IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Insurance Fruad Data Feature Selection Using Hybrid Random Forest Feature Importance (Rffi) With Recursive Feature Elimination (Rfe) Model

R.Nisha

Research Scholar,
Department of Computer Science,
Hindusthan College of Arts and Science,
Coimbatore, Tamilnadu, India.

Dr.G.Dalin,
Associate Professor,
Department of Computer Science,
Hindusthan College of Arts And Science,
Coimbatore, Tamilnadu, India.

Abstract - Quality and relevance of features used in training are key factors in machine learning models' efficacy and dependability. Overfitting, higher computational complexity, and worse predictive performance can result from the redundant or unnecessary features that are frequently present in high-dimensional datasets. This work suggests a thorough architecture for feature selection that blends embedding, filtering, and wrapping methods to find the most important predictors while reducing noise. For the best subset selection, wrapper-based recursive feature elimination (RFE) is used after filter techniques like correlation analysis and chi-square tests are used to eliminate redundant features. To guarantee robust selection during model training, embedded strategies are employed, such as Lasso and tree-based feature significance algorithms. The suggested approach increases accuracy across a variety of machine learning techniques, decreases training time, and improves model interpretability. Comparing experimental outcomes to baseline models with all characteristics, and the F1-score show notable increases. This study highlights how crucial systematic feature selection is as a first step in creating scalable and effective predictive modelling pipelines.

Keywords: Feature Selection, Dimensionality Reduction, Machine Learning, Recursive Feature Elimination, Lasso Regression, Model Optimization, Predictive Modeling.

1.Introduction

The development of effective machine learning models depends on feature selection. Many variables or features are gathered from real-world datasets, but not all of them are equally pertinent to the prediction task. Adding superfluous or unnecessary features frequently results in overfitting, higher computational cost, and subpar model generalization. Consequently, the methodical process of choosing the most informative aspects aids in the creation of models that are not only more accurate but also easier to understand and use. By ensuring that the model only concentrates on the variables that significantly contribute to the job at hand, feature selection acts as a link between the collection of raw data and predictive modeling. In light of contemporary uses like fraud detection, healthcare analytics, and financial forecasting, feature selection has emerged as one of the key steps that determine the success of a machine learning pipeline.

1.2 Role of Dimensionality Reduction in Model Optimization

The "curse of dimensionality" is another term for high-dimensional information. presents a significant challenge for machine learning. Most algorithms lose their discriminative power when the number of features rises because the data becomes sparser. This problem is mitigated by dimensionality reduction via feature selectionby focusing only on a smaller subset of the most relevant attributes, thereby reducing noise and improving the signal-to-noise ratio. This not only accelerates model training and prediction but also lowers storage requirements and computational costs. Importantly, models trained on reduced feature sets tend to generalize better to unseen data, thereby improving real-world applicability. By eliminating irrelevant attributes, the model avoids fitting random fluctuations in the data and instead captures the underlying patterns that truly matter.

1.3 Categories of Feature Selection Methods

Filter methods, wrapper methods, and embedding methods are the three categories of feature selection techniques. Filter methods are model-independent strategies that use statistical measures like correlation coefficients, chi-square tests, or mutual information scores to rank features independently of the learning process. In contrast, wrapper approaches use iterative training and testing of a model to evaluate subsets of features, choosing the subset that produces the greatest performance metric. Despite their higher computational cost, wrapper approaches frequently yield feature sets that are ideal for the selected model. During model training, embedded approaches perform feature selection using regularization techniques such as Lasso (L1) and Ridge (L2), which penalize the presence of irrelevant variables. This integrated approach balances computational efficiency with predictive power and is widely used in modern machine learning applications.

1.4 Feature Selection for Interpretability and Trust

Contributing to interpretability is another advantage of feature selection that is frequently disregarded. In domains where transparency is crucial—such as healthcare, economics, and legal decision-making—being able to explain why a model produces a given forecast is just as important as the prediction itself. Reducing the feature set to the most important variables allows data scientists and domain experts to better understand the drivers behind model decisions. This improves trust and facilitates regulatory compliance in sensitive applications. For example, in medical diagnosis, knowing that a small number of biomarkers are primarily responsible for a classification decision provides actionable insights for doctors and patients alike. Thus, feature selection is not only a technical step but also a tool for enhancing the human interpretability of machine learning models.

1.5 Challenges and Research Opportunities

Despite its importance, feature selection remains a challenging problem, particularly in the age of deep learning and huge data. Extensive search for the ideal subset is nearly impossible in high-dimensional datasets with millions of attributes, such as those used in text analytics, genomics, and picture processing. Additionally, the presence of noisy or irrelevant features, missing values, and multicollinearity complicates the selection process. Hybrid strategies that combine filter, wrapper, and embedding techniques to maximize their advantages and minimize their disadvantages are the subject of recent research. Additionally, the emergence of explainable artificial intelligence (XAI) has rekindled interest in feature selection strategies that enhance model accuracy while offering explanations that are understandable to humans. Future work in this area is likely to involve automated machine learning (AutoML) systems that integrate feature selection as a built-in, adaptive process, A crucial stage in creating optimal machine learning models is feature selection. It supports interpretability by concentrating on the most significant predictors, increases generalization by decreasing overfitting, and boosts performance by eliminating superfluous variables. The need for efficient, effective, and interpretable feature selection techniques will only grow as machine learning develops and is used in mission-critical applications. This research attempts to investigate the theoretical underpinnings, practical techniques, and experimental results of feature selection, highlighting its significance in building robust, scalable, and trustworthy predictive models.

