Adaptive Arctan kernel: a generalized kernel for support vector machine

SELÇUK BAŞ[1], SERHAT KILIÇARSLAN[2,*], ABDULLAH ELEN[2] and
CEMIL KÖZKURT[3]

[1]Department of Accounting and Tax, Bandirma Vocational School, Bandirma Onyedi Eylul University,
Bandirma, Balıkesir, Turkey
[2]Department of Software Engineering, Faculty of Engineering and Natural Sciences, Bandirma Onyedi Eylul
University, Bandirma, Balıkesir, Turkey
[3]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Bandirma Onyedi Eylul
University, Bandirma, Balıkesir, Turkey
e-mail: sbas@bandirma.edu.tr; skilicarslan@bandirma.edu.tr; aelen@bandirma.edu.tr;
ckozkurt@bandirma.edu.tr

Abstract. Support Vector Machines (SVMs) can learn from high-dimensional and small amounts of training data, thanks to effective optimization methods and a diverse set of kernel functions (KFs). The adaptability of SVM to numerous real-world problems has increased interest in the SVM method, and studies conducted with this method carry significant weight in various fields. The fixed parameter "AtanK" for KFs must be specified before the SVM model training process. Therefore, determining the appropriate kernel parameter values can be timE−consuming and may lead to slow convergence of the SVM model. On the other hand, the method provides faster and more robust convergence due to the adaptive parameter in the SVM model. In this study, a new Adaptive Arctan (AA) KF, tailored to the characteristics of different datasets, is proposed as an enhancement to the AtanK KF for the SVM algorithm. The proposed AA KF is compared with experimental results on 30 public KEEL and UCI datasets, alongside AtanK, adaptive Gaussian, Radial Basis Function, linear, and polynomial KFs. The results demonstrate that the proposed AA KF outperforms the other KFs, and it exhibits superior learning ability.

1. Introduction

The support vector machine (SVM) is a machine learning model developed by Vapnik. It is used in clustering and regression problems, especially in classification [1–3]. SVM is a non-parametric, supervised machine learning method based on statistical learning theory [4]. Researchers widely prefer SVM for classification algorithms due to its successful performance, robust mathematical infrastructure, and relatively fast speed [5, 6]. In SVM, mapping is performed to transfer data samples to a dimensional space, and the classification process begins with the hyperplane that separates the two classes. Samples close to the boundaries are defined as support vectors. The separation of SVM training samples into different subgroups using support vectors is determined by the choice of kernel functions (KFs). Although there are various types of KFs, the basic ones for classification are Linear, Polynomial, Gaussian, Radial Based Function (RBF), and Sigmoid kernels [7–10]. The choice of KFs has a significant impact on the classification performance in SVM. SVM can mitigate the shortcomings of existing models, such as overfitting, reliance on a single consideration factor, and the inability to generate predictions for lengthy periods. As a supervised learning model, SVM is an effective statistical machine learning tool for classifying datasets of both linear and nonlinear separable types. One of the key features that makes SVM advantageous is its use of the Lagrange binary optimization problem. This allows for the optimal number of training iterations, resulting in a significant time advantage compared to other machine learning methods. SVM excels on datasets that require substantial memory and datasets with many features but few samples.

A Kernel Function (KF) transforms the input data into Hilbert space within the Support Vector Machine (SVM), creating intervals for the hyperplane to classify the data. While there are several well-known KFs, new proposals continue to emerge every day. The most renowned and
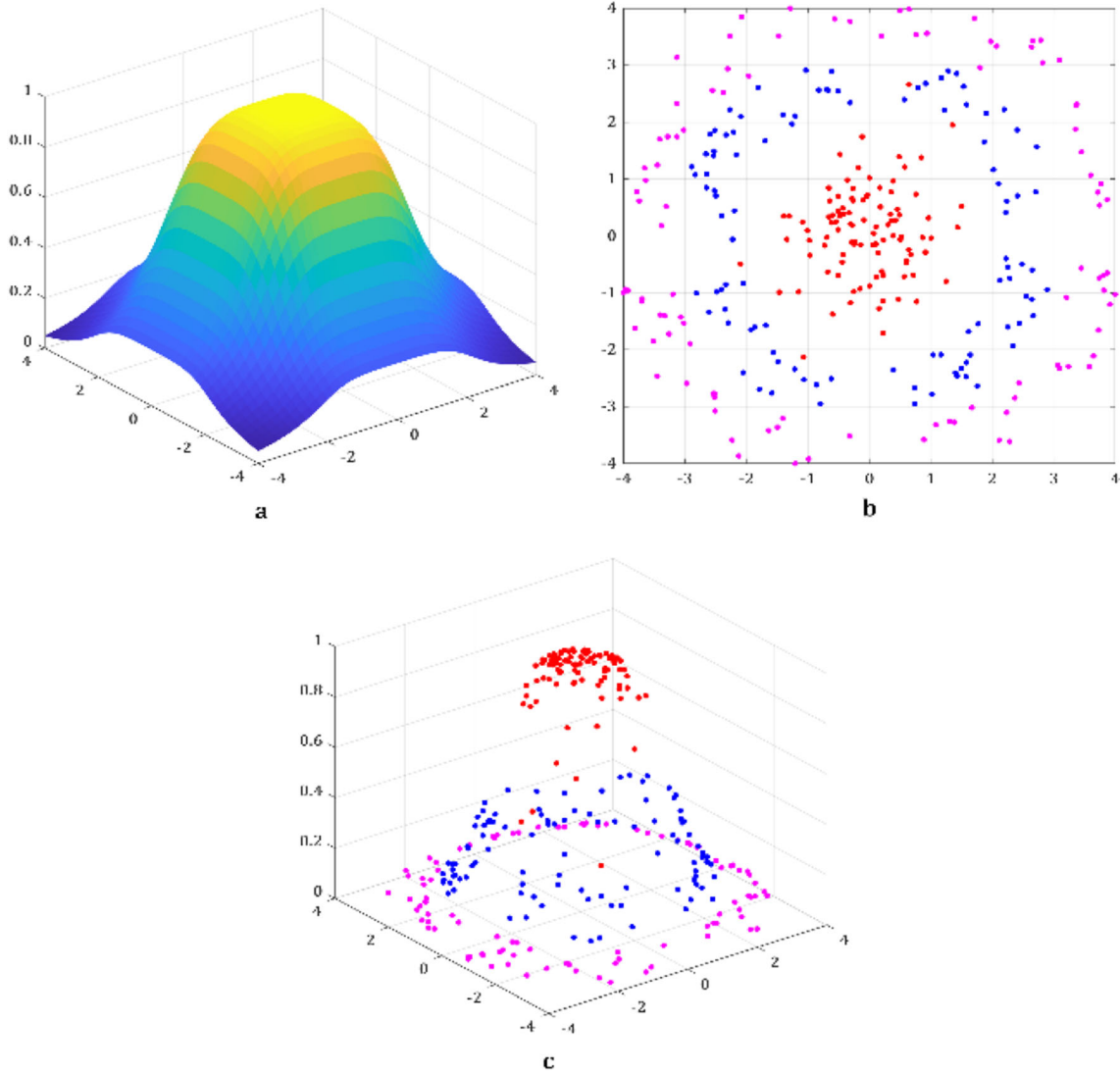
---

*For correspondence

Figure 1. Behavior of the SVM with AtanK on nonlinear separable synthetic 3 class dataset; (**a**) 3D AtanK surface with $\sigma = 5$, (**b**) Nonlinearly separable synthetic 3 class dataset, (**c**) Dataset mapped to Hilbert space by AtanK.

successful KFs can be categorized as linear, *Gaussian, RBF, polynomial, and AtanK. Some of these functions include a parameter with an arbitrary value. This arbitrary value can be manually adjusted according to the dataset, and classification in the SVM is performed using the most suitable version of the KF for the dataset. FinE−tuning the optimal arbitrary value based on the dataset requires time and results in a high processing load. Although the Gaussian KF is one of the most well-known KFs with parameters, recent studies have also introduced parameters for AtanK. In Eqs. (1) and (2), the Gaussian KF and AtanK KF are presented, respectively.

$$K(s,q) = \exp\left(-\frac{\|s-q\|_2^2}{2\sigma^2}\right) \qquad (1)$$

$$K(s,q) = \frac{2}{\pi}\arctan\left(\frac{\sigma^2}{\|s-q\|_2^2}\right) \qquad (2)$$

where s and q are input vectors. These vectors are expressed in KFs with the Euclidean norm. σ is a real number, which is an arbitrary value. These KFs vary smoothly in the input feature space in all directions around the support vector, resulting in hyper-spherical contours of the KF.

