International Journal of Safety and Security Engineering
Vol. 14, No. 6, December, 2024, pp. 1679-1687
Journal homepage: http://iieta.org/journals/ijsse

Comparative Analysis of SVM Classifiers in Criminal Profiling Using a Hybridized Algorithm

Adeleke J. Adeyiga[1], Adedayo F. Adedotun[2*], Oluwatosin E. Adebisi[1], Olasumbo O. Agboola[2]

[1] Department of Computer Science and Information Technology, Bells University of Technology, Ota 112102, Nigeria
[2] Department of Industrial Mathematics, Covenant University, Ota 112101, Nigeria

Corresponding Author Email: adedayo.adedotun@covenantuniversity.edu.ng

https://doi.org/10.18280/ijsse.140603

Received: 29 April 2024
Revised: 17 July 2024
Accepted: 28 October 2024
Available online: 31 December 2024

Keywords:
criminal, profiling, support vector machine, Fuzzy C-Means, classifier, cluster

ABSTRACT

The difficult jobs of investigating illegal activity and identifying offenses have fallen to the Law Enforcement Agencies (LEA). To help the LEA solve crimes, a number of criminal profiling systems have been created; however, the methods used in most systems do not allow for the clustering of criminals according to their behavioral traits. In order to get the best SVM classifier with the best kernel function that best fits our crime data set, this study hybridized the Fuzzy C-Means technique with the support vector machine algorithm in criminal profiling. To reduce the intra-cluster variances, the Fuzzy C-Means (FCM) method was altered by hybridizing it with Support Vector Machine (SVM). This was accomplished by substituting the SVM inner-product distance norm for the Euclidean distance used in the current FCM method to calculate the similarity and dissimilarity measure. Next, additional characteristics were added to the data along with the hybridized algorithm. MATLAB scripts were used to implement the developed strategies. The following criteria assessed the performance: execution time, sensitivity, precision, accuracy, and specificity. The result shows that the RBF kernel function performed best with both OAA and OAO classifiers. The OAA classifier performed best with 93.53% Specificity, 96.89% Precision, and 95.44% Accuracy over OAO and the pairwise classifier (BSVM). Therefore, the RBF kernel function using the OAA classifier is recommended to best suit our crime data set for criminal profiling, contributing to Sustainable Development Goal (SDG) 16.

1. INTRODUCTION

Clustering is a typical way of analyzing criminals by way of association and the behavioral nature of humans. Clustering algorithms are then expected to profile criminals based on association and relationship. One of the main shortcomings of criminal profiling is its inability to cluster criminals based on their behavioral features appropriately. The Euclidean distance function's limitation is that it can only measure data sets with a Euclidean form and without noise [1]. Being fully aware that the Fuzzy C-Means (FCM) algorithm uses the Euclidean distance for measurement, the more advanced FCM, or Kernelized Fuzzy C-Means (KFCM), also has limitations, including the inability to support data from multiple sources and the difficulty in choosing the best kernel for the given situation. The kernel function in use must also adhere to the learning objectives to get relevant results. Three classifiers make up the kernel function, a variation on the Support vector machine (SVM): the pairwise classifier (BSVM), the one against all (OAA) classifier, and the multiclass ranking SVM, also known as the one against one (OAO) classifier. Also, each classifier has a different kernel function. The linear, polynomial, radial basis function, and the sigmoid kernel functions.

Originally designed for binary classification, support vector machines may now be utilized for multiclass classification. To accommodate nonlinear classification issues where a maximum separation hyper plane is generated, the SVM algorithm maps input space to a higher-dimensional space. The largest margin of the hyperplane, a linear pattern, indicates the greatest distance between the decision classes.

Crimes have depressed trade and weaken investors' confidence in the economy, and to that extent it is a clear danger to the national security and the prosperity of citizens [2]. This work thus suggested the hybridization of the support vector machine (SVM) with the FCM method to calculate the distance between the data point and the cluster center in order to produce a better clustering result, based on the above-identified difficulty. A novel approach of profiling criminal was developed that would help the Law Enforcement Agencies to improving decision making in the various law enforcing agencies, reduce process time of crime analysis to enable quick completion of a crime investigation, and also to reduce the difficulties in managing the large volume of data involved in criminal activities, in a situation where a crime is committed in the absence of witnesses or any clue for forensics analysis by an expert from the crime scene.

This aligns with SDG 16's objective of promoting peaceful and inclusive societies for sustainable development, as it seeks to improve law enforcement methodologies and strengthen justice institutions (United Nations, 2015). Finding the best SVM classifier with the best kernel function is the main goal of this article.

2. REVIEW OF RELATED WORKS

This research work used Fuzzy C-Means algorithm to profile criminals and the algorithm used was modified by hybridizing it with a Support Vector Machine. The Euclidean distance function used in the existing FCM algorithm to measure the distance between the data point and cluster center. The limitation in using the Euclidean distance is that it measures only noise free data and Euclidean Shaped data set [1]. Also, the existing Kernelized Fuzzy C-Means (KFCM) has the limitation of not flexible enough to support data from different sources and the challenges of selecting the optimal kernel for the problem at hand. The kernel function in use must conform to the learning objectives in order to obtain meaningful results. Therefore, the SVM inner product distance norm was introduced in this research to measure the distance between the data point and the cluster center to provide a better clustering result.

After a training phase, support vector machines are an effective tool for binary classification because they can provide classifier functions extremely quickly. When using SVMs to classify issues involving three or more classes, there are many methods to consider. To tackle classification issues in the field of crime analysis, for instance, Rao et al. [3] offered a unique method of categorizing different crimes based on temporal and temporal physical variables. The study provides a way for users to employ a support vector machine (SVM)-based cybercrime classifier to carry out an easy and efficient classification conclusion. To create a model to prepare over a preparation set and provide the most precise results, the effort included categorizing the dataset using either random forests or decision trees. It is a low-key and effective way to categorize cybercrimes so that those impacted may determine the type of incident and take appropriate action. Convicted offenders are categorized by the model's design into three risk categories: low, medium, and high. This promotes the welfare and well-being of the community's residents by reducing the rising crime rates in the area. Additionally, three data mining approaches were examined against test crimes and criminal databases in the previous research [4], which employed the SVM algorithm for crime detection and forecasting. The highest-performing algorithm was then deployed against test crimes and criminal databases to identify prospective suspects in the crime.

