DOI: https://doi.org/10.52756/ijerr.2024.v42.026 *Int. J. Exp. Res. Rev.*, Vol. 42: 298-311 (2024)

Original Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in

Kruskal Wallis and mRMR Feature Selection based Online Signature Verification System using Multiple SVM and KNN

Bhimraj Prasai Chetry* and Biswajit Kar

Department of Instrumentation Engineering, Central Institute of Technology Kokrajhar, Kokrajhar-783370, Assam, India

E-mail/Orcid Id:

BPC, bp.chetry@cit.ac.in, https://orcid.org/0009-0002-5072-4739;
BK, b.kar@cit.ac.in, https://orcid.org/0000-0003-2686-5814

Article History:

Received: 21th May, 2024
Accepted: 14th Aug., 2024
Published: 30th Aug., 2024

Keywords:
Feature extraction, Feature selection, Kruskal Wallis, mRMR, Multiple SVM, Multiple KNN, Signature verification

How to cite this Article:
Bhimraj Prasai Chetry and Biswajit Kar (2024). Kruskal Wallis and mRMR Feature Selection based Online Signature Verification System using Multiple SVM and KNN. *International Journal of Experimental Research and Review*, *42*, 298-311.
DOI:
https://doi.org/10.52756/ijerr.2024.v42.026

Abstract: Signature verification is a very important research area. Signature has been widely accepted as a person authentication method for centuries. It is mostly used in financial transactions, document authentication and agreements. It is more susceptible to being forged than any other biometrics. Online signature verification is used in real-time applications like e-commerce, online resource access, online financial transactions, physical access into a restricted area and many more. In order to achieve high efficiency in online signature verification systems, feature extraction and feature selection play a significant role. A suitable signature verification system is needed to prevent forgery and accept the genuine signer. We have extracted 30 global features from all 40 signers for verification. Here, k fold cross-validation technique is used to enhance the model's performance on unseen data. User-specific feature selection and ranking are done using Kruskal Wallis and Minimum Redundancy Maximum Relevance (mRMR) algorithm to hunt which performs better in our case. Kruskal-Wallis method tests if two or more classes have an equal median and gives the value of P based on which discriminative features are selected, whereas the mRMR algorithm ranks the whole feature set according to its importance. It evaluates the relevance of a feature and penalizes redundancy. Finally, multiple SVM and KNN classifiers are trained and tested with various selected features using Kruskal Wallis and mRMR to determine which combination performs best for the online signature verification system. Our model is trained, validated and tested on the SVC 2004 Task 1 database, which consists of skill forgery signatures. Here, one to one verification is done using each user's genuine and skill forgery signatures, which is the hardest to detect. Best average testing accuracy achieved in our case is 90.25% using Weighted KNN and Kruskal Wallis selected 15 features.

Introduction

Signature verification is a biometric authentication method we need to deal with in our daily lives in a wide range of practical applications, including fraud prevention in financial transactions, e-commerce, e-delivery and other important documentation. Generally, signature verification systems are divided into online and offline systems. Offline systems refer to static images of signatures, whereas online system signatures are characterized and analyzed as time sequences of the dynamic writing process (e.g., Velocity, Acceleration, Time, Pressure, etc.). Online signature verification methods have been proven to achieve better accuracy than offline verification methods (Napa Sae-Bae and Memon, 2014). Therefore, in this work, we propose an online signature verification system. Biometric verification system automatically identifies a person's identity based on its behavioral or physiological traits (Kar et al., 2018). More stable traits like fingerprint and iris are generally available for verification because of their high accuracy still, handwritten signature-based verification is a trending research field because of its social and legal acceptance and its presence in contracts, wills and other important documents since time

***Corresponding Author:** bp.chetry@cit.ac.in

immemorial (Diaz et al., 2018). Based on its application, signature biometrics can be used for identification or verification purpose. In verification, the system confirms claimed identity by comparing the biometric identifier presented by the user with a reference template for the claimed identity stored in the system during enrolment. This is done by carrying out a one-to-one matching process. In identification, the system compares the biometric identifier with all the templates stored in the system database. This is done by conducting a one-to-many comparison process. Biometric verification has gained popularity due to the unpredictability and inconvenience of traditional verification techniques. The main job of any signature verification system is to verify whether the signature is genuine or forgery. Among all the skilled forgeries, signatures are the hardest to detect because skilled forgeries are signatures in which forgers know the signer's name and style of original signatures (Parmar et al., 2020). In this world of emerging technology, online signature verification will play a very important role in the field of biometrics with good user acceptance and will be very helpful in preventing possible imitation by the forgers while dealing with e-commerce, e-transactions, e-delivery and many more. A forger can easily forge the shape/pattern of the signature, but it is not possible for him to forge the dynamic information of the signature, which is hidden in the writing process and is very personal to each user.

Literature Review

Thorough research is available in the area of online signature verification, which can be seen in (Parmar et al., 2020; Impedovo and Pirlo, 2008; Plamondon and Lorette, 1989). Automatic signature verification by Herbst and Liu in 1977 summarizes the state-of-the-art prior to that date (Herbst and Liu, 1977). Their analysis of existing methodologies was later updated in the year 2000 (Plamondon and Srihari, 2000). Online signature verification systems use more advanced techniques such as Dynamic time wrapping (Nalwa, 1997). And the hunt for global features was ongoing (Lee et al., 1996). Online signature verification methods, which are difficult exercises, are of two types: The first type is based on the use of global features, and the second type is called the temporal function-based approach (Kar et al., 2018). Generally, function-based features show better discriminating ability than the parameter-based features but require a time-taking algorithm for comparison (Kar and Dutta, 2012). However, the work done by Aguilar et al. shows that the parametric approaches also compete equally with the function-based approaches (Fierrez-Aguilar et al., 2005). Hence, the author uses only global features. In this work, we have extracted 30 global features for implementation. A set of different number of global features is selected using Kruskal Wallis and mRMR techniques for support vector machine (SVM) and K-nearest neighbours (KNN) based enrollment and verification. We have used K-fold cross-validation to enhance the machine learning model's performance on unseen data and to overcome problems like selection bias or overfitting (Rao and Wu, 2005). We have seen commonly used values of k is 5(five), as this value is observed to provide test error rate assessment that suffers neither from extremely high bias nor very high variance (Nti et al., 2021). So, we have used the value of k as 5(five) in our work. SVM was expanded in the 1990s to create nonlinearly separating functions and to estimate real-valued functions (Chamasemani and Singh, 2011). Most of the mathematical ideas that underlie the implementation of multiclass SVM are found in the following (Abe, 2010; Hong and Cho, 2008). KNN performance depends upon the optimum value of K and the distance. Researchers have used various methods to determine the distance. The Euclidean distance method is more common and famous (Kotsiantis et al., 2006). We have implemented the algorithm using the SVC 2004 Task 1 benchmark database, which includes skilled forgeries. The mRMR algorithm, first proposed by Peng et al. (Hanchuan Peng et al., 2005), is the most widely used filter method for feature selection. It uses mutual information to calculate measures of relevance and redundancy between the different features and the class label. As seen in (Ali Khan et al., 2014), Kruskal Wallis algorithm selects the more discriminative face features by reducing the search space greatly. Kruskal Wallis algorithm is simple and less time-consuming. Majority of systems presented handling the handwritten signatures choose verification over identification. So, we have developed an online signature verification system that incorporates the abovementioned techniques in the literature. The outcomes of this work are quite promising.

