Sinop University
Positive Science International
e-ISSN: 3108-3862

Positive Science International, 2(1), 43-52
Research Article, Year: 2026
DOI: 10.71340/psi.1709547

Evaluation of Kernel and Parameter Effects on the Performance of Classifiers for Differentiated Thyroid Cancer Recurrence

Mertcan TUTUM[1*], Yavuz ÜNAL[2]

[1]* *Graduate School of Natural and Applied Sciences, Amasya University, 05100, Amasya, Türkiye*

[2]*Department of Computer Engineering, Faculty of Engineering and Architecture, Sinop University, Sinop, Türkiye*

Article Info

Research article
Received: *30.05.2025*
Accepted: *10.06.2026*
Published: *30.06.2026*
***Corresponding Author*:** *Mertcan TUTUM*
mertutum@gmail.com
0009-0000-1292-8517
Keywords

Cancer recurrence prediction
Machine Learning
Thyroid cancer

Abstract

Thyroid cancer causes tens of thousands of deaths every year in the world. Treatment of this disease is possible with today's medical conditions. However, as in other types of cancer, recurrence is possible. The Differentiated Thyroid Cancer Recurrence dataset, which includes data from 383 patients and was obtained from the UCI platform, was used in the study. The aim of this study is to compare the classification performances of machine learning algorithms according to various parameters and Kernel functions in order to predict recurrence in thyroid cancer patients. For this purpose, three machine learning algorithms were used to predict differentiated thyroid cancer recurrence. In the K-Nearest Neighbors (K-NN) algorithm, the most appropriate k parameter was determined by trying different neighbor number (k) values. The Support Vector Machines (SVM) algorithm examined various kernel functions and hyperparameter settings. In addition, model performance was compared by optimizing different tree numbers and depth parameters with the Random Forest algorithm. These methods aimed to determine the model with the highest accuracy and generalization capacity.

Farklılaşmış Tiroid Kanseri Nüksü İçin Sınıflandırıcıların Performansı Üzerindeki Çekirdek ve Parametre Etkilerinin Değerlendirilmesi

Makale Bilgisi

Araştırma makalesi
Başvuru: *30.05.2025*
Kabul: *10.06.2026*
Yayın: *30.06.2026*
***Sorumlu Yazar*:** *Mertcan TUTUM*
mertutum@gmail.com
0009-0000-1292-8517
Anahtar Kelimeler

Kanser tekrarlama tahmini
Makine Öğrenimi
Tiroid kanseri

Özet

Tiroid kanseri, dünyada her yıl on binlerce kişinin ölümüne neden olmaktadır. Günümüzün tıbbi imkânlarıyla bu hastalığın tedavisi mümkündür. Ancak, diğer kanser türlerinde olduğu gibi, nüksetme olasılığı da mevcuttur. Çalışmada, UCI platformundan elde edilen ve 383 hastanın verilerini içeren "Farklılaşmış Tiroid Kanseri Nüksü" veri seti kullanılmıştır. Bu çalışmanın amacı, tiroid kanseri hastalarında nüksü tahmin etmek için çeşitli parametrelere ve çekirdek fonksiyonlarına göre makine öğrenimi algoritmalarının sınıflandırma performanslarını karşılaştırmaktır. Bu amaçla, farklılaşmış tiroid kanseri nüksünü tahmin etmek üzere üç makine öğrenimi algoritması kullanılmıştır. K-En Yakın Komşu (K-NN) algoritmasında, farklı komşu sayısı (k) değerleri denenerek en uygun k parametresi belirlenmiştir. Destek Vektör Makineleri (SVM) algoritmasında çeşitli çekirdek fonksiyonları ve hiperparametre ayarları incelenmiştir. Ayrıca, Rastgele Orman algoritmasıyla farklı ağaç sayısı ve derinlik parametreleri optimize edilerek model performansı karşılaştırılmıştır. Bu yöntemler, en yüksek doğruluk ve genelleme kapasitesine sahip modeli belirlemeyi amaçlamıştır.