In fraud detection systems, feature selection is a crucial stage since irrelevant and duplicated attributes can lower prediction accuracy and increase model complexity. This paper proposes a novel hybrid approach, FS-Hybrid, which combines Random Forest Feature Importance (RFFI) with To find the best representative subset of characteristics in insurance claims data, use Recursive Feature Elimination (RFE). To create an initial relevance ranking of all 45 available traits, the suggested framework initially uses RFFI. The RFE procedure is then guided by this ranking, iteratively removing weak features while keeping an eye on

performance indicators like accuracy, precision, recall, and F1-score. Trials carried out on a publicly accessible insurance fraud detection dataset demonstrate that FS-Hybrid consistently selects the top 10 most influential features, achieving superior classification performance compared to individual feature selection methods. The results confirm that the hybrid methodology not only improves accuracy and robustness but also reduces overfitting and computational overhead, thereby providing a more efficient foundation for fraud prediction models.

2. Related Work

Ewen Hokijuliandy, HerlinaNapitupulu (2023) proposed chi-square feature selection. Feature selection improves the model's performance by employing fewer features and more effective calculations. The Chi-Square approach is a popular feature selection technique. Using statistical theory, the Chi-Square feature selection approach determines if a term is irrespective of its class. To choose which phrase or words will be utilized as features, the Chi-Square values for each term are arranged in descending order.

Taha, A, Hadi (2022) proposed algorithm for greedy feature selection (GFSA). Finding the set of features with the least amount of correlation between them is the foundation of the Greedy Feature Selection Algorithm (GFSA). This algorithm looks for a wide range of qualities that are not overly dependent on one another. An optimization problem is used to formulate the feature selection problem. GFSA handles both categorical and numerical features. By following a methodical approach, the GFSA method determines the required dimension that minimizes the numerous relationships among them. Selecting the least dependent pair of features is the first stage. Until the required number of features is achieved, it then iteratively adds one feature to the solution. The GFSA algorithm assumes that one feature in the solution set has the smallest pairwise relationship with the incoming feature. After that, it calculates each candidate set's multiple association and chooses the set with the lowest multiple association.

He, X., Cai et.al (2005) proposed The Laplacian Score (LS) is used to select features. One filter-based feature selection technique is the Laplacian Score (LS). The selection of LS features is limited to continuous data. The relevance of each characteristic is quantified using the Eigensystem of the Laplacian matrix. The Laplacian matrix is computed using the elements' similarities. The characteristics are arranged in order of their Laplacian scores, which represent each feature's capacity to preserve locality. The LS score is dependent on an object's geometric structure. Initially, a closest neighbor graph G is constructed by the LS algorithm. A feature is represented by each node. One node is one of the other node's k nearest neighbors, and two nodes are IFF. Then, using the k nearest neighbor graph as a gauge of its power of locality preservation, it calculates a Laplacian score for every feature. The LS searches for characteristics that adhere to this graph topology.

Zhao, Z.; Liu, H et.al (2007) proposed Spec, the spectral graph theory serves as the foundation for a technique for choosing spectral features. The Spec method creates a similarity graph, G, to represent the dataset using the built-in features. If a feature aligns with the object-to-object similarity graph's structure, it

is deemed significant. Furthermore, a similarity score is computed for every characteristic. This score ought to align with the structure of the graph. To assess the coherence between a feature and the nontrivial eigenvectors of the Laplacian matrix, the Spec method assigns a consistency score to each feature.

Features for mixed data are selected using the Unsupervised Spectral Feature Selection Method (USFSM) in the absence of supervision. Using a kernel function is its foundationto construct a similarity matrix between objects. Distance functions are used for both nominal and numeric characteristics in the kernel function. When the components of the Laplacian matrix are separated, the spectrum of the data contains the data's structural information. This technique calculates the shift in the Normalized Laplacian matrix's spectrum distribution after a feature is removed in order to assess the feature's relevance. The spectral gap score is iteratively calculated for each feature using a leave-one-out procedure. The spectral gap score is then used to rank the features.

3. Proposed Methodology

High-dimensional datasets, such as insurance claims with 45 enriched attributes, often contain redundant or weakly informative variables. Overfitting, noise introduction, and longer training times can all result from incorporating all features at once, which can impair model performance. Feature selection tackles these issues by identifying the most important characteristics that support fraud detection.