Using the traditional machine learning approach of SVM Classification, Krysovatyy et al. [5] was able to create an algorithm that identifies fraudulent enterprises; a unified software environment was created for the quick identification of economic crimes. Data from 1,100 businesses that operate in Ukraine were utilized to construct the approach. 355 fake firms provide the data that are displayed in the set of logical binary values. Three approaches were used to model the SVM: radial, polynomial, and linear. The polynomial technique to classification yields the best results. Evaluation results were 100% for the training sample and 99.7% for the test sample.

The confusion matrix also produced some excellent findings. Support vector machines (SVM) and neural networks were employed in the previous study [6], for each data set, the SVM model had the greatest effectiveness among the classifiers. While He and Liu [7] offered insight into the usage of stochastic gradient descent algorithms for big data applications, such as boosting the efficacy of online learning or real-time forecasting (control) or speeding up SVM or controlled regression on a huge scale. Kim et al. [8] suggested using Twitter postings and vector-based filtering to remove noise in a machine learning method to crime detection and localization.

Okonkwo and Enem [9] employed a variety of data mining techniques to fight crime and terrorism in Nigeria. They also looked at the limitations of data mining in preventing crime in Nigeria and how law enforcement agencies can use it to track terrorists' criminal activity. Zulfadhilah et al. [10] used K-Means methods to analyze user behavior based on their logs of online usage. For the experiment, a WEKA software package was utilized. Their findings demonstrate that higher education institutions' internet users have easier access to websites for information searches. Additionally, social media receives more traffic than search engines.

To demonstrate the superiority of the fuzzy clustering algorithm over the hard clustering algorithm, Adesina et al. [11] conducted a comparative analysis of the three clustering approaches utilized in criminal profiling. The study looked at the integration and analysis of real-world data from Nigerian law enforcement organizations to create "profiles" of criminal activities and behavior. The Nigeria Police Department in Eleyele, Ibadan, and the Nigeria Police Force Headquarters in Abuja provided the data utilized in the project. The WEKA software program was used to construct the result, which examined three different clustering techniques—two hard clustering algorithms and one fuzzy clustering algorithm. The algorithms' performance was assessed based on time complexity and cluster accuracy. According to their findings, Expectation Maximization produced 90.5% accurate clusters in 8.5 seconds, K-Means produced 62.6% in 0.09 seconds, and the Hierarchical clustering technique produced 51.9% in 0.11 seconds. However, their analysis was constrained by the availability of the data they used.

The literature pertinent to this research has been thoroughly reviewed, with additional sources provided in previous research [12-22]. The efficacy of the Fuzzy C-Means (FCM) algorithm in managing data uncertainty and imprecision is limited by its dependence on the Euclidean distance measure. In order to overcome these constraints, the suggested approach introduces the Support Vector Machine (SVM) inner product distance norm, which is more suitable for handling diverse and intricate data structures. The kernelized Fuzzy C-Means (KFCM) is an extension of the Fuzzy Comparison Method (FCM) that incorporates a kernel function to effectively handle non-linear data structures. However, KFCM faces challenges in choosing the most suitable kernel function and obtaining data from various sources. The integration of FCM with SVM optimises the advantages of both approaches, resulting in enhanced clustering precision and decreased computational complexity. The hybrid FCM-SVM method suggested here is a notable improvement compared to current methods. It provides a more adaptable, precise, and computationally efficient algorithm for criminal profiling, which is especially beneficial for law enforcement agencies that handle intricate and diverse datasets.

3. RESEARCH APPROACH

This paper applied a hybridized FCM algorithm and SVM to profile criminals with the aim of determining the optimal SVM classifier with the best kernel function that best fits into the classification of behavioral activities. The data used to test the validity of the system was downloaded online and it is available at https://portal.chicagopolice.org/portal/page/portal/ClearPath/News/Crime%20 statistics from the city of Chicago Police Department. The data acquired was preprocessed by first performing data cleaning routines to fill in the missing value, identify outliers and correct data inconsistencies. The raw data was then replaced with higher level concepts utilizing concept hierarchies, and the data was then translated into the proper forms required using the generalization concept. It was further lowered by discretization, which replaced the raw data with higher conceptual levels, and feature selection, which involved using the vector slicer idea to remove as much redundant and useless information as feasible. Feature that was selected include; crime nature, mode of operation, time of commission, crime location etc.

The feature scaling approach was used to normalize the data, converting all of the characteristics to numeric values and narrowing the range of the data's features to a scale between 0 and 1. Eq. (1) calculates *z*, the normalized value of a member of the set of observed values of *x*, was used to do this.

$$Z=\frac{x-min(x)}{max(x)-min\,(x)} \tag{1}$$

The *Cp* value which is gotten from Eq. (2) gives the various points of the data items, that is, the representation of the various data items in order to generate a cluster analysis based on the severity of the criminals.

Step 1:

$$Cp = (Cv * i)\,(w) + CFv \tag{2}$$

where, *i* is the seriousness of the offense committed and *Cp* is the criminal profile for each offender; *w* is the weight of the explanation of how the crime was done, and *Cv* is the criminal value with an associated value of one; The number for crime frequency is *CFv*. Table 1 is used to compute and get the *CFv*.

Step 2: Use fuzzy partition to compute the membership matrix (*U*).

$$\sum_{i=1}^{c} Uij = 1\ \forall\ j=1, 2,\ldots, n \tag{3}$$

Step 3. Determine the centroids, the cluster analysis system's central point.

