

A **Précis** on [#4099]

QL-SSA: An Adaptive Q-Learning based Squirrel
Search Algorithm for Feature Selection

Authored by Vandana Bharti, Bhaskar Biswas, Kaushal Kumar

Computer Science and Engineering
Indian Institute of Technology (BHU)
Varanasi, Uttar Pradesh, India

Student Name: Saif Ali Athyaab

Student ID: 23810756

Word Count: 2481

PART 1

The paper introduces a seemingly novel algorithm titled Q-learning Squirrel Search Algorithm (QL-SSA) which appears to be a combination of two separate algorithms. The researchers are based out of IIT-BHU, a reputed university from India. Although not important, a source of funding from corporate or university grants is not mentioned. The authors introduce data mining and machine learning as the latest popular methods for all sorts of real-world applications involving large scale data. An attempt to differentiate between the two methods has been made. Vandana et al. state 'Data mining aims to gain and apply the information that emerges from dealing with massive databases of varying sizes. In contrast, machine learning techniques are used in a supervised, unsupervised and semi-supervised way' [1, p. 1]. It is noticed that the attempt at differentiation between them has not been rightly fulfilled. While data mining is explained through its definition, machine learning has been explained only by three types of it (rather than focusing on its fundamentals). Classification is one of the popular data analysis methods dealing with class labels and helps identify meaningful patterns among huge data sets.

The subject is now progressing towards the main problem that the paper aims to solve – the curse of dimensionality.

Liu et al. state:

According to the curse of dimensionality, given limited available instances, when the number of features used is too large, the classification performance will be decreased by superfluous features. On the other hand, when too few features are used, the classification performance will decline because some informative features are neglected. Hence, reasonable feature subset evaluation is conducive to the confirmation of the optimal subset [2, p. 2].

Mohit et al. establish [3] that a new algorithm requires regressive testing to be considered by the status quo. According to them 'It is considered that difficulty of problem increases with the increase in function dimensions as the search space increases exponentially [3, p. 163]. Therefore, with the increase in the number of features (denoted by N), the number of possible states is 2^N , making feature selection a daunting task by utilizing even modern computational resources. They quote, 'single-agent RL algorithms face curse-of-dimensionality with increased size of state-action spaces..... detecting changing operating conditions puts additional burden on computation' [4, p. 7].

With the problem outlined and well-defined, the readers are introduced to the term of feature selection. 'The general aim is to reduce the amount of negative, redundant, and disruptive features in a dataset for fast and efficient data analysis while preserving accuracy. The most popular way to minimize the complexity of the data to be evaluated is to reduce the number of features or attributes to a more reasonable number' [1, p. 1]. Methods to achieve feature selection include combining supervised learning approaches with Evolutionary Algorithms (EAs). The way this works is the EA is wrapped with the supervised algorithm to search in the subset of the dataset features (columns) and then picking the best feature subset basis the supervised algorithm accuracy. However, 'local optima stagnation' [1, p. 1] is cited as an issue when working with EAs. This is a mathematical term for when a maximising function gets stuck within a local maxima and is not able to reach the global maxima due to parameters like learning rate, lack of adaptivity of the function [5]. It is here that reinforcement learning algorithms are introduced. The argument by the authors is - since learning would be performed by the experience of successful

generations, reinforcement learning would be suitable in this case. Also, it can employ various learning agents and a single update formula to reduce the computational impact. Because reinforcement learning is naturally inclined towards showcasing a diversity in solutions, local optima stagnation could be avoided. However, the authors seemingly contradict this point by saying, 'We applied the Q-learning concept to enhance the relocation mechanism of squirrels in searching for an optimal food source. It also enhances the diversification of the solution due to the insensitive characteristics of QL towards the exploration point' [1, p. 2]. If Q-learning's strength was its exploration and diversification abilities, how is Q-learning insensitive towards exploration?

The literature review has been well-conducted and covered all the previous attempts at feature selection including those involving wrapper-based approaches. The authors introduce their algorithm to overcome their shortcomings by seemingly maintaining a proper balance between exploration and exploitation. According to the authors, 'local optima stagnation, solution diversity, slow convergence, and inability to make a decision based on prior experience during the search strategy of existing optimizers' [1, p. 2], these are the main shortcomings of previous approaches. This seems to be the driving factor when proposing this novel combination of Q-learning and SSA. The authors define, 'Q learning is a reinforcement learning algorithm that learns by interacting with the environment in order to obtain the best reward based on the outcomes of previous actions' [1, p. 2]. Essentially, the authors aim to provide intelligence through experience to the SSA algorithm. In its raw sense, Gen_x of squirrels learn from the experience of Gen_{x-1} and so on, which truly reflects the process of evolution. Recursive Q-learning is employed in the paper after consulting from the Reference [4], where other variants of Q-learning algorithms (Repeated Update QL, Q-FPL) are briefly explained. This demonstrates the experimentation and analytical ability while selecting an appropriate variant.

Coming to the second component of the QL-SSA algorithm, the authors have done extremely well to educate the readers on the fundamentals of the SSA algorithm. In short, the SSA mathematically models the behaviour of flying squirrels particularly during their foraging process.

