International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
IJCAI 2021; 2(2): 06-11
Received: 04-05-2021
Accepted: 06-06-2021

Afrid V
Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

Boyella Mala Konda Reddy
Assistant Professor, Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

Corresponding Author:
Afrid V
Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

An efficient feature reduction approach for dermatology disease detection utilizing neural network approach

Afrid V and Boyella Mala Konda Reddy

DOI: https://doi.org/10.33545/27076571.2021.v2.i2a.31

Abstract
Skin infections are a significant worldwide medical issue related with high number of individuals. With the quick advancement of advances and the use of different AI strategies lately, the advancement of dermatological prescient grouping has gotten increasingly prescient and precise. Hence, improvement of AI strategies, which can adequately separate skin infection order, is critical. This paper manages the development and preparing of a fake neural organization for skin infection conclusion (SDD) in view of patients' side effects and causative living beings. This investigation is led to group the sort of skin illness in six distinct classes like incorporate psoriasis, seborrheic dermatitis, lichen planus, pityriasis rosea, constant dermatitis, and pityriasis rubra. The examination is done on the dataset taken from the University of California at Irvine Machine Learning Data Repository. The dataset contains an enormous volume of highlight measurements which are decreased utilizing SVM-RFE based element determination procedure. The dataset contains an enormous list of capabilities which is decreased utilizing an improved component choice strategy named as covering technique. The proposed covering technique is based on a MLP-RFE calculation to choose the main features from the given dataset. The chose subset of features then, at that point goes through a preprocessing step to present a consistency in the appropriation of information. Since MLP is perceived to have the advantage of giving an eminent execution in characterization stage.

Keywords: ANN, MLP, SVM-RFE, feature selection and dermatology

1. Introduction
The skin is the main piece of human body. The skin shields the body from UV radiation diseases, wounds, heat and destructive radiation, and furthermore helps in the production of nutrient D. The skin assumes a significant part in controlling internal heat level, so keep up great wellbeing and shield the body from skin infections [1]. The part of medication that is managing the skin and its illnesses is the dermatology. It intends to clarify, to analyze, to treat infectious and non-infectious skin sicknesses and to mind the patients with these illnesses. The skin sicknesses have different causes: parasitic contaminations, bacterial diseases, unfavorably susceptible responses, and even bug chomps. They can likewise happen as a result of different illnesses (of the skin or not) or in light of the climate. Once in a while the hereditary angles have a significant impact in the event of a skin illness [2].
The quick advancement of PC innovation in present many years, the utilization of AI innovation assumes a urgent part in the investigation of skin infections. This paper centers on the uses of fake neural organizations in skin infection. The clinical informatics field developed around the construction, the putting away and the handling of clinical data for different purposes.
One of these designs is to foster calculations ready to make expectations in regards to the conclusion, the therapy or the clinical development of a patient. In this paper, we have proposed a productive framework, which can group the sort of skin illness in six unique classes.

1.1 The objectives of this paper are
- To study various features of dermatology dataset
- To apply the SVM Recursive Feature Elimination (SVM-RFE) for reducing the number of attributes
- To classify attacks using Multilayer Perceptron (MLP)

2. Feature Selection

Feature determination strategies diminish the dimensionality of highlight space, eliminate repetitive, unimportant or loud information. It brings the prompt impacts for application: accelerating an information mining calculation, improving the information quality and the exhibition of information mining and expanding the understandability of the mining results [6]. Feature determination has been a functioning exploration region in information mining networks since it permits essentially improving the understandability of the subsequent classifier models [7]. It comprises to pick a subset of info factors from a dataset with exceptionally huge of properties by taking out highlights with almost no prescient data.

Feature choice procedures have their most extreme importance in information mining, AI, and example acknowledgment, particularly for enormous datasets [3]. The fundamental point of these procedures is to eliminate superfluous or repetitive highlights from the dataset. Highlight determination strategies have two classes: covering and channel [8, 10]. The covering assesses and chooses credits dependent on exactness gauges by the objective learning calculation. Utilizing a specific learning calculation, covering fundamentally look through the component space by excluding a few features and testing the effect of feature exclusion on the forecast measurements [4]. The element that has huge effect in learning measure infers it does matter and ought to be considered as an excellent element.