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_{ij}}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{4}$$

Step 4. The threshold value is checked in Eq. (5) and the process ends if the proper threshold value is discovered; if not, it moves on to next step. The dissimilarity function is used to determine the differences between the centroid and the data points. It gauges how well the data match the clustering.

$$J(U, c_1, c_2, \ldots, c_c) = \sum_{i=1}^{c} j_i = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2} \tag{5}$$

where,

$$d_{ij}^{2} = min\,\frac{1}{2}\left\|Cp_j - v_i\right\|^{2}\omega + \gamma\sum_{i=1}^{n}\lambda_i = min\,\frac{1}{2}\left(Cp_{j-}v_i\right)^{T}\omega + \gamma\sum_{i=1}^{n}\lambda_i \tag{6}$$

is a SVM inner-product distance norm, such that:

$$q_i\left(Cp_{j-}v_i\right)^{T}\psi\left(Cp_j\right) + b = 1 - \lambda_i,\ \lambda_i \geq 0, i = 1, 2, \ldots n \tag{7}$$

where, *n* is the training datasets $[(Cp_j, q_i)]_{i=1}^{n}$; $Cp_i \in R^d$ is the input vector and $q_i \epsilon [-1, +1]$ is the corresponding class label for the point Cp_j.

Here, the non-linear mapping $\psi(Cp_j)$ maps the input vector into higher dimensional distance space and "b" denotes the bias whereas ω denotes weight vector of the same dimension as the distance space.

λ_i is the slack variables to tolerate misclassification and the regulation parameter γ is a constant to tradeoff between the maximization of the cluster margin and minimization of the classification error. The large the value of γ, the more the error term is emphasized and the small the value means that the large classification margin is encouraged.

$m \in [1, \infty]$ is a parameter that, when set to 2, controls how fuzzy the generated clusters are.

$$\text{If } ||U^{(k+1)} - U^{(k)}|| < 0.01 \tag{8}$$

Eq. (8) compares the difference between the threshold value and the values of the current and subsequent classes in the membership function.

Restart step 2 until it converges if the threshold value is not met.

Table 1. Criminal frequency value

Category	No. of Crimes (#)	Assigned Value (CFv)
1	1	1
2	2-4	2+(#-2)/3
3	5-10	3+(#-5)/5
4	>10	4

Source: Hammouri [2].

4. EXPERIMENTAL SETUP

Three experiments were carried out using the hybridized algorithm on each SVM classifier; the first experiment used the multiclass ranking SVMS also known as one against one (OAO) classifier, the second experiment used the one against all (OAA) classifier and the third approach used the pair wise classifier (BSVM). Also, different kernel function was selected for each experiment in order to know the best kernel function that most describe our crime data. since the classification problem at hand involves more than two classes. in each experiment, different kernel functions were adopted and compared with each other and the mean computed metrics was then further compared with the existing methodologies used. the constant C parameter was set to 10 to control the tradeoff between the margin and the misclassification errors.

4.1 Result analysis

The results obtained from the implementation of the modified Fuzzy C-Means Clustering Algorithm using a Support Vector Machine are shown in Figures 1-12. Since more than two clusters were formed, three different approaches were adopted for the problem; the first approach uses the multiclass ranking SVMs, also known as one against one (OAO); the second approach uses the one against all (OAA) classification, and the third approach used the pairwise classification (BSVM2). Also, different kernel function was selected for each approach in order to know the best kernel function that best describes our crime data.

Kernels are the most tricky and important part of using SVM because they create the kernel matrix, which summarizes all of the data points. The reason is that the data points appear in the dot product, and the kernel function is able to compute the inner products of these points. So, there is no need to map the points explicitly in feature space.

Therefore, the use of different kernel functions directly computes the data points through the inner product and finds the equivalent points on the hyperplane. The data points on the boundaries of the hyperplane are called the support vectors, and they basically determine and differentiate the clusters formed. The larger number of SVs shows that the feature space poorly discriminates the clusters. The smaller number of SVs shows that the clusters are well separated, which is the goal of the clustering algorithm. The three clusters formed also categorized the criminals into groups based on their behavioral characteristics in terms of their relationship to the three clusters created, such as light, intermediate, and heavy-weight criminals alongside the numbers of support vectors formed are shown on the different clusters formed in each of the different approaches. The higher the number of support vectors, the noisier the data, and this shows that the cluster is not well-defined.

