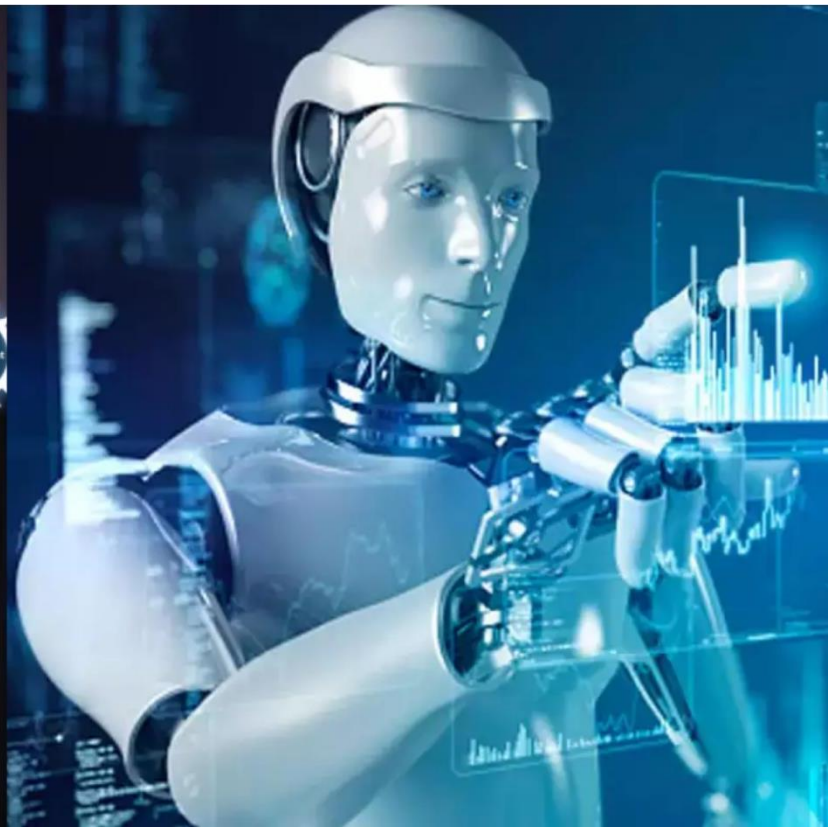
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 4, April 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504107|

An Automated Feature Optimization System using Genetic Algorithm

Dr.Atul Kumar Ramotra, Tejasri Velishala, Varun Makthala, Madanu Maria Williams

Associate Professor, Department of CSE (AIML), ACE Engineering College, Hyderabad, India

III B.Tech Students, Department of CSE (AIML), ACE Engineering College, Hyderabad, India

ABSTRACT: Feature selection is an essential step in improving the accuracy and efficiency of machine learning models, especially when handling high-dimensional datasets. Manual and traditional feature selection methods are time-consuming and often inefficient. This project presents an automated feature optimization system using Genetic Algorithms integrated with Machine Learning techniques. Genetic Algorithms intelligently search for optimal feature subsets through evolutionary operations such as selection, crossover, and mutation.

I. INTRODUCTION

In recent years, the amount of data being generated has increased rapidly, leading to the growth of advanced techniques in Machine Learning and Data Science. With such large datasets, it has become very important to extract meaningful information efficiently. One of the key steps in building a good machine learning model is selecting the right features that actually contribute to making accurate predictions. Feature selection helps improve not only the model's performance but also its efficiency and ease of understanding.

In real-world datasets, there are often many features, but not all of them are useful. Some may be irrelevant, some may repeat the same information, and others may just add noise. Using all these features can slow down the model and sometimes even reduce its accuracy. It can also lead to overfitting, where the model works well on training data but fails to perform on new, unseen data. That's why choosing the most relevant features is an important step in building reliable models.

