# AN EFFECTIVE PRE-PROCESSING METHOD FOR CLASSIFICATION OF SKIN DISEASE USING MACHINE LEARNING TECHNIQUES

**Mrs.K.SHUNMUGA PRIYA,** Research Scholar, Department of Computer science, Mother Teresa Women's University, Kodaikanal.

**DR(Mrs) SELVI V,** Assistant Professor, Department of Computer science, Mother Teresa Women's University, Kodaikanal.

**Abstract:**

The medical industry is making strides through modern technical advances and the development of newer healthcare and treatment techniques. Biotechnology is the basis of all these technical advancements. The health of individuals is declining each day with the introduction of many toxins, materials and chemicals in our everyday lives. It affects not only physical or psychological health, but also our lifestyle. The type and stage of skin cancer treatment depends on the size and location of the tumour, and the overall health and medical history of your patient. The purpose of treatment in most cases is to complete cancer removal or destruction. If detected and treated early, the majority of skin cancers can be cured. Treating skin cancer can harm healthy tissues and cells, with often unwanted side effects. Side effects primarily depend on the treatment type and scope. For each person, side effects cannot be the same. We proposed an efficient pre-processing approach for finding the most relevant and eradicating the features which are irrelevant from the dermatology dataset to increase the accuracy in the decision-making process. In this research, an optimization-based Feature Selection method is proposed using Information Gain, Feature Extraction techniques and Genetic Algorithm. The proposed feature selection methods performance is compared with the existing techniques through various classifiers.

**Keywords:** Dermatology, Pre-Processing, Data Mining, Machine Learning, Feature Selection, Classification

## 1. INTRODUCTION

In order to be more effective and productive, today's healthcare firms are migrating from volume-based to value-based business models, which mandates doctors and nurses working longer hours [1]. This will improve healthcare by modifying people's lifestyles, encouraging them to live longer lives, and preventing infections, illnesses and diseases. Healthcare data is more complex in the last few years as a result of the vast amount of data that has just become available, as well as the quick change of technology and mobile applications, and the discovery of new diseases.

Skin infections are a major health concern that affects a vast number of people around the world. In recent years, there is a rapid advancement of technology and the application of different data mining approaches, predictive dermatological classification is becoming more accurate and predictive. [2]. Finally, the advances in technological learning approaches are efficiently capable in the classification of skin disease.

## 2. IMPORTANCE OF DATA PRE-PROCESSING

In the majority of real-world data that is incomplete with aggregate and missing values, noisy data with inconsistencies, outliers, and duplication information, pre-processing eliminates errors and preserves the accuracy and accountability of missing values. The quality of mining is determined by the quality of data used to make decisions. As a consequence, by normalising the attributes in the given dataset, this contribution addresses the value of the pre-processing mission. Pattern recognition, machine learning, and statistics all use feature selection as part of their processes. The major benefits of feature selection methods are: Improving classification efficiency and a deeper understanding of the data in different mining applications by reducing computation time.

### Feature Selection Technique

It is a technique for reducing the dimensionality of a dataset by picking the available a subset characteristic [6] [7]. In FS, features that are redundant (have the same value) or useless (contain unwanted information) are removed. FS is a powerful machine learning technique that aids in the development of a classification system. With a smaller function subset, the time complexity of a classifier is decreased, and the accuracy of the classifier increases [8]. Feature selection can be done in one of three ways: embedded, filter, or wrapper. In the embedded technique, FS emerges as a data mining algorithm. The wrapper approach chooses features for the classifier deliberately, whereas the filter technique includes features irrespective of the classifier. The filter method selects features using some statistical method, while the wrapper seeks the best features of subset using learning algorithm. The wrapper method is more computationally costly and slower than the filter approach, but it yields more reliable performance.

The evaluation functions and the search algorithm are the two key components of the FS algorithm. The evaluation function describes which selection method is used. Search methods may be categorised as exponential, sequential, or randomised, according to their results. The exponent method is exponentially complex, the randomised method selects features at random, resulting in high accuracy, and the sequential method adds or subtracts features in a linear manner.

## 3. RELATED WORKS ON FEATURE SELECTION TECHNIQUES

Surjeet Kumar, Anurag Kumar, Verma and Saurabh Pal [3] developed an advanced model for predicting several kinds of skin illness that employs six different data mining classification approaches along with ensemble utilizing AdaBoost, gradient and bagging classifiers. The feature value method is frequently used to choose essential factors that influence prediction. A subset of the original dataset is created after only 15 features from the original dataset are selected to examine the effects of six machine learning approaches and an ensemble method on the complete dataset. A new subset of the original dataset is produced using the feature selection method, and it is analysed to the prediction model on the skin disease dataset.

Kasturi DewiVarathan, Amin, Mohammad Shafenoor, Yin Kia Chiam, and Mohammad Shafenoor [4] aimed to identify key features and data mining techniques that could improve cardiovascular disease prediction accuracy. To develop prediction models, various sets of features and seven categorization techniques were used: Neural Network, Naive Bayes, Support Vector Machine (SVM), k-NN, Logistic Regression (LR), Decision Tree and Vote (a hybrid technique combining Nave Bayes and Logistic Regression).

Bera, Tanmoy, et al [5] discussed about several image processing and data mining techniques in identifying multiple rice plant diseases. The image processing mechanism is used to remove relevant disease-related features from the disease portion of the image. Data mining methods are used to retrieve relevant secret information required for detecting the disease using the extracted features. Farmers can use the strategies to rescue their crops from illness by taking corrective action in a short period of time. The authors looked at a number of articles on rice plant disease detection in order to assess the existing state of the agricultural decision support system.

Selvakumar, B., and Karuppiah Muneeswaran [6] In order to reduce the amount of time spent identifying anomalies from high-dimensional network traffic features, dimensionality reduction is essential in IDS. The proposed work uses a filter and wrapper-based approach with the firefly algorithm in the wrapper to pick the features, which affects the speed of the analysis. With the KDD CUP 99 dataset, a C4.5 and Bayesian Networks (BN) based classifier is used to classify the generated features.

Mengjie Zhang, Bing Xue and Nguyen [7] presented a systematic survey of current work using swarm intelligence for achieving classification and feature selection, with an emphasis on representation and search methods. The aim is to provide a broad overview of various types of state-of-the-art techniques, as well as their benefits and drawbacks, to inspire researchers to look at more advanced methods, to give practitioners advice on how to select the best methods to use in real-world situations, and to address possible shortcomings and concerns for future study.