KFs are expressed as a set of mathematical functions used to process input data [11]. KFs play a crucial role in solving nonlinear problems with the aid of a linear classifier. In the context of the SVM architecture, KFs are employed to achieve accurate data separation by transforming the nonlinear structure of the input dataset from a higher-dimensional space to a linear structure [12]. Furthermore, KFs are recommended to simplify complex
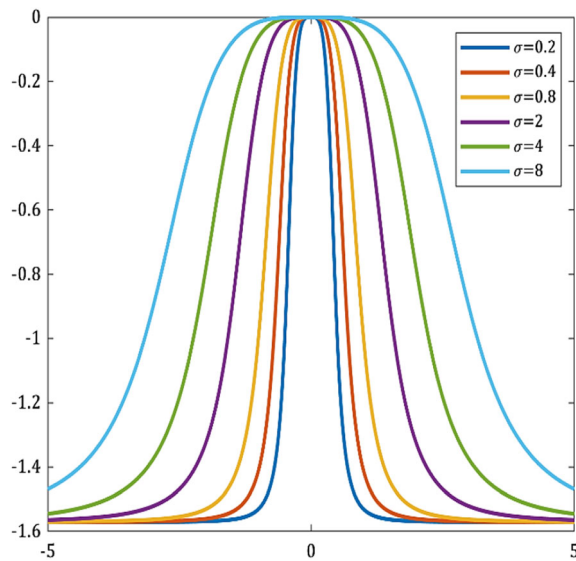
Figure 2. Plot of the Arctan Kernel function at various σ values.

computations, enabling the creation of hyperplanes in higher dimensions without increasing complexity. In summary, KFs are utilized to address nonlinear problems. Additionally, given the unique characteristics of each dataset, identifying the most appropriate KF for optimal data classification has become a significant research focus in recent years [13].

Selecting appropriate KFs can be a challenging process. To choose the most suitable KF for the characteristics of the dataset in use, the performance of SVM algorithms with different KFs is compared. Additionally, parameter tuning for each KF is a timE−consuming process. Therefore, proposing an adaptive KF that can adjust to the data, rather than attempting to find the optimal function by experimenting with KFs having various parameter values, reduces a substantial workload. The adaptive approach enables SVM to diminish parameter dependency in KFs during classification and regression processes [14].

The Linear, Polynomial, Sigmoid, Gaussian, and RBF kernels are widely used in the literature [15]. However, many studies in the literature have observed that new functions are proposed to outperform existing KFs. Elen and colleagues introduced a new KF called Adaptive Gaussian (AG), inspired by the Gaussian kernel used in the SVM [13]. In their study, Anna *et al* proposed a new KF by combining three types of KFs: polynomial, RBF, and Fourier [7]. As a result of the experimental evaluation, they reported that the combined KF offers better performance than other KFs. Ding *et al* introduced a new KF, the random radial basis function (RRBF). The proposed RRBF outperforms traditional RBF and other KFs [16]. Negri *et al*

presented two new KFs for classifying remotely sensed images in their study [17]. They reported that the proposed new KFs perform well on both remote sensing and real-world datasets compared to other functions. In the study by Tian and Li, the ArcTan KF (AtanK) is proposed as an alternative to the Gaussian KF to increase the practicality of the SVM [18]. They reported that the proposed AtanK has better efficiency and robustness than the polynomial kernel, Gaussian kernel, and exponential radial basis functions. Khan *et al* introduced a new KF, a combination of polynomial and RBF, to improve the performance of the SVM used to classify protein structures. They reported that they achieved better performance than other machine learning algorithms with the proposed hybrid KF SVM [19]. In the study by Li and He, an adaptive KF is proposed to enhance the classification performance of the SVM [20]. In the study by Ba *et al*, a soft sensing method based on least square the SVM regression with multi-core learning was proposed. In the proposed model, they reported that they used the RBF, polynomial, and sigmoid KFs to construct the multiple kernel structure [21]. The adaptive arctan KF proposed in this study outperformed the KFs suggested in the literature in the experimental evaluations conducted on 30 public UCI datasets.

In this study, a new Adaptive Arctan (AA) KF is proposed to align with the characteristics of various datasets in the SVM classification algorithm. It often consumes a significant amount of time to determine the most suitable parameter value when applying KFs to the SVM problems. However, the proposed AA KF SVM can ascertain the most appropriate parameter during the initial run, enabling quicker classification with optimal performance during the learning process. Instead of a fixed σ parameter value used in the AtanK function, the AA KF, capable of adjusting itself to each dataset, is suggested. Consequently, the SVM model delivers faster and more accurate classification. The proposed AA KF permits us to construct hyperplanes in higher dimensions without increasing complexity. Furthermore, the overfitting issue commonly observed in the SVM can be effectively addressed with the proposed AA KF. The proposed AA KF is compared with AtanK, AG, RBF, LIN, and POL KFs through experiments conducted on 30 public UCI datasets. The results indicate that the proposed AA KF outperforms other KFs and exhibits superior learning capabilities.

The contributions of this study are listed as follows:

- The new adaptive arctan KF is proposed.
- The proposed adaptive arctan KF has a better learning ability.
- Since it does not need problem-specific parameters, it can be applied directly and practically to all classification problems.
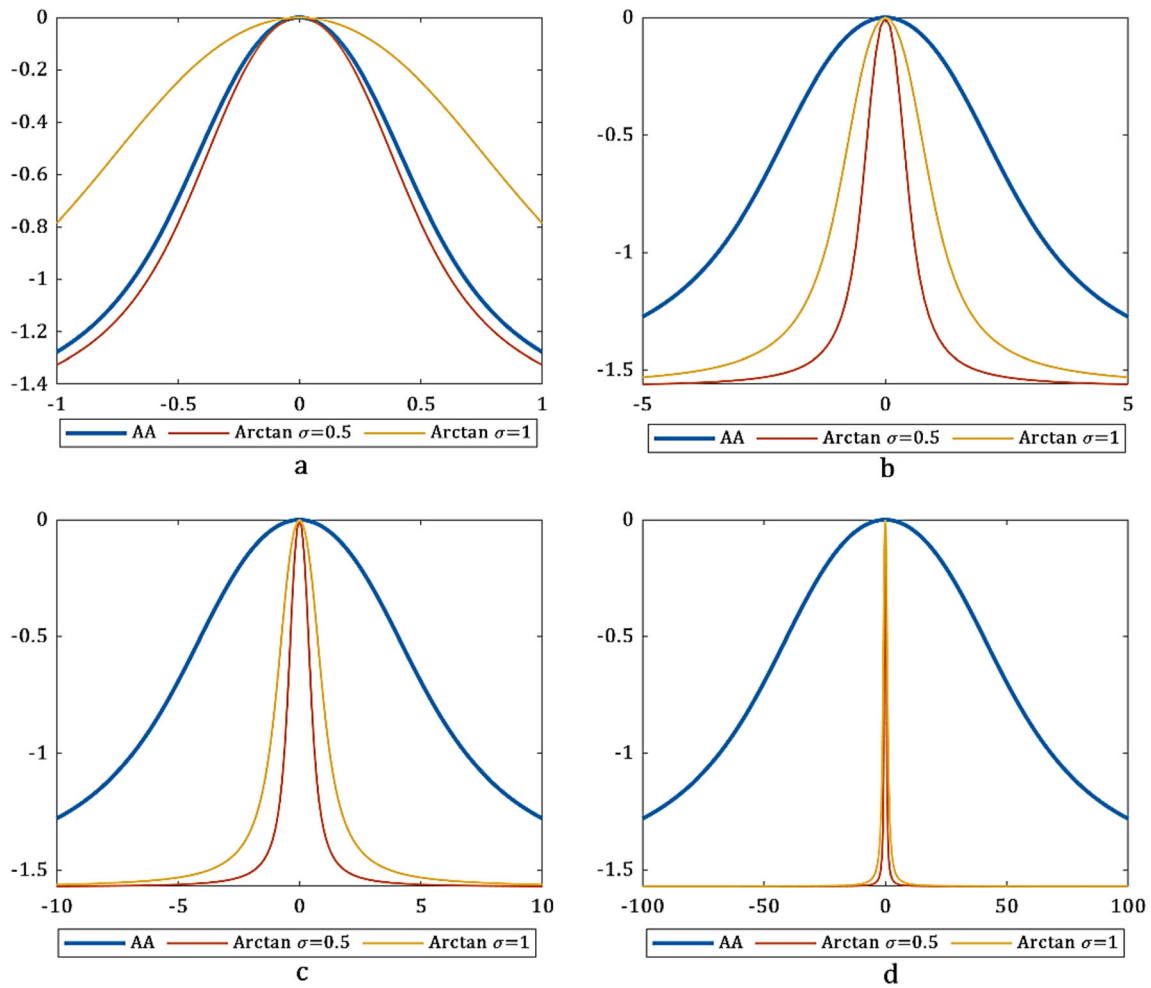
Figure 3. The proposed AA vs. arctangent function on various input data (x) intervals. (**a**) $-1 \leq x \leq 1$, (**b**) $-5 \leq x \leq 5$, (**c**) $-10 \leq x \leq 10$, (**d**) $-100 \leq x \leq 100$.
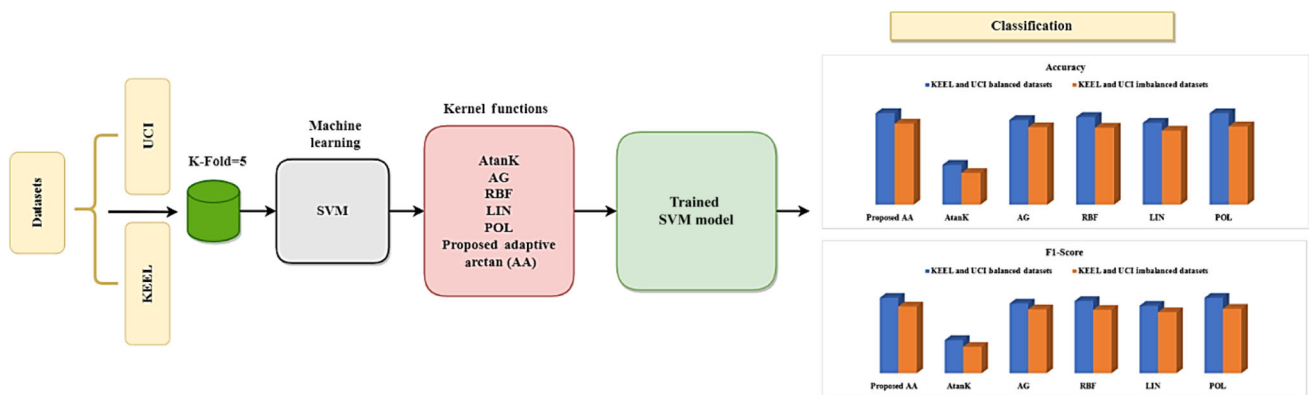
Figure 4. Process steps of experimental studies for KFs.