In this paper, we focus on systematically narrowing the enriched feature space to retain the most relevant variables. Specifically, two complementary approaches are used:

1. **Random Forest Feature Importance (RFFI)** – ranks characteristics according to how well they reduce classification impurity.
2. **Recursive Feature Elimination (RFE)** – eliminates the least significant traits iteratively in order to arrive at an ideal subset.

This hybrid strategy ensures both global importance ranking and fine-grained elimination, resulting in a condensed, superior feature collection for the prediction model.

3.1 Random Forest Feature Importance (RFFI)

An ensemble learning technique called Random Forest determines the significance of a feature by calculating the impurity reduction at each split. The importance score of feature f is:

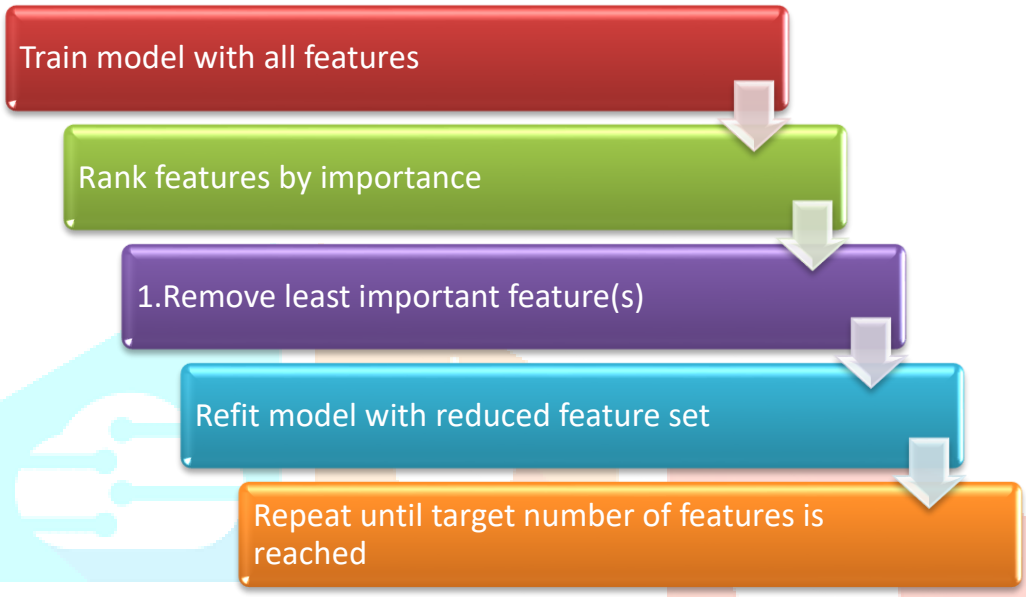
$$\text{Importance}(f) = \frac{1}{T}\sum_{t=1}^{T}\sum_{n\in \text{Nodes}(t)} \Delta I(f,n)$$

Where:

- T: total number of trees,
- $\Delta I(f, n)$: impurity reduction (e.g., Gini decrease) at node n using feature f.

2. Recursive Feature Elimination (RFE)

RFE uses a base estimator (e.g., logistic regression or SVM) to recursively prune less important features.



Recursive Feature Elimination (RFE) eliminates weak features by building models iteratively. The model is trained using the feature set at iteration $k.F^{(k)}$. Each feature f is assigned a ranking score:

$$\text{Rank}(f) = \arg \min_{f \epsilon F^{(k)}} \text{Score}\,(\text{Model} \setminus f)$$

Where $Score(\cdot)$may represent accuracy, F1-score, or loss function value. The feature with the least impact when removed is discarded. By following the RFFI ranking, RFE prioritizes low-ranked features for elimination, ensuring a structured reduction.

3. Hybrid RFFI + RFE Strategy

The first step appliesRandom Forest Feature Importance (RFFI)to the dataset containing 45 features. Random Forest evaluates each attribute based on its contribution to lowering impurity across decision trees (e.g., Gini index). This procedure produces animportance scorefor every feature, allowing them to be sorted in descending order of relevance. The result a ranked list that highlights which variables have the strongest initial influence in identifying fraudulent claims. Once the features are ranked,is Recursive Feature Elimination (RFE)is employed to refine the selection. RFE uses a machine learning estimator to iteratively

train models while removing the least important attributes at each step. Importantly, this elimination isguided by the RFFI ranking**,** ensuring that low-rankedfeatures are prioritized for removal first. This combination prevents random elimination and makes the process more systematic and efficient. The elimination process continues until the model reaches an ideal collection of characteristics, usually the top ten in this research. The stopping criterion is determined by monitoring validation accuracy and F1-score during the iterative process. When removing additional features no longer improves or starts reducing these performance metrics, the algorithm stops. The final selected subset represents the most informative features that balance predictive accuracy, computational efficiency, and reduced overfitting risk.

The process stops when model performance stabilizes. Let $\mathrm{Acc}(F)$ be accuracy using feature set F , and $\mathrm{F1}(F)$ the F-measure. The stopping condition is:

$$F^* = \arg\max_{F \subseteq \{1,\ldots,45\}} \left(\beta \cdot \mathrm{Acc}(F) + (1-\beta) \cdot \mathrm{F1}(F)\right)$$

Where β balances accuracy and F1-score (often set to 0.5). The final subset F^* (e.g., top 10 features) maximizes validation performance while reducing dimensionality.