To cite this article:
Tutum, M., Ünal Y. (2026). Evaluation of Kernel and Parameter Effects on the Performance of Classifiers for Differentiated Thyroid Cancer Recurrence, Positive Science International, 2(1), 43-52, https://doi.org/10.71340/psi.1709547

1. Introduction

In recent years, the incidence of thyroid cancer has been rapidly increasing. The advancements in imaging technologies have enabled the early diagnosis of this disease. The survival rate is quite high compared to other types of cancer. Thyroid cancer is a cancerous tumor that starts in the thyroid gland in the front of the neck [1]. It happens when cells in this gland, which are vital for making thyroid hormones, grow out of control. Papillary thyroid carcinoma is the most frequent type of thyroid cancer. It makes up around 80% of all cases and usually has a decent prognosis [2]. Follicular thyroid cancer is not as frequent as papillary thyroid cancer, but it can be treated in the same way. The C cells in the thyroid give rise to medullary thyroid carcinoma, which may be linked to genetics [3, 4].

In the last few years, a growing number of people have been getting thyroid cancer. By modelling the stages of thyroid cancer spreading to the body, the risk of death for the patient can be lowered. When the main risk factors for the disease are known, early diagnosis and treatment planning can help steer the disease in the right direction.

Differentiated Thyroid Carcinoma (DTC), notably the papillary and follicular subtypes, is a kind of cancer that usually has a good prognosis and a five-year survival rate of over 95%. But even if this is a benign course, the rate of recurrence (reappearance) is still quite high [5, 6].

Machine learning is changing a lot of things in health care, such how doctors diagnose, treat, and follow up with patients. It works well for finding diseases early and figuring out risk factors, especially when it finds meaningful patterns in vast data sets. Image processing techniques have made it possible to automatically find cancer and other disorders. Also, machine learning algorithms play a big role in making personalized treatment plans and medication development processes. Analyzing electronic health records makes health services better and more efficient. So, machine learning is now a very important tool in the health field for lowering costs and improving patient outcomes [7].

The aim of this study is to predict recurrence in patients with DTC by comparing the classification performance of machine learning algorithms, specifically Support Vector Machines (SVM) and K-Nearest Neighbors (K-NN). The study further investigates the impact of different kernel functions (e.g., RBF, Polynomial) and parameter settings on model accuracy to identify the most effective predictive configuration.

Machine learning has become a common tool for predicting thyroid cancer outcomes, with studies focusing on both survival analysis and recurrence prediction. The following studies illustrate the range of approaches applied to clinical and pathological data in this field.

Mourad et al. [8] proposed a machine learning approach to predict survival in thyroid cancer patients, combining 34 clinical variables from the SEER database with feature selection methods such as Fisher discriminant ratio, Kruskal-Wallis analysis, and Relief-F. They trained multilayer perceptron-based supervised neural networks with patient groups ranging from 6,756 to 20,344 and achieved 94.5%

accuracy in distinguishing patients who lived longer than 10 years from diagnosis from those who died within five years.

Park and Lee [1] analyzed clinicopathological data from 1040 patients diagnosed with papillary thyroid carcinoma (PTC) between 2003 and 2009 to predict disease recurrence and developed a machine learning-based prediction model. They compared the performance of five different machine learning models using parameters such as age, gender, tumor size, ETE, ENE, pT, pN, number of metastatic lymph nodes, and lymph node ratio (LNR). The Decision Tree model achieved the highest accuracy at 95%, while the lightGBM and stacking models showed 93% accuracy.

Kim et al. [2] proposed an inductive logic programming-based approach to predict recurrence in papillary thyroid cancer patients who underwent thyroidectomy. They used data from 785 patients who underwent bilateral total thyroidectomy and received radioactive iodine therapy; 624 (79.5%) were allocated for rule generation and 161 (20.5%) for validation, and the analysis was performed using the DELMIA Process Rules Discovery tool. They identified three rules that predicted recurrence and determined that postoperative thyroglobulin levels were the strongest predictor. The rules generated in the validation set correctly predicted 10 out of 14 recurrence cases, achieving a 71.4% success rate.