2.1 SVM Recursive Feature Elimination (SVM-RFE)

SVM-RFE highlight determination technique was proposed in [5] to lead quality choice for disease characterization. Settled subsets of highlights are chosen in a consecutive in reverse end way, what begins with all the component factors and eliminates each element variable in turn. At each progression, the coefficients of the weight vector of a straight SVM are utilized to process the component positioning score. SVM-RFE technique positions every one of the highlights as per some score work and kills at least one highlights with the most minimal scores. This cycle is rehashed until the most elevated characterization precision is acquired. Because of its effectively use in choosing useful qualities for malignant growth order, SVM-RFE acquired an extraordinary ubiquity and is notable as quite possibly the best element determination strategy [9]. Notwithstanding, the SVM-RFE is a voracious technique that lone desires to track down the most ideal mix for characterization.

2.1.1 The SVM -RFE calculation [5] can be broken into four stages

1. Train a MLP on the preparation set;
2. Request highlights utilizing the loads of the subsequent classifier;
3. Wipe out highlights with the littlest weight;
4. Rehash the cycle with the preparation set limited to the excess highlights.

3. Proposed Methodology

The framework introduced here was utilized a multi-facet feed-forward counterfeit neural organization was picked for this framework; it was prepared in an administered way, utilizing the back propagation calculation.

3.1 Artificial Neural Network (ANN)

A counterfeit neural organization (ANN) is an information getting ready perspective that is pushed by the way wherein a characteristic tactile framework in human psyche works. ANNs are used comprehensively for the course of action of different issues, including portrayal, vision, talk, plan affirmation, control systems, etc. A colossal number of neurons present in the human frontal cortex structures the vital segment of the neural framework perspective and go probably as simple dealing with segments [10, 11]. These neurons are significantly interconnected and work in relationship to handle complex issues. A fake neuron is a little planning unit and plays out a clear computation that is essential to the action of a neural framework. The model of a neuron contains the crucial parts like wellsprings of data, synaptic burdens, inclination, adding crossing point, and activation work.

From a customer point of view, ANN can be parceled into two huge classes reliant upon their affiliation geography: feed forward and feed backward neural frameworks. Feed forward neural frameworks empower the sign to stream the forward path figuratively speaking. The sign from any neuron doesn't stream to some other neuron in the main layer. In feed backward neural frameworks the sign from a neuron in a layer can stream to whatever other neuron whether it might go previously or succeeding layers. Figure-1 depicts the construction of this class of neural organize.

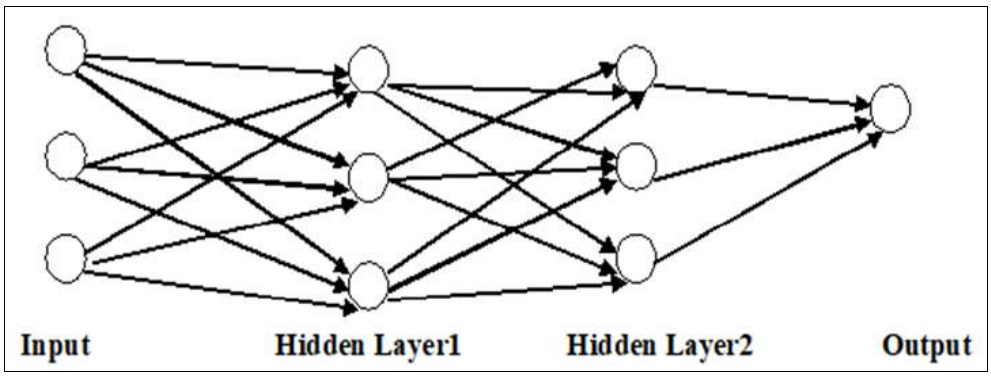


Fig 1: Structure of neural network

3.2 Multilayer Perceptron (MLP)

MLP is a champion among the most broadly perceived Neural Network Design that has been used for various applications. The MLP organize is commonly made out of different center points or taking care of units, and it is figured out into a movement of at any rate two layers [4].

