An overlapping conscious relief-based feature subset selection method

Nishat Tasnim Mim[1], Md. Eusha Kadir[2], Suravi Akhter[3], Muhammad Asif Hossain Khan[4]
[1]Institute of Information Technology, University of Dhaka, Dhaka, Bangladesh
[2]Institute of Information Technology, Noakhali Science and Technology University, Noakhali, Bangladesh
[3]Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, Dhaka, Bangladesh
[4]Department of Computer Science and Engineering, Faculty of Engineering and Technology, University of Dhaka, Dhaka, Bangladesh

Article Info

Article history:

Received Apr 21, 2023
Revised Sep 6, 2023
Accepted Nov 4, 2023

Keywords:

Class overlapping
Distance-based method
Feature selection
Relief-based method
Reward-penalty

ABSTRACT

Feature selection is considered as a fundamental prepossessing step in various data mining and machine learning based works. The quality of features is essential to achieve good classification performance and to have better data analysis experience. Among several feature selection methods, distance-based methods are gaining popularity because of their eligibility in capturing feature interdependency and relevancy with the endpoints. However, most of the distance-based methods only rank the features and ignore the class overlapping issues. Features with class overlapping data work as an obstacle during classification. Therefore, the objective of this research work is to propose a method named overlapping conscious MultiSURF (*OMsurf*) to handle data overlapping and select a subset of informative features discarding the noisy ones. Experimental results over 20 benchmark dataset demonstrates the superiority of *OMsurf* over six existing state-of-the-art methods.

Corresponding Author:

Muhammad Asif Hossain Khan
Department of Computer Science and Engineering, Faculty of Engineering and Technology, University of Dhaka
Nilkhet Road, Dhaka 1000, Bangladesh
Email: asif@du.ac.bd

1. INTRODUCTION

We are now living in the age of modern technologies. The rapid growth and wide use of technologies generate a huge amount of data which imposes a challenge for the data scientists and engineers to manage these data in an effective and efficient way. Data engineers usually find patterns and relationships after analyzing the data with the assistance of data mining and machine learning techniques. However, an issue named curse of dimensionality can be found in high-dimensional data which may mislead the learning phase of machine learning techniques. To achieve the desired performance in a variety of application domains (e.g., bioinformatics, life science, health care, and cyber security), the high-dimensional data often need to be pre-processed before applying machine learning techniques [1]–[4]. One of such data pre-processing techniques is feature selection.

Feature selection is the process of identifying informative features and eliminating the noisy and un-useful features from the original feature set. Over the years, several feature selection methods have been proposed which can be broadly grouped into two types namely wrapper and filter method. The wrapper methods search for an optimal feature set based on a specific machine learning technique [5]–[7]. And thus, this feature selection procedure is highly dependent on the nature of the machine learning technique. On the contrary, filter methods use statistical approaches to measure the dependency between feature and target, and the relations among features [8]. Among the filter methods, mutual information based feature selection approaches (such as minimum-redundancy maximum-relevance [9], Joint mutual information [10], mDSM [11], min-redundancy and max-dependency) [12] and distance-based approaches (such as *ReliefF* [13], spatially uniform *ReliefF* (*SURF*) [14], multiple threshold spatially uniform *ReliefF* (*MultiSURF*) [15], modified relief (*mRelief*) [16], statistical inference relief (*STIR*) [17]) have gained more popularity as these methods can capture both linear and non-linear relationship among the features.

The limitation of mutual information based methods are observed in capturing feature interaction whereas distance methods efficiently handle it both in binary and multiclass problems [11], [18]. Distance methods consider all the features of a particular (target) instance at a time and compute the distance between that target and other instances [13], [16], [19]–[21]. Instead of considering all other instances, these methods apply k-nearest neighbor (kNN) search strategy. With the assistance of k closest neighbors of an arbitrary instance I_n, the density of I_n, denoted by $p(I_n)$, can be estimated using (1) where N is the total number of instances and V is the volume.