Lee and Park [9] compiled the recent use of machine learning methods in the early diagnosis of thyroid diseases according to different data types. They determined that random forest and gradient boosting methods were the most suitable approaches for numerical data, while random forest was the most suitable approach for genomic and ultrasound data. In the studies examined, they reported that accuracy values ranged from 64.3% to 99.5%, sensitivity from 66.8% to 90.1%, specificity from 61.8% to 85.5%, and the area under the ROC curve from 64.0% to 96.9%. They identified clinical stage, histological type, age, tumor diameter, extrathyroidal spread, various RNA and microRNA expression levels, and ultrasound characteristics as the most important variables for early diagnosis.

Zhang et al. [10] systematically compiled machine learning-based applications used in the pathogenesis, diagnosis, and prognosis of thyroid cancer and created a comprehensive taxonomy. By reviewing a total of 758 studies, they classified artificial intelligence techniques in the field and revealed the organizational structure of the existing literature. They identified key challenges encountered in thyroid cancer research and suggested future research opportunities for less studied topics.

Duan et al. [11] applied machine learning algorithms to predict hypothyroidism developing in Graves' disease patients after radioactive iodine therapy (RAI) at an early stage. They randomly divided 471 patients who underwent RAI between 2016 and 2019 into two groups: 310 for training and 161 for validation. They extracted 138 clinical and laboratory features from electronic medical records. They developed a multivariate model including patient age, thyroid mass, 24-hour radioactive iodine uptake, aspartate aminotransferase, thyroid antibodies, and neutrophil count. They obtained an AUROC value of 0.72 in the training set and 0.74 in the validation set, along with an F1 score of 0.74 and an MCC score of 0.63 in both sets and created a nomogram for clinical use.

2. Methods

2.1. Dataset

In our study, the Differentiated Thyroid Cancer Recurrence dataset obtained from the University of California Irvine (UCI) Machine Learning Repository website was used [12]. The features and their types included in the dataset are provided in Table 1.

Table 1. Differentiated thyroid cancer recurrence dataset and features

Feature name	Role	Type
Age	Feature	Integer
Gender	Feature	Categorical
Smoking	Feature	Categorical
Hx Smoking	Feature	Categorical
Hx Radiothreapy	Feature	Categorical
Thyroid Function	Feature	Categorical
Physical Examination	Feature	Categorical
Adenopathy	Feature	Categorical
Pathology	Feature	Categorical
Focality	Feature	Categorical
Risk	Feature	Categorical
T	Feature	Categorical
N	Feature	Categorical
M	Feature	Categorical
Stage	Feature	Categorical
Response	Feature	Categorical
Recurrend	Target	Categorical

The dataset consists of 16 features and one target. These include information such as the patient's age, gender, whether they smoke or not, whether they undergo radiotherapy or not, and whether they engage in physical activity or not. Out of these 16 features, 15 are categorical, and one is an integer.

2.2. Parameter Optimization

In machine learning, parameter optimization refers to modifying a model's learnable parameters or hyperparameters to maximize its performance throughout the training phase. The weights and bias values that the model learns during training are called parameters; the coefficients in linear regression are one example. The learning rate, the number of network layers, and the batch size are examples of hyperparameters, which are the values that are set before training and that dictate the model's structure.

To reduce the model's loss, the Gradient Descent method enables the parameters to be modified in accordance with the gradients. While hyperparameter optimization can employ trial-and-error techniques like Grid Search and Random Search, Bayesian Optimization uses a probabilistic approach to determine the optimal hyperparameters [13]. By automating this process, autoML tools Optuna and Hyperopt are two examples that make hyperparameter optimization easier. Generally speaking, the goal is to improve performance, reduce training and test error, or prevent overfitting in order to increase the model's capacity for generalization [13].

2.3. Support Vector Machine

One supervised learning method that is well-known for its excellent performance, particularly in classification tasks, is Support Vector Machines (SVM). Finding a "hyperplane" (decision boundary) that best divides data points from different classes is its primary objective; in the process, it concentrates on maximizing the "margin," or distance, between the data points closest to this hyperplane, known as "support vectors." By improving the model's capacity for generalization, maximizing the margin enables it to produce predictions on fresh, untested data that are more accurate. Because SVMs can use a technique known as the "kernel trick" to transfer complex data structures that cannot be separated linearly to a higher dimensional space and make linear distinctions there, they also provide flexibility [14] [15] [16].

