

A Predictive Model for the Risk of Infertility in Men Using Machine Learning

Senem Koç¹, Leman Tomak², Erdem Karabulut³

¹Nişantaşı University Faculty of Medicine, Department of Medical Statistics, İstanbul, Türkiye

²Ondokuz Mayıs University Faculty of Medicine, Department of Biostatistics, Samsun, Türkiye

³Hacettepe University Faculty of Medicine, Department of Biostatistics, Ankara, Türkiye

Abstract

Infertility is a worldwide problem and causes considerable social, emotional and psychological stress between couples and among families. This study is aimed at determining the machine learning classifier capable of developing the most effective predictive model to determine the risk of infertility in men by genetic and external factors. The data set was collected at Ondokuz Mayıs University in the department of Urology. The model was formulated using supervised learning methods and by algorithms like Decision Tree, K Nearest Neighbor, Naive Bayes, Support Vector Machines, Random Forest and Superlearner. Performances of the classifiers were assessed with the area under curve. Results of the performance evaluation showed that Support Vector Machines and Superlearner algorithm had an area under curve of 96% and 97% respectively and this performance outperforming the remained classifier. According to the results for importance of variables sperm concentration, FSH and LH and some genetic factors are the important risk factors for infertility. These findings whenever applied to any patient's record of infertility risk factors can be used to predict the risk of infertility in men. The predictive model developed can be integrated into existing health information systems which can be used by urologist to predict patients' risk of infertility in real time.

Keywords: Classification, superlearner, prediction model, infertility, genetic factors

Introduction

The World Health Organization defined infertility as 12 months of frequent, unprotected intercourse without pregnancy (1). Infertility is a medical and social problem effects about 15% of couples and 40% of these couples are infertile because of male factor (2). Infertility is a worldwide problem and is estimated that only in Turkey 10-15% couples are infertile (3). Male infertility is highly heterogeneous disorder and genetic causes play an important role in male infertility. Karyotypic abnormalities, cystic fibrosis transmembrane conductance regulator gene mutations and microdeletions on Y chromosome are well known genetic causes in azospermic or severely oligozoospermic men (4,5). There are diverse external factors for infertility including age, smoking, obesity etc (3).

Prediction contains variables in the data set to make analysis and find patterns which describes the data structure that can be interpreted by humans (6). Machine learning is a fast-growing field which explores how computers can automatically learn to recognize complex data structures and make a conclusion based on a set of observed data (7).

Nowadays, machine learning applications are a part of our daily life in different areas, for example web searches, spam/email filtering, face recognition programs, and speech recognition programs (8). Machine learning has been used for the classification of different medical data and these results show that the performance of this study was given promising results for different data sets. However, gathering and inventorying of more complex data types, the discovery of new diseases, and the development of new diagnostic methods have raised the need for machine learning methods in the medical area which provides new ways for interpreting complex data sets that researcher are faced (9,10).

Machine learning has been separated into different subfields which deal with different types of learning tasks. Supervised learning is the most common one used in practice and can be grouped into classification and regression. There are many algorithms for classification task with an increasing number and different features day by day, some classification algorithms commonly used are Decision trees (DT), K Nearest Neighbor (KNN), Naive Bayes (NB), Support Vector Machines (SVM), Random Forest (RF) (11,12).

Correspondence: Senem Koç, Nişantaşı University Faculty of Medicine, Department of Medical Statistics, İstanbul, Türkiye

Phone: +90 532 300 34 88

E-mail: koc.senem@nisantasi.edu.tr **ORCID-ID:** orcid.org/0000-0002-8357-3431

Received: 09.12.2021

Accepted: 17.04.2022

Cite this article as: Koç S, Tomak L, Karabulut E. A Predictive Model for the Risk of Infertility in Men Using Machine Learning.

There are different algorithms which can be used for research. The main question is which algorithm will fit on your data well? For statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the learning algorithms. Superlearner (SL) allows researchers to use multiple algorithms to outperform a single algorithm in nonparametric statistical models. So, there is no need to decide before which single technique to use for prediction. Instead, there is the way to use several candidate learners together in different weights by incorporating cross-validation. Cross-validation is an important evaluation technique used to assess the generalization performance of a machine learning model (13-15).