Proposed System

In the online signature verification system, input is the online signature which is collected by a pen-sensitive tablet PC or other online signature-capturing devices like a camera or touchscreen. The raw signatures are processed using different filters and normalization techniques. Our paper used a benchmark database (SVC 2004 Task 1), which had already been captured and processed for research. These signatures are further processed for feature extraction and selection. The selected features are used to generate the classification model. The model's template is kept in database for signature verification. The block diagram as shown in Figure 1 depicts the various stages of our proposed model.

Touchscreen-based online signature collection is shown in Figure 2 below

Global Feature Extraction

Feature extraction is a very important step for an online signature verification. Features can be global or

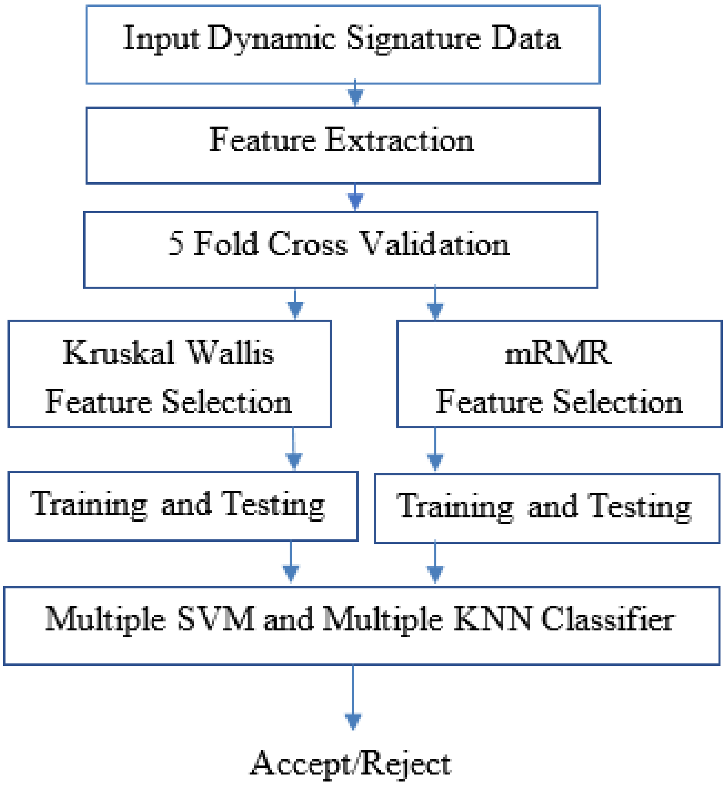

Figure 1. Block diagram of the proposed system.

Our Proposed system is divided into five sections explained as follows:

Database Used

We have used the SVC 2004 Task 1 dynamic signature database (Yeung et al., 2004). SVC 2004 Task 1 is a signature modality having 40 sets of signatures from each user. The first twenty signatures represent the genuine signatures and the remaining twenty represent the skilled forgeries furnished by the other users. The SVC 2004 Task 1 contains 40 users with 40 signatures, each amounting to a total 1600 signatures (Najda and Saeed, 2022). Each genuine or forged signature is kept as a separate text file. “UxSy.txt” is the file name format, where x is the user and y is the one signature instance of the corresponding user.

Where,

$x = [1,2,3, \ldots..40]$ ……………………………….(1)

$y = [1,2,3, \ldots..40]$ …………………....................(2)

In every file, the signature is simply described as a sequence of points. The first line is a single integer, which denotes the total number of points in the signature. Each of the following lines corresponds to one point characterized by four features as X-Coordinate, Y-Coordinate, Time Stamp and Button Status.

local, where global features represent signature properties in general and local features correspond to properties specific to a sampling point. The selection of features to consider for extraction is a very difficult task as it is directly related to the efficiency of the particular signature verification system. The features extracted must be able to describe the signature in such a way that it should have large inter-class variations and negligible intra-class variations.

Figure 2. Touchscreen-based online signature collection.

From the given SVC 2004 signature database, we can create 30 new global features vectors, such as mean absolute velocity X coordinates, total signing duration,

DOI: https://doi.org/10.52756/ijerr.2024.v42.026

etc. The global features characterized signatures' overall gross properties. In this work, 30 global features are extracted as predictors for classification. The description of each global feature is shown in Table 1.

The global feature vector G, comprising of 30 global features, is represented as

$G = [G_{01}, G_{02}, \ldots.. G_{30}]^T \in R^{30}$ …………………..(3)

Here $G_i \in R$ is the i[th] global feature.

Feature Selection Techniques applied

Kruskal Wallis Algorithm

The efficiency of verification system may be degraded by using all the features of input data as it increases the complexity. Selection of the optimized features is very important as some features play a very important role in verification and are more relevant. Many methods are developed and used for feature selection, but most are computationally expensive and complex. Kruskal Wallis technique (Ali Khan et al., 2014) is used in our work to select relevant features that are computationally less expensive and very simple in use. Kruskal-Wallis method tests are selected if two or more classes have equal median and give the value of P. Features with discriminative information. If the value of P is close to "0, " the feature contains discriminative information; otherwise, it will not be selected.

mRMR Algorithm

The feature selection technique aims to select an excellent feature subset by removing irrelevant information from the original feature space according to certain criteria (Mary and Nagarajan, 2024). Feature selection decreases the dimension of data by selecting only a subset of measured features to create a model.

The mRMR algorithm ranks the whole feature set according to its importance. To perform this, it evaluates the relevance of a feature and penalizes redundancy. The objective is to find the maximum dependency between the set of features X and class C, taking mutual information (I) (Hermo et al., 2024). The result of global feature selection using the mRMR and Kruskal Wallis algorithm is shown. Out of 30 global features, 10 selected features for user1 using Kruskal Wallis and mRMR method are shown in Table 2 and Table 3, respectively.