Proposed Hybrid Feature Selection using RFFI + RFE Algorithm

Input:

- D: Insurance dataset with 45 features (39 original + 6 extracted).
- K: Target number of final features (e.g., 10).
- M: Base machine learning model for evaluation (e.g., Logistic Regression, SVM).

Output:

- F^*: Optimal reduced feature subset.

1. Compute feature importance scores using Random Forest on D
2. Sort all features in descending order of importance to form ranking R
3. Initialize F ← R (all features).
4. While |F| > K do
5. Apply Recursive Feature Elimination (RFE) using model M.
6. Identify least important feature(s) from F guided by ranking R.
7. Remove selected feature(s) from F
8. Compute performance measures (Accuracy, Recall, F1).
9. Identify F^* as the feature subset that achieves best validation results.

10. Return F^*

4. Experimental result

4.1 Accuracy

Accuracy is the degree to which a measurement closely reflects its true value. The accuracy formula is:

$$\boldsymbol{Accuracy} = \frac{(\boldsymbol{truevalue} - \boldsymbol{measuredvalue})}{\boldsymbol{truevalue}} * \mathbf{100}$$

Dataset	Chi-square	Greedy Feature Selection Algorithm (GFSA)	Proposed Hybrid RFFI + RFE Algorithm
100	60	80	90
200	73	78	96
300	77	70	87
400	85	76	98
500	89	72	99

Table 1.Comparison Table of Accuracy

The comparison Table 1 of accuracy explains the disparities between the Greedy Feature Selection Algorithm (GFSA), the proposed hybrid RFFI + RFE algorithm, and the current Chi-square. It is clear that the suggested approach yields superior outcomes when compared to the current approaches and the Proposed HYBRID RFFI + RFE ALGORITHM. The Filter Method (CHI-SQUARE) values range from 60 to 89, the Greedy Feature Selection Algorithm (GFSA) values range from 70 to 80, whereas the Proposed HYBRID RFFI + RFE ALGORITHM values start from 87 and go up to 99. This clearly demonstrates that the suggested hybrid RFFI + RFE algorithm works noticeably better than conventional feature selection methods, offering higher accuracy and more reliable results across all dataset sizes.

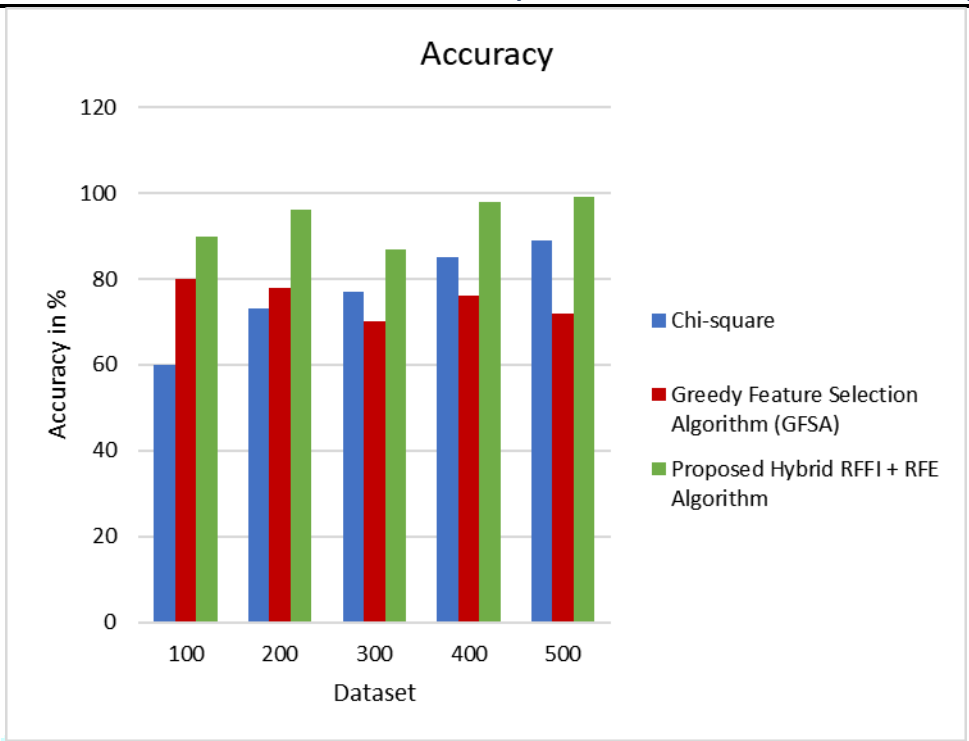


Figure 1.Comparison Chart of Accuracy

The Figure 1 displays the accuracy comparison chart, showing how well the Greedy Feature Selection Algorithm (GFSA), the proposed hybrid RFFI + RFE algorithm, and the current CHI-SQUARE algorithm perform. The accuracy percentage is displayed on the Y-axis, and the dataset size is displayed on the X-axis. RFE displays a range of 70% to 80%, but CHI-SQUARE accuracy values range from 60% to 89%. In contrast, the Proposed HYBRID RFFI + RFE ALGORITHM method achieves superior results, with accuracy values ranging from 87% to 99% across all dataset sizes. These outcomes make it abundantly evident that the suggested hybrid RFFI + RFE algorithm technique consistently outperforms both CHI-SQUARE and RFE techniques, offering a substantial improvement in predictive performance and demonstrating its effectiveness for feature selection in machine learning.

