Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 818–827

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine

Mandakini Priyadarshani Behera, Archana Sarangi*, Debahuti Mishra , Shubhendu Kumar Sarangi

ITER, Faculty of Engineering & Technology
Siksha 'O' Anusandhan Deemed to be University,
Bhubaneswar, Odisha, India

Abstract

Machine learning is now extensively applied in a variety of fields. Machine learning is employed like an efficient assistance mechanism in clinical diagnostics since vast amounts of data are readily available. Owing to heavy alcohol consumption, inhalation of contaminated gas, narcotics, food contamination, unhealthy life style people suffering from heart and liver disease has been significantly growing. Both heart and liver disease cause high mortality rate worldwide. It is critical to discover these diseases at an early stage in order to save people's lives. Incorporating machine learning classification algorithms into health-care organizations yields remarkable outcomes, allowing health-care practitioners to diagnose diseases more quickly and accurately. Machine learning techniques and tools aid in the extraction of useful information from datasets, resulting in more exact findings. In this study for heart and liver data classification, a hybrid model is created by combining support vector machine (SVM) approach and modified particle swarm optimization model. The data sets are collected from UCI machine learning repository. The results are calculated based on classification accuracy, error, correctness, recall as well as F1 score. The results obtained is compared with SVM, hybrid particle swarm optimization support vector machine algorithm (PSOSVM), hybrid Crazy particle swarm optimization support vector machine algorithm (CPSOSVM).

* Corresponding author. Tel.: 919438670340.
*E-mail address:*archanasarangi24@gmail.com

1877-0509 © 2023 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the International Conference on Machine Learning and Data Engineering
10.1016/j.procs.2023.01.062

Keywords: Machine learning; Support vector machine; particle swarm optimization; heart disease; liver disease

1. Introduction

Machine learning is a subset of computer science which focuses on improving innovation of computers. Machine learning has several applications in our daily lives, particularly in the domain of healthcare. Machine learning has many aspects, including feature extraction, feature selection, algorithm selection, training, and testing. Machine learning is important in the healthcare area because of its powerful data analysis capabilities. Typically, scientists express their interest about prediction and diagnosis by employing machine learning techniques that 1) minimize diagnosis time and 2) improve accuracy and efficiency. With the use of supervised machine learning techniques, any illness category can be diagnosed. By the use of supervised machine learning algorithms, any category of disease can be diagnosed, although the focus of this study will be on heart and liver disease detection. Heart and liver disease is now considered a leading cause of death all over the world, necessitating correct diagnosis in a timely manner or appropriate diagnosis of heart disease to save lives. The heart and liver are considered to be the most significant organs in the human body. As a result, heart and liver disease is seen as a major health concern in everyday life. According to several reports, heart and liver disorders are the most common cause of sudden death in developed countries. Heart as well as liver illnesses now impact new-born babies as well. As a result, testing for heart and liver disorders is quite common in everyday life [1].