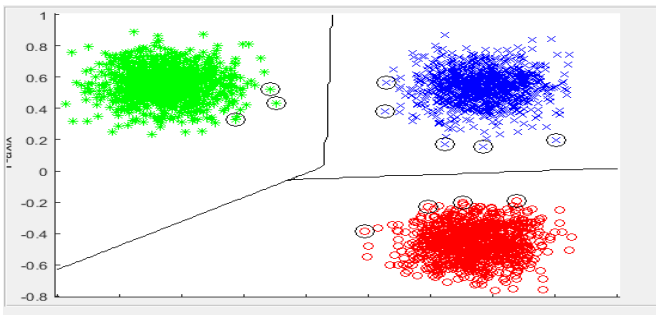

Figure 1. MFCM cluster using OAO with linear kernel

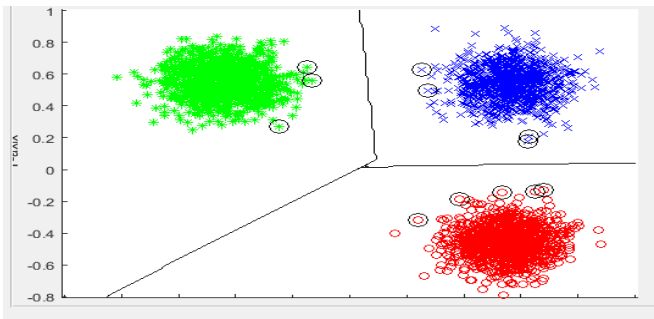

Figure 2. MFCM cluster using OAO with polynomial kernel

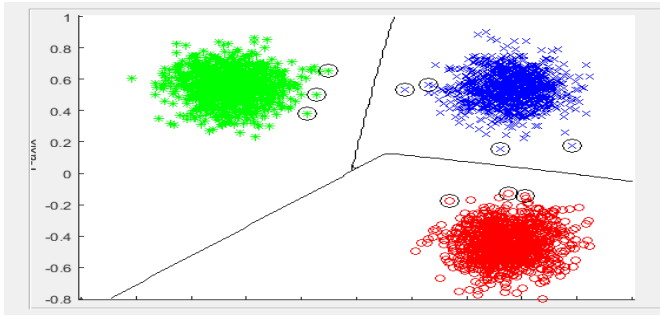

Figure 3. MFCM cluster using OAO with RBF kernel

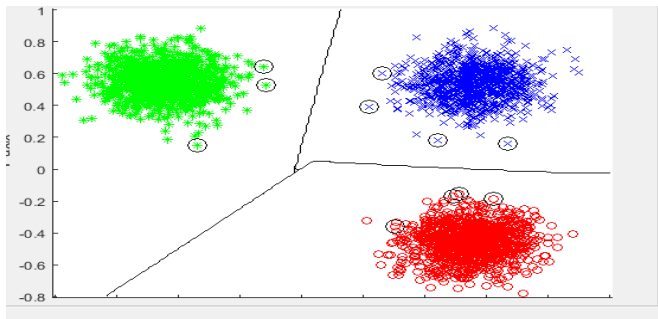

Figure 4. MFCM cluster using OAO with sigmoid kernel

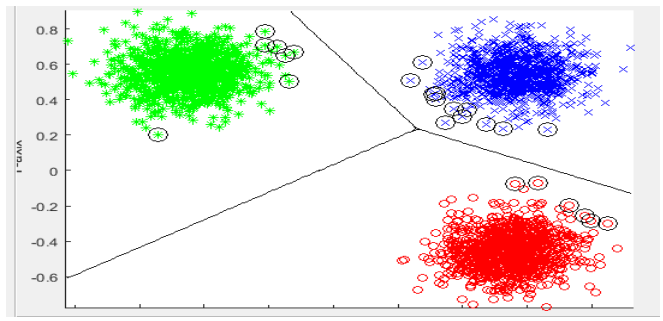

Figure 5. MFCM cluster using OAA with linear kernel

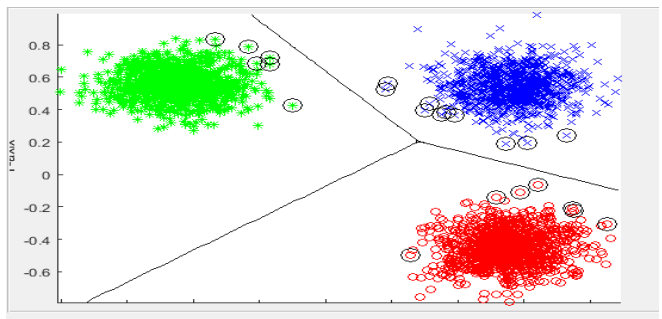

Figure 6. MFCM cluster using OAA with polynomial kernel

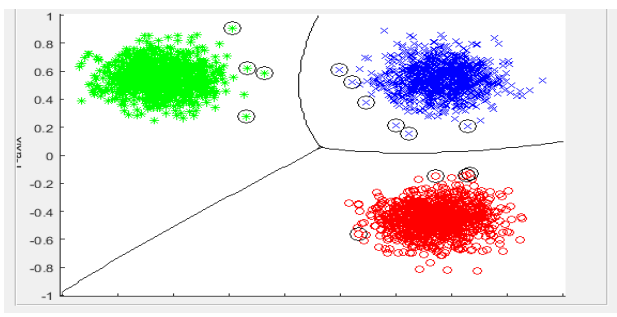

Figure 7. MFCM cluster using OAA with RBF Kernel

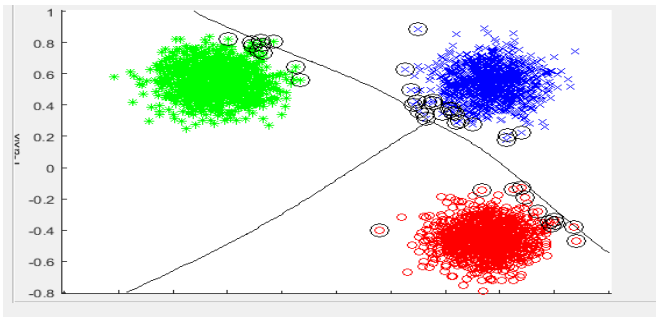

Figure 8. MFCM cluster using OAA with sigmoid kernel

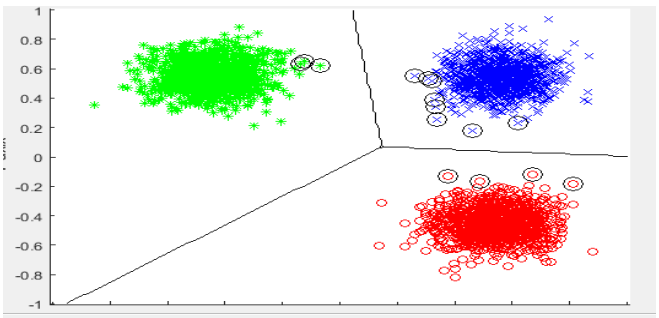

Figure 9. MFCM cluster using BSVM2 with linear kernel

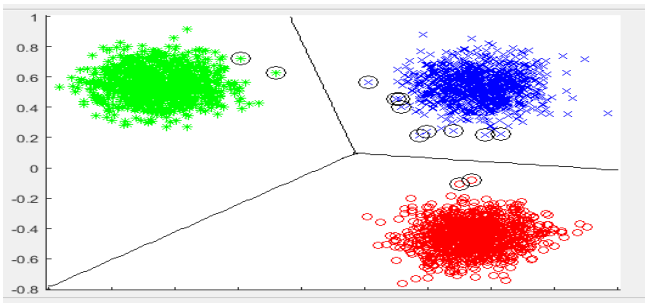

Figure 10. MFCM cluster using BSVM2 with polynomial kernel

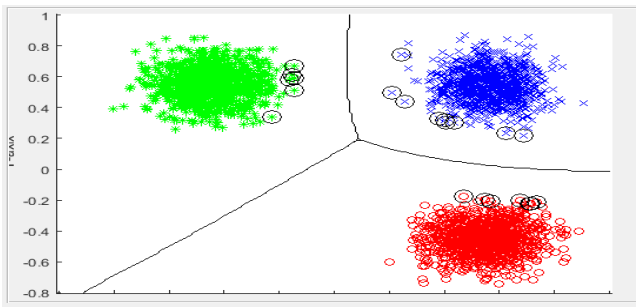

Figure 11. MFCM cluster using BSVM2 with RBF kernel

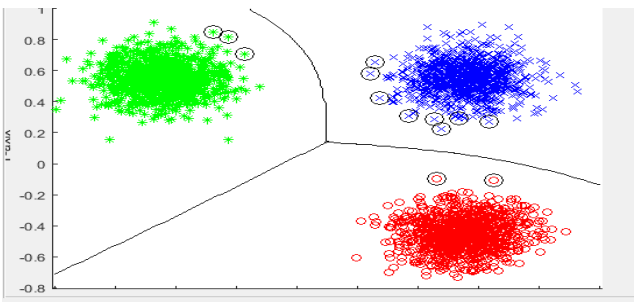

Figure 12. MFCM cluster using BSVM2 with sigmoid kernel

The results gotten from the three experiments are shown in Tables 2-4. From Table 2, it shows that RBF kernel outperform others with ten numbers of Support Vectors (SV) with a running time of 0.15s, for the OAO classifier while Table 3 also shows that RBF kernel outperform others with fourteen numbers of Support Vectors with a running time of 1.39s for the OAA classifier and lastly Table 4 shows that RBF kernel outperform others with ten numbers of Support Vectors and a running time of 0.17s for the BSVM classifier.

It is expected that the larger the number of the SVs, shows that the feature space poorly discriminates the clusters. While the smaller the number of SVs shows that the clusters are well separated which is the goal of clustering algorithm. The Radial Basis Function (RBF) kernel was chosen over other kernels since it has the smallest number of SVs in all the three experiments performed, this shows that it has a better boundary response. Hence, the RBF perfectly describe our crime data set over the others.

Table 2. Result for different kernel functions using OAO

Kernel Function	Time-Taken	No. of Support Vector
Linear	0.34	12
RBF	0.15	10
Poly	0.59	12
Sigmoid	0.25	11

Table 3. Result for different kernel functions using OAA

Kernel Function	Time-Taken	No. of Support Vector
Linear	2.12	24
RBF	1.39	14
Poly	1.32	23
Sigmoid	0.78	44