The essential layer is named as an information layer where it gets the external information while the last layer is a yield layer where the response for the issue is gotten. The covered layer is the widely appealing layer in the data layer and the yield layer, and may shape with in any event one layer. The objective of Multilayer Perceptron learning is to find the best loads that limit the differentiation between the information and the yield. Most of planning computation are used in Neural Network is Back Propagation and it has been used in dealing with various issues in model affirmation and portrayal [11]. This computation depends on a couple of boundaries like Learning Rate, Momentum Rate and Activation work and so forth. In this proposed structure, include decrease technique utilizing SVM-RFE is directed as an underlying advance towards diminishing the quantity of traits without losing the principle reason and target data from the first information. The following stage is fostering an indicator with an improved exactness to order informational index. There are various stages in the proposed design for an efficient Arrhythmia arrangement. We are proposing another model for proficient element determination and Arrhythmia expectation. This methodology is of the accompanying strides as follows:

Stage 1: Read the Arrhythmia dataset.
Stage 2: Preprocess the dataset.
Step 2.1: Imputer the missing qualities utilizing mean
Stage 3: Select the huge highlights utilizing SVM-RFE calculation.
Stage 4: perform Classification utilizing MLP calculation on the dataset to choose the best highlights.
Stage 5: Evaluate execution of the classifier.

4. Experimental Details and Results

The examinations have been directed by utilizing Python programming language. It is an open-source programming language give amazing usage of various information investigation and Visualization strategies. It is a ground-breaking library that gives many AI grouping calculations, proficient apparatuses for information mining and information examination. The Python Scikit-learn is a bundle for information order, relapse, bunching and representation. We have considered the Dermatology data from UCI Machine Learning Repository datasets [12], for assessing the productivity and adequacy of neural network systems.

4.1. Dataset

The dataset was formed to examine skin disease and classify type of erythemato-squamous diseases. This dataset contains 35 variables, in this dataset 34 variables are linear and 1 variable is nominal. There are six classes of erythemato-squamous disease, with 366 instances and 35 attributes in the domain. In dermatology, erythemato-squamous disease identification and diagnosis is a difficult because all the classes contribute to the same clinical properties of scaling and erythema, with minor changes. These six classes of skin disease include 1: psoriasis, 2: seborrheic dermatitis, 3: lichen planus, 4: pityriasis rosea, 5: chronic dermatitis, 6: pityriasis rubra. Initially patients were first examined with 12 clinical features, after which the assessment of 22 histopathological attributes was performed using skin disease samples.

4.2 Data Preprocessing

The methodology proposed in this research paper starts with data preprocessing. Data preprocessing step includes (i) a data driven method to select patient records and selecting important variables for analysis and (ii) The collected data from patient records are not clean and may include noise, incorrect, missing values, or inconsistent data. In this data, age attribute 8 instances were missed, so we have to apply mean imputation technique to filling the missing records to clean such anomalies.

As the dataset has a large set of features, feature selection is applied to select the most relevant and significant features containing useful information required for data classification. We use 70% of records as the training data and the other 30% as the testing data.

4.3 Results

In the first stage MLP algorithm is trained on the original set of features was used in the experiment. In the second stage we implement a SVM-RFE algorithm for obtaining the adequate number of features to identify the features selected. The results that we got for MLP algorithm for the full dimension data and also after the feature reduction with SVM-RFE technique are shown in the figure-1. The dermatology dataset is labeled as one of 6 different skin disease types. From 34 attribute we have filtered to 24 feature vectors by using SVM-RFE technique to get an optimum selection from complete dataset for training as well as for testing experiments.