Liao, Xiufeng, et al [8] An optimization technique was used to enhance the recognition of drilling parameters and improve them. 618 data sets were calculated and collected for this purpose, including RPM, strength parameters, compressive and flushing media. After an initial assessment, the concrete strength function of samples, which is a significant property of rocks, was used as an appropriate criterion for categorization. Then, with the help of intelligent algorithms, three various levels of rock strength, as well as all of the specifics, were modelled. The results showed that systems categorised based on compressive strength performed much better for ROP evaluation because to the proximity of features. As a result, these three levels were used to classify the data. This problem was solved using a modern artificial bee colony algorithm. Under various optimization conditions, optimizations were applied to the selected models, and optimal states were calculated. Because determining drilling machine characteristics is critical, these values were computed based on ideal conditions.

Namous, Feras, and others [9] A metaheuristic-based technique for feature selection was examined in binary classification issues. Multiple significantly skewed datasets are involved in the problem. The standard fitness function with suitable classification accuracy is replaced with two more ideal fitness functions: the area under the ROC curve and the geometric mean, to deal with the problem of imbalanced data. Two common metaheuristic approaches are examined utilising the three fitness functions for classifying six imbalanced datasets to establish the effectiveness of the suggested approach.

Ibrahim, Rehab Ali, et al [10] Slap Swarm Algorithm (SSA) is a combined hybrid optimization with particle swarm optimization method for handling FS problem. Slap Swarm Algorithm – Particle Swarm Optimization (SSAPSO) is a hybrid algorithm that increases the effectiveness of the exploration and exploitation phases by integrating both approaches. To test the proposed algorithms efficiency, two experimental series are used. Benchmark functions are utilized in the first to compare it to other related techniques. Meanwhile, the SSAPSO is used in the second experiment sets to evaluate the best feature set using various UCI datasets. When redundant or misleading features are removed from the original dataset, the accuracy is maintained or improved.

Bashir, Saba, et al [11] The emphasis is on feature selection methods, algorithms, and numerous heart disease datasets are used for research analysis for demonstrating the process of accuracy. Decision Tree, Support Vector Machine (SVM), Logistic Regression, Random Forest and Nave Bayes are used as feature selection methods with the Rapid miner as a guide, and progress is seen in the results by showing the accuracy.

Nalić, Jasmina, Goran Martinović, and Drago Žagar [12] A new hybrid data mining model combining ensemble classification and features selection approaches was developed to aid decision-making. The model is constructed in stages. Initially, dataset is pre-processed in the first step, and the authors paid special attention to feature selection in addition to using various pre-processing techniques. In this approach about 5 various feature selection algorithms are used and the obtained result is combined based on the different voting types using ROC and accuracy measurements from the logistic regression algorithm. It has been also introduced if any, a new voting mechanism is more efficient than other voting methods as well as the outcome of a single feature selection algorithm. On the dataset obtained during the feature selection process, four different classification algorithms were performed, including generalised linear model, support vector machine, naive Bayes and decision tree model.

Mafarja, Majdi, et al [13] proposed a Grasshopper Optimisation Algorithm (GOA) with wrapper-based system for choosing the optimal function subset for executing the classification. A binary GOA is constructed using two mechanisms: initially with V-shaped transfer and Sigmoid functions represented by BGOA-V and BGOA-S. Secondly, a novel strategy that incorporates the best solutions found thus far. Additionally, the BGOA algorithm incorporates a mutation operator to enhance the exploration process.

Abhishek, Kuntal Mukherjee, Bhattacharya and Radha Tamal Goswami [14] proposed for feature selection in permission-based malware detection on Android. The fundamental contribution of this research is to present a novel random key encoding approach (PSORS-FS) for converting the classical Optimization method to the discrete domain, which is used in the suggested work (PSORS-FS). It also removes problems with particle maximum velocity and the sigmoid function, all of which are correlated with binary PSO. PSORS-FS guarantees that the search process is diverse, as well as reducing the chance of premature convergence.

Taradeh, Mohammad, et al [15] proposed a new GSA-based algorithm that includes evolutionary crossover and mutation operators. As an NP-hard problem, FS seeks out the best subset of characteristics from a given set. Both Decision Tree (DT) and K-Nearest Neighbours (KNN) classifiers are used as evaluators in the wrapper FS procedure in the proposed study. Eighteen well-known UCI datasets are used to evaluate the output of the proposed techniques. The proposed algorithms' findings are compared to the efficiency of certain common nature-inspired algorithms (e.g., Particle Swarm Optimizer (PSO), Genetic Algorithm (GA), and Grey Wolf Optimizer (GWO)).

Sankhwar, Shweta, et al[16] Improved Grey Wolf Optimization (IGWO) and Fuzzy Neural Classifier(FNC) were combined to create a novel predictive system for the FCP model. The convergence of the GWO algorithm and the tumbling effect yields an IGWO algorithm. The IGWO-based FS method presented here is used to extract the best features from financial data. For classification purposes, FNC is employed. The proposed methodology is validated on two benchmark data sets, Australian Credit and German Credit, using a number of performance metrics.

## 4. PROPOSED OPTIMIZATION BASED FEATURE SELECTION TECHNIQUE

### Genetic Algorithm

Genetic Algorithms (GA) are a form of natural selection-based adaptive heuristic search technique [17]. It is based on Darwin's theory of evolution, which employs "survival of the fittest," one of the randomised search tactics. The algorithm starts with a population, which is a set of individuals (chromosomes). A chromosome is a collection of genes that can take the form of bits, integers, or characters. Individuals are picked depending on their ability to reproduce. The higher individual fitness ratings, is more probable and they will be selected [18].

A new population is created by cross-pollination and mutation. Crossover speeds up the quest early in a population's evolution, while mutation restores missing knowledge to the population through global or local movement in the search room. The procedure is replicated iteratively until the stopping conditions are met or the best solution is found [19].

### Pseudo Code for GA:

***Step 1***: Create the population P by selecting individuals at random from the search space S.

***Step 2:*** Evaluate each individual's fitness f(xi) in P.

***Step 3:*** Repeat (until stopping condition satisfied)

- Selection – according to the fitness value individuals are selected.
- Crossover – according to predetermined crossover probability, crossover the selected individuals.
- Mutation – according to mutation probability, newly generated in individuals are mutated $P_{new}$
- Update - $P \leftarrow P_{new}$
- Evaluate – compute the fitness $f(x_i)$ of each individual in P.

***Step 4:*** Return the most fitted individuals from P.

### Information Gain based Feature Selection technique

Information Gain determines the entropy value for what function? (i.e. total information it will delivers). A random variable's entropy is a measure of its uncertainty. We may decide the most useful function for classification using this value. The entropy value is higher, then more knowledge the function contains [20]. The entropy value decreases as the data becomes smaller. If the target attribute has k different values, the feature's entropy in relation to this k-wise classification is described as:

$$Entropy\ (S) = \sum_{c=1}^{k} -[p_c * (log_2 * p_c)]$$

 

Where $p_c$ denotes the percentage of S that relates to class c. Because encoding length is defined in bits, the base of the logarithm is 2.

$$IG(S,a) = E(S) - E(S|a)$$

A variable value is represented by the letter a. The Information Gain gained by training example S from the observation that a random variable V has a value is calculated using the preceding equation.

