Educational Administration: Theory and Practice
2024, 30(5), 13025 -13032
ISSN: 2148-2403
https://kuey.net/

Research Article

Enhanced SVM-CRFE-GK integrating Optimized Chi-RFE Feature Selection and Greedy Kernel for COVID-19 Prediction

C. Dharmadevi[1*], Dr. S. Thaddeus[2]

[1*]Assistant Professor, Sacred Heart College (Autonomous), Tirupattur, udharmadevi@gmail.com
[2]Assistant Professor, Don Bosco College (Co-Ed), Yelagiri, principal@dbcyelagiri.edu.in

Citation: C. Dharmadevi et al. (2024), Enhanced SVM-CRFE-GK Integrating Optimized Chi-RFE Feature Selection And Greedy Kernel For Prediction. *Educational Administration: Theory and Practice*, *30*(5), 13025-13032
Doi: 10.53555/kuey.v30i5.3445

ARTICLE INFO

ABSTRACT

In the Pandemic Situation, Health Care data are accumulating at faster rate, but interpreting, predicting and classifying them is still a challenging one. Machine Learning Models has different algorithms to simplify the problem of interpretation, prediction and classification of data set either it is structured or unstructured. Mainly, for prediction, among many feature selection algorithms simple Support Vector Machine (SVM), Support Vector Machine with Recursive Feature Elimination (SVM-RFE) identifies dependent features and improves the prediction rate and accuracy. But the SVM, SVM-RFE models are not enough when the data size is very huge, unstructured and having high dimensions. To overcome the limitation, this research work propose an Enhanced SVM-CRFE-GK (SVM-Chi-Square-RFE with Greedy Kernel) model which coupled Chi-Square feature selection algorithm to identify the dependent features, RFE to remove the irrelevant features based on the weight vectors with reduced number of iterations without losing the accuracy and uses Greedy Kernel method to optimize performance and to improve the classification accuracy in less time consumption and memory utilization.

***Keywords*:** *Machine Learning Models, Support Vector Machine (SVM), Feature Selection Algorithms, Recursive Feature Elimination (RFE), Chi-Square Feature Selection, Greedy Kernel.*

I.Introduction

In recent years, SVM has been widely used in diverse fields such as healthcare, epidemiology and bioinformatics [11] owing to its ability to handle high-dimensional data and nonlinear relationships effectively [10]. Despite its effectiveness, SVM's performance heavily relies on the selection of relevant features from the input data. Feature selection plays a crucial role in improving model interpretability, reducing computational complexity and enhancing prediction accuracy [16]. Various feature selection techniques [4, 5, 6] have been proposed among that Recursive Feature Elimination (RFE) has been extensively employed to iteratively remove irrelevant features to enhance classification task [2, 9, 16].
In this context, this research proposes SVM-CRFE-GK an optimized Support Vector Machine (SVM) that integrates two feature selection techniques, Chi-Square and Recursive Feature Elimination (Chi-RFE) with Greedy Kernel Prediction Strategy. The combination of these techniques aims to enhance the SVM model's performance by selecting the most discriminative features using weight vectors and improving prediction accuracy in reduced amount of time and memory. Building upon the work of previous studies [1, 2, 9], this research seeks to explore the effectiveness of Chi-RFE feature selection in conjunction with a greedy kernel approach to address classification challenges in predicting the COVID existence. Through empirical analysis and experimentation, the proposed algorithm efficiency in handling huge dataset and improving prediction accuracy will be evaluated and compared against existing methodologies.

This paper is organized as follows, Section 2 presents the review of research works carried out using Enhanced SVM models to understand the features and limitations in the different approaches used. In Section 3, the proposed SVM-CRFE-GK algorithm, an enhanced Chi-square-RFE feature selection method

with greedy kernel is described in detail. Finally, the computational experiment to evaluate the efficiency of our approach is presented and discussed in Section 4.