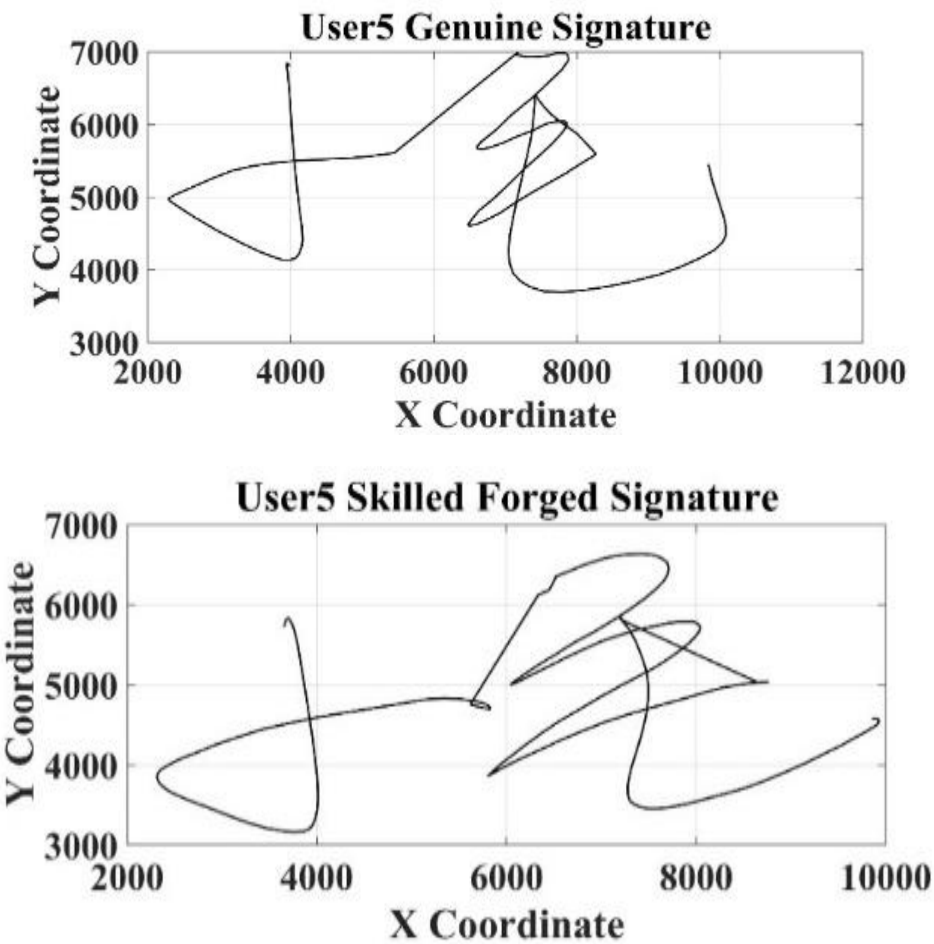

Figure 3. Genuine and Skilled forgery shape of the online signature of the User5 from the SVC 2004 database.

Table 1. All 30 Global Features Extracted.

Sl No.	Features No.	Name of Global Features	Mathematical Descriptions
01	G_{01}	Maximum Value of X Coordinates	X_{max}
02	G_{02}	Maximum Value of Y Coordinates	Y_{max}
03	G_{03}	Minimum Value of X Coordinates	X_{min}
04	G_{04}	Minimum Value of Y Coordinates	Y_{min}
05	G_{05}	Mean Value of Y Coordinates	mean{Y}
06	G_{06}	Width	$w=(X_{max} - X_{min})$
07	G_{07}	Height	$h=(Y_{max} - Y_{min})$
08	G_{08}	Aspect ratio	$ar=(X_{max} - X_{min})/(Y_{max} - Y_{min})$
09	G_{09}	Width of the Signature	$W=\text{mean}\{\lvert X-U(X)\rvert\}$
10	G_{10}	Height of the Signature	$H=\text{mean}\{\lvert Y-U(Y)\rvert\}$
11	G_{11}	Aspect Ratio	$AR=\text{mean}\{\lvert X-U(X)\rvert\}/\text{mean}\{\lvert Y-U(Y)\rvert\}$
12	G_{12}	Distance travelled in X direction	$D_x=\sum\{\Delta\lvert X\rvert\}/W$
13	G_{13}	Distance travelled in Y direction	$D_y=\sum\{\Delta\lvert Y\rvert\}/H$
14	G_{14}	Total distance travelled in XY direction	$D_{xy}=\sum\lvert\sqrt{(\Delta X^2+\Delta Y^2)}\rvert/(W+H)$
15	G_{15}	Number of Pen up sequences	N_{up}
16	G_{16}	Number of Pen down sequences	N_{dw}
17	G_{17}	Ratio of Pen Up to total signing time Rut	$R_{ut}=(N_{up}/T_t)$
18	G_{18}	Total Number of samples	T_s
19	G_{19}	Total Signing Durations	T_t
20	G_{20}	Mean Absolute Displacement X Coordinates	$\text{mean}\{\lvert D_x\rvert\}$
21	G_{21}	Mean Absolute Velocity X Coordinates	$\text{mean}\{\lvert V_x\rvert\}$
22	G_{22}	Standard Deviation of X Coordinates	std(X)
23	G_{23}	Standard Deviation of Absolute Displacement X Coordinates	$\text{std}\{\lvert D_x\rvert\}$
24	G_{24}	Mean Absolute Displacement Y Coordinates	$\text{mean}\{\lvert D_y\rvert\}$
25	G_{25}	Mean Absolute Velocity Y Coordinates	$\text{mean}\{\lvert V_y\rvert\}$
26	G_{26}	Standard Deviation of Y Coordinates	std(Y)
27	G_{27}	Standard Deviation of X directional Absolute Velocity	$\text{std}\{\lvert Vx\rvert\}$
28	G_{28}	Standard Deviation of XY Velocity	$\text{std}[\lvert\sqrt{(\Delta X^2+\Delta Y^2)}\rvert]/\Delta T$
29	G_{29}	Pen Down Duration While Signing	T_{pu}
30	G_{30}	Pen Up Duration While Signing	T_{pd}

Table 2. 10 Selected Features Ranked using Kruskal Wallis for User1 Model.

Sl No.	Signers	Feature No	Ranking	Mathematical Descriptions	Scores
01	User1	G_{29}	1st	T_{pu}	8.776
		G_{16}	2nd	N_{dw}	8.776
		G_{18}	3rd	T_s	8.770
		G_{28}	4th	$\text{std}[\lvert\sqrt{(\Delta X^2+\Delta Y^2)}\rvert]/\Delta T$	8.758
		G_{27}	5th	$\text{std}\{\lvert Vx\rvert\}$	8.758
		G_{25}	6th	$\text{mean}\{\lvert V_y\rvert\}$	8.758
		G_{24}	7th	$\text{mean}\{\lvert Dy\rvert\}$	8.758
		G_{23}	8th	$\text{std}\{\lvert D_x\rvert\}$	8.758
		G_{22}	9th	std(X)	8.758
		G_{21}	10th	$\text{mean}\{\lvert V_x\rvert\}$	8.758

DOI: https://doi.org/10.52756/ijerr.2024.v42.026

Table 3. 10 Selected Features Ranked using mRMR for User1 Model.

