

Optimized Feature Selection and Enhanced SVM for Accurate Classification of Medical Datasets

Ahmed Hassan
ElSewedy University of
Technology
Cairo, Egypt
ahmed.hassan@sut.edu.eg

Khaled M, Fouad
New Mansoura
University
Mansoura, Egypt
Khaled.foad@nmu.edu.eg

Alaa Eldin Abdalla Yassin
Information System department
Benha University
Benha, Egypt
alaaelden.yassin@fci.bu.edu.eg
Alaaeldin.yassin@must.edu.eg

Somia Mohamed
Information System
department
Benha University
Benha, Egypt
somia.aboelnaga@fci.bu.edu.eg

Abstract- the increasing volume of medical data presents both opportunities and challenges for enhancing classification methods. While Support Vector Machines (SVMs) are widely recognized for their effectiveness in various classification tasks, traditional SVM approaches exhibit significant limitations when applied to large-scale datasets. This paper presents a hybrid feature selection framework designed to enhance the classification performance of medical data by identifying the most relevant features for Support Vector Machine (SVM) classifiers. The proposed framework consists of two phases. The first phase combines Information Gain with several optimization techniques, including Particle Swarm Optimization (PSO), Bat Search Method, Elephant Search Algorithm (ESA), and Firefly Algorithm, to effectively select key features for medical datasets. Our results show that the hybrid framework significantly improves classification accuracy compared to traditional feature selection methods, with the ESA Algorithm excelling in their respective categories. In the second phase, we integrate enhancements into the traditional SVM model to address uncertainties in medical data, resulting in a more adaptable and accurate classifier. The enhanced SVM refines both training and testing sets, ensuring that the model is trained on the most relevant and accurate data, thus improving its performance in uncertain or noisy conditions. Overall, the proposed hybrid framework provides a more robust and efficient solution for medical data classification, improving both the accuracy and adaptability of SVM classifiers in real-world applications.

Keywords— Machine learning, Medical data classification, KNN, feature selection, optimized method, SVM, Data uncertainty, Hybrid model, Classification accuracy

I. INTRODUCTION

The rapid expansion of medical data generated through electronic health records, imaging technologies, and wearable devices presents both significant opportunities and challenges for the healthcare sector. Accurate classification of this data is crucial for enhancing diagnostic accuracy, personalizing treatment plans, and improving patient outcomes [1]. Among various machine learning techniques, Support Vector Machines (SVMs) have garnered attention for their effectiveness in handling classification tasks across different domains, including medicine [2-3]. However, traditional SVM methods face notable limitations when confronted with large-scale datasets and uncertain data scenarios [4-5].

Large-scale datasets can complicate the training process of SVMs due to increased computational demands and the risk of overfitting [6]. Furthermore, uncertain data, characterized by noise, missing values, or conflicting class labels, can significantly hinder the classification performance of standard SVM models. Addressing these issues is essential for harnessing the full potential of SVMs in medical applications.

Data are growing not only in terms of the sheer number of data objects but also in the dimensionality of features. This increase in both volume and complexity can significantly impact the accuracy and efficiency of most learning algorithms [7]. High-dimensional datasets often lead to the "curse of dimensionality," where the distance between

data points becomes less meaningful, making it challenging for algorithms to find patterns and make accurate predictions [8]. In medical data classification, these challenges are particularly pronounced. The variety of features derived from clinical measurements and patient history can create rich datasets that hold valuable insights. However, when algorithms are confronted with high dimensionality, the risk of irrelevant or redundant features can dilute the signal of meaningful patterns, leading to reduced classification performance [9]. Feature selection methods are essential tools for identifying and removing irrelevant attributes from datasets, which do not contribute to the accuracy of a classification model. By focusing on the most pertinent features, these methods help streamline the data, reducing dimensionality and enhancing model performance [10]. Previous studies examining the use of feature selection in conjunction with classification models consistently conclude that incorporating feature selection as a preprocessing step yields significant benefits [11]. As a result, models that use feature selection often show higher accuracy, improved generalization to unseen data, and greater interpretability. In the context of medical data classification, these advantages are critical, as they lead to more reliable diagnostic results.

Classification is a supervised machine learning method that involves building models capable of assigning objects to one of several predefined classes or categories. The process typically begins with the preparation of a dataset, which is divided into two distinct subsets: the training set and the testing set. The training set is used to develop the classification model; it contains examples of the input features along with their corresponding class labels. During this phase, the model learns to recognize patterns and relationships within the data, effectively capturing the underlying structure that distinguishes one class from another. Once the model has been trained, it is validated using the testing set, which contains new, unseen data. This set is critical for assessing the model's performance and generalizability. By evaluating how accurately the model classifies the testing set, researchers can determine its effectiveness and reliability in making predictions on real-world data [12]. In medical data classification, effective training and validation are paramount, as they directly influence the quality of diagnostic decisions and patient outcomes.

