Volume 25 Issue 04, 2022 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

APPLICATION OF SUPPORT VECTOR MACHINE (SVM) FOR DIAGNOSIS OF CANCER

Prathap. S

Department of Mathematics, S. A. College of Arts & Science, Thiruverkadu, Chennai, India

Jamberi. K

Department of Computer Science, S. A. College of Arts & Science, Thiruverkadu, Chennai, India

Anthony Raj. A

Department of Mathematics, Panimalar Engineering College, Chennai, India

Thiyagarajan, G

Department of Mathematics, S. A. College of Arts & Science, Thiruverkadu, Chennai, India

Email: [1]prathapmaths@gmai.com, [2]kjamberi111@gmail.com, [3]rajaloyola16@gmail.com, [4]thyaga179@gmail.com

Abstract- In this research, four linear SVM classifiers are used to rapidly diagnose breast cancer. SVM classifiers were used to analyze breast cancer using the WBCD informative index. There are various modified classifiers, such as Linear Programming SVM and Lagrangian SVM, that are compared to SVM in terms of classification performance. SVM surpassed all other algorithms for all exhibition lists, with an accuracy of 97.71 percent, whereas Lagrangian SVM had the lowest accuracy, at 95.61 percent. A distant second place goes to linear programming SVM (with an accuracy of 97.33 percent). In order to get the best results, it is essential that the classifier and kernel functions be determined. During the validation stage, Linear Programming SVM achieved an overall accuracy of 97.14 percent, outperforming Lagrangian SVM (95.43%), Proximal SVM (96 %), and SVM (94.86 %). The overall sensitivities of Linear Programming SVM accomplished 98.25%, which is better than Lagrangian SVM (96.52%), Proximal SVM (97.37%) and SVM (95.65%). The overall specificities of Linear Programming SVM accomplished (95.08 %), which is better than Lagrangian SVM (93.33%), Proximal SVM (93.44%) and SVM (93.33%). Estimation of AUC for LPSVM accomplished 99.38%, individually, which beat different classifiers. The outcomes firmly recommend that Linear Programming SVM can help in the analysis of cancer data.

Keywords- Classification, SVM, Proximal SVM, Lagrangian SVM, Linear Programming SVM

I. INTRODUCTION

From World Health Organization, Carcinoma is the most continuous malignant growth among ladies, affecting 2.1 million ladies every year, and furthermore causes the best number of malignancy related passing's among ladies. In 2018, it is evaluated that Six lakhs twenty even ladies passed on from breast cancer – that is around 15% of all disease passing's among ladies.

Submitted: 17/10/2022 **Accepted**: 14/11/2022

While Carcinoma rates are greater among women who live in more developed places, rates are increasing in practically every region of the world. Early detection and treatment of breast cancer are essential for improving the outcome and endurance of the disease. Early detection techniques for breast cancer are classified as either early determination or screening procedures.

Artificial intelligence technologies have been studied to accurately categories breast cancer cases. Liu et al [9]. Vapnik [23] introduced SVM, a relatively new AI system. Its goal is to lower the upper limit of the speculative error in the structural risk minimization (SRM) head as much as possible.SVM training is analogous to resolving a linear quadratic programming issue. [4]. It has been used in several domains such as handwritten digit recognition [3, 23, 17, 18], object recognition [18], and face recognition in photos [12]. Two considerations must be made while using SVM for practical concerns [4].A good SVM model relies heavily on the right kernel parameter configuration. The key parameter, C, sets the tradeoff between fitting error reduction and model complexity. γ is the kernel function.

The aim is to compare the performance of SVM, Proximal SVM, Lagrangian SVM and Linear Programming SVM

II. MATERIALS ANDMETHODS

2.1 Support Vector Machine (SVM)

Vapnik's SVM technique is based on the SRM concept and is one of the most extensively used machine learning algorithms because it delivers exceptional generalisation performance for both classification and regression applications [3, 18, 23]. When training samples are separable, SVM solve the classification issue by locating the hyperplane in the feature space that maximises sample margin. SVM has been successfully used to a variety of complicated, real-world issues, including handwriting recognition, object identification, data mining, bio informatics, pharmaceuticals, financial forecasting, and stock market trading.