This study focused on diagnosing the risk factors for male infertility disease by machine learning algorithms. The present study aims to compare different machine learning classifiers with different training and testing proportions. Additionally, the results were used to compare SL algorithm and see the advantages of this algorithm.

Materials and Methods

Ethics approval for this study was given by the medical research ethical committee (2017/208, issued June 22, 2017). The data set for the study is collected from 587 infertile and 57 fertile patients between 2007-2018 and published partially in two separate studies (16,17). A total of eleven attributes (ten attributes and one class attribute), age, hormone analysis, FSH (Follicular Stimulating Hormone) level, LH (Luteinizing Hormone) level, routine semen parameters, total testosterone level, sperm concentration, and genetic variations. A total of five categorical and five numerical values are present in the dataset.

In the pre-processing step the data set was checked for missing values. The attribute $gr/gr+b2/b3$ is dropped out from the analysis and for numerical data z-score normalization is used to scale the data. First of all, 80% of the collected data was used to train the algorithms and the remaining 20% was used for testing the performances also the following split ratio used in the study are 70%-30% and 60%-40%.

After removing the missing values, the final data set was composed of 329 (85.5%) infertile and 56 (15.5%) fertile patients. We performed the classification using R, which is open-source statistical software. In the pre-processing step "Plyr" and "ggplot2" were used and analyses were carried out by "caret", "Superlearner", "e071" and "rpart" packages for classification purposes. A 10-fold cross-validation method was used to test the validity of the system.

Machine Learning Algorithms Used for Classification

This study focused on 6 different machine learning algorithms which are DT, RF, NB, KNN, SVM and an ensemble method called SL.

I. Decision Tree

The algorithm uses a tree-like model which starts at the root and builds the tree by choosing the most informative attribute at each step (18). The internal node and the root node pointed with the name of the attribute; the sides are labelled by the most informative attribute values and a leaf node is displayed with different classes. The leaf corresponds to the decision outcomes (19). For attribute selection measure the decision is chosen by the highest gain ratio. The training data set is used while creating decision trees with the C4.5 algorithm. For each node in the tree, the class that divides the sample into the best subsets is determined and this is the highest gain ratio. For continuous variables this algorithm can be used as well (20). The rpart (recursive partitioning and regression trees) package is used for classification trees (21).

II. Random Forest

This algorithm is a kind of ensemble learning that uses a combination of tree estimators. The principle is random sampling by building sub-trees and random subsets of features considered when splitting nodes. The samples are drawn by replacement which is known as bootstrapping and the final model is the majority vote from the creating trees in the forest (22).

From the original data set a sample of N is drawn to construct each tree. When the attributes have been selected, the algorithm forms a repetitive partitioning of the covariates. The best split is chosen as the one optimizing the CART (Classification and Regression Tree) splitting criterion which is the gini index along with the mtry preselected directions (23). This process is repeated until each branch contains less than a pre-specified number node size of observations. After this step, the prediction at a new point is computed by averaging observations falling into the branch of the new point. Each of the M trees gives a prediction which is simply the majority class of the M predicted three (22).

III. Naive Bayes

NB classifier is based on applying bayesian theorem. A probabilistic model estimates conditional probabilities of the dependent variable from the training data and uses them for classification. This classifier assumes that the attributes are independent between the features and are equally important (24). This classifier predicts class membership probability of examples by using the naive conditional independence assumption (25). The bgglm (Bayesian generalized linear model)

is a Bayesian function for generalized linear modelling by different distributions (26).

IV. K Nearest- Neighbor

This classifier is a method on learning by comparing a given test data set with the training data set which resembles it. Here the samples of training data set are defined by n attributes where each example indicates a point in n -dimensional space. This algorithm searches for the K training data samples that are nearest to the unknown example (7).