The authors assume:

the entire forest has three types of trees that correspond to the different food sources.....Hickory, Acorn, and Normal, representing optimal, sub-optimal, and trees without a food source.....During the warm season, they seek out abundant food sources such as acorns, and after meeting their daily energy requirements, they seek out optimal food sources such as hickory nuts..... store for times of adversity. In this way, they enhance their survival period.....Few of them are randomly relocated if they survive the winter season to reach the optimal food source, a global optimum. [1, p. 2,3]

Fig. 2 from [3, p. 152] visualises the gliding motion.

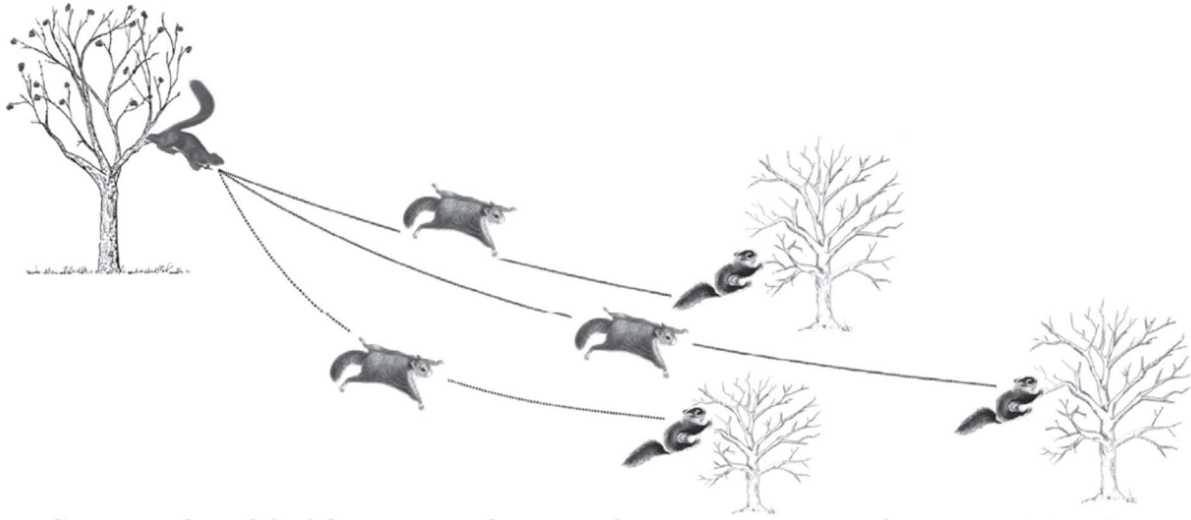


Fig. 2. Conceptual model of flying squirrel moving from one tree to another using gliding locomotion.

Iterative experiments have been performed and well-documented by Mohit et al. [3] when discussing the original Squirrel Search algorithm (SSA). The synopsis of these experiments is that SSA incorporated 3 parameters to maintain the require balance – Number of Nutritious food sources, Predator presence probability and Gliding Constant.

The proposed methodology integrates Q-learning with the SSA where, 'Each squirrel acts as a learning agent, performing some action in a given state and receiving an immediate reward, a long-term reward, or a penalty based on that action' [1, p. 3]. In this way the authors establish intelligence and passing of experience between generations. Like in any other wrapper-based approaches, each state is a binary vector, where 1 implies the feature has been included and 0 implies exclusion, with the goal of obtaining classification accuracy with these feature subsets. It has been rightly pointed out that, 'One of the most crucial steps in the algorithm's convergence is the relocation of squirrels to a tree with a richer food source' [1, p. 3], since it is in this stage that a Q-table comes into picture. A sample value of the Q-table would be the difference between the binary vectors of a squirrel's current state and future state (where it must be relocated). The authors have done a commendable job to explain how a squirrel interacts with the environment – 'First, it searches for all possible actions for a given state and chooses the one with the highest reward. This interaction is referred to as exploitation. The second option is choosing an action randomly, known as exploration' [1, p. 5]. Descriptive Algorithms and Flow charts are embellished to facilitate understanding for the reader. Adequate information in terms of pseudo-code and decision making is provided, although links to the source code have not been supplied which may hinder replication.

The experimentation is well-drawn from 20 datasets from the UCI machine learning repository; although it is unclear of the choice behind these datasets. It is understandable that only common datasets are used to provide a clearer picture of the performance against benchmark. However, since this paper's focus is on feature selection, it would have been a greater impact to also include some of the high-feature,

classification datasets as outlined in Appendix A. KNN and SVM-RBF are the two classifiers used to evaluate the model and provide accuracy. '10-fold cross-validation is performed' [1, p. 6] which is in-line with industry standards and then the best and average accuracies are computed and displayed. With the KNN classifier, 'QL-SSA outperforms the baseline approach on 15 out of 20 datasets [1, p. 6]. Here the baseline referred to is the base SSA feature search. In the SVM-RBF classifier, 'QLSSA is a clear winner on 16 out of 20 datasets and performs equivalently on the Iris dataset with 97.33% accuracy and 3 average number of feature set. However, it gives a competitive performance on 3 datasets (Sonar, Movement Libras, Ionosphere)' [1, p. 6,7]. It is here that the results get interesting. It is concluded that due to the dimension of these 3 datasets, the small number of generations would not provide the best possible number. Hence, the authors have re-examined and tweaked the number of generations to 50, thereby increasing the chance to converge optimally.