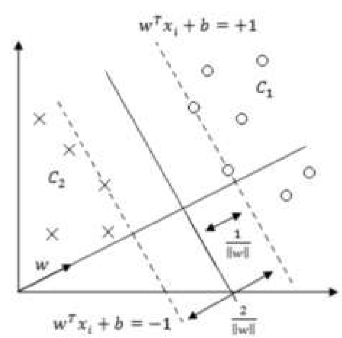

Figure 2: Linear SVM

For linear SVM, Consider the training set D be (x_i, y_i), $i = 1,2,..n$, where $x_i \in R^n$ and the output label $y_i \in \{+1, -1\}$ for the hyperplane $w^T.x + b = 0$, which separates into two classes by satisfying the constraints $(w^T.x_i + b) \geq 1, \forall\, i$ where $y_i = 1$ and $(w^T.x_i + b) \leq -1, \forall\, i$ where $y_i = -1$. Combining both the constraints, $y_i(w^T.x_i + b) \geq 1, \forall\, i = 1,2,..n$. The distance between two hyperplanes, $w^T.x_i + b = 1$ and $w^T.x_i + b = -1$ is $\frac{2}{\|w\|}$ and is known as margin of the classifier. Hence optimization problem which maximizes the margin

Minimize $= \frac{1}{2} \| w \|^2$

subject to $y_i(w^T.x_i + b) \geq 1, \forall\, i = 1,2,..n.$

With Lagrangian multiplierα_i, the objective function becomes

Min $= \frac{1}{2} \| w \|^2 - \sum_{i=1}^{n} \alpha_i (y_i(w^T.x_i + b) - 1)$, subject to $\alpha_i \geq 0$.

The Lagrangian dual formulation of the above is

Max $L_D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=0}^{n} \sum_{j=0}^{n} \alpha_i \alpha_j y_i\, y_j x_i{}^T x_j$

subject to $\alpha_i \geq 0$ and $\sum_{i=1}^{n} \alpha_i y_i = 0$.

The decision function is $f(x) = sign(w^T.x + b)$. If D is not linearly separable, the optimization problem will be unsolvable. To address these situations, the slack variables ξ_i 's are added to quantify the degree of misclassification. The optimization problem for soft margin SVM is Minimize $= \frac{1}{2} \| w \|^2$+C$\sum_{i=1}^{n} \xi_i$ subject to $y_i(w^T.x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall\, i$

Where C is a regularization parameter that determines the tradeoff between the margin size and training error. Let α_i's be the Lagrangian multipliers for $y_i(w^T.x_i + b) - 1 + \xi_i \geq 0$ and μ_i's be the Lagrangian multipliers for $\xi_i \geq 0$. The Lagrangian primal objective function is

Minimize $L_P(\alpha) = \frac{1}{2} \| w \|^2$+C$\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i [y_i(w^T.x_i + b) - 1 + \xi_i] - \sum_{i=1}^{n} \mu_i \xi_i$

The dual of this is

Maximize $L_D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=0}^{n} \sum_{j=0}^{n} \alpha_i \alpha_j y_i\, y_j x_i{}^T x_j$ subject to $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^{n} \alpha_i y_i = 0$.

With the exception of the upper constraint for the Lagrangian multipliers, this is the same as the previous example. In this scenario, the value of w is the same as before, and the value of b may be determined by applying the Karush Kuhn Tucker criteria to the primal. Because of this, the ideal hyperplane $w^T.x+b=0$ has been established. The decision-making function is referred to as

$f(x) = sign(w^T.x + b) = sign(\sum_{i \in SV} \alpha_i y_i\, x_i.x + b)$.