Motivation to the SVM and KF are presented in Section 2 of the study. Proposed adaptive arctan KF is presented in Section 3, experimental results are presented in Section 4, and conclusions are discussed in section 5.
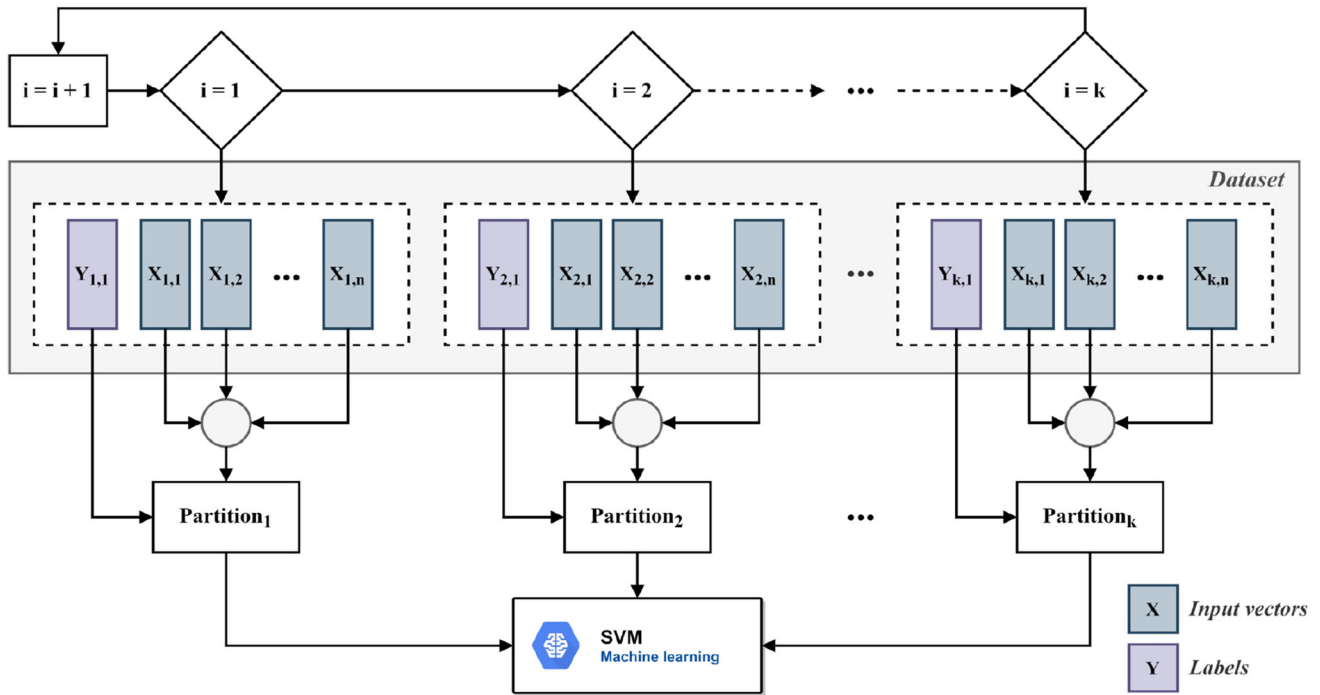
Figure 5. k-fold cross validation of the dataset [13].

Table 1. Benchmark balanced datasets for evaluation of the SVM kernels.

| #	Repository	Dataset	Features	Instances	Classes
1	KEEL	Appendicitis	7	106	2
2		Bupa	6	345	2
3		Phoneme	5	5404	2
4		Ring	20	7400	2
5		Spam base	57	4597	2
6		Wine	13	178	3
7		Zoo	16	101	7
8	UCI	Banknote authentication	4	1372	2
9		User knowledge modeling	5	403	4
10		Vertebral column 2C	6	310	2
11		Vertebral column 3C	6	310	3
12		Breast cancer	9	277	2
13		Monks-1	6	124	2
14		Monks-2	6	169	2
15		Monks-3	6	122	2
16		Plrx	12	182	2
17		Transfusion	4	748	2
18		WDBC	30	569	2
19		WPBC	33	198	2

2. Motivation

2.1 *Support vector machine*

The AtanK function is used in SVM, where the input vectors of the kernel are in their raw form. This function is unimodal and symmetric about the third axis. For small values of σ, the function exhibits a Gaussian-like bell-shaped curve, while at large values, it assumes a plateau-like shape. Although the variation in this shape scale is effective in separating the classes in the Hilbert space, it is

Table 2. Benchmark imbalanced datasets for evaluation of the SVM kernels.

#	Repository	Dataset	Features	Instances	Classes
1	KEEL	Bands	19	365	2
2		Cleveland	13	303	5
3		German	20	1000	2
4		Haberman	3	306	2
5		House votes	16	435	2
6		Ionosphere	33	351	2
7		New thyroid	5	215	3
8		Yeast	8	1483	10
9	UCI	South German credit	20	1000	2
10		Glass	9	214	6
11		Satimage	36	4435	6

Table 3. Synthetic datasets for evaluation of the SVM kernels.

#	Dataset	Features	Instances	Classes
1	3-Spiral	2	312	3
2	Atom	3	800	2
3	Cassini	2	1000	3
4	Chain-link	3	1000	2
5	Compound	2	399	6
6	R-15	2	600	15

determined through testing the parameter σ, which provides the optimum plateau width. This experimental process involves an iterative method that is timE−consuming and power-intensive. Therefore, there is a need for an efficient method to determine the optimum value of σ based on the data. In this study, drawing inspiration from the AtanK function and the Adaptive Gaussian KF, a statistical model that adapts to the data within the function instead of σ is proposed. A snapshot of the AtanK function can be represented as a plateau in the context of threE−class synthetic data distributions and a threE−dimensional surface, as shown in figure 1.

2.2 *Kernel trick*

Let $\{s_k\}$, be a sequence in the (S, d) metric space. If there is M number such as

$$d(s_k, s_l) < \varepsilon \tag{3}$$

is called Cauchy sequence for $\forall \varepsilon > 0$ and $\forall k, l \geq M.$. Every Cauchy sequence in the full metric space is convergent. A distance in the norm space is defined by

$$d(s_k, s_l) = \|s_k - s_l\| \tag{4}$$

Also, the complete inner product space in metric space is called Hilbert space H.

$$K(s, q) = \langle \gamma(s), \gamma(q) \rangle_H \tag{5}$$

is defined as $K : S \times S \to R$ KF by means of inner product. Here, H is feature space of K and $\gamma : S \times S \to H$ is feature map. Further, with the inner product's property, $f, g \in H$ begin by

$$f(s) = \sum_{i=1}^{k} \eta_i K(s_i, s),$$
$$g(q) = \sum_{j=1}^{l} \lambda_j K(q_j, q) \tag{6}$$

Then

$$\langle f, g \rangle = \sum_{i=1}^{k} \sum_{j=1}^{l} \eta_i \lambda_j K(s_i, q_j)$$
$$= \sum_{i=1}^{k} \eta_i g(s_i) = \sum_{j=1}^{l} \lambda_j f(q_j) \tag{7}$$

Mercer Theorem. For the kernel $K(s, q)$ with $\gamma(s)$, integral in Eq. (8) should always provide non-negativity, the purpose of being a valid SVM KF [1].