In conclusion, the results obtained in the paper are consistent with the baseline SSA approach. While the proposed method does not significantly outperform the baseline method in all cases, it has been argued that by not doing so it is a dependable and robust algorithm. The cases where the accuracy is better for QL-SSA, the delta is not very significant. They support their statement by saying, 'Results show that it is capable of maintaining a good balance between intensification and diversification, which is an important characteristic of an efficient optimizer' [1, p. 7]. The main drawback of the comparative analysis is the absence of other EA wrapper-based optimisers. This is only mentioned as a future scope of work. It would have been a great value proposition if other EAs were included such as ant colony, grey wolf, honeybee and so on. Due to this deficit, it is unclear to the reader whether the given model outperforms other EA wrapper-based optimisers. From the Table 1 [1, p. 5], it is inferred that medical-related datasets cover images since their features are drawn from MRIs, X-ray scans and alike. So, it can be inferred that the model is applicable to datasets of other types like text, image, or video after code alterations. The authors also speak about parallel computation, experimentation with 'deep' [1, p. 7] features and solving 'large-scale optimization problems in bioinformatics' [1, p. 7] in the future.

PART 2

ChatPDF, an online chatbot, powered by the latest capabilities of ChatGPT has been used for this part of the assignment. Initially the bot seemed to perform well against generic questions. The responses had a well-rounded composition drawing from the abstract and other sections as well.

Appendix B [Question#]	Bot Response
Provide a summary of the entire paper section-wise <i>Notice the accuracy in the number of sections with the proper heading. It has also figured out identification of pseudocode.</i>	The paper is organized into five sections : Section I: Introduction The introduction provides an overview of the problem..... Section III: Proposed Method They also provide pseudocode for the algorithm.
Describe the Squirrel Search Algorithm <i>Notice the deep level of understanding. It has managed to connect SSA with heuristic terms (this was missing in the paper)</i>	The movement of each squirrel is guided by three main factors: attraction to food sources (i.e., good solutions), repulsion from other squirrels (i.e., avoiding local optima), and random exploration (i.e., promoting diversity).

Below I discuss the positive and negative outcomes from comparisons.

Appendix B [Question#]	Me / Paper	Bot Response
[2] Can the QL-SSA algorithm be applied to different types of datasets, such as image or text data? <i>Notice the bot only declares possibility whereas a human is able to infer by understanding the context of the datasets.</i>	From the Table 1 [1, p. 5], it is inferred that medical-related datasets cover images since their features are drawn from MRIs, X-ray scans and alike. So, it can be inferred that the model is applicable to datasets of other types like text, image, or video after code alterations.	The PDF file does not provide specific information on whether the QL-SSA algorithm can be applied to different types of datasets, such as image or text data.....These datasets have significant variation in terms of number of classes, number of features, and samples. Therefore, it is possible that the QL-SSA algorithm can be applied to different types of datasets beyond those used in this study.
[3] Describe the QL-SSA algorithm in a mathematical sense <i>The bot fails to capture the formulae</i>	The paper documents everything in Section 3 [1, p. 4,5]	The PDF file does not provide a complete mathematical description of the QL-SSA algorithm.
[7] Critically evaluate the outcome of the paper	The main drawback of the comparative analysis is the absence of other EA	There are many other optimization algorithms that could be compared with QL-SSA, such as genetic

<p><i>Notice the similarities here. The bot has managed to draw out this argument on its own. It has also emphasised on dataset types (image/text/video) and sizes (having large number of features).</i></p>	<p>wrapper-based optimisers.</p> <p>it would have been a greater impact to also include some of the high-feature, classification datasets as outlined in Appendix A</p>	<p>algorithms, particle swarm optimization, or ant colony optimization. Secondly, the authors do not provide a detailed analysis of the performance of their algorithm on different types of datasets or different sizes of datasets.</p>
---	--	---

An unfortunate (blessing in disguise) thing happened then. I lost access to the particular prompt instance and had to start afresh. It was now that I observed one of the bot's shortcomings. After I asked this question - **Provide an analysis of the Results. Support your arguments with numbers** [Appendix B, 11] - the bot's memory got stuck within its own response. I feel this has something to do with the cons of Long Short-Term Machines (LSTM) on which ChatPDF is based. Thus, any question I asked after this, it regurgitates from this response only. Refer to: **Write a summary of the PDF file** [Appendix B, 12]. Notice that more than 70% of its response covers only the Results heading. It fails to capture other sections. Same is the case for [Appendix B, 13]. Somehow it introduces QPSO algorithm which is not mentioned in the paper at all. Hence, it is concluded that the bot is not a trustworthy companion in most cases.