2.2 Proximal SVM

The SVM issue may be thought of as categorizing m points in R^n, the mxn matrix A, which can be represented as follows: In this case, consider the diagonal matrix D, which has one or two values along its diagonal, depending on the classification of the data. The issue may be written as follows in matrix notation, using the value υ>0 as a parameter:

$$\min_{w,\gamma,y \in R^{n+1+m}} \upsilon e' y + \frac{1}{2}(w'w)$$

$$\text{such that} \quad D(Aw - e\gamma) + y \geq e, \quad y \geq 0$$

$$\min_{w,\gamma,y \in R^{n+1+m}} \upsilon \frac{1}{2} \| y \|^2 + \frac{1}{2}(w'w + \gamma^2)$$

$$\text{such that} D(Aw - e\gamma) + y \geq e, \quad y \geq 0$$

Replace inequality constraint, with equality constraint,

$$\min_{w,\gamma,y\in R^{n+1+m}} \upsilon\frac{1}{2}\parallel y\parallel^2+\frac{1}{2}(w'w+\gamma^2)$$
$$\text{such that} D(Aw-e\gamma)+y=e$$

These alteration changes the nature of the optimization problem essentially, as the explicit accurate solution can be given as $w=A'Du, \gamma=-e'Du, \upsilon y=u$ with $u=\left(\frac{I}{\upsilon}+HH'\right)^{-1}e$, where H is defined as,$H=D[A-e]$ by Lagrangian formulation and Karush Kuhn Tucker optimality conditions [7].

2.3 Lagrangian SVM

According to Mangasarian and Musicant (2000, 2000), a well-known Lagrangian for the dual of a basic quadratic LSVM reformulation was proposed, which they dubbed "the well-known Lagrangian for the dual." It was used in the development of a very successful iterative technique. An unconstrained differentiable convex function with the same dimension as the number of specified points is minimized as a consequence of this in a three-dimensional space with the same dimension as the number of specified points. The KKT necessary and sufficient requirements for the dual issue are legally relied on by the Lagrangian Support Vector Machine

$$0\leq u\perp Qu-e\geq 0$$

The optimal condition, for any positive α,

$Qu-e=((Qu-e)-\alpha u)_+,\ 0<\alpha<\frac{2}{\upsilon}$.

Setting the slope as for u of this convex and differentiable Lagrangian to zero gives

$$(\alpha I-Q)\left((Qu-e)-\left((Q-\alpha I)u-e\right)_+\right)=0.$$

2.4 Linear Programming SVM

Using an articulated element choice property for linear classifiers, Fung and Mangasarian (2001) devised a rapid method for resolving a basic classification issue in data mining. When nonlinear kernels are utilized, the calculation includes a decision in the dual factor's high-dimensional space, resulting dependent on just a few components of the kernel function. Feature selection and fast nonlinear kernel classifiers, such as those needed for online decision making, misrepresentation detection, and interruption identification, are all made possible with this method. A linear equation solver is all that is needed to do the Linear Programming SVM computation, making it simple, fast, and open. There are several categorization difficulties in gene expression microarray data that may be solved well using linear programming and support vector machines (SVMs). Additionally, SVMs may be used to categorize large informative indexes in a more compact information space using Linear Programming SVM.

2.5 Performance Evaluation Tools

There are a several measuring tools to assess the execution of the classifiers. They are sensitivity, specificity, accuracy and AUC. Formula for measuring tools are given below ([1], [8], [16], and [19])

$$Sensitivity(\%) = \frac{TP}{FN + TP} \times 100$$

$$Specificity(\%) = \frac{TN}{TN + FP} \times 100$$

$$Accuracy(\%) = \frac{TP + TN}{TP + FN + TN + FP} \times 100$$

$$AUC(\%) = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \times 100$$

III. APPLICATION TO CANCER DATA

A review and discussion of the medical records of breast cancer patients is carried out as part of this research in more depth. Dr. William H. Wolberg of the University of Wisconsin Clinic in Madison, Wisconsin, was the driving force behind the creation of this database. With the use of full field advanced mammography technology, a total of 699 instances have been identified, 458 of which have been confirmed to be benign and 241 of which have been determined to be malignant. Clams have traits such as a thick layer of cell tissue, uniformity of cell size and shape, marginal adhesion between cells, the size of single epithelial cells, bare nuclei and proper nucleoli, and mitosis, among other things. It is impossible to distinguish between benign and malignant cells without considering these properties. Using the models' sensitivity, specificity, and accuracy, as well as the Area Under the ROC Curve, the efficiency of the models was assessed (AUC). In order to generate the SVM classification model, LIBSVM is used to run the classifiers in Matlab and to construct the classification model.

