Musik in bayern

ISSN: 0937-583x Volume 88, Issue 10 (Dec -2023)
https://musikinbayern.com

DOI https://doi.org/10.15463/gfbm-mib-2023-*257*

EXPLORATION OF DISCRETE OPTIMIZATION ALGORITHM FOR FEATURE SELECTION ESTABLISHED ON MACHINE LEARNING(ML) FOR PREDICTION OF CHRONIC KIDNEY DISEASE (CKD)

T.SAROJA [1],

[1]Research Scholar,

Department Of Computer Science,

VISTAS, Pallavaram,

Chennai.

tsaroja11@gmail.com

DR.Y.KALPANA [2]

[2]Professor,

Department Of Information Technology,

VISTAS, Pallavaram, Chennai

kalpana.scs@velsuniv.ac.in

To Cite this Article

T.SAROJA , DR.Y.KALPANA ," EXPLORATION OF DISCRETE OPTIMIZATION ALGORITHM FOR FEATURE SELECTION ESTABLISHED ON MACHINE LEARNING(ML) FOR PREDICTION OF CHRONIC KIDNEY DISEASE (CKD) *"Musik In Bayern, Vol. 88, Issue 10, Dec 2023, pp225-257*

Article Info

Received: 24-11-2023 Revised: 04-12-2023 Accepted: 14-12-2023 Published: 24-12-2023

ABSTRACT

CKD is a worldwide health problem that causes morbidity and death. People typically ignore CKD because there are no obvious early symptoms. Individuals who are diagnosed with CKD early on can prevent the disease from progressing. Due of their speedy and exact determination capabilities, machine learning (ML) approaches can greatly assist clinicians in accomplishing this goal. In most databases, there are incorrect features that can degrade the effectiveness of the classifier. These features are identified using Feature Selection Algorithms. Following the identification of features, ML techniques are used to solve the classification task.

In this paper a wide variety of Optimization Algorithms are studied for feature selection(FS).

KEYWORDS:

Chronic Kidney Disease(CKD), Feature Selection(FS), Ant – colony Optimization Algorithm(ACO),Machine Learning(ML), Fruit- Fly Optimization(FFO) Algorithm**,** Teacher Learner Based Optimization (TLBO) Algorithm, Improved Teacher Leaner Based Optimization(ITLBO), Particle Swarm Optimization (PSO), Butterfly Optimization Algorithm(BOA), Whale Optimization Algorithm(WOA):

1. INTRODUCTION

CKD is currently seen as serious threat to society's health. CKD is a blanket term encompassing a number of disorders that affect the shape and functionality of the kidneys. It has a significant fatality ratio. It is characterised as the existence of renal injury (albumin,uria) or impaired renal function over a three-month or longer period of time. [1] Renal failure is the most serious complication of CKD, and its symptoms are primarily caused by renal function problems.. Dialysis and transplantation are the sole alternatives for treatment when symptoms are severe; End-stage renal disease is the name for this type of

ISSN: 0937-583x Volume 88, Issue 10 (Dec -2023)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2023-*257*

kidney failure. CKD failure affects four persons in every 1000 people in the UK. It is difficult to detect CKD in its early stages due to the limited availability of practising nephrologists and their incapacity to fully administrate individuals with CKD. As a result, developing an accurate, convenient, and automated CKD diagnosis approach is critical. Machine Learning contribute in the early identification of the disease.

The dataset contains irrelevant data. Those data has to be removed to improve the accuracy. This selection of the relevant attributes is performed by various Feature Selection(FS) algorithm. One among those is the Optimization Algorithm. There are a lot of Optimization Algorithm.

In this paper, the various optimization algorithm used for the FS of the attribute from the dataset in diagnosis of CKD is discussed.

2. METHODS

2.1 FS:

Knowledge discovery, pattern identification, and statistical research all rely on FS. Goal of FS is to eliminate a subtype of non-essential code. Information concerning predicted classes has no bearing on features. By minimizing the dimensionality of attributes and removing un -important features, a thorough classification pattern can be created. Identifying the optimal subset of characteristics to obtain the greatest results is the main issue in FS.

The fundamental goal of FS is to achieve the best categorization results possible. FS not only simplifies data realisation, but it reduces the overfitting problem and data storage size, lowering costs and increasing accuracy.

2.1.1TYPES:

Filter algorithm

There are no learning algorithms used in the filter algorithms. The filter method chooses the features with the highest ranks, and the subset of features chosen can then be used in any prediction algorithm. Filter won't provide efficiency but solves overfitting problem.