Sl No.	Signers	Feature No	Ranking	Mathematical Descriptions	Scores
01	User1	G_{06}	1st	w	0.693
		G_{30}	2nd	T_{pd}	0.052
		G_{26}	3rd	std(Y)	0.049
		G_{17}	4th	R_{ut}	0.049
		G_{18}	5th	T_s	0.049
		G_{21}	6th	mean{ |Vx| }	0.049
		G_{27}	7th	std{ |Vx| }	0.049
		G_{19}	8th	T_t	0.044
		G_{22}	9th	std(X)	0.044
		G_{15}	10th	N_{up}	0.044

Classifiers Used

Multiple Support Vector Machine (SVM)

Support vector machine (SVM) is a supervised machine learning algorithm used for linear and nonlinear classification (Haloi et al., 2023). In our case, we used SVM for classification to perform signature verification. Since SVM can handle high-dimensional data and nonlinear relationships. It effectively finds the maximum hyperplane that separates the available classes. The main goal is to find the best hyperplane in an N-dimensional space that can be used to separate data points into different classes in the feature space. The hyperplane attempts to maintain the maximum possible margin between the nearest points of various classes. If the data are not possible to separate linearly separable SVM resolves this by creating a new variable using a kernel. The SVM uses Kernel's mathematical function to sketch the original input data into high-dimension feature spaces. So that the hyperplane can be easily found even if the data are not linearly separable in the original input space. Therefore, to get the best signature verification results and for comparison we have used the following kernel functions: Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian and Coarse Gaussian.

Multiple K-Nearest Neighbours (KNN)

The sample data is classified using the K-nearest neighbour (KNN) classifier by allocating it to the class label that more frequently corresponds to its nearest neighbour value, which is k. Decision-making is based on computed distance if a draw situation arises between test samples. The sample will be classified into classes with a smaller distance in comparison to the test sample. KNN performance depends upon the optimum value of k and the distance. Researchers have used various techniques to determine the distance. KNN classifier learns fast, but its classification accuracy is poor. KNN is the smallest classifier compared to all other machine learning algorithms (Kotsiantis et al., 2006).

Experiments Done

We have used 30 global features of all 40 signers. Here, 50% of all 40 signers' data are utilized to train the model using multiple SVM for all 40 users. Once with all 30 feature sets first, Kruskal Wallis selected 25 feature sets, Kruskal Wallis selected 20 feature sets, Kruskal Wallis selected 15 feature sets, and Kruskal Wallis selected 10 feature sets. Similarly, it is done by combining multiple SVM and mRMR feature selection algorithms. Again, the same procedure as above is repeated for the combination of Multiple KNN and Kruskal Wallis feature selection algorithm and the combination of Multiple KNN and mRMR feature selection algorithm. The K-fold Cross Validation (KCV) method is used to select model and estimate the error of the classifiers. The dataset is splitted by KCV into k subsets; then, iteratively, some are used to learn the model, while others are used to assess its performance (Lee et al., 2012). It is seen that the most commonly used value of k is 5(five), as this value provides test error rate assessment that suffer neither from extremely high bias nor from high variance (Nti et al., 2021). In this paper, we have used the numerical value of k as 5 for validation purposes for all 40 signers and evaluated the model.

Here, 100% of all the 40 signers' data are used to test all four combinations of classifiers and feature selection techniques. And finally, accuracy of validation and accuracy of testing for each model with various combinations are shown in Table 8, Table 9, Table 10 and Table 11.

DOI: https://doi.org/10.52756/ijerr.2024.v42.026

Table 4. Model Linear SVM Hyperparameters.

	Parameters	Hyper parameters
40 signers	Preset	Linear SVM
	Multiclass method	One-vs-One
	Box constraint level	1
	Kernel scale	Automatic
	Kernel function	Linear
	Standardize data	Yes

Table 5. Model Quadratic SVM Hyperparameters.

	Parameters	Hyper parameters
40 signers	Preset	Quadratic SVM
	Multiclass method	One-vs-One
	Box constraint level	1
	Kernel scale	Automatic
	Kernel function	Quadratic
	Standardize data	Yes

Table 6. Model Fine KNN Hyperparameters.

	Parameters	Hyper parameters
40 signers	Preset	Fine KNN
	Number of neighbors	1
	Distance metric	Euclidean
	Distance weight	Equal
	Standardize data	Yes

Table 7. Model Weighted KNN Hyperparameters.

	Parameters	Hyper parameters
40 signers	Preset	Weighted KNN
	Number of neighbors	10
	Distance metric	Euclidean
	Distance weight	Squared inverse
	Standardize data	Yes

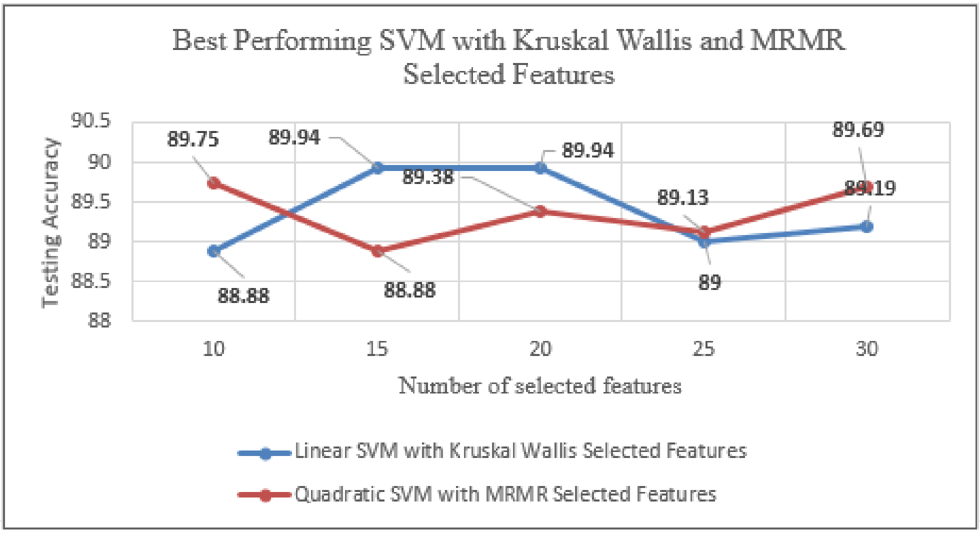

Figure 4. Best performing SVM with Kruskal Wallis and mRMR Selected Features.

Observation and Analysis

Four best-performing model hyper-parameters are shown in Table 4, Table 5, Table 6 and Table 7. Similarly combinations of various classifiers with different feature selection techniques and their performances, comparisons and analysis are shown in Figure 4, Figure 5, Figure 6 and Figure 7. Bar diagram in Figure 8, Figure 9, Figure 10 and Figure 11 clearly depicts that we get high performance with fewer selected features, hence will save computational time and model size.