II.Literature Review

Support vector machine (SVM) [1, 7] is an efficient and popular classification technique and has been widely applied in many fields such as biological data processing [19]. SVM-recursive feature elimination (SVM-RFE) [8, 16] is a feature selection algorithm based on SVM. While the SVM learning model is built, the weights of the features are also computed. SVM-RFE iteratively removes the features with the lowest weights. The removing sequence of the features represents the feature importance ranking [11, 16]. SVM-RFE has been adapted in many applications, such as signal processing [22], genomics [20, 21], proteomics [23] and metabolomics [24, 25], due to its superiority.
Huang et al. [2] pioneered an SVM-RFE-based approach for feature selection combined with Taguchi parameter optimization, tailored specifically for multiclass SVM classification tasks. Alcaraz et al. [3] proposed a multi-objective approach to SVM feature selection, striving to strike equilibrium between diverse classification objectives for medical applications accentuate the significance of SVMs in disease diagnosis and prognosis. For instance, Behnood et al [8] uses RFE to identify optimal features for brain tumor classification and Theerthagiri and Siddalingaiah [9] devised a recursive Gaussian SVM-based feature selection algorithm for liver disease classification, demonstrating the relevance of sophisticated machine learning techniques in medical diagnostics. Hector Sans [10] in his research work identifies and visualizes the most relevant features using SVM-RFE through non-linear kernels. Similarly, Alshanbari et al. [14] harnessed a weighted radial kernel SVM coupled with recursive feature elimination to predict and classify COVID-19 admissions to intensive care units, showcasing the adaptability of SVM based methodologies in addressing healthcare complexities.

Tang et al. [26] proposed a two-stage SVM-RFE. In the first stage, multiple SVM-RFEs with different parameters were applied to remove the noise and non-informative data; in the second stage, the final feature subset was selected by a fine SVM-RFE. Li et al. [27] combined SVM-RFE with the T-statistic to define the genes associated with CRC development or metastasis. mRMR-SVM [28] tries to select an important and non-redundant feature subset by means of SVM-RFE and mRMR. R-SVM [29] is also a recursive feature selection method based on SVM, which combines SVM weights and class means to evaluate feature discriminative abilities. There are also some studies on determining how many features with the low weights are removed in each iteration of SVM-RFE [30, 31, 32].
Besides, from the research review, a comparative analysis on performance of SVM with different feature selection techniques using SVM classification model is given in Table. I.

Author	Feature Selection used	Limitation
[2] Mei-Ling	RFE (Recursive Feature Elimination) and Taguchi parameters	• Limited to multiclass SVM classifier. • Might not be applicable to other classification problems.
[3] Javier Alcaraz	Incorporates feature selection as a Multiobjective optimization problem	• Complexity may increase due to multiobjective optimization. • Effectiveness depends on the choice of objectives and their weights.
[4] J. Vijayashree	Integration of PSO for feature selection	• Performance heavily depends on the quality of features and parameters of PSO.
[5] Sahaya Sheela	Hybrid approach combining PSO with SVM	• Performance relies on the quality and quantity of training data. • May require ample parameter tuning for optimal performance.
[6] Vamshidhar	Combines deep learning (ResNet) and SVM	• Hybrid models might introduce additional complexity in deployment and interpretation.
[7] Dana Bazazeh	Comparison of ML models	• Limited to breast cancer detection performance vary based on dataset and feature selection techniques.
[8] Behnood	RFE	• Limited to brain tumor classification effectiveness may vary based on the quality and quantity of the image.
[15] Xiaojuan	Integrates clustering and RFE.	• Performance depends on the quality of clustering and RFE. • Limited to gene selection tasks, might not be directly applicable to other domains.
[10] Hector Sanz	RFE with non-linear kernels	• Limited discussion on the effectiveness of non-linear kernels compared to linear kernels.
[11] Xiaohui Lin	RFE with the overlapping ratio to identify feature subsets	• Might not be directly applicable to other domains than bioinformatics. • Effectiveness of the overlapping ratio in feature selection may depend on the dataset.

Table. I. A Comparative Analysis on Performance of SVM with different Feature Selection Techniques

The analysis reveals that many research works prefer RFE as a feature selection technique. Though using RFE with SVM is good in predication and classification, the limitation column exposes still lacking in

performance for huge data set with different characteristics and hence this feature selection technique has to be optimized.

III.Proposed Work

The reviews, and in general, SVM-RFE, rank the features according to the feature deletion order during the iterations. It is subjective to specify the number of features to be selected in advance in applications. Consequently, due to this specific entry of feature count, there are chances of irrelevant features existing or relevant features being missed in the feature subset selected with the specified number of features. This also affects the prediction rate and accuracy level of the classifier. Hence, this study proposes an algorithm, SVM-CRFE-GK, an enhanced SVM Chi Square-Recursive Feature Elimination with Greedy Kernel for prediction.
In this SVM-CRFE-GK algorithm, the SVM-CRFE feature selection technique first identifies the dependent features using chi-square. The feature subset selected from chi-square applies optimized RFE technique, where the optimal features, Fs, are iteratively selected by calculating the weight vectors |w| of current feature subsets, finding the rank, and removing the feature with the minimum rank. The feature elimination iteration resumes until the cross-validation accuracy is maximum. Additionally, this research work contributes SVM-CRFE-GK algorithm to improve the accuracy rate of the prediction.
As the efficiency of the SVM model highly varies based on the choice of the kernel [2, 9], especially Gaussian, greedy kernel shows better accuracy [10, 14] for datasets with multi dimensions and huge size. Using the optimal features, sFs, the algorithm SVM-CRFE-GK identifies the Greedy kernel hyper parameters C and gamma to improve the accuracy of prediction. The flow diagram for the proposed algorithm is given in Fig.1. The flow starts with the input COVID-19 dataset, X, with n instances and m features, and y, the target variable. After preprocessing the data for the removal of noise and outliers, it is split into training (*X_train*) and testing (*X_test*) datasets.