The performance of a KNN classifier depends on the choice of K and the distance metric. Without foreknowledge, this classifier applies Euclidean distances as the measurement of the closeness between examples. As like in other classifiers majority vote assign the class label (27). Usually, the K parameter in the classifier is chosen experimentally. For each model different numbers of nearest neighbors are chosen and the parameter with the best accuracy is given to defining the classifier (28).

V. Support Vector Machines

This algorithm is mostly used for classifying linear and non-linear patterns. Linear patterns can be easily separated in low dimensions whereas non-linear patterns can't be easily separated. For this task, a set of mathematical functions known as kernels are used. The basic idea for SVM is the use of an optimal hyperplane, which can be used for classification, to solve linearly separable patterns. The optimal hyperplane is selected from the set of hyperplanes for classifying patterns that maximize the margin of the hyperplane. That is the distance from the hyperplane to the closest point of each pattern by maximizing the margin it can correctly classify the given patterns (29).

For non-linear separable patterns, the kernel functions return the inner product between two points in a higher feature space. The training takes place in the feature space, and the data points just appear inside dot products with other points. This is called the "kernel trick", where the non-linear pattern becomes linearly separable (30). The kernel function converts the data into the desired format and for this different kernel are used for non-linear patterns (31).

VI. Superlearner

This algorithm is a cross-validation based method which chooses one or weight of more optimal learners that perform asymptotically as well or better than any of the candidate learners. This prediction algorithm, which applies a set of candidate learners to the observed data can include as many candidate learners to the model if computationally feasible (13,14). Different algorithms can be adapted to SL algorithm such as RF, SVM, NB (14).

The training set train the estimators and the validation set estimate the performance of these estimators. The cross-validation selector is selecting the best performance for the learner on the validation set for SL algorithm. In v -fold cross-validation, the training set is divided into v mutually sets of as nearly equal size. The v set and its complementary validation and training sample give v splits of the learning sample into a training and corresponding validation sample. For every v splits the predictor is applied to the training set, and its risk is estimated by the corresponding validation set. Each learner, risks, and the validation sets are averaged resulting in cross-validated risk. The predictor is selected by the minimum cross-validated risk. The calculated risk is a measure of performance and the model getting minimized risk is the model with the minimum errors in prediction. This algorithm provides a weighted model by using candidate learners. If the model is obtained with a single learner, this gives the discrete SL algorithm. There is no limitation for candidate learners which is the main advantage of this algorithm (15).

Performance Evaluation

The performance of the algorithms selected for the study is evaluated by using Area Under Curve (AUC). The reason is to find common criteria to compare the performances of all algorithms. AUC measures the area underneath the entire area under the Receiver Operating Characteristic (ROC) curve. ROC curve is a graph showing the performance of a classification model at all classification thresholds (32). Also, the performance metric which can be adopted by confusion matrix like accuracy, sensitivity and specificity values have been evaluated for the algorithms (33).

$$\text{Accuracy} = (TP+TN)/(TP+FP+FN+TN) \times 100 \quad (1)$$

$$\text{Sensitivity} = TP/(TP+FN) \times 100 \quad (2)$$

$$\text{Specificity} = TN/(FP+TN) \times 100 \quad (3)$$

In the equations, TP defines the number of true positives; FN defines the number of false negatives; TN defines the number of true negatives; the last is FP which defines the number of false positives (34).

Results

The genetic data for the diagnosis of infertility was evaluated in terms of supervised machine learning algorithms. The C4.5, KNN, NB, SVM and RF algorithms were used as classifiers and compared with SL algorithm according to AUC performance criteria. The C4.5 decision tree algorithm was implemented using the J48 decision tree algorithm, KNN algorithm was implemented using Euclidean distance, Naïve Bayes algorithm

was implemented using the naïve bayes classifier, SVM algorithm was implemented using Radial Basis Kernel, RF algorithm was implemented using bootstrapping while SL algorithm was implemented using different weights simultaneously all available on R program. The models were trained using 10-fold cross-validation and different split ratios.

