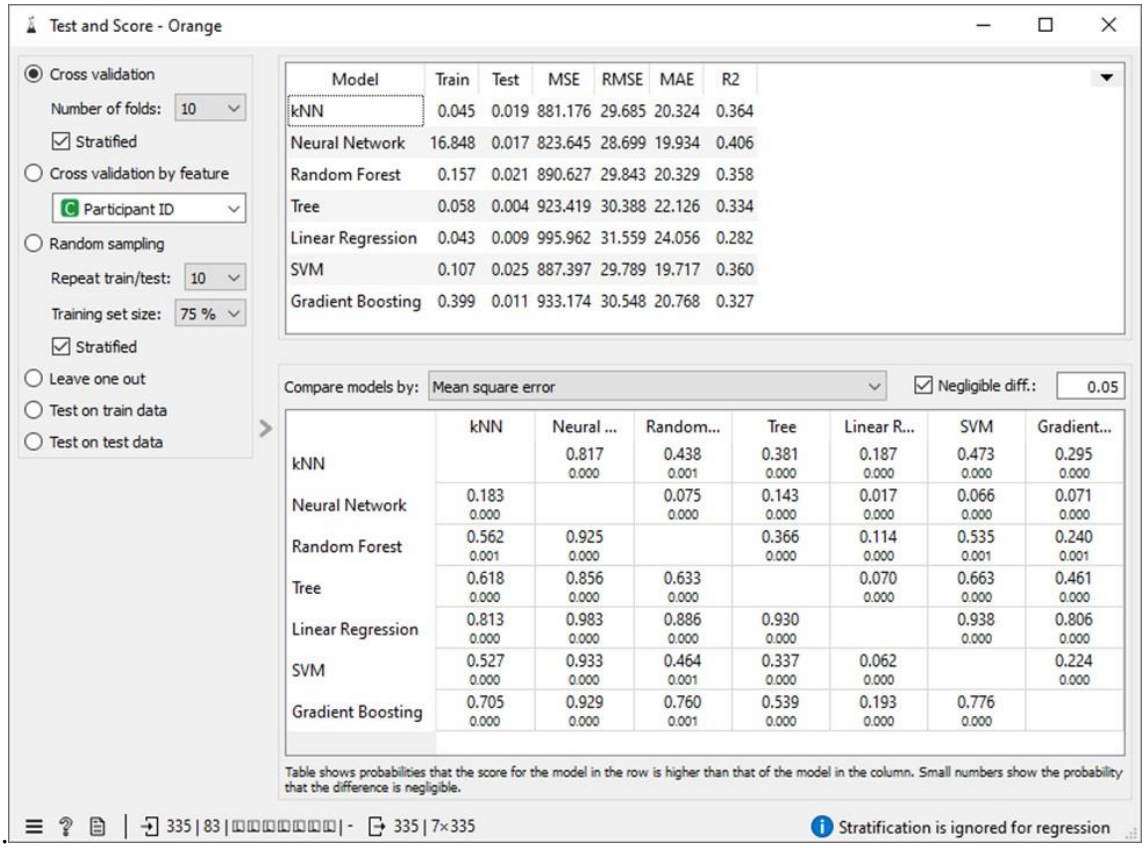


Figure 4: Test and Score showing the model evaluation metric for the MLAs under scenario 2

<u>Scenario 3: Removing rows with missing values option</u>

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Missing values in the dataset were handled by applying 10-fold cross-validation. We allocated 80% of the dataset for training machine learning algorithms (MLA) and reserved 20% for testing the MLA models used to predict free IgE concentration. This process led to a reduction in the dataset size from 717 instances to 418. The parameters of each MLA were fine-tuned to achieve optimal performance metrics. The results of this approach, including the performance metrics, are available in the Test and Score widget window. Figure 5 displays the performance metrics for the MLAs trained and tested in this context. Among the considered MLAs, the neural network emerged as the best-performing model, as illustrated in Figure 6. Its optimal configuration included three hidden layers with 50, 100, and 50 neurons, utilizing ReLu activation, and a regularization parameter (α) set to 0.05. The algorithm was trained and tested for a maximum of 500 iterations. Conversely, the least effective MLA model among those considered was linear regression, specifically the ridge regression variant, which achieved its best performance with an alpha value of 30, as indicated in Figure 6.