Wrapper algorithm

The space of all highlights of subsets is picked up by the wrapped search algorithm around the classifier. Wrapper solves efficiency but not solves the outfitting problem.

Embedded algorithm

Embedded algorithm is combination of Filter and Wrapper method. It solves outfitting problem and also provide efficiency.

ACO Algorithm:

ACO is a meta-information strategy for identifying solutions to conjunctional optimization problems that is repeatable and likely. It is a method of exploration used by biological ants to find a little error on their way from their colony to the food supply. During the forage, the ants make indirect contact with each other thanks to their scent, which helps to record the corresponding training samples (paths) and entices the remaining ants. Artificial ants (agents) perform the virtual smell in the ACO algorithm to alter their course using the decision graph.[2]

ISSN: 0937-583x Volume 88, Issue 10 (Dec -2023)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2023-*257*

In the proposed study, The ACO-based feature selection approach minimzes the quantity of features in the CKD dataset. Classifier with Extreme Learning Machine (ELM) is formulated for the diagnosis of CKD which increases the classification accuracy.

The ELM classifier with the ACO algorithm achieves a precision of 90.03%, but other approaches like KNN and NN only achieve 74.85 % and 80.89 %, respectively.

FFO ALGORITHM:

FFOA is a different idea based on the fruit fly's food identification principle. When compared to other species, it is better at detecting and analysing olfactory and visual signals. Fruit flies' olfactory organs are capable of detecting a greater range of pleasant odours in the environment, as well as pointing to a food supply even from a great distance away. It can detect food using any sensitive vision and fly toward the location once the food in the immediate vicinity has been consumed. As a result, a fruit fly's finding food is summarised as follows: (a) to start with, using the olfactory organ to analyse the food source while also flying to a desired position; (b) otherwise, using sensitive looks to get closer to the food location; and (c) finally, switching the location of the fruit flies flocking and flying in the direction. FFOA is used to choose the best features from a wide number of options. UCI CKD dataset's selected characteristics are processed and sent into the Multi-kernel SVM (MKSVM)for classification purposes. [3] The study reaches 98.5% classification accuracy.

BAT ALGORITHM(BA):

Bats are enthralling creatures. They are the only animals that have van and can use new echolocation. [5]. Insectivores make up the majority of microbats. Microbats use echolocation, a sort of sonar, to recognise prey, avoid obstacles, and discover roosting niches in the dark. These bats emit a loud rhythm and hear for the resounding it receives from their surroundings. Pulses have diverse characteristics based on the species and it be linked to stalking strategies.. Most bats use short, recurrence-modulated waves that sweep through an octave for echolocation, however some prefer constant-recurrence waves. Their signal bandwidth varies depending on the classification, and it is typically expanded by adding extra harmonics. BA is created to idealise some of the echolocation features of microbats. [4]

HYBRID FILTER WRAPPER EMBEDDED (HFWE) FS:

FS based on HFWE finds the best selection of attributes from the CKD dataset. The filter, wrapper, and embedding algorithm techniques are combined in the HFWE-FS algorithm. One-R, Relief, Gini Index(GI) and Gain Ratio (GR), are some of the important functions used in the filter technique. The Wrapper approach selects analytical characteristics from the CKD dataset using the Improved Bat Algorithm. (IBA). The embedded method use the SVMachine-t-statistics(SVM-t) to choose analytical features. The findings of all FS based on strategies are merged in the HFWE- FS method. [5]

In contast, to the SVM-HFWE-FS and ANN-HFWE-FS algorithms, the suggested SVM, HFWE-FS approach achieves 93.33 percent accuracy, which is 6.66 percent and 8.1 percent higher, respectively.

TLBO ALGORITHM

There are no algorithm-specific parameters that must be modified with TLBO. The TLBO algorithm is based on educating and studying as a natural activity. [7].

Musik in bayern

ISSN: 0937-583x Volume 88, Issue 10 (Dec -2023)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2023-*257*

ITLBO :

ITLBO uses the Chebyshev distance calculation to evaluate the capability task and ordinary command factors like population density and generation count to calculate the optimum attribute subgroup for timely detection of persistent diseases. The suggested FS mechanism was applied to the CKD dataset and resulted in a 36 percent attribute deplition, compared to a 25 percent attribute depliction using the original TLBO algorithm.