$$\int \int K(s, q) \gamma(s) \gamma(q) ds dq \geq 0 \tag{8}$$

If κ is the kernel matrix of $K(s, q)$ KF with $\kappa_{i,j} = K(s, q)$,

$$\rho^T \kappa \rho = \sum_{i,j} \rho_i \rho_j K(s, q) = \langle \gamma(s), \gamma(q) \rangle$$
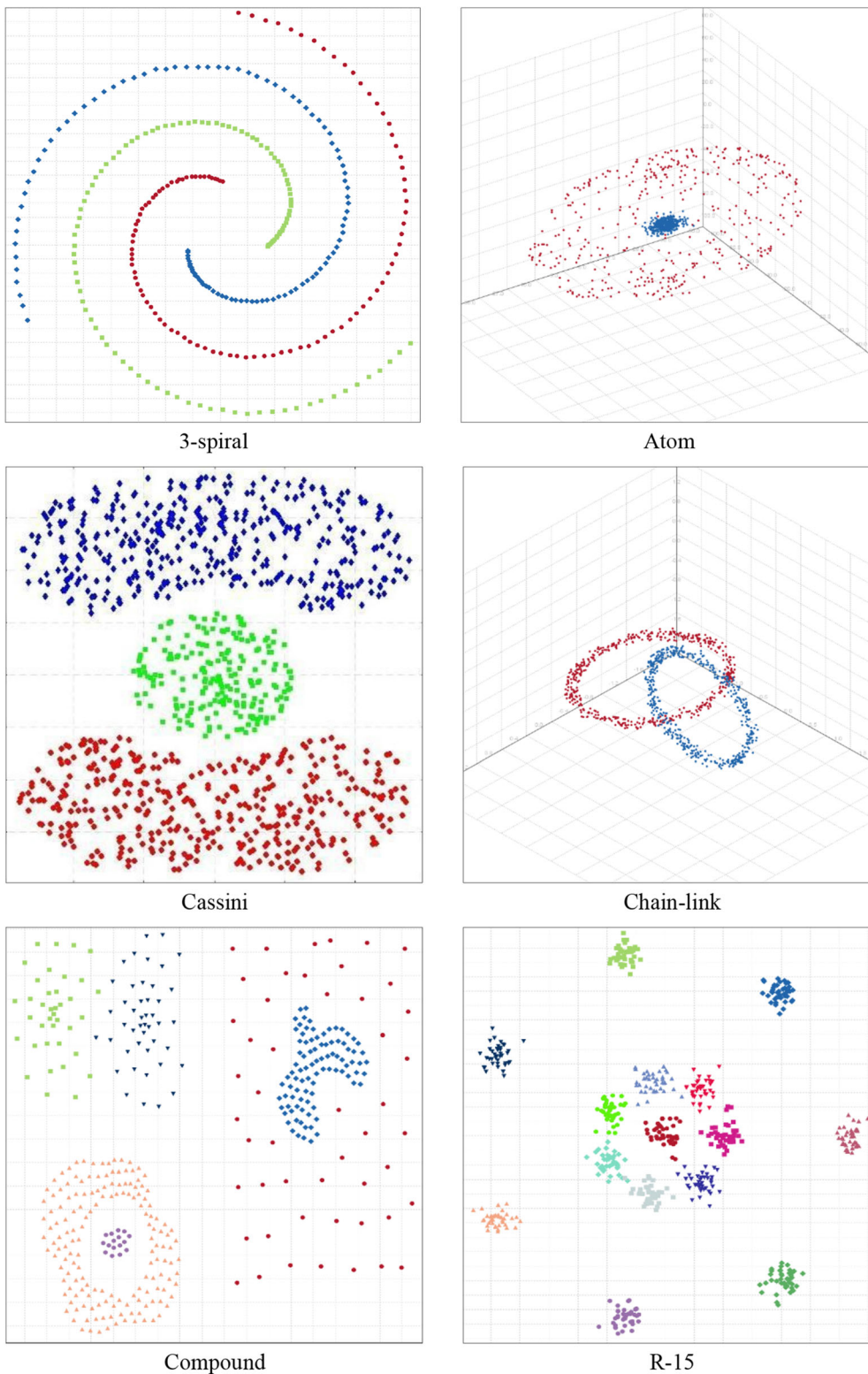$$= \left\| \sum_i \rho_i \gamma(s) \right\|^2 \geq 0 \tag{9}$$

Figure 6. Visualizations of the synthetic datasets.

Table 4. Accuracy rates of kernels for the KEEL and UCI balanced datasets.

Datasets	Proposed AA	AtanK	AG	RBF	LIN	POL
Appendicitis	0.860	0.802	0.879	**0.888**	0.860	0.879
Bupa	0.661	0.580	0.580	0.609	0.580	**0.693**
Phoneme	**0.800**	0.707	0.645	0.792	0.771	0.794
Ring	0.961	0.505	0.627	**0.978**	0.764	0.973
Spam base	**0.931**	0.606	0.741	0.922	0.901	0.926
Wine	0.989	0.599	0.985	0.989	**0.992**	0.977
Zoo	**0.985**	0.824	0.981	0.962	0.980	0.978
Banknote authentication	**1.000**	0.555	0.961	0.985	0.980	0.999
User knowledge modeling	**0.978**	0.660	0.973	0.966	0.965	0.974
Vertebral column 2C	**0.840**	0.677	0.810	0.813	0.765	0.839
Vertebral column 3C	**0.881**	0.656	0.852	0.856	0.830	0.877
Breast cancer	0.751	0.708	**0.755**	**0.755**	0.729	0.722
Monks-1	0.742	0.476	0.734	0.718	0.661	**0.775**
Monks-2	0.674	0.621	**0.716**	0.650	0.621	0.657
Monks-3	0.853	0.508	0.845	**0.893**	0.763	0.844
Plrx	0.709	**0.714**	0.654	**0.714**	**0.714**	0.687
Transfusion	**0.775**	0.762	0.521	0.763	0.762	0.767
WDBC	**0.978**	0.627	0.912	0.975	0.974	0.977
WPBC	0.763	0.753	0.703	**0.778**	0.758	0.773
Mean values:	**0.8490**	**0.6495**	**0.7825**	**0.8424**	**0.8089**	**0.8479**

Table 5. Accuracy rates of kernels for the KEEL and UCI imbalanced datasets.

Datasets	Proposed AA	AtanK	AG	RBF	LIN	POL
Bands	**0.712**	0.630	0.710	0.710	0.674	0.649
Cleveland	0.818	0.816	0.818	**0.826**	0.825	0.821
German	**0.758**	0.700	0.747	0.754	0.753	0.693
Haberman	0.732	0.735	0.663	0.732	**0.739**	0.726
House votes	**0.966**	0.535	0.944	0.927	**0.966**	0.927
Ionosphere	0.92	0.641	0.897	**0.949**	0.877	0.858
New thyroid	**0.972**	0.798	0.960	0.941	0.919	0.969
Yeast	**0.918**	0.862	0.888	0.915	0.911	**0.918**
South German credit	**0.761**	0.700	0.756	0.760	0.751	0.696
Glass	0.880	0.785	0.885	0.877	0.860	**0.889**
Satimage	0.960	0.747	0.927	0.965	0.954	**0.966**
Mean values:	**0.8332**	**0.6905**	**0.8114**	**0.8236**	**0.8086**	**0.8133**

3. Adaptive Arctan kernel function

KFs in the SVM are similarity scales between two samples. The selection of a kernel function is an important process for the SVM and has a significant impact on the results. The kernel function expresses the inner product in the feature space. The Gaussian, as given in Eq. (1), is one of the most preferred and efficient classifying KFs. In this study, we drew inspiration from the Gaussian KF and the AtanK function provided in Eq. (2). The arctan part in the AtanK function creates the curve shape, while the $2/\pi$ factor sets the upper limit on the vertical axis to 1. Although $2/\pi$ is convenient for visualization, it does not affect the SVM classification process. According to AtanK, the arctan KF given in Eq. (10) is proposed as a simpler and non-adaptive function.

$$K_A(s, q) = \arctan\left(-\frac{\|s - q\|^2}{\sigma^2}\right) \tag{10}$$

Negativity in parentheses prevents the inversion of the curve. This negativity actually aids in visualization. Although the highest value of the proposed arctan KF is 0, it does not introduce any negativity in the SVM classification process. Figure 2 displays the graphs of the proposed

Table 6. Confidence Interval Profile of KFs for balanced and imbalanced datasets (for accuracy rates).

Kernel function	95% Confidence intervals	
	Balanced dataset	Imbalanced dataset
Proposed AA	0.8490 ± 0.002865	0.8332 ± 0.002983
AtanK	0.6495 ± 0.003818	0.6905 ± 0.003699
AG	0.7825 ± 0.003301	0.8114 ± 0.003130
RBF	0.8424 ± 0.002916	0.8236 ± 0.003050
LIN	0.8089 ± 0.003146	0.8086 ± 0.003148
POL	0.8479 ± 0.002874	0.8133 ± 0.003118

arctan KF at various σ values. As the value of σ increases, it becomes evident that the arctan KF transitions from a bell-shaped form to a plateau.