Model	Train	Test	MSE	RMSE	MAE	R2
kNN	0.045	0.019	881.176	29.685	20.324	0.364
Neural Network	16.848	0.017	823.645	28.699	19.934	0.406
Random Forest	0.157	0.021	890.627	29.843	20.329	0.358
Tree	0.058	0.004	923.419	30.388	22.126	0.334
Linear Regression	0.043	0.009	995.962	31.559	24.056	0.282
SVM	0.107	0.025	887.397	29.789	19.717	0.360
Gradient Boosting	0.399	0.011	933.174	30.548	20.768	0.327

	kNN	Neural ...	Random...	Tree	Linear R...	SVM	Gradient...
kNN		0.817 0.000	0.438 0.001	0.381 0.000	0.187 0.000	0.473 0.000	0.295 0.000
Neural Network	0.183 0.000		0.075 0.000	0.143 0.000	0.017 0.000	0.066 0.000	0.071 0.000
Random Forest	0.562 0.001	0.925 0.000		0.366 0.000	0.114 0.000	0.535 0.001	0.240 0.001
Tree	0.618 0.000	0.856 0.000	0.633 0.000		0.070 0.000	0.663 0.000	0.461 0.000
Linear Regression	0.813 0.000	0.983 0.000	0.886 0.000	0.930 0.000		0.938 0.000	0.806 0.000
SVM	0.527 0.000	0.933 0.000	0.464 0.001	0.337 0.000	0.062 0.000		0.224 0.000
Gradient Boosting	0.705 0.000	0.929 0.000	0.760 0.001	0.539 0.000	0.193 0.000	0.776 0.000	

Figure 5: Test and Score showing the model evaluation metric for the MLAs under scenario 3

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

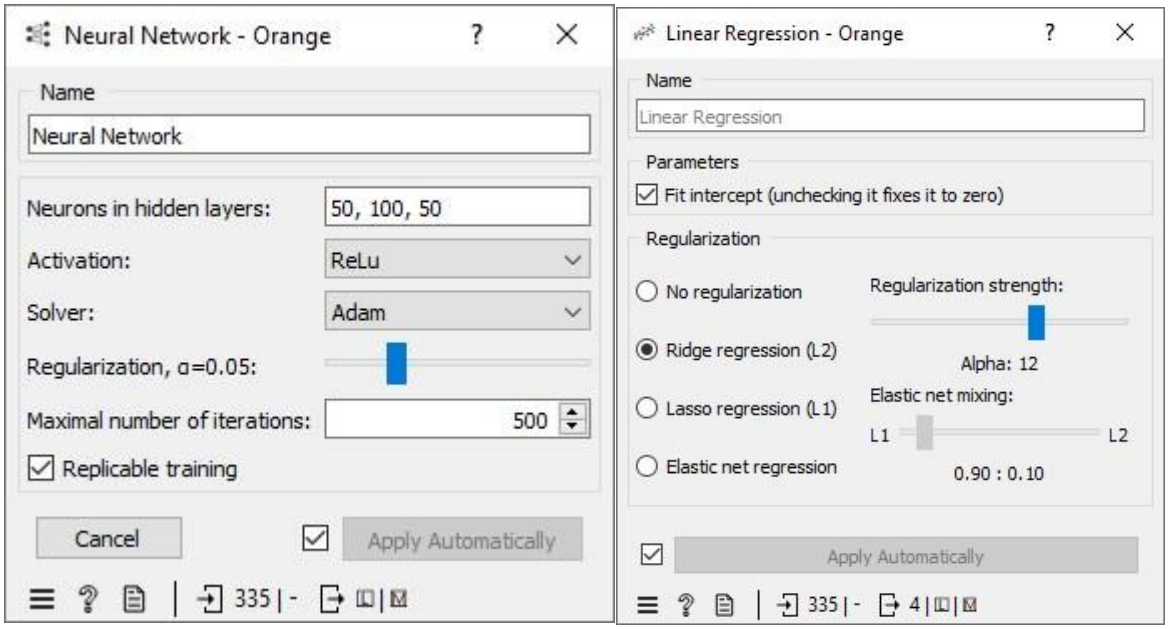


Figure 6. Neural Network and Linear Regression parameter windows under scenario 3

