International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
***Article Received:** 15 April 2023 **Revised:** 25 July 2023 **Accepted:** 18 August 2023*

Data Mining with Supervised Instance Selection Improves Artificial Neural Network Classification Accuracy

Mr. S. Srinivas Reddy
Department of Computer Science & Engineering
Dr. A.P.J. Abdul Kalam University
Indore, India
srinivasreddy.sarvigari@gmail.com

Dr. Rajeev G. Vishwakarma
Department of Computer Science & Engineering
Dr. A.P.J. Abdul Kalam University
Indore, India
rajeev@mail.com

Abstract—IDSs may monitor intrusion logs, traffic control packets, and assaults. Nets create large amounts of data. IDS log characteristics are used to detect whether a record or connection was attacked or regular network activity. Reduced feature size aids machine learning classification. This paper describes a standardised and systematic intrusion detection classification approach. Using dataset signatures, the Naive Bayes Algorithm, Random Tree, and Neural Network classifiers are assessed. We examine the feature reduction efficacy of PCA and the fisheries score in this study. The first round of testing uses a reduced dataset without decreasing the components set, and the second uses principal components analysis. PCA boosts classification accuracy by 1.66 percent. Artificial immune systems, inspired by the human immune system, use learning, long-term memory, and association to recognise and v-classify. Introduces the Artificial Neural Network (ANN) classifier model and its development issues. Iris and Wine data from the UCI learning repository proves the ANN approach works. Determine the role of dimension reduction in ANN-based classifiers. Detailed mutual information-based feature selection methods are provided. Simulations from the KDD Cup'99 demonstrate the method's efficacy. Classifying big data is important to tackle most engineering, health, science, and business challenges. Labelled data samples train a classifier model, which classifies unlabeled data samples into numerous categories. Fuzzy logic and artificial neural networks (ANNs) are used to classify data in this dissertation.

Keywords- IDS , Principal Component Analysis, ANN, KDD Cup'99, Artificial Neural Networks.

I. INTRODUCTION

The framework for intrusion detection entails analysing data from several perspectives and then gathering the outcomes. As per the literature, it refers to the significant task of identifying authentic, novel, and potentially significant data patterns. Knowledge discovery, often known as data mining, is the systematic process of extracting meaningful and relevant information from a vast database [1]. Data mining uses a diverse range of statistical methodologies to reveal previously unidentified relationships between datasets and provide exceptionally precise forecasts on these relationships. Data mining requires the integration of several disciplines such as statistics, machine learning, pattern recognition, neural networks, image processing, and database systems. Knowledge discovery in databases and its associated methodologies are often seen in many situations. The key terms of a KDD process are visually shown. The KDD method is divided into many stages:

The chosen option.

- Pre-processing involves transforming the data, if necessary.
- The identification and extraction of patterns and connections.
- The analysis and assessment of discovered structures.
- Data mining has the capacity to enhance many facets of a firm, including financial performance, client interactions, and operational efficiency.

CRM has become an essential business strategy in today's fiercely competitive corporate landscape. The objective of customer relationship management (CRM) is to ensure client satisfaction and loyalty, hence encouraging repeat business. CRM, or customer relationship management, refers to the administration of business processes and technologies that are focused on understanding customers inside an organisation. Recent successful data mining applications often include identifying the most profitable customers for product sales, determining the most lucrative market categories for a company, devising strategies to increase product market share, reducing costs without affecting the production process, and optimising inventory management. The process of data mining may be divided into three equally significant phases.

- To predict future behaviour, it is essential to first extract previous data for the purpose of training patterns and models.

- This can involve predicting consumer interest in new products, the probability of consumers buying specific products, and maximising profits.
- New transactions are evaluated based on the probability of exhibiting the predicted behaviour, and these evaluations are then used to optimise a business goal.

Customer segmentation, a fundamental principle of CRM, enables businesses to reduce advertising costs. It facilitates the achievement of market growth success, making it more evident and gratifying. Data mining has the ability to provide several results, such as customer retention, upselling, and cross-selling. In 2006. Predictive modelling is the most often used data mining technique for creating corporate intelligence and operational support systems [2]. Prediction systems may be developed in data-rich contexts using reliable and automated methods. Companies may achieve a more equitable distribution of their highest-paying clients by effectively managing their interactions with them. CRM is crucial for collecting and converting operational data into meaningful insights to provide proactive customer service.

II. LITERATURE REVIEW

The publication by White in 2003 The user's text is "[1]". In order to mitigate the impact of data-snooping, we used a methodology to ascertain the statistical significance of portfolios formed during training. The portfolio chosen by SVR exhibits a Value at Risk (VaR) of -6.87% and a cumulative return of 374.40% due to the use of a reverse multiquadric kernel. The findings of this research support the notion that the new SVR technique of portfolio development is superior due to its robust predictive capabilities and ability to manage highly complex interactions.

The study conducted by Gunasundari et al. in 2012 The user's text is "[2]". An analogous automated method for early diagnosis of liver issues using CT imaging has been developed by A computed tomography (CT) image of the liver may be used to separate the liver and lesion using a hepatoma or hemangioma histogram analyzer and morphological operation. In this case, a Grey Level Co-occurrence Matrix (GLCM) is used to extract biorthogonal wavelets using the Fast Discrete Cosine Transform. Various neural networks, such as the Back Propagation Neural Network (BPN), the Probabilistic Neural Network (PPN), and the Cascade feed Forward Neural Network (CFBPN), are trained utilising textual input and extracted attributes. The neural networks' outputs are compared to identify the optimal mix of characteristics and neural networks.