Instead of the arbitrary σ value in Eq. (10), the adaptive arctan (AA) KF given in Eq. (11) is proposed to adapt to the data.

$$K_{AA}(s, q) = \arctan\left(-\frac{\|s-q\|^2 - \delta\left(\|s-q\|^2\right) + \omega}{\delta\left(\|s-q\|^2\right)} \right) \quad (11)$$

While the σ value is determined manually in the AtanK KF, the proposed kernel function (AA KF) adapts to different problems by calculating the inverse tangent of the input vector with a coefficient obtained depending on the

standard deviation in the input vector. In order to handle division by a value in undefined or infinite results, which can occur if the standard deviation is equal to zero, division by a very small α value is provided in Eq. (12).

$$K_{AA}(s, q) = \arctan\left(-\frac{\|s-q\|^2 - \delta\left(\|s-q\|^2\right) + \omega}{\delta\left(\|s-q\|^2\right) + \alpha} \right) \quad (12)$$

where $\delta(\cdot)$ is a function of the standard deviation, α is an extremely small number to eliminate dividing by zero and ω is as shown in Eq. (13).

$$\omega = \begin{cases} \left| \|s-q\|^2 - \delta\left(\|s-q\|^2\right) \right|, & \|s-q\|^2 < \delta\left(\|s-q\|^2\right) \\ 0, & otherwise \end{cases} \quad (13)$$

The proposed AA KF's denominator in parentheses responds to the input vectors. Simultaneously adaptation is supplied by computing the square norm of the input vector differences and their standard deviations in the numerator. As a result, the proposed AA kernel is regarded as adaptable. When the numerator portion of the piecewise function ω is utilized, if the numerator part's minimum value is less than zero, the numerator part's absolute value is computed, and an offset operation is then performed on the whole dataset.
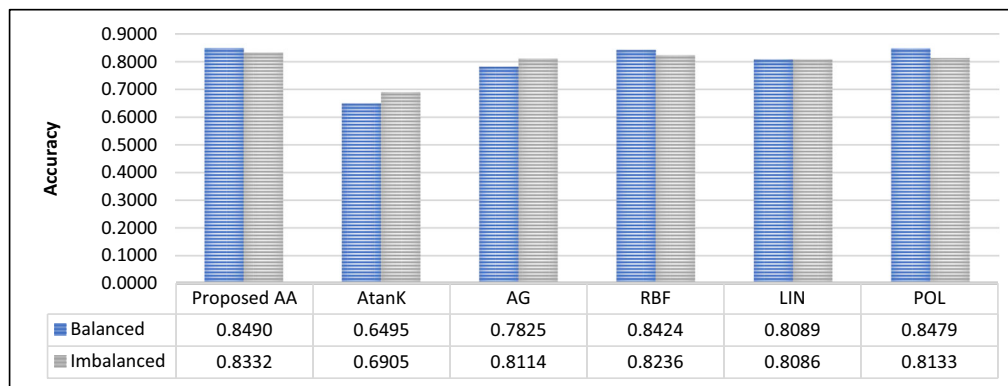
A dataset with a two-dimensional property space can be used for visualization to simplify the comprehension of the

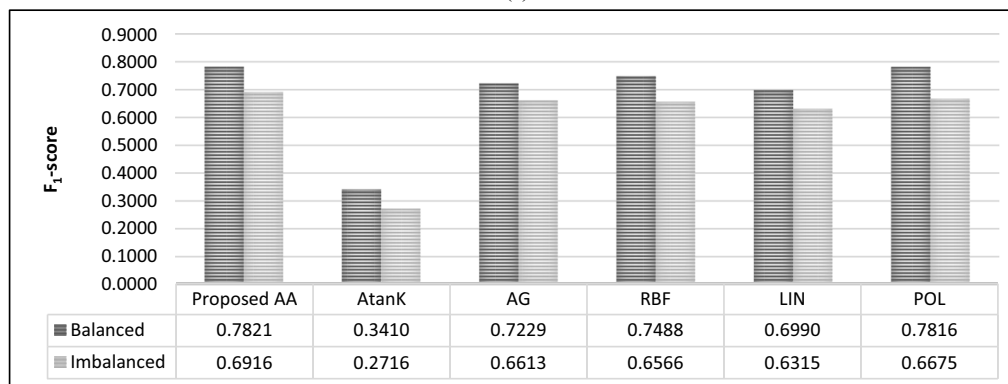**Table 7.** F_1-scores of KF for the KEEL and UCI balanced datasets.

Dataset	Proposed AA	AtanK	AG	RBF	LIN	POL
Appendicitis	0.728	0.445	0.775	0.793	0.733	**0.797**
Bupa	0.577	0.367	0.548	0.459	0.367	**0.654**
Phoneme	**0.751**	0.414	0.621	0.749	0.716	0.748
Ring	0.961	0.335	0.573	**0.978**	0.762	0.973
Spam base	**0.925**	0.377	0.689	0.917	0.895	**0.922**
Wine	0.982	0.19	0.977	0.983	**0.988**	0.966
Zoo	0.829	0.085	**0.877**	0.709	0.819	0.825
Banknote authentication	**1.000**	0.357	0.96	0.985	0.98	0.999
User knowledge modeling	**0.957**	0.121	0.946	0.914	0.91	0.954
Vertebral column 2C	**0.821**	0.404	0.788	0.769	0.677	0.81
Vertebral column 3C	**0.8001**	0.217	0.728	0.737	0.652	0.78
Breast cancer	0.609	0.414	0.623	0.622	**0.634**	0.615
Monks-1	0.736	0.349	0.727	0.714	0.657	**0.774**
Monks-2	0.638	0.383	**0.706**	0.591	0.383	0.634
Monks-3	0.852	0.337	0.843	**0.892**	0.762	0.843
Plrx	0.415	0.417	0.395	0.417	0.417	**0.437**
Transfusion	**0.707**	0.432	0.479	0.449	0.432	0.461
WDBC	0.982	0.386	0.909	0.973	0.971	0.975
WPBC	0.59	0.449	0.572	0.578	0.526	**0.685**
Mean values:	**0.7821**	**0.3410**	**0.7229**	**0.7488**	**0.6990**	**0.7816**

Table 8. F_1-scores of KF for the KEEL and UCI imbalanced datasets.

Dataset	Proposed AA	AtanK	AG	RBF	LIN	POL
Bands	**0.682**	0.387	0.672	0.628	0.568	0.604
Cleveland	0.235	0.14	0.239	0.262	0.276	**0.308**
German	**0.682**	0.412	0.669	0.655	0.679	0.628
Haberman	**0.529**	0.424	0.505	0.433	0.436	0.451
House votes	**0.968**	0.348	0.944	0.925	0.966	0.926
Ionosphere	0.912	0.391	0.886	**0.944**	0.855	0.842
New thyroid	**0.942**	0.274	0.903	0.862	0.792	0.931
Yeast	**0.545**	0.048	0.468	0.542	0.495	0.539
South German credit	**0.680**	0.412	0.675	0.664	0.675	0.636
Glass	0.581	0.087	0.551	0.439	0.382	**0.603**
Satimage	0.852	0.065	0.763	0.869	0.823	**0.874**
Mean values:	**0.6916**	**0.2716**	**0.6613**	**0.6566**	**0.6315**	**0.6675**

	Proposed AA	AtanK	AG	RBF	LIN	POL
■ Balanced	0.8490	0.6495	0.7825	0.8424	0.8089	0.8479
▨ Imbalanced	0.8332	0.6905	0.8114	0.8236	0.8086	0.8133

(a)

	Proposed AA	AtanK	AG	RBF	LIN	POL
▥ Balanced	0.7821	0.3410	0.7229	0.7488	0.6990	0.7816
▥ Imbalanced	0.6916	0.2716	0.6613	0.6566	0.6315	0.6675

(b)

Figure 7. Mean classification results of the SVM kernels on KEEL and UCI repositories. (**a**) Accuracy, (**b**) F1-score.

adaptive property. Distributed vectors on a 3D bell or plateau-shaped surface are formed due to all the vectors entering the kernel in alignment with the central vector or in the central static state. These Hilbert space vectors can now be quickly categorized using the SVM. In figure 3, bell and plateau curves of the AtanK function, and the bell and plateau curves of the AA KF corresponding to inputectors in various value ranges, are contrasted. The AA KF, which creates a smooth bell curve distribution regardless of the data range, offers an adaptive feature. Thus, the problem of

Table 9. Confidence interval profile of KFs for balanced and imbalanced datasets (for F$_1$-scores).