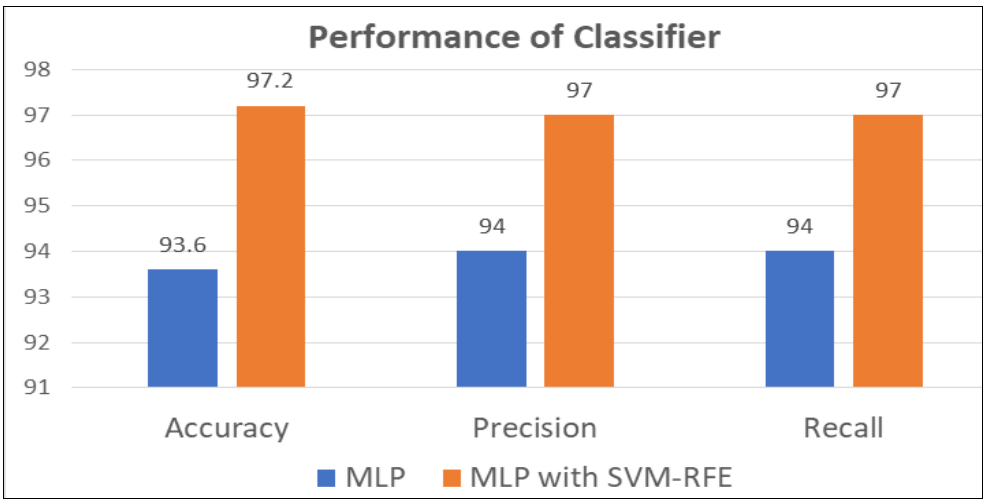
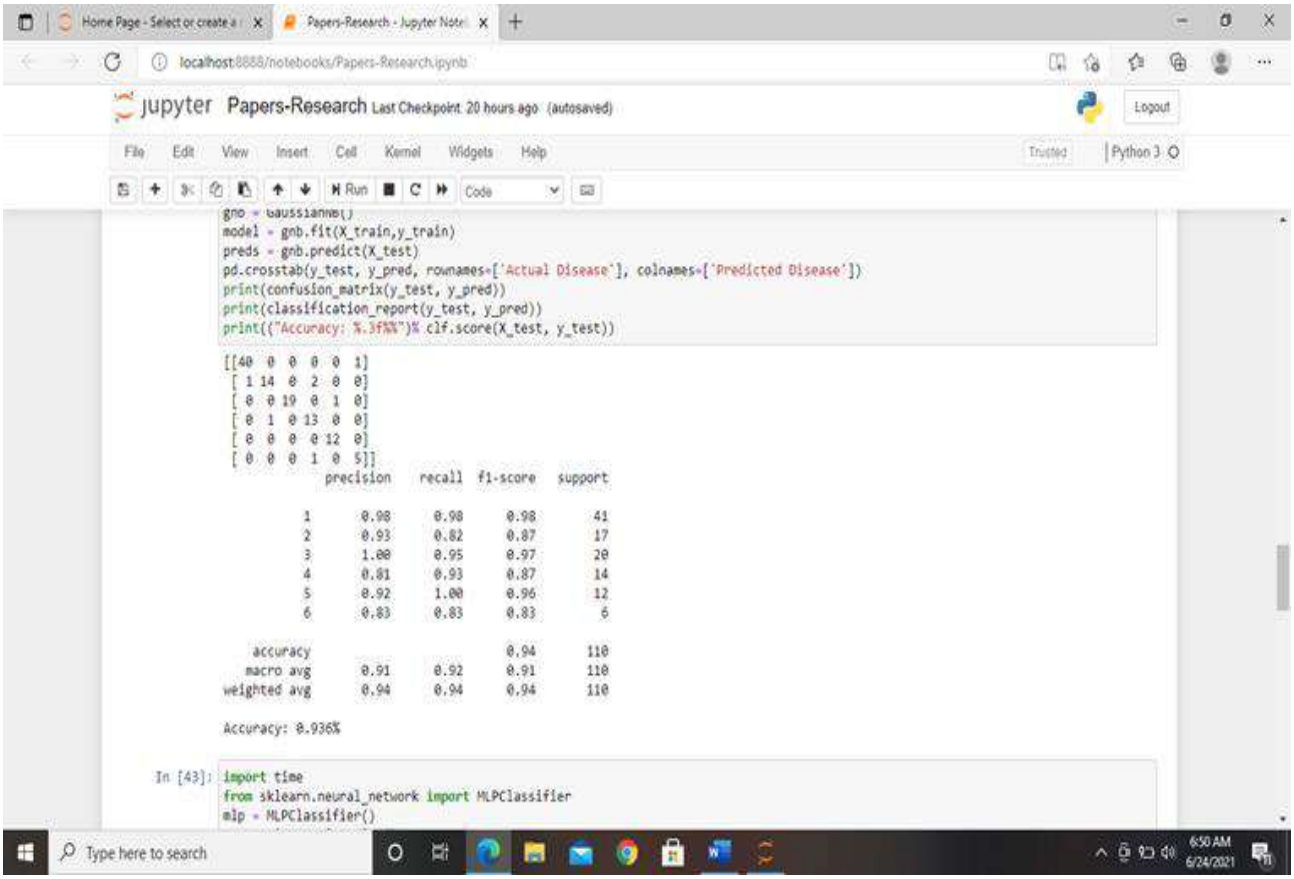