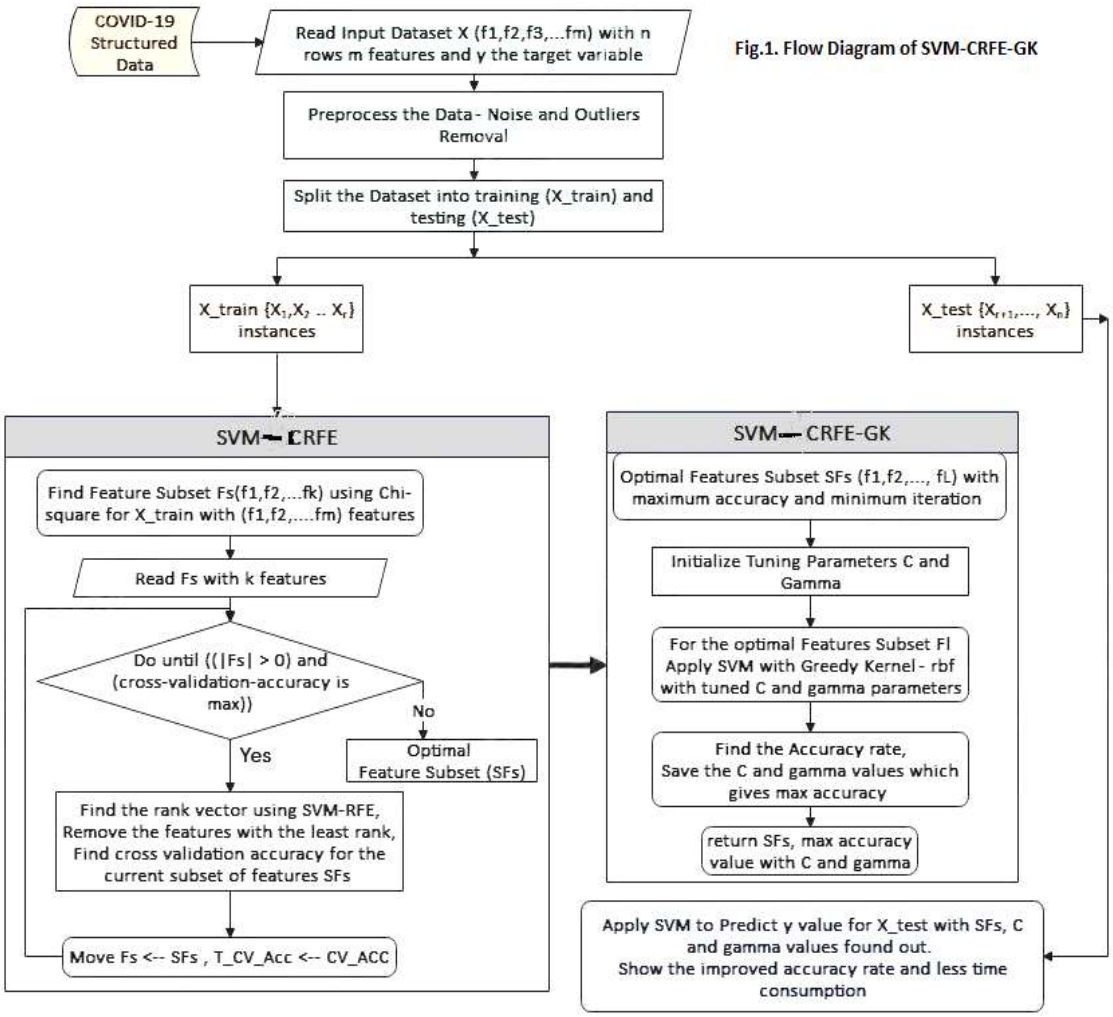


Fig.1. Flow Diagram of SVM-CRFE-GK

From the *X_train* dataset with all the m input features, the optimal features (*Fs*) are filtered using Algorithm-1-SVM-CRFE. Furthermore, the SVM-CRFE-GK trains the data with *SFs* features and identifies the fine hyper-tuning parameters that produce high prediction accuracy, as mentioned in Algorithm-2. The

X_test dataset with the resulting feature subset *SFs* and hyper parameters C, *gamma*, are used for the prediction of COVID-19 with improved accuracy rate in less time and memory.