IV. RESULT AND DISCUSSION

4.1 Training stage of classifier

In Table 4.1, the results of the SVM classifier training experiments (which are also shown in Fig. 4.1) are presented. Gather a collection of data and split it into two groups: one for training and one for validation. This will serve as the starting point for the classification process. During our study, we employed cross-validation to gather 25 percent of the data, which was a considerable quantity of information. An example of a confusion matrix is presented below, which depicts the classification results of the four SVM classifiers during the training phase. As a result of these experiments, linear programming SVM, Lagrangian SVM, Proximal SVM, and SVM were found to be accurate in 97.33 percent of cases with three hundred and thirty three correct classifications, 95.61 percent accuracy with three hundred and twenty seven correct classifications, 96.18 percent accuracy with three hundred thirty-one correct classifications, and 97.71 percent accuracy with three hundred thirty-three correct classifications, respectively. Lagrangian SVM has the highest minimum accuracy, and SVM surpasses all other models across the board, including minimum accuracy. In terms of performance, Linear Programming SVM is a close second to the first. Individual AUC estimations attained an accuracy of 99.51 percent, outperforming all other classifiers examined in this investigation.

Table 4.1:Performance of classifiers during Testing on WBCD

Measuring Tools	SVM	Proximal SVM (PSVM)	Lagrangian SVM (LSVM)	Linear Programming SVM (LPSVM)
Accuracy (%)	97.71	96.18	95.61	97.33
Sensitivity(%)	97.08	95.39	95.61	96.52
Specificity (%)	98.90	97.74	95.60	98.88
AUC (%)	99.51	97.77	97.27	99.49

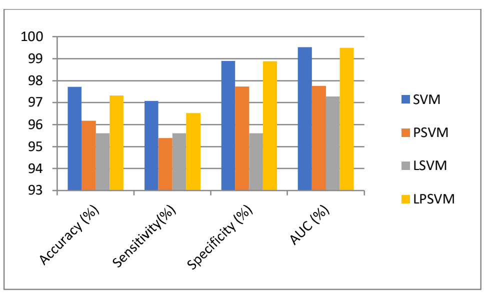

Fig. 4.1: Performance of classifiers during Testing on WBCD

4.2 Validation stage of classifier

The results of the demonstration examinations of the validation step using SVM classifiers are reported in Table 4.2 and illustrated visually in Fig. 4.2.

Table 4.2: Performance of classifiers during Validation on WBCD

Measuring Tools	SVM	Proximal SVM (PSVM)	Lagrangian SVM (LSVM)	Linear Programming SVM (LPSVM)
Accuracy (%)	94.86	96	95.43	97.14
Sensitivity (%)	95.65	97.37	96.52	98.25
Specificity (%)	93.33	93.44	93.33	95.08

