International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Performance Analysis of Poly Cystic Ovary Syndrome (PCOS) using Broyden's Kernel Import Point (BKIP) Classifier

Rakshitha Kiran [1], Dr. Naveen NC [2]

Submitted: 10/09/2022 **Accepted:** 20/12/2022

Abstract: Machine Learning (ML) is a progressive and immensely approached technological application which became an enormous trend within the industry. ML is widely utilized in various applications and can be utilized by healthcare companies to acquire valuable data that can be used to diagnose diseases in the earlier stage. In this paper, a new classification approach Broyden's Kernel Import Point (BKIP) classifier is proposed and summarizes the use and its application in healthcare domain. BKIP classifier is a new approach used for classification which is built on Kernel Logistic Regression and Import Vector Machine (IVM). Data for the current study project was gathered from the ESIC Hospital in Bengaluru, India, and it was found that the BKIP algorithm's computational cost was significantly lower than that of the Support Vector Machine (SVM) and the IVM. The paper provides the implementation of the BKIP classification algorithm and it is noted that when applied to Polycystic Ovary Syndrome (PCOS) data, the model produced an accurate result of 89.1 %.

Keywords: Broyden; Import Vector Machine; Support Vector Machine; Machine Learning.

1. Introduction

Healthcare prediction using ML techniques has been prominent nowadays. Popular algorithms such as SVM and IVM has gained attention for a decade [1] [2]. BKIP is a new approach that is built upon IVM. BKIP classifier can perform efficiently as it employed a small part of the training data to index the kernel basis function [3]. As a result, the BKIP classifier has the advantage over IVM and in the current proposed methodology Broyden's is used as the core function which makes the algorithm less expensive. IVM is built on the Newton-Rampshon methodology which is computationally expensive. Electronic Health Record's (EHR) [4], store patient records acquired during healthcare decisions as a form of clinical data. It's very difficult to work with this raw data manually and ML has come to prominence as a sophisticated tool for analysis. In healthcare, ML can help doctors to perceive and understand the patient's situation as well as identify better treatment based on the present disease condition [5]. It is observed [6] that IVM is mainly applied to the spatial datasets and it gives potentially good outcomes, but its application and performance when applied to healthcare datasets is limited. To address this drawback BKIP classifier is applied which uses Genetic Algorithm in the initial stage for processing the datasets and selecting the best feature set [7-8]. BKIP classifier is an improvised version of the IVM classifier and has a greater potential to handle healthcare data. This paper summarizes the implementation of the BKIP classifier for the prediction of infertility in obese women [9]. The ladies in the age range of 19 to 40 years deal with the very serious issue of PCOS and if left untreated can lead to several other health issues and will finally lead to infertility [10]. Obesity is the primary cause of infertility and the fact that PCOS causes obesity or obesity causes PCOS is still under research.

[1]*Assistant Professor, Department of MCA, Research scholar at ISE, Dayananda Sagar College of Engineering,*
VTU University, Bengaluru, Karnataka, India.
[2]*Professor and Head, Department of Computer Science & Engineering, JSS Academy of Technical Education, VTU University, Bengaluru, Karnataka, India.*
Corresponding Author Email: naveennc656@gmail.com

2. Theoretical Background

2.1 Logistic Regression (Lr) And Kernel Logistic Regression (Klr)

(A)*LR*: Let us consider the training set (p_n, q_n) where k=1,…,K where N is the labeled samples with features vectors p_n€ IR^M and class labels q_n € P= {P_1, . . ., P_K}. These interpretations are processed in the form of a matrix which is represented as P, Where S= {s_1...,s_N}, and the labels are abbreviated in the vector form with is represented as q=[q_1,…,q_N].

The posterior probability pb_n for a two-class classification of p_n is supposed to obey the LR method which is represented as follows:

$$pb_n = pb(q_n = B_1|p_n;z) = \frac{1}{1+exp\,(-z^T x p_n)} \qquad (1)$$

Where the extended feature vector is represented as $P^T{}_n = [1, p^T{}_n]$ € IR^M and the extended parameters is represented as $Z^T = [z_0, z^T]$ € IR^{M+1} here z_0 is biased and the z is weight vector.