A. *SVM-CRFE algorithm for Optimal Feature* Selection

The SVM-CRFE algorithm is provided in ALGORITHM-1. The *X_train (f1, f2, f3,…., fm)* dataset contains $\sum_{i=1}^{i=r} Xi$, COVID data, where r represents number of instances and m represents number of input features. Initially, the Chi-square values for the m features are calculated using the following equation:

$$\chi 2 = \sum \frac{(Ob_i - Ex_i)^2}{Ex_i}$$

where *Ob* – observed values, *Ex* – expected values
The features with high Chi value are selected and named as *Fs {f1,f2,f3,…. Fk}*.

ALGORITHM-1: SVM-CRFE

Input : Training dataset X_train : ${}_1^r Xi \subseteq {}_1^n X \{f1, f2, f3, \ldots . fm\}$ and d
Output: Selected Feature Subset *Fs* with *{f1,f2,f3,….,fl}* features

```
function SVM-CRFE(X_train, d)
{
Fs { f1, f2, f3, …. fk}= Chi-Square(r1Xi {f1, f2, f3, …. fm}
SFS = Fs;
CV_Acc = 0;
Flag = 'Y'
i = 1
Fs = {k input features selected from chi-square};
while (|Fs| > 0) do {
Construct an SVM based on X_train and Fs;
Find T_CV_Acc = d-fold cross validation accuracy rate of SVM;
Find |w| - absolute weight vector of features;
Rank the features in Fs by |w| in descending order;
if ((T_CV_Acc  > CV_Acc) and (flag ='Y') Then
        {    CV_Acc = T_CV_Acc;
             SFS = Fs;
             Print iteration-i, T_CV_Acc, Fs;
   Fs = Fs– {s × |Fs| - remove bottom ranked features in Fs;
             Increment i by 1;
      } else
        {        SFS = Fs;
                 Flag = 'N';
        }
     endif;
} end while;
return SFS – feature set with l features;
}
s (0 < s < 100%) is the filter factor. In each iteration of SVM-CRFE,
(s × |F|) bottom ranked features are removed from the current feature subset F.
```

The Feature subset *Fs* is then given as input for enhanced RFE feature selection. The mathematical formula to calculate the weight vector $|\mathbf{w}|$ in an SVM model is given by:

$|w| = \sum_{i=1}^{i=n} \alpha_i y_i X_i$

Where,

- α_i lagrange multiplier corresponding to the i-th support vector
- y_i class label of the i-th support vector
- X_i class label of the i-th support vector

The iteration starts with *X_train* data contains k features (*Fs)*. In each iteration of enhanced RFE, the d-fold cross-validation accuracy and weight vector (|w|) of each feature in the current feature subset *Fs* are calculated according to the support vectors on the hyper-plane of the SVM classifier. If the accuracy rate (T_CV_Acc) is greater than the previous accuracy (CV_Acc), the features are ranked based on |w|, and the bottom ranked features are removed from *Fs*. The iteration continues until the accuracy rate (T_CV_Acc) is maximized. The procedure terminates with the optimal feature subset *SFs* with {f1,f2,…., fl}features which retains maximal accuracy rate.

B. *SVM-CRFE-GK Algorithm*

ALGORITHM-2: SVM-CRFE-GK

Input : Training dataset X_train : ${}_1^r Xi \subseteq {}_1^n X \{f1, f2, f3, \ldots . fl\}$ and
Test dataset X_test : ${}_{r+1}^n Xi \subseteq {}_1^n X \{f1, f2, f3, \ldots . fl\}$
Output: C and gamma tuning parameters best suited for SVM
using X-dataset with Fs having l features

```
Begin
function SVM-CRFE-GK(X_train)
{     Choose Iteration_Parameters :
            C – any positive value range from 1 to positive integer;
            gamma – any positive value range from 0 to 1;
Do for all combination of C, gamma parameters
    svm = SVC(C=C, gamma=gamma, kernel='rbf');
    Compare and store the C, gamma values which gives the maximum
        cross-validation accuracy;
End do;
     Return the selected C, gamma values;}end
```

The SVM-CRFE-GK algorithm, as shown in ALGORITHM-2, utilizes the greedy kernel to enhance the performance of the prediction model. The input data, *X_train*, consists of an optimal subset of *L* features, *SFs*, selected using ALGORITHM-1. The hyper-tuning parameters for the greedy kernel, *C* and *gamma*, accept positive values. When *C* is small, the margin is large, and all vectors become support vectors. Conversely, with a high positive value of C, the margin becomes small, reducing the number of support vectors and enhancing the efficiency of the training model.
Hence, the preferred values for *C* are {1, 10, 100, 1000}, and for *gamma*, {0, 0.001, 0.01, 0.1, 1}. For each combination of *C* and *gamma* for the RBF kernel, the accuracy value is determined. In each combination of the hyper-tuning parameters, the accuracy value is compared, and the C and gamma values that provide the maximum accuracy are identified.