Guo Yuanyuan (2012) The user's text is enclosed in tags. Utilising half-supervised learning techniques may enhance classification accuracy when there is a shortage of labelled samples. Self-training and co-training are two learning paradigms that include choosing and labelling scenarios where the current classifier has a high level of confidence in order to update the classification system. Unlike the samples that have been self-identified and grouped into categories, the original examples that have been labelled are more dependable. The accuracy of classification may be compromised if incorrect labels are supplied to unlabeled cases and then the classifier is altered. This study presents a novel approach, referred to as ISBOLD, for the selection of instances from the original dataset. ISBOLD assesses the performance of a classifier not just on freshly labelled data but also on previously categorised data. ISBOLD use the enhanced accuracy of the newly acquired classifier compared to the initial labelled data to determine whether unlabeled instances are eligible to advance to the subsequent iteration. We used Naive Bayes's base classification to evaluate both self-training and co-training situations. The findings from studies done on 26 UCI datasets demonstrate that the use of self-training and co-training ISBOLD may greatly enhance both accuracy and AUC.

Elena Marchiori published a work in 2012. The user's text is "[4]". Supervised learning involves employing an algorithm to build a classification model based on a training set that contains labelled examples. This model is then used to predict the class label of future instances for generalisation. The presence of noise in the data and the use of small sample sizes are both factors that might impact the training process and, therefore, the accuracy of the generalisation. The suggested hit misses' network (HMN) in this study is a distinctive network model that describes the closest neighbour link between pairs of data from each pair of classes. We demonstrate a correlation between the structural characteristics of HMNs and various attributes of training points, such as the border and centre points, that are connected with the nearest (1-NN) choice rule. This is the reason why we use HMN (Heterogeneous Multimodal Network) to enhance the accuracy of 1-NN (1-Nearest Neighbour) classification by removing cases from the training set using a process known as instance selection. We provide three novel HMN-based methodologies, exemplifying our capabilities. There are three different versions of HMN that have been created: HMN-C, which eliminates examples from the original training set without affecting the accuracy of the nearest neighbour algorithm; HMN-E, which uses a more drastic decrease in storage; and HMN-EI, which applies HMN-E several times. A total of 22 datasets with diverse attributes, such as input size, cardinality, class balance, class number, noise content, and the existence of duplicated variables, are used to evaluate their performance. The experimental findings demonstrate a substantial improvement in the accuracy of 1-NN classification when including HMN-EI into these datasets. The datasets demonstrate the superior generalisation performance of HMN-EI, with no significant disparity in storage needs compared to state-of-the-art approaches. The findings indicate that HMN provides a robust visual representation of a training set, which may effectively reduce noise and redundancy in instance-based learning.

Matteo Di Benedetti, born in 2012. The user's text is "[5]". Three separate investigations were undertaken for the dissertation. The main objectives of both the first (study 1) and subsequent (study 2) investigations are to identify and assess damage, with a particular focus on detecting corrosion via the use of Acoustic Emission (AE) techniques. The latest discoveries (Study 3) provide a novel AE technique for evaluating the extent of damage in reinforced concrete elements during load testing.

The author of the publication is N. Selvadorai, and it was published in 2012. The user's text is "[6]". By incorporating Acoustic Emission (AE) data into mathematical models, efforts are made to prevent fatigue cracking in metal beams. Nine specimens of a steel I beam were subjected to acoustic emission (AE) testing after undergoing cyclic loading that caused bending at three specific places. The data obtained during these tests has been purged of long-term trends, excessive occurrences, and clear anomalies. The data that had undergone filtration were then categorised into several metal failure modes. This categorization was based on the length, energy, amplitude, and average frequency of the acoustic emission (AE) events. A neural map, using the Neural Works® Professional II/Plus (SOM) network programme, was used for this classification. Plastic deformation predictions were generated using data from mathematical models that are grounded on the features of the AE failure process. Subsequently, we use constrained Johnson and Weibull distributions to provide a mathematical representation of the amplitude data derived from the categorised plastic deformation data.

The research conducted by Ireneusz Czarnowski in 2012 [7] Data reduction, also known as instance selection in supervised machine learning, refers to the process of determining which instances from the training set will be used in the subsequent stages of the learning process. The choice of configuration may result in accelerated learning times, enhanced scalability for large data sources, or increased robustness. This inquiry presents four iterations of an agent-based study methodology for the purpose of example selection, with each iteration using a clustering algorithm as its foundation. The main conjecture is that examples are selected after the clustering of training data. The simulation aimed to evaluate the suggested methodology and quantify the impact of the clustering technique on the accuracy of classification.

In their study, EvginGoceri et al. (2012) [8] introduced a segmentation technique that integrated a modified version of the k-means algorithm with a specific kind of localised contouring. This technology enables the achievement of high accuracy in 3D rendering and liver segmentation. The authors have developed a unique approach for completely automated liver segmentation. This approach uses a dual-series methodology that only relies on integer operations and circumvents the need to solve partial differential equations. The proposed approach is efficient since it applies the algorithm uniformly to all slices, using the same amount of iterations. Additionally, the contour development is achieved without requiring a pre-established starting contour.

The study conducted by Deng et al. in 2012 [9] In order to determine the actual decrease in the decision-making system utilising the given attribute subset, a novel Blindly Deletion Algorithm using Inverse Ordering (BDAIO) was devised. The suggested approach was built upon the Blindly Deleting Algorithm (BDA). BDAIO's performance has been validated using several datasets sourced from the UCI library. Experimental results showed that classifiers exhibited improved performance when trained using the BDAIO reduction set.

Liang et al. (2012) developed an innovative and efficient feature selection approach to handle large-scale decision-making based on rough theory. The E-FSA identified an appropriate set of attributes by decomposing a complex decision table into smaller ones and then combining the characteristics of these smaller tables. The experimental results shown that the E-FSA is a rapid and efficient approach for handling large decision tables.