Scenario 4: Replacing with missing values with machine-learning-algorithm-generated data

To handle missing values in the datasets, we employed a method of imputation by replacing the missing free IgE values with predictions made by an MLA (Machine Learning Algorithm) specifically designed for this task. These predictions were generated using measured features that corresponded to the missing free IgE concentrations, serving as predictor variables. Notably, the predictive model used for this purpose was a neural network MLA. The machine pipeline for this scenario can be seen below in Figure 7. Our approach involved training the MLAs using 80% of the dataset with a 10-fold cross-validation strategy, while the remaining 20% of the dataset was reserved for testing. This division allowed us to identify the most proficient MLA for predicting free IgE concentrations. Furthermore, we meticulously fine-tuned the parameters of each MLA to optimize various performance metrics, including coefficient determination (R2), mean absolute error (MAE), root mean square error (RMSE), and mean square error. The performance metrics resulting from this method can be observed in the Test and Score widget window, and Figure 8 visually represents the performance metrics for the MLAs trained and tested within this context. The decision tree MLA emerged as the top-performing model, as depicted in Figure 8. Its optimal configuration featured a minimum of 29 instances in leaves, with split subsets not smaller than 120. The neural network, which ranked as the next best-performing model, was configured with three hidden layers comprising 50, 100, and 50 neurons, utilizing ReLu activation and a regularization parameter (α) set to 0.05. Additionally, the maximum number of iterations employed during the algorithm's training and testing phases was limited to 500. Conversely, the least effective MLA model among those considered was linear regression, particularly the ridge regression variant, which demonstrated its best performance with an alpha value of 12, as displayed in Figure 9.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