IV.Results and Discussion

To experiment and evaluate the proposed algorithm SVM-CRFE-GK, we utilized the COVID-19 dataset comprising 5435 instances with 21 features, including 20 input features and a target variable. The dataset was split into *X_train* and *X_test*, which were then passed to the SVM-CRFE-GK algorithm.
Chi-values were computed for all features, and those with chi scores above 0, specifically 13 features, were selected and stored in *Fs*, as depicted in Fig. 2.

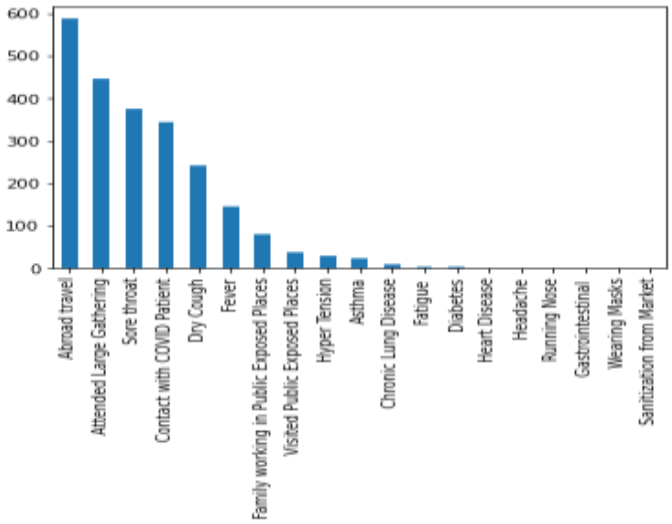


Fig. 2. Chi-values of all the m Features

Features were iteratively eliminated, and the optimal features selected, *SFs*, at the end of the enhanced SVM-CRFE technique are detailed in Fig. 3.

```
Input Feature Index - [0 1 2 3 5 6 10 11 13 14 15 16 17]
Iteration- 1
0.9664136185875316
11
[0, 1, 2, 3, 5, 6, 10, 13, 14, 15, 17]
Iteration- 2
0.9664136185875316
9
[0, 1, 2, 3, 5, 13, 14, 15, 17]
Iteration- 3
0.9664136185875316
8
```

[0, 1, 2, 3, 5, 14, 15, 17]
Iteration- 4
0.9484702093397744
Selected Feature Subset(SFS) : [0, 1, 2, 3, 5, 14, 15, 17]

Fig.3. Accuracy, Count and List of Feature Subset iteratively selected in enhanced SVM-CRFE

Parameters: C=0.1, gamma=1, Cross-Validation Accuracy=0.9326008227186753
Parameters: C=1, gamma=1, Cross-Validation Accuracy=0.9500818750578681
Parameters: C=10, gamma=1, Cross-Validation Accuracy=0.9498517254606298
Parameters: C=100, gamma=1, Cross-Validation Accuracy=0.9498517254606298
Parameters: C=100, gamma=0.1, Cross-Validation Accuracy=0.9473214025898443
Parameters: C=1000, gamma=1, Cross-Validation Accuracy=0.9498517254606298
Parameters: C=1000, gamma=0.1,Cross-Validation Accuracy=0.9496215758633916

Fig.4. A Sample of Cross-Validation Accuracy for combination of C and gamma values using enhanced SVM-CRFE-GK

Subsequently, the C and gamma values were predicted using ALGORITHM-2: SVM-CRFE-GK, as illustrated in Fig. 4, with the best parameters highlighted.
The optimal features, selected SFs, and C and gamma parameters are substituted into X_*test* and y values to predict the existence of COVID. The results of the SVM models with existing feature techniques and the proposed enhanced SVM-CRFE-GK are presented in Table 2 and Table 3. Additionally, graphical representations comparing accuracy, time, memory, and classification results are shown in Fig. 5, Fig. 6, Fig. 7, and Fig. 8.