Many researchers in the medical field have experimented with various strategies to increase data classification accuracy. Techniques that offer better accurate classification will give more evidence to find possible patients and enhance diagnosis accuracy. Many machine learning classifiers like decision tree (DT), random forest (RF), K-nearest neighbour (KNN) and logistic regression (LR) and optimisation algorithms like PSO, Firefly algorithm (FA), GA has been applied for heart and liver disease prediction. In [2], a new method for classifying cardiac arrhythmia beats is proposed. Pan-Tompkins's algorithm for recognition of R-peak, discrete orthogonal Stockwell transform (DOST) for extraction of features in ECG signal in addition to support vector machines (SVMs) for classification for automatic cardiac arrhythmia beat classification, the parameters are modified using the particle swarm optimization (PSO) technique are among those algorithms used in this method. In [3] classification algorithms Nave Bayes and support vector machine techniques were employed to predict liver disorders. The performance metrics used to compare these classifier algorithms are classification accuracy and execution time. According to the results of the experiments, the SVM is a better classifier for predicting liver disorders. In [4]. The major idea of this research paper is to use different categorization techniques for prediction of liver disease. Logistic Regression, K-Nearest Neighbour, and Support Vector Machines are the algorithms employed in this project. This classification system is compared using an accuracy score and a confusion matrix. It is concluded that Logistic Regression is effective in predicting liver illness. In [5] Researchers examine the accuracy of machine learning approaches for predicting cardiac illness, such as kNN, decision tree, linear regression in addition to support vector machine, by means of the UCI repository dataset for training as well as testing of SVM. In [6] opinion mining uses decision tree-based feature selection methods for heart disease prediction with lesser number of features and the experimental results are more accurate. In [7] unsupervised rough set techniques have been proposed for web opinion text clustering task to get superior results. However, the classification accuracy of existing approaches for liver illnesses is still poor and insignificant enough to be used in practical applications. As the severity of heart and liver diseases are growing rapidly and the mortality rate can be decreased with proper treatment so strong diagnosis is needed. However, the classification accuracy of existing approaches for liver and heart illnesses is still poor and insignificant enough to be used in practical applications. In this study our main objective is to progress a hybrid model according to modified particle swarm optimisation method along with support vector machine for heart as well as liver disease prediction. In this study our main objective is to progress a hybrid model according to modified particle swarm optimisation method along with support vector machine for heart as well as liver disease prediction. The existing hybrid algorithms and the machine learning classifiers are unable to give accurate results for disease prediction. To obtain a better accuracy disease prediction result by overcoming the drawbacks we proposed a hybrid model by combine the modified PSO algorithm with SVM. The remaining part of the study is arranged like. In **Section 2**linked with this study has been described. **Section 3** clearly states about the proposed methodology. **In Section 4** contains the evidence and outcomes. Subsequently, conclusion is framed in **Section 5**.

2. Related Work

This segment contains the machine algorithms used for disease prediction has been discussed

2.1. Support Vector Machine (SVM)

SVM refers for machine learning in supervisory type process that analyses as well as identifies patterns in input data to perform classification or regression analysis. SVM is used in a variety of applications, including digit recognition, handwriting recognition, face detection, cancer classification, time series forecasting, and so on [8].

2.2. Particle Swarm Optimization (PSO)

The PSO is a population-oriented, metaheuristic for optimizing the systems based on swarm or collective social behavior. Because to its simplicity, ease of application, and effectiveness in resolving optimization challenges, this optimization method is quite popular. Individuals move randomly with different velocities throughout the search process, and these velocities are used to alter each individual's position. The position as well as velocity of every swarm particle will be modified using objective function in this approach to attain the best simulation outcomes [9].

2.3. Cauchy mutation operator PSO algorithm(CPSO)

The Cauchy mutation operator can be integrated with PSO algorithm to obtain better optimized results than the standard PSO algorithm. The Cauchy distribution is a continuous probability distribution. By mutating the standard PSO algorithm with the Cauchy mutation operator the diversity can be maintained properly as well as the converge will be faster than the standard PSO. In addition to this the search space can be expanded to have better optimized results.

2.4. Hybrid Particle Swarm Optimization Support Vector Machine Algorithm (PSOSVM)

In this PSOSVM classification algorithm PSO is integrated with SVM for better classification results than the standard SVM classifier. The process of finding the optimal solution to a problem is called optimization. In contrast to other optimization algorithms PSO is simple, faster and easy to implement.SVM is extremely sensitive to changes in its parameter values, hence in this hybrid algorithm PSO is applied as an optimizer which can optimize the parameters of SVM to find better classification accuracy. This will ensure the hybrid algorithm's reliability by searching for the best SVM parameter values. [10].

2.5. Hybrid Crazy Particle Swarm Optimization Support Vector Machine Algorithm (CPSOSVM)

In CPSOSVM algorithm a modified version of PSO namely crazy PSO (CPSO) is integrated with SVM for data classification task. Many swarm intelligence methods employ the term "craziness" to describe an unusual shift of searching direction in an optimization method. By adding a crazy operator to the standard PSO algorithm, the global searching capabilities of the standard PSO can be improved, and this modification is considered the best way to ensure that each particle has a set probability of craziness for better diversity maintenance [11].