$$p(I_n) = \frac{k/_N}{V} \tag{1}$$

The density, $p(I_n)$ can be approximated in two ways. A way is to fix the number of neighborhood instances (k) and find the volume (V) that contains k points inside. Another way specifies the volume (V) and determines the number of points (k) inside V. Instead of $p(I_n)$, we are more interested in approximating the posterior $p(y_n|I_n)$ where y_n is the actual class label of I_n. This posterior gives the assurance of similar data points belonging to the same class which can be expressed using (2).

$$p(y_n|I_n) \approx \frac{k_n/_N}{V\sum_{j=1}^{n^c}\frac{k_j/_N}{V}} = \frac{k_n}{K} \tag{2}$$

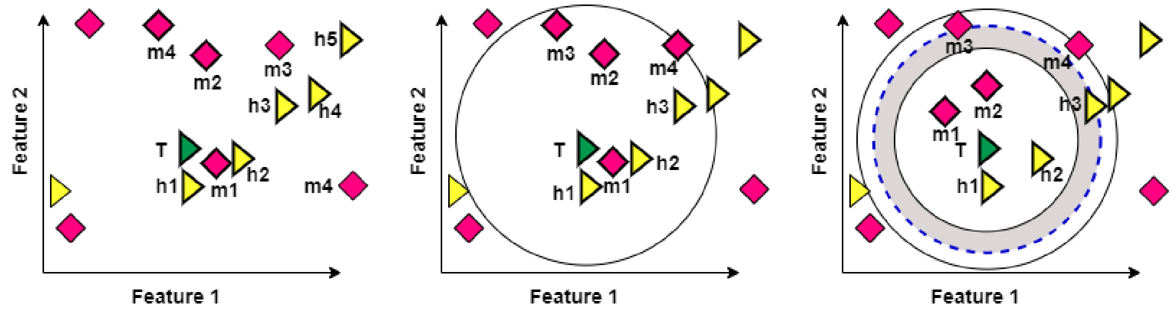
where k_n is the number of neighbors of I_n belonging in y_n (same class as I_n), K is the total number of neighbors inside a fixed volume, $I_n \leq K$, and n^c denotes the number of class labels.

However, probability approximation from k_n/K is somewhat questionable as it cannot explain the compactness of the same class instances as well as the diversity of different class instances. Probability approximation from distance will be more appropriate to measure compactness and disperseness. The distance-based methods such as *ReliefF* and *mRelief* calculate nearest hit and miss to estimate $p(y_n|I_n)$ and $P(y_{\neq n}|I_n)$ respectively. Here, a hit is a neighborhood instance belonging to the same class as I_n and a miss is a neighborhood instance belonging to the other class from I_n.

One limitation of k neighborhood search methods is that these methods often discard some informative instances as well as select some noisy instances to meet the condition of k. The traditional distance-based methods such as *ReliefF* and *mRelief* exhibit limitations as depicted in Figure 1(a), when dealing with distant instances ($h4$ and $h5$) relative to the target (T) in order to satisfy the condition of k ($k = 5$). However, it is important to note that these instances do not provide meaningful information as these instances cannot assure the effective class separation for the target. Both kNN and threshold-based methods work fine in case of clearly separable scenarios which is rarely practical. Although threshold-based methods [14], [15] can accumulate informative instances by discarding the uninformative ones, these cannot prioritize the features in data overlapping scenario.

Let us consider Figures 1(b) and 1(c) where T is the target instance, $h1$, $h2$, $h3$ and $m1$, $m2$, $m3$, and $m4$ are hit and miss neighbors of T respectively. In Figure 1(b) miss neighbor $m1$ is inside hit data which indicates data overlap issue. In feature selection data overlapping problem should be addressed and therefore a solution should be provided for radius-based algorithms as shown in Figure 1(c). Let us consider the moves $m1$ and $m2$, depicted distantly from the target as in Figure 1(c). The radius-based methods give importance in case of Figure 1(b) compared to Figure 1(c) although Figure 1(c) deserves higher score for its class separability.