AUC (%)	96.61	97.75	97.18	99.38

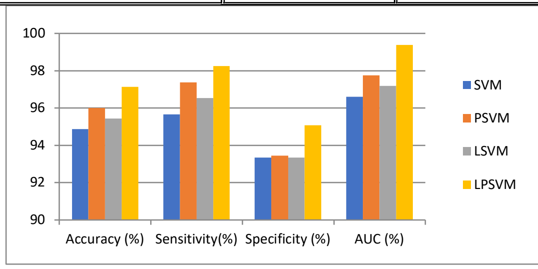

Fig. 4.2: Performance of classifiers during Validation on WBCD

SVM uses a four-fold cross-validation approach because it is accurate and can be easily implemented. Results of classification and execution for each SVM strategy are shown in Table 4.2 It tend to be seen from Tables 4.2 that the overall accuracies of Linear Programming SVM accomplished 97.14 % with correct classification 112, which is better than Lagrangian SVM (95.43 %), Proximal SVM (96 %), and SVM (94.86 %). The overall sensitivities of Linear Programming SVM accomplished 98.25 %, which is better than Lagrangian SVM (96.52 %), Proximal SVM (97.37 %) and SVM (95.65 %). The overall specificities of Linear Programming SVM accomplished (95.08 %), which is better than Lagrangian SVM (93.33 %), Proximal SVM (93.44 %) and SVM (93.33 %). Estimation of AUC for Linear Programming SVM achieved 99.38 %. After being trained, SVM outperformed all other techniques in terms of accuracy and efficiency. During the validation phase, it was able to achieve the greatest reduction in accuracy. It has been shown that SVM with Linear Programming has exceptional performance in the detection of breast cancer tumors.

V. **CONCLUSION**

It is proposed in this work that four linear SVM classifiers be employed to give incredibly quick, simple, and successful breast cancer diagnosis. A set of experiments on the WBCD informational index was carried out with the use of SVM classifiers to explore breast cancer. In comparison to other modified classifiers, such as Linear Programming SVM (LPSVM), Lagrangian SVM (LSL), and Proximal SVM, it performs much better (PSVM). Based on the training results, Lagrangian SVM has the lowest accuracy of 95.31%, while SVM was superior to other techniques in all exhibition lists (accuracy of 97.71%) and followed closely by Linear Programming SVM (accuracy of 97.33%). These findings indicated that selecting the most appropriate classifier and kernel function is crucial for attaining the best possible outcomes in classification. During the validation process, the overall accuracy of the models is checked for accuracy. Linear Programming SVM achieved 97.14 percent, which was higher than Lagrangian SVM (95.43 %), Proximal SVM (96%), and SVM using a random walk (94.86%). In terms of total sensitivity, Linear Programming SVM achieved 98.25%, which is higher than those achieved by Lagrangian SVM (96.52%), Proximal SVM (97.37%), and SVM (95.65%).

The overall specificities of Linear Programming SVM accomplished (95.08%), which is better than Lagrangian SVM (93.33%), Proximal SVM (93.44%) and SVM (93.33%). Estimation of AUC for Linear Programming SVM accomplished 99.38%, individually, which beat different classifiers. The results strongly support the use of Linear Programming SVM in breast cancer analysis. The physicians may utilize SVM classifier Linear Programming SVM for their formal conclusions on their patients. By using such a powerful model, they can make incredibly precise decisions.

REFERENCES

[1]. Azmi M. S., and Z. C. Cob (2010), Breast Cancer Prediction Based on Back propagation Algorithm.Student Conference on Research and Development (SCOReD). 164–168.
[2]. Blanz V, Scholkopf B, Bulthoff H et al (1996) Comparison of view–based object recognition algorithms using realistic 3d models. In: von der Malsburg C, von Seelen W, Vorbruggen JC,Sendhoff B (eds) Artificial Neural Networks—ICANN'96, Springer Lecture Notes in Computer Science, Berlin, vol 1112, pp 251–256.
[3]. Burges CJC, Scholkopf B (1997) Improving the accuracy and speed of support vector learning machines. In: Mozer M, Jordan M, Petsche T (eds) Advances in neural information processing systems 9. MIT Press, Cambridge, pp 375–381.
[4]. Chen HL, Yang B, Wang G et al (2011) Support vector machine based diagnostic system for breast cancer using swarm intelligence. J Med Syst. doi:10.1007/s10916-011-9723-0.
[5]. Cortes C, Vapnik V (1995) Support vector network. Mach Learn 20:273–297.
[6]. Fung G, Mangasarian OL (2004) A feature selection Newton method for support vector machine classification. Comput Optim Appl 28(2):185–202.
[7]. Fung.G, Mangasarian. O. L(2001),Proximal Support Vector Machine Classifiers, Proceedings of Knowledge Discovery and Data Mining, San Francisco.
[8]. Keyvanfard, M. A. Shoorehdeli, and M. Teshnehlab (2011), Feature selection and classification of breast cancer on dynamic magnetic resonance imaging using ANN and SVM, American Journal of Biomedical Engineering, vol. 1, pp. 20-25.
[9]. Liu HX, Zhang RS, Luan F et al (2003) Diagnosing breast cancer based on support vector machines. JChem Inf Comput Sci 43(3):900–907.
[10]. Mangasarian OL, Setiono R, Wolberg WH (1990) Pattern recognition via linear programming: theory and application to medical diagnosis. Proceedings of the workshop on large-scale.
[11]. Mangasarian OL, Musicant DR (2000) Lagrangian Support Vector Machine Classification. Tec. Report, Data Mining Institute, Computer Sciences Department, University of Wisconsin.
[12]. Osuna E, Freund R, Girosit F (1997) Training support vector machines: an application to face detection. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern
Recognition, Jun 17–19, pp 130–136.