There are several traditional methods used for feature selection, such as filter methods, wrapper methods, and embedded methods. Filter methods use statistical techniques to rank features, while wrapper methods test different combinations of features using machine learning algorithms. Embedded methods perform feature selection during the model training process itself. Although these approaches are useful, they don't always find the best set of features, especially when dealing with datasets that have a very large number of features.

II. EXISTING SYSTEM

In many existing systems, feature selection is usually done using traditional approaches like filter, wrapper, and embedded methods. Filter methods rely on statistical techniques such as correlation and chi-square to choose important features. While these methods are simple and fast, they do not take into account how well a machine learning model actually performs with those selected features.

Another challenge with these methods is scalability. As the size of the dataset and the number of features increase, the computational cost also rises significantly, making the process time-consuming and less efficient. In addition, traditional techniques often get stuck in local optimal solutions, meaning they fail to explore all possible combinations of features. Because of this, the selected feature subset may not always be the best one.

Moreover, most of these existing systems are not fully automated. They often require human effort, domain knowledge, and manual tuning to achieve good results. The presence of noisy or irrelevant data can further reduce the model's accuracy.These limitations clearly show the need for a better approach. To overcome these issues, a more advanced and automated feature optimization system is required that can efficiently handle large datasets and identify the most relevant features.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504107|

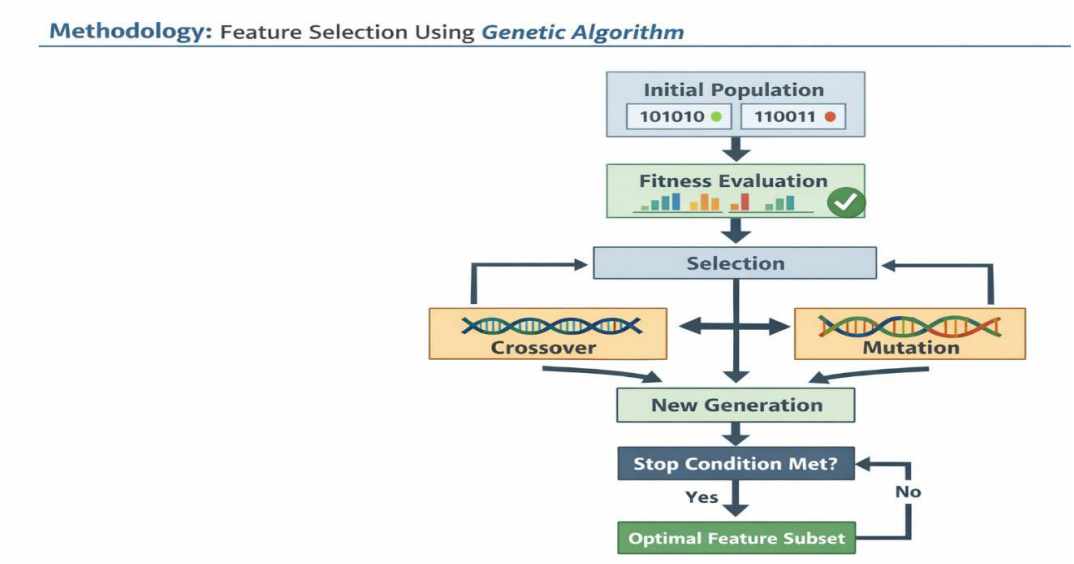
III. PROPOSED SYSTEM

The proposed system presents an automated method for feature selection using a Genetic Algorithm. In this approach, each possible set of features is represented as a chromosome using binary encoding, where each bit indicates whether a particular feature is selected or not. The process starts by generating an initial population of random feature combinations.

Each of these feature sets is then evaluated using a fitness function, which is usually based on the accuracy of a machine learning model. The better-performing feature subsets are selected and used to create the next generation. Genetic operations such as crossover and mutation are applied during this stage. Crossover helps combine useful features from different parent solutions, while mutation introduces small random changes to maintain diversity and avoid premature convergence.