3. Proposed Methodology

In this section the proposed algorithm for heart and liver disease prediction along with the heart and liver datasets is described in detail.

3.*1. Hybrid Cauchy Crazy Particle Swarm Optimization Support Vector Machine Algorithm (CCPSOSVM)*

In CCPSOSVM algorithm a The Cauchy mutation operator is used in association with CPSOSVM method to obtain more accurate classification results on heart and liver disease datasets. It is not uncommon to use the mutation operator in an optimization technique. It is believed that the Cauchy distribution is a continuous probability distribution. The main goal is to introduce little fluctuation in the total population periodically in order to maintain its diversity [12]. Prior to the convergence; particles in PSO should flip amongst the former finest particle as well as the finest global particle as perceived by every particle, according to specified experiment results. The exploration space for the finest particle would be expanded if exploration neighbors of the most excellent overall particle were added in every iteration, which will help every particle move to the finest possible point. At each iteration, this can be achieved using a Cauchy mutation operator on the overall most excellent particle.

i. The course of action of the proposed CCPSOSVM algorithm is summarized as follows:
ii. Start
iii. Read the dataset.
iv. Normalization of the input dataset.
v. 80% of data is taken for training where as 20% of data is taken for testing.
vi. The SVM model is constructed.
vii. Train the SVM model with training data structure.
viii. The SVM model is optimized by using CPSO.
ix. Cauchy mutation operator is applied.
x. Model constructed by using the muted optimized weight and bias.
xi. Simulate and testing the muted optimized hybrid model.
xii. End