Table 4. Result for different kernel functions using BSVM

Kernel Function	Time-Taken	No. of Support Vector
Linear	0.15	15
RBF	0.17	10
Poly	0.20	11
Sigmoid	0.21	13

Also, the result for the confusion matrix table for the first experiment is shown in Table 5 that used the OAO classifier with the different kernel functions from a to d. And its performance evaluation results are shown in Table 6 also for the different kernel function using OAO classifier for easy analysis. While Table 7 and Table 8 show the Confusion matrix table and the performance evaluation results for the second experiments. Lastly, the confusion matrix table for the third experiment is shown in Table 9 and the performance evaluation result is shown in Table 10, respectively. The mean percentage of the various metrics is also calculated.

Table 5. Experiment one: Confusion matrix using OAO classifier with different kernel functions

Kernel Function	Cluster No.	TP	FN	FP	TN
Linear Kernel	1	232	8	29	88
	2	230	7	5	140
	3	213	8	3	61
	Average	**225**	**8**	**12**	**96**
RBF Kernel	1	191	7	39	130
	2	247	9	2	122
	3	161	7	3	114

	Average	**200**	**8**	**15**	**122**
Polynomial Kernel	1	214	9	19	109
	2	258	10	1	113
	3	237	8	2	38
	Average	**236**	**9**	**7**	**87**
Sigmoid Kernel	1	208	8	50	113
	2	219	8	4	151
	3	175	7	4	99
	Average	**201**	**8**	**19**	**121**

Table 5 summarizes the true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) for each cluster across different kernel functions, along with their averages.

Table 6. Experiment one: Performance analysis result using OAO classifier with different kernel functions

Kernel Function	Cluster No.	SPEC (%)	SENS (%)	PREC (%)	ACC (%)	Time (s)
Linear Kernel	1	91.85	96.57	94.71	94.69	7.75
	2	96.48	96.67	97.89	96.59	7.75
	3	98.00	96.20	99.56	96.51	7.75
	Average	**95.44**	**96.48**	**97.39**	**95.93**	**7.75**
RBF Kernel	1	97.10	97.35	99.23	97.29	7.78
	2	98.29	96.60	99.22	97.12	7.78
	3	98.77	96.12	99.49	96.86	7.78
	Average	**98.05**	**96.69**	**99.32**	**97.09**	**7.78**
Polynomial Kernel	1	90.40	96.33	94.59	94.17	7.78
	2	95.65	96.17	99.01	96.07	7.78
	3	97.01	96.29	99.05	96.47	7.78
	Average	**94.36**	**96.26**	**97.55**	**95.57**	**7.78**
Sigmoid Kernel	1	66.94	96.38	85.41	86.59	7.78
	2	96.23	97.46	98.53	97.12	7.78
	3	95.62	95.94	95.94	95.79	7.78
	Average	**86.26**	**96.59**	**93.29**	**93.17**	**7.78**

Table 6 presents the specificity (SPEC), sensitivity (SENS), precision (PREC), accuracy (ACC), and time (TIME) across different kernel functions and their averages for each function.