Support Vector Machines (SVM) is supervised learning algorithms used primarily for classification tasks. They work by finding the optimal hyperplane that separates data points of different classes in a high-dimensional space. Support Vector Machines (SVM) has gained significant traction in the medical field for various classification tasks due to their robustness and effectiveness. Medical datasets often contain a large number of features, such as genetic markers or imaging pixels. SVM has ability to work effectively even when the number of features exceeds the number of samples [13].

In this paper, an enhanced classification method for medical data is introduced. The presented approach consists of two phases. The initial phase employs a hybrid method to identify and select an optimal subset of features. By integrating various feature selection techniques. This phase aims to enhance the model's efficiency and accuracy, which is particularly crucial in medical datasets where irrelevant or redundant features can hinder performance. In the second phase, an enhanced Support Vector Machine (SVM) classifier is developed to manage class imbalances and dealing with uncertain data. Our contributions are as follows:

- Introducing a hybrid feature selection approach that combines optimized techniques to generate an optimal set of features. This method significantly enhances model accuracy by focusing on the most relevant data, thus improving overall performance.
- Developing an improved Support Vector Machine (SVM) method capable of addressing uncertainties in the data. This approach involves updating both the training and testing sets to better manage ambiguities, resulting in more robust classification outcomes.

The structure of this paper is as follows. The related works are reviewed in Section 2, and then the proposed model is showed in Section 3. The experimental platform is shown in Section 4. The used medical datasets are presented in Section 5. The experimental results and a comparison between other models are discussed in Section 6. Finally, Section 7 concludes this study.

II. LITERATURE REVIEW

Several recent works have addressed the issue of feature selection and classification for medical datasets using hybrid and optimized methods. Eva Tuba et al. [14] proposed a feature selection method for medical datasets based on the Brain Storm Optimization (BSO) algorithm. The algorithm was used to select the best features and optimize the parameters of the Support Vector Machine (SVM) classification method. The results showed that the proposed method not only achieved higher accuracy but also reduced the number of features, making it more efficient.

B. Subanya et al. [15] focused on the classification of cardiovascular disease using the Artificial Bee Colony (ABC) algorithm. The ABC algorithm was used for feature selection, effectively eliminating irrelevant features. The SVM method was employed for classification, and when compared to the reverse ranking method, the proposed method demonstrated superior performance.

Tianyun Xiao et al. [16] proposed a model for diagnosing breast cancer. The proposed framework combined feature selection method with hyper parameter optimization to improve the accuracy for classification methods. Multiple linear regression method was used for feature selection. Whale optimization algorithm (WOA) was applied to optimize the hyper parameters of the model to improve their performance. Comparing with existing model, the proposed method gave best prediction performance.

Yılmaz Kaya et al. [17] developed a hybrid model for diagnosing breast cancer, lymphography, and erythematous-squamous diseases. Their approach involved using Factor Analysis for preprocessing to extract relevant features, followed by classification using an Extreme Learning Machine (ELM). The findings revealed that this hybrid model significantly outperformed the use of ELM alone.

In [18], a feature selection method based on clustering with the Artificial Bee Colony (ABC) algorithm was proposed to identify the most relevant features for liver diseases, hepatitis, and diabetes datasets. SVM was then employed for classification, and the results showed that the proposed method outperformed other existing approaches in terms of classification accuracy.

Saba Bashir [19] utilized various feature selection methods, including filters, wrappers, and embedded approaches, to identify the most relevant features from medical datasets. The SVM classification method was then applied to the selected features, resulting in high classification performance. The proposed model demonstrated exceptional effectiveness.

In [20], Particle Swarm Optimization (PSO) was used as a wrapper method for feature selection to identify the most informative features and eliminate redundant ones, thereby improving the classification accuracy for medical datasets. The experiments, which used five medical datasets from the UCI repository, showed that the proposed method yielded high accuracy across different classification algorithms.

Zhou Tao et al. [21] proposed an optimized feature selection method that combined K-means clustering and hybrid techniques, such as Genetic Algorithm (GA) and Principal Component Analysis (PCA), along with Particle Swarm Optimization (PSO) and PCA. The method was used to select optimal features and tune the parameters of the SVM classification model. The proposed method achieved superior classification results compared to other methods.

Moloud Abdar et al. [22] conducted experiments on four microarray datasets with high dimensionality and relatively small sample sizes. They applied five different feature selection algorithms to reduce the dimensionality of the datasets. The proposed technique incorporated Bagging in the feature selection process, resulting in significant improvements in selection stability. Finally, the work of Saba Bashir et al. [23] aimed to enhance breast cancer detection using a hybrid ensemble classifier combining Confidence Weighted Voting (CWV) with Boosting Artificial Neural Networks (BANN) and SVM. The CWV-BANN-SVM model demonstrated improved accuracy in breast cancer detection when compared to traditional ANN or SVM methods alone, highlighting the potential of hybrid ensemble classifiers in medical data classification.

III. METHODOLOGY

In this section, we introduce two phases of the proposed algorithm to address the unique challenges presented by medical data. One significant challenge is the small number of samples relative to the large number of features, which necessitates careful feature selection to enhance model performance. The first phase of the algorithm focuses on producing the optimal features, which serve as inputs for the second phase. By identifying and selecting the most relevant features, this phase lays the groundwork for more effective analysis.