Proposed methodology flow diagram

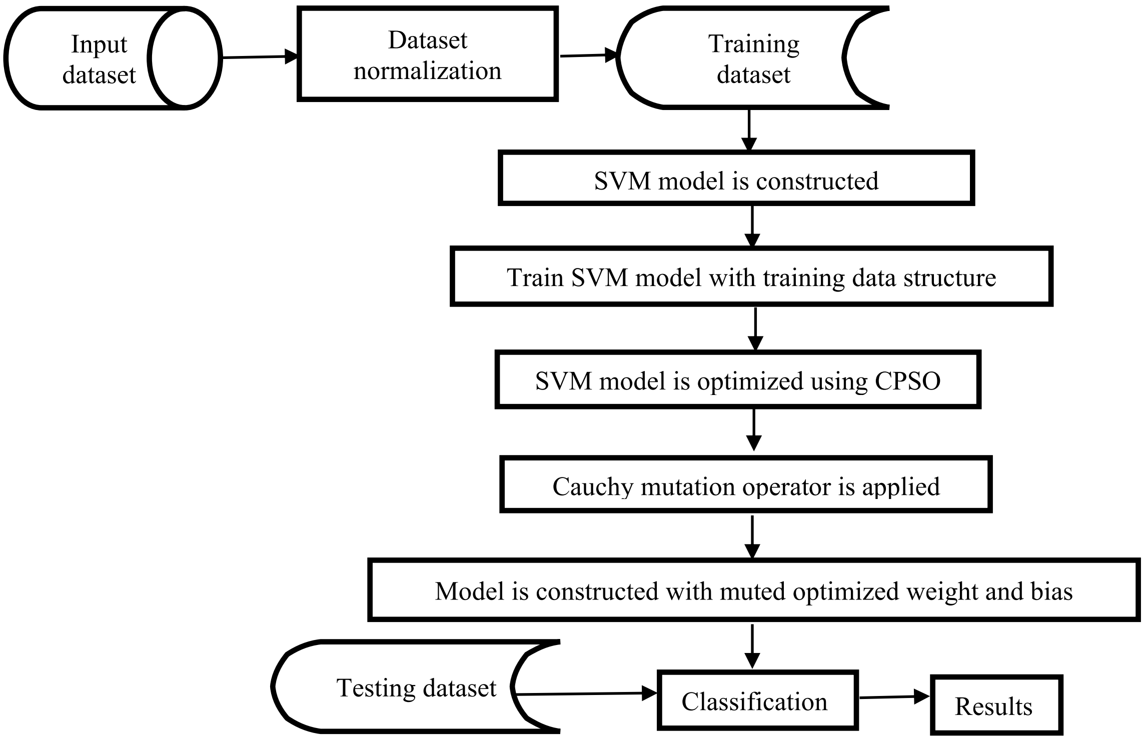

Fig. 1. Proposed methodology flow diagram

3.2. Heart Dataset

The heart disease dataset is obtained from the University of California repository of machine learning at the, Irvine (UCI). There are 13attributes and 270 instances present in the heart dataset. The variable to be predicted is 1 for heart disease absence or 2 for presence of heart disease [13].

3.3. Liver Dataset

The liver dataset is obtained from UCI repository of machine learning. There are 10 attributes and 583 instances present in this dataset [14].

4. Result Analysis

The outcomes are presented in this section from SVM, PSOSVM, CPSOSVM and CCPSOSVM on heart and liver dataset will be analyzed based on accuracy, error, precision, recall, and F1 score. The dataset is divided into two groups in the experiment: training and testing. The training set's ratio is 80 percent and 20 percent, respectively. The experiment is carried out using the Python programming language and the pandas, sci-kit learn and pyswarm libraries.

Performance Measures:

Accuracy: The most common metric for evaluating the performance of classification algorithms. It's calculated as a proportion of overall correct predictions to total correct forecasts.

Classification Error: A kind of measuring fault in which the respondents do not offer a true response to a surveyed item is known as classification error.

Precision: It is used in document access and is expressed as the number of documents accurately produced by the machine learning algorithm.

Recall: The number of times your machine learning model gives a beneficial outcome is called recall.

F1 score: This is estimated as the weighted average of precision and recall.

The Scikit learn library as well as the Python programming language were used to conduct all of the research on machine learning approaches discussed in this study. SVM classifier is trained with 200 iterations and to optimize the SVM classifier modified PSO algorithm is used. The primary goal of this research is to find the best classification results for heart and liver disease dataset. A simple rule is to split the 0.8 databases into training set 0.2 databases for testing randomly.

Table 1. Parameters used in the proposed CCPSOSVM.

Parameters Used	CCPSOSVM
W	0.5
C1	0.5
C2	0.9
Vdcraziness	0.95
Pr	1
Sgnr	-1
N	1
Iteration	200
Target Error	1
Number of Particles	20

Table 1 represents the parameters used for the experiment on heart and liver datasets of the proposed CCPSOSVM algorithm. In this Table the parameters used along with values is represented. W is 0.5, C1 is 0.5, C2 is 0.9, Vdcraziness is 0.95, Pr is 1, Sgnr is -1, N is 1, no of iterations is 200, target error is 1 and 20 particles.

Table 2. Classification results for Heart dataset.

Algorithms	Precision	Recall	F1 Score
SVM	90.91	88.24	89.55
PSOSVM	93.94	88.57	91.17
CPSOSVM	96.97	88.89	92.75
CCPSOSVM	91.89	97.14	94.44

Table 3. Classification results for Liver dataset.

Algorithms	Precision	Recall	F1 Score
SVM	100	62.93	77.24
PSOSVM	100	83.62	91.07
CPSOSVM	100	96.55	98.24
CCPSOSVM	100	97.41	98.68

Table 4. Table showing accuracy and error (in %) of various classifiers used for Heart dataset.

Sl.No.	Algorithms	Accuracy	Error
1	SVM	87.04	12.96
2	PSOSVM	88.89	11.11
3	CPSOSVM	90.74	09.26
4	CCPSOSVM	92.59	07.41

Table 5. Table showing accuracy and error (in %) of various classifiers used for Liver dataset.