2.4. Random Forest

One supervised learning algorithm that can yield good results, particularly in complex data sets, is Random Forest. It is essentially predicated on the idea that a "forest" is created when numerous decision trees converge. Using distinct subsets and features chosen at random from the primary data set, each tree is trained separately throughout the training process. Every tree in the forest makes a prediction when one is needed, and the class that receives the most votes (in classification problems) or the average of the tree predictions (in regression problems) is the final outcome. By decreasing the propensity of a single decision tree to overfit the data (overfitting), this collective decision-making mechanism improves the model's capacity for generalization and typically yields more reliable and accurate predictions [17] [18].

2.5. K-Nearest Neighbours (K-NN) algorithm

The k-Nearest Neighbours (k-NN) algorithm is one of the easiest and most common ways to classify things in supervised learning. The main idea behind it is to find the distances between the sample to be classed and the data as the number of nearest neighbours k in the feature space. The Euclidean distance metric is the most common one, although for some sorts of problems, other distance metrics like Manhattan and Minkowski may be better. k-NN is quite sensitive to how its parameters are adjusted. The right k value has a direct effect on how well the model works. Small k values can make the model too specific, while very large k values can make it harder for the model to generalize. Also, the k-NN technique is easy to train, and all the math is done when the data is classified, therefore it can be expensive to run on large data sets. The k-NN technique is good for medical classification issues like

predicting the recurrence of thyroid cancer because it makes proper classifications by looking at the local structure of the instances in the data set [19].

2.6. WEKA

The University of Waikato in New Zealand created the well-known open-source machine learning program Weka (Waikato Environment for Knowledge Analysis). Weka provides an extensive toolkit for data mining and machine learning algorithms and is written in the Java programming language. Numerous machine learning tasks, including feature selection, regression, clustering, classification, and data preprocessing, are supported. Its intuitive graphical user interface makes data analysis simple, even for users without programming experience. Weka, which is widely used in both academia and business, can work with various data formats, including CSV and Excel, and supports the ARFF (Attribute-Relation File Format) format. Numerous algorithms, including k-means clustering, decision trees, neural networks, support vector machines, and others, are included, and the results can be visualized [20] [21].

2.7. Performance Metrics

The study's performance has been evaluated using a variety of criteria. These criteria include F1-score, recall, accuracy, and precision. The ideas of true positive, true negative, false positive, and false negative must be understood before the formulas for these terms can be explained. True positives are examples that are correctly classified as positive, true negatives are examples that are correctly classified as negative, false positives are examples that are incorrectly classified as positive, and false negatives are examples that are incorrectly classified as negative [22] [24]. Consequently, the following formulas are given for the performance metrics:

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (1)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (2)$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (3)$$

$$f1 - score = 2x\frac{\mathrm{PrecisionxRecall}}{\mathrm{Precision+Recall}} \quad (4)$$

3. Results and Discussion

The analyses were performed on the WEKA program. CSV format is one of the formats that can be used in the WEKA program. Model parameters and kernel types were tested randomly and in various combinations for performance evaluation. First, the Support Vector Machine (SVM) classifier was used and the kernel functions were tested as RBF, Polynomial Kernel, Normalized Polynomial Kernel and PUK, respectively. The results are given in Table 2.

Table 2. Comparison of Support Vector Machine (SVM) classifier performance using different kernel functions

Classifier	Kernel	Accuracy	Precision	Recall	F1-score
SVM	RBF	94.57%	0.94	0.94	0.94
	PolyKernel	95.56%	0.95	0.95	0.95
	NormalizedPolyKernel	94.77%	0.94	0.94	0.94
	Puk	93.22%	0.93	0.93	0.93

When the accuracy results obtained by using different kernel functions in the Support Vector Machine (SVM) algorithm are examined, Polynomial Kernel has shown superior performance than other kernel types with the highest success rate (95.56%). While RBF and Normalized Polynomial Kernel have similar accuracy values (94.57% and 94.44%), PUK kernel has lower performance compared to the others (93.22%). These findings show that the kernel selection has a significant effect on the model success. The classification results obtained with K-NN and different k values are given in Table 3

Table 3. Performance evaluation of the K-Nearest Neighbors (K-NN) classifier with varying values of K.