4.2 Precision

A model's precision indicates how effectively it can forecast a value given an input.

$$Precision = \frac{truepositive}{(truepositive + falsepositive)}$$

Dataset	Chi-square	Greedy Feature Selection Algorithm (GFSA)	Proposed Hybrid RFFI + RFE Algorithm
100	87.12	82.37	96.67
200	80.69	86.82	95.26
300	77.62	84.54	97.21
400	73.55	80.63	94.58
500	75.94	78.72	89.87

Table 2.Comparison Table of Precision

The comparison Table 2 of Precision illustrates the disparities between the Greedy Feature Selection Algorithm (GFSA), the existing CHI-SQUARE, and the proposed HYBRID RFFI + RFE ALGORITHM approach. When comparing the existing algorithms with the Proposed HYBRID RFFI + RFE ALGORITHM, It has been noted that the suggested strategy regularly yields superior outcomes. RFE values range from 73.55 to 87.12, whereas CHI-SQUARE values range from 78.72 to 86.82, whereas the Proposed HYBRID RFFI + RFE ALGORITHM values range from 89.87 to 97.21. This clearly shows that the proposed method achieves superior precision across all dataset sizes, significantly outperforming the traditional Filter and Wrapper approaches.

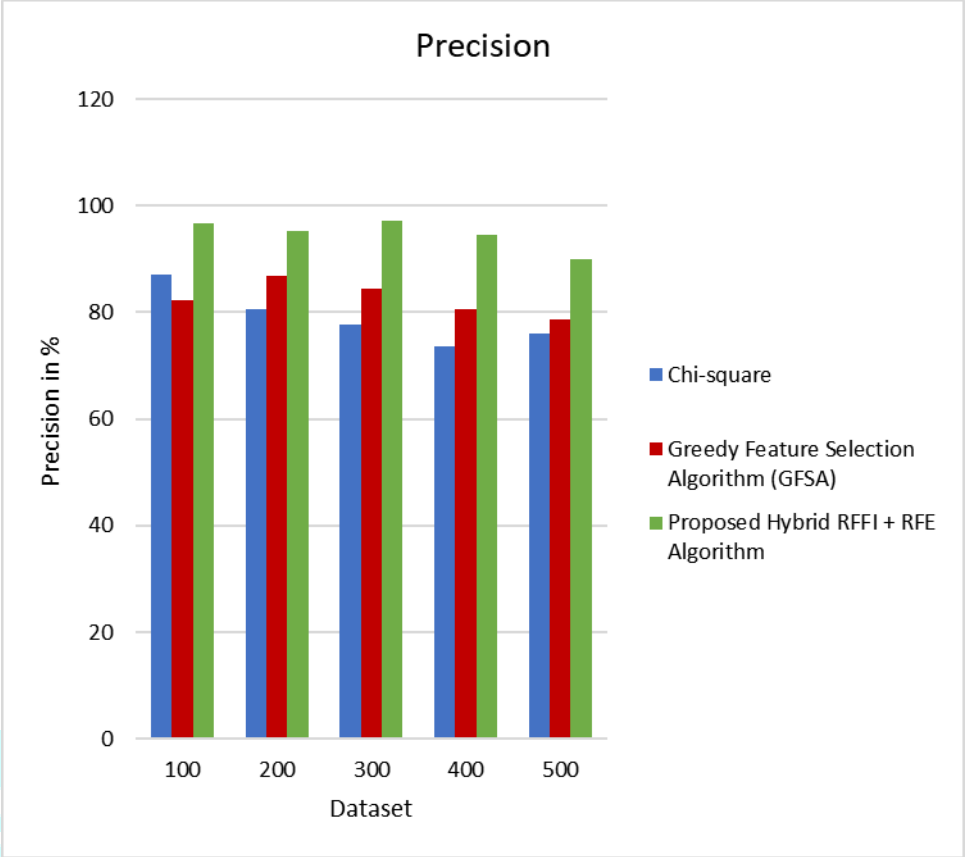


Figure 2.Comparison Chart of Precision

Figure 2 shows the Greedy Feature Selection Algorithm (GFSA), Chi-square, and the proposed comparison chart for the hybrid RFFI + RFE algorithm. The precision ratio is shown on the Y-axis, and the dataset size is shown on the X-axis. The proposed hybrid RFFI + RFE algorithm consistently outperforms the existing feature selection techniques. The scope of the CHI-SQUARE values is 73.55 to 87.12, while the RFE values range from 78.72 to 86.82. In contrast, the Proposed HYBRID RFFI + RFE ALGORITHM achieves considerably more precision, with numbers between 89.87 and 97.21. This clearly demonstrates that the proposed approach provides superior predictive performance and robust feature selection compared to conventional methods.