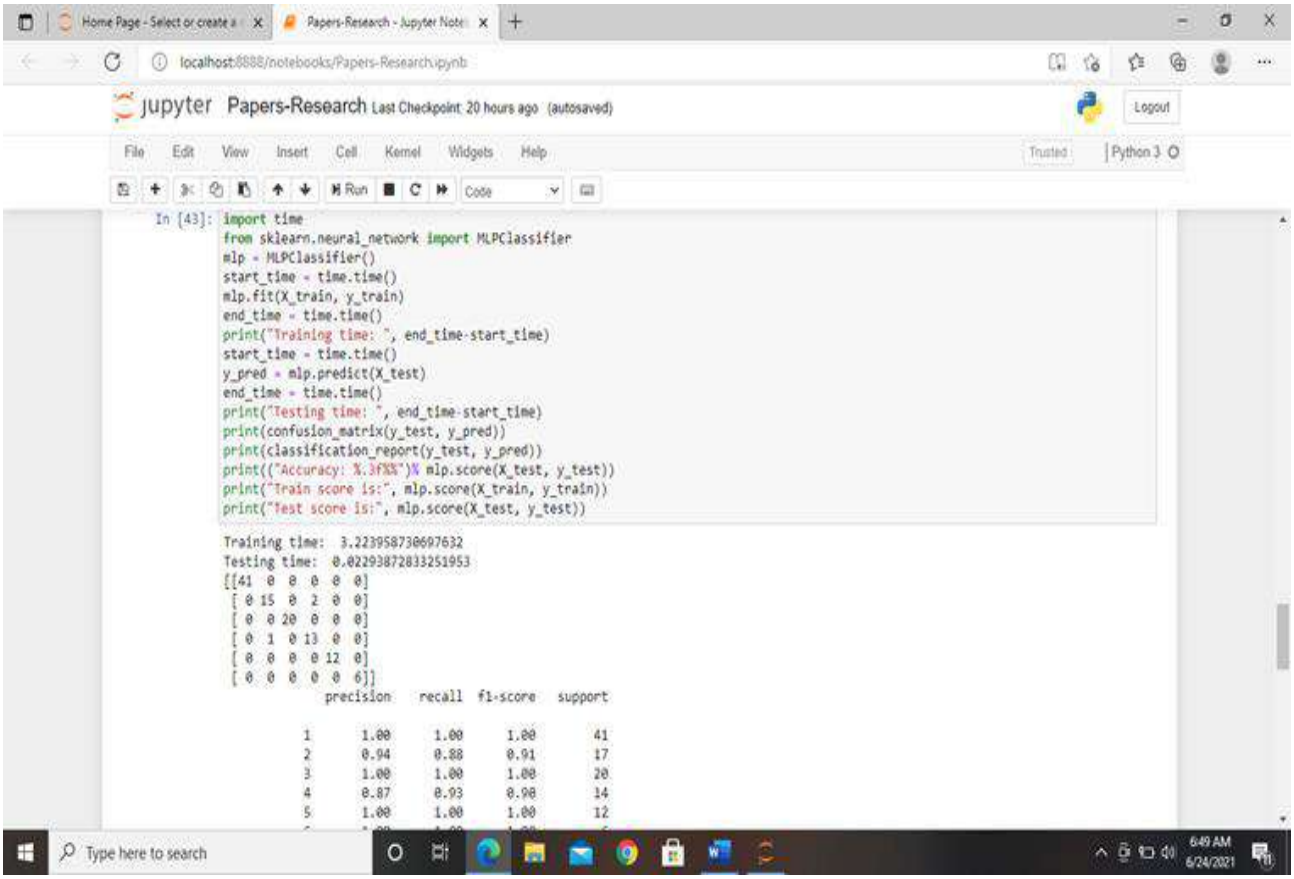