Kernel function	95% Confidence intervals	
	Balanced dataset	Imbalanced dataset
Proposed AA	0.7821 ± 0.003303	0.6916 ± 0.003695
AtanK	0.3410 ± 0.003793	0.2716 ± 0.003559
AG	0.7229 ± 0.003581	0.6613 ± 0.003789
RBF	0.7488 ± 0.003470	0.6566 ± 0.003800
LIN	0.6990 ± 0.003670	0.6315 ± 0.003860
POL	0.7816 ± 0.003306	0.6675 ± 0.003770

finding the optimum value of the parameter σ in Arctan KF and AtanK function, which is an arbitrary real constant value, is eliminated with the AA KF.

4. Experimental results

In experimental studies, two repository datasets, namely KEEL, and UCI, were used to evaluate the performance of the proposed kernel function. The Gaussian KFs σ parameter value was taken default value 1. Additionally, the polynomial kernel degree value was taken default value 2.

The steps used in the experimental studies for the AA KF are shown in figure 4. In order to test the performances of machine learning algorithms, the k-fold crossover method is used as shown in figure 5.

In the study, the k-fold value was determined as five for the experimental evaluation. The cross-validation method is a widely used method to evaluate the predictive performance of models. In the k-fold cross-validation method, the data set is randomly divided into k subgroups of approximately equal size. Thus, k runs are performed for each data set in the training and testing process of machine learning models.

Then, the average accuracy values were calculated and the overall success rate of the model was obtained. The two repository datasets, namely KEEL, and UCI used for the experimental evaluation are presented in tables 1 and 2 as balanced, imbalanced. To measure how the proposed kernel function generalizes to different data distributions, six different synthetic datasets are given in table 3. Visualizations of the synthetic datasets are given in figure 6.

The performances of classification algorithms were compared according to the criteria of accuracy (ACC), and F$_1$-score, respectively Eqs. (14) and (15) [22–25].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (14)$$

Table 10. Test accuracies of the synthetic datasets.

	Accuracy	RBF	LIN	POL	AG	AtanK	Proposed AA
3-Spiral	*Mean*	0.5629	0.5438	0.6026	0.6401	0.5608	**0.8152**
	Std. Dev.	0.0271	0.0164	0.0205	0.0370	0.0211	0.0214
	Min	0.5054	0.5161	0.5699	0.5873	0.5269	0.7849
	Max	0.6085	0.5699	0.6237	0.7097	0.5979	0.8413
Atom	*Mean*	**1.0000**	0.6469	**1.0000**	0.6163	0.4988	0.9981
	Std. Dev.	0.0000	0.0236	0.0000	0.0357	0.0040	0.0042
	Min	1.0000	0.6125	1.0000	0.5750	0.4875	0.9875
	Max	1.0000	0.6813	1.0000	0.6813	0.5000	1.0000
Cassini	*Mean*	0.9997	0.9937	0.9993	0.9980	0.3333	**1.0000**
	Std. Dev.	0.0011	0.0037	0.0014	0.0023	0.0000	0.0000
	Min	0.9967	0.9867	0.9967	0.9933	0.3333	1.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.3333	1.0000
Chain-link	*Mean*	0.9350	0.6520	0.9370	0.8355	0.5655	**1.0000**
	Std. Dev.	0.0191	0.0435	0.0151	0.0287	0.0673	0.0000
	Min	0.9000	0.5750	0.9050	0.7900	0.5000	1.0000
	Max	0.9600	0.7050	0.9600	0.8800	0.6650	1.0000
Compound	*Mean*	0.9545	0.9428	0.9607	0.9808	0.7987	**0.9858**
	Std. Dev.	0.0053	0.0051	0.0045	0.0112	0.0018	0.0074
	Min	0.9458	0.9375	0.9542	0.9625	0.7958	0.9747
	Max	0.9667	0.9536	0.9667	0.9958	0.8000	0.9958
R-15	*Mean*	0.9989	**0.9993**	0.9992	0.9986	0.8667	0.9989
	Std. Dev.	0.0010	0.0008	0.0007	0.0011	0.0000	0.0012
	Min	0.9967	0.9978	0.9978	0.9967	0.8667	0.9967
	Max	1.0000	1.0000	1.0000	1.0000	0.8667	1.0000

Table 11. Test F1-scores of the synthetic datasets.

	Accuracy	RBF	LIN	POL	AG	AtanK	Proposed AA
3-spiral	*Mean*	0.3418	0.3134	0.3977	0.4574	0.2073	**0.7218**
	Std. Dev.	0.0419	0.0255	0.0342	0.0543	0.0597	0.0328
	Min	0.2566	0.2696	0.3444	0.3768	0.1500	0.6774
	Max	0.4221	0.3527	0.4387	0.5571	0.2993	0.7615
Atom	*Mean*	**1.0000**	0.5957	1.0000	0.5478	0.3328	0.9981
	Std. Dev.	0.0000	0.0349	0.0000	0.0551	0.0018	0.0042
	Min	1.0000	0.5440	1.0000	0.4813	0.3277	0.9875
	Max	1.0000	0.6452	1.0000	0.6452	0.3333	1.0000
Cassini	*Mean*	0.9994	0.9880	0.9987	0.9963	0.0000	**1.0000**
	Std. Dev.	0.0020	0.0070	0.0027	0.0043	0.0000	0.0000
	Min	0.9937	0.9747	0.9937	0.9877	0.0000	1.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000
Chain-link	*Mean*	0.9347	0.6516	0.9369	0.8352	0.4762	**1.0000**
	Std. Dev.	0.0195	0.0437	0.0151	0.0288	0.1454	0.0000
	Min	0.8990	0.5745	0.9050	0.7895	0.3333	1.0000
	Max	0.9599	0.7050	0.9600	0.8799	0.6650	1.0000
Compound	*Mean*	0.6918	0.6200	0.7257	0.9295	0.0945	**0.9474**
	Std. Dev.	0.0361	0.0266	0.0255	0.0423	0.0009	0.0303
	Min	0.6191	0.5861	0.6858	0.8725	0.0931	0.9012
	Max	0.7500	0.6889	0.7521	0.9867	0.0952	0.9867
R-15	*Mean*	0.9917	**0.9950**	0.9941	0.9891	0.0000	0.9916
	Std. Dev.	0.0078	0.0058	0.0058	0.0079	0.0000	0.0089
	Min	0.9753	0.9833	0.9826	0.9749	0.0000	0.9747
	Max	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000
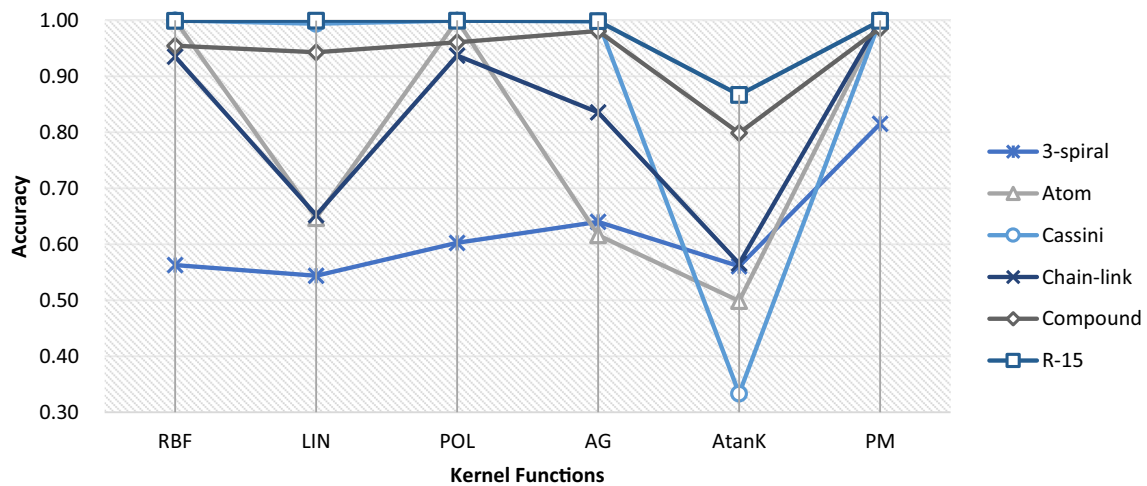
Figure 8. Mean accuracy of the kernel functions on synthetic datasets.

$$F_1 = \frac{2TP}{2TP + FP + FN} \qquad (15)$$

The experimental results of the proposed Adaptive Arctan (AA) KF and the AtanK, AG, RBF, LIN, and POL KFs on the 30 datasets are presented in tables 4, 5, 6 and 7.