All classifiers and different split ratio of the overall performance of the data set is shown in Table 1. The split ratio of 80% - 20% using RF algorithm showing better accuracy among all other classifiers whereas SVM showing an AUC of 95% which is the best classifier. The split ratio of 70%-30% using SVM showing performance of 95% whereas split ratio of 60%-40% using RF showing performance of 94% among all other classifiers. The results of sensitivity and specificity show a good performance for all different proportions as well.

According to these results in Figure 1, showing the importance of variables after analysing the data set. Here the first line is sperm concentration following by FSH and LH hormones in the line. Genetic factors sy1291, gr/gr2 and b2/b3 are the important genetic factors according to these findings.

Using the SL algorithm, the predictive model developed using risks of different algorithms and coefficients it was yield an AUC of %96 following by Discrete SL and RF of an AUC of % 95. The coefficient is how much weight SL puts on that model in the weighted-average. The lowest risk is yield by RF given below in Table 2. As seen from the table Bglm will not give any contribution to the analysed model. The weighted model consists of RF, KNN and rpart. Therefore, SL performed as the best algorithm as AUC 97% (Table 3). The following performance is Discrete SL and RF as AUC 96%, respectively.

Discussion

In this study, a machine learning-based prediction model for infertility in men was developed based on genetic data. This study demonstrated that the RF algorithm has higher accuracy than NB, SVM, DT and KNN algorithm irrespective of different split ratio. According to the results, it was discovered that different split ratios can change the classifier used for analysis. The accuracy was highest for RF for a split ratio of 80%-20% whereas the NB classifier showed poor accuracy of 89%. In a study by Noi et al, it was shown that the larger the training sample size, the higher the accuracy (35). Our findings support this results too, where they obtained 90% -95% accuracy for analyzing different data sizes and split ratios for balanced and unbalanced data sets in their studies. The highest performance is adopted by SVM, RF and KNN for the split ratio of 60% -40%. The results of our data set showed that the highest performance was yield by SVM using RBF as kernels and RF classifiers which supports findings in literature (36). In conformity with the results obtained the performance is increased by using RF algorithm for the genetic data set. RF is an important algorithm for medical data sets (37,38). One of the biggest problems in machine learning is which algorithm to use and the ideal split ratio for training and testing data. This study answers these questions by using different classifiers that are compared to the SL algorithm which applies weighted candidate learners to the model.

The SL algorithm picks one or more optimal learners which are called candidate learners to build the algorithm. RF algorithm is a candidate learner which put the biggest weight because of the lowest risk in the model for SL algorithm. KNN and rpart put the next important weights in the model as candidate learners.

Table 1. Performance metrics for infertility data set

		Infertility Data Set				
Split ratio	Performance Metrics	C4.5	KNN	NB	SVM	RF
80% - 20%	Accuracy	0.9342	0.9079	0.8947	0.9342	0.9605*
	Sensitivity	0.9385	0.9385	0.8923	0.9692	0.9846
	Specificity	0.9091	0.7273	0.9091	0.7273	0.8182
	AUC	0.9244	0.9069	0.8727	0.9594*	0.9209
70%-30%	Accuracy	0.9123	0.9035	0.8596	0.9561*	0.9386
	Sensitivity	0.9082	0.9388	0.8469	0.9694	0.9694
	Specificity	0.9375	0.6875	0.9375	0.8750	0.7500
	AUC	0.9237	0.9435	0.8903	0.9534*	0.9298
60%- 40%	Accuracy	0.8954	0.8824	0.8889	0.8889	0.9346*
	Sensitivity	0.8855	0.9313	0.8779	0.9618	0.9695
	Specificity	0.9545	0.5909	0.9545	0.4545	0.7273
	AUC	0.9200	0.9221	0.9302	0.9323	0.9458*

*The best performance of the model

According to these findings, the best performance is yielded by SL algorithm of 97% AUC. In a previous study by Van der Laan et al different candidate learners like RF, Least squares method, Least Angle Regression and Delete / Change / Addition set to the model for the diabetes data set and smallest risk was obtained by Delete / Change / Addition (15).