**Feature Extraction Techniques**

Prior to data classification, three feature extraction techniques, PCA [21], LDA [12], and LSR [23], are used to extract data features. In general, feature extraction is done with the goal of minimizing the resources required to denote a large amount of data.

**Principle Component Analysis**

It is an unsupervised learning method for reducing the dimensionality of a data set while preserving the original variability of the data. The orthogonal transformation is used in PCA to obtain linearly uncorrelated variables, which are referred to as theory components from the correlated variables. The number of principal components is equal to or less than the number of initial variables. The statistical analysis of a matrix, such as mean, Standard Deviation (SD), Covariance, Eigen values, and Eigen vectors, are evaluated in PCA.

**Mean:** It states the average values of variables throughout the distribution. This measure is also referred to as central tendency. The random variable mean value is represented in Eq. (1) in which $F_k = F_1, F_2, \dots, F_n$ represents the random variables. The size for the random variables is specified as k.

$$Mean\ (\bar{F}) = \frac{1}{l}\sum_{k=1}^{l} F_k \qquad (1)$$

**Standard Deviation:**It's used to find out how much scatter there is. The average distance between the mean and the point at which the data is set is calculated by squaring them in order to measure the spread out of the data.The mathematical equation for SD is represented in Eq. (2) in which the mean is denoted as $\bar{F}$.

$$SD = \sqrt{\frac{1}{l}\sum_{k=1}^{l}(F_k - \bar{F})^2} \qquad (2)$$

**Covariance:** It is measured between 2 dimensions. This measurement predicts the quantity of the variations in dimension from the mean. The covariance is calculated using equation (3)

$$Cov(F,G) = \frac{\sum_{k=1}^{l}(F_k - \bar{F})(G_k - \bar{G})}{l} \qquad (3)$$

Eigen values & Eigen vectors of a matrix: The rectangular array of numbers, symbols, or expressions is termed as a matrix, and each and every individual item belonging to the matrix is called as elements. The term B is a $n$ X $n$ matrix and eigen value of B is represented in Eq. (4). Moreover, $\lambda$ is indicated as a scalar parameter. For attaining the distinct eigen values, the eigen vector of a symmetric matrix is orthogonal and it is symmetric for real values.

$$[B][F] = \lambda[F] (4)$$

The extracted features obtained from PCA model is specified as $F_i^{PCA}$ and it is represented in Eq. (5) in which the count of the feature PCA is specified as NP.

$$F_i^{PCA} = \{F_1^{PCA}, F_2^{PCA}, \dots, F_{NP}^{PCA}\} \qquad (5)$$

**Linear Discriminant Analysis**

It's a modern feature extraction and dimension reduction technique that's been used in speech recognition, face recognition, multimedia information retrieval, and other applications. LDA's main goal is to predict the best transformation from high-dimensional data that has been divided into groups. In order to solve the issues related to the optimal discrimination projection matrix, the within-class scatter matrix and between-class scatter matrix are projected.The mathematical equation to find optimal discrimination projection matrix is shown in Eq. (6) in which $B_{class}\ and\ W_{class}$ represents the between class scatter matrix and within-class scatter matrix, respectively. The formulation for calculating$B_{class}\ and\ W_{class}$ is shown in the equation (7) and equation (8), respectively. Further, the eigenvectors of the projection matrix S are shown in Eq. (9) in which $D_r =$

 

$B_{class} + W_{class}$, $T_j$ is the feature vector of the data, $\alpha_{NJ}$ and NJ are the data vector and samples in the data class J.

$$S_{odp} = \arg\max_{S} \frac{S^T B_{class} S}{S^T W_{class} S} \quad (6)$$

$$W_{class} = \sum_{j=1}^{H} (T_j - \mu_{NJ})(T_j - \mu_{NJ})^T \quad (7)$$

$$B_{class} = \sum_{j=1}^{H} Q_J (\mu_J - \alpha)(\mu_J - \alpha)^T \quad (8)$$

$$S = eig(D_r^{-1} B_{class}) \quad (9)$$

Further, $F_i^{LDA}$ is the feature extracted from LDA model and it is represented in Eq. (10), in which ND specifies the count of the features of LDA.

$$F_i^{LDA} = \{F_1^{LDA}, F_2^{LDA}, \ldots, F_{ND}^{LDA}\} \quad (10)$$

**Least Square Regression**

LSR is a supervised dimensionality reduction technique that has been revised. The LSR is used to derive information from data in the majority of cases.The optimization problem of LSR is expressed in Eq. (11) in which the corresponding label of the data is represented as $L_e$ and the class indicator matrix is $Y_o = \{y_1, y_2, \ldots, y_n\}$. The matrix with kth columns having the dimensionality vector as $d^* + 1$ is shown in Eq (12). The optimal transformation matrix is $Z^*$. Furthermore, $(Z^*, Z^{T^*}) \pm$ is pseudo-inverse of $Z^*, Z^{T^*}$

$$I(V^*) = \min_{V^*} \frac{1}{2} \left\| V^{T^*} Z^* - Y_0 \right\| (11)$$

$$V_{ls}^* = (V^* Z^{T^*}) Z^* Y_0^T \quad (12)$$

The extracted features obtained from LSR model is shown in Eq. (13) in which the count of the LSR feature are represented as NS.

$$F_i^{LSR} = \{F_1^{LSR}, F_2^{LSR}, \ldots, F_{NS}^{LSR}\} \quad (13)$$

Further, the extracted features got from all the three models (PCA, LDP, and LSR) are represented in Eq. (14). The combined form of the extracted features $F_j$ is modified as per eq (15):

$$F_i = F_i^{PCA} + F_i^{LDR} + F_i^{LSR} \quad (14)$$

$$F_i = \{F_1, F_2, \ldots, F_n\} \quad (15)$$

The combined features $F_i$ is subjected to hybrid algorithm, which further provides the optimal features$F_i^*$. For data categorization, next to the optimal feature's selection it is provided as an input to the Neural Network classifier.