In 2012, Kanik [11] introduced a method for rebuilding the original decision table and implementing a two-step attribute reduction procedure. The subsequent procedure calculates the proportion of the initial data that has to be removed. A discretization approach was used to narrow down the scope of an irreducible and optimum attribute, while maintaining the consistency of the dataset's categorization. The reasoning is based on the combination of the matrix effect and the reduction of individual features. When subjected to rigorous testing utilising actual biometric data, the projected prediction technique emerged as the most successful. By using the rough set theory and implementing a novel preprocessing methodology, the issue is handled with superior effectiveness compared to previous methods.

The study conducted by Bayram et al. in 2012 The user's text is "[12]". Servet conducted a comparative analysis of dyslexic and normal readers while reading texts and pseudowords. Fifteen Turkish teenagers diagnosed with dyslexia, the majority of whom have difficulties in reading, participated in the study. Fifteen Turkish kids also encountered no obstacles in their reading ability. The participants' eye movements were captured as they read passages and pseudowords. The dyslexic readers had more and longer fixations while reading both text and pseudowords, although there were very few instances of regression. Individuals with dyslexia had a higher number of fixations due to the presence of lengthier words and pseudowords.

The authors Dong ping and Tian published a paper in 2013 [13]. The extraction and encoding of features is a vital stage in multimedia processing, as mentioned by. They concluded that more inquiry into the relationship between the number of features and performance results was necessary, based on their findings. Increased functionality should presumably result in enhanced performance. Furthermore, an intriguing and demanding subject is in the correlation between the portrayal of characteristics and the ultimate achievement. The inclusion of feature representation techniques, such as global, block, and area features, is included. The size of the division or segmentation plays a crucial role in determining the ultimate performance of blocks and regions. It is crucial to analyse the relationship between the best mix of factors and the overall effectiveness to determine the possibility of further improving performance.

The authors Ali-Khashashneh and Al-Radaideh published a paper in 2013, referenced as [14]. The Johnson's reduction technique and the reduction of objects employing the feature weighting technology for reduction computation were offered. Both techniques seek to reduce the dimensionality of the distinguishability matrix by concentrating on a smaller

***Article Received:** 15 April 2023 **Revised:** 25 July 2023 **Accepted:** 18 August 2023*

number of data sets. When dealing with data sets that have a high number of characteristics, it is advised to use a reduction strategy that requires significant processing resources. Matrix-based reduction computation algorithms are often used for function selection due to their heuristic nature and ease of identification. The use of the attribute weighting algorithm and the Johnson reduction algorithms facilitated the development of feature selection techniques that effectively minimise redundant data points and accurately identify the most essential ones for information classification.
The study conducted by Govindaraj et al. in 2013 is referenced as [15]. A diverse range of image filtering methods has been examined in a research report. Image filtering is an essential component of image processing that aims to remove background noise from displays. Multiple filters may be applied to a picture. There are advantages to using each unique filtering technique during picture processing. Applying the hybrid median filter and the trimmed alpha filter to the picture might potentially enhance the quality by reducing the salt and pepper noise, surpassing the performance of existing filters.
The study conducted by Prabhdeep et al. in 2013 is referenced as [16]. Smoothing pictures is an essential and prevalent task in the field of image processing, with a multitude of techniques available for this purpose. This study highlights the significance of the non-linear filtering approach, namely median filtering, in the realm of picture filtering, where many algorithms and techniques are used. The research conducted by Govindaraj et al. (2013) has examined several techniques for picture filtering. The findings indicate that early stages of vision processing may address multiple problems, including random intensity changes, light fluctuations, and low contrast. Image filtering is an essential component of image processing that aims to remove background noise from displays. Multiple filters may be applied to a picture. There are advantages to using each unique filtering technique during picture processing. Applying the hybrid median filter and the trimmed alpha filter to the picture might potentially enhance the quality by reducing the salt and pepper noise more effectively compared to existing filters.
The study conducted by Ilakkiya et al. in 2013 [17] Classification, as advocated by experts, is crucial for precise cancer diagnosis and efficient therapy. Despite current challenges, investigating the gene expression patterns of liver tumours has the potential to expedite patient categorization if accomplished well. The current methodologies use morphological and clinical attributes to categorise distinct forms of cancer. The use of microarray technology has facilitated the iterative examination of many genes, hence contributing to the generation of gene expression data that may be employed for cancer classification. This concept has been widely used in precise cancer classification trials, where data mining and machine learning techniques are combined to pick genes. The microarray data sets are classified using Fuzzy Neural Network (FNN) and Principal Component Analysis (PCA), which identify the characteristic features of self-production, restriction augmentation, and rules-based popularisation.
Huang et al. (2013) [18] devised a voxel-based method to segment tumours in three dimensions. The ELM is a rapid-learning voxel classifier created by the Extreme Learning Machine. Through the random selection of data in three-dimensional space (ROI), the Extreme Learning Machine (ELM) is capable of identifying parts of the liver that are healthy and comparing them to a two-class ELM in that specific area.
The study conducted by Selvathi et al. in 2013 is referenced as [19]. The study used a differential analysis approach to examine liver cancer pictures and introduced a novel segmentation method called Contour and Extreme Machine Learning hybrid segmentation. The liver segmentation is created using adaptive thresholding and morphological processing. Clustering using Fuzzy C mean (FCM) algorithm uses statistical and linguistic data from tumour outlines to aid in the conversion of the tumour into a liver tumour. The radiologist gains advantages from enhanced accuracy and a different viewpoint offered by the segmented tumour picture.