DOI: https://doi.org/10.52756/ijerr.2024.v42.026

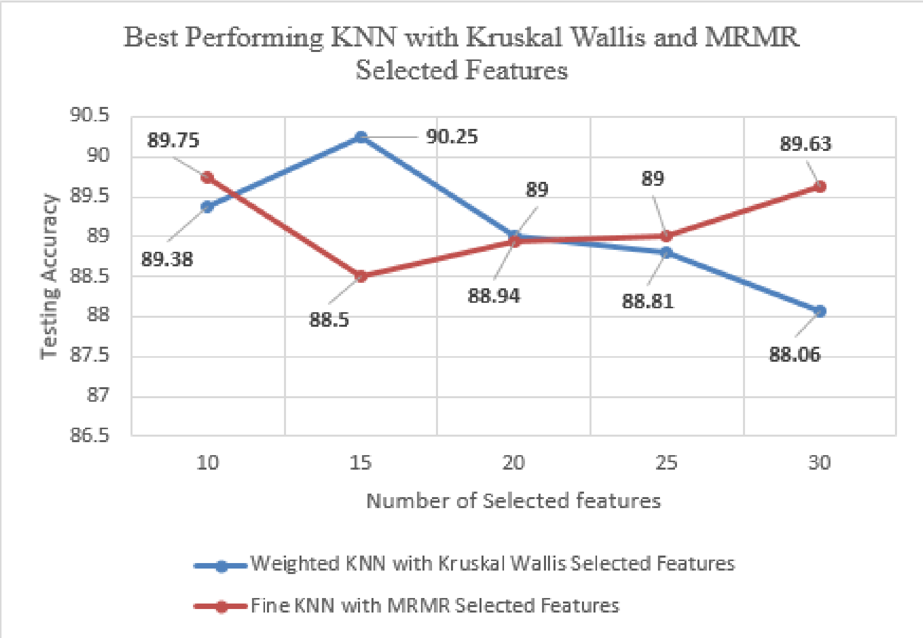

Figure 5. Best Performing KNN with Kruskal Wallis and mRMR Selected Features.

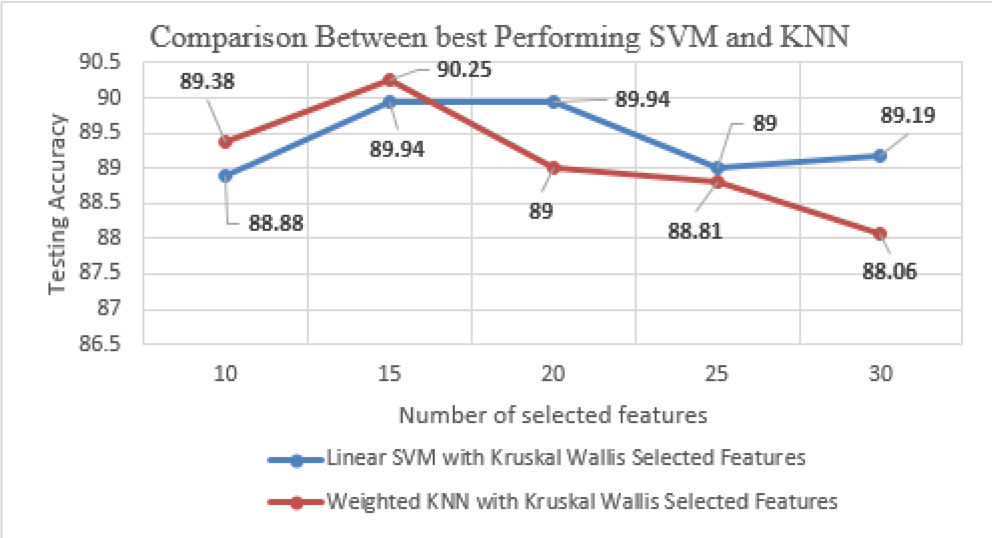

Figure 6. Comparison between best Performing SVM and KNN.

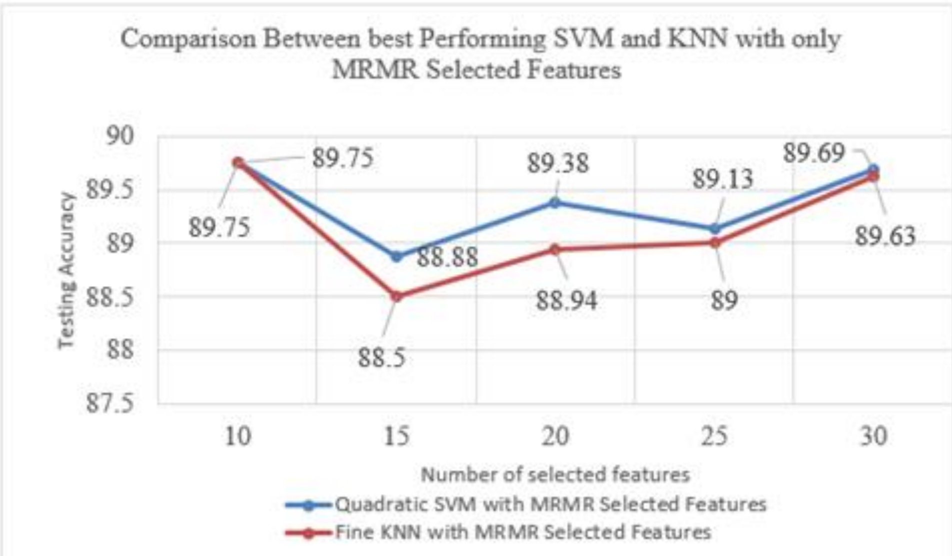

Figure 7. Comparison between best Performing SVM and KNN with only mRMR Selected Features.

DOI: https://doi.org/10.52756/ijerr.2024.v42.026

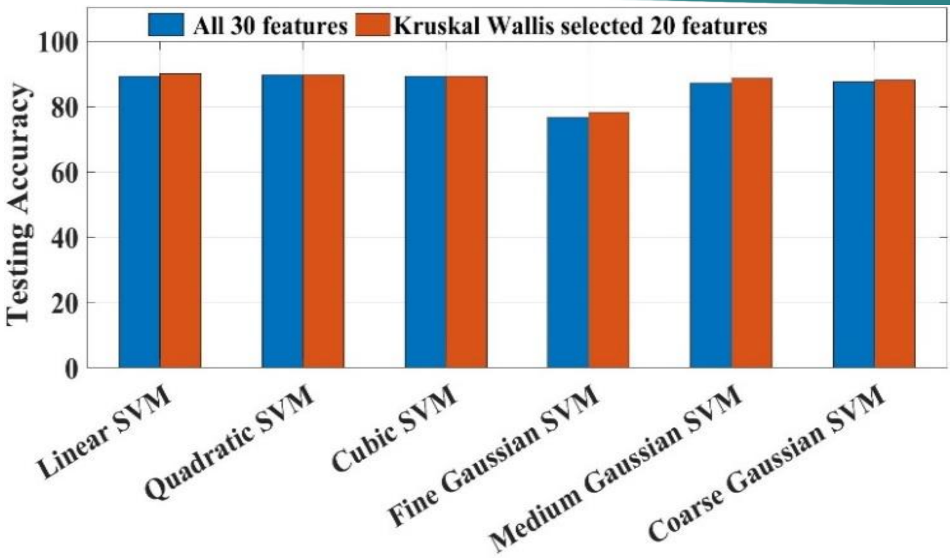

Figure 8. A bar diagram of multiple SVMs representing their difference in testing accuracy using all 30 extracted features, and Kruskal Wallis selected 20 features.