The variable importance analyses show that sperm concentration is the most important variable. The Polymorphism genes are respectively in order of sy1291, gr/gr and b2/b3. As a matter of fact, in reference by Kumar (39), it is stated that the important factor for infertility is due to semen parameter values not within normal limits. Information on the importance of variables and as a result of infertility data analysis the results supports the literature. For example, Hicks et al. (40), a male infertility prediction study used sperm videos. As mentioned, sperm parameters play a big role in infertility. The reported algorithms

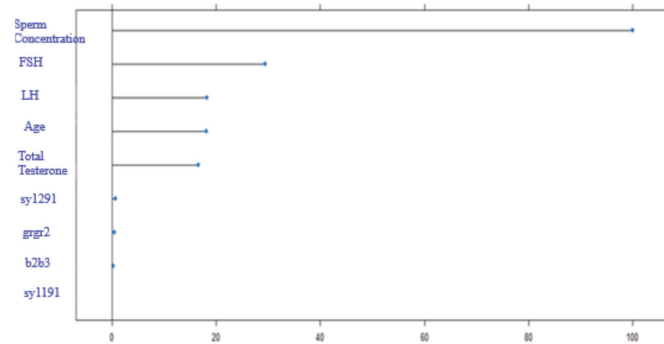


Figure 1. The importance of variables for infertility data set

Table 2. Risk and Coefficient results for Superlearner		
	Infertility Data Set	
	Risk	Coefficient
SLRF_All	0.0589*	0.5810*
SLKNN_All	0.0625	0.2327
SLbgIm_All	0.0788	0.0000
SLSVM_All	-	-
SLrpart_All	0.0613	0.1862
* Minimum risk and Coefficient		

Table 3. The area under curve results for Superlearner	
	Infertility Data Set
SL	0.9653*
Discrete SL	0.9597
SLRF_All	0.9597
SLKNN_All	0.9321
SLbgIm_All	0.9465
SLSVM_All	-
SLrpart_All	0.9304
*The best performance of the model	

used in this study are Simple Linear Regression, Random Forests, Gaussian Process, SMOreg (Sequential Minimal Optimization Regression), Elastic Net, and Random Trees. Here the error rate for RF make difference according to other mentioned algorithms. RF algorithm is an ensemble learning algorithm in which multiple models are combined to solve a particular problem (41).

One in six couples worldwide experiences infertility (42). It has been reported that the emotional status of couples who apply to a physician with infertility is deteriorated and their susceptibility to depression increases (42). It is also known that about a quarter of couples cannot continue their infertility treatments due to the burden of treatment (43). We think that the prediction of infertility, which has a complex nature and affects many areas such as the emotional conditions of couples, other health problems and the health system expenditures of the states, is of great importance for clinicians. Therefore, the development of models with high predictive ability will also improve clinical approaches for infertility treatment. These study findings whenever applied to any patient's record of infertility risk factors can be used to predict the risk of infertility in men. The predictive model developed can be integrated into existing health information systems which can be used by urologist to predict patients' risk of infertility in real time.

Conclusions

The results of the study show that different split ratios have an effect on the performance also it can change the algorithm that be used. The SL algorithm is a weighted model which consist of different candidate learners. According to the results the algorithm with the highest performance and minimum risk is linked to each other.

A researcher builds a model, by using different algorithms while different classifiers show different performances. However, there are too many algorithms in the literature. Choosing the best algorithm requires time and expertise. At this stage, SL is an important tool and recommended for achieving high performance and is a guide to the researcher. In this study, the model was obtained by using five candidate learners and the performances were compared. Superlearner gives the researcher time and expertise of solving data sets. However, different models can be established by evaluating different algorithms. In later studies, it is planned to conduct studies by trying combinations of different algorithms and using bigger sample data sizes. Simulation finding could be a good study to conduct.

Ethics

Ethics Committee Approval: Ethics approval for this study was given by the medical research ethical committee (2017/208, issued June 22, 2017).

Informed Consent: Written informed consent was taken from all patients.

Peer-review: Externally peer-reviewed.