III. PROPOSED METHOD

3.1 ANN based classifier with reduced features

In order for the neural network technique to be feasible for a substantial engineering challenge, it is imperative to reduce the dimensions. Both the cost of measurements and the precision of classifications require minimising the number of parameters used to describe patterns. Dimensionality reduction may be achieved by many methods, such as feature extraction and feature selection. Feature extraction is a technique that converts the original collection of features into a reduced set of features, either by linear or nonlinear methods. The process of selecting a function entails the deliberate choice of a subset from a pool of accessible qualities. Both strategies strive to enhance the speed and efficiency of estimation by minimising the number of characteristics that a network has to take into account.

This chapter presents the notion of mutual information and a reciprocal information technique to aid in the selection of suitable functions for neural network classifiers. This approach is used to ascertain the underlying factors behind issues related to computer network intrusion detection.

3.1.1 Dimensionality reduction techniques

The performance of the classifier is determined by the connection between the sample size, the number of features, and the complexity of the classifier. To assign a class label to each cell of the function space using a standard table-lookup approach, a feature dimension is needed that grows exponentially with the amount of training data points. The curse of dimensionality is the ultimate consequence of the categorization process, representing the greatest level of phenomena in its formation. The input data should be concise for two primary reasons: to facilitate the measurement of expenditures and to guarantee precise classification.

A concise set of features is used to streamline the representation of the pattern and the classifiers that utilise it. Therefore, the resultant classifier is more efficient and demands less memory. Additionally, having a small number of features may help mitigate the problem of high dimensionality when there are only a limited number of training examples available. Conversely, decreasing the amount of characteristics may negatively impact the accuracy of the data categorization system.

The main difficulty in dimensionality reduction is in the selection of a criteria function. The classification error of the feature subset is a widely used metric to assess its quality. Nevertheless, properly predicting the classification error becomes unattainable when the sample size is very small compared to the number of attributes.

When selecting a criteria function, it is crucial to assess the appropriateness of the resultant reduced feature space. Size may be reduced using two main methods: feature extraction and feature selection. The specific details of these approaches are as follows:

Feature Extraction: The process of transforming the original collection of features into a smaller set of features, either using linear or nonlinear methods. Functional extraction methods aim to discover a solution in the original space d (m d) by using an appropriate space m, either in a linear or nonlinear manner. Principal analysis, component analysis, linear discrimination, and projection are often used linear transformations in the pattern recognition field to extract features and reduce dimensions.

The principle components analysis (PCA) or Karhunen-Loève expansion is a method that identifies the m biggest intrinsic vectors of the d x d covariance matrix of n d-dimensional patterns. It is widely recognised as the most popular linear feature extractor. Principal component analysis uses the most influential functions (eigenvectors with the greatest eigenvalue) to accurately estimate data using the mean squared error criteria inside a linear subspace. Projection pursuit and independent component analysis (ICA) are particularly suitable for non-Gaussian distributions since they do not rely on second-order data features.

Multidimensional scaling (MDS) is a non-linear extraction approach. The objective is to map a characteristic space of d dimensions onto a 2- or 3-dimensional space, while preserving as much of the original distance matrix as possible. Due to the absence of an explicit mapping function in MDS, it is not possible to apply a new model to a precomputed map that was generated for a specific training set.

3.1.2 Selection of feature

Function Selection chooses a subset from a given collection of features. To precisely define the function selection issue, we may state it as follows: Select a subset of size M, consisting of characteristics D, in order to minimise the estimation error. The issue of subset selection may be effectively answered by doing an exhaustive search of all potential subsets of functions. The total of all subsets that need to be taken into account (Ns).

$$N_s = \left(\frac{D}{M}\right) = \frac{D!}{M!(D-M)!} \qquad (1)$$

Indeed, the optimal subset can only be determined by the exhaustive search method, but this approach usually requires an unreasonably lengthy amount of time.

Various suboptimal feature selection techniques have been proposed in the literature. These approaches include searching for the optimal combination of attributes based on certain criteria. We may categorise these models into two categories based on whether feature selection is conducted independently of the learning strategy used to train the network. If feature selection is performed independently from the learning process, the technique is said to use a filter approach 1. Initially, the data is assessed, maybe via statistical techniques, to extract the features of the dataset that are most relevant to the current challenge. The network is trained using the relevant variables. The attributes are selected autonomously, without considering the network's future performance.

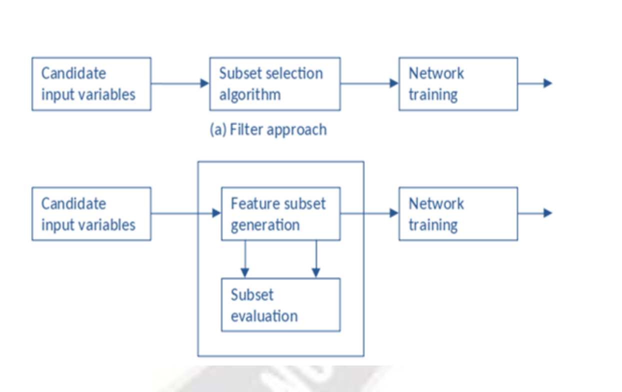
Candidate input variables
Subset selection algorithm
Network training
(a) Filter approach
Candidate input variables
Feature subset generation
Subset evaluation
Network training

Figure 1. Feature Selection

Conversely, the wrapper technique integrates the learning algorithm into the function selection process. Through comprehensive exploration of all potential traits, we can identify the optimal ones to use in our study. In other words, a certain set of characteristics is selected and assessed. The assessment is conducted by training and testing the network using various subsets of features. Wrapper's technique typically surpasses the filter because to the dynamic interaction between the learning algorithm and the training data. It is not as efficient as filtering, however.