4.3 Recall

The capacity of a model to accurately identify good examples from the test set is measured by recall:

$$Recall = \frac{TruePositives}{(TruePositives + FalseNegatives)}$$

Dataset	Chi-square	Greedy Feature Selection Algorithm (GFSA)	Proposed Hybrid RFFI + RFE Algorithm
100	82	62	83
200	78	72	93
300	85	66	95
400	82	76	91
500	87	72	97

Table 3.Comparison Table of Recall

The analogy that the Greedy Feature Selection Algorithm (GFSA) and the proposed hybrid RFFI + RFE algorithm's performance method, and CHI-SQUARE is shown in table 3 of Recall. When comparing the existing feature selection techniques (CHI-SQUARE and RFE) with the Proposed HYBRID RFFI + RFE ALGORITHM, It is clear that the suggested approach produces better outcomes. The chi-square values range from 0.78 to 0.87, while the RFE values range from 0.62 to 0.76, whereas the Proposed HYBRID RFFI + RFE ALGORITHM achieves significantly higher values ranging from 0.83 to 0.97. This clearly shows that the proposed method provides more accurate and robust feature selection, resulting in enhanced model performance and improved predictive accuracy.

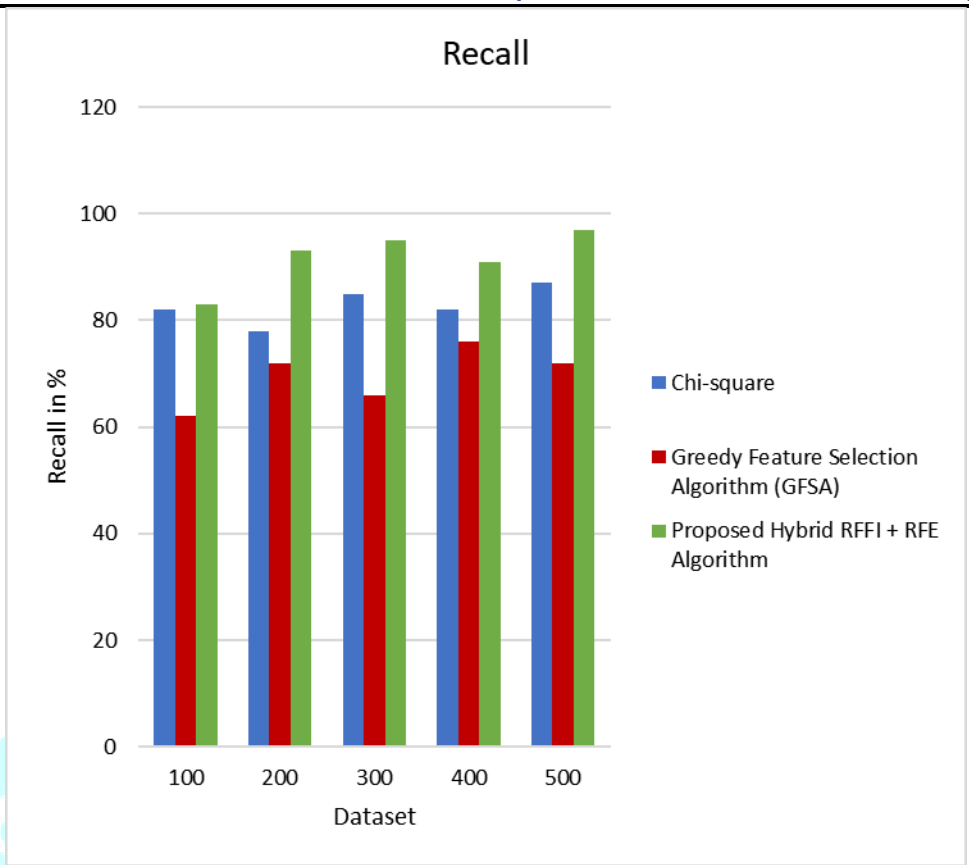


Figure 3.Comparison Chart of Recall

The Figure 3 shows the recall comparison graphic, which shows how well the current (CHI-SQUARE), Greedy Feature Selection Algorithm (GFSA), and the Proposed HYBRID RFFI + RFE ALGORITHM approach. The The dataset size is indicated by The recall ratio is displayed on the X-axis and the Y-axis. The proposed hybrid RFFI + RFE algorithm consistently outperforms the existing methods in terms of values. Specifically, CHI-SQUARE values range from 0.78 to 0.87, RFE values range from 0.62 to 0.76, while the Proposed HYBRID RFFI + RFE ALGORITHM values range from 0.83 to 0.97. This clearly the dataset size is indicated by the X-axis, and the Y-axis shows the recall ratio. The recommended hybrid RFFI + RFE algorithm's valuesare continuously greater than those of the current techniquesbaseline feature selection methods, achieving superior recall and thereby enhancing the overall predictive performance of the model.