Sl.No.	Algorithms	Accuracy	Error
1	SVM	62.93	37.07
2	PSOSVM	83.62	16.38
3	CPSOSVM	96.55	03.45
4	CCPSOSVM	97.41	02.59

Table 2 represents the classification results of SVM, PSOSVM, CPSOSVM and CCPSOSVM on heart dataset respectively. In Table 2 the precision along with recall as well as F1 score for all above discussed classification algorithm for heart disease prediction is presented. From Table 2 we observed that the planned algorithm CCPSOSVM is enhanced in comparison to other algorithms with 91.89% of precision, 97.14% of recall and 94.44% of F1 score for heart dataset. Table 3 represents the classification results of the techniques such as SVM, PSOSVM, CPSOSVM and CCPSOSVM on liver dataset respectively. In Table 3 the precision along with recallas well as F1 score for all the above discussed classification algorithm for liver disease prediction is presented. Table 3 gives 100% precision, 97.41% of recall and 98.68% of F1 score for liver datasets. Table 4 shows the best accuracy produced by each algorithm employing the best optimal hyper-parameters on heart dataset. From Table 4 the proposed CCPSOSVM achieved 92.59% of accuracy with 07.41% error rate on heart dataset. Table 5 shows the best accuracy produced by each algorithm employing the best optimal hyper-parameters liver dataset. From Table 5 the proposed CCPSOSVM gets 97.41% accuracy with 02.59% of error rate for liverdataset.

The confusion matrix is a tabular form of error matrix showing results of classification algorithms. In which each row denotes to the targeted class whereas each column is for output class or vice versa. Figure 2 is the confusion matrix for SVM algorithm for heart disease prediction. A confusion matrix is applied to know the results of the proposed model. From this figure the TN value is 30, FN is 4, TP is 17 whereas FP is 3.