Method	Time in ms	Accuracy	Memory Utilization (in bytes)
SVM	**144.65**	**95.21**	**390265, 1640973**
SVM- chi Square	**171.33**	**95.54**	**352822, 1184325**
SVM – RFE	**230.21**	**96**	**726432, 1613005**
SVM – C-RFE	**316**	**98.16**	**214013, 710229**
SVM-CRFE-GK	**108.12**	**98.22**	**11267, 141065**

TABLE. 2. Classification Comparison of Accuracy, Time and Memory

Fig.5 Accuracy of Existing SVM methods and Proposed SVM-CRFE-GK

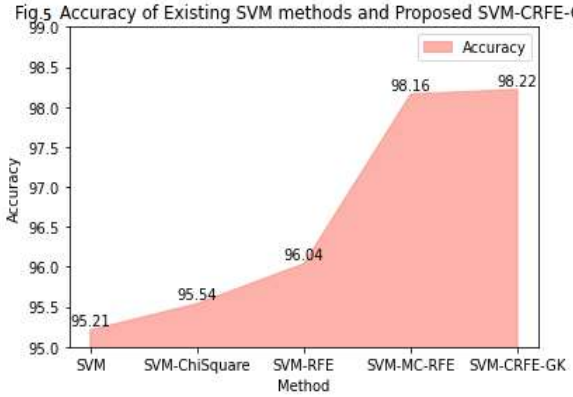
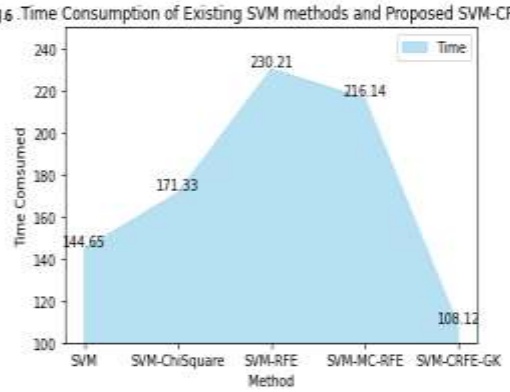


Fig.6 .Time Consumption of Existing SVM methods and Proposed SVM-CRFE-GK



Method	Accuracy	Precision	Recall	f1-score
SVM	**95.21**	**98**	**96**	**97**
SVM- chi Square	**95.54**	**96**	**99**	**98**
SVM – RFE	**96**	**98**	**98**	**99**
SVM – C-RFE	**97.16**	**97**	**99**	**98**
SVM-CRFE-GK	**98.22**	**99**	**99**	**99**

TABLE. 3. Classification comparison of Accuracy, Precision, Recall and f1-score

Fig.7 Memory Utilization of Existing SVM methods and Proposed SVM-CRFE-GK

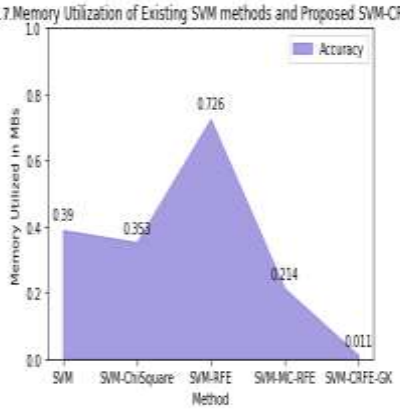


Fig.8 Comparison of Classification Results of Existing SVM mehtods and Proposed SVM-CRFE-GK

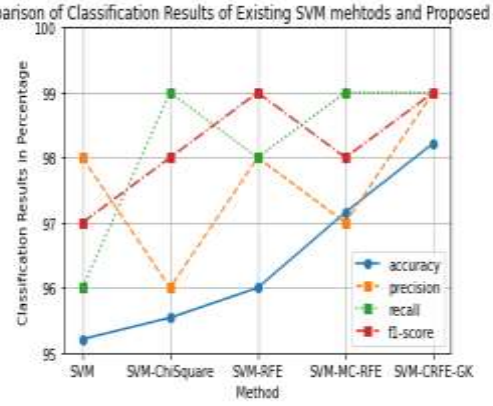


Fig. 2 identifies 13 features selected as dependent factors for predicting COVID. In Fig. 3, the elimination of features at each level is depicted, comparing cross-validation accuracy and identifying 8 optimal features with a maximum cross-validation accuracy of 96.64%. This figure illustrates that instead of 20 features, these 8 optimal features are appropriate for predicting COVID, maintaining maximum accuracy. Additionally, Fig. 4 highlights C and gamma values of 1, 1 with a high accuracy of 95.01%, which best suits this dataset and enhances the accuracy rate of the prediction model.

Table 2 provides insights into the computational aspects, highlighting significant increase in the accuracy level (98.22%) and reductions in both time consumption and memory utilization of proposed enhanced SVM when compared to SVM models with existing feature techniques, Furthermore, Table 3 illustrates the comparative performance in terms of accuracy, precision, recall, and F1-score, showcasing notable improvements achieved by the SVM-CRFE-GK approach.