Table 7. Experiment two: Confusion matrix using OAA classifier with different kernel functions

Kernel Function	Cluster No.	TP	FN	FP	TN
Linear Kernel	1	268	7	4	57
	2	246	8	4	124
	3	228	9	1	49
	Average	**247**	**8**	**3**	**77**
RBF Kernel	1	220	8	5	104
	2	247	9	4	122
	3	198	8	1	80
	Average	**222**	**8**	**13**	**102**
Polynomial Kernel	1	274	11	22	47
	2	244	10	1	127
	3	208	8	2	65
	Average	**242**	**10**	**8**	**80**
Sigmoid Kernel	1	237	9	68	87
	2	325	10	1	46
	3	142	6	6	131
	Average	**235**	**8**	**25**	**88**

Table 7 presents the true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) across different kernel functions and their averages for each function.

Table 8. Experiment two: Performance analysis using OAA classifier with different kernel functions

Kernel Function	Cluster No.	SPEC (%)	SENS (%)	PREC (%)	ACC (%)	Time (s)
Linear Kernel	1	91.85	96.57	94.71	94.69	7.75
	2	96.48	96.67	97.89	96.59	7.75
	3	98.00	96.20	99.56	96.51	7.75
	Average	**95.44**	**96.48**	**97.39**	**95.93**	**7.77**
RBF Kernel	1	97.10	97.35	99.23	97.29	7.78
	2	98.29	96.60	99.22	97.12	7.78
	3	98.77	96.12	99.49	96.86	7.78
	Average	**98.05**	**96.69**	**99.32**	**97.09**	**7.78**
Polynomial Kernel	1	90.40	96.33	94.59	94.17	7.78
	2	95.65	96.17	99.01	96.07	7.78
	3	97.01	96.29	99.05	96.47	7.78
	Average	**94.36**	**96.26**	**97.55**	**95.57**	**7.78**
Sigmoid Kernel	1	66.94	96.38	85.41	86.59	7.78
	2	96.23	97.46	98.53	97.12	7.78
	3	95.62	95.94	95.94	95.79	7.78
	Average	**86.26**	**96.59**	**93.29**	**93.17**	**7.78**

Table 8 displays the specificity (Spec.), sensitivity (Sens.), precision (Prec.), accuracy (Acc.), and time (s) across different kernel functions, including the averages for each kernel function.

Table 9. Experiment three: Confusion matrix using BSVM classifier with different kernel functions

Kernel Function	Cluster No.	TP	FN	FP	TN
Linear Kernel	1	235	9	15	87
	2	219	9	4	150
	3	200	6	2	77
	Average	**218**	**8**	**7**	**105**
RBF Kernel	1	197	8	22	123
	2	299	6	2	75
	3	217	6	2	60
	Average	**238**	**7**	**9**	**86**
Polynomial Kernel	1	203	7	10	118
	2	259	6	5	110
	3	200	7	1	77
	Average	**221**	**7**	**5**	**102**
Sigmoid Kernel	1	240	9	50	83
	2	287	10	2	85
	3	164	6	4	111
	Average	**230**	**8**	**19**	**93**

Table 9 summarizes the true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) for each kernel function, along with their average values across clusters.

Table 10. Experiment three: Performance analysis result using BSVM2 classifier with different kernel functions

Kernel Function	Cluster No.	SPEC (%)	SENS (%)	PREC (%)	ACC (%)	Time (s)
Linear Kernel	1	82.12	96.08	87.89	90.14	7.60
	2	96.38	96.31	97.92	96.34	7.60
	3	97.47	97.09	99.01	97.19	7.60
	Average	**91.99**	**96.49**	**94.94**	**94.56**	**7.60**
RBF Kernel	1	80.83	96.58	90.76	91.24	7.77
	2	95.74	95.53	98.58	96.34	7.77
	3	96.77	97.31	99.09	94.19	7.77
	Average	**91.12**	**96.81**	**96.14**	**94.92**	**7.77**
Polynomial Kernel	1	91.74	96.10	96.10	94.71	7.53

	2	96.03	97.29	98.05	96.87	7.53
	3	98.72	96.61	99.50	97.19	7.53
	Average	**95.49**	**96.67**	**97.88**	**96.26**	**7.53**
Sigmoid Kernel	1	89.31	97.17	93.64	94.17	7.70
	2	95.95	97.01	97.42	96.59	7.72
	3	96.52	96.47	97.61	96.49	7.72
	Average	**93.93**	**96.88**	**96.23**	**95.75**	**7.72**

Table 10 provides a clear comparison of the performance metrics (Specificity, Sensitivity, Precision, Accuracy, and Time) across different kernel functions for each cluster.

4.2 Comparative analysis of the different classifiers

The mean percentage of all the metrics used were computed and selected for each classifier using different kernel functions. Table 11 shows the mean percentage of the different kernel functions used under the OAO classifier, with the RBF kernel function performing best with 96.32% accuracy and 96.24% specificity over other kernel functions. In the same way, Table 12 also shows the mean percentage of the different kernel function used under the OAA classifier with RBF kernel function performing best with 97.09% accuracy and 96.24% specificity over other kernel functions. Lastly, Table 13 depicts the mean percentage of the different kernel function used under the BSVM classifier with the polynomial kernel function performing best with 96.26% accuracy over other kernel functions. Hence, RBF kernel function performs better for the crime data set used over others. Since its percentage accuracy is the highest and has been seen to have performed better than the other two using two different classifiers from the three classifiers used.