Authorship Contributions

Surgical and Medical Practices: S.K., L.T., E.K., Concept: S.K., L.T., E.K., Design: S.K., L.T., E.K., Data Collection or Processing: S.K., L.T., E.K., Analysis or Interpretation: S.K., L.T., E.K., Literature Search: S.K., L.T., E.K., Writing: S.K., L.T., E.K.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declare that they have no relevant financial

References

- World Health Organization. Infertility: A Tabulation of Available Data on prevalence of Primary and Secondary Infertility Programme on Maternal and Child Health and Family Planning, Division of Family Health, Geneva; 1999.
- Al-Rekabe BKK, Jazaa Al-Wayelli DAO, Khadim AHA. Evaluation Of Serum FSH, LH and Testosterone Levels In Infertile Patients Affected With Different Male Infertility Factors After Iui Technique. *Thi-Qar Journal*, 2010; 4 (2).
- Mahat RK, Arora M, Bhale DV, Holkar S, Kumar S, Yadav T. Risk Factors and Causes of Male Infertility-A Review. *Biochem Anal Biochem*. 2016; 5:2.
- Deyhoul N, Mohamaddoost T, Hosseini M. Infertility Related Risk Factors: A Systematic Review, *International Journal of Women's Health and Reproduction Sciences*. 2017; 5 (1).
- Gunes S, Sevgili E, Asci R. Sperm DNA Hasarı Mekanizmaları ve Değerlendirme Yöntemleri. *Türkiye Klinikleri J Urology*. 2013; 4 (3).
- Idowu PA, Balogun JA, Babawale OT. A predictive model for the risk of Infertility in Women using Supervised Learning Algorithms- A Comparative Analysis. *Computer Reviews Journal*. 2018; (2).
- Han J, Kamber M, Pei J. *Data mining concepts and techniques*. 3rd edition, Waltham, MA 2012.
- Shwartz S, Ben-David S. *Understanding machine learning: From Theory to Algorithms*. NY, USA.
- Al-Aidaros KM, Bakar AA, Othman Z. Medical data classification with Naive Bayes approach. *Information Technology Journal*. 2012; 11: 9 1166-1174.
- Sajda P. Machine learning for detection and diagnosis of disease. *Annu. Rev. Biomed. Eng*. 2006; 8:8.1-8.29.
- Fumo D. Types of machine learning algorithms you should know [Internet]. 2017 [updated 2017 June 15; cited 2021 September 21]. Available from: <https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861>.
- Chakraborty C. *Advanced Classification Techniques for Healthcare Analysis*. Mesra, 2019; 3-12.
- Van der Laan MJ, Dudoit S. Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: finite sample oracle inequalities and Examples. *The Berkeley Electronic Press*. 2003; 130, 1-103.
- Van der Laan MJ, Dudoit S, Van der Waart AW. The cross-validated adaptive epsilon-net estimator. *Statistics and Decisions*. 2006; 24 3: 373-395.
- Van der Laan MJ, Polley EC, Hubbard AE. *Superlearner*. U.C. Berkeley Division of Biostatistics Working Paper Series. 2007; 222: 1-20.
- Abur U, Gunes S, Asci R, Altundag E, Akar OS, Ayas B, Alpaslan MK, Ogur G. Chromosomal and Y-chromosome microdeletion analysis in 1,300 infertile males and the fertility outcome of patients with AZFc microdeletions. *Andrologia*, 2019; 51(11).
- Beyaz CC, Gunes S, Onem K, Kulac T, Asci R. Partial Deletions of Y-Chromosome in Infertile Men with Non-obstructive Azoospermia and Oligoasthenoteratozoospermia in a Turkish Population. Partial Deletions of Y-Chromosome in Infertile Men with Non-obstructive Azoospermia and Oligoasthenoteratozoospermia in a Turkish Population. *In Vivo*, 2017;31(3):365-371.
- Marsland S. *Machine learning an algorithmic perspective*. Boca Raton, FL; 2015.
- Lantz B. *Machine Learning with R*. 2nd Edition. Birmingham, UK; 2015.
- Hssina B, Merbouha A, Ezzikouri H, Erritali M. A comparative study of decision Tree ID3 and C4.5. *International Journal of Advanced Computer Science and Applications, Special Issue on Advances in Vehicular Ad Hoc Networking and Applications*. 2014; 13-19.