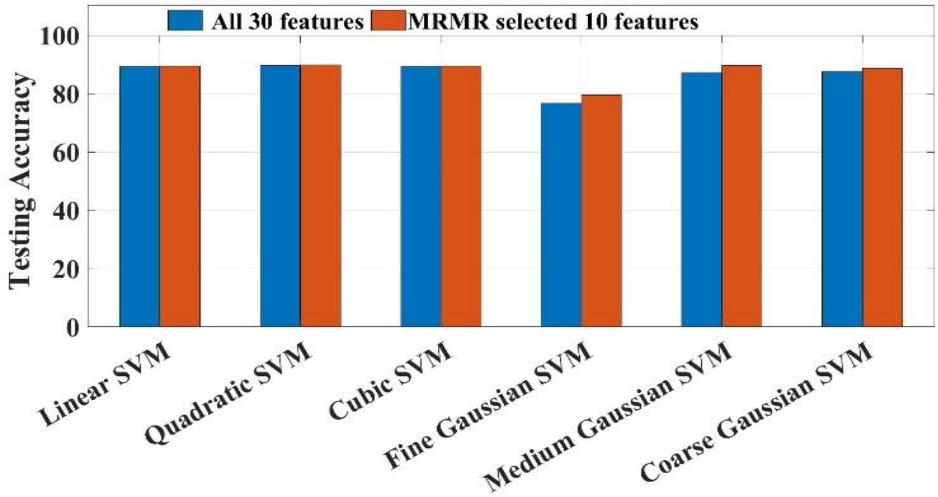

Figure 9. A bar diagram of multiple SVM representing their difference in testing accuracy using all 30 features extracted and 10 features selected by mRMR.

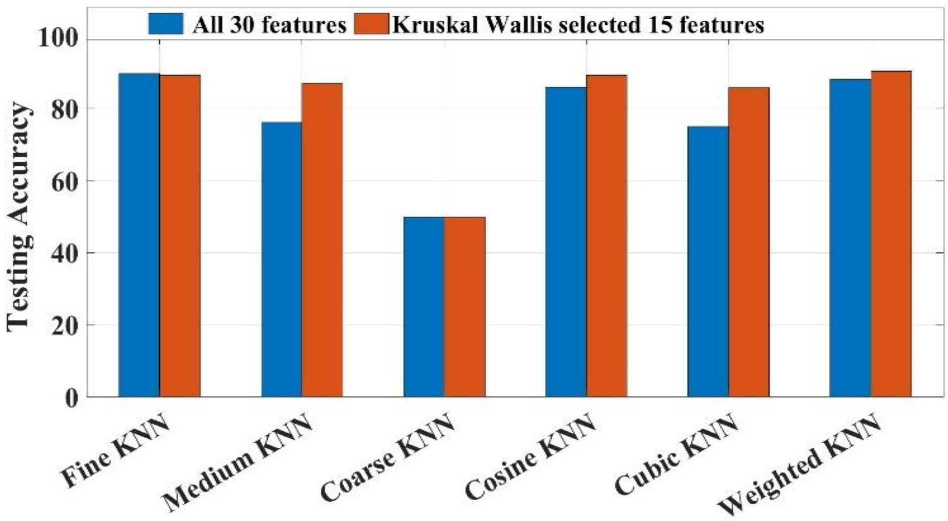

Figure 10. A bar diagram of multiple KNNs representing their difference in testing accuracy using all 30 extracted features, and Kruskal Wallis selected 15 features.

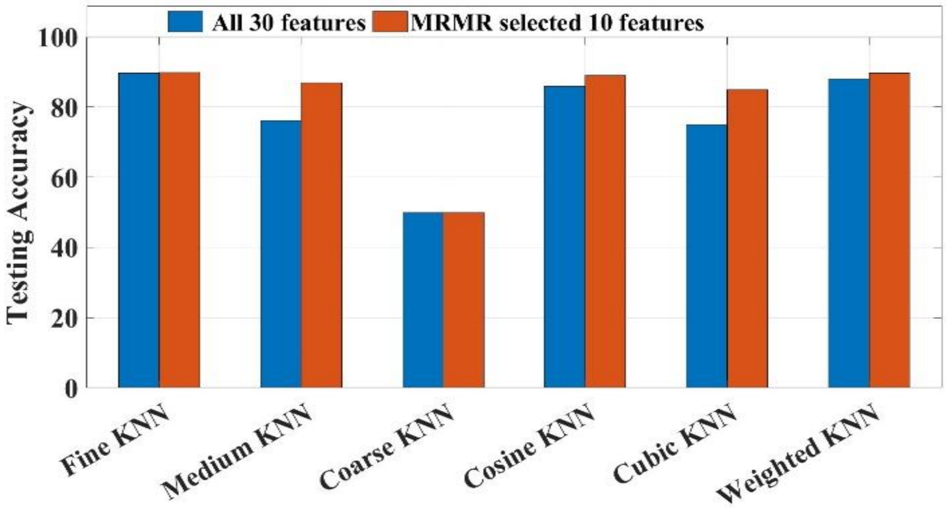

Figure 11. A bar diagram of multiple KNN representing their difference in testing accuracy using all 30 features extracted and 10 features selected by mRMR.

Table 8. Verification Results of Multiple SVM using Different Numbers of Kruskal Wallis Selected Features.

	Model	Using All 30 Features		Selected 25 Features (Using Kruskal Wallis)		Selected 20 Features (Using Kruskal Wallis)		Selected 15 Features (Using Kruskal Wallis)		Selected 10 Features (Using Kruskal Wallis)	
	SVM	*Accuracy% (Validation)*	*Accuracy% (Test)*	*Accuracy% (Validation)*	*Accuracy% (Test)*	*Accuracy% (Validation)*	*Accuracy% (Test)*	*Accuracy% (Validation)*	*Accuracy% (Test)*	*Accuracy% (Validation)*	*Accuracy% (Test)*
40 signers	Linear SVM	96.63	89.19	96.50	89.00	96.50	**89.94**	97.00	**89.94**	97.00	88.88
	Quadratic SVM	97.25	89.69	97.50	89.63	97.38	89.69	97.50	89.56	97.50	89.13
	Cubic SVM	97.38	89.25	97.38	89.31	97.50	89.31	97.50	88.88	97.50	89.00
	Fine Gaussian SVM	77.75	76.63	78.13	77.56	79.00	78.25	83.38	78.75	84.63	81.13
	Medium Gaussian SVM	97.63	87.13	97.75	86.88	97.88	88.69	97.88	88.63	97.25	88.75
	Coarse Gaussian SVM	92.50	87.50	92.13	88.44	93.88	88.19	94.50	89.06	95.13	88.25