Kernel Logistics Regression: Classical logistic regression will not classify accurately for non-linearly separable data, hence the kernel version. In the case of linear non-separable data, the original observations which are represented as P. The kernel matrix ƙ= ƙ_{nm} where the kernel function is ƙ_{nm} = ƙ(p_n, p_m). In kernel-based approach parameters are referred as ά, and works in an iterative way:

$$\acute{\alpha}_{(i)}=(\frac{1}{N}\,\text{ƙ}^T\,\text{Я}\text{ƙ} + \text{ƛ}\,\text{ƙ})^{-1}\,\text{ƙ}^T\text{Я}z \qquad (2)$$

$$\acute{Z}=\frac{1}{N}(\text{ƙ}\,\acute{\alpha}_{(i-1)} + \text{Я}^{-1}(\text{ƥ} - t)) \qquad (3)$$

By Using the Newton-Raphson procedure object function is optimized:

$$\bar{Q}_{(i)} = -\frac{1}{N}\sum[t_n \log \beta_n + (1-t_n)\log(1-\beta_n)] + \frac{\lambda}{2}\alpha T_{(i)} K \acute{\alpha}_{(i)} \quad (4)$$

The elements of (N x N)-dimensional diagonal matrix Я is $\check{r}_{mn} = \beta_n(1-\beta_n)$ where $\beta_n = \frac{1}{(1+expexp\,(-kn\acute{\alpha}))}$ and k_n refers to the n^{th} rows of the kernel matrix ƙ.

The binary target vector of length N codes where t € {0, 1} with the labels t_n=0 for $y_n = C_1$ and $t_n = 1$ for $y_n = C_2$ Additionally parameter λ is added to prevent overfitting.

2.2 Import Vector Machines (IVM)

Zhu, T. Hastie [12] proposed IVM, which is based on KLR. IVM algorithm relies on a classic LR model which has been enhanced with kernel features. The subset 'v' is termed as set of import vectors. Up until the convergence principle is satisfied, the import vectors are originally determined using data samples for the empty set. The selection of the data samples is based on the observation that include the subset "v" reduces the objective function. The initial import vector is chosen as the data point at first. This site results in the best one-class categorization, which indicates that it is located in the region with the greatest concentration of class-specific data [13]. After that, the second import vector that produces the best linear reparability to one of the other classes is picked.