Accuracy of SVM, Convolution Neural Network, and Gradient Boosting classification algorithms is used to assess the TLBO algorithm's derived optimal feature subgroup and the ITLBO algorithm's feature subgroup. The experimental findings reveal that when the three techniques from the suggested FS algorithm for the gained attribute subgroup are compared to the native TLBO algorithm, the suggested FS technique completely improves classification accuracy. The ITLBO algorithm was used to adjust the native TLBO approach while computing the fitness function. The ITLBO-CNN approach had a higher accuracy of 95.25 percent than the other classification algorithms. [8].

GREY WOLF OPTIMIZATION (GWO):

Grey wolves have a greater rate of fortunate replication than pack stalking, according to GWO, a recently released inprogress algorithm.. Male and Female grey wolves hold a superior rank in the pack and are in charge of the other wolves [9].

GWO is one of the most cutting-edge superlative strategies for simulating the trapping behaviour of a bunch of grey wolves in the wild.

PSO :

PSO [10] is a fascinating global expansion method that replicates the social behaviour of bird flocks in a multidimensional environment in order to achieve specified goals. In the search space, PSO employs a population of particles (dubbed a swarm) [11,12]. The state is found by the location and acceleration of each piece. The PSO-KDE model has a better average performance in diagnosing kidney illness, according to the experimental results[13].

BOA:

To tackle worldwide expansion issues, the BOA is based on the food search and mating character of butterflies. The design is influenced by butterflies' foraging strategy, which entails using their sense of smell to find nectar or a mate. Three classic engineering problems are also solved with BOA (spring scheme, welded beam scheme, and gear train scheme). The proposed BOA is more well-organised than previous multiobjective algorithms, according to the results.

Each fragrance in the BOA collection has its grant distinct odour and unique nudge. It's among the main characteristics that distinguishes BOA from other multiobjective. To comprehend by what means odour is estimated in BOA we have to find how a stimulus processes a modality such as odour, light, warmth, sound and so on, The whole notion of feeling and fining the paradigm is built around three key terms: sensory paradigm (c), input intensity (I), and power exponent (P) (a). Sensory paradigm represents to the raw input used by the sensors. Sensory means to measure the form of energy and process it in similar ways.
Various paradigm now include odour, sound, light, warmth, and aroma in BOA. If a butterfly transmits a greater quantity of smell, other butterflies in the location can identify it and get attracted to it.

ISSN: 0937-583x Volume 88, Issue 10 (Dec -2023)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2023-*257*

Butterfly places are produced at random in the search location, and its scent and fitness usefulness are measure and saved. Thus early step is finished, and the algorithm moves on to the repetation stage, where the artificial butterflies are used to do the search.
Butterflies can look for its nourishment and companion on a confined and overall range. When physical presence and other conditions such as wind, rain and other elements are taken into account, the quest for nourishment can account for a large part p of a butterfly's overall companion-searching or food-searching activity. BOA uses a switch probability p , to move from common overall search to intensive confined forage.

The BOA's performance in simulations appears to be promising[14].

DBOA:

DBOA is an improved variant to BOA for FS problems. To reduce the defects of the original BOA, two significant developements are introduced in the original BOA: the use of LSAM to improve BOA solutions diversity.[15] and the development of a Local Search Algorithm Based on Mutation (LSAM) operator to avoid local optima problem.

WOA :

WOA is a new metaheuristic algorithm that simulates baleen whale foraging. Baleen whales hunt small fish near the surface by swimming in a decreasing circle and blowing distinctive bubbles in a circle or '9'-shaped pattern.WOA [16] is a new optimization algorithm that simulates humpback whales' smart food searching behaviour. WOA has a lot of advantages, including a small number of specifications to regulate (it only has two main inner specifications to alter), ease of execution, and great adaptability. Based on a single specification, the WOA algorithm easily shifts from exploration to exploitation..

2.2MACHINE LEARNING ALGORITHM

A lot of classification algorithms are chosen for the purpose of benchmarking the results. They are Back Propagation Neural Network (ANN), NB, DT, Decision trees, KNN, RF, Log- Reg and SVM.

3.DATA SET

The CKD database [17], which includes 400 patients, was obtained from the UCI ML Repository. The database comprises 24 attributes divided into eleven different attributes and 13 categorical attributes and 2 classes. Results are divided into patients with CKD disease and without CKD disease. Anemia, appetite, diabetes mellitus, RBC count, WBC count, haemoglobin, bacteria, potassium, packed cell volume, sodium, blood urea, coronary artery disease, pus cell clumps, RBCs, albumin,specific gravity, blood pressure, and age are some of the characteristics. The database is not equal, in which 250 samples specify "ckd" class (62.5 percent) and 150 of "notckd" (37.5 percent). It's worth mentioning that the data comprises a significant amount of information.