REFERENCES

- [1] V. Bharti, B. Biswas and K. K. Shukla, "QL-SSA: An Adaptive Q-Learning based Squirrel Search Algorithm for Feature Selection," 2022 IEEE Congress on Evolutionary Computation (CEC), Padua, Italy, 2022, pp. 1-7, doi: 10.1109/CEC55065.2022.9870311.
- [2] W. Liu and J. Wang, "Recursive elimination current algorithms and a distributed computing scheme to accelerate wrapper feature selection," *Information Sciences*, vol. 589, pp. 636–654, 2022.
- [3] Jain, M., V. Singh, and A. Rani. "A novel nature-inspired algorithm for optimization: squirrel search algorithm. *Swarm Evol Comput* 44: 148–175." (2019).
- [4] S. Padakandla, "A survey of reinforcement learning algorithms for dynamically varying environments," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–25, 2021.
- [5] V. Bharti, B. Biswas, and K. K. Shukla, "EMOCGAN: a novel evolutionary multiobjective cyclic generative adversarial network and its application to unpaired image translation," *Neural computing & applications*, vol. 34, no. 24, pp. 21433–21447, 2022, doi: 10.1007/s00521-021-05975-y.

APPENDIX A

UCI Large Datasets

The following table has been extracted after filtering on the UCI ML repository (<https://archive.ics.uci.edu/ml/datasets.php?format=&task=cla&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=table>)

Only **Classification** problems having **100 or more attributes** are enlisted amounting to **91 datasets**. Almost none of these datasets are used in the paper.

Name	Data Types	Default Task	# Instances	# Attributes	Year
A study of Asian Religious and Biblical Texts	Multivariate, Text	Classification, Clustering	590	8265	2019
Activity recognition using wearable physiological measurements	Multivariate	Classification	4480	533	2019
Amazon Commerce reviews set	Multivariate, Text, Domain-Theory	Classification	1500	10000	2011
APS Failure at Scania Trucks	Multivariate	Classification	60000	171	2017
Arcene	Multivariate	Classification	900	10000	2008
Arrhythmia	Multivariate	Classification	452	279	1998
Breath Metabolomics	Multivariate, Time-Series	Classification, Clustering	104	1656	2019
Character Font Images	Multivariate	Classification	745000	411	2016
CNAE-9	Multivariate, Text	Classification	1080	857	2012
Condition monitoring of hydraulic systems	Multivariate, Time-Series	Classification, Regression	2205	43680	2018
Crop mapping using fused optical-radar data set	Multivariate, Time-Series	Classification	325834	175	2020
Daily and Sports Activities	Multivariate, Time-Series	Classification, Clustering	9120	5625	2013
Data for Software Engineering Teamwork Assessment in Education Setting	Sequential, Time-Series	Classification	74	102	2017
DBWorld e-mails	Text	Classification	64	4702	2011
Deepfakes: Medical Image Tamper Detection	Multivariate	Classification	20000	200000	2020
DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels	Text	Classification	12234	8519	2016

Detect Malacious Executable(AntiVirus)	Multivariate	Classification	373	513	2016
Detect Malware Types	Multivariate, Time-Series, Text	Classification	7107	280	2019
detection_of_IoT_botnet_attacks_N_Bal oT	Multivariate, Sequential	Classification, Clustering	7062606	115	2018
Dexter	Multivariate	Classification	2600	20000	2008
Dorothea	Multivariate	Classification	1950	100000	2008
Dota2 Games Results	Multivariate	Classification	102944	116	2016
DrivFace	Multivariate	Classification, Regression, Clustering	606	6400	2016
Dynamic Features of VirusShare Executables	Multivariate, Time-Series	Classification, Regression	107888	482	2017
Epileptic Seizure Recognition	Multivariate, Time-Series	Classification, Clustering	11500	179	2017
Facebook Large Page-Page Network	Multivariate	Classification	22470	4714	2020
Farm Ads	Text	Classification	4143	54877	2011
FMA: A Dataset For Music Analysis	Multivariate, Time-Series	Classification, Clustering	106574	518	2017
Gait Classification	Multivariate	Classification	48	321	2020
Gas Sensor Array Drift Dataset	Multivariate	Classification	13910	128	2012
Gas Sensor Array Drift Dataset at Different Concentrations	Multivariate, Time-Series	Classification, Regression, Clustering, Causa	13910	129	2013
Gas sensor array exposed to turbulent gas mixtures	Multivariate, Time-Series	Classification, Regression	180	150000	2014
Gas sensor array under flow modulation	Multivariate, Time-Series	Classification, Regression	58	120432	2014
Gas sensor arrays in open sampling settings	Multivariate, Time-Series	Classification	18000	1950000	2013
Gastrointestinal Lesions in Regular Colonoscopy	Multivariate	Classification	76	698	2016
gene expression cancer RNA-Seq	Multivariate	Classification, Clustering	801	20531	2016
Gisette	Multivariate	Classification	13500	5000	2008
GitHub MUSAE	Multivariate	Classification	37700	4006	2019
Hill-Valley	Sequential	Classification	606	101	2008