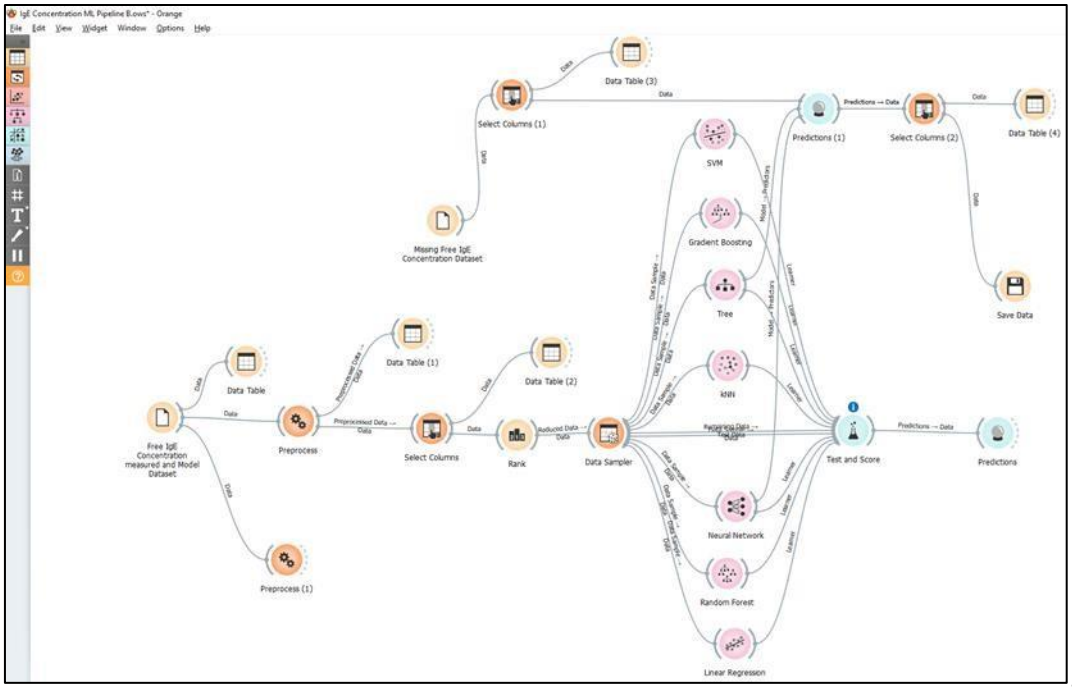
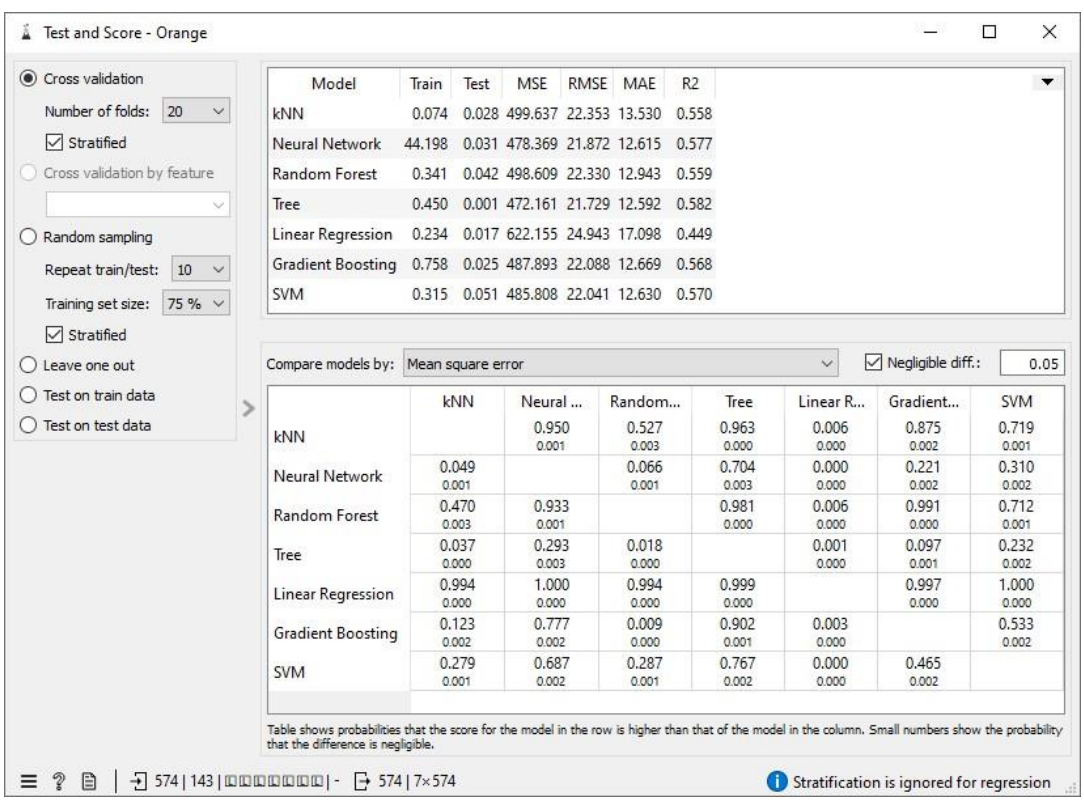


Figure 7: Machine Learning Pipeline for the Prediction of Free Immunoglobin E (IgE)



Test and Score - Orange

Cross validation
Number of folds: 20
Stratified
Cross validation by feature
Random sampling
Repeat train/test: 10
Training set size: 75 %
Stratified
Leave one out
Test on train data
Test on test data