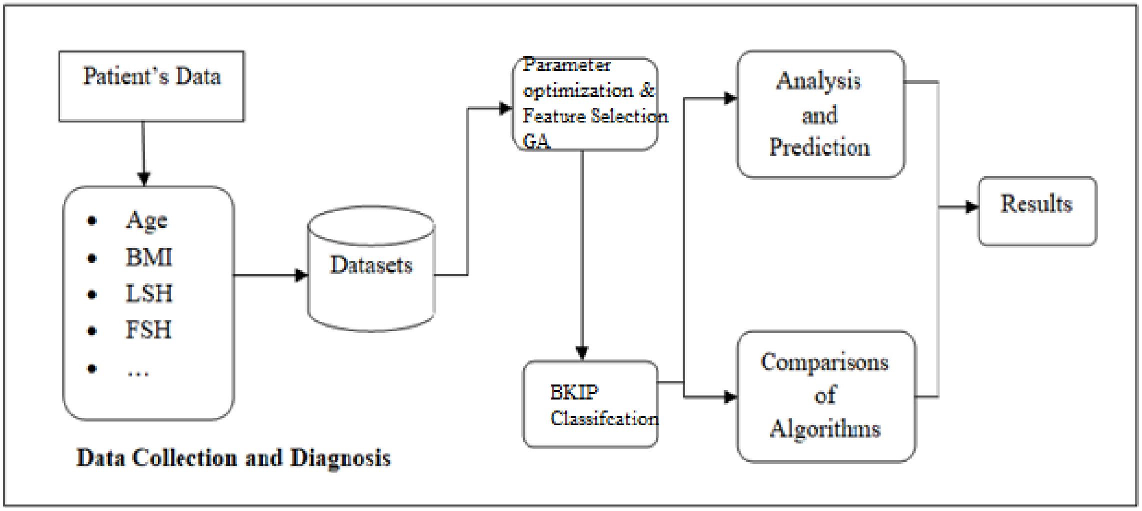

Fig. 1. Overall working of BKIP Classifier

All of the training data are used to train the classifier in kernel logistic regression. This requires a lot of memory and processing resources for data sets with a lot of training examples. Similar to the IVM method and SVM [12], the kernel logistic regression is solved spatially by selecting a subset of feature vectors from the training set with Ṿ = |v| samples P_v = [$P_{v,m}$] where m = 1,...v. Import vectors are the name given to these feature vectors.

The parameters in iteration I are established via (2) and (3).

$$\acute{\alpha}_{(i)} = \left(\frac{1}{N} k^T_v Я k_v + \lambda k Я\right)^{-1} k_v^T Я z \quad (5)$$

$$\acute{Z} = \frac{1}{N}\left(k_v\, \acute{\alpha}_{(i-1)} + Я^{-1}(\beta - t)\right) \quad (6)$$

3. Research methodology BKIP Classifier:

The PCOS dataset is preprocessed using the standard preprocessing technique to remove all types of unwanted data, ambiguous data, and missing values [14]. The dataset is optimized using a Genetic Algorithm where the feature extraction takes place. The top 10 best features are extracted and then sent to the BKIP classifier.

GAs are a type of heuristic remedy or optimization technique that was inspired by Darwin's idea of evolution by selection. A GA evolves solutions to problems using a very abstract version of evolutionary processes. The successor population is produced by GA using fitness-based selection and recombination.

3.1 BKIP Classification Algorithm working methodology:

This paper proposes a unique classification technique BKIP [15], which overcomes the drawbacks of existing ML algorithm. Table 1 shows the summary of SVM, IVM, and BKIP Classification algorithm comparing the objective_ function, optimization Technique, Sparse, Training_ time, Testing_ time, and Probabilistic properties.

Table 1: Summary of The Svm, IVM, and BKIP Algorithm concerning a few Characteristics.

Approaches	Objective_ function and Optimization Technique	Sparse	Training_ Time	Testing_ time	Probabilistic
IVM	convex, IRLS with greedy FSS*1	+	0/+	+	+
BKIP Classifier	Convex, IRLS*2 and search-based and Genetics and Natural Selection protocols optimization procedure	+	+	+	+
SVM	convex, greedy with SMO algorithm	0	+	0	0

*1Forward Stepwise Selection.

*2 Iterated Re-weighted least squares

*Note: Here "–" Means the Algorithm "Barely" satisfies the mentioned property, "0" indicates that the Algorithm "Partially" fulfils the mentioned condition and "+" Means the Algorithm "Completely" satisfies the property. The SVM algorithm has quadratic as Objective function and therefore its property is convex. Here the optimization problem is competently solved with sequential minimal optimization Technique [7]. Similarly, IVM has a concave Objective function that can be solved using the IRLS technique and a greedy forward selecting of import vectors even though it is non-quadratic [16].

The BKIP classifier's objective function is also symmetric and employs both a natural selection method and the IRLS technique. The BKIP algorithm works as follows: The datasets $\{d_1, d_2, \ldots, d_n\}$ are preprocessed and processed using a Genetic Algorithm which performs parameter optimization and Feature Selection [17]. The GA model selects the top five features which actually contribute to the prediction $\{d_1, d_2, \ldots, d_5\}$.