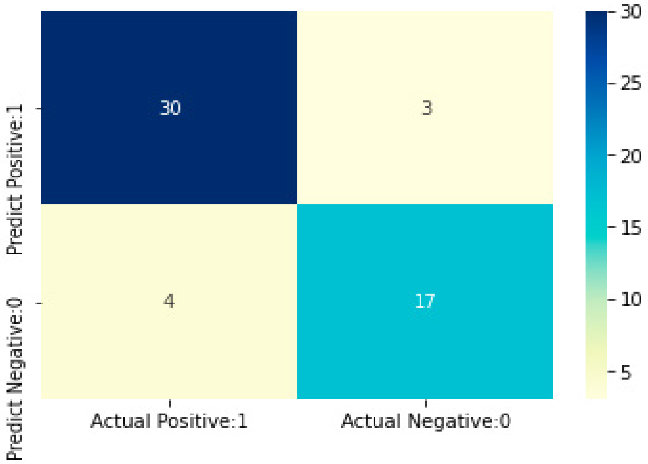

Fig. 2. Confusion matrix for SVM on Heart dataset

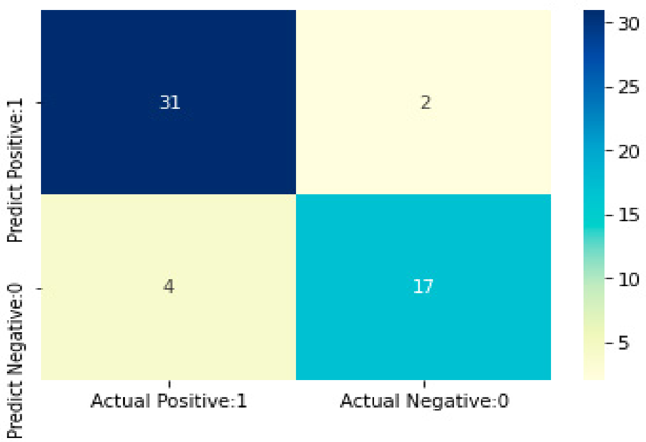

Fig. 3. Confusion matrix for PSOSVM on Heart dataset

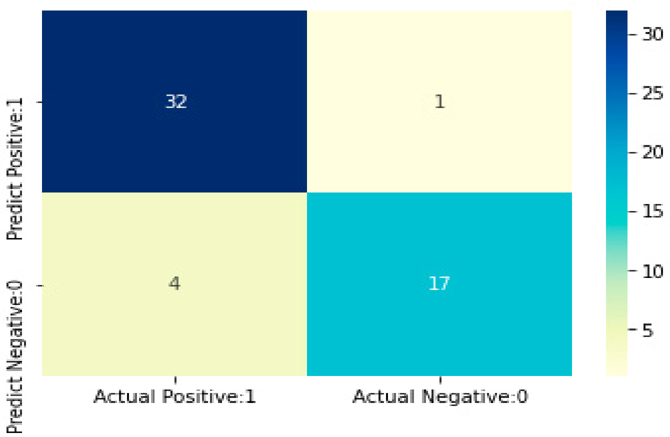

Fig. 4. Confusion matrix for CPSOSVM on Heart dataset

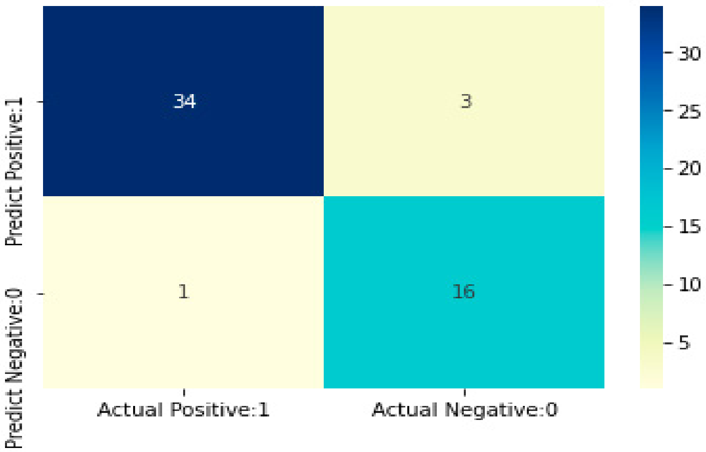

Fig. 5. Confusion matrix for CCPSOSVM on Heart dataset

Figure 3 is representing the confusion matrix for PSOSVM algorithm for heart disease prediction. The concept of confusion matrix is applied for the result analysis of the proposed prototypical. From the figure the TN value is 31, FN is 4, TP is 17 whereas FP is 2. Figure 4 is the confusion matrix for CPSOSVM algorithm for heart disease prediction. The confusion matrix is used to know the performance of the proposed model. From this figure the TN value is 32, FN is 4, TP is 17 whereas FP is 1. Figure 5 is the confusion matrix concept for CCPSOSVM algorithm for heart disease prediction. The confusion matrix is applied to know the proposed model performance. From this Figure the TN value is 34, FN is 1, TP is 17 whereas FP is 3.