Model	Train	Test	MSE	RMSE	MAE	R2
kNN	0.074	0.028	499.637	22.353	13.530	0.558
Neural Network	44.198	0.031	478.369	21.872	12.615	0.577
Random Forest	0.341	0.042	498.609	22.330	12.943	0.559
Tree	0.450	0.001	472.161	21.729	12.592	0.582
Linear Regression	0.234	0.017	622.155	24.943	17.098	0.449
Gradient Boosting	0.758	0.025	487.893	22.088	12.669	0.568
SVM	0.315	0.051	485.808	22.041	12.630	0.570

Compare models by: Mean square error Negligible diff.: 0.05

	kNN	Neural ...	Random...	Tree	Linear R...	Gradient...	SVM
kNN		0.950 0.001	0.527 0.003	0.963 0.000	0.006 0.000	0.875 0.002	0.719 0.001
Neural Network	0.049 0.001		0.066 0.001	0.704 0.003	0.000 0.000	0.221 0.002	0.310 0.002
Random Forest	0.470 0.003	0.933 0.001		0.981 0.000	0.006 0.000	0.991 0.000	0.712 0.001
Tree	0.037 0.000	0.293 0.003	0.018 0.000		0.001 0.000	0.097 0.001	0.232 0.002
Linear Regression	0.994 0.000	1.000 0.000	0.994 0.000	0.999 0.000		0.997 0.000	1.000 0.000
Gradient Boosting	0.123 0.002	0.777 0.002	0.009 0.000	0.902 0.001	0.003 0.000		0.533 0.002
SVM	0.279 0.001	0.687 0.002	0.287 0.001	0.767 0.002	0.000 0.000	0.465 0.002	

Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.

574 | 143 | - 574 | 7×574 Stratification is ignored for regression

Figure 8: Test and Score showing the model evaluation metric for the MLAs under scenario 4

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

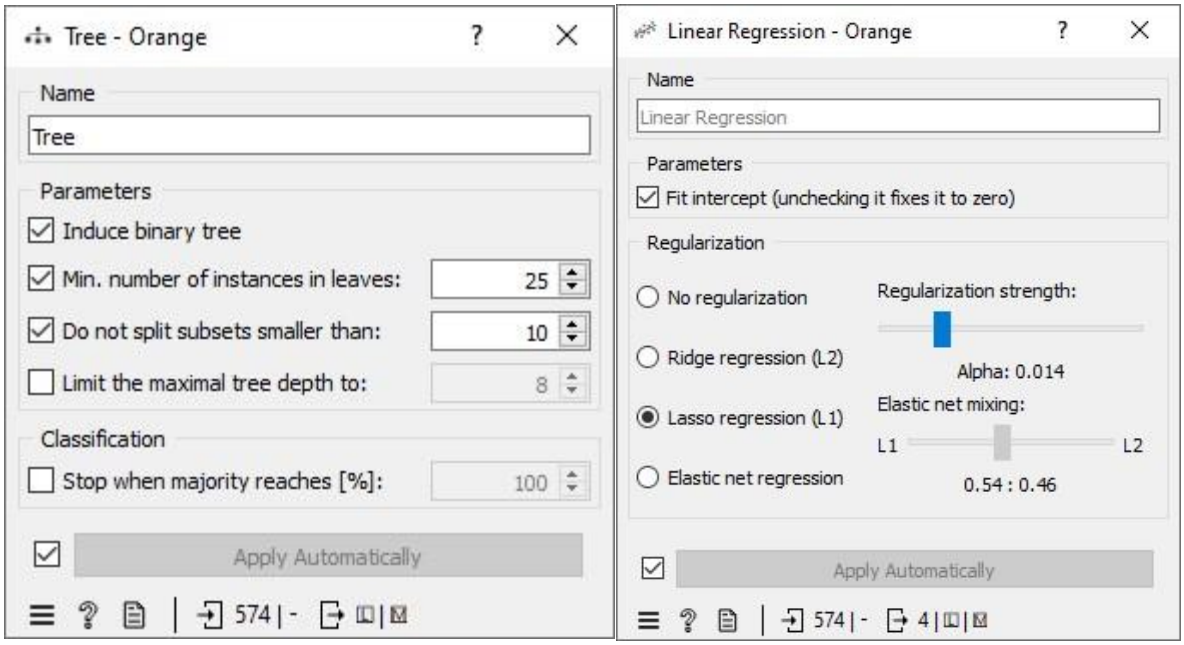


Figure 9. Neural Network and Linear Regression parameter windows under scenario 4