Tables 4 and 5 show the experimental results performed on the KEEL and UCI datasets. When analyzing table 4, we observe that the proposed AA KF achieves an average best accuracy value of 0.8488 on the balanced distributed data sets. The closest average accuracy value to the proposed KF is seen as polynomial (PO) 0.8479. The worst success rate
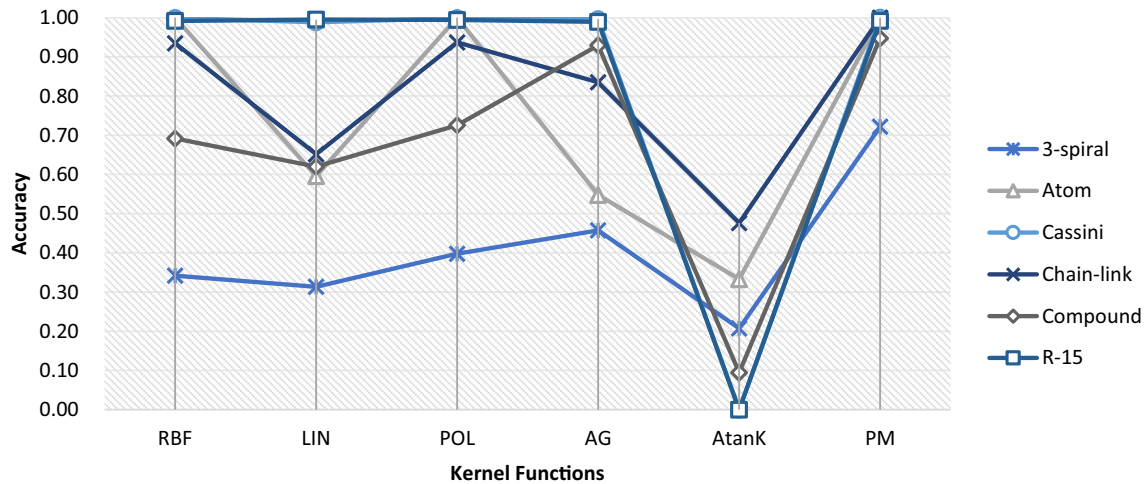
Figure 9. Mean F1-scores of the kernel functions on synthetic datasets.

Table 12. Comparison of the Wilcoxon signed-rank test results ($p = 0.05$) for balanced datasets.

Dataset	AA vs. AtanK	AA vs. AG	AA vs. RBF	AA vs. LIN	AA vs. POL
Appendicitis	3.7756E−04	6.7756E−01	5.3879E−01	1.0000E+00	2.3894E−01
Bupa	2.1951E−13	6.3709E−08	1.5709E−04	2.1951E−13	7.9065E−04
Phoneme	0.0000E+00	5.6441E−120	4.3536E−09	5.0007E−03	1.4259E−05
Ring	0.0000E+00	0.0000E+00	1.6946E−02	2.5085E−14	4.6258E−02
Spam base	0.0000E+00	3.3852E−54	9.6543E−01	2.8672E−01	8.6243E−01
Wine	1.7159E−01	8.3970E−01	8.9123E−01	9.4834E−01	8.3466E−01
Zoo	1.5836E−19	8.4242E−01	1.1916E−01	9.5798E−01	7.5791E−01
Banknote authentication	1.5612E−172	7.0069E−01	4.2065E−01	3.0069E−01	9.3880E−01
User knowledge modeling	1.6815E−36	8.7133E−01	7.7763E−01	8.1704E−01	9.1708E−01
Vertebral column 2C	8.7166E−27	3.0675E−01	6.0307E−02	1.4922E−05	6.6314E−01
Vertebral column 3C	3.4741E−06	4.0022E−01	1.8392E−01	8.1900E−05	9.7042E−01
Breast cancer	5.8236E−09	6.9647E−01	8.9544E−01	5.3265E−03	1.7952E−02
Monks-1	3.7550E−01	5.2464E−01	5.2681E−01	1.0000E+00	4.4753E−01
Monks-2	2.4613E−15	1.8556E−02	2.2871E−01	2.4613E−15	8.9379E−02
Monks-3	1.4864E−20	5.2189E−01	2.5045E−01	3.7164E−01	5.2357E−01
Plrx	3.1998E−01	3.3907E−03	3.1998E−01	3.1998E−01	1.0467E−02
Transfusion	6.7785E−06	2.4643E−94	2.4937E−03	6.7785E−06	5.6296E−03
WDBC	8.2596E−70	9.5238E−01	1.8156E−02	7.5735E−03	1.5353E−02
WPBC	9.0458E−05	2.9196E−02	1.2976E−01	1.2976E−01	1.7942E−03

on balanced datasets was obtained with the AtanK KF. It is given in table 4 that the multi-class dataset Zoo achieved higher success compared to other KFs. In table 5, it can be observed that the suggested AA KF achieved an average best accuracy value of 0.8832 on imbalanced distributed data sets. In addition, the closest average success rate to the proposed AA KF was observed as 0.8236 with the RBF kernel. It has been observed that the best success rate on the multiclass dataset Yeast is associated with the proposed AA KF. Consequently, tables 4 and 5 display that a high success rate was achieved on multi-class datasets. The AA KF, when tested on both balanced and imbalanced datasets,

yielded significantly better results than the original AtanK function. In table 6, confidence intervals at a 95% confidence level were also calculated for each kernel. The proposed AA KF exhibits a more favorable confidence interval profile with a higher mean and a narrower interval compared to other kernels.

Tables 7 and 8 present the results of the F_1-score criterion from experiments performed on the KEEL and UCI datasets. Table 7 shows that the average best F_1-score, which is 0.7821, belongs to the proposed AA KF on balanced distributed datasets. The mean F_1-score value closest to AA KF is 0.7816 in PO KF. Additionally, the worst F_1-

Table 13. Comparison of the Wilcoxon signed-rank test results ($p = 0.05$) for imbalanced datasets.

Dataset	AA vs. AtanK	AA vs. AG	AA vs. RBF	AA vs. LIN	AA vs. POL
Bands	7.6628E−22	1.5018E−02	3.1191E−02	2.1053E−03	1.5018E−02
Cleveland	3.1480E−29	9.7778E−01	9.2331E−01	2.8165E−01	3.2130E−02
German	5.0725E−47	1.3240E−01	2.4249E−01	8.5202E−02	5.5838E−07
Haberman	3.5408E−08	1.6163E−03	2.3955E−06	1.6296E−07	8.1745E−04
House votes	6.7244E−34	7.8089E−01	2.2617E−01	8.5311E−01	7.8089E−01
Ionosphere	1.4314E−34	6.3414E−01	8.7455E−01	1.1414E−02	4.2574E−01
New thyroid	5.9311E−18	2.1773E−01	7.1559E−02	9.2387E−03	4.9336E−01
Yeast	7.3414E−52	6.3309E−17	6.7154E−05	6.3120E−04	9.6747E−05
South German credit	4.0782E−44	1.5387E−01	5.9440E−01	2.4564E−02	2.0663E−10
Glass	5.0129E−01	6.3142E−02	2.7590E−02	3.7958E−01	4.7455E−02
Satimage	0.0000E+00	5.0933E−02	3.8217E−01	7.4414E−02	4.6759E−02

Table 14. Execution times (in sec.) of KF for the KEEL and UCI imbalanced datasets.

Dataset	AAtanK	AtanK	AG	RBF	LIN	POL
Appendicitis	0.00514	0.00169	0.00204	0.00116	0.00104	0.00109
Bands	0.01603	0.01781	0.01710	0.00116	0.00091	0.00141
Bupa	0.00534	0.01306	0.00840	0.00117	0.00109	0.00133
Cleveland	0.02172	0.01623	0.02821	0.00472	0.00271	0.00428
German	0.08093	0.11009	0.08909	0.00160	0.00118	0.00163
Haberman	0.00357	0.00658	0.00492	0.00112	0.00086	0.00112
House votes	0.00257	0.00746	0.00345	0.00133	0.00087	0.00127
Ionosphere	0.00736	0.02166	0.00622	0.00108	0.00096	0.00120
New Thyroid	0.00422	0.00423	0.00587	0.00189	0.00130	0.00174
Phoneme	0.39817	0.91874	0.34395	0.01664	0.00136	0.01567
Ring	0.55989	2.42200	0.56229	0.00312	0.00181	0.00226
Spam base	0.76060	1.42186	0.75576	0.01513	0.00208	0.00821
Wine	0.00413	0.00614	0.00527	0.00159	0.00127	0.00164
Yeast	0.28272	0.63290	0.36073	0.02281	0.01473	0.02086
Zoo	0.01435	0.01334	0.01994	0.00670	0.00472	0.00697
Banknote authentication	0.00730	0.09589	0.01056	0.00153	0.00102	0.00124
South German credit	0.07842	0.11270	0.08613	0.00137	0.00108	0.00135
User knowledge modeling	0.01261	0.03431	0.01775	0.00310	0.00195	0.00276
Vertebral column 2C	0.00329	0.00719	0.00444	0.00099	0.00081	0.00101
Vertebral column 3C	0.00640	0.01361	0.00880	0.00178	0.00127	0.00178
Breast cancer	0.00328	0.00595	0.00464	0.00126	0.00100	0.00106
Glass	0.01715	0.01620	0.02310	0.00539	0.00353	0.00538
Monks-1	0.00203	0.00254	0.00284	0.00095	0.00085	0.00095
Monks-2	0.00306	0.00299	0.00382	0.00093	0.00080	0.00097
Monks-3	0.00263	0.00249	0.00304	0.00114	0.00086	0.00092
Plrx	0.00280	0.00243	0.00368	0.00112	0.00107	0.00104
Satimage	1.47356	10.83170	1.45625	0.01696	0.01188	0.01162
Transfusion	0.00818	0.02965	0.00962	0.00122	0.00089	0.00130
WDBC	0.00652	0.06096	0.00854	0.00108	0.00106	0.00119
WPBC	0.00469	0.00293	0.00493	0.00107	0.00105	0.00120
Mean	*0.12662*	*0.56118*	*0.12871*	*0.00404*	*0.00220*	*0.00348*

score on balanced datasets was 0.3410 in AtanK KF. Table 8 shows that the best average F_1-score value is 0.6916 for AA SF over imbalanced distributed datasets. Furthermore, the average success rate closest to AA KF is observed as 0.6675 in PO CF. Additionally, the worst F_1-score on imbalanced datasets was obtained as 0.2716 in the