Results and Discussion

Results

Here in the developed online signature verification system, training, validation and testing are done for all 40 users once with all the 30 global features extracted and subsequently with various numbers of features selected using Kruskal Wallis and mRMR feature selection algorithm using multiple SVM and multiple KNN classifier. The verification results are shown in Table 8, Table 9, Table 10 and 11. Among all the combinations of SVM best performing SVM is Linear SVM with Kruskal Wallis selected 15 and 20 features which yielded 89.94 % average testing accuracy and Quadratic SVM with mRMR selected 10 features yielded 89.75% average testing accuracy, as shown in Figure 4. Among all the combinations of KNN best performing KNN is Weighted KNN with Kruskal Wallis selected 15 features yielded the highest average testing accuracy of 90.25% and Fine KNN with mRMR selected 10 features yielded 89.75% average testing accuracy as

DOI: https://doi.org/10.52756/ijerr.2024.v42.026

Table 9. Verification results of multiple SVM using different numbers of mRMR selected features.

	Model	Using All 30 Features		Selected 25 Features (Using mRMR)		Selected 20 Features (Using mRMR)		Selected 15 Features (Using mRMR)		Selected 10 Features (Using mRMR)	
	SVM	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***
40 signers	Linear SVM	96.63	89.19	96.63	89.38	96.63	89.50	96.25	88.44	96.25	89.50
	Quadratic SVM	97.25	89.69	97.50	89.13	98.13	89.38	97.75	88.88	97.13	**89.75**
	Cubic SVM	97.38	89.25	97.75	89.13	98.00	89.13	97.38	89.06	97.38	89.50
	Fine Gaussian SVM	77.75	76.63	78.00	78.06	76.38	78.50	77.88	78.88	83.63	79.44
	Medium Gaussian SVM	97.63	87.13	98.13	88.44	98.00	88.69	98.25	88.94	98.25	89.63
	Coarse Gaussian SVM	92.50	87.50	92.88	87.25	94.00	87.81	93.63	87.06	93.50	88.69

Table 10. Verification results of multiple KNN using different numbers of Kruskal Wallis selected features.

	Model	Using All 30 Features		Selected 25 Features (Using Kruskal Wallis)		Selected 20 Features (Using Kruskal Wallis)		Selected 15 Features (Using Kruskal Wallis)		Selected 10 Features (Using Kruskal Wallis)	
	KNN	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***
40 signers	Fine KNN	97.25	89.63	97.50	89.31	97.63	89.06	97.75	89.25	97.63	89.31
	Medium KNN	74.38	76.19	80.88	80.44	85.63	83.88	90.50	86.94	92.88	88.00
	Coarse KNN	50.00	50.00	51.25	50.69	50.00	50.00	50.00	50.00	50.00	50.00
	Cosine KNN	88.38	85.88	89.75	86.94	93.25	87.94	94.25	89.13	96.00	88.38
	Cubic KNN	73.13	75.00	79.50	79.69	84.38	83.31	88.75	85.75	92.88	87.00
	Weighted KNN	93.63	88.06	93.63	88.81	95.63	89.00	96.38	**90.25**	97.25	89.38

DOI: https://doi.org/10.52756/ijerr.2024.v42.026

shown in Figure 5. But if we compare among the best performing SVM and KNN, Weighted KNN with Kruskal Wallis selected 15 features yielded highest average testing accuracy of 90.25% as shown in Figure 6, which outperforms all other combinations proving the combination to be best for signature verification. As shown in Figure 7 SVM and KNN, in combination with mRMR selected best 10 features perform equally well, yielding an accuracy 89.75%. It is pertinent to mention here the importance of feature selection as it is directly related to the performance of the system, as shown in the form of bar diagram in Figure 8, Figure 9, Figure 10 and Figure 11, we get high performance with less number of selected features hence it will save computational time as well as consume less model size.

Table 11. Verification results of multiple KNN using different numbers of mRMR selected features.

	Model	Using All 30 Features		Selected 25 Features (Using mRMR)		Selected 20 Features (Using mRMR)		Selected 15 Features (Using mRMR)		Selected 10 Features (Using mRMR)	
	KNN	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***	***Accuracy% (Validation)***	***Accuracy% (Test)***
40 signers	Fine KNN	97.25	89.63	97.88	89.00	97.38	88.94	98.00	88.50	98.00	**89.75**
	Medium KNN	74.38	76.19	76.50	78.00	78.63	80.94	82.38	82.31	86.38	86.94
	Coarse KNN	50.00	50.00	50.00	50.00	50.00	50.00	51.25	50.75	50.00	50.00
	Cosine KNN	88.38	85.88	89.13	85.69	89.63	86.88	90.38	86.88	91.88	89.00
	Cubic KNN	73.13	75.00	75.13	76.44	76.88	79.13	80.75	81.63	84.88	85.00
	Weighted KNN	93.63	88.06	93.75	88.69	94.50	88.19	95.75	88.00	96.00	89.69

features by Linear SVM and using mRMR, which selected 10 features by Quadratic SVM.

In the case of KNN, the best testing average accuracy for all 40 signers is seen if we apply Kruskal Wallis feature selection to select 15 best global features out of 30 and perform verification using Weighted KNN, yielding highest average testing accuracy of **90.25%** in the SVC 2004 Database (Task 1) and if we apply mRMR feature selection to select only 10 best global features out of 30 and perform verification using Fine KNN yielding average testing accuracy of 89.75% in the SVC 2004 Database (Task 1). It is also seen that the said system even outperforms the system using all 30 global features, and the other system uses various numbers of Kruskal Wallis/mRMR selected global features, as shown in Table 10 and Table 11. Verification results with multiple SVMs combined with Kruskal Wallis and mRMR feature selection are shown in Table 8 and 9. And Verification results with multiple KNN combined with Kruskal Wallis and mRMR feature selection are shown in Table 10 and 11. It is very clear from the observation that mRMR feature selection techniques with the best selected 10 features perform equally well in combination with both SVM and KNN classifiers, yielding similar highest average testing accuracy of 89.75% in both cases, as shown in Figure 7. However, in our case, Kruskal Wallis performed well in combination with Weighted KNN, yielding the highest average testing accuracy of 90.25%. Verification here was a difficult task because of the

Discussion

In the proposed system, in case of SVM best testing average accuracy for all 40 signers is seen if we apply Kruskal Wallis feature selection to select 15 and 20 best global features out of 30 and perform verification using Linear SVM in the SVC 2004 Database (Task 1) and if we apply mRMR feature selection to select only 10 best global features out of 30 and perform verification using Quadratic SVM in the SVC 2004 Database (Task 1). It is also seen that the said system even outperforms the system using all 30 global features, and the other system uses various numbers of Kruskal Wallis/mRMR selected global features, as shown in Table 8 and Table 9. Alternatively, we can conclude that in this case, the best result is shown using Kruskal Wallis, which selected 15 presence of the skilled forgeries, which are hardest to detect.