Scenario 4 yields the most favorable performance evaluation results when predicting free IgE concentration based on three crucial predictor variables: treatment group (TRT_Grp), weekly visits without treatment administered (V_visit_num), and weekly visits with treatment administered (V_visit_common). The coefficient of determination (R2) for scenario 4 which is 0.582 is significant for a dataset involving humans. Additionally, we conducted an assessment of information gain for these three predictor variables, which measures their impact on predicting the free IgE concentration. This analysis revealed that the treatment groups (TRT_Grp) exert a substantial influence on predicting free IgE concentration. In contrast, the impact of weekly visits, both with and without treatment administered, was found to be negligible in terms of predicting free IgE concentration. These findings align with the insights provided in the Rank widget of the Orange Data Mining platform, as illustrated in Figure 10. It reinforces the statement made on the Free IgE Dataset Properties page of the ITN TrialShare, which indicates that the dataset lacks a clear relationship with specific visits.

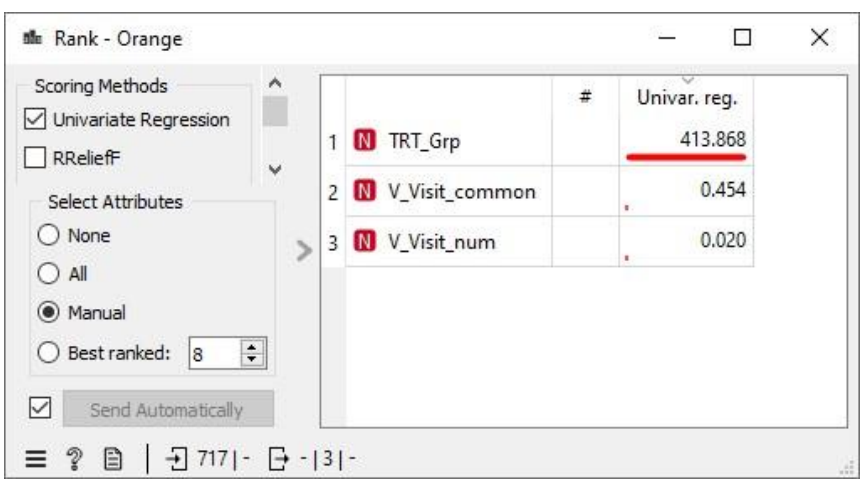


Figure 10. Impact of the three attributes in predicting the Free IgE concentration

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1 shows assessment metrics for all the machine learning algorithms (MLAs) trained and tested with the Free IgE concentration dataset. The neural network algorithm spent almost 45 seconds in learning the dataset and 0.031 seconds in testing while 0.074 seconds was used to train the kNN model and 0.028 seconds was used to the test data. The best performing MLA model in Table 1 is the decision tree with coefficient of determination, $R^2 = 0.582$; and the lowest errors measured relative to other MLA models. These errors include mean square error (MSE =472.161), root mean square error (RMSE= 21.79), and mean absolute error (MAE=12.592).

Table 1. Showing the metric for the scenario 4

	Scenario 4					
MLA Model	**Train time (s)**	**Test time (s)**	**MSE**	**RMSE**	**MAE**	**R^2**
kNN	0.074	0.028	499.637	22.353	13.530	0.558
Neural Network	44.198	0.031	478.369	21.872	12.615	0.577
Random Forest	0.341	0.042	498.609	22.330	12.943	0.559
Tree	**0.450**	**0.450**	**472.161**	**21.729**	**12.592**	**0.582**
Linear regression	0.234	0.017	622.155	24.943	17.098	0.449
Gradient Boosting	0.785	0.025	487.893	22.088	12.699	0.568
SVM	0.315	0.051	485.808	22.041	12.630	0.570