Additionally, the system is scalable and works well with large datasets. It can also be easily integrated with different machine learning algorithms, making it flexible for various applications. Overall, this approach provides a reliable and efficient solution for feature optimization, leading to better performance and more consistent results.

IV. METHODOLOGY



i. Initial Population
The process starts by creating random combinations of features. Each combination represents a possible solution.
ii. Fitness Evaluation
Each feature set is tested using a machine learning model to check how well it performs.
iii. Selection
The best-performing feature sets are chosen so they can be used to create better solutions.
iv. Crossover
Selected feature sets are combined to mix their features and produce new combinations.
v. Mutation
Small random changes are made to some feature sets to introduce variety and avoid repeating the same solutions.
vi. New Generation
A new set of feature combinations is formed using selection, crossover, and mutation.
vii. Stop Condition
The process checks whether the solution is good enough or if the maximum number of iterations is reached.
viii. Optimal Future Subset
Finally, the best set of features is selected, which gives good accuracy with fewer features.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

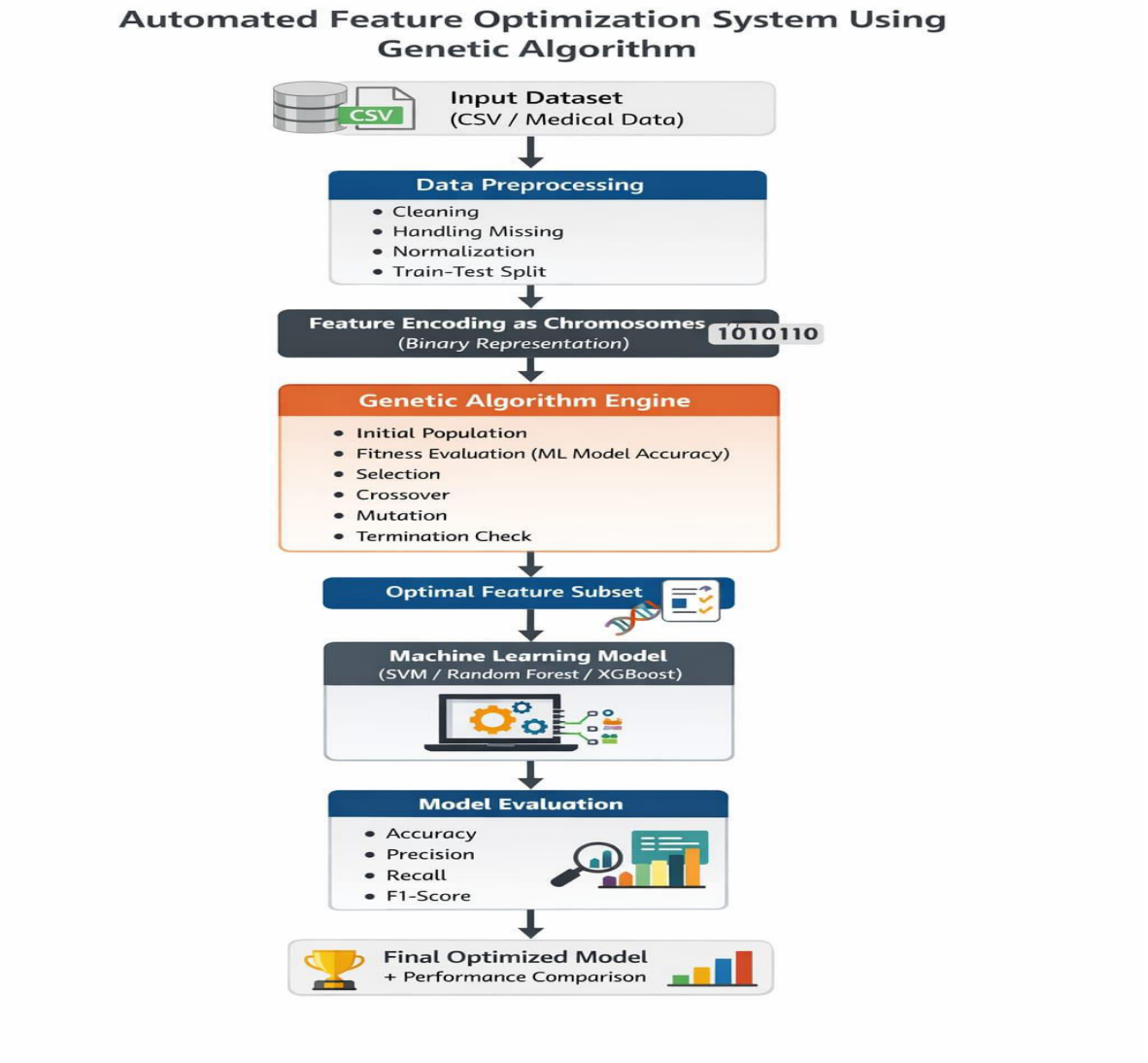
|DOI: 10.15680/IJIRSET.2026.1504107|

V. SYSTEM ARCHITECTURE