Description	Values
No. of Instances	400
No. of Features	24
No. of Class	2

ISSN: 0937-583x Volume 88, Issue 10 (Dec -2023)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2023-*257*

Percentage of Positive Samples	62.50%
Percentage of Negative Samples	37.50%
Data source	UCI

Dataset Description.

4.CONCLUSION

In this paper variety of Feature selection Optimization Algorithms are discussed. FS mainly concentrates on eliminating non-important features from the dataset. By reducing the quantity of attibutes efficiency increased and implementation time is decreased. Results are fed to the classifiers of ML for finding the performance which is measured using the metrics like F1 Score, Accuracy, Specificity,Sensitivity,Recall, Precision.

REFERENCES

[1]. Stevens, Levey, "Current status and future perspectives for CKD testing". American Journal of Kidney Diseases, Vol. 53, No. 3, Supplement 3,2009.
[2]. Belina, Sara "Ant Colony Optimization (ACO) Based Feature Selection And Extreme Learning Machine (ELM) For Chronic Kidney Disease Detection", Int. Journal of Advanced Studies of Scientific Research (IJASSR) ISSN 2460 4010
[3]. L. J. Rubini, E. Perumal, " Efficient classification of chronic kidney disease by using multi-kernel support vector machine and fruit fly optimization algorithm", DOI: 10.1002/ima.22406
[4] Komarasamy, Wahi, "An optimized K-means clustering technique using bat algorithm", European Journal of Scientific Research, 2012; 84(2), pp.263-273
[5]. R.Parthiban , S.Usharani., " Prognosis of Chronic Kidney Disease (CKD) using Hybrid Filter Wrapper Embedded Feature Selection method ",European Journal of Molecular & Clinical Medicine ISSN 2515-8260 Volume 07, Issue 09, 2020
[6]. H. Polat, H. Danaei, "Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods", DOI 10.1007/s10916-017-0703-x
[7]. Mohan Allam , M. Nandhini, "Feature Optimization using Teaching Learning Based Optimization for Breast Disease Diagnosis." Int. Journal of Recent Technology and Engineering. 78-85.
[8]. Manonmani , Sarojini Balakrishnan ,"Feature Selection Using Improved Teaching Learning Based Algorithm on Chronic Kidney Disease Dataset", 10.1016/j.procs.2020.04.178
[9]. S. Shoghian, M. Kouzehgar, "A comparison among wolf pack search and four other optimization algorithms", World Acad. Sci., Eng. Technol., Vol.6, 2012.

ISSN: 0937-583x Volume 88, Issue 10 (Dec -2023)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2023-*257*

[10]. Kennedy, Eberhart, "Particle swarm optimization", Proc. of the IEEE Int. Conference on Neural Networks, Vol. 4, Pp.1942–1948, 1995
[11]. Del Valle, Venayagamoorthy, "Particle swarm optimization: basic concepts, variants and applications in power systems", IEEE Trans. Evoln. Comput., Vol.12, No.2, Pp.171–195, 2008
[12]. Krusienski, Jenkins, "Nonparametric density estimation based independent component analysis via particle swarm optimization", IEEE Int. Conference on Acoustics, Speech, and Signal Processing, Proceedings, Vol. 4, 2005, March
[13]. Thamil Selvi, Soundharamanikandan," Feature Selection Using Binary Grey Wolf Optimization Algorithm on Chronic Kidney Disease Dataset ",Journal on Science Engineering and Technology, Volume 5.
[14]. S Arora, Satvir Singh,"Butterfly optimization algorithm: a novel approach for global optimization", https://doi.org/10.1007/s00500-018-3102-4
[15]. Mohammad Tubishat , Mohammed Alswaitti, "Dynamic Butterfly Optimization Algorithm for Feature Selection ",Digital Object Identifier 10.1109/ACCESS.2020.3033757
[16]. Majdi Mafarja , Seyedali Mirjalili , "Hybrid Whale Optimization Algorithm with Simulated Annealing for Feature Selection ",http://dx.doi.org/10.1016/j.neucom.2017.04.053
[17].Dua and Graff, ``UCI ML repository,". http://archive.ics.uci.edu/ml.