Table 11. Comparative analysis of the different kernel functions using OAO classifier

Kernel Function	SPEC (%)	SENS (%)	PREC (%)	ACC (%)	Time (s)
Linear	90.18	96.49	95.56	94.23	7.7
RBF	96.24	96.44	97.72	96.32	7.78
Polynomial	94.39	96.99	96.94	96.01	7.77
Sigmoid	92.87	96.41	94.80	94.70	7.78
Average	93.42	94.08	96.26	95.36	7.78

Table 12. Comparative analysis of the different kernel functions using OAA classifier

Kernel Function	SPEC (%)	SENS (%)	PREC (%)	ACC (%)	Time (s)
linear	95.44	96.48	97.39	95.93	7.75
RBF	98.05	96.69	99.32	97.09	7.78
Polynomial	94.36	96.26	97.55	95.57	7.78
sigmoid	86.26	96.59	93.29	93.17	7.79
Average	93.53	96.51	96.89	95.44	7.78

Table 13. Comparative analysis of the different kernel functions using BSVM classifier

Kernel Function	SPEC (%)	SENS (%)	PREC (%)	ACC (%)	Time (s)
Linear	91.99	96.49	94.94	94.56	7.60
RBF	91.12	96.81	96.14	94.92	7.77
Polynomial	95.49	96.67	97.88	96.26	7.53
Sigmoid	93.93	96.88	96.23	95.75	7.72
Average	93.13	96.71	96.29	95.37	7.65

The mean percentage of the different metrics under each classifier was also computed and selected and further compared in Table 14 to really ascertain the classifier that best suits the crime data set used in the experiment. It can be seen that the OAA Classifier performed best with 93.53%. Specificity, 96.89% Precision, and 95.44% Accuracy. Also, the mean percentages of the three classifiers used were lastly computed to obtain the mean performance of the hybridized algorithm on all the different classifiers.

Table 14. Comparative analysis of different classifier

Classifier	SPEC (%)	SENS (%)	PREC (%)	ACC (%)	Time (s)
OAO	93.42	94.08	96.26	95.34	7.78
OAA	93.53	96.51	96.89	95.44	7.78
BSVM2	93.13	96.71	96.29	95.37	7.65
Average	93.44	95.84	96.57	95.46	7.74

4.3 Discussion of results

While it is established that the RBF kernel function outperformed others in terms of accuracy, specificity, and the number of support vectors, the significance of these findings can be further elaborated. The smaller number of support vectors associated with the RBF kernel indicates that the data clusters are well-separated, which is a critical factor in efficient and accurate classification. A lower number of support vectors typically leads to faster model training and prediction times, which is particularly important when scaling the approach to larger datasets.

Moreover, the running time of the RBF kernel was consistently lower across the experiments. This suggests that the RBF kernel not only provides better classification boundaries but also does so more efficiently. This efficiency is crucial for practical applications, especially in real-time crime analysis, where timely decision-making is essential. The implications of these findings extend beyond mere performance metrics. The efficiency and scalability of the RBF kernel in processing crime data highlight its suitability for large-scale applications, where quick and accurate profiling can significantly impact law enforcement operations. The reduction in processing time, combined with the algorithm's ability to handle large datasets with fewer support vectors, underscores its potential to streamline criminal investigations, enhance decision-making processes, and ultimately contribute to more effective crime prevention strategies.

The proposed hybrid methodology advances the state-of-the-art by integrating the Fuzzy C-Means (FCM) algorithm with Support Vector Machine (SVM) to overcome the limitations of traditional clustering and classification approaches, such as sensitivity to noise, high-dimensional data, and inflexible kernel selection, thus improving clustering accuracy and robustness in real-world crime data. This combination leverages the strengths of both techniques—FCM's capability to handle overlapping clusters and SVM's effectiveness in high-dimensional spaces—resulting in a more adaptive and computationally efficient model that addresses the shortcomings of previous approaches.

These aspects are aligned with the broader goals of Sustainable Development Goal 16, which emphasizes the importance of just, accountable, and inclusive institutions. By leveraging an advanced algorithm like the RBF kernel, law enforcement agencies can more effectively uphold the rule of

law and ensure equal access to justice, as envisioned by the United Nations. This discussion enriches the results analysis by not only identifying the best-performing kernel function but also by contextualizing its practical benefits in real-world crime analysis and profiling.

5. CONCLUSIONS

The major challenge with the use of kernel functions in problem solving is selecting the optimal kernel for the problem at hand; this paper therefore was able to establish that the RBF kernel function performs better for the crime data set used over others. Since its percentage accuracy is the highest and has been seen to have performed better than the other two using two different classifiers from the three classifiers used. This research also focuses on the investigating the three different SVM classifiers in criminal profiling using crime data set. It was evident in the result that the OAA Classifier performed best with 93.53%. Specificity, 96.89% Precision, and 95.44% Accuracy. It is therefore recommended that the RBF kernel function using the OAA classifier best suits our crime data set for criminal profiling.

Through the development of a novel approach of profiling criminal to improving decision making in the various law enforcing agencies, reducing process time of crime analysis to enable quick completion of a crime investigation and also to reduce the difficulties in managing large volume of data involved in criminal activities, in a situation where a crime is committed in the absence of witnesses or any clue for forensics analysis by expert from the crime scene, this research underscores the importance of leveraging advanced algorithms to support SDG 16's vision of fostering just, accountable, and inclusive institutions that uphold the rule of law and promote equal access to justice for all.

ACKNOWLEDGMENT

The authors hereby acknowledge Covenant University Centre for Research, Innovation and Discovery (CUCRID) for their support toward the completion of this research.

REFERENCES

[1] Sheeba, M.S., Sathya, A. (2015). Hybrid approach of kernelized fuzzy c-means and support vector machine for breast medical image segmentation. Journal of Chemical and Pharmaceutical Research, 7(2): 281-291.