The system architecture of the Automated Feature Optimization System is designed to improve the performance of machine learning models by combining Genetic Algorithms with standard learning techniques. The process starts with an input dataset, such as a CSV file or medical data, which is first preprocessed. This includes cleaning the data, handling missing values, normalizing it, and splitting it into training and testing sets.

After preprocessing, the features are converted into a binary format, where each feature is represented as a bit indicating whether it is selected or not. These binary representations are then passed into the Genetic Algorithm, which creates an initial set of random feature combinations. The algorithm repeatedly evaluates these combinations based on model performance and improves them using operations like selection, crossover, and mutation.

Over several iterations, the system identifies the most effective set of features. This optimized feature subset is then used to train machine learning models such as SVM, Random Forest, or XGBoost. Finally, the model is evaluated using performance metrics like accuracy, precision, recall, and F1-score. The end result is a model that performs better while using fewer and more relevant features, making it both efficient and reliable.



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

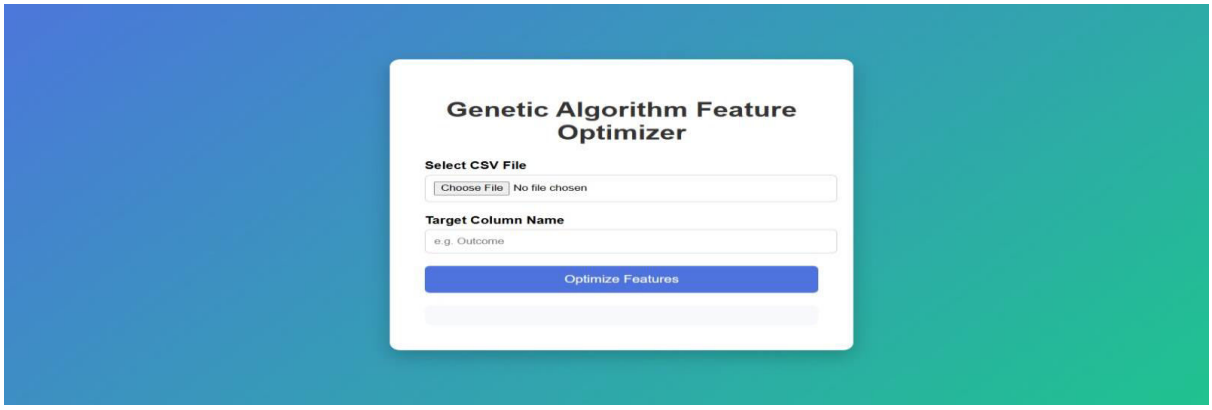
Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