V.Conclusion

The study evaluates the effectiveness of enhanced SVM-CRFE-GK, integrating Optimized Chi-RFE feature selection and Greedy Kernel in enhancing prediction accuracy, computational efficiency, and memory utilization. These findings underscore the efficacy of our approach in achieving superior predictive performance while minimizing computational overhead, thereby offering a promising solution for real-world applications in various domains. The combination of feature selection techniques with a greedy kernel not only improves predictive accuracy but also streamlines computational resources, making it well-suited for large-scale datasets and resource-constrained environments. Additionally, the robustness of our method suggests its potential for deployment across diverse healthcare domains, such as diabetes management and breast cancer diagnosis, where accurate predictions and efficient resource utilization are paramount.

REFERENCES

[1] C. Dharmadevi and S. Thaddeus, “Prediction of COVID-19 Infections using Classification Algorithms in Machine Learning”, *Indian Journal of Natural Sciences, 2023, Vol.14.*

[2] Mei-Ling Huang, Yung-Hsiang Hung, W. M. Lee, R. K. Li and Bo-Ru Jiang, “SVM-RFE Based Feature Selection and Taguchi Parameters Optimization for Multiclass SVM Classifier”, *The Scientific World Journal, Hindawi Publicaions, 2014, Vol.2014, 10 pages.*

[3] Javier Alcaraz, Martine Labbé and Mercedes Landete, “Support Vector Machine with feature selection: A multiobjective approach”, *Expert Systems with Applications, Elsevier, 2022, Vol.204.*

[4] J. Vijayashree & H. Parveen Sultana, “A Machine Learning Framework for Feature Selection in Heart Disease Classification Using Improved Particle Swarm Optimization with Support Vector Machine Classifier”, *Programming in Computer Software, Springer Link, 2019, Vol.44, pages 388-397.*

[5] M. Sahaya Sheela1, C. A. Arun, “Hybrid PSO–SVM algorithm for Covid-19 screening and quantification”, *International Journal of Information Technology, Springer Link, 2022, Vol.14, pages 2049-2056.*

[6] Vamsidhar Enireddy, Mathe John Kenny Kumar, Babitha Donepudi and C Karthikeyan, “Detection of COVID-19 using Hybrid ResNet and SVM”, *IOP Conference Series: Materials Science and Engineering, IOPScience,2020.*

[7] Dana Bazazeh and Raed Shubair, “Comparative Study of Machine Learning Algorithms for Breast Cancer Detection and Diagnosis”, *The 2016 IEEE 5th International Conference on Electronic Devices, Systems, and Applications (ICEDSA’2016), 2016.*

[8] Behnood Gholami, saiah Norton, Allen R. Tannenbaum, and Nathalie Y. R. Agar, “Recursive Feature Elimination for Brain Tumor Classification using Desorption Electrospray Ionization Mass Spectrometry Imaging”, *Conf Proc IEEE Eng Med Biol Soc, 2012, PMC, pages 5258-5261.*