To address the data overlapping issue, *mRelief* has been proposed. However, it selects a fixed kNN which might consider uninformative instances for feature performance estimation. Besides this, most of the aforementioned methods are feature ranking methods and depend on the user input to adjust the number of features. Addressing these aforementioned issues, a new distance-based feature subset selection method named Overlapping conscious MultiSURF (*OMsurf*) has been proposed in this paper having the following key contributions. i) *OMsurf* introduces a reward and penalty mechanism to handle data overlapping while ranking features; ii) Rather than only ranking the features, *OMsurf* selects an optimal feature subset from top ranked features; and iii) Rigorous experiments on numerous datasets are performed to demonstrate the superiority of *OMsurf* over the existing state-of-the-art methods.



(a) (b) (c)

Figure 1. Limitations of distance-based method in (a) traditional kNN, (b) overlapped scenario, and (c) non-overlapped scenario

2. METHOD

The primary objective of the proposed method is to select a subset of features having the capability of discriminating actual class. Our proposed *OMsurf* is a class overlap conscious distance-based method prioritizes and selects discriminating feature subset. In this section, the major two parts of *OMsurf* namely prioritization of features, and feature subset selection will be described.

2.1. Prioritization of features

Feature prioritization is considered as an initial step of feature selection process. The objective is to minimize the within class scatter and increase between class scatter in the neighborhood section using (3).

$$S[f_k] \coloneqq S[f_k] + \frac{1}{N}\left(\frac{1}{|M_n|}\sum_{c=1,c\neq yn}^{n^c}\frac{|M_n^C|}{|M_n|}\,\Delta\big(I_n(f_k), M_n(f_k)\big) - \frac{1}{|H_n|}\,\Delta\big(I_n(f_k), H_n(f_k)\big)\right) \tag{3}$$

where n^c is the number of classes, $|\cdot|$ counts the number of hit and miss instances and $\Delta(\cdot)$ measures the distance between two vectors, f_k denotes the k^{th} feature of a feature vector. In relation to the problem addressed in Figure 1, a new feature weighting technique has been introduced where features with overlapping data will get penalty and non-overlapping data will get reward. The concept of reward is defined in order to raise the feature weight based on exemplary boundary where the boundary is defined as the absolute difference between the max hit boundary with the standard deviation of hit distances (as can be seen in Figure 1(c). The penalty term is defined based on the number of miss inside hit and total number of miss instances. The definition of reward and penalty can be stated as follows.

Definition 1. *Reward measures the presence of target class instance in the exemplary boundary with respect to the overall count of hit instances which is defined in (4).*

$$R_n = \frac{|H_{exemplar}|}{|H_n|} \tag{4}$$

where $|\cdot|$ denotes the cardinality of a vector, Hn denotes a vector containing the hit instances of an instance (I_n) and $H_{exemplar}$ represents the vector of hit instances located within the exemplary boundary.

Definition 2. *Penalty measures the number of overlapping miss instances in the exemplary boundary in relation to the total number of miss instances which is defined in (5) where $M_{exemplar}$ represents the miss instances located inside the exemplary boundary.*

$$P_n = \frac{|M_{exemplar}|}{|M_n|} \tag{5}$$

By incorporating reward and penalty in (3), we propose a scoring mechanism for feature ranking which is defined in (6) and (4) is formed such a way that the greater value of reward assigned to a feature (indicating the compactness of hit instances) will lead to a large value in feature score.

$$S[f_k] \coloneqq S[f_k] + \frac{1}{N}\left(\frac{1}{|M_n|}\sum_{c=1,c\neq yn}^{n^c}\frac{|M_n^C|}{|M_n|}\,\Delta\big(I_n(f_k), M_n(f_k)\big) \times R_n - \frac{1}{|H_n|}\,\Delta\big(I_n(f_k), H_n(f_k)\big) \times P_n\right) \tag{6}$$

Besides, an increased penalty value (measuring the overlapping situation) will result in the reduction of the overall feature score. After computing all the feature scores, scores are sorted in a descending order to obtain the ranked feature set (F_R). The detailed approach of feature prioritization is given in Algorithm 1.