Conclusion

This study aims to develop a machine-learning model for predicting free IgE concentration in patients with allergic rhinitis treated with AIT and omalizumab. The dataset was extracted from Immune Tolerance/TrialShare, which contains data from clinical trials of immunotherapy. The datasets were preprocessed to address inconsistency, duplicity, noise, and missing data. Several machine learning algorithms were trained and tested on the preprocessed dataset. The algorithm with the best-performing metrics was chosen as the Machine Learning Algorithm Model for predicting the Free IgE concentration. Potential applications of the machine learning model include diagnosis, monitoring, and personalized treatment of allergic rhinitis. Four scenarios which include replacing the missing values with average/most frequent, random values, removing instances with missing free IgE concentration, and replacing missing IgE concentration with MLA predicted values were adopted to address the missing free IgE concentration in the dataset. Scenario 4 gave the best performance for the prediction of free IgE concentration based on the three predictor variables. The coefficient of determination (R2) for scenario 4 is 0.582, which is significant for a dataset involving humans. The treatment group has a major impact on the free IgE concentration prediction, while weekly visits with and with no treatment administered have negligible impact on predicting the free IgE concentration. The decision tree algorithm is the bestperforming algorithm with the lowest errors measured relative to other MLA models.

References

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1. J. Eckl-Dorna, S. Villazala-Merino, N.J. Campion, M. Byazrova, A. Filatov, D. Kudlay, A. Karsonova, K. Riabova, M. Khaitov, A. Karaulov, V. Niederberger-Leppin, and R. Valenta, "Tracing IgE-Producing Cells in Allergic Patients," *Cells*, Vol. 8, 9, Aug 28 2019, pp. 994. doi: 10.3390/cells8090994.

2. L.C. Wu and A.A. Zarrin, "The production and regulation of IgE by the immune system," *Nature Reviews Immunology*, Vol. 14, 4 2014, pp. 247-259.

3. P. Stoffersen, P.S. Skov, L.K. Poulsen, and B.M. Jensen, "The Allergen-Specific IgE Concentration Is Important for Optimal Histamine Release From Passively Sensitized Basophils," *Frontiers in allergy*, Vol. 3 2022, pp. 875119.

4. M. Carlsson, L. Thorell, A. Sjölander, and S. Larsson-Faria, "Variability of total and free IgE levels and IgE receptor expression in allergic subjects in and out of pollen season," *Scandinavian Journal of Immunology*, Vol. 81, 4 2015, pp. 240-248.

5. J. Corren, G. Shapiro, J. Reimann, Y. Deniz, D. Wong, D. Adelman, and A. Togias, "Allergen skin tests and free IgE levels during reduction and cessation of omalizumab therapy," *Journal of allergy and clinical immunology*, Vol. 121, 2 2008, pp. 506-511.

6. P.P. Belliveau, "Omalizumab: a monoclonal anti-IgE antibody," *Medscape General Medicine*, Vol. 7, 1 2005, pp. 27.

7. R.G. Hamilton, "Assessment of human allergic diseases," *Clinical Immunology*, Elsevier, 2019, pp. 1283-1295. e1.

8. R.G. Hamilton, W. Hemmer, A. Nopp, and J. Kleine-Tebbe, "Advances in IgE testing for diagnosis of allergic disease," *The Journal of Allergy and Clinical Immunology: In Practice*, Vol. 8, 8 2020, pp. 24952504.

9. T.B. Casale, W.W. Busse, J.N. Kline, Z.K. Ballas, M.H. Moss, R.G. Townley, M. Mokhtarani, V. Seyfert-Margolis, A. Asare, and K. Bateman, "Omalizumab pretreatment decreases acute reactions after rush immunotherapy for ragweed-induced seasonal allergic rhinitis," *Journal of Allergy and Clinical Immunology*, Vol. 117, 1 2006, pp. 134-140.

10. F. Dall'Antonia and W. Keller, "SPADE web service for prediction of allergen IgE epitopes," *Nucleic acids research*, Vol. 47, W1 2019, pp. W496-W501.