[2] Hammouri, J.A.A. (2023). Modeling the performance of criminal law functions in the context of safety and security development. International Journal of Safety and Security Engineering, 13(3): 395-401. https://doi.org/10.18280/ijsse.130302

[3] Rao, P.S., Sudheer, D., Babu, M.R. (2021). Support vector machine algorithm for analysis of FBI crime data. Journal of Cardiovascular Disease Research, 12(4): 1694.

[4] Bhavani, D. (2019). The data mining support vector machine algorithm used for detecting and forecasting of crimes. International Journal of Engineering and Advanced Technology (IJEAT), 9(1): 395. http://dx.doi.org/10.35940/ijeat.A9386.109119

[5] Krysovatyy, A., Lipyanina-Goncharenko, H., Sachenko, S., Desyatnyuk, O. (2021). Economic crime detection using support vector machine classification. MoMLeT+ DS, 2917: 830-840.

[6] Costa, N.L., Llobodanin, L.A.G., Castro, I.A., Barbosa, R. (2019). Using support vector machines and neural networks to classify Merlot wines from South America. Information Processing in Agriculture, 6(2): 265-278. https://doi.org/10.1016/j.inpa.2018.10.003

[7] He, W., Liu, Y. (2018). To regularize or not: Revisiting SGD with simple algorithms and experimental studies. Expert Systems with Applications, 112: 1-14. https://doi.org/10.1016/j.eswa.2018.06.026

[8] Kim, S., Joshi, P., Kalsi, P.S., Taheri, P. (2018). Crime analysis through machine learning. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 415-420. https://doi.org/10.1109/IEMCON.2018.8614828

[9] Okonkwo, R.O., Enem, F.O. (2011). Combating crime and terrorism using data mining techniques. In 10th International Conference IT People Centred Development, Nigeria Computer Society, Nigeria, pp. 80-89.

[10] Zulfadhilah, M., Riadi, I., Prayudi, Y. (2016). Log classification using K-means clustering for identify Internet user behaviors. International Journal of Computer Applications, 154(3): 34-39.

[11] Adesina, O.S., Adedotuun, A.F., Adekeye, K.S., Imaga, O.F., Adeyiga, A.J., Akingbade, T.J. (2024). On logistic regression versus support vectors machine using vaccination dataset. Journal of the Nigerian Society of Physical Sciences, 6(1): 1092. https://doi.org/10.46481/jnsps.2024.1092

[12] Fauzi, A., Butar, J.B, Budi, I., Ramadiah, A., Putra, P.K., Santoso, A.B. (2024). Supervised machine learning entity sentiment analysis: Prediction of support for 2024 Indonesian presidential candidates. Revue d'Intelligence Artificielle, 38(2): 587-594. https://doi.org/10.18280/ria.380222

[13] Eweoya, I.O., Odetunmibi, O.A., Odun-Ayo, I.A., Agbele, K.K., Adedotun, A.F., Akingbade, T.J. (2023). Machine learning approach for the prediction of COVID-19 spread in Nigeria using SIR model. International Journal of Sustainable Development and Planning, 18(12): 3783-3792. https://doi.org/10.18280/ijsdp.181210

[14] Singh, P., Rathee, N., Sharda, S., Kumar, S. (2023). Comparative study of Rough Set-based FCM and K-Means clustering for tumor segmentation from brain MRI images. Revue d'Intelligence Artificielle, 37(4): 921-927. https://doi.org/10.18280/ria.370412

[15] Jonathan, O., Misra, S., Osamor, V. (2021). Comparative analysis of machine learning techniques for network traffic classification. IOP Conference Series: Earth and Environmental Science, 655(1): 012025 https://doi.org/10.1088/17551315/655/1/012025

[16] Obagbuwa, I.C., Abidoye, A.P. (2021). South Africa crime visualization, trends analysis, and prediction using machine learning linear regression technique. Applied Computational Intelligence and Soft Computing, 2021: e5537902. https://doi.org/10.1155/2021/5537902

[17] Aldossari, B.S., Alqahtani, F.M., Alshahrani, N.S., Alhammam, M.M., Alzamanan, R.M., Aslam, N., Irfanullah. (2020). A comparative study of decision tree and naive bayes machine learning model for crime

category prediction in Chicago. In Proceedings of 2020 6th International Conference on Computing and Data Engineering, pp. 34-38. https://doi.org/10.1145/3379247.3379279

[18] Adesina, O.S., Adedotun, A.F., Ayoola, F.J., Adesina, T.F., Alayande, S.A., Onayemi, O.O. (2024). Statistical learning insights on Nigerian aviation service quality. International Journal of Transport Development and Integration, 8(1): 1-7. https://doi.org/10.18280/ijtdi.080101

[19] Oladipo, O., Omidiora, E.O., Osamor, V.C. (2024). Comparative analysis of features extraction techniques for black face age estimation. AI & SOCIETY, 39(4): 1769-1783. https://doi.org/10.1007/s00146-022-01407-0

[20] Omidiora, E.O., Oladele, M.O., Adepoju, T.M., Sobowale, A.A., Olatoke, O.A. (2016). Comparative analysis of back propagation neural network and self organizing feature map in estimating age groups using facial features. British Journal of Applied Science & Technology, 15(1): 1-7. https://doi.org/10.9734/bjast/2016/24303

[21] Omodero, C.O. (2023). Effects of insecurity, terrorism and political instability on foreign direct investment inflows in Nigeria. International Journal of Safety and Security Engineering, 13(6): 1091-1098. https://doi.org/10.18280/ijsse.130612

[22] Kontiainen, L.E., Koulu, R., Sankari, S.E. (2022). Research agenda for algorithmic fairness studies: Access to justice lessons for interdisciplinary research. Frontiers in Artificial Intelligence, 5: 882134. https://doi.org/10.3389/frai.2022.882134