Classifier	K	Accuracy	Precision	Recall	F1-score
K-NN	1	91.38%	0.91	0.91	0.91
	2	92.95%	0.93	0.93	0.92
	3	93.21%	0.93	0.93	0.93
	4	93.47%	0.93	0.93	0.93

The effect of different k values on model performance in the K-Nearest Neighbors (K-NN) algorithm was investigated. As a result of the experiments, 91.38% accuracy rates were obtained for k=1, 92.95% for k=2, 93.21% for k=3 and 93.47% for k=4. These findings show that there is a certain improvement in classification accuracy with increasing k values; however, it should be noted that the increase may stop or start to decrease after a certain point. Therefore, the optimal k value should be selected carefully. The classification results obtained with the Random Forest classifier using different numbers of trees are presented in Table 4.

Table 4. Impact of the number of trees on the performance of the Random Forest classifier

Classifier	Num. of Trees	Accuracy	Precision	Recall	F1-score
Random Forest	10	96.34%	0.96	0.96	0.96
	50	96.86%	0.96	0.96	0.96
	100	96.34%	0.96	0.96	0.96
	200	96.34%	0.96	0.96	0.96

The effect of different tree numbers on model success was evaluated in the Random Forest algorithm. Accuracy rates of 96.34% were obtained with 10 trees, 96.86% with 50 trees, and 96.34% with 100 and 200 trees. The results show that a small performance increase was achieved by increasing the number of trees to 50, but adding more trees did not lead to a significant improvement in accuracy. This situation reveals that when the optimal number of trees is exceeded, the model complexity increases and the performance becomes stagnant.

Random Forest gave the highest accuracy at 96.86% with 50 trees, followed by SVM with Polynomial kernel at 95.56% and K-NN at 93.47%. The dataset structure explains this ordering. Sixteen of seventeen features are categorical, and tree-based methods split categorical attributes directly while SVM and K-NN depend on distance metrics that distort under encoded categories.

Borzooei et al. [25] reported similar accuracy ranges on the same 383 patient cohort, which suggests classical machine learning has likely reached its ceiling on this dataset. Adding trees beyond 50 produced no gain, and kernel differences in SVM stayed within one percentage point. Smaller models are preferable for clinical use because they train and run faster.

Several limitations apply. The class distribution was not reported, and no resampling or class weighting was used, so accuracy may overstate performance on the recurrence class. ROC AUC, sensitivity, specificity, and confusion matrices would have given a fuller picture. The K-NN experiments stopped at k=4 and used Euclidean distance, which is suboptimal for categorical data. External validation on independent cohorts is also missing, so the results describe one dataset rather than a clinical tool.

4. Conclusion

This study looked at how well three machine learning algorithms Support Vector Machine (SVM), K-Nearest Neighbours (K-NN), and Random Forest could predict the return of differentiated thyroid carcinoma in individuals. We looked at multiple kernel functions for SVM and varied parameter settings for K-NN and Random Forest to find the best model accuracy. The results showed that Polynomial Kernel had the best accuracy of all the SVM kernels. K-NN, on the other hand, got better as the K values went up to 4. Random Forest worked well with 50 trees, but there were no big improvements after that.

These results show how important it is to use the right kernel and tune the hyperparameters to make medical classification tasks more accurate. The study shows that machine learning models could be useful for finding cancer recurrences early, which could help doctors give each patient the best care possible. Different research might look at different algorithms and use bigger, more varied datasets to make the model even more useful in the real world and for more people.

Next steps include comparing gradient boosting methods (XGBoost, LightGBM, CatBoost), since CatBoost handles categorical variables natively. SHAP analysis will quantify which clinical features drive predictions, which is more actionable for physicians than accuracy alone. External

validation on multi center Turkish cohorts will test transferability. Deep tabular models such as TabNet and FT Transformer are worth comparing against the ensemble baselines established here.