[9] Prasannavenkatesan Theerthagiri & Sahana Devarayapattana Siddalingaiah, "Recursive Gaussian support vector machine based feature selection algorithm for liver disease classification", *Multimedia Tools and Applications, Springer Link, 2023.*
[10] Hector Sanz , Clarissa Valim, Esteban Vegas , Josep M. Oller and Ferran Reverter, "SVM-RFE: selection and visualization of the most relevant features through non-linear kernels", *Bioinformatics, BMC, 201.*
[11] Xiaohui Lin, Chao Li, Yanhui Zhang, Benzhe Su, Meng Fan and Hai Wei, "Selecting Feature Subsets Based on SVM-RFE and the Overlapping Ratio with Applications in Bioinformatics", *Molecules, MDPI, 2017*
[12] Zhixian Yao, Xinyi Zheng, Zhong Zheng, Ke Wu1, Junhua Zheng, "Construction and validation of a machine learning-based nomogram: A tool to predict the risk of getting severe coronavirus disease 2019 (COVID-19)", *Immunity, Inflammation and Disease, Wiley, 2021.*
[13] Huda M. Alshanbari, Tahir Mehmood, Waqas Sami, Wael Alturaiki, Mauawia A. Hamza and Bandar Alosaimi, "Prediction and Classification of COVID-19 Admissions to Intensive Care Units (ICU) Using Weighted Radial Kernel SVM Coupled with Recursive Feature Elimination (RFE)", *Life, MDPI, 2022.*
[14] Robinson Joel M , Manikandan G, Bhuvaneswari G, Shanthakumar P, "SVM-RFE enabled feature selection with DMN based centroid update model for incremental data clustering using COVID-19, *Computer Methods in Biomechanics and Biomedical Engineering,2023.*
[15] Xiaojuan Huang, Li Zhang, Bangjun Wang, Fanzhang Li & Zhao Zhang, "Future clustering based support vector machine recursive feature elimination for gene selection, Applied Intelligence, *Springer Link, 2017, Vol.48, pages 594-607.*
[16] Xiaojian Ding, Fan Yang, Fuming Ma, "An efficient model selection for linear discriminant function-based recursive feature elimination", *Journal of Biomedical Informatics, Elsevier, 2022, Vol.129.*
[17] Ebrahime Mohammed Senan, Mosleh Hmoud Al-Adhaileh, "Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques", *Journal of Healthcare Engineering, Hindawi Publications, 2021, Vol.2021.*
[18] Viktor Wottschela, Declan T. Chard, Christian Enzingerd , Massimo Filippie, "SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of multiple sclerosis", *NeuroImage: Clinical, Elsevier, 2019, Vol.24.*
[19] Butkiewicz, M, Lowe E, Mueller R, Mendenhall J, Teixeira P, Weaver C, Meiler J, "Benchmarking ligand-based virtual high-throughput screening with the pubchem database", *Molecules 2013, 18, 735–756. [CrossRef] [PubMed]*
[20] Guyon I, Weston J, Barnhill S, Vapnik V, "Gene selection for cancer classification using support vector machines", *Machine Learning, 2002, 46, 389–422. [CrossRef]*
[21] Duan K.B, Rajapakse J.C, Wang H, Azuaje F, "Multiple SVM-RFE for gene selection in cancer classification with expression data", *IEEE Trans. Nanobiosci. 2005, 4, 228–234. [CrossRef].*
[22] Hidalgo-Muñoz A.R, López M.M, Pereira A.T, Tomé A, "Spectral turbulence measuring as feature extraction method from EEG on affective computing", *Biomed. Signal Process. Control 2013, 8, 945–950. [CrossRef]*
[23] Dao F.Y, Yang H, Su Z.D, Yang W.R.T, Wu Y, Ding H, Chen W, Tang H, Lin H, "Recent advances in conotoxin classification by using machine learning methods", *Molecules 2017, 22, 1057. [CrossRef] [PubMed].*
[24] Mahadevan S, Shah S.L, Marrie T.J, Slupsky C.M, "Analysis of metabolomic data using support vector machines", *Anal. Chem. 2008, 80, 7562–7570. [CrossRef] [PubMed].*
[25] Lin X.H, Yang F.F, Zhou L.N, Yin P.Y, Kong H.W, Xing L, Jia, L, Wang Q.C, Xu G.W, "A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information", *Journal of Chromatography. B 2012, 910, 149–155. [CrossRef] [PubMed]*
[26] Tang Y, Zhang Y.Q, Huang Z, "Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis", *IEEE/ACM Trans. Comput. Biol. Bioinform. 2007, 4, 365–381. [CrossRef] [PubMed]*
[27] Li X.B, Peng S.H, Chen J, Lü B.J, Zhang H.H, Lai M.D, "SVM-T-RFE: A novel gene selection algorithm for identifying metastasis-related genes in colorectal cancer using gene expression profiles", *Biochem. Biophys. Res. Commun. 2012, 419, 148–153. [CrossRef] [PubMed].*
[28] Mundra P.A, Rajapakse J.C, "SVM-RFE with MRMR filter for gene selection", *IEEE Trans. Nanobiosci. 2010, 9, 31–37. [CrossRef] [PubMed].*
[29] Zhang X.G, Lu X, Shi Q, Xu X.Q, Hon-Chiu E.L, Harris L.N, Iglehart J.D, Miron A, Liu J.S, Wong W.H, "Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data", *BMC Bioinform. 2006, 7, 1–13. [CrossRef]*
[30] Bolón-Canedo V, Sánchez-Marono N, Alonso-Betanzos A, Benítez J.M, Herrera F, "A review of microarray datasets and applied feature selection methods", *Inf. Sci. 2014, 282, 111–135. [CrossRef]*
[31] Ding Y, Wilkins D, "Improving the performance of SVM-RFE to select genes in microarray data", *BMC Bioinform. 2006, 7, S12. [CrossRef] [PubMed]*
[32] Lin X, Song H, Fan M, Ren W, Li L, Yao W, "The feature selection algorithm based on feature overlapping and group overlapping", *In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, Shenzhen, China, 15–18 December 2017; pp. 619–624. [CrossRef.*