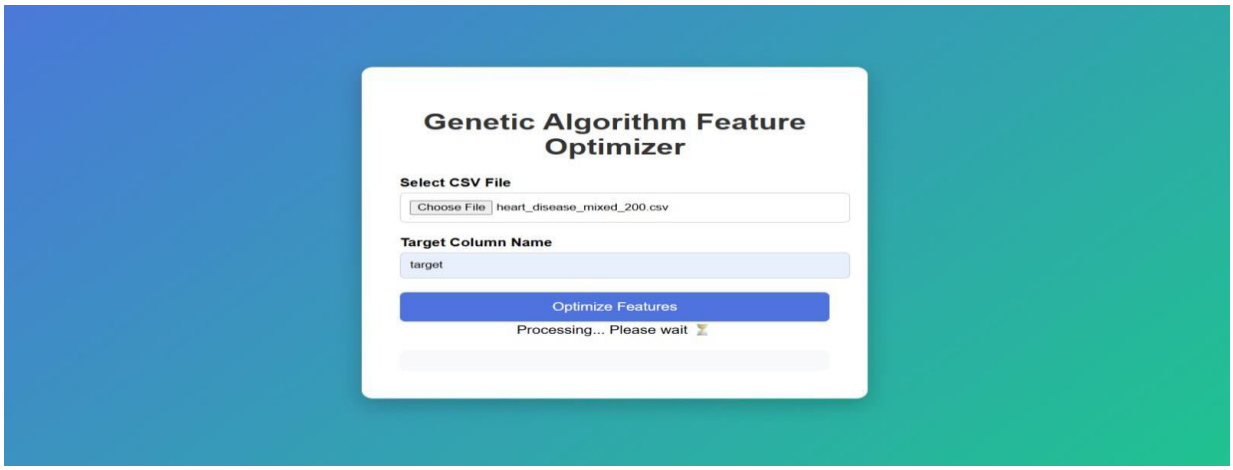
Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504107|

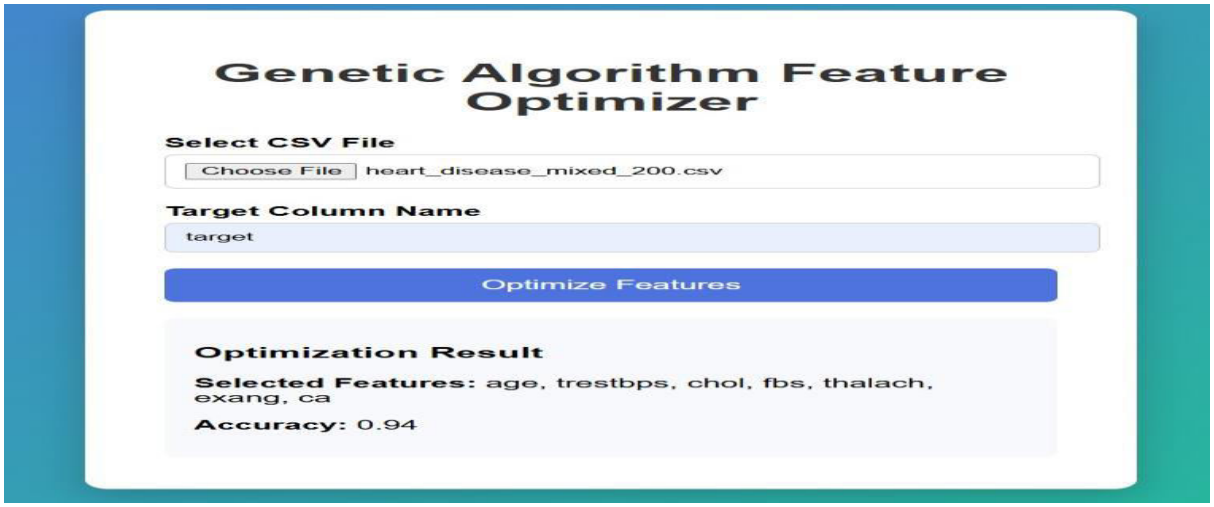
VI. RESULTS AND OUTPUT



Final Output



Sample DataSet



Result

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504107|

VII. CONCLUSION

In conclusion, this project presents an effective approach for improving machine learning model performance through automated feature optimization using a Genetic Algorithm. One of the main challenges in working with large datasets is identifying which features actually contribute to accurate predictions. By addressing this problem, the proposed system helps in reducing unnecessary complexity and improving overall efficiency.