The BKIP classifier identifies the subset of Š of {c1, c2, … cn} such that the sub-model provides a reasonable approximation of the entire model. A computationally costly, greedy forward approach is used here to avoid the combinatorial difficulty that can arise from trying to find all subsets. Initially the null model Š =ø is considered and then built up iteratively. The main idea followed here is finding a data point among $\{c_1, c_2, \ldots c_n\} \backslash$ Š, such that the new sub-reduced Š model's regularized negative log-likelihood will result from adding it to the existing.

Step 1: Let Š = $\emptyset$, £ = $\{C_1, C_2, \ldots C_n\}$, p=1

Step 2: For each C_0 € £, Let

$$u(C) = \sum_{ci \in \check{S}\, U\, \{c1\}} Zi\, P(c, ci) \qquad (7)$$

Step 3: Use Broyden's Method to Minimize 'a': Here is the Broyden's equation:

$$J_n = J_{n-1} + (\Delta F_n - J_{n-1}\, \Delta C_n / |C_n|^2) * \Delta C_n^T \qquad (8)$$

Where C_0 is the initial parameter and C_2..Cn uses the Secant method uses the finite difference approximation.

$$\check{S}(C_l) = \frac{1}{n}\sum_{i=0}^{n} \ln(1+\exp(-b \int l\,(Ci))) + \frac{\lambda}{2} \| g_l(x) \|^2\, \hat{H}_K \qquad (9)$$

$$= \frac{1}{n}\sum_{i=0}^{n} \ln(1+\exp(-b.\, P^l{}_1.J_n a) + \frac{\lambda}{2} a^{T.} J(n).P_2{}^l a \qquad (10)$$

where the regressor matrix

$$P^l{}_1 = (P(C_i, C_{i`}))_{n*p},\ C_i \in \{C_1, C_2, \ldots C_n\},\ C_{i`} \in \check{S}\ U\ \{C_l\} \qquad (11)$$

the regularization matrix

$$P^l{}_1 = (P(C_i, C_{i`}))_{n*p}, C_i, C_{i`} \in \check{S}\ U\ \{C_l\}; \qquad (12)$$

and p=| Š | +1;

Step 4: Now find the value of

$$C_l = (\check{S}\,(Cl))\ Cl \in £ .$$

Let Š = Š ∪{ C_l ∗ },O = O\{ C_l ∗ }, $\check{S}_k$ = Š (C_l ∗), p = p + 1.

Step 5: Once Šk iterates, execute Processes 2 and 5 once more.

The stopping criteria is that if the ratio $\frac{\check{S}k - \check{S}k - \Delta k}{\check{S}k}$ is less than some prechosen small number, adding new import points to Š is stoped.

Choosing The Regularization Parameter Lambda (λ) and Sigma (Σ)

Regularization parameters of BKIP Classifier Sigma and Lambda are not fixed. To acquire precise results, the best and most ideal value must be selected. The complete collection of data is arbitrarily divided into a training set and a testing set. Misclassification error on the testing set serves as the primary factor for selecting λ & Σ.

For this experiment the value of λ is assigned as exp(-11) and value of Σ is assigned as 12, based on the misclassification error. Every iteration the value keeps changing and the figure 2 and 3 below shows the Optimal Lambda and Sigma value.