Firstly, the model undertakes several preprocessing steps to enhance the accuracy of the results. This includes removing any missing data to ensure a complete dataset and scaling the features to normalize all values. These preprocessing steps are crucial for preparing the data, as they help to improve the overall performance of the model. Then, we generate ten subsets of data using 10-fold cross-validation, ensuring that the subsets are divided equally and randomly. This technique is commonly used to assess the model's performance and reduce bias. For each subset, we apply a hybrid feature selection method combined Information Gain with an optimization technique to identify the best features (f). The optimized methods use accuracy of SVM method as fitness function. This approach allows us to effectively evaluate the performance of the proposed method while ensuring that only the most relevant features are utilized in the analysis. This hybrid method aims to reduce the number of input features while maintaining high classification accuracy. It begins with a filter method, such as information gain, to rank the features based on their relevance. Following this, various optimized search methods are employed to select the best subset of features. This combination ensures that we retain only the most informative features, ultimately enhancing the model's performance without compromising accuracy. Finally, the best features are aggregated using a voting method, which consolidates the most relevant features identified in the previous steps. These selected features are

then used as inputs for the second phase of the algorithm, ensuring that the subsequent analysis is based on the most informative and impactful data Figure 1.

The second phase implements an enhanced Support Vector Machine (SVM) algorithm to effectively address uncertain data while performing classification.

The process of developing an enhanced Support Vector Machine (SVM) model begins with the collection of features from Phase 1, which consists of pre-processed and relevant data attributes. The dataset is partitioned into two subsets: the training set (Train_set) and the testing set (Test_set). This partitioning is crucial for evaluating the model's performance on unseen data.

Initially, an SVM model is trained using the selected features from the Train_set, establishing a baseline model. The algorithm then identifies uncertain objects in the Test_set, where uncertainty is quantified by a classification confidence score. Objects with a confidence score below a predefined threshold are flagged as uncertain.

These uncertain objects are incorporated into the training data by adding them to the Train_set. This augments the learning process with challenging examples, enhancing the model's ability to generalize. To preserve the integrity of the testing phase, these uncertain objects are removed from the Test_set.

Subsequent to the data update, the SVM model is retrained on the augmented Train_set. This retraining allows the model to refine its decision boundaries based on the newly included uncertain objects. The model's performance is then reassessed on the updated Test_set, providing insights into the impact of active learning on the model's generalization capability. Finally, the performance of the enhanced SVM model is compared with other methods to validate the effectiveness see Figure 2.