- Therneau T, Atkinson B, Ripley B. *Package' rpart'*, 2019.
- Breiman L. Random forests. *Machine learning*. 2001; 45:1, 5-32.
- Aldrich C. Process variable importance analysis by use of random forests in a shapley regression framework. *Minerals*, 2020.
- Karakoyun M, Hacıbeyoğlu M. Biyomedikal veri kümeleri ile makine öğrenmesi sınıflandırma algoritmalarının istatistiksel olarak karşılaştırılması. *Dokuz Eylül Üniversitesi, Mühendislik Fakültesi Mühendislik Bilimleri Dergisi*. 2014; 16 (48): 30-41.
- Domingos P, Pazzani M. On the optimality of the simple bayesian classifier under zero-one loss. *Machine Learning*. 1997; 29: 103-130.
- Molano-Gonzalez N, Cepeda-Cuervo E. *Package' bglm'*, 2014.
- Song Y, Huang J, Zhou D, Zha H, Giles CL. Informative K-nearest neighbor pattern classification. *Heidelberg*; 2007.
- Hassanat AB, Abbadi MA, Altarawneh GA, Alhasanat AA. Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach. *International Journal of Computer Science and Information Security*. 2014; 12, 8: 33-39.
- Pradhan A. Support vector machines: a survey. *International Journal of Emerging Technology and Advanced Engineering*. 2012; 2 8: 82- 85.
- Karatzoglou A, Meyer D, Hornik K. Support vector machines in R. *Journal of Statistical Software*. 2006; 15 9: 1-25.
- Alpaydın E. *Introduction to machine learning*. 2nd Edition. USA; 2010.
- Cosgun E, Karaagaoglu E. Veri madenciliği yöntemleriyle mikrodizilim gen ifade analizi. *Hacettepe Tıp Dergisi*. 2011; 42: 180-189.
- Sunasra M. Performance metrics for classification problems in machine learning- Part I [Internet]. 2017 [updated 2017 November 11, cited 2021 September 21]. Available from: <https://medium.com/greyatom/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b>.
- Kaynar O, Yıldız M, Görmez Y, Albayrak A. Makine öğrenmesi yöntemleri ile duygu analizi. *International Artificial Intelligence and Data Processing Symposium, Malatya*, 2016; 234-241.
- Noi PT, Kappas M. Comparison of random forest, K-nearest neighbor and support vector machine classifiers for land cover classification using sentinel-2 imagery; 2017.
- Kaur H, Kumari V. Predictive modelling and analytics for diabetes using a machine learning approach. *Applied Computing and Informatics*. DOI: 10.1016/j.aci.2018.12.004.
- Ali J, Khan R, Ahmad N, Maqsood I. Random Forests and Decision Trees. *IJCSI International Journal of Computer Science*. 2012; 9:5(3).

38. Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. *Medical Informatics and Decision Making*. 2011; 11:5(1).
39. Kumar N, Singh AK. Trends of male factor infertility, an important cause of infertility: A review of literature. *Journal Human Reprod Science*. 2015; 8 4: 191–196.
40. Hick's SA, Andersen JM, Witczak O, Thambawita V. Machine Learning-Based Analysis of Sperm Videos and Participant Data for Male Fertility Prediction, 2019.
41. Dimitriadis S I, Liparas D. How random is the random forest ? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI) database. *Neural Regen Res* 2018;13(6):962-970.
42. Liao S, Pan W, Dai W D, Jin L, Huang G, Wang R, Hu C, Pan W, Tu H. Development of a Dynamic Diagnosis Grading System for Infertility Using Machine Learning. *Jama Network Open*. 2020;2,3(11).
43. Brandes M, Van der Steen J O M, Bokdam S B, Hamilton C J C M, De Bruin J P, Nelen W L D M, Kremer J A M .When and why do subfertile couples discontinue their fertility care? A longitudinal cohort study in a secondary care subfertility population. *Hum Reprod*. 2009;24(12),3127-35.