**Objective Function**

The primary goal of this work is to minimize the association between data features when choosing the best features. The objective function's mathematical equation is shown in Eq (16). Eq (17) expresses the association between two data features $F_1$ and $F_2$where n denotes the number of data features.

$$N = \min[Correlation] \quad (16)$$

$$Correlation = \frac{n\sum F_1F_2 - \sum F_1 \sum F_2}{(n\sum F_1^2 - (\sum F_1)^2) - n\sum F_2^2 - (\sum F_2)^2} (17)$$

**Feature Encoding**

The solution to the proposed Optimization based feature selection algorithm for selecting the optimal features is given as the combined features after combining the PCA, LDA, and LSR, which results in Eq. (15).According, the extracted data features $F_i$ where $F_i = \{F_1, F_2, \ldots, F_n\}$ are selected optimally by the proposed Optimization based Feature Selection algorithm. Then, the selected feature vector $F_i^* = F_1^*, F_2^*, \ldots, F_n^*$is fed as input to classifier, and the classified data is obtained in terms of some labels.

**Proposed Feature Selection Technique using Feature Extraction, GA and IG**

GA is an adaptive mutation strategy that uses a heuristic search and is inspired by the genetics evolution process. A population of competing solutions is maintained, which evolves and converges to the best solution through selection, crossover, and mutation. The solution space is searched in parallel to find the best solution without getting trapped in a local optimum. Because of its robustness to the underlying search space size and multivariate distributions, GA can generate promising feature selection methods in a high-dimensional space.

In general, a search engine with an initial state, a state space, and a termination condition are required for searching for an optimal solution in the entire feature space.Given $n$ number of features, the size of search space is $2^n$-1.As every feature hastwo possible states: "1" or "0", an $n$ bit string will have $2^n$possible combinations. Assume τ features, which are not important to decision making in terms of the values oftheir SU, be removed. The length of a binary string becomes $n$-$\tau$. Even inthe reduced search space ($2n$-τ), a brute-force search for a large space of $2^n$-τ is stillin feasible. Of course, such a reduction in space is beneficial to the GA Wrapper quest. The proposed Cultural Algorithm dependent Feature Selection Method has the following steps.

***Stage 1: Initialization of Population:*** GA maintains a diverse population$x_{1 \ldots n} = \langle x_1 \ldots x_n \rangle$of $n$ individuals $x_i$, the candidate solutions. The fitness of these individuals is evaluated by calculating anobjective function$F(x_i)$ that is to be optimized for a given problem. Individual solutions are described as chromosomes, which encompass the entire range of options.

***Stage 2: Selection Operator:*** The process of assessing an individual's fitness and choosing them for reproduction is known as selection. Selecting can be done in a variety of ways. Rank Selection, Hierarchical Selection, Roulette-Wheel Selection, Tournament Selection and Elitist Selection are some of the most widely used techniques. Tournament selection was used in this study to find individuals who were strong enough to mate.

***Stage 3: Crossover Operator:*** By swapping sections of the genomes of the two parent chromosomes, the crossover operator produces two offspring. The process of removing the best genes from parents and reconstructing them into potentially superior children is known as crossover. Single-point crossover is the most basic form of crossover. Two-Point Crossover and Uniform Crossover are two other varieties. Single point crossover was used in this project.

***Stage 4: Mutation Operator:*** Mutation preserves population genetic diversity from generation to generation, chromosome to chromosome, and increases the algorithm's chances of producing more fit individuals. At each location in the string, a character is modified at random using a small mutation probability. The bits in bit strings are flipped at random locations with a limited probability when they are mutated. Uniform mutation was used in this study.

***Stage 5: Termination Condition:*** Three possible termination conditions for the GA includes achieving a satisfactory solution, reaching a predetermined extreme number of generations, and converting the population to a particular degree of genetic variety. The mutation probability affects the algorithm's convergence: a maximum mutation rate prevents the search from converging, while a low rate causes the search to converge prematurely. Maximum count of generations = 20 to 50 is the work's termination criteria.

**Step by Step Procedure for Genetic Algorithm based Feature Selection Method**

 

***Step 1:*** Measure Information Gain of individual features from the dataset.

***Step 2:*** Rank the features in the dataset according to their importance: $F = (f_1 > f_2 > f_3 \ldots.)$ and applied Feature Extraction techniques.

***Input:*** Top m features set $f_r$ and class label C.

***Output:*** S

***Step 3:*** S← null

***Step 4:*** Procedure GA

***Input:*** PopSize Ps, GenSize, GenomeLength N, ProbMutation Pm

***Output:*** The best individuals in all generations.

***Step 4.1:*** Population Initialization: Ps *N.

***Step 4.2:*** Retain $f_1$ from $f_r$

***Step 4.3:*** Ps← Random binary chromosome

***Step 4.4:*** for each chromosome do

***Step 4.5:*** compute fitness according to classifier ANN.

***Step 4.6:*** end for

***Step 4.7:*** repeat

***Step 4.8:*** Select parents p1,p2 from population based on the fitness

***Step 4.9:*** for all new children do

***Step 4.10:*** retain $f_1$ from $f_r$

***Step 4.11:*** Crossover p1,p2;

***Step 4.12:*** Mutate each gene in new child chromosome with probability Pm;

***Step 4.13:*** end for

***Step 4.14:*** Evaluate fitness of new individuals according to the classifier.

***Step 4.15:*** Replace least-fit population with new best individuals

***Step 4.16:*** Stopping criteria

***Step 5:*** End Procedure

## 5. RESULT AND DISCUSSION

### Description of the Dataset

In this research work, the dermatology dataset is considered from UCI repository [24]. This dataset contains 35 characteristics, with the family history feature having a value of 1indicates these diseases has been detected in the family, and if not detected then its 0. The patient's age is represented by the age element. Every other characteristic (clinical and histopathological) is graded on a scale of 0 to 3. The absence of the feature is represented as 0, maximum feasible quantity is represented as 3 and relative intermediate levels are represented as 1,2. The class labels values are indicated as 1 – psoriasis, 2- seboreic dermatitis, 3 - lichen planus, 4 - pityriasis rosea, 5 - cronic dermatitis, 6 - pityriasis rubra pilaris. Table 1 depicts the dermatology dataset used in this research work.