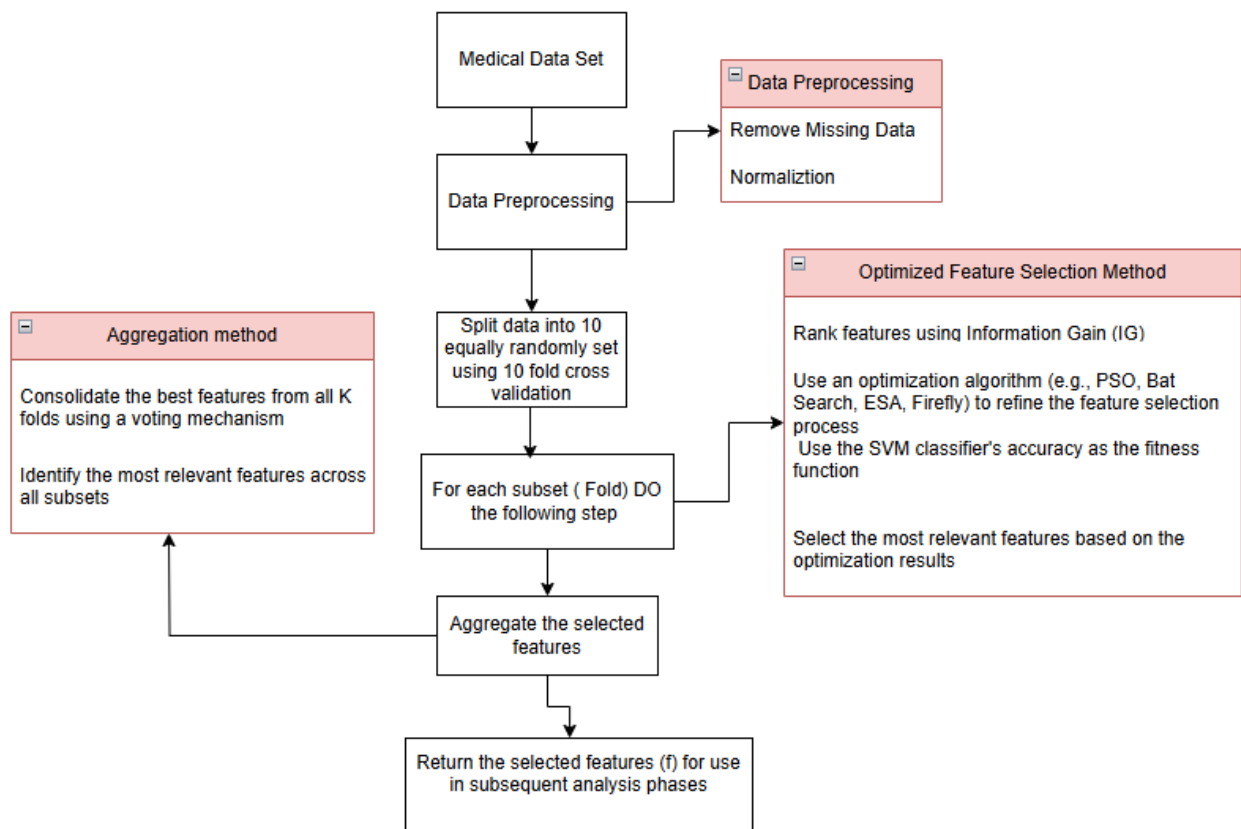


Figure. 1. Steps for first phase in the proposed methods

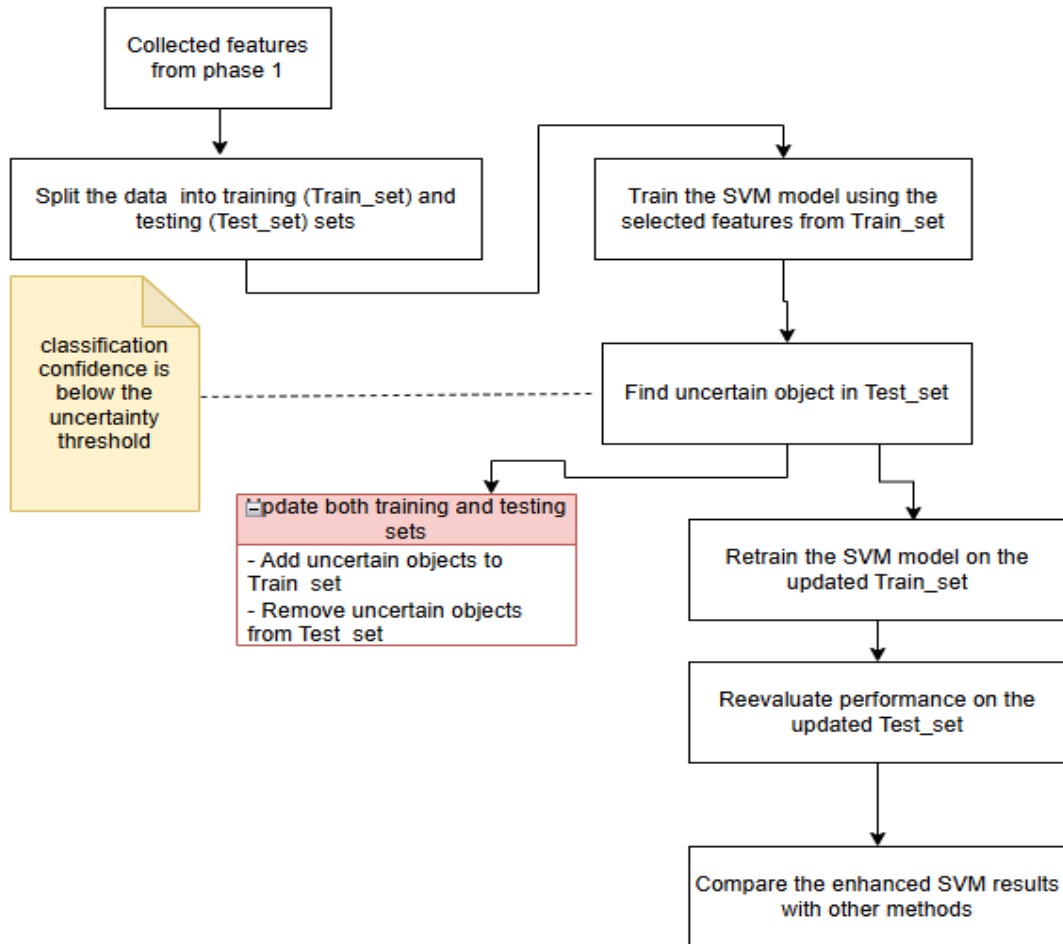


Figure. 2. Steps for second phase in the proposed methods

IV. EXPERIMENTAL PLATFORM

The proposed method was implemented using the R programming language to generate results. The implementation was conducted on a 64-bit operating system, specifically Windows 7 Ultimate, with 4 GB of RAM and an Intel® Core™ i5 processor.

V. DATA SAMPLE

In this study, we utilize various medical datasets to evaluate the performance of the proposed method. These datasets, characterized by high dimensionality and small sample sizes, are critical for assessing the robustness and effectiveness of our approach. They are publicly available in the Feature Selection @ ASU repository. A summary of the datasets used is provided in TABLE I.