Algorithm 1. Prioritization of features
Input: Instances $I=\{I_1, I_2, \ldots, I_N\}$, features, $F=\{f_1, f_2, \ldots, f_n\}$ and label, $Y=\{y_1, y_2, \ldots, y_n\}$
Output: Ranked feature set, (F_R)

Set each k^{th} feature score, $S[f_k] := 0$
 for $n = 1$ to N do
  $H_n \leftarrow \Phi$, $M_n \leftarrow \Phi$
Compute Θ_n following [15]
For j=1 to N do
  If $\Delta(I_n, I_j) < \Theta_n$ then
   If $y_n = y_n$ then
    $H_n \leftarrow H_n \cup I_j$
   else
    $M_n \leftarrow M_n \cup I_j$
   end if
  end if
  for k=1: M do
   Compute $S[f_k]$ following (6)
  end for
 end for
end for
$F_R \leftarrow$ sort S in descending order
return F_R

2.2. Feature subset selection

We utilize a greedy forward selection approach to select relevant features with class discrimination capability. Initially the top-ranked feature, denoted by $F_R(1)$, is chosen into selected subset (F_S), and then the combined performance of remaining features in F_R are taken sequentially to measure their combined performance ($F_{S,k}$) following (7).

$$S[F_{S,k}] := S[F_{S,k}] + \frac{1}{N}\left(\frac{1}{|M_n|}\sum_{c=1,c\neq yn}^{n^c}\frac{|M_n^c|}{|M_n|}\,\Delta\left(I_n(F_{S,k}), M_n(F_{S,k})\right)\times R_n - \frac{1}{|H_n|}\,\Delta\left(I_n(F_{S,k}), H_n(F_{S,k})\right)\times P_n\right) \quad (7)$$

When assessing the combined performance of a candidate feature f_k and the already selected feature subset F_S for a target instance (i.e., $S[F_{S,k}]$), it obtains a higher score if its neighboring instances from the same class are in close proximity to the target, while instances from other classes remain distant from the target. We choose the feature f_k in a way that maximizes the value of $S[F_{S,k}]$, which enhances the class separability. The detail of the feature subset selection process is provided in Algorithm 2.

Algorithm 2. Feature subset selection
Input: Instances $I=\{I_1, I_2, \ldots, I_N\}$, label, $Y=\{y_1, y_2, \ldots, y_N\}$, all features F, and ranked feature set, F_R
Output: Selected subset of feature, (F_S), where $F_S \subseteq F$

$F_S \leftarrow F_R$ (1)
Threshold value, $\alpha \leftarrow \infty$
for k=2: $|F|$ do
 for n=1 to N do
  Compute $S[F_S, F_R(k)]$ following (7)
 end for
 if $S[F_s, F_R(k)] > \alpha$ then
  $F_S \leftarrow F_S \cup F_R(k), \alpha \leftarrow S[F_S, F_R(k)], F_R \leftarrow F_R \backslash F_R(k)$
 end if
end for
return F_S

3. RESULT AND DISCUSSION

In this section, we primarily present the experimental results. Before presenting the obtained results, we briefly describe the dataset used in this work for performance evaluation along with the experimental setup. Finally, we analyze the obtained result compared to existing state-of-the-art methods.

3.1. Dataset description and implementation detail

We use twenty benchmark datasets collected from UCI machine learning repository [22] that are widely used in the existing literature [20], [23], [24], to evaluate the performance of *OMsurf* and other existing state-of- the-art methods. A detailed description of datasets is given in Table 1 in the first four columns. The proposed *OMsurf* is compared with other distance-based methods such as *ReliefF*, *SURF*, *SURF**, *MultiSURF** (*M.SURF**), *MultiSURF* (*M.SURF*) and *mRelief* to show the excellency of *OMsurf*.

We conduct 10-fold cross validation and apply support vector machine (SVM) using linear kernel as classification algorithm.