11. H. Zhou, W. Fan, D. Qin, P. Liu, Z. Gao, H. Lv, W. Zhang, R. Xiang, and Y. Xu, "Development, validation and comparison of artificial neural network and logistic regression models predicting eosinophilic chronic rhinosinusitis with nasal polyps," *Allergy, Asthma & Immunology Research*, Vol. 15, 1 2023, pp. 67.

12. X. Yu, M. Liu, X. Li, Y. Zhang, and B. Tao, "Qualitative and quantitative prediction of food allergen epitopes based on machine learning combined with in vitro experimental validation," *Food Chemistry*, Vol. 405 2023, pp. 134796.

13. S. Klunker, L.R. Saggar, V. Seyfert-Margolis, A.L. Asare, T.B. Casale, S.R. Durham, J.N. Francis, and Immune Tolerance Network Group, "Combination treatment with omalizumab and rush

medRxiv preprint doi: https://doi.org/10.1101/2023.09.29.23296326; this version posted October 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

immunotherapy for ragweed-induced allergic rhinitis: Inhibition of IgE-facilitated allergen binding," *Journal of allergy and clinical immunology*, Vol. 120, 3 2007, pp. 688-695.

14. G. Hoffmann, A. Bietenbeck, R. Lichtinghagen, and F. Klawonn, "Using machine learning techniques to generate laboratory diagnostic pathways—a case study," *J Lab Precis Med*, Vol. 3, 6 2018.

15. A.L. Beam and I.S. Kohane, "Big data and machine learning in health care," *Jama*, Vol. 319, 13 2018, pp. 1317-1318.

16. D. Goodman-Meza, A. Rudas, J.N. Chiang, P.C. Adamson, J. Ebinger, N. Sun, P. Botting, J.A. Fulcher, F.G. Saab, and R. Brook, "A machine learning algorithm to increase COVID-19 inpatient diagnostic capacity," *Plos one*, Vol. 15, 9 2020, pp. e0239474.

17. S.J. Kim, K.J. Cho, and S. Oh, "Development of machine learning models for diagnosis of glaucoma," *PloS one*, Vol. 12, 5 2017, pp. e0177726.

18. S. Trebeschi, Z. Bodalal, T.N. Boellaard, T.M. Tareco Bucho, S.G. Drago, I. Kurilova, A.M. CalinVainak, A. Delli Pizzi, M. Muller, and K. Hummelink, "Prognostic value of deep learning-mediated treatment monitoring in lung cancer patients receiving immunotherapy," *Frontiers in Oncology*, Vol. 11 2021, pp. 609054.

19. J. Kong, D. Ha, J. Lee, I. Kim, M. Park, S. Im, K. Shin, and S. Kim, "Network-based machine learning approach to predict immunotherapy response in cancer patients," *Nature Communications*, Vol. 13, 1 2022, pp. 3703.

20. R. Kavya, J. Christopher, S. Panda, and Y.B. Lazarus, "Machine learning and XAI approaches for allergy diagnosis," *Biomedical Signal Processing and Control*, Vol. 69 2021, pp. 102681.

21. G. Hurault, E. Domínguez-Hüttinger, S.M. Langan, H.C. Williams, and R.J. Tanaka, "Personalized prediction of daily eczema severity scores using a mechanistic machine learning model," *Clinical & Experimental Allergy*, Vol. 50, 11 2020, pp. 1258-1266.

22. M. Sebastiani, C. Vacchi, A. Manfredi, and G. Cassone, "Personalized medicine and machine learning: A roadmap for the future," *Journal of Clinical Medicine*, Vol. 11, 14 2022, pp. 4110.

23. Immune Tolerance Network, "ITN TrialShare: /home,", Vol. 2023, Sep 16, September 16, 2023.

24. J. Demšar, T. Curk, A. Erjavec, Č Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, and A. Starič, "Orange: data mining toolbox in Python," *the Journal of machine Learning research*, Vol. 14, 1 2013, pp. 2349-2353.