*Table 1: Dermatology Dataset used in this research work*

| Sl.No | Feature Name | Values |
|---|---|---|

| | | |
|---|---|---|
| 1 | Erythema | 0,1,2,3 |
| 2 | Scaling | 0,1,2,3 |
| 3 | definite borders | 0,1,2,3 |
| 4 | Itching | 0,1,2,3 |
| 5 | koebner phenomenon | 0,1,2,3 |
| 6 | polygonal papules | 0,1,2,3 |
| 7 | follicular papules | 0,1,2,3 |
| 8 | oral mucosal involvement | 0,1,2,3 |
| 9 | knee and elbow involvement | 0,1,2,3 |
| 10 | scalp involvement | 0,1,2,3 |
| 11 | family history, (0 or 1) | 0 or 1 |
| 12 | melanin incontinence | 0,1,2,3 |
| 13 | eosinophils in the infiltrate | 0,1,2,3 |
| 14 | PNL infiltrate | 0,1,2,3 |
| 15 | fibrosis of the papillary dermis | 0,1,2,3 |
| 16 | Exocytosis | 0,1,2,3 |
| 17 | Acanthosis | 0,1,2,3 |
| 18 | Hyperkeratosis | 0,1,2,3 |
| 19 | Parakeratosis | 0,1,2,3 |
| 20 | clubbing of the rete ridges | 0,1,2,3 |
| 21 | elongation of the rete ridges | 0,1,2,3 |
| 22 | thinning of the suprapapillary epidermis | 0,1,2,3 |
| 23 | spongiform pustule | 0,1,2,3 |
| 24 | munromicroabcess | 0,1,2,3 |
| 25 | focal hypergranulosis | 0,1,2,3 |
| 26 | disappearance of the granular layer | 0,1,2,3 |
| 27 | vacuolisation and damage of basal layer | 0,1,2,3 |
| 28 | Spongiosis | 0,1,2,3 |
| 29 | saw-tooth appearance of retes | 0,1,2,3 |
| 30 | follicular horn plug | 0,1,2,3 |
| 31 | perifollicular parakeratosis | 0,1,2,3 |
| 32 | inflammatory monoluclearinfiltrate | 0,1,2,3 |
| 33 | band-like infiltrate | 0,1,2,3 |
| 34 | Age | Linear |
| 35 | Class | 1,2,3,4,5,6 |

**Performance Metrics**

Table 2 depicts the metrics used in this research work to evaluate the proposed optimization-based feature selection method.

**Table 2: Performance Metrics**

| Metrics | Equation |
|---|---|
| Accuracy | $\frac{TP+TN}{TP+FN+TN+FP}$ |
| True Positive Rate (TPR) (Sensitivity or Recall) | $\frac{TP}{TP+FN}$ |
| False Positive Rate (FPR) | $\frac{FP}{FP+TN}$ |
| Precision | $\frac{TP}{TP+FP}$ |
| True Negative Rate (Specificity) | 1- False Positive Rate (FPR) |
| Miss Rate | 1-True Positive Rate (TPR) |
| False Discovery Rate | 1- Precision |

**Number of Features obtained**

Table 3 depicts the features attained by existing feature selection techniques like Genetic Algorithm, Information Gain and proposed Optimization based Feature Selection method. Table 3 clearly shows the proposed Optimization based Feature Selection method gives less number of features when it is compared with existing feature selection techniques.

**Table 3: Features obtained by Existing Feature Selection techniques like GA, IG and proposed Feature Selection method**

| Sl.No | Genetic Algorithm | Information Gain | Proposed Feature Selection method |
|---|---|---|---|
| 1 | Clubbing of the rete ridges | Erythema | melanin incontinence |
| 2 | Elongation of the rete ridges | Scaling | Exocytosis |
| 3 | Thinning of the suprapapillary epidermis | definite borders | Acanthosis |
| 4 | Spongiform pustule | Itching | Hyperkeratosis |
| 5 | Munro microabcess | koebner phenomenon | Parakeratosis |
| 6 | Erythema | polygonal papules | Erythema |
| 7 | Scaling | follicular papules | Scaling |
| 8 | definite borders | oral mucosal involvement | Itching |
| 9 | Itching | knee and elbow involvement | koebner phenomenon |
| 10 | koebner phenomenon | scalp involvement | polygonal papules |
| 11 | family history, (0 or 1) | family history, (0 or 1) | follicular papules |
| 12 | melanin incontinence | melanin incontinence | eosinophils in the infiltrate |
| 13 | eosinophils in the infiltrate | eosinophils in the infiltrate | spongiform pustule |
| 14 | PNL infiltrate | PNL infiltrate | munromicroabcess |
| 15 | fibrosis of the papillary dermis | fibrosis of the papillary dermis | focal hypergranulosis |
| 16 | Exocytosis | Exocytosis | disappearance of the granular layer |
| 17 | Acanthosis | Acanthosis | vacuolisation and damage of basal layer |
| 18 | Hyperkeratosis | Hyperkeratosis | Spongiosis |
| 19 | Parakeratosis | Parakeratosis | perifollicular parakeratosis |
| 20 | disappearance of the granular layer | clubbing of the rete ridges | inflammatory monoluclearinfiltrate |
| 21 | vacuolisation and damage of basal layer | elongation of the rete ridges | band-like infiltrate |
| 22 | Spongiosis | thinning of the suprapapillary epidermis | follicular horn plug |
| 23 | saw-tooth appearance of retes | spongiform pustule | |
| 24 | follicular horn plug | munromicroabcess | |
| 25 | perifollicular parakeratosis | focal hypergranulosis | |
| 26 | inflammatory monoluclearinfiltrate | | |
| 27 | band-like infiltrate | | |

 

| | | | |
|---|---|---|---|
| 28 | Age | | |

**Performance of the Proposed Feature Selection Method**

The popular classification techniques used for measuring the proposed system performances are Support Vector Machine (SVM), Artificial Neural Network (ANN) and Naïve Bayes (NB) with existing feature selection techniques using the above-mentioned metrics.

Table 4 gives the classification accuracy (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using ANN, NB and SVM classification techniques. Figure 1 depicts the graphical representation of the classification accuracy (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using SVM, NB and ANN classification techniques. From the table 4 and figure 1, it is clear that the proposed FS method with ANN gives more accuracy than other classifiers.

**Table 4: Classification Accuracy (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques**

| Feature Selection Techniques | Accuracy (in %) by Classification techniques | | |
|---|---|---|---|
| | **SVM** | **NB** | **ANN** |
| **Original Dataset** | 45.67 | 48.23 | 53.76 |
| **GA** | 66.17 | 67.49 | 79.43 |
| **IG** | 57.42 | 58.39 | 60.42 |
| **Proposed FS Method** | 79.34 | 73.78 | **89.56** |