The decision between feature selection and feature extraction is contingent upon the issue domain as well as the amount and quality of the training data that is accessible.

By eliminating some attributes throughout the selection process, the costs of measurement are minimised, and the remaining traits maintain their original physical relevance.

The conserved characteristics may further provide insight into the underlying physical mechanism at play in this context. The modified characteristics generated by feature extraction may provide superior differentiation compared to the optimal subset of characteristics, although these novel characteristics (a linear or non-linear amalgamation of specific traits) may lack any apparent physical significance.

To address the issue of excessive dimensionality in the data, this study employs a method known as mutual feature selection. The next paragraphs will provide a thorough examination of the specifics.

3.2 Mutual information-based feature selection

3.2.1 Concept of mutual information

Entropy quantifies the average level of randomness in an experiment. Consider a discrete random variable Y with possible values yi, where I ranges from 1 to NY. Let Prob(Y=yi) = Pi represent the probability distribution function that describes the outcomes of a random experiment.

Subsequently, the entropy of the experiment may be precisely described as

$$H(Y) = -\sum_{i=j}^{N_y} P_i logp_i$$

The entropy of a random experiment may be reduced by acquiring more knowledge X. The conditional entropy of a random experiment may be determined if you have knowledge about X.

$$H(Y|X) = \sum_{j-1}^{N_y} P_j \left(\sum_{j-1}^{N_y} p\left(y_{i\text{¿}}\, x_j\right) \log P\ \left(y_{i\text{¿}}\, x_j\right) \right)$$

Where P_j is the probability distribution function of X with possible values x_j, j=1,2,…,N_x and $P(y_i/x_j)$ is the conditional probability for y_i given x_j. The conditional entropy is always less than or equal to the original entropy.

The mutual information I(Y; X) between Y and X is, by definition, the amount by which the entropy (uncertainty) is reduced:

$$I(Y;\, X) = H(Y) - H(Y/X)$$

As a result, the mutual information level lowers the typical degree of uncertainty regarding the experiment's random outcome Y. Mutual information is the symmetrical metric. Seeing X is similar to seeing Y in terms of the amount of knowledge gained about Y. X is the raw data, and Y is the final class label for this function selection problem.

3.2.2 Computation of mutual information

In order to calculate mutual information, it is necessary to possess the probability distribution of variables that are currently unavailable. Consequently, we must rely on the histogram of the data as a substitute.

In order to get the inverse data from the histogram of the training data, the following procedures need to be followed:
Step 1: Arrange the output patterns in a descending order and then split the sorted patterns into Ny equal groups.
Step 2: In the case when the input variable is not known, it is necessary to compute the initial entropy of the output Y, which has a value of 4.2.
Step 3: Arrange the patterns in the first input variable in a certain sequence, dividing X1 into groups of Nx in a decreasing manner.
Step 4: Compute the conditional entropy of Y given the input variable X1.
Step 5: Calculate the reciprocal of Y's information on X1.
Step 6: Duplicate Steps 3 to 5 for the remaining variables.

3.3 Mixed genetic algorithm approach for fuzzy classifier design

The development of a fuzzy categorization system is strongly dependent on the formulation of if-then rules and membership functions. In the previous chapter, we saw the use of a binary coded genetic algorithm in the development of a fuzzy classifier. In the binary genetic algorithm, strings are employed to represent solution variables such as rullset and membership function. However, this can result in problems related to Hamming Cliff membership. Additionally, using a fixed length binary coder to represent all the possible options becomes difficult when dealing with discrete variables that do not have a value equal to 2 raised to the power of k (where k is an integer).

In this study, we propose a Mixed Genetic Algorithm (MGA) methodology for optimising the selection of rules and membership functions in a fluid-coated classifier. The suggested MGA uses a revised set of genetic operators and a hybrid kind of representation to construct the classificatory. The effectiveness of the suggested technique is shown by creating a fuzzy classification for data obtained from Iris, Wine, and Tcpdump.

3.3.1 Mixed genetic algorithm

To enhance the performance of the classifier, this chapter proposes the implementation of the following enhancements to the mixed genetic algorithm.The mixed chromosomes reflect both the rule set and genetic population membership, with the membership functions represented by actual values and the rule set represented by a binary string.Employing a hybrid representation does not modify the selection procedure, rather, it necessitates the implementation of novel crossover and mutation operators. The specific crossover representation and mutation operators used in this investigation will be outlined in the next section.

3.3.2 Representation

The first and vital aspect when using evolutionary algorithms for the design of a fluid system is the choice of representation technique. The ruleset is represented using binary strings. A substring of two bits represents each variable in the rule set. Two extra bits are used to represent the output class. The genetic population has a limited number of 'Nr' rules, enabling the construction of succinct regulations. A rule selection bit is used to determine the optimal rules to apply inside the 'Nr' ruleset.

The membership function is shown as a count of decimal places. Each variable may have one of three potential input levels: low, medium, or high. The trapezoidal membership function is used to indicate the median input variable value, as well as lower and higher values. Figure 2 displays a representative chromosomal structure.

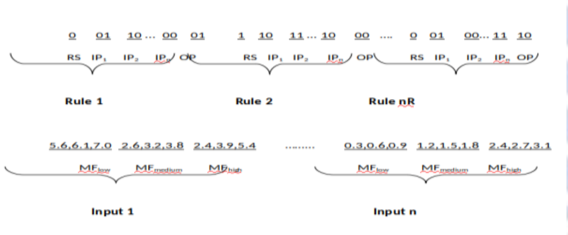

Figure 2: Chromosome Structure

Utilising this format instead of binary encoding when handling membership functions has several advantages. By eliminating the necessity to convert input variables to a binary type, the performance of the GA is enhanced.