Fig: Performance of MLP

From the figure-1, we observe the performance of MLP without SVM-RFE based on accuracy has got 93.6%, whereas the performance of SVM with SVM-RFE feature selection based on accuracy has achieved 97.2%. However, there is an improvement in the accuracy with feature selection. The accuracy rate is increased 6.4% with feature selection.

Screen Shots





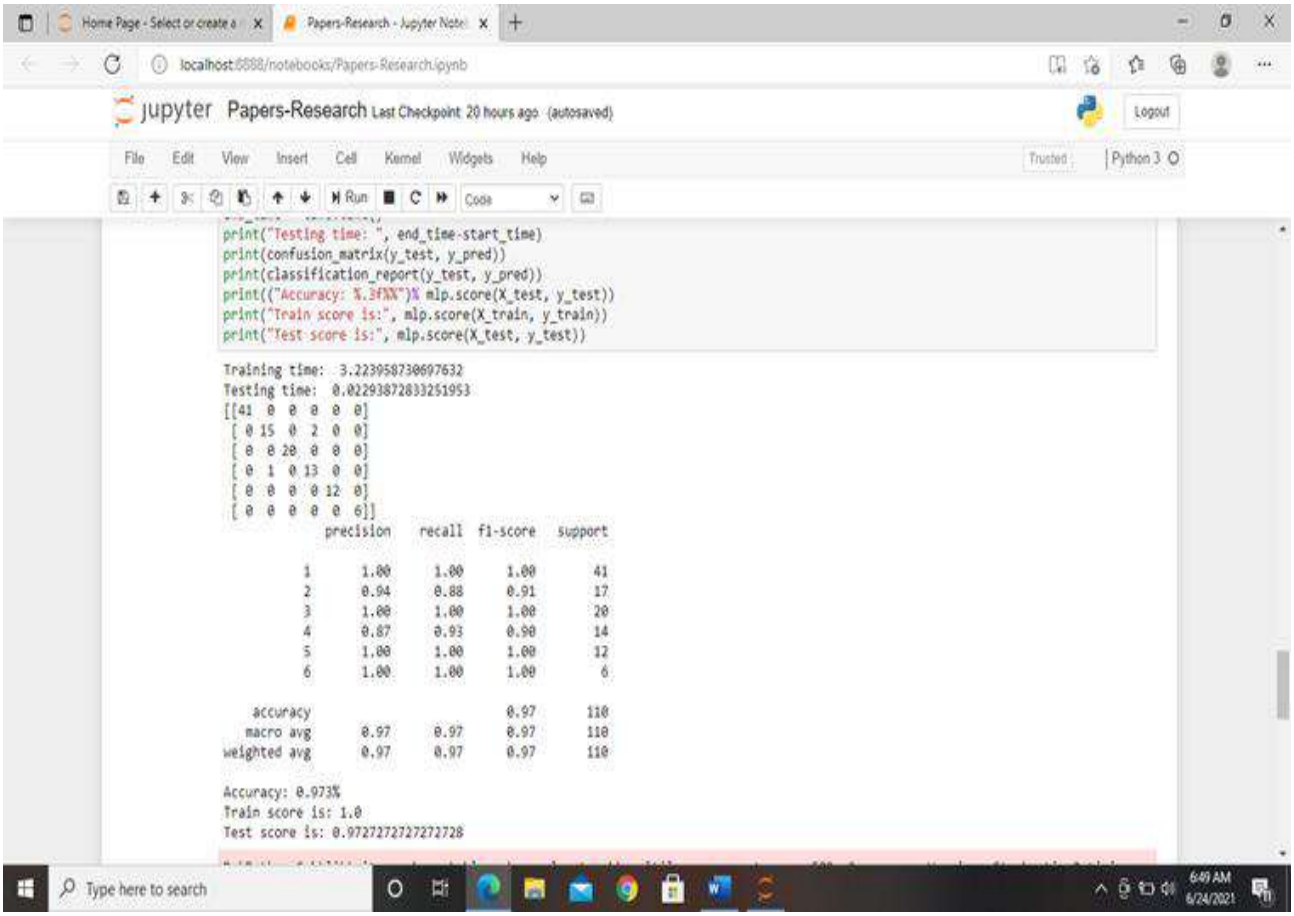


Fig 2: Results Screen shots

In our experimental result the MLP with SVM-RFE algorithm shows the highest accuracy compared with MLP without SVM-RFE. With the improvement the accuracy, the proposed model demonstrated that it performs well after selecting relevant features. This result provided new insight using a classification learning algorithm and reduction technique to selection relevant and important feature in order to improve the accuracy of the system and to identify possible features which may contribute to this improvement. Most of the proposed research system could effectively utilize feature selection process to improve detection rate of their system and minimize considerably the false alarm rate.

5. Conclusion

In this paper, the meaning of utilizing a bunch of pertinent highlights with the MLP arrangement learning calculation for Dermatology sickness forecast has been illustrated. This paper proposes a technique for multiclass order of Dermatology records with MLP based methodology. A show and suggestion of an element determination strategy which comprise of a recursive component disposal utilizing a SVM classifier to distinguish significant highlights have been finished. The component choice, preprocessing, and arrangement strategies have delivered a blend which gives promising outcomes to sickness grouping. The assessment the viability of the technique utilizing distinctive arrangement metric estimation has been made and it has been demonstrated that by decreasing the quantity of highlights, the precision of the model was improved. To recognize Dermatology sickness from enormous dataset, recognition calculation, and highlight determination technique have excessively more proficient.