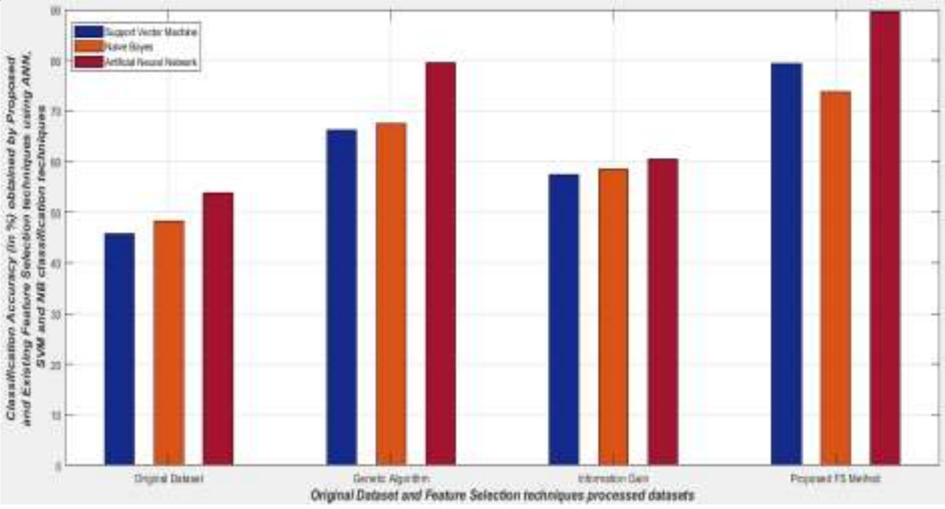


***Figure 1: Graphical representation of the Classification Accuracy (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques***

Table 5 gives the sensitivity (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using ANN, NB and SVM classification techniques. The Figure 2 illustrates the graphical representation of the sensitivity (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using SVM, NB and ANN classification techniques. Figure 2 and table 5 clearly shows the proposed FS method with ANN gives more sensitivity than other classifiers.

**Table 5: Sensitivity (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques**

| Feature Selection Techniques | Sensitivity (in %) by Classification techniques | | |
|---|---|---|---|
| | **SVM** | **NB** | **ANN** |
| **Original Dataset** | 42.26 | 45.85 | 49.81 |
| **GA** | 60.62 | 61.30 | 63.46 |
| **IG** | 51.84 | 52.53 | 54.39 |
| **Proposed FS Method** | 69.73 | 70.39 | 85.32 |

 

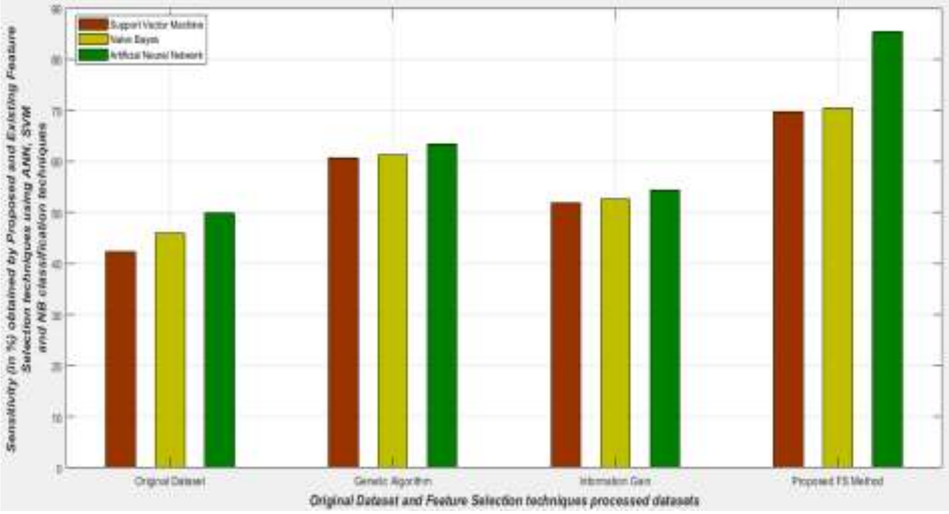


**Figure 2: Graphical representation of the Sensitivity (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques**

In table 6, the False Positive Rate (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using ANN, NB and SVM classification techniques. Figure 3 depicts the graphical representation of the False Positive Rate (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using SVM, NB and ANN classification techniques. Figure 3 and table 6 clearly shows the proposed FS method with ANN gives reduced FPR than other classifiers.

**Table 6: False Positive Rate (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques**

| Feature Selection Techniques | False Positive Rate (in %) by Classification techniques | | |
|---|---|---|---|
| | SVM | NB | ANN |
| **Original Dataset** | 70.35 | 68.78 | 65.51 |
| **GA** | 54.64 | 56.22 | 44.81 |
| **IG** | 62.73 | 61.15 | 54.59 |
| **Proposed FS Method** | 42.58 | 30.43 | 22.47 |

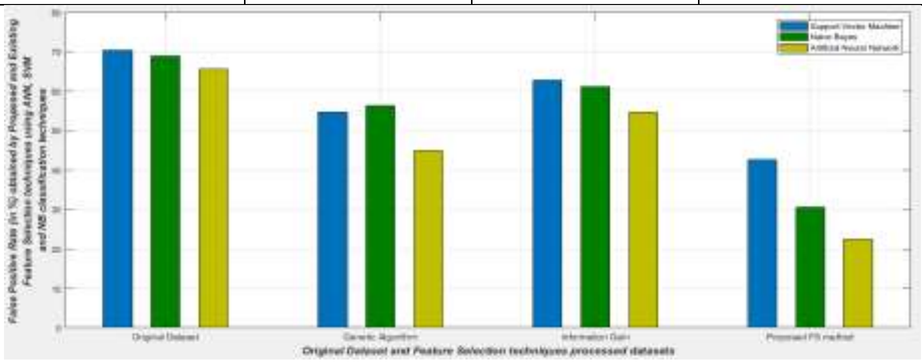


**Figure 3: Graphical representation of the False Positive Rate (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques**

Table 7 gives the Precision (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using ANN, NB and SVM classification techniques. Figure 4 depicts the graphical representation of the Precision (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using SVM, NB and ANN classification techniques. Figure 4 and table 7 clearly shows the proposed FS method with ANN gives reduced FPR than other classifiers.

**Table 7: Precision (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques**

| Feature Selection Techniques | Precision(in %) by Classification techniques | | |
|---|---|---|---|
| | SVM | NB | ANN |
| **Original Dataset** | 48.77 | 50.53 | 57.85 |
| **GA** | 74.26 | 75.73 | 78.8 |

| | | | |
|---|---|---|---|
| **IG** | 61.13 | 64.47 | 63.68 |
| **Proposed FS Method** | **72.57** | **75.42** | **85.72** |