TABLE I. USED DATASETS

Data set	#objects	#features	Keywords
Blood	89	2759	Discrete, binary
Colon	62	2000	Discrete, binary
Leukemia	72	12584	Discrete, binary

VI. EXPERIMENTAL RESULTS

The proposed framework is evaluated and discussed in this section. Methods consists of two phases, the first phase is to select the optimal features from many irrelevant. The second phase is to classify the data with enhanced SVM classification method. In order to show enhanced method, we carried out an empirical study of some of chosen medical datasets.

A. Results from Optimized Feature selection with traditional SVM

In the first phase of our analysis, we used feature selection method to enhance the accuracy of our model. We employed a combination of Information Gain and optimized search methods such as PSO, Bat search method; Elephant search method and Firefly search method to identify the most significant features from our dataset.

First, some preprocessing steps are performed such as handling missing values to ensure integrity of data and normalization to make features more comparable. Then data are divided into ten subsets, and then for each subset, best features are selected using the hybrid methods. Finally, features using voting are shown in TABLE II, TABLE III and TABLE IV.

TABLE II. BLOOD DATASET RESULTS

Method used for features	# features
PSO + IG	29
Bat + IG	31
Elephant + IG	32
FireFly +IG	32

TABLE III. COLON DATASET RESULTS

Method used for features	# features
PSO + IG	28
Bat + IG	28
Elephant + IG	30
FireFly +IG	33

TABLE IV. LEUKEMIA DATASET RESULTS

Method used for features	# features
PSO + IG	197
Bat + IG	212
Elephant + IG	228
Firefly +IG	231

After selecting the most significant features, traditional SVM classification method is used through 10-fold cross-validation to evaluate the classification performance.

The summarized results of the classification accuracy for each four feature selection methods are presented for three medical datasets in Figure 3, Figure 4 and Figure 5.

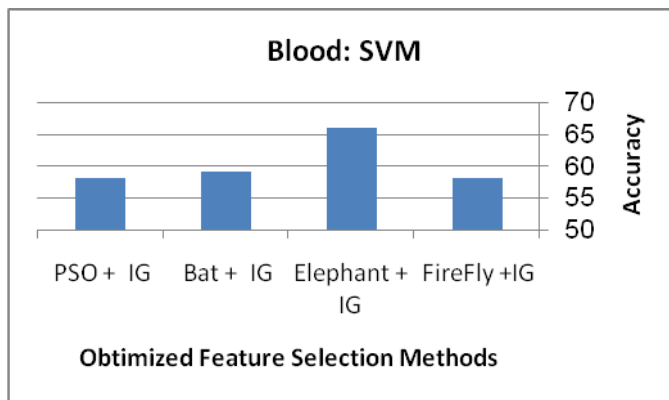


Figure. 3. Blood datasets results

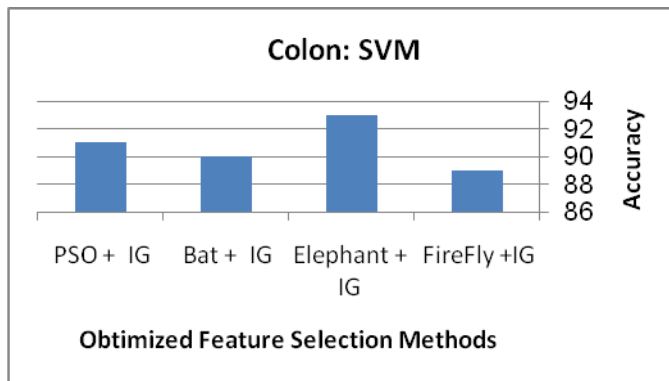


Figure. 4. COLON datasets results

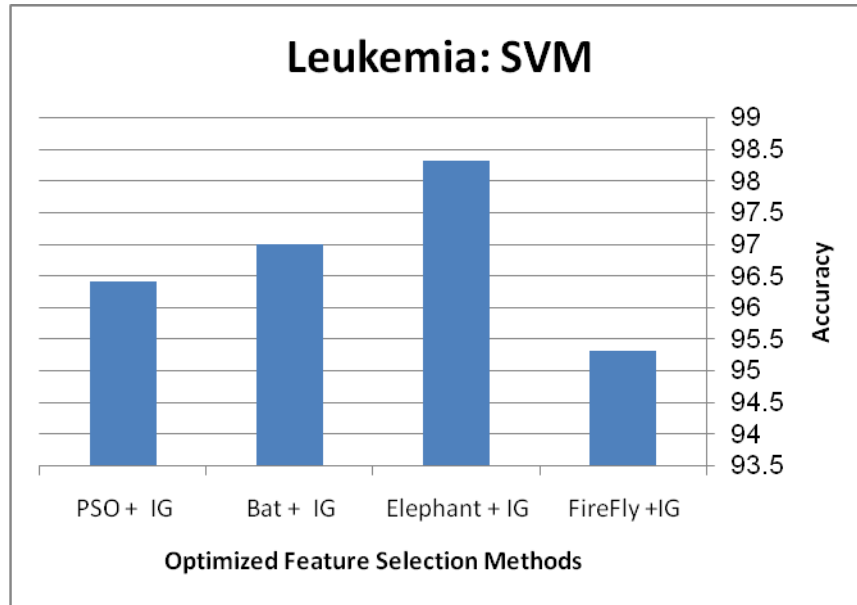


Figure. 5. Leukemia datasets results

By conducting a comparison between the four optimized feature selection methods using SVM classification, the results show that the Elephant algorithm outperforming the other methods.