AtanK function. It has been observed that much better results are obtained with AA KF when tested on both balanced and unbalanced datasets compared to the original AtanK function. A graphic is shown in figure 7 for the accuracy and F_1-score results of the KFs on the KEEL and UCI datasets.

In table 9, the proposed AA KF exhibits a better confidence interval profile with a higher mean and a smaller interval compared to the other kernels. Figure 7 shows the general classification accuracies of the SVM kernel according to dataset repositories. It is seen that the proposed method (AA) has higher classification accuracy on both the KEEL and UCI dataset repositories compared to other kernels.

The experimental results of the proposed Adaptive Arctan (AA) and AtanK, AG, RBF, LIN, and POL kernels on the six synthetic datasets are presented in tables 10 and 11. The mean accuracy and F_1-scores of the kernel functions on the synthetic datasets are shown in figures 8 and 9 respectively.

Tables 10 and 11 present the results of the accuracy and F1-score criterion from experiments performed on the synthetic datasets. It is observed that the proposed AA KF gives better accuracy and F1-score results in four out of six synthetic datasets compared to other KFs.

In tables 12 and 13, Wilcoxon signed-rank test was used to determine whether there is a difference between the measurement findings of two kernels via results obtained from the data sets in the experimental evaluations. The level of significance for the hypothesis test was determined as 0.05. The proposed adaptive arctan (AA) KF can be clearly distinguished from other KFs, as shown in the test results obtained from the hypothesis test.

When table 14 is analyzed, it is observed that the proposed AA KF on average takes shorter time than AtanK and AG, but takes longer time than the other kernels.

In this study, we propose a new AA KF that adapts to the characteristics of various datasets. The AA KF is obtained by hybridizing the AG and AtanK KFs. When using KFs in SVM, a significant amount of time is required to determine the most suitable σ value for real-world problems. Our proposed SVM with the AA KF streamlines the learning process, as it employs a statistical model in the initial run instead. Instead of using a fixed σ parameter as in the AtanK function, we introduce the AA KF with an adaptive structure tailored to each dataset. This modification results in faster and more efficient convergence for the SVM model. In the results we obtained, it is evident that the SVM model, when used with the proposed KF, performs exceptionally well on the KEEL and UCI datasets. No single method consistently delivers accurate F_1-score values across all datasets, as each dataset possesses unique characteristics. However, the proposed AA KF outperforms most of the previously suggested KFs in terms of average accuracy for the KEEL and UCI repositories.

5. Conclusion

The SVM is preferred by researchers in classification algorithms due to its successful performance, robust mathematical infrastructure, and relatively fast speed. Kernel functions (KFs) have a significant impact on both the performance and computational costs of SVMs. During the training process of the SVM model, manually determining the constant arbitrary value parameter of the AtanK KF can be timE−consuming and may lead to slow convergence of the SVM model. On the other hand, by introducing an adaptive KF into the SVM model, it's possible to obtain the most suitable Hilbert space decompositions for the datasets and the SVM during the training process. As a result, the SVM can achieve faster and better convergence. In this study, a new Adaptive Arctan (AA) KF is proposed for the SVM algorithm, which adapts to the characteristics of different datasets. Experimental evaluations were conducted on 30 public KEEL and UCI datasets using the proposed AA KF as well as AtanK, AG, RBF, linear, and polynomial kernels. The results indicate that the AA KF outperforms the other KFs and exhibits superior learning ability. Future work aims to finE−tune hyper parameters, including computational cost, kernel parameters, and the penalty constant. Additionally, in future studies, performance evaluation will extend to hybrid models involving real-time data and transfer learning. There is also an intention to introduce a new SVM kernel by incorporating an optimized ANN weight vector with a meta-heuristic optimization method, transformed into a Bayesian-optimized the SVM kernel.

Declarations

References

[1] Cortes C and Vapnik V 1995 Support vector machine. *Mach. Learn.* 20(3): 273–297

[2] Kilicarslan S, Adem K and Celik M 2020 Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network. *Med. Hypotheses* 137: 109577

[3] Cetin O and Temurtas F 2021 A comparative study on classification of magnetoencephalography signals using probabilistic neural network and multilayer neural network. *Soft Comput.* 25: 2267–2275

[4] Ozguven M, Yilmaz G, Adem K and Kozkurt C 2019 Use of support vector machines and artificial neural network

methods in variety ımprovement studies: potato example. *Curr. Investig. Agric. Curr. Res.* 6(1): 706–712

[5] Melenk J and Babuška I 1996 The partition of unity finite element method: basic theory and applications. *Comput. Methods Appl. Mech. Eng.* 139(1–4): 289–314

[6] Baş S and Körpinar T 2019 Modified roller coaster surface in space. *Mathematics* 7(2): 195

[7] An-na W, Yue Z, Yun-tao H and Yun-lu L 2010 A novel construction of SVM compound kernel function. In: *2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM)* vol. 3, pp. 1462–1465

[8] Lin H and Lin C 2003 A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. *Neural Comput.* 3(1–32): 16

[9] Müller K, Mika S, Tsuda K and Schölkopf K 2018 An introduction to kernel-based learning algorithms. In: *Handbook of Neural Network Signal Processing*. pp. 4–1

[10] Fan L and Sung K 2003 Model-based varying pose face detection and facial feature registration in colour images. *Pattern Recognit. Lett.* 24(1–3): 237–249

[11] Deng N and Tian Y 2004 *Support Vector Machines: A New Method in Data Mining*. Science Press

[12] Chapelle O, Vapnik V, Bousquet O and Mukherjee S 2002 Choosing multiple parameters for support vector machines. *Mach. Learn.* 46: 131–159

[13] Elen A, Baş S and Közkurt C 2022 An adaptive Gaussian kernel for support vector machine. *Arab. J. Sci. Eng.* 47(8): 10579–10588

[14] Batuwita R and Palade V 2013 Class imbalance learning methods for support vector machines. In: *Imbalanced Learning: Foundations, Algorithms, and Applications*. pp. 83–99

[15] Hofmann T, Schölkopf B and Smola A 2008 Kernel methods in machine learning. *Ann. Stat.* 36(3): 1171–1220

[16] Ding X, Liu J, Yang F and Cao J 2021 Random radial basis function kernel-based support vector machine. *J. Frankl. Inst.* 358(18): 10121–10140

[17] Negri R, Da Silva E and Casaca W 2018 Inducing contextual classifications with kernel functions into support vector machines. *IEEE Geosci. Remote Sens. Lett.* 15(6): 962–966

[18] Tian M and Li H 2021 AtanK-A new SVM kernel for classification. *Appl. Math. Nonlinear Sci.* 8(1): 135–146

[19] Khan I, Fattah H A and Shill P C 2021 A hybrid kernel support vector machine for predicting protein structure. In: *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*. pp. 1–5

[20] Liu X and He W 2022 Adaptive kernel scaling support vector machine with application to a prostate cancer image study. *J. Appl. Stat.* 49(6): 1465–1484

[21] Ba W, Lu Z and Li Q 2021 Prediction of product quality based on multiple kernel learning least square support vector machine regression. In: *2021 China Automation Congress (CAC)*. pp. 3197–3201

[22] Kiliçarslan S and Celik M 2022 KAF+ RSigELU: a nonlinear and kernel-based activation function for deep neural networks. *Neural Comput. Appl.* 34(16): 13909–21392

[23] Kilicarslan S, Celik M and Sahin Ş 2021 Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification. *Biomed. Signal Process. Control* 63: 102231

[24] Kiliçarslan S and Celik M 2021 RSigELU: a nonlinear activation function for deep neural networks. *Expert Syst. Appl.* 174: 114805

[25] Pacal I and Kilicarslan S 2023 Deep learning-based approaches for robust classification of cervical cancer. *Neural Comput. Appl.* 35(25): 18813–21882