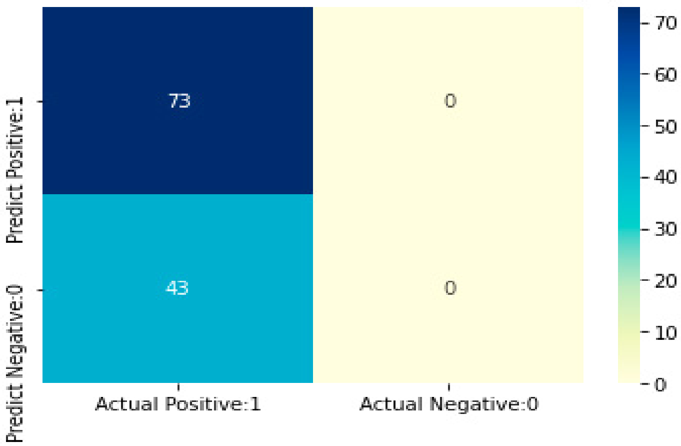

Fig. 6. Confusion matrix for SVM on Liver dataset

Figure 6 is the confusion matrix for SVM algorithm for Liver disease prediction. The confusion matrix is applied to know the performance of the proposed prototypical. From this figure the TN value is 73, FN is 43, TP is 0 whereas FP is 0. Figure 7 is the confusion matrix for PSOSVM algorithm for Liver disease prediction. The confusion matrix is applied to know the proposed modelling performance. From this figure the TN value is 97, FN is 19, TP is 0 whereas FP is 0. Figure 8 is the confusion matrix for CPSOSVM algorithm for Liver disease prediction. From this figure the TN value is 112, FN is 4, TP is 0 whereas FP is 0. Figure 9 is the confusion matrix concept for CCPSOSVM algorithm for Liver disease prediction. The confusion matrix is applied to know the performance of the proposed model. The TN value is 113, FN is 3, TP is 0 whereas FP is 0.

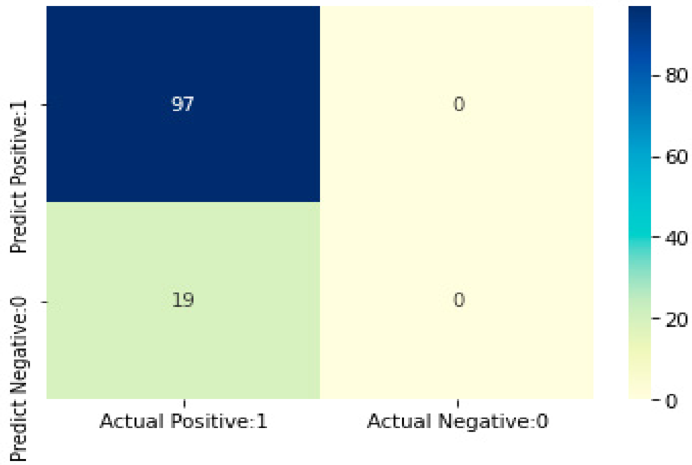

Fig. 7. Confusion matrix for PSOSVM on Liver dataset

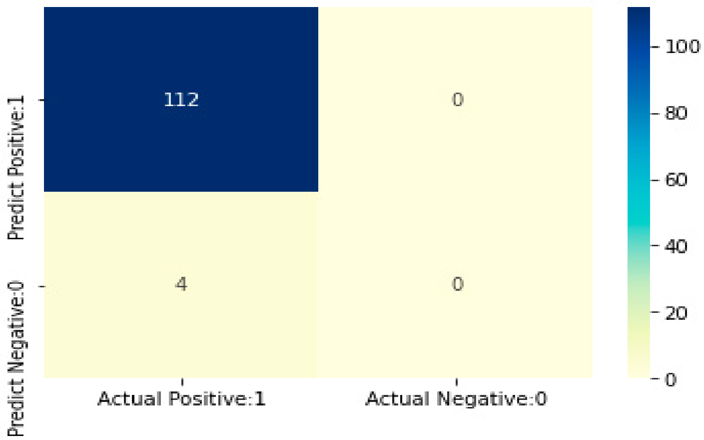

Fig. 8. Confusion matrix for CPSOSVM on Liver dataset

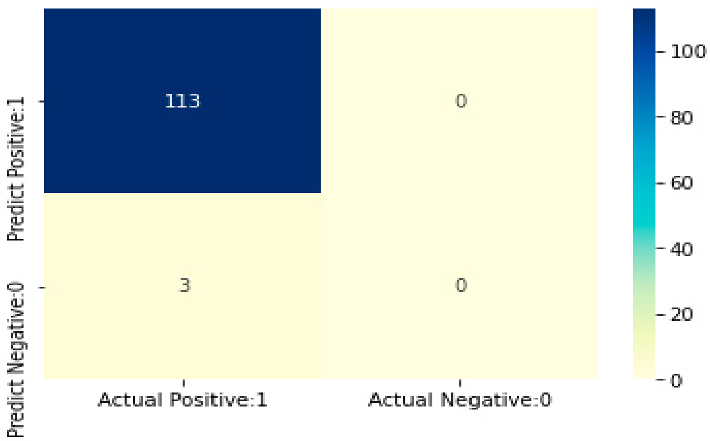

Fig. 9. Confusion matrix for CCPSOSVM on Liver dataset

Figure 2 to Figure 9 represents the confusion matrix for SVM, PSOSVM, CPSOSVM and CCPSOSVM algorithms on heart and liver dataset. Every confusion matrix is a tabular representation of an error matrix including classification algorithm results. These confusion matrix figures has been used for the performance analysis of the above discussed classification algorithms. From these confusion matrix figures the classification accuracy, error, precision, recall and F1 score is calculated.