B. Results for enhanced SVM

In this section, we explore enhancements to the traditional SVM algorithm to address its limitation. Traditional SVM method can't handle uncertain data. This can do by updating both training set and testing set. If uncertain object found, both training set and testing set are rebuild. This strategy can help SVM handle uncertain data more effectively, improving its accuracy and reliability. By comparing the obtained results from enhanced SVM with other methods listed in [22] also with enhanced KNN [24], and with SVM with features selection obtained in phase 1, we find that the value of accuracy of enhanced SVM get the best performance as shown in Figure. 6.

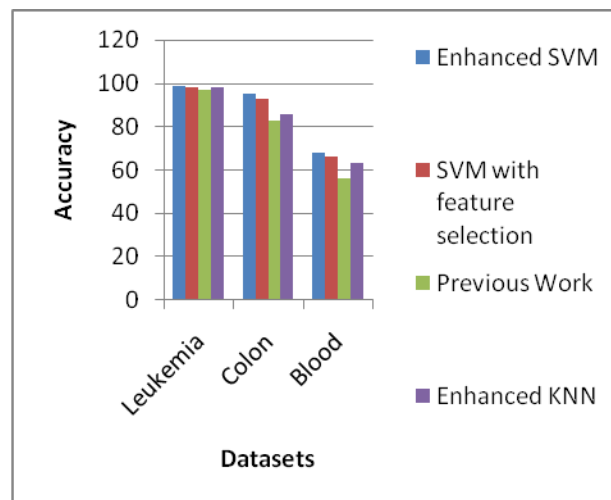


Figure. 6. Comparison with other methods

From the results provided, the enhanced Support Vector Machine (SVM) model demonstrates superior performance compared to other methods across all three datasets—blood, colon, and leukemia.

For blood dataset, the enhanced SVM outperforms the other methods for the blood dataset, with a slight improvement over SVM with feature selection (2%) and a more noticeable improvement over KNN (5%) and previous work (9%). For the colon dataset, the enhanced SVM achieves a higher accuracy (95.3%) compared to SVM with feature selection (93%) and shows a considerable advantage over the previous work (83%) and KNN method (86%). In the case of the leukemia dataset, the enhanced SVM achieves the highest accuracy (98.7%), which is slightly better than SVM with feature selection (98.3%), and considerably better than the previous work (97%) and KNN method (97.9%).

This indicates that the enhanced SVM model, especially in terms of its adaptability and feature selection process, provides more robust and accurate results compared to other compared methods like SVM with feature selection obtained from phase1, previous work, and KNN for these datasets.

VII. DISCUSSION

In this study, we proposed a hybrid feature selection framework aimed at enhancing the classification performance of medical data by identifying the most relevant features for SVM classifier. We employed a combination of Information Gain with several optimization techniques, including PSO, Bat Search Method, Elephant Search Algorithm, and Firefly Algorithm, to identify and select the most significant features for the datasets.

Our results demonstrated that the performance of the feature selection methods varied depending on the classification algorithm used. Specifically, the Elephant Search Algorithm outperformed the other optimization methods when coupled with the SVM classifier, while the Firefly Algorithm achieved the best results with the KNN classifier. These findings suggest that different optimization techniques may be more suitable for different classification models, highlighting the importance of selecting the right feature selection method based on the specific classification task.

The use of 10-fold cross-validation for model evaluation provided a robust assessment of the classification performance, ensuring that our results were not biased by overfitting. By comparing the accuracy, stability, and efficiency of the optimized feature selection methods, we were able to show that the proposed hybrid framework significantly improves classification accuracy compared to traditional feature selection methods, with the Elephant Search Algorithm and Firefly Algorithm leading the performance in their respective categories.

The improvements in classification accuracy observed in our experiments can be attributed to the effective feature selection process, which reduced the dimensionality of the data while retaining the most informative features. This is particularly important in medical datasets, where the number of features often exceeds the number of samples, creating challenges for traditional machine learning algorithms. Our approach helps mitigate the curse of dimensionality and enhances the classifier's ability to generalize from the data.

In second phase, the results presented highlight the significant improvements made to the traditional Support Vector Machine (SVM) model through the incorporation of enhancements that address the challenges posed by uncertain data. These results suggest that the enhanced SVM is particularly effective when handling uncertain or noisy data. The improvements in performance can likely be attributed to the model's ability to adaptively refine both training and testing sets, ensuring that the classifier is trained on more relevant and accurate information. The combination of uncertainty handling and feature selection boosts the performance of the SVM. While SVM with feature selection alone provides good results, the enhanced SVM leverages an additional layer of adaptability that allows it to better handle varying data conditions.