3.3.3 Crossover operator

There are two types of crossing over methods used when dealing with rule sets stored in binary. The Gene Cross Swap Operator (GCSO), an improved operator, is used after the implementation of the two-point crossover operator. The GCSO identifies variant alleles by using data found in low-fit strings, while two-point crossover enables the transfer of information across highly-fit chromosomes.

The real-valued membership function uses a simple arithmetic crossing. Figure 3 displays the Arithmetic crossover for the one-dimensional scenario. The variables Pi1 and Pi2 denote the two parents, whereas ai and bi indicate the lower and upper limits for the ith variable, respectively.

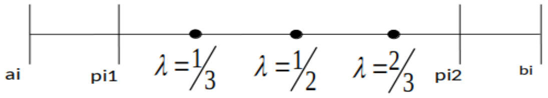

Figure 3: Schematic Representation of Arithmetic Crossover

The arithmetic crossover technique produces two offspring by randomly selecting two chromosomes, pi1 and pi2, from a given variable representing the parents.

$$C_i^1 = \lambda'^{p_i^1} + (1 - \lambda)P_i^2$$

$$C_i^2 = \lambda'^{P_i^2} + (1 - \lambda)P_i^1$$

Where is a uniform random number [0,1].

3.3.4 Mutation operator

The mutation operator improves the population by promoting quicker and more precise convergence, while also preventing stagnation at a local maximum. The Rule set prioritises bitwise mutation, which involves randomly changing a small number of bits from 1 to 0 or from 0 to 1, with a low probability of mutation (Pm). The "Uniform Mutation" operator is used to alter the Membership function.

IV. RESULTS AND DISCUSSION

5.1 Experimental results

Our dataset is split into 15:85, 25:75, and 30:70 subsets since we're testing the AIRS algorithm on condensed training sets. Table 1 displays the results of the categorization accuracy.

Table 1. Classification Accuracy of Algorithms

Techniques	Classification Accuracy of Algorithms
Naïve Bayes	90.32%
Neural Network	97.55%
AIRS 25% Training set	99.47%
AIRS 15% Training set	99.12%
AIRS 30% Training set	99.66%
AIS + Fisher score feature selection	97.90%
Genetic algorithm	98.65%

According to Table 1, when compared to Naive Bayes and Genetic Algorithm, AIRS with a 30% training set produced a better classification accuracy of 9.8326% and 1.0186%, respectively.

Principal Component Analysis (PCA) and the Fisher score are used to reduce the number of characteristics to be utilised. The classification accuracy data are shown in Table 2.

Table 2. Displays the results of a calculation of the categorization accuracy.

Techniques	Classification Accuracy
Genetic algorithm	98.65%
AIS	99.35%
AIS with PCA	99.54%
AIS + Fisher score feature selection	97.90%

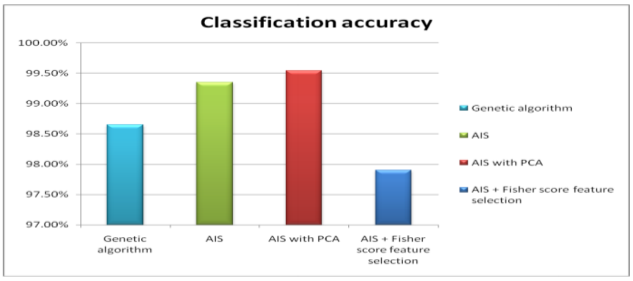

Figure 4. Feature Selection Graph

Figure 4 demonstrates that compared to the Genetic algorithm, the AIS with PCA-based feature selection yields a 0.8981& improvement in accuracy, while the AIS with Fisher score-based feature selection yields a 0.7632% increase in accuracy

5.2 Iris data classification

The iris dataset comprises 150 four-dimensional vectors, representing 50 distinct plants, with each plant belonging to one of three species: iris setosa, iris versicolor, or iris virginica. The four input characteristics are the sepal length, sepal width, petal length, and petal width. All of these input characteristics are immutable variables. Each input variable requires three fuzzy sets, resulting in a total of seven points. The domain of each membership function is dynamically defined.

Each preceding element of the genetic population's rules is denoted by a binary substring consisting of two bits: 01 represents "low," 10 represents "middle," 11 represents "high," and 00 represents "not caring." Iris Virginica is represented by the binary string '01', Iris Versicolor by '10', Iris Setosa by the binary string '11', and No Iris Setosa by '00'. These binary strings all correspond to the same output class. The genetic population requires a total of 77 bits, with 11 bits allocated for rule selection, 11 bits for input variables, and 11 bits for output classes.

All 150 samples are used as training patterns, and the fuzzy classification is employed to assess the potential of the research. The most effective genetic algorithm configurations are shown in Table 5.18. Achieving a 98% accuracy in classification, a newly developed fuzzy system presently operates with three rules. Subsequently, the input is partitioned into training and test sets to evaluate the lavonoids' capabilities. Out of the total of 150, half, or 75, are allocated for teaching, with 25 being assigned to each class. The remaining 75 are designated for testing, also at a rate of 25 per class. To determine the optimal membership feature and ruleset, it is necessary to use the training data set. The fuse classifier that was built is assessed using the test data set. All 75 data points were accurately categorised throughout the testing phase. The classifier adheres to three specific guidelines:

Iris setosa may be accurately recognised by the presence of petals with a moderate length.
Iris virginica flowers are characterised by short petal length and small petal breadth.
Iris flowers are classified as versi-colored when their sepal breadth is somewhat smaller than their petal width.
Throughout 25 generations, the functional points of the membership are evenly spread throughout the continually created and updated range. Figure 5 illustrates the ideal membership function for the petal-length dependent variable.