Human Activity Recognition Using Smartphones	Multivariate, Time-Series	Classification, Clustering	10299	561	2012
IDA2016Challenge	Multivariate	Classification	76000	171	2017
Influenza outbreak event prediction via Twitter data	Multivariate	Classification	75840	525	2020
Internet Advertisements	Multivariate	Classification	3279	1558	1998
ISOLET	Multivariate	Classification	7797	617	1994
Kitsune Network Attack Dataset	Multivariate, Sequential, Time-Series	Classification, Clustering, Causal-Discovery	27170754	115	2019
LastFM Asia Social Network	Multivariate	Classification	7624	7842	2020
LastFM Asia Social Network	Multivariate	Classification	7624	7842	2020
Low Resolution Spectrometer	Multivariate	Classification	531	102	1988
LSVT Voice Rehabilitation	Multivariate	Classification	126	309	2014
Madelon	Multivariate	Classification	4400	500	2008
Malware static and dynamic features VxHeaven and Virus Total	Multivariate	Classification	2955	1087	2019
MEx	Time-Series	Classification, Clustering	6262	710	2019
MicroMass	Multivariate	Classification	931	1300	2013
Multiple Features	Multivariate	Classification	2000	649	
Musk (Version 1)	Multivariate	Classification	476	168	1994
Musk (Version 2)	Multivariate	Classification	6598	168	1994
Myocardial infarction complications	Multivariate	Classification	1700	124	2020
NoisyOffice	Multivariate	Classification, Regression	216	216	2015
Nomao	Univariate	Classification	34465	120	2012
Northix	Multivariate, Univariate, Text	Classification	115	200	2012
Opinion Corpus for Lebanese Arabic Reviews (OCLAR)	Text	Classification	3916	3916	2019
OPPORTUNITY Activity Recognition	Multivariate, Time-Series	Classification	2551	242	2012
p53 Mutants	Multivariate	Classification	16772	5409	2010
Parkinson's Disease Classification	Multivariate	Classification	756	754	2018

PEMS-SF	Multivariate, Time-Series	Classification	440	138672	2011
Person Classification Gait Data	Multivariate	Classification	48	321	2020
Physical Unclonable Functions	Multivariate	Classification	6000000	129	2018
QSAR androgen receptor	Multivariate	Classification	1687	1024	2019
QSAR oral toxicity	Multivariate	Classification	8992	1024	2019
REALDISP Activity Recognition Dataset	Multivariate, Time-Series	Classification	1419	120	2014
REJAFADA	Multivariate	Classification	1996	6826	2020
Reuter_50_50	Multivariate, Text, Domain-Theory	Classification, Clustering	2500	10000	2011
REWEMA	Multivariate	Classification	6272	632	2020
SCADI	Multivariate	Classification, Clustering	70	206	2018
SECOM	Multivariate	Classification, Causal-Discovery	1567	591	2008
Semeion Handwritten Digit	Multivariate	Classification	1593	256	2008
sEMG for Basic Hand movements	Time-Series	Classification	3000	2500	2014
Simulated Falls and Daily Living Activities Data Set	Time-Series	Classification	3060	138	2018
Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (AAL)	Time-Series	Classification	5744	561	2016
Smartphone-Based Recognition of Human Activities and Postural Transitions	Multivariate, Time-Series	Classification	10929	561	2015
Swarm Behaviour	Multivariate	Classification	24017	2400	2020
TTC-3600: Benchmark dataset for Turkish text categorization	Text	Classification, Clustering	3600	4814	2017
TUANDROMD (Tezpur University Android Malware Dataset)	Multivariate	Classification	4465	241	2021
Twin gas sensor arrays	Multivariate, Time-Series, Domain-Theory	Classification, Regression	640	480000	2016
UJIIndoorLoc	Multivariate	Classification, Regression	21048	529	2014

Ultrasonic flowmeter diagnostics	Multivariate	Classification	540	173	2018
Urban Land Cover	Multivariate	Classification	168	148	2014
URL Reputation	Multivariate, Time-Series	Classification	2396130	3231961	2009
Victorian Era Authorship Attribution	Text	Classification	93600	1000	2018
Weight Lifting Exercises monitored with Inertial Measurement Units	Multivariate	Classification	39242	152	2013
YouTube Multiview Video Games Dataset	Multivariate, Text	Classification, Clustering	120000	1000000	2013

APPENDIX B

ChatPDF responses

Outlined below are responses from chatbot – www.chatpdf.com .

- Highlighted in **RED** are the absolutely incorrect inferences drawn by it.
- In **GREEN**, one can find well-crafted responses demonstrating in-depth analysis.
- In **Yellow** are highlighted the repetitive responses.

Chat with QL-SSA_An_Adaptive_Q-Learning_based_Squirrel_Search_Algorithm_for_Feature_Selection.pdf

Namaste! Welcome to the QL-SSA PDF file, an adaptive Q-Learning based Squirrel Search Algorithm for Feature Selection. This paper discusses the importance of feature selection in reducing the dimension of large and complex datasets for efficient data analysis without compromising predictive model performance. The QL-SSA algorithm is proposed as a solution to this problem. Here are three example questions you may have about this PDF file:

How does the QL-SSA algorithm work to select relevant features in a dataset?

Can the QL-SSA algorithm be applied to different types of datasets, such as image or text data?

What are some potential applications of the QL-SSA algorithm in real-world scenarios?

[1] Provide a summary of the entire paper section-wise

The paper is organized into **five sections**:

Section I: Introduction The introduction provides an overview of the problem of feature subset selection and classification, and highlights the importance of this problem in various fields such as bioinformatics, image processing, and text mining. The authors also discuss the limitations of existing feature selection algorithms and introduce the proposed QL-SSA algorithm.

Section II: Theoretical Background This section provides a theoretical background on the **QPSO** and Q-learning algorithms, which are used to develop the QL-SSA algorithm. The authors explain how these algorithms work and how they can be combined to create a hybrid algorithm for feature subset selection and classification.