4.4 F -Measure

The F1-measure is a test accuracy metric that integrates recall and precision.The precision and recall harmonic means are used in the computation.

$$F1 - Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)}$$

Dataset	Chi-square	Greedy Feature Selection Algorithm (GFSA)	Proposed Hybrid RFFI + RFE Algorithm
100	86	75	96
200	88	76	98
300	82	68	96
400	77	66	94
500	78	67	92

Table 4.Comparison Table of F -Measure

This analogy F-Measure values in Table 4 demonstrate the effectiveness of the current (CHI-SQUARE), Greedy Feature Selection Algorithm (GFSA), and the Proposed HYBRID RFFI + RFE ALGORITHM approach. When comparing the traditional feature selection techniques (CHI-SQUARE and RFE) with the proposed hybrid RFFI + RFE algorithm method, the proposed approach consistently delivers superior results. The CHI-SQUARE values range from 0.77 to 0.88, RFE values range from 0.66 to 0.76, whereas the Proposed HYBRID RFFI + RFE ALGORITHM values range from 0.92 to 0.98. These results show unequivocally that the suggested hybrid RFFI + RFE algorithm performs better than the conventional approaches by providing higher F-Measure values across all dataset sizes, thereby enhancing the overall predictive performance.

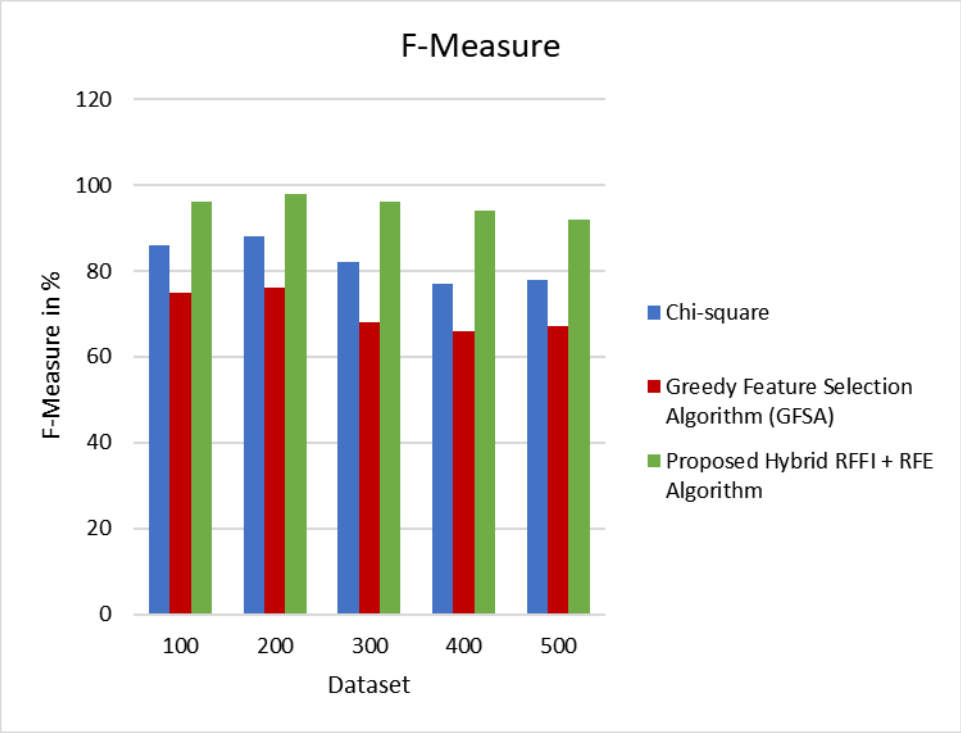


Figure 4.Comparison Chart of F -Measure

The Figure 4 displays the F-Measure comparison graphic, which illustrates the effectiveness of the current (CHI-SQUARE), Greedy Feature Selection Algorithm (GFSA), and the suggested RFFI + RFE Algorithm hybrid method. The dataset's size is displayed on the X-axis, while the F-Measure ratio is displayed on the Y-axis. The proposed hybrid RFFI + RFE algorithm consistently outperforms the existing methods in terms of values. Specifically, CHI-SQUARE values range from 0.77 to 0.88, RFE values range from 0.66 to 0.76, whereas the Proposed HYBRID RFFI + RFE ALGORITHM values range from 0.92 to 0.98, clearly indicating superior performance. The proposed feature selection strategy thus provides significant improvement in predictive accuracy and delivers more robust and reliable results compared to traditional filter and wrapper approaches.