Table 1. Accuracy comparison of proposed *OMsurf* and other existing state-of-the-art methods. Bold faces represent the overall best performing method for a particular dataset

Dataset	Instance	feature	class	*ReliefF*	*SURF*	*SURF**	*M.SURF**	*M.SURF*	*mRelief*	*OMsurf*
Hepatitis	80	19	2	77.78	74.44	**81.11**	**81.11**	74.44	77.78(6)	**81.11(7)**
Appendicitis	106	7	2	82.50	80.00	80.00	85.83	80.83	**87.50(6)**	**87.50(4)**
Iris	150	4	3	**96.00**	**96.00**	**96.00**	**96.00**	**96.00**	**96.00(4)**	**96.00(4)**
Hayes-Roth	160	4	3	51.00	51.11	51.11	**53.33**	**53.33**	51.67(3)	52.22(3)
Wine	178	13	3	96.50	97.00	97.89	96.84	98.95	97.37(11)	**98.95(10)**
Parkinsons	195	22	2	84.50	84.50	84.50	84.50	84.50	85.00(11)	**85.60(12)**
Sonar	208	60	2	75.00	77.27	**78.64**	**78.64**	75.45	75.00(18)	78.27(17)
Newthyroid	215	5	3	96.36	96.36	96.36	93.18	96.36	94.55(3)	**96.53(3)**
Spectfheart	267	44	2	78.93	80.00	79.64	79.64	77.86	78.57(13)	**81.07(19)**
Cleveland	297	13	5	54.38	51.88	51.88	51.88	54.38	55.31(10)	**55.39(10)**
Ecoli	336	7	8	83.00	83.63	82.63	83.16	83.16	83.16(6)	**83.68(6)**
Ionosphere	351	34	2	83.61	80.83	80.83	80.83	80.83	85.00(16)	**86.28(22)**
Dermatology	358	34	6	96.84	97.37	97.63	97.63	97.63	95.50(15)	**98.63(31)**
Led7digit	500	7	10	69.64	69.82	69.82	68.00	69.82	66.55(7)	**70.36(5)**
Wdbc	569	30	2	96.55	97.41	96.55	96.55	97.59	97.07(14)	**97.09(19)**
Australian	690	14	2	**87.57**	**87.57**	69.14	69.14	**87.57**	87.14(9)	**87.57(7)**
Pima	768	8	2	73.40	75.19	75.19	67.66	75.32	**75.58(6)**	**75.58(4)**
Breast	990	30	2	96.21	96.72	95.86	95.86	96.55	96.03(13)	**97.24(15)**
Contraceptive	1473	9	3	44.77	46.64	46.78	46.78	51.34	**51.75(5)**	51.70(6)
Yeast	1484	8	10	**58.63**	**58.63**	**58.63**	58.30	58.30	57.84(7)	55.69(7)
Win/tie/loss	**-**	**-**	**-**	**17/2/1**	**17/2/1**	**16/2/2**	**15/2/3**	**15/3/2**	**15/3/2**	**-**

To evaluate performance, we use accuracy as performance metric. However, accuracy alone cannot determine the excellency of feature subset selection method [11]. For better performance approximation, a new measure called *Score* is computed from accuracy and number of selected features which is defined in (8).

$$Score = \alpha \times \frac{|F|-|F_S|}{|F|} + \beta \times Accuracy \tag{8}$$

where $|F|$, $|F_S|$ and α, β are the number of total features, selected features and tuning parameters respectively. Here, we use $\alpha = \beta = 0.5$ to give equal importance on both number of selected features and accuracy.

3.2. Result analysis

The classification accuracy of *OMsurf* and competitive state-of-the-art methods are reported in Table 1 along with the number of selected features. Note that, the number of features for competitive method are kept same as *OMsurf* except *mRelief* as it is feature subset selection method. The last line of Table 1 presents win/tie/loss which indicates the number of datasets in which *OMsurf* performs better, equal, and worse than the compared method respectively. From Table 1, it can be observed that *OMsurf* performs better in 15 datasets and worse in two datasets compared to *MultiSURF* and *mRelief* in least case, whereas *OMsurf* has won in 17 cases and defeated in only one dataset in the best case compared to *ReliefF* and *SURF*. In a few cases, *OMsurf* performs a bit lower compared to state-of-the-art methods.