On the methodological side, class imbalance will be handled with SMOTE and class weighted loss, hyperparameter search will move from manual WEKA testing to Optuna under nested cross validation, and the evaluation protocol will expand to ROC AUC, precision recall curves, and Matthews correlation coefficient.

Declarations

Funding/Financial Disclosure The authors have no received any financial support for the research, authorship, or publication of this study.

Ethics Committee Approval and Permissions The work does not require ethics committee approval and any private permission.

Conflict of Interests The authors stated that there is no conflict of interest in this article.

References

[1] Y. M. Park and B.-J. Lee, "Machine learning-based prediction model using clinico-pathologic factors for papillary thyroid carcinoma recurrence," *Sci. Rep.*, vol. 11, no. 1, p. 4948, Mar. 2021, doi: 10.1038/s41598-021-84504-2.

[2] S. Y. Kim et al., "New approach of prediction of recurrence in thyroid cancer patients using machine learning," *Medicine*, vol. 100, no. 42, p. e27493, Oct. 2021, doi: 10.1097/MD.0000000000027493.

[3] İ. Kara, M. G. Yildiz, and İ. Orhan, "Tiroid Nodülü," *Kahramanmaraş Sütçü İmam Üniversitesi Tıp Fakültesi Dergisi*, vol. 15, no. 3, pp. 94–99, Oct. 2020, doi: 10.17517/ksutfd.685884.

[4] H. Wang et al., "Development and validation of prediction models for papillary thyroid cancer structural recurrence using machine learning approaches," *BMC Cancer*, vol. 24, no. 1, p. 427, Apr. 2024, doi: 10.1186/s12885-024-12146-4.

[5] B. Bektas Gunes, R. Samlı, M. B. Dogan, and D. Yildirim, "SEGMENTATION OF THYROID NODULES ON ULTRASOUND IMAGES," *J. Naval Sci. Eng.*, vol. 20, no. 2, pp. 191–211, Nov. 2024, doi: 10.56850/jnse.1507140.

[6] L.-R. Li, B. Du, H.-Q. Liu, and C. Chen, "Artificial Intelligence for Personalized Medicine in Thyroid Cancer: Current Status and Future Perspectives," *Front. Oncol.*, vol. 10, p. 604051, Feb. 2021, doi: 10.3389/fonc.2020.604051.

[7] Z. Aytaç, İ. Iseri, and B. Dandil, "Derin Öğrenme Kullanarak Tiroid Kanseri Teşhisi," *Eur. J. Sci. Technol.*, Dec. 2021, doi: 10.31590/ejosat.1011166.

[8] M. Mourad et al., "Machine Learning and Feature Selection Applied to SEER Data to Reliably Assess Thyroid Cancer Prognosis," *Sci. Rep.*, vol. 10, no. 1, p. 5176, Mar. 2020, doi: 10.1038/s41598-020-62023-w.

[9] K.-S. Lee and H. Park, "Machine learning on thyroid disease: a review," *Front. Biosci. (Landmark Ed)*, vol. 27, no. 3, p. 101, Mar. 2022, doi: 10.31083/j.fbl2703101.

[10] X. Zhang, V. C. Lee, and F. Liu, "From Data to Insights: A Comprehensive Survey on Advanced Applications in Thyroid Cancer Research," 2024, *arXiv*, doi: 10.48550/ARXIV.2401.03722.

[11] L. Duan et al., "Machine learning identifies baseline clinical features that predict early hypothyroidism in patients with Graves' disease after radioiodine therapy," *Endocr. Connect.*, vol. 11, no. 5, p. e220119, May. 2022, doi: 10.1530/EC-22-0119.

[12] S. Borzooei, G. Briganti, M. Golparian, J. R. Lechien, and A. Tarokhian, “Machine learning for risk stratification of thyroid cancer patients: a 15-year cohort study,” *Eur. Arch. Otorhinolaryngol.*, vol. 281, no. 4, pp. 2095–2104, Apr. 2024, doi: 10.1007/s00405-023-08299-w.