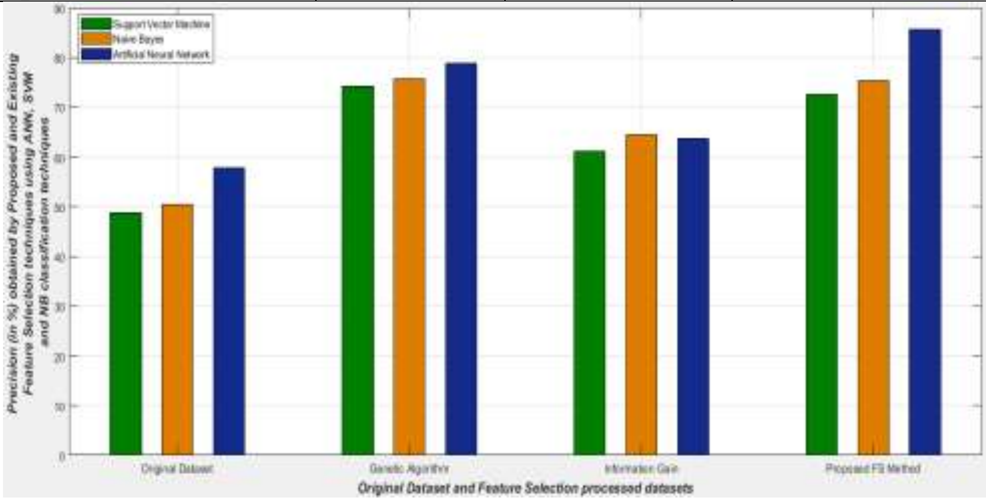


**Figure 4: Graphical representation of the Precision (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques**

Table 8 gives the Specificity (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using ANN, NB and SVM classification techniques. Figure 5 depicts the graphical representation of the Specificity (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using SWM, NB and ANN classification techniques. Figure 5 and table 8 clearly shows the proposed FS method with ANN gives reduced FPR than other classifiers.

***Table 8: Specificity (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques***

| **Feature Selection Techniques** | **Specificity (in %) by Classification techniques** | | |
|---|---|---|---|
| | **SVM** | **NB** | **ANN** |
| **Original Dataset** | 29.65 | 31.22 | 34.49 |
| **GA** | 45.36 | 43.78 | 55.19 |
| **IG** | 37.27 | 38.85 | 45.41 |
| **Proposed FS Method** | 57.42 | 69.57 | 77.53 |

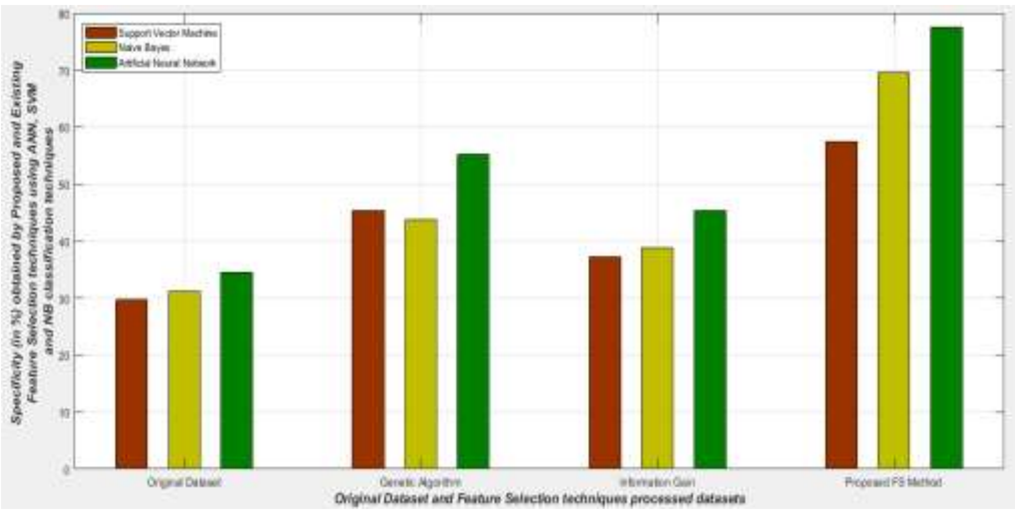


**Figure 5: Graphical representation of the Specificity (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques**

Table 9 gives the Miss Rate (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using ANN, NB and SVM classification techniques. Figure 6 depicts the graphical representation of the Miss Rate (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using SVM, NB and ANN classification techniques. Figure 6 and table 9 clearly shows the proposed FS method with ANN gives reduced FPR than other classifiers.

**Table 9: Miss Rate(in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques**

| Feature Selection Techniques | Miss Rate (in %) by Classification techniques | | |
|---|---|---|---|
| | SVM | NB | ANN |
| **Original Dataset** | 57.74 | 54.15 | 50.19 |
| **GA** | 39.38 | 38.7 | 36.54 |
| **IG** | 48.16 | 47.47 | 45.61 |
| **Proposed FS Method** | 30.27 | 29.61 | 14.68 |

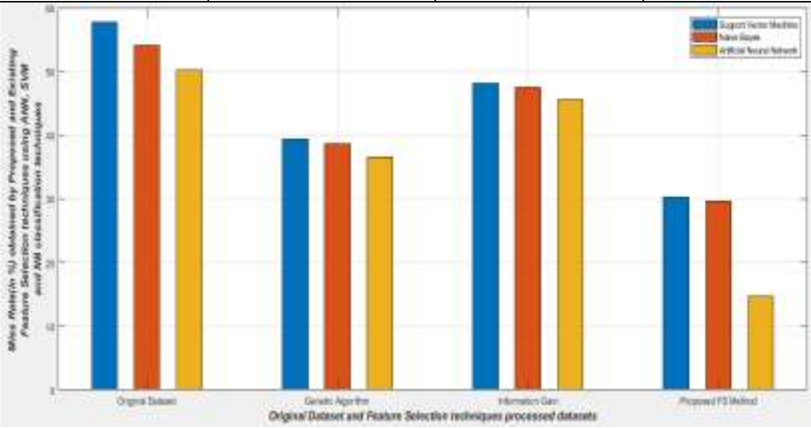

**Figure 6: Graphical representation of the Miss Rate (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques**

Table 10 gives the False Discovery Rate (in %) obtained by Proposed FS and existing IG, GA feature selection techniques using ANN, NB and SVM classification techniques. Figure 7 depicts the graphical representation of the False Discovery Rate(in %) obtained by Proposed FS and existing IG, GA feature selection techniques using SVM, NB and ANN classification techniques. Figure 7 and table 10 clearly shows the proposed FS method with ANN gives reduced FPR than other classifiers.