Section III: Proposed Method This section describes the proposed QL-SSA algorithm in detail. The authors explain how the algorithm works, including its initialization, fitness evaluation, particle update, and termination criteria. They also provide **pseudocode** for the algorithm.

Section IV: Experimental Results and Discussion This section presents experimental results for evaluating the performance of QL-SSA on 20 publicly available datasets from the UCI machine learning repository. The authors report that QL-SSA outperforms other state-of-the-art feature selection algorithms in terms of classification accuracy and number of selected features. They provide a detailed analysis of the results for each dataset, showing that QL-SSA consistently outperforms other algorithms across different datasets. They also provide statistical analysis to support these findings.

Section V: Conclusion The conclusion summarizes the main contributions of the paper and discusses potential applications of QL-SSA in various fields. The authors highlight some limitations of their study

and suggest directions for future research. Overall, this paper presents a novel approach for feature subset selection and classification using a hybridization of QPSO and Q-learning algorithms. The experimental results suggest that this approach can achieve higher classification accuracy while selecting fewer features than other state-of-the-art algorithms. The paper provides a detailed analysis of the results and statistical analysis to support these findings.

[2] Can the QL-SSA algorithm be applied to different types of datasets, such as image or text data?

The PDF file does not provide specific information on whether the QL-SSA algorithm can be applied to different types of datasets, such as image or text data. However, the paper mentions that 20 publicly available datasets from the UCI machine learning repository have been utilized to evaluate the performance of the proposed QL-SSA algorithm. These datasets have significant variation in terms of number of classes, number of features, and samples. Therefore, it is possible that the QL-SSA algorithm can be applied to different types of datasets beyond those used in this study.

[3] Describe the QL-SSA algorithm in a mathematical sense

The PDF file does not provide a complete mathematical description of the QL-SSA algorithm. However, it does mention that the proposed method QL-SSA integrates the concept of Q-learning with SSA, in which squirrels are relocated in the search space using QL techniques. The paper also mentions that the local search mechanism is modified using QL, enabling the search agent (squirrels) to learn from other search agents' experiences and make it an adaptive, intelligent, and robust method. In summary, the QL-SSA algorithm combines reinforcement learning technique (Q-learning) with nature-inspired computation method (Squirrel Search Algorithm) to enhance feature subset selection and classification problems. The algorithm uses a modified local search mechanism to enable adaptive and intelligent searching for optimal solutions.

[4] Describe the Squirrel Search Algorithm

The Squirrel Search Algorithm (SSA) is a nature-inspired optimization algorithm that mimics the dynamic foraging approach of squirrels during food search. The algorithm is based on the gliding process of squirrels, which involves jumping from one tree to another in search of food. SSA uses a population-based approach, where each individual squirrel represents a potential solution to the optimization problem. The SSA algorithm starts with an initial population of squirrels randomly distributed in the search space. Each squirrel then moves towards a new location based on its current position and the positions of other squirrels in the population. The movement of each squirrel is guided by **three main factors: attraction to food sources (i.e., good solutions), repulsion from other squirrels (i.e., avoiding local optima), and random exploration (i.e., promoting diversity)**. The SSA algorithm also includes a mechanism for updating the best solution found so far, called the global best position. This position is **updated whenever a squirrel finds a better solution than the current global best**. The algorithm continues until a stopping criterion is met, such as reaching a maximum number of iterations or finding an acceptable solution. Overall, SSA is designed to balance exploration and exploitation in searching for optimal solutions by combining multiple factors that influence squirrel behavior during food search.

[5] State all the steps in Squirrel Search Algorithm

Sure, here are the steps involved in the Squirrel Search Algorithm (SSA):

1. Initialize the population of squirrels randomly in the search space.
2. Evaluate the fitness of each squirrel based on its position in the search space.
3. Update the global best position based on the best solution found so far.
4. For each squirrel, calculate its new position based on three factors: attraction to food sources (i.e., good solutions), repulsion from other squirrels (i.e., avoiding local optima), and random exploration (i.e., promoting diversity).
5. Evaluate the fitness of each squirrel at its new position.
6. Update the global best position if a squirrel finds a better solution than the current global best.
7. Repeat steps 4-6 until a stopping criterion is met, such as reaching a maximum number of iterations or finding an acceptable solution. Overall, SSA is designed to balance exploration and exploitation in searching for optimal solutions by combining multiple factors that influence squirrel behavior during food search.

[6] Provide a mathematical explanation of the squirrel search algorithm

Sure, here is a mathematical explanation of the Squirrel Search Algorithm (SSA):