Conclusion

Here, it is seen that the result shown by a combination of Linear SVM with Kruskal Wallis selected 15 features gives an average testing accuracy of 89.94%, which is the best result yielded when we combine all types of SVM with Kruskal Wallis feature selection technique. On the other hand, the combination of Quadratic SVM with mRMR selected 10 features yielded an average testing accuracy of 89.75%, which is the best result yielded when we combine all types of SVM with mRMR feature selection technique. In combination with Weighted KNN, Kruskal Wallis yielded the highest testing average accuracy of **90.25%** among all other combinations. mRMR feature selection techniques with the best selected 10 features perform equally well when combined with both SVM and KNN classifiers, yielding a similar average testing accuracy of 89.75%. This type of suitable online signature verification system is highly needed to prevent forgery as well as accept genuine users.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

Abe, S. (2010). Multiclass support vector machines. *Support Vector Machines for Pattern Classification*, pp.113-161. https://doi.org/10.1007/978-1-84996-098-4_3

Ali Khan, S., Hussain, A., Basit, A., & Akram, S. (2014). Kruskal-Wallis-Based Computationally efficient feature selection for face recognition. *The Scientific World Journal*, *2014*, 1-6. https://doi.org/10.1155/2014/672630

Chamasemani, F. F., & Singh, Y. P. (2011). Multi-class support vector machine (SVM) classifiers -- An application in hypothyroid detection and classification. *2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications*. https://doi.org/10.1109/bic-ta.2011.51

Diaz, M., Fischer, A., Ferrer, M. A., & Plamondon, R. (2018). Dynamic signature verification system based on one real signature. *IEEE Transactions on Cybernetics*, *48*(1), 228-239. https://doi.org/10.1109/tcyb.2016.2630419

Fierrez-Aguilar, J., Krawczyk, S., Ortega-Garcia, J., & Jain, A. K. (2005). Fusion of local and regional approaches for on-line signature verification. *Advances in Biometric Person Authentication*, pp. 188-196. https://doi.org/10.1007/11569947_24

Haloi, R., Chanda, D., Hazarika, J., & Barman, A. K. (2023). Statistical feature-based EEG signals classification using ANN and SVM classifiers for Parkinson's disease detection. *International Journal of Experimental Research and Review*, *31*(Spl Volume), 141-149. https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.014

Hanchuan Peng, Fuhui Long, & Ding, C. (2005). Feature selection based on mutual information criteria of Max-dependency, Max-relevance, and MIN-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226-1238. https://doi.org/10.1109/tpami.2005.159

Herbst, N. M., & Liu, C. N. (1977). Automatic signature verification based on Accelerometry. *IBM Journal of Research and Development*, *21*(3), 245-253. https://doi.org/10.1147/rd.213.0245

Hermo, J., Bolón-Canedo, V., & Ladra, S. (2024). Fed-mrmr: A lossless federated feature selection method. *Information Sciences*, *669*, 120609. doi:10.1016/j.ins.2024.120609

Hong, J., & Cho, S. (2008). A probabilistic multi-class strategy of one-vs.-rest support vector machines for cancer classification. *Neurocomputing*, *71*(16-18), 3275-3281. https://doi.org/10.1016/j.neucom.2008.04.033

Impedovo, D., & Pirlo, G. (2008). Automatic signature verification: The state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *38*(5), 609-635. https://doi.org/10.1109/tsmcc.2008.923866

Kar, B., & K. Dutta, P. (2012). SVM based signature verification by fusing global and functional features. *International Journal of Computer Applications*, *60*(16), 34-39. https://doi.org/10.5120/9778-4359

Kar, B., Mukherjee, A., & Dutta, P. K. (2018). Stroke point warping-based reference selection and verification of online signature. *IEEE Transactions on Instrumentation and Measurement*, *67*(1), 2-11. https://doi.org/10.1109/tim.2017.2755898

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, *26*(3), 159-190. https://doi.org/10.1007/s10462-007-9052-3

Lee, J. A., Schneider, P., & Quinn, J. (2012). Advances in artificial neural networks, machine learning, and computational intelligence (ESANN 2011).

Neurocomputing, *90*, 1-2. https://doi.org/10.1016/j.neucom.2012.03.002

Lee, L., Berger, T., & Aviczer, E. (1996). Reliable online human signature verification systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(6), 643-647. https://doi.org/10.1109/34.506415

Mary, P. N., & Nagarajan, R. (2024). Feature selection in Microarray using proposed hybrid minimum redundancy-maximum relevance (MRMR) and modified genetic algorithm (MGA). *International Journal of Experimental Research and Review*, *39*(Spl Volume), 82-91. https://doi.org/10.52756/ijerr.2024.v39spl.006

Najda, D., & Saeed, K. (2022). Impact of augmentation methods in online signature verification. *Innovations in Systems and Software Engineering*. https://doi.org/10.1007/s11334-022-00464-4

Nalwa, V. (1997). Automatic on-line signature verification. *Proceedings of the IEEE*, *85*(2), 215-239. https://doi.org/10.1109/5.554220

Napa Sae-Bae, & Memon, N. (2014). Online signature verification on mobile devices. *IEEE Transactions on Information Forensics and Security*, *9*(6), 933-947. https://doi.org/10.1109/tifs.2014.2316472

Nti, I. K., Nyarko-Boateng, O., & Aning, J. (2021). Performance of machine learning algorithms with different K values in K-fold Cross Validation. *International Journal of Information Technology and Computer Science*, *13*(6), 61-71. https://doi.org/10.5815/ijitcs.2021.06.05

Parmar, M., Puranik, N., Joshi, D., & Malpani, S. (2020). Image processing based signature duplication and its verification. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3645426

Plamondon, R., & Lorette, G. (1989). Automatic signature verification and writer identification — the state of the art. *Pattern Recognition*, *22*(2), 107-131. https://doi.org/10.1016/0031-3203(89)90059-9

Plamondon, R., & Srihari, S. (2000). Online and off-line handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 63-84. https://doi.org/10.1109/34.824821

Rao, C., & Wu, Y. (2005). Linear model selection by cross-validation. *Journal of Statistical Planning and Inference*, *128*(1), 231-240. https://doi.org/10.1016/j.jspi.2003.10.004

Yeung, D., Chang, H., Xiong, Y., George, S., Kashi, R., Matsumoto, T., & Rigoll, G. (2004). SVC2004: First international signature verification competition. *Biometric Authentication*, 16-22. https://doi.org/10.1007/978-3-540-25948-0_3

How to cite this Article:

Bhimraj Prasai Chetry and Biswajit Kar (2024). Kruskal Wallis and mRMR Feature Selection based Online Signature Verification System using Multiple SVM and KNN. *International Journal of Experimental Research and Review*, *42*, 298-311.

DOI : https://doi.org/10.52756/ijerr.2024.v42.026