***Table 10: False Discovery Rate (in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques***

| Feature Selection Techniques | False Discovery Rate (in %) by Classification techniques | | |
|---|---|---|---|
| | SVM | NB | ANN |
| **Original Dataset** | 51.23 | 49.47 | 42.15 |
| **GA** | 25.74 | 24.27 | 21.2 |
| **IG** | 38.87 | 35.53 | 36.32 |
| **Proposed FS Method** | 27.43 | 24.58 | 14.28 |

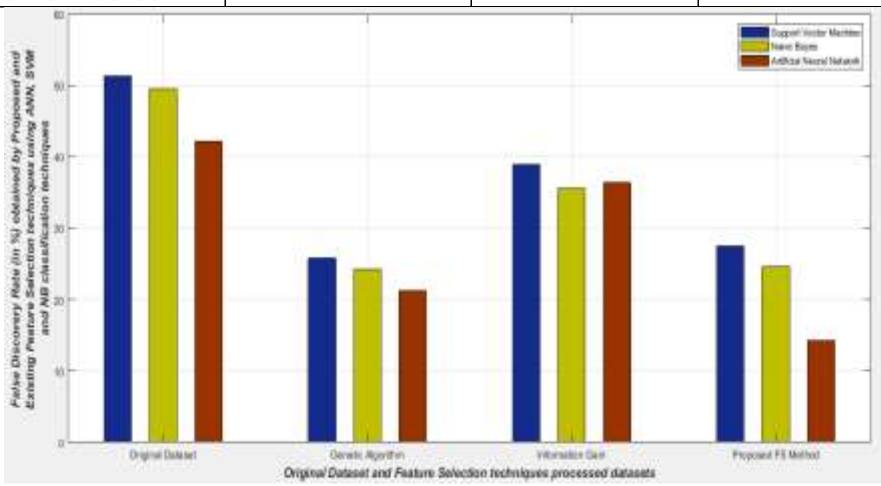

***Figure 7: Graphical representation of the False Discovery Rate(in %) obtained by Proposed and Existing Feature Selection techniques using ANN, SVM and NB classification techniques***

## 6. CONCLUSION

When there is just one feature, alternatives can be completed quickly, but when there are numerous metrics, selecting among many different ways might be complex. The feature selection algorithms and its strength are largely determined by various dataset. On one type of dataset, one technique may perform well, while on another type of dataset, it may underperform. Through this research work, an effective feature selection method is proposed using Genetic Algorithm optimization, Information Gain based feature selection and feature extraction techniques. From the results obtained, the proposed feature selection method is more sensitivity, accurate, specificity and precision with ANN classifier than other classifiers and it also reduced the error rates like FPR, Miss rate and FDR.

## REFERENCES


[1] Prasad, KalliSrinivasaNageswara, and Shameena Begum. "Enhancing Healthcare Systems Using Advanced Data Analytics Techniques."

[2] Nguyen, Sau Huu, et al. "Health-Related quality of life impairment among patients with different skin diseases in Vietnam: a cross-sectional study." International journal of environmental research and public health 16.3 (2019): 305.

[3] Verma, Anurag Kumar, Saurabh Pal, and Surjeet Kumar. "Prediction of skin disease using ensemble data mining techniques and feature selection method—a comparative study." Applied biochemistry and biotechnology 190.2 (2020): 341-359.

[4] Amin, Mohammad Shafenoor, Yin Kia Chiam, and Kasturi DewiVarathan. "Identification of significant features and data mining techniques in predicting heart disease." Telematics and Informatics 36 (2019): 82-93.

[5] Bera, Tanmoy, et al. "A survey on rice plant disease identification using image processing and data mining techniques." Emerging Technologies in Data Mining and Information Security. Springer, Singapore, 2019. 365-376.

[6] Selvakumar, B., and Karuppiah Muneeswaran. "Firefly algorithm based feature selection for network intrusion detection." Computers & Security 81 (2019): 148-155.

[7] Nguyen, Bach Hoai, Bing Xue, and Mengjie Zhang. "A survey on swarm intelligence approaches to feature selection in data mining." Swarm and Evolutionary Computation 54 (2020): 100663.

[8] Liao, Xiufeng, et al. "Effects of a proper feature selection on prediction and optimization of drilling rate using intelligent techniques." Engineering with Computers 36.2 (2020): 499-510.

[9] Namous, Feras, et al. "Evolutionary and swarm-based feature selection for imbalanced data classification." Evolutionary Machine Learning Techniques. Springer, Singapore, 2020. 231-250.

[10] Ibrahim, Rehab Ali, et al. "Improved salp swarm algorithm based on particle swarm optimization for feature selection." Journal of Ambient Intelligence and Humanized Computing 10.8 (2019): 3155-3169.

[11] Bashir, Saba, et al. "Improving heart disease prediction using feature selection approaches." 2019 16th international bhurban conference on applied sciences and technology (IBCAST). IEEE, 2019.

[12] Nalić, Jasmina, Goran Martinović, and Drago Žagar. "New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers." Advanced Engineering Informatics 45 (2020): 101130.

[13] Mafarja, Majdi, et al. "Binary grasshopper optimisation algorithm approaches for feature selection problems." Expert Systems with Applications 117 (2019): 267-286.

[14] Bhattacharya, Abhishek, Radha Tamal Goswami, and Kuntal Mukherjee. "A feature selection technique based on rough set and improvised PSO algorithm (PSORS-FS) for permission based detection of Android malwares." International journal of machine learning and cybernetics 10.7 (2019): 1893-1907.

[15] Taradeh, Mohammad, et al. "An evolutionary gravitational search-based feature selection." Information Sciences 497 (2019): 219-239.

[16] Sankhwar, Shweta, et al. "Improved grey wolf optimization-based feature subset selection with fuzzy neural classifier for financial crisis prediction." Soft Computing 24.1 (2020): 101-110.


 

[17] Khan, Muhammad Attique, et al. "An optimized method for segmentation and classification of apple diseases based on strong correlation and genetic algorithm based feature selection." IEEE Access 7 (2019): 46261-46277.
[18] Al Imran, Abdullah, Md Rifatul Islam Rifat, and Rafeed Mohammad. "Enhancing the classification performance of lower back pain symptoms using genetic algorithm-based feature selection." Proceedings of International Joint Conference on Computational Intelligence. Springer, Singapore, 2020.
[19] Farissi, Al, and Halina Mohamed Dahlan. "Genetic Algorithm Based Feature Selection for Predicting Student's Academic Performance." International Conference of Reliable Information and Communication Technology. Springer, Cham, 2019.
[20] Li, Mengmeng, et al. "Fast hybrid dimensionality reduction method for classification based on feature selection and grouped feature extraction." Expert Systems with Applications 150 (2020): 113277.
[21] Adegbola, Oluwole A., et al. "A principal component analysis-based feature dimensionality reduction scheme for content-based image retrieval system." Telkomnika 18.4 (2020): 1892-1896.
[22] Yang, Libo, et al. "Robust and efficient linear discriminant analysis with l 2, 1-norm for feature selection." IEEE Access 8 (2020): 44100-44110.
[23] Sefidian, Amir Masoud, and NeginDaneshpour. "Missing value imputation using a novel grey based fuzzy c-means, mutual information based feature selection, and regression model." Expert Systems with Applications 115 (2019): 68-94.
[24] https://archive.ics.uci.edu/ml/datasets/dermatology

 