5. Conclusion

Four algorithms like SVM, PSOSVM, CPSOSVM and CCPSOSVM employed on heart and liver disease prediction. The performance of each individual algorithm has been calculated as well as analyzed with respect to the confusion matrix, classification accuracy, classification error rate, precision, recall and F1 score. The conclusion can be drawn with proper experimental analysis is that planned CCPSOSVM gives improved classification results, with highest classification rate and lowest error rate for heart and liver disease prediction. The experiments are carried out only heart and liver prediction in the future; the presented hybrid algorithm could be used to forecast various diseases. We plan to use our proposed hybrid algorithm with additional parameters on larger data sets with a wider range of diseases to enhance accuracy in our further study.

References

[1] Vijayashree, J., Sultana, H.P. A Machine Learning Framework for Feature Selection in Heart Disease Classification Using Improved Particle Swarm Optimization with Support Vector Machine Classifier. Program Comput Soft 44, 388–397 (2018). https://doi.org/10.1134/S0361768818060129

[2] Raj, S., Ray, K. C., & Shankar, O. (2016). Cardiac arrhythmia beat classification using DOST and PSO tuned SVM. Computer methods and programs in biomedicine, 136, 163-177.

[3] Vijayarani, S., &Dhayanand, S. (2015). Liver disease prediction using SVM and Naïve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4(4), 816-820.

[4] k. Thirunavukkarasu, A. S. Singh, M. Irfan and A. Chowdhury, "Prediction of Liver Disease using Classification Algorithms," 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-3, doi: 10.1109/CCAA.2018.8777655.

[5] A. Singh and R. Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," 2020 International Conference on Electrical and Electronics Engineering (ICE3), 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958.

[6]Senbagavalli, M., & Arasu, G. T. (2016). "Opinion Mining for Cardiovascular Disease using Decision Tree based Feature Selection." Asian Journal of Research in Social Sciences and Humanities, 6(8), 891-897.

[7] Valli, M. S., and G. T. Arasu. "An Efficient Feature Selection Technique of Unsupervised Learning Approach for Analyzing Web Opinions." (2016).

[8] D. R, K. R, P. G, P. S and S. R. A K, "Breast Cancer Classification using the Supervised Learning Algorithms," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1492-1498, doi: 10.1109/ICICCS51141.2021.9432293.

[9] J. Kennedy and R. Eberhart, "Particle swarm optimization," Proceedings of ICNN'95 - International Conference on Neural Networks, 1995, pp. 1942-1948 vol.4, doi: 10.1109/ICNN.1995.488968.

[10] A. Kumar, A. Ashok and M. A. Ansari, "Brain Tumor Classification Using Hybrid Model of PSO And SVM Classifier," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2018, pp. 1022-1026, doi: 10.1109/ICACCCN.2018.8748787.

[11] S. K. Sarangi, R. Panda and A. Sarangi, "Crazy firefly algorithm for function optimization," 2017 2nd International Conference on Man and Machine Interfacing (MAMI), 2017, pp. 1-5, doi: 10.1109/MAMI.2017.8307875.

[12] H. Wang, C. Li, Y. Liu and S. Zeng, "A Hybrid Particle Swarm Algorithm with Cauchy Mutation," 2007 IEEE Swarm Intelligence Symposium, 2007, pp. 356-360, doi: 10.1109/SIS.2007.367959.

[13] K. Bache and M. Lichman. (2013). UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA, USA. [Online]. Available: http://archive.ics.uci.edu/ml/

[14] K. Bache and M. Lichman. (2013). UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA, USA. [Online]. Available: http://archive.ics.uci.edu/ml/