Conclusion

The foundation of creating dependable and successful machine learning models is feature selection. In order to increase fraud detection, this phase created FS-Hybrid, a hybrid feature selection technique in insurance claims by Combining Recursive Feature Elimination and Random Forest Feature Importance. The integration of global feature importance ranking with iterative elimination ensured that only the most relevant and discriminative attributes were retained for modeling. Experimental findings revealed that the reduced subset of 10 features outperformed the full 45-feature dataset regarding F1-score, recall, accuracy, and precision, demonstrating that dimensionality reduction leads to both improved model interpretability and predictive efficiency. Furthermore, FS-Hybrid provides a generalizable framework that can be applied to other domains where data redundancy and noise are prevalent. Future work will extend this approach by exploring

ensemble-based feature selection and adaptive thresholding strategies to further enhance scalability and robustness.

References

1. Hokijuliandy, E., Napitupulu, H., &Firdaniza. (2023). Application of SVM and Chi-Square Feature Selection for Sentiment Analysis of Indonesia's National Health Insurance Mobile Application. Mathematics, 11(17), 3765. https://doi.org/10.3390/math11173765.
2. Taha, A.; Hadi, A.S.; Cosgrave, B.; Mckeever, S. A Multiple Association-Based Unsupervised Feature Selection Algorithm for Mixed Data Sets. *Expert Syst. Appl.* **2022**, 1–31. [**Google Scholar**]
3. He, X.; Cai, D.; Niyogi, P. Laplacian score for Feature Selection. *Adv. Neural Inf. Process. Syst.* **2005**, *18*, 507–514. [**Google Scholar**]
4. Zhao, Z.; Liu, H. Spectral feature selection for supervised and unsupervised learning. In Proceedings of the 24th International Conference on Machine Learning, New York, NY, USA, 20–24 June 2007; pp. 1151–1157. [**Google Scholar**]
5. Solorio-Fernández, S.; Martínez-Trinidad, J.F.; Carrasco-Ochoa, J.A. A new unsupervised spectral feature selection method for mixed data: A filter approach. *Pattern Recognit.* **2017**, *72*, 314–326.
6. Bouchlaghem, Y., Akhtat, Y., & Amjad, S. (2022). Feature Selection: A Review and Comparative Study. E3S Web of Conferences, *351*, 01046. https://doi.org/10.1051/e3sconf/202235101046.
7. Sanz, H., Valim, C., Vegas, E., Oller, J. M., &Reverter, F. (2018). SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. BMC Bioinformatics, *19*, 432. https://doi.org/10.1186/s12859-018-2451-4.
8. Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. Neural Computing and Applications, *24*(1), 175–186. https://doi.org/10.1007/s00521-013-1368-0.
9. Wang, F., Zain, A. M., Ren, Y., Bahari, M., Samah, A. A., Ali Shah, Z. B., Yusup, N. B., Jalil, B. A., Mohamad, A., & Azmi, N. F. M. (2025). Navigating the microarray landscape: a comprehensive review of feature selection techniques and their applications. *Frontiers in Big Data*, 8:1624507. doi: 10.3389/fdata.2025.1624507.
10. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, *3*, 1157-1182.
11. Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers and Electrical Engineering, 40(1), 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.
12. Beraha, M., Metelli, A. M., Papini, M., Tirinzoni, A., & Restelli, M. (2019). Feature Selection via Mutual Information: New Theoretical Insights. ArXiv Preprint. Retrieved from https://arxiv.org/abs/1907.07384v1.

13. Sosa-Cabrera, G., Gomez-Guerrero, S., Garcia-Torres, M., & Schaerer, C. E. (2023). Feature Selection: A perspective on inter-attribute cooperation. *ArXiv Preprint*. Retrieved from.

14. Mielniczuk, J. (2022). Information Theoretic Methods for Variable Selection—A Review. Entropy, *24*(8), 1079. https://doi.org/10.3390/e24081079.

15. Molina, L. C., Belanche, L., & Nebot, A. (2002). Feature Selection Algorithms: A Survey and Experimental Evaluation. In Proceedings of the 2002 IEEE International Conference on Data Mining (pp. 306-313). IEEE.

16. Johnson, T.F.; Isaac, N.J.; Paviolo, A.; González-Suárez, M. Handling missing values in trait data. *Glob. Ecol. Biogeogr.* **2021**, *30*, 51–62.

17. Taha, A.; Hadi, A.S. A general approach for automating outliers identification in categorical data. In Proceedings of the ACS International Conference on Computer Systems and Applications (AICCSA), Ifrane, Morocco, 27–30 May 2013; pp. 1–8.

18. Tang, C.; Liu, X.; Li, M.; Wang, P.; Chen, J.; Wang, L.; Li, W. Robust unsupervised feature selection via dual self-representation and manifold regularization. *Knowl. Based Syst.* **2018**, *145*, 109–120.

19. Taha, A.; Hadi, A.S. Pair-wise association measures for categorical and mixed data. *Inf. Sci.* **2016**, *346*, 73–89.

20. Gomes, C.; Jin, Z.; Yang, H. Insurance fraud detection with unsupervised deep learning. *J. Risk Insur.* **2021**, *88*, 591–624.

21. Scriney, M.; Nie, D.; Roantree, M. Predicting customer churn for insurance data. In *International Conference on Big Data Analytics and Knowledge Discovery*; Springer: Cham, Switzerland, 2020; pp. 256–265.