To understand the class separability, we consider the t-distributed stochastic neighbor embedding (tSNE) visualization of Pima dataset illustrated in Figure 2. Figure 2(a) depicts the two-dimensional representation of Pima dataset using tSNE in the original feature space, encompassing all available features. Conversely, Figure 2(b) illustrates the dataset using tSNE within the selected feature space after applying *OMsurf*. We can see that *Pima* has a substantial amount of data overlapping between classes in the original space. In contrast, *OMsurf* selects only four features from the original feature space and discards the noisy features. In this way, *OMsurf* achieves a better class separability that can be clearly observed in Figure 2(b). *OMsurf* achieves a better performance compared to *ReliefF*, and *mRelief*. As kNN based methods such as *ReliefF* and *mRelief* have to select a fixed number of neighbors to meet the value of k, these methods may select less informative distant instances. On the contrary, *OMsurf* finds neighbors from a short distance by employing a radius-based approach which leads it to enhance class separability.

As we have previously mentioned, only accuracy cannot determine the excellence of feature subset selection method. In Table 2, we report the average ranking of feature subset selection based on score defined in (8) that measures the quality of a feature subset selection method. Analyzing Table 2, we find that *OMsurf* achieves the highest rank (as minimum rank is the highest rank) for achieving better performance using optimal feature subset. *mRelief* is the closest competitor of *OMsurf*. Although *mRelief* selects less features in many cases compared to *OMsurf*, *mRelief* discards some informative features which is a reason of the compromise of its performance. And thus, *OMsurf* can be considered superior over it.

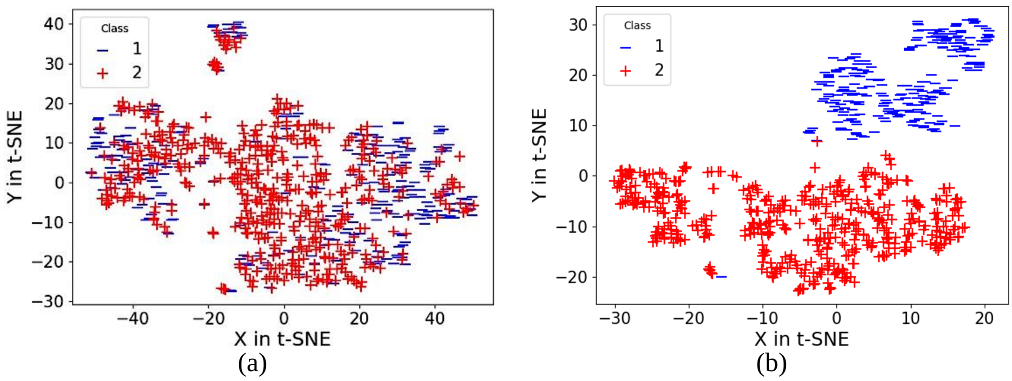


Figure 2. Data visualization using tSNE [25] for Pima dataset (a) before applying *OMsurf* and (b) after applying *OMsurf*; class-1 and class-2 represent the two classes

Table 2. Average ranking of methods based on score

	OMsurf	*mRelief*	*M.SURF*	*ReliefF*	*SURF*	*SURF**	*M.SURF**
Score	0.5818	0.5804	0.5750	0.5734	0.5732	0.5701	0.5688
Rank	1	2	3	4	5	6	7

4. CONCLUSION

In this article, we have proposed *OMsurf* which is an overlapped conscious distance-based feature subset selection method. The objective of this method is to prioritize features based on their relevance, and class separation capability. To achieve this, a mechanism applying the concept of reward and penalty is introduced as feature weighting techniques. Finally, *OMsurf* finds the optimal feature subset based on combined feature performance. The experimental results over twenty datasets show the superiority of *OMsurf* over existing state-of-the-art methods where *OMsurf* outperforms in most datasets. However, *OMsurf* suffers in case of large sample size data due to computing pairwise distances for all instances. To deal with large sample size data, sample selection techniques can be introduced which will be addressed in future.