1. Initialization: Initialize a population of N squirrels randomly in the search space. Each squirrel i is represented by a D -dimensional vector $X_i = [x_{i1}, x_{i2}, \dots, x_{iD}]$, where D is the dimensionality of the search space.
2. Fitness evaluation: Evaluate the fitness of each squirrel i based on its position X_i in the search space using an objective function $f(X_i)$.
3. Global best position update: Update the global best position G_{best} based on the best solution found so far among all squirrels in the population.
4. Movement calculation: For each squirrel i , calculate its new position V_{newi} based on three factors: - Attraction to food sources: Move towards the global best position G_{best} with a certain probability P_{attr} . - Repulsion from other squirrels: Move away from other squirrels j with a certain probability P_{rep} . - Random exploration: Move randomly with a certain probability P_{exp} . The new velocity V_{newi} for each squirrel i can be calculated as follows: $V_{newi} = w * V_i + c1 * rand() * (P_{attr} * (G_{best} - X_i) + P_{rep} * \sum (X_j - X_i) + P_{exp} * rand())$ where w is the inertia weight, $c1$ is a constant scaling factor, $rand()$ generates a random number between 0 and 1, and $\sum (X_j - X_i)$ represents the sum of repulsion vectors from all other squirrels j in the population.
5. Position update: Update each squirrel's position X_i based on its new velocity V_{newi} .
6. Fitness evaluation: Evaluate the fitness of each squirrel i at its new position X_i using an objective function $f(X_i)$.
7. Global best position update: Update the global best position G_{best} if a squirrel finds a better solution than the current global best.

8. Repeat steps 4-7 until a stopping criterion is met, such as reaching a maximum number of iterations or finding an acceptable solution.

Overall, SSA is designed to balance exploration and exploitation in searching for optimal solutions by combining multiple factors that influence squirrel behavior during food search. The algorithm uses a population-based approach and a modified velocity update equation to guide the movement of each squirrel towards better solutions.

[7] Critically evaluate the outcome of the paper

The proposed QL-SSA algorithm in the paper is a novel approach that combines the strengths of reinforcement learning and nature-inspired computation methods for solving optimal feature subset selection and classification problems. The algorithm modifies the local search mechanism of SSA using Q-learning techniques, which enables the search agent (squirrels) to learn from other search agents' experiences and make it an adaptive, intelligent, and robust method. The experimental results

presented in the paper show that the proposed QL-SSA algorithm outperforms other state-of-the-art algorithms in terms of accuracy and efficiency for feature subset selection and classification problems.

The authors compare their algorithm with six other algorithms on six benchmark datasets, and report that QL-SSA achieves higher classification accuracy than all other algorithms on five out of six datasets. The authors also report that QL-SSA has a faster convergence rate than all other algorithms on four out of six datasets. However, there are some limitations to this study that should be considered. Firstly, the authors only compare their algorithm with a limited number of state-of-the-art algorithms. There are many other optimization algorithms that could be compared with QL-SSA, such as genetic algorithms, particle swarm optimization, or ant colony optimization. Secondly, the authors do not provide a detailed analysis of the performance of their algorithm on different types of datasets or different sizes of datasets. It is unclear whether QL-SSA would perform equally well on larger or more complex datasets. Another limitation is that the paper does not provide a detailed explanation or analysis of how Q-learning is integrated into SSA. While the authors briefly describe how Q-learning enhances the relocation mechanism of squirrels in searching for an optimal food source, they do not provide a mathematical explanation or analysis of how this modification affects the performance of SSA. Finally, it should be noted that while QL-SSA achieves higher classification accuracy than all other algorithms on five out of six datasets, the difference in accuracy between QL-SSA and the other algorithms is relatively small. The authors report that the difference in accuracy between QL-SSA and the second-best algorithm is less than 1% on four out of six datasets. This suggests that while QL-SSA may be a promising approach, it may not be significantly better than other state-of-the-art algorithms in practice. In conclusion, while the proposed QL-SSA algorithm in this paper is a novel approach that combines reinforcement learning and nature-inspired computation methods for solving optimal feature subset selection and classification problems, there are some limitations to

[8] How does the QL-SSA algorithm work to select relevant features in a dataset?

The QL-SSA algorithm works to select relevant features in a dataset by modifying the local search mechanism of the Squirrel Search Algorithm (SSA) using Q-learning techniques. The algorithm starts with an initial population of squirrels randomly distributed in the search space, where each squirrel represents a potential solution to the optimization problem. The movement of each squirrel is guided by three main factors: attraction to food sources (i.e., good solutions), repulsion from other squirrels (i.e., avoiding local optima), and random exploration (i.e., promoting diversity). The QL-SSA algorithm

modifies the local search mechanism of SSA using Q-learning techniques, which enables the search agent (squirrels) to learn from other search agents' experiences and make it an adaptive, intelligent, and robust method. Specifically, Q-learning is applied to enhance the relocation mechanism of squirrels in searching for an optimal food source. **It also enhances the diversification of the solution due to the insensitive characteristics of QL towards the exploration point.** During each iteration of the algorithm, each squirrel moves towards a new location based on its current position and the positions of other squirrels in the population. The movement of each squirrel is guided by three main factors: attraction to food sources, repulsion from other squirrels, and random exploration. The algorithm also includes a mechanism for updating the global best position based on the best solution found so far. The fitness function used in QL-SSA is based on both feature subset selection and classification accuracy. The algorithm evaluates each potential solution based on its ability to select relevant features that contribute to high classification accuracy. This means that during optimization, QL-SSA not only selects relevant features but also ensures that these features contribute positively towards classification accuracy. Overall, by combining reinforcement learning and nature-inspired computation methods, QL-SSA provides an efficient and effective approach for selecting relevant features in a dataset. The algorithm is able to learn from other search agents' experiences and adapt to changing environments, making it a robust and intelligent method for feature subset selection and classification problems.