[13]. Prathap, S., Vallinayagam, V. and Venkatesan, P. (2016), A Comparative Study of Artificial Neural Network Architectures and Proportional Hazard Model using Heat Attack Data, Journal of Chemical and Pharmaceutical Sciences, Vol.9, No.3, 2339-2343.ISSN- 0974-2115.

[14]. Prathap, S., Vallinayagam, V. and Venkatesan, P. (2018), A Comparative Study of Artificial Neural Network Architectures and Support Vector Machine using cancer Data, International Journal of Current Research and Modern Education, Vol.3, No.1, 63-67. ISSN-2455-5428.

[15] Prathap, S (2018), A Comparative Study of Cox Proportional Hazard model and Bayesian Cox Proportional hazard model using Cancer data, IAETSD Journal for Advanced Research in Applied Sciences, Vol.5, No.4, 821-826. ISSN- 2394-842.

[16]. Ren J. (2012), ANN vs. SVM: Which One Performs Better in Classification of MCCs in Mammogram Imaging. Knowledge-Based Systems. 26: 144–153.

[17]. Scholkopf B, Burges C, Vapnik V (1995) Extracting support data for a given task. In: Fayyad UM,Uthurusamy R (eds) Proceedings, first international conference on knowledge discovery & data mining. AAAI Press, Menlo Park.

[18]. Scholkopf B, Burges C, Vapnik V (1996) Incorporating invariances in support vector learning machines. In: von der Malsburg C, von Seelen W, Vorbruggen JC, Sendhoff B (eds) Artificialneural networks- ICANN'96, vol 1112. Springer Lecture Notes in Computer Science, Berlin, pp 47–52.

[19]. Subashini T. S., V. Ramalingam, and S. Palanivel (2009), Breast Mass Classification Based on Cytological Patterns using RBFNN and SVM. Expert Systems with Applications. 36: 5284–5290.

[20].UCI (2012) Machine learning repository. http://archive.ics.uci.edu/ml/index.html.

[21]. Vallinayagam, V., Prathap, S., and Venkatesan, P. (2014), Parametric Regression Models in the Analysis of Breast Cancer Survival Data, International Journal of Science and Technology, Vol. 3, No. 3, 163-166. ISSN 2049-7318.

[22]. Vallinayagam, V., Prathap, S., and Venkatesan, P. (2014), Non-Linear Regression Models for Heart Attack Data – An Emprical Comparison, Indian Journal of Applied Research, Vol.4, No. 6, 332-334. ISSN -2249-555X.

[23]. Vapnik VN (1995) The nature of statistical learning theory. Springer, New York.

[24]. Wolberg WH, Mangasarian OL (1990) Multi surface method of pattern separation for medical diagnosis applied to breast cytology. Proc Natl Acad Sci 87:9193–9196.