VIII. CONCLUSIONS

This study proposes a hybrid method to enhance the performance of the Support Vector Machine (SVM) classification method for large-scale medical data. The proposed approach addresses the challenges faced by traditional SVMs when working with extensive medical datasets. The technique consists of two main phases. In the first phase, we aim to identify the optimal subset of features from the available data. This is achieved through a combination of four optimized methods along with Information Gain, ensuring the selection of features that contribute the most to model performance. In the second phase, the SVM method is improved to effectively handle uncertain data. The proposed model addresses potential confusion arising from class imbalances by updating both the training and testing sets with instances that contribute to this imbalance. Our results demonstrate that the proposed framework outperforms existing methods, showcasing its effectiveness in managing uncertainty in medical data classification.

While accuracy is the primary evaluation metric used in this study, in future work, it would be valuable to assess the proposed model using other metrics such as precision, recall, F1-score. Also, future research could investigate more advanced uncertainty management techniques, such as fuzzy logic to further enhance the model's robustness in the presence of noisy or imbalanced data.

REFERENCES

- [1] E. M. F. El Houby, "A survey on applying machine learning techniques for management of diseases," *Journal of Applied Biomedicine*, vol. 16, no. 3, pp. 165–174, 2018. doi: 10.1016/j.jab.2018.01.002W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
- [2] M. Khalaf, A. Hussain, O. Alfandi, D. Al-Jumeily, M. Alloghani, M. Alsaadi, O. A. Dawood, and D. Abd, "An application of using support vector machine based on classification technique for predicting medical data sets," in *Intelligent Computing Theories and Applications*, pp. 636–644, Springer, 2019. doi: 10.1007/978-3-030-26969-2_55.
- [3] P. Janardhanan, L. Heena, and F. Sabika, "Effectiveness of support vector machines in medical data mining," *Journal of Communications Software and Systems*, vol. 11, no. 1, pp. 25–30, 2015. doi: 10.24138/jcomss.v11i1.114.
- [4] X. Wu, X. Zhu, G. Wu, and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97–107, 2014. doi: 10.1109/TKDE.2013.109
- [5] M. Panda, "Elephant search optimization combined with deep neural network for microarray data analysis," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 8, pp. 940–948, 2020. doi: 10.1016/j.jksuci.2017.12.002
- [6] X.-J. Hen, L. Mu, Z. Li, and H.-X. Wu, "Large-scale support vector machine classification with redundant data reduction," *Neurocomputing*, vol. 172, pp. 188–194, 2015. doi: 10.1016/j.neucom.2014.10.102
- [7] J. C. Pamila, R. Selvi, P. Santhi, and T. M. Nithya, "Ensemble classifier-based big data classification with hybrid optimal feature selection," *Advances in Engineering Software*, vol. 173, p. 103183, 2022. doi: 10.1016/j.advengsoft.2022.103183.
- [8] A. L'Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, "Machine learning with big data: Challenges and approaches," *IEEE Access*, vol. 5, pp. 7776–7797, 2017. doi: 10.1109/ACCESS.2017.2696365.
- [9] A. Y. Mahmoud, "Novel efficient feature selection: Classification of medical and immunotherapy treatments utilizing Random Forest and Decision Trees," *Intelligence-Based Medicine*, vol. 10, p. 100151, 2024. doi: 10.1016/j.ibmed.2024.100151.
- [10] Z. Ma, L. T. Yang, and Q. Zhang, "Support multimode tensor machine for multiple classification on industrial big data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3382–3390, 2021. doi: 10.1109/TII.2020.2999622