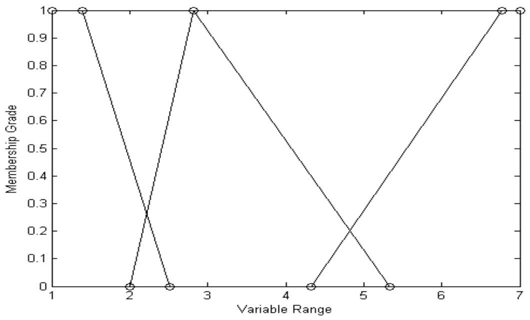

Figure 5. Optimal Membership Function for The Variable Petal Length

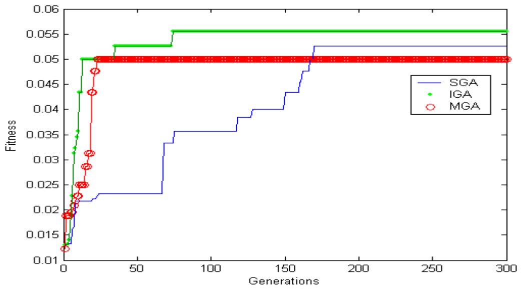

Figure 6 Convergence Behavior of MGA Compared with SGA and IGA

Article Received: *15 April 2023* ***Revised:*** *25 July 2023* ***Accepted:*** *18 August 2023*

Based on the graph, the Mixed GA model demonstrates faster convergence compared to other GA models that use binary format. The SGA required around 175 generations to get an optimal solution.

The fitness value for the IGA scenario undergoes a significant improvement over the first 20 generations, followed by a consistent and gradual improvement until reaching the optimal value around the 100th generation. During the first 25 generations, the proposed MGA exhibits a significant increase in fitness. Subsequently, convergence occurs rapidly, leading to the attainment of an optimal value. Table 3 presents a comparison between the recommended Mixed GA and two other GA models that use binary representation.

Table 3. Mixed GA Compared with Binary Coded GA for Iris Data

Performance Parameter	**LEARNING ABILITY** **Full Data: 150**			**GENERALIZATION ABILITY** **Training: 75, Testing: 75**		
	SGA	**IGA**	**MGA**	**SGA**	**IGA**	**MGA**
Generations	300	240	**200**	300	75	**50**
Population Size	30	30	**30**	30	30	**30**
Cross Over Rate	0.8	0.8	**0.9**	0.8	0.8	**0.9**
Mutation Rate	0.05	0.05	**0.1**	0.05	0.05	**0.1**
Swap Probability Rate	***	0.01	**0.02**	***	0.02	**0.02**
Correctly Classified data	146	147	**148**	73	74	**75**
No. of Rules	3	3	**3**	3	3	**3**
Accuracy in percentage	97.3%	98%	**98.6%**	97.3%	98.6%	**100%**
CPU Time in Seconds	302	257	**167.48**	247.6	56.81	**29.17**

Table 4 demonstrates that the proposed mixed genetic algorithm outperforms the binary coded genetic algorithm in terms of both learning and generalisation capabilities. Table 4 presents a comparison of the generalisation performance of the suggested strategy with other techniques found in the literature for developing the fuzzy classifier for the Iris data set. The leave-one-out approach is used for comparing the performance against the actual results.

Table 4. Performance Comparison with Other Approaches for Iris Data

Approach	**Classification Rate**	**Rules**	**Generations**
Fuzzy Rule Selection	94.67%	13	1000
SLAVE	95.7%	4	1000
Multi Objective GA	96.4%	3	690
Gaussian Fuzzy Classifier	96.7%	3	300
Proposed Method (MGA)	**100%**	**3**	**50**

According to the findings shown in table 4, it seems that the proposed Mixed Genetic Algorithm has successfully created a Fuzzy Classifier that attains a high level of accuracy using a limited number of rules. The recommended approach requires a relatively small number of generations compared to other methods mentioned in the literature.

V. CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

Data sets may be grouped using algorithm-based categorization models. Examples of data mining algorithms for intrusion classification include decision trees, naïve Bayesian classifiers, neural networks (NN), and support vector machines (SVM). Attackers regularly vary between old and new patterns, making attack detection difficult. Current network intrusion detection algorithms have many false positives. The intrusion detection model's DR and FP false positive rates determine its efficacy (FP). The number of intrusions the system successfully identified divided by the total number of incursions is used to compute DR. The FP warning may or may not imply an assault. IDS is meant to boost DR and lower FP. Signature-based methods fail to mitigate anonymous attack patterns. Discovering a new attack, creating a signature, and deploying the network takes time. To address these weaknesses, new intrusion detection systems are needed. AIS is cutting-edge and may address similar difficulties. Its rise led to new problem-solving methods.

This study examines data mining classification of normal/abnormal traffic and suggests improvements. After mining KDD 99 for UDP data streams, a multi-class dataset was generated to highlight their multiple risks. The dataset's signatures were accurately classified using Naive Bayes, Random Tree, and NN. Random tree-based classification was 99.88% accurate. In this work, we compare PCA to Fisher Score for dimensionality reduction. PCA uses data reduction to find and describe patterns to show similarities and contrasts. Fisher Score discriminates using models. It's a simple way to test your label-trait distinction.

Our analysis selected examples based on 18 criteria. Feature reduction improved classification accuracy. PCA and Fisher Score improve categorization equally. PCA's output is good, however computing the transformation matrix takes too long. Fisher Score reduces investigative qualities. GA analysis may detect abnormal network connections. Normal packets with the Satan effect are accurately detected. Network connectivity and time/location data were encoded throughout GA's development, making it efficient. This improves network anomaly detection. GA may preprocess IDS data in the future. Multi-cellular information systems organised by AIS safeguard complicated systems from malicious faults.