ACKNOWLEDGEMENT

This research is supported by a grant 56.00.0000.052.33.004.22-34 from the ICT Division, Ministry of Posts, Telecommunications and Information Technology, Bangladesh. The authors are thankful to Mohammad Shoyaib for his sincere observations and insightful comments on the manuscript.

REFERENCES

[1] Z. Chen *et al.*, "Feature selection may improve deep neural networks for the bioinformatics problems," *Bioinformatics*, vol. 36, no. 5, pp. 1542–1552, Mar. 2020, doi: 10.1093/bioinformatics/btz763.

[2] R. Sagban, H. A. Marhoon, and R. Alubady, "Hybrid bat-ant colony optimization algorithm for rule-based feature selection in health care," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, pp. 6655–6663, Dec. 2020, doi: 10.11591/ijece.v10i6.pp6655-6663.

[3] L. Zheng, H. Wang, and S. Gao, "Sentimental feature selection for sentiment analysis of Chinese online reviews," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 1, pp. 75–84, Jan. 2018, doi: 10.1007/s13042-015-0347-4.

[4] Z. Halim *et al.*, "An effective genetic algorithm-based feature selection method for intrusion detection systems," *Computers & Security*, vol. 110, Nov. 2021, doi: 10.1016/j.cose.2021.102448.

[5] D. Banerjee, B. Chatterjee, P. Bhowal, T. Bhattacharyya, S. Malakar, and R. Sarkar, "A new wrapper feature selection method for language-invariant offline signature verification," *Expert Systems with Applications*, vol. 186, Dec. 2021, doi: 10.1016/j.eswa.2021.115756.

[6] G. Sahebi, P. Movahedi, M. Ebrahimi, T. Pahikkala, J. Plosila, and H. Tenhunen, "GeFeS: A generalized wrapper feature

selection approach for optimizing classification performance," *Computers in Biology and Medicine*, vol. 125, Oct. 2020, doi: 10.1016/j.compbiomed.2020.103974.

[7] R. Sawhney, P. Mathur, and R. Shankar, "A firefly algorithm based wrapper-penalty feature selection method for cancer diagnosis," in *Computational Science and Its Applications–ICCSA 2018: 18th International Conference, Melbourne, VIC, Australia, July 2-5, 2018, Proceedings, Part I 18*, Melbourne, 2018, pp. 438–449.

[8] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of Biomedical Informatics*, vol. 85, pp. 189–203, Sep. 2018, doi: 10.1016/j.jbi.2018.07.014.

[9] H. Peng, Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.

[10] H. Yang and J. Moody, "Data visualization and feature selection: New algorithms for nongaussian data," *Advances in neural information processing systems*, 1999.

[11] S. Sharmin, M. Shoyaib, A. A. Ali, M. A. H. Khan, and O. Chae, "Simultaneous feature selection and discretization based on mutual information," *Pattern Recognition*, vol. 91, pp. 162–174, Jul. 2019, doi: 10.1016/j.patcog.2019.02.016.

[12] W. Gao, L. Hu, and P. Zhang, "Feature redundancy term variation for mutual information-based feature selection," *Applied Intelligence*, vol. 50, no. 4, pp. 1272–1288, Apr. 2020, doi: 10.1007/s10489-019-01597-z.

[13] I. Kononenko, "Estimating attributes: Analysis and extensions of relief," in *European conference on machine learning*, 1994, pp. 171–182.

[14] C. S. Greene, N. M. Penrod, J. Kiralis, and J. H. Moore, "Spatially uniform relieff (SURF) for computationally-efficient filtering of gene-gene interactions," *BioData Mining*, vol. 2, no. 1, Dec. 2009, doi: 10.1186/1756-0381-2-5.

[15] R. J. Urbanowicz, R. S. Olson, P. Schmitt, M. Meeker, and J. H. Moore, "Benchmarking relief-based feature selection methods for bioinformatics data mining," *Journal of Biomedical Informatics*, vol. 85, pp. 168–188, Sep. 2018, doi: 10.1016/j.jbi.2018.07.015.