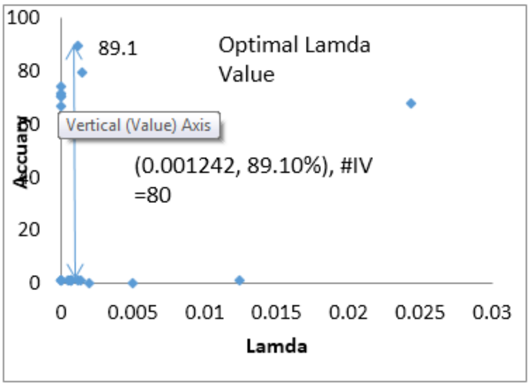

Fig. 2. Optimal Lambda Value

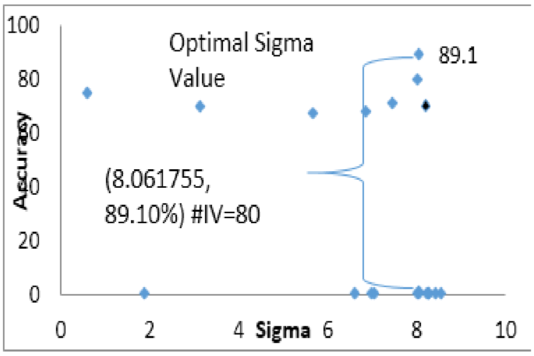

Fig. 3. Optimal Sigma Value

Table 2. Comparison Chart of IVM and BKIP in terms of various parameters.

Algorithm	n	Lambda	Sigma	Fval*	Rat*	Err_train/N*	Accuracy
IVM	1600	Exp(-11)	12	0.693147	INF	0.32	67.2
BKIP	1600	0.000651(optimal)	8.278592(optimal)	0.494179	0	0.22	84.61

Note: Rat represents Ratio of Negative log-likelihood and fval represents Negative log-likelihood, N represents Dimension, n represents no of datasets, Lambda and Sigma are the regularization parameter which is constant for IVM but BKIP its variable. Err_ train represents the training error.

4. Results and Discussions

The BKIP and the IVM's results on PCOS datasets are contrasted in this section. An overview of these datasets is provided in Table 2. These datasets all employ a radial kernel (2.4). For the BKIP, the variable is based on generalization performance, whereas the parameters and are fixed at particular values that are best for the IVM. The table infers that the performance of BKIP is better when compared to IVM.

Table 3 compares the performance of the IVM and BKIP classifiers based on the total number of datasets. It is observed that IVM doesn't show much progress in Accuracy when the total number of datasets are increased but whereas in BKIP Classifier, the performance significantly increases as the as the no of datasets increases.

4.1 Optimization Results

The Table 4 shows the output of BKIP Classification Algorithm output. It is clear that when the IP (Import Point) value is 80 the model has achieved highest Accuracy with 89.1% with the Ratio of Negative log-likelihood is 0.0013 and Negative log-likelihood is 0.396787 and the training error is 0.11.

Table 3: Comparison of IVM and BKIP based on the count of datasets

No of datasets	IVM	BKIP
400	**73.7**	79.2
600	84.6	80.1
800	67.8	80.7
1000	60.8	88.2
1200	75.9	88.4
1600	73.2	89.1

Table 4. Output of BKIP Classification Algorithm

ambda	Sigma	Accuracy	#IP	Ratio of Negative log-likelihood	Negative log-likelihood	training error
0.000049	0.591828	74.7	150	0.0027	0.495131	0.25
0.00065	8.025803	84.7	150	0.0015	0.500105	0.15
0.00141	**8.059192**	**88.2**	**111**	**0.001**	**0.443757**	**0.11**
0.000469	8.252924	82.9	150	0.001	0.471621	0.17
0.000001	1.510942	65.3	16	0	0.644957	0.33
0.00007	7.029014	84.2	30	0.048	0.400413	0.14
0.001242	**8.061755**	**89.1**	**80**	**0.0013**	**0.396787**	**0.11**
0.000001	5.078608	77.3	23	0.0009	0.530792	0.23
0.012384	6.987424	81.8	18	0.0006	0.515698	0.18
0.000651	8.241517	87.3	150	0.0016	0.435532	0.13
0.000001	3.125	70.4	15	0.0085	0.554224	0.29
0.00151	8.025318	79.8	65	0.001	0.557704	0.2
0.000766	8.336958	84.90	150	0.0014	0.502627	0.15
0.000003	8.534278	79.8	35	0.009	0.443074	0.2
0.000651	8.277776	85.1	150	0.0017	0.465015	0.15
0.000625	8.262028	86.4	150	0.0015	0.470594	0.14
0.005043	6.591216	71.3	19	0.001	0.629474	0.29
0.00193	1.875	73.8	51	0.0009	0.596405	0.26
0	8.442832	82.9	34	0.0042	0.375633	0.17

Figure 4 shows the output of Genetic Algorithm when combined with BKIP Classifier. The figure 4a. Shows the penalty value: Best: -0.868132 Mean: -0.693407 for 15 generations. The second figure 4b demonstrates how the Fitness function measures how closely a particular solution adheres to the ideal answer to the intended issue. It establishes how suitable a solution is.