[13] C. Parlak, “Konuşma Duygu Tanıma Uygulamalarında Hiper Parametre Optimizasyonu ile Derin Öğrenme Metotlarının Geliştirilmesi,” *Karadeniz Fen Bilimleri Dergisi*, vol. 14, no. 4, pp. 1955–1975, Dec. 2024, doi: 10.31466/kfbd.1508578.

[14] S. Metlek and K. Kayaalp, “Derin Öğrenme ve Destek Vektör Makineleri İle Görüntüden Cinsiyet Tahmini,” *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, vol. 8, no. 3, pp. 2208–2228, Jul. 2020, doi: 10.29130/dubited.707316.

[15] G. Harman, “Destek Vektör Makineleri ve Naive Bayes Sınıflandırma Algoritmalarını Kullanarak Diabetes Mellitus Tahmini,” *Eur. J. Sci. Technol.*, Dec. 2021, doi: 10.31590/ejosat.1041186.

[16] E. Efeoğlu, “Destek Vektör Makinelerinin Wi-Fi Tabanlı İç Mekan Lokalizasyon Tespitinde Kullanımı ve Çekirdek Fonksiyon Seçiminin Sınıflandırma Performansına Etkisi,” *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 5, no. 3, pp. 1370–1382, Dec. 2022, doi: 10.47495/okufbed.1057825.

[17] O. K. M. Salman and B. Aksoy, “RASGELE ORMAN VE İKİLİ PARÇACIK SÜRÜ ZEKÂSI YÖNTEMİYLE KALP YETMEZLİĞİ HASTALIĞINDAKİ ÖLÜM RİSKİNİN TAHMİNLENMESİ,” *Int. J. 3D Print. Technol. Dig. Ind.*, vol. 6, no. 3, pp. 416–428, Dec. 2022, doi: 10.46519/ij3dptdi.982670.

[18] H. Y. Dalkiliç, S. N. YeşiLyurt, and P. Samui, “Daily Flow Modeling With Random Forest and K-Nearest Neighbor Methods,” *Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 14, no. 3, pp. 914–925, Dec. 2021, doi: 10.18185/erzifbed.949126.

[19] Ö. Türk, “MR GÖRÜNTÜLERİNDEN ALZHEİMER TESPİTİNDE BOYUT AZALTMA VE DERİN ÖĞRENME YAKLAŞIMLARININ KARŞILAŞTIRILMASI,” *DÜMF MD*, Sep. 2022, doi: 10.24012/dumf.1141233.

[20] M. Hacibeyoglu, M. ÇeliK, and Ö. Erdaş ÇiÇek, “K En Yakın Komşu Algoritması ile Binalarda Enerji Verimliliği Tahmini,” *neufmbd*, no. 2, Dec. 2023, doi: 10.47112/neufmbd.2023.10.

[21] A. V. Ağlarcı and F. Karakurt, “K En Yakın Komşu Makine Öğrenme Algoritmasına Dayalı Diabetes Mellitus Tahmini,” *Turk J. Diab. Obes.*, vol. 8, no. 3, pp. 265–276, Dec. 2024, doi: 10.25048/tudod.1549498.

[22] Ö. F. Akmeşe, “Diagnosing Diabetes with Machine Learning Techiques,” *Hittite J. Sci. Eng.*, vol. 9, no. 1, pp. 9–18, Mar. 2022, doi: 10.17350/HJSE19030000250.

[23] M. DiRiK, “Machine learning-based lung cancer diagnosis,” *Turk. J. Eng.*, vol. 7, no. 4, pp. 322–330, Oct. 2023, doi: 10.31127/tuje.1180931.

[24] Y. Unal and M. Bolat, “Detecting Wheat Leaf Diseases: A Deep Feature-Based Approach with Machine Learning Classification,” *SJAFS*, no. 3, Dec. 2024, doi: 10.15316/SJAFS.2024.041.

[25] D. Bhende et al., “Machine Learning-Based Classification of Thyroid Disease: A Comprehensive Study on Early Detection and Risk Factor Analysis,” in *Proc. IEEE Int. Students’ Conf. Electr. Electron. Comput. Sci. (SCEECS)*, Bhopal, India, Feb. 2024, pp. 1–6, doi: 10.1109/SCEECS61402.2024.10481980.