[16] S. Akhter, S. Sharmin, S. Ahmed, A. A. Sajib, and M. Shoyaib, "*mRelief*: A reward penalty based feature subset selection considering data overlapping problem," *International Conference on Computational Science*, pp. 278–292, 2021, doi: 10.1007/978-3-030-77961-0_24.

[17] T. T. Le, R. J. Urbanowicz, J. H. Moore, and B. A. McKinney, "Statistical inference relief (STIR) feature selection," *Bioinformatics*, vol. 35, no. 8, pp. 1358–1365, Apr. 2019, doi: 10.1093/bioinformatics/bty788.

[18] X. Gu, J. Guo, L. Xiao, and C. Li, "Conditional mutual information-based feature selection algorithm for maximal relevance minimal redundancy," *Applied Intelligence*, vol. 52, no. 2, pp. 1436–1447, Jan. 2022, doi: 10.1007/s10489-021-02412-4.

[19] D. R. Munirathinam and M. Ranganadhan, "A new improved filter-based feature selection model for high-dimensional data," *The Journal of Supercomputing*, vol. 76, no. 8, pp. 5745–5762, Aug. 2020, doi: 10.1007/s11227-019-02975-7.

[20] X. Cui, Y. Li, J. Fan, and T. Wang, "A novel filter feature selection algorithm based on relief," *Applied Intelligence*, vol. 52, no. 5, pp. 5063–5081, Mar. 2022, doi: 10.1007/s10489-021-02659-x.

[21] K. Kira and L. Rendell, "The feature selection problem: traditional methods and a new algorithm," *Proceedings of the tenth national conference on Artificial intelligence*, pp. 129–134, 1992, doi: 10.5555/1867135.1867155.

[22] A. Asuncion and D. Newman, "UCI machine learning repository," UCI, 2017.

[23] B. Draper, C. Kaito, and J. Bins, "Iterative relief," in *2003 Conference on Computer Vision and Pattern Recognition Workshop*, Jun. 2003, pp. 62–62, doi: 10.1109/CVPRW.2003.10065.

[24] Z. Deng, F.-L. Chung, and S. Wang, "Robust relief-feature weighting, margin maximization, and fuzzy optimization," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 4, pp. 726–744, Aug. 2010, doi: 10.1109/TFUZZ.2010.2047947.

[25] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

BIOGRAPHIES OF AUTHORS

Nishat Tasnim Mim received her bachelor's degree in computer science and engineering from East West University in 2021. Currently, she is an M.S student at the Institute of Information Technology, University of Dhaka. Her research interests include machine learning, data science and pattern recognition. She can be contacted at email: msse1756@iit.du.ac.bd.

Md. Eusha Kadir received his B.Sc. and M.Sc. degree in software engineering from the Institute of Information Technology, University of Dhaka, Dhaka, Bangladesh. He is working as a lecturer in the Institute of Information Technology at Noakhali Science and Technology University. He currently teaches and conducts research programs in the areas of artificial intelligence. He has publications in IEEE, ACM, and Springer. His research interests include machine learning, data mining, and software engineering. He can be contacted at eusha@nstu.edu.bd.

Suravi Akhter received her B.Sc. and M.Sc. degree in software engineering from the Institute of Information Technology, University of Dhaka, Dhaka, Bangladesh in 2019 and 2022 respectively. She is a lecturer in the Department of Computer Science and Engineering at the University of Liberal Arts Bangladesh. Her research interests include machine learning, data science and natural language processing. She can be contacted at email: suravi.akhter@ulab.edu.bd.

Muhammad Asif Hossain Khan received the B.Sc. and M.Sc. degrees in computer science from University of Dhaka, Dhaka, Bangladesh, in 1999 and 2001 respectively. He received the Ph.D. degree in information science and technology from University of Tokyo in Japan. He has been a professor in the Department of Computer Science and Engineering at University of Dhaka, since 2019. His research interests include the applications of artificial intelligence, machine learning, natural language processing, text mining, and information retrieval. He can be contacted at asif@du.ac.bd.