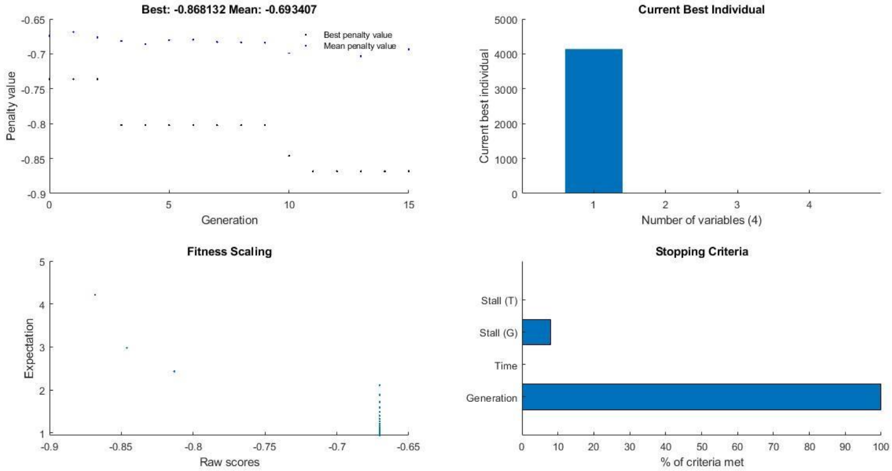

Fig. 4. 4a) Best penalty value, 4b) Optimal-Random seed Generation 4c) Fitness Scaling and 4d) Stopping Criteria.

5. Conclusion

The main outcome of this work is to show the implementation of Broydon's Kernel Import Point (BKIP) Classification methodology in Healthcare domain. This paper introduces BKIP Algorithm and its importance in the initial stage followed by the working Algorithm. A working model is proposed and is applied on PCOS datasets. The Datasets has been optimized and the Feature Selection is carried out using Genetic Algorithm. The proposed BKIP model accurately performs classification and prediction giving the accuracy of 89.1% which is better when compared to IVM. Through this research it is evident that BKIP performs better in predicting infertility with increase in training data set. The research work has been carried out for 1600 PCOS Datasets obtained from ESIC Hospital Bengaluru. The limitation of this research is that the datasets are focused on Bengaluru (Urban), India region.

Conflict of interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

Funding: Not applicable

Availability of data and material: Not applicable

References:

[1] Zhu, T. Hastie, Kernel logistic regression and the import vector machine, Comput.Graph. Stat. 14 (2005) 185–205.278R. Roscher et al. / Image and Vision Computing 30 (2012) 263–278.

[2] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. Benediktsson, "SVMand MRF-Based Method for Accurate Classification of Hyperspectral Images," IEEE Geosci. Remote Sens. Lett., vol. 7, pp. 736–740, 2010.

[3] M. Pal and P. Mather, "Some Issues in the Classification of DAIS Hyperspectral Data," Int. J. Remote Sens., vol. 27, no. 14, pp. 2895–2916, 2006.

[4] Ehrmann DA: Polycyctic ovary syndrome. N. Engl. J. Med. 352, 1223–1236 (2005).

[5] Professor Cindy Farquhar , Associate Professor Neil Johnson, Dept of Obstetrics and Gynaecology, University of Auckland "*Understanding polycystic ovary syndrome*"Issue 12 April 2008, ISSN 1177-5645 (Print) ISSN 2253-1947 (Online).

[6] B. Waske, J. Benediktsson, K. Arnason, and J. Sveinsson, "Mapping of Hyperspectral AVIRIS Data using Machine-Learning Algorithms," Can. J. Remote Sensing, vol. 35, pp. 106–116, 2009.