AIS requires self-maintenance, dispersion, and computational system flexibility. AIS optimises and finds answers. If the solution or product is important, AIS excels. An comparable concept awaits resolution. Evolving algorithms like AIS need frequent responses. Data showed that the IDS AIS classifier worked. Naive Bayes and IMMUNOS can't match Clonalg's irregular data packet detection. Compared to Naive Bayes' prediction that 38% of anomalous packets are normal, our technique predicts 12%. An IDS modifies the human immune system to safeguard networks via apoptosis. The final IDS architecture uses genetic search to select traits. Apoptosis categorization supported AIS. Experiments proved the approach worked. Classification accuracy was 0.9038% without GA function selection and 0.9988% with.

5.2 Future scope of the study

The future scope of data mining using supervised instance selection for better classification accuracy in artificial neural networks is promising and can have a significant impact on various domains. Here are some key areas of potential growth and development:

1. **Improved Classification Accuracy**: The primary goal of supervised instance selection is to improve the accuracy of classification models like artificial neural networks. As the demand for accurate predictive models continues to rise in fields like healthcare, finance, and marketing, the application of instance selection techniques will become increasingly important.
2. **Reduced Computational Complexity**: Instance selection methods can help reduce the computational complexity of training and inference in neural networks. This can lead to faster model training and inference, making real-time or near-real-time applications more feasible.
3. **Enhanced Model Interpretability**: Instance selection can help in identifying the most informative instances, which in turn can lead to more interpretable neural network models. This is crucial in domains where model transparency and explainability are essential, such as healthcare and legal applications.

References

1. White J. A, Garrett S. M, ‖ Improved pattern recognition with artificial clonal selection‖, Springer, pp. 181-193. (2003),
2. Gunasundari, S, Suganya, M &Ananthi „Comparison and Evaluation of Methods for Liver Tumor Classification from CT Datasets‟, International Journal of Computer Applications (0975 – 8887), vol. 39, no. 18. 2012,
3. Yuanyuan Guo Instance Selection in Semi-supervised Learning," Advances in Artificial Intelligence pp 158-169 (2012),"
4. Elena Marchiori" Hit Miss Networks with Applications to Instance Selection," Journal of Machine Learning Research Submitted 8/07; Revised 1/08; Published 6/08 2012,
5. Matteo Di Benedetti 'Acoustic Emission in Structural Health Monitoring of Reinforced Concrete Structures', Electronic Theses and Dissertations, University of Miami. 2012,
6. Prathikshen Nambiar Selvadorai 'Neural Network Fatigue Life Prediction in Steel I-Beams using Mathematically Modeled Acoustic Emission Data', Dissertations and Thesis, Embry-Riddle Aeronautical University - Daytona Beach, Florida. 2012,
7. Ireneusz Czarnowski Cluster-based instance selection for machine classification," Department of Information Systems, Gdynia Maritime University, Gdynia, Poland 2012,"
8. Goceri, E, Unlu, MZ, Guzelis, C &Dicle, O „An automatic level set based liver segmentation from MRI data sets‟, Image Processing Theory, Tools and Applications (IPTA), 2012 3rd International Conference on 2012. 2012,
9. Lu X, Peng X, Liu P, Deng Y, Feng B, Liao B, ‖A novel feature selection method based on CFS in cancer recognition‖, pp. 226-231. (2012),
10. Liang Jin, Suzhen Liu, Chuang Zhang & Yang Li 'Amplitude Characteristics of Acoustic Emission Signals Induced by Electromagnetic Exciting', IEEE Transactions on Magnetics, Article. 2012,
11. Kanik, T. Hepatitis disease diagnosis using rough set, Proceedings in Conference of Informatics and Management Sciences, No. 01, Slovak Republic (2012).
12. ServetBayram, MücahitCamnalbur&EsadEsgin ‗Analysis of dyslexic students ', reading disorder with eye movement tracking, Cypriot Journal of Educational Sciences, vol. 7, issue. 2, pp. 129-148 2012,
13. Dong ping Tian ‗A Review on Image Feature Extraction and Representation Techniques ', International Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 4, pp. 16-20. 2013,
14. Ali-Khashashneh, E. A. and Q. A. Al-Radaideh Evaluation of discernibility matrix based reduct computation techniques, 5th International Conference on Computer Science and Information Technology - IEEE, Amman, pp. 76-81, Jordan. (2013).
15. Govindaraj, V &Sengottaiyan, G „Survey of Image Denoising using Different Filters‟, International Journal of Science, Engineering and Technology Research (IJSETR), vol. 2, issue. 2. 2013,
16. Prabhdeep Singh &Amanarora „Analytical Analysis of Image Filtering Techniques‟, International Journal of Engineering and Innovative Technology (IJEIT), vol. 3, issue. 4. 2013,
17. Ilakkiya, G & Jayanthi, B „Liver Cancer Classification Using Principal Component Analysis and Fuzzy Neural Network‟, International Journal of Engineering Research & Technology (IJERT) vol. 2, issue. 10. 2013,
18. Weimin Huang, Ning Li &Ziping Lin &Guang-Bin Huang „Liver tumor detection and segmentation using kernel-based extreme learning machine‟, 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 3-7. 2013,
19. Selvathi D,Malini D and Shanmugavalli P, "Automatic segmentation and classification of liver tumor in CT images using adaptive hybrid technique and Contour based ELM classifier", Recent Trends in Information Technology (ICRTIT), 2013 International Conference on 25-27 July 2013