- [11] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "A comparison of multi-label feature selection methods using the problem transformation approach," *Electronic Notes in Theoretical Computer Science*, vol. 292, pp. 135–151, 2013. [doi: 10.1016/j.entcs.2013.02.010](https://doi.org/10.1016/j.entcs.2013.02.010).
- [12] K. J. D'souza and Z. Ansari, "Big data science in building medical data classifier using Naïve Bayes model," in *2018 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*, pp. 76–80, 2018. [doi: 10.1109/CCEM.2018.00020](https://doi.org/10.1109/CCEM.2018.00020)
- [13] P. Janardhanan, L. Heena, and F. Sabika, "Effectiveness of support vector machines in medical data mining," *Journal of Communications Software and Systems*, vol. 11, no. 1, pp. 25–30, 2015. [doi: 10.24138/jcomss.v11i1.114](https://doi.org/10.24138/jcomss.v11i1.114)
- [14] E. Tuba, I. Strumberger, T. Bezdán, N. Bacanin, and M. Tuba, "Classification and feature selection method for medical datasets by brain storm optimization algorithm and support vector machine," *Procedia Computer Science*, vol. 162, pp. 307–315, 2019. [doi: 10.1016/j.procs.2019.11.289](https://doi.org/10.1016/j.procs.2019.11.289).
- [15] B. Subanya and R. R. Rajalaxmi, "Feature selection using Artificial Bee Colony for cardiovascular disease classification," in *2014 International Conference on Electronics and Communication Systems (ICECS)*, pp. 1–6, 2014. [doi: 10.1109/ECS.2014.6892729](https://doi.org/10.1109/ECS.2014.6892729).
- [16] T. Xiao, S. Kong, Z. Zhang, F. Liu, A. Yang, and D. Hua, "FS-WOA-stacking: A novel ensemble model for early diagnosis of breast cancer," *Biomedical Signal Processing and Control*, vol. 95, Part B, p. 106374, 2024. [doi: 10.1016/j.bspc.2024.106374](https://doi.org/10.1016/j.bspc.2024.106374).
- [17] Y. Kaya and F. Kuncan, "A hybrid model for classification of medical data set based on Factor Analysis and Extreme Learning Machine: FA+ELM," *Biomedical Signal Processing and Control*, vol. 78, p. 104023, 2022. [doi: 10.1016/j.bspc.2022.104023](https://doi.org/10.1016/j.bspc.2022.104023).
- [18] M. S. Uzer, N. Yilmaz, and O. Inan, "Feature selection method based on artificial bee colony algorithm and support vector machines for medical datasets classification," *The Scientific World Journal*, vol. 2013, p. 419187, 2013. [doi: 10.1155/2013/419187](https://doi.org/10.1155/2013/419187).
- [19] S. Bashir, I. Khattak, A. Khan, F. Khan, A. Gani, and M. Shiraz, "A novel feature selection method for classification of medical data using filters, wrappers, and embedded approaches," *Complexity*, vol. 2022, p. 8190814, 2022. [doi: 10.1155/2022/8190814](https://doi.org/10.1155/2022/8190814).
- [20] T. Saw, "Feature selection to classify healthcare data using wrapper method with PSO search," *International Journal of Information Technology and Computer Science*, vol. 11, pp. 31–37, 2019. [doi: 10.5815/ijitcs.2019.09.04](https://doi.org/10.5815/ijitcs.2019.09.04).
- [21] T. Zhou, H. Lu, W. Wang, and Y. Xia, "GA-SVM based feature selection and parameter optimization in hospitalization expense modeling," *Applied Soft Computing*, vol. 75, pp. 323–332, 2019. [doi: 10.1016/j.asoc.2018.11.001](https://doi.org/10.1016/j.asoc.2018.11.001).
- [22] S. Alelyani, "Stable bagging feature selection on medical data," *Journal of Big Data*, vol. 8, p. 10, 2021. [doi: 10.1186/s40537-020-00385-8](https://doi.org/10.1186/s40537-020-00385-8).
- [23] M. Abdar and V. Makarenkov, "CWV-BANN-SVM ensemble learning classifier for an accurate diagnosis of breast cancer," *Measurement*, vol. 146, pp. 557–570, 2019. [doi: 10.1016/j.measurement.2019.05.022](https://doi.org/10.1016/j.measurement.2019.05.022).
- [24] A. Hassan, K. M. Fouad, A. E. A. Yassin and S. Mohamed, "Enhanced Classification method for Large-Scale Medical Data," *2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, Cairo, Egypt, 2024, pp. 209-213, [doi: 10.1109/MIUCC62295.2024.10783601](https://doi.org/10.1109/MIUCC62295.2024.10783601)

Call for Papers: International Journal of Computer Science and Information Security (IJCSIS)

Scope and Topics: The International Journal of Computer Science and Information Security (IJCSIS) invites researchers, practitioners, and academicians to submit original, unpublished contributions covering a wide range of topics in the field of computer science and information security. We welcome submissions that include but are not limited to:

- Computer and Network Security
- Cryptography and Data Security
- Information Assurance
- Artificial Intelligence and Machine Learning in Security
- Cybersecurity Policies and Standards
- Security in Cloud Computing
- Internet of Things (IoT) Security
- Blockchain and Distributed Ledger Technologies
- Secure Software Development
- Privacy-Enhancing Technologies

Submission Guidelines:

- Manuscripts must be original and not currently under consideration for publication elsewhere. All papers should be submitted in English.
- The manuscript should follow the IJCSIS formatting guidelines, available on our website: <https://sites.google.com/site/ijcsis/ijcsis>
- Submissions must include the title of the paper, abstract, keywords, and full contact information for all authors.
- Papers should be submitted via our submission system here: <https://sites.google.com/site/ijcsis/submit-paper>

Important Dates:

- **Paper Submission Deadline:** Monthly, 2025
- **Notification of Acceptance:** Within TWO weeks, 2025
- **Final Manuscript Due:** Monthly, 2025
- **Publication Date:** Monthly Issue

Review Process: All submitted papers will undergo a rigorous peer-review process by the IJCSIS editorial board and selected external reviewers. Authors will be notified of the review results by the notification of acceptance date.

Contact Information: For further inquiries, please contact the editorial office at:

Email: ijcsiseditor@gmail.com

We look forward to receiving your submissions!

Website: <https://sites.google.com/site/ijcsis/ijcsis>