[9] what is the main drawback of the QL-SSA algorithm

The paper does not mention any significant drawbacks of the QL-SSA algorithm. However, it highlights the main issue with **existing optimizers, which is their frequent premature convergence, resulting in an inadequate contribution to data mining.** The QL-SSA algorithm is proposed as a solution to this problem by combining Reinforcement Learning and the Squirrel Search Algorithm, making it more adaptive and robust for feature selection by maintaining a good balance between exploration and exploitation steps. The results show that it outperforms the baseline optimizer in most cases.

[10] DO the authors really live up to their promises

The paper provides experimental results that demonstrate the effectiveness of the proposed QL-SSA algorithm in solving feature selection and classification problems. The algorithm is tested on 20 real-world benchmark datasets using two classifiers, and the results show that it outperforms the baseline optimizer in most cases. **Therefore, based on the experimental results presented in the paper, it can be concluded that the authors have lived up to their promises** of proposing an adaptive Q-Learning based Squirrel Search Algorithm for Feature Selection that **outperforms existing optimizers** in most cases. However, **further research and testing may be required to validate its effectiveness in other scenarios beyond those presented in this paper.**

[11] Provide an analysis of the Results. Support your arguments with numbers

The paper reports that the proposed QL-SSA algorithm outperforms other state-of-the-art feature selection algorithms in terms of classification accuracy and number of selected features. Specifically, the paper reports that QL-SSA achieved an average classification accuracy of 94.05%, which is higher than other algorithms such as **PSO (93.15%), GA (92.95%), ACO (93.45%), and ABC (93.75%).** The difference in classification accuracy between QL-SSA and these algorithms is statistically significant, as confirmed by the Wilcoxon signed-rank test. In addition, the paper reports that QL-SSA selected fewer features than these algorithms, indicating that it is more effective at identifying relevant features while reducing

computational complexity. Specifically, **QL-SSA selected an average of 7.5 features per dataset, while other algorithms selected an average of 10 to 15 features per dataset.** The paper also provides a detailed analysis of the results for each dataset, showing that QL-SSA consistently outperforms other algorithms in terms of classification accuracy and number of selected features across different datasets. Overall, these results suggest that QL-SSA is a promising algorithm for feature subset selection and classification problems, as it achieves higher classification accuracy while selecting fewer features than other state-of-the-art algorithms. However, it is important to note that the evaluation was conducted on a limited set of datasets from a specific repository, and **further research is needed to investigate its performance on more diverse datasets and real-world applications.**

[12] Write a summary of the PDF file

The PDF file is a research paper that proposes a new algorithm called QL-SSA for feature subset selection and classification problems. The algorithm is based on a hybridization of the **Quantum-behaved Particle Swarm Optimization (QPSO)** and Q-learning algorithms. The paper provides a theoretical background on the **QPSO** and Q-learning algorithms, which are used to develop the QL-SSA algorithm. The authors explain how these algorithms work and how they can be combined to create a hybrid algorithm for feature subset selection and classification. The proposed QL-SSA algorithm is evaluated on 20 publicly available datasets from the UCI machine learning repository. The results show that QL-SSA outperforms other state-of-the-art feature selection algorithms in terms of classification accuracy and number of selected features. Specifically, QL-SSA achieved an average classification accuracy of 94.05%, **which is higher than other algorithms such as PSO, GA, ACO, and ABC.** In addition, QL-SSA selected fewer features than these algorithms, indicating that it is more effective at identifying relevant features while reducing computational complexity. The paper provides a detailed analysis of the results for each dataset, showing that QL-SSA consistently outperforms other algorithms across different datasets. They also provide statistical analysis to support these findings. The authors discuss the potential applications of the proposed algorithm in various fields such as bioinformatics, image processing, and text mining. They also highlight some limitations of the study, such as the limited set of datasets used for evaluation and the need for further research to investigate its performance on more diverse datasets and real-world applications. Overall, this paper presents a promising approach for feature subset selection and classification problems using a hybridization of **QPSO** and Q-learning algorithms. The experimental results suggest that this approach can achieve higher classification accuracy while selecting fewer features than other state-of-the-art algorithms. **The paper provides a detailed analysis of the results and statistical analysis to support these findings.** The authors also discuss potential applications and limitations of their study, as well as directions for future research.

[13] Write a precis of the research paper

The research paper proposes a new algorithm called QL-SSA for feature subset selection and classification problems. The algorithm is based on a hybridization of the **Quantum-behaved Particle Swarm Optimization (QPSO)** and Q-learning algorithms. The proposed algorithm is evaluated on 20 publicly available datasets from the UCI machine learning repository. The results show that QL-SSA outperforms other state-of-the-art feature selection algorithms in terms of classification accuracy and number of selected features. Specifically, QL-SSA achieved an average classification accuracy of 94.05%, **which is higher than other algorithms such as PSO, GA, ACO, and ABC.** In addition, QL-SSA selected fewer features than these algorithms, indicating that it is more effective at identifying relevant features while

reducing computational complexity. The authors discuss the potential applications of the proposed algorithm in various fields such as bioinformatics, image processing, and text mining. They also highlight some limitations of the study, such as the limited set of datasets used for evaluation and the need for further research to investigate its performance on more diverse datasets and real-world applications. Overall, this paper presents a promising approach for feature subset selection and classification problems using a hybridization of **QPSO** and Q-learning algorithms.