6. References

1. Ahmed K, Jesmin T, Rahman MZ. Early prevention and detection of skin cancer risk using data mining. Int J Comput App 2013;62:1-6.
2. Bakpo FS, Kabari LG. In Tech Publisher. Diagnosing skin diseases using an Artificial Neural Network, in Artificial Neural Networks - Methodological Advances and Biomedical Applications 2011, 253-70.
3. Ravi Kumar G, Nagamani K, Anjan Babu G. A Framework of Dimensionality Reduction Utilizing PCA for Neural Network Prediction, Lecture Notes on Data Engineering and Communications Technologies, ISBN 978-981-15-0977-3, Springer Nature Singapore Pte Ltd 2020;37:173-180.
4. Ravi Kumar G, Venkata Sheshanna Kongara, Dr Ramachandra GA. An Efficient Ensemble Based Classification Techniques for Medical Diagnosis, International Journal of Latest Technology in Engineering, Management and Applied Sciences 2013;2(8):5-9. ISSN-2278-2540.
5. Guyon, Weston, Barnhill, Vapnik. Gene selection for cancer classification using support vector machines," Machlearn: Machine Learning 2002, 46.
6. Liu H, Yu L. Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng 2005;17(4):491-502.
7. Liu H, Sun J, Liu L, Zhang H. Feature selection with dynamic mutual information, Pattern Recognition 2009;42(7):1330-1339.
8. Witten H, Frank E. Data mining: practical machine learning tools and techniques with Java

implementations, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc 2000.
9. Guyon, Elisseeff A. An introduction to variable and feature selection, J Mach. Learn. Res 2003;3:1157-1182.
10. Han J, Kamber M. Data Mining concepts and Techniques, the Morgan Kaufmann series in Data Management Systems, 2nd ed. San Mateo, CA; Morgan Kaufmann 2006.
11. Michael N. Artificial Intelligence – A Guide to Intelligent Systems, 2nd Edition, Addison Wesley 2005.
12. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/.