[7] J. Platt, Advances in Kernel Methods - Support Vector Learning. MIT Press, 1999, ch. Fast Training of Support Vector Machines using Sequential Minimal Optimization, pp. 185–208.

[8]] B. Waske, J. Benediktsson, K. Arnason, and J. Sveinsson, "Mapping of Hyperspectral AVIRIS Data using Machine-Learning Algorithms," Can. J. Remote Sensing, vol. 35, pp. 106–116, 2009.

[9] Ahmed Cherif Megri, Sameer Hamoush, Ismail Zayd Megri, Yao Yu. (2021). Advanced Manufacturing Online STEM Education Pipeline for Early-College and High School Students. Journal of Online Engineering Education, 12(2), 01–06. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/47

[10] Kelley, C. T., and E. W. Sachs. "Broyden's Method for Approximate Solution of Nonlinear Integral Equations." *Journal of Integral Equations*, vol. 9, no. 1, 1985, pp. 25–43, http://www.jstor.org/stable/26164266. Accessed 19 Apr. 2022.

[11] Yang, J.-P. . "A Novel Storage Virtualization Scheme for Network Storage Systems". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 1, Jan. 2022, pp. 08-13, doi:10.17762/ijritcc.v10i1.5514.

[12] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt, "Spatio- Spectral Remote Sensing Image Classification with Graph Kernels," IEEE Geosci. Remote Sens. Lett., vol. 7, no. 4, pp. 741–745, 2010.

[13] J. Mu˜noz-Mar´ı, F. Bovolo, L. G´omez-Chova, L. Bruzzone, and G. Camp- Valls, "Semisupervised One-Class Support Vector Machines for Classification of Remote Sensing Data," IEEE Trans. Geosci. Remote Sens., vol. 48, no. 8, pp. 3188–3197, 2010.

[14] Cristianini, N., & Shawe-Taylor, J. (2000). Support Vector Machines. In *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* (pp. 93-124). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511801389.008.

[15] A. Mathur and G. Foody, "Multiclass and binary SVM classification: implications for training and classification users," IEEE Geosci. Remote Sens. Lett., vol. 5, no. 2, pp. 241–245, 2008.

[16] G. Foody, "RVM-Based Multi-Class Classification of Remotely Sensed. Data," Int. J. Remote Sens., vol. 29, no. 6, pp. 1817–1823, 2008.

[17] J. Platt, N. Cristianini, and J. Shawe-Taylor, "Large Margin DAGs for Multiclass Classification," Adv. Neural Inf. Process. Syst., vol. 12, no. 3, pp. 547–553, 2000.

[18] Ghazaly, N. M. . (2022). Data Catalogue Approaches, Implementation and Adoption: A Study of Purpose of Data Catalogue. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(1), 01–04. https://doi.org/10.17762/ijfrcsce.v8i1.2063

[19] G. Foody, "RVM-Based Multi-Class Classification of Remotely Sensed Data," Int. J. Remote Sens., vol. 29, no. 6, pp. 1817–1823, 2008.

[20] M. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," J. of Mach. Learn. Research, vol. 1, pp. 211–244, 2001.

[21] Chawla, A. (2022). Phishing website analysis and detection using Machine Learning. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 10–16. https://doi.org/10.18201/ijisae.2022.262

[22] F. Giacco, C. Thiel, L. Pugliese, S. Scarpetta, and M. Marinaro, "Uncertainty Analysis for the Classification of Multispectral Satellite Images Using SVMs and SOMs," IEEE Trans. Geosci. Remote Sens., no. 99, pp. 1–11, 2010.

[23] Vapnik, V.(1995).:"Support-vector networks".Machine Learning. 20 (3): 273–297. doi:10.1007/BF00994018.

[24] Christopher Houck, Jeffery Joines, Michael Kay, A Genetic Algorithm for Function Optimization: A Matlab Implementation, NCSU-IE Technical Report 95-09, 1996.