The use of a Genetic Algorithm allows the system to explore different combinations of features in an intelligent way, rather than relying on traditional methods that may miss the best solution. Through processes like selection, crossover, and mutation, the system gradually improves the quality of feature subsets and avoids getting stuck in less optimal solutions. This makes it more reliable, especially when dealing with high-dimensional data.

Another important advantage of this approach is automation. Unlike conventional methods that require manual effort and domain knowledge, the proposed system reduces human intervention and makes the feature selection process more systematic. It also helps in minimizing overfitting by selecting only the most relevant features, which improves the model's ability to perform well on unseen data.

The system is flexible and can be used with different machine learning algorithms such as SVM, Random Forest, and XGBoost. It is also scalable, making it suitable for handling large and complex datasets. The results show that the optimized feature subsets lead to better accuracy and improved model performance compared to using all features.

Overall, this work demonstrates that combining Genetic Algorithms with machine learning is a powerful solution for feature optimization. It not only enhances performance but also reduces computational cost and model complexity. This makes the system useful for real-world applications where efficiency and accuracy are equally important.

VIII. FUTURE SCOPE

The proposed Automated Feature Optimization System using a Genetic Algorithm has delivered effective results in selecting important features and improving model performance. However, there is still a wide scope for future enhancements that can make the system more powerful, efficient, and applicable to a broader range of problems.

One possible direction for future work is the integration of other optimization techniques along with the Genetic Algorithm. Hybrid approaches that combine Genetic Algorithms with methods like Particle Swarm Optimization or Ant Colony Optimization can be explored to improve the search capability and find better feature subsets. These hybrid models may help in achieving faster convergence and more accurate results.

Another important area of improvement is the use of advanced machine learning and deep learning models. Currently, the system can be integrated with models like SVM, Random Forest, and XGBoost, but in the future, it can be extended to work with neural networks and deep learning architectures. This will make the system more suitable for handling complex data such as images, text, and time-series data.

The system can also be enhanced to support real-time data processing. In many real-world applications like healthcare monitoring, stock market prediction, and IoT systems, data is generated continuously. Adapting the system to work with streaming data can significantly increase its practical usefulness.

Scalability is another area where improvements can be made. As datasets continue to grow in size, it becomes important to handle large volumes of data efficiently. Techniques such as parallel processing, distributed computing, and cloud integration can be used to reduce computation time and improve performance.

In addition, the user interface of the system can be improved to make it more interactive and user-friendly. Providing visualization tools for feature importance, model performance, and optimization progress can help users better understand the results and make informed decisions.Future work can also focus on improving the fitness function by considering multiple evaluation metrics such as precision, recall, F1-score, and computational cost, instead of relying only on accuracy. This will ensure a more balanced and reliable model performance.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504107|

Overall, the future scope of this project is quite broad. With continuous improvements and integration of new technologies, the system can evolve into a highly efficient and intelligent tool for feature optimization. It has the potential to play an important role in solving complex real-world problems and supporting better decision-making in various fields.

REFERENCES

[1] Raghad Aldamani et al.“ Recent Advances in Genetic Algorithm-Based Feature Selection: A Comprehensive Survey”.
[2] Zhila Yaseen Taha et al. “Optimizing Feature Selection with Genetic Algorithms: A Review of Methods and Applications”.
[3] H. Alirezapour et al. **“**A Comprehensive Survey on Feature Selection with Grasshopper Optimization Algorithm”.
[4] Hussein M. Alshamlan, Gheith A. Abandah, Majed H. Anbar **“**Quantum and Evolutionary Feature Selection: Survey of Techniques and Applications”.
[5] Yvan Saeys, Inaki Inza “Evolutionary Feature Selection Techniques for High-Dimensional Data: A Survey”.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details