International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 03, 2021

An Adaptive Graph Based Feature Selection And Deep Learning Classification Framework For Rice Disease Prediction

T.P.Senthilkumar[1], Dr.P.Prabhusundhar[2]

[1]Assistant Professor of Computer Science, Gobi Arts & Science College, Gobichettipalayam.
[2]Assistant Professor of Computer Science, Gobi Arts & Science College,Gobichettipalayam.

Email: [1]t.p.senthilkumar@gmail.com, [1]tps@gascgobi.ac.in,
[2]drprabhusundhar@gascgobi.ac.in

Abstract: The volume of information in every application necessitates the use of deep learning approaches for data analysis. Designing a deep learning-based classifier for decision-making and, as a result, upgrading the whole system's knowledge base is one such data investigation task in a dynamic context. The building of a classifier represents the retrieval of interesting patterns from a huge database of data and the prediction of future trends based on those patterns. The classification system's time consumption rises with time, and the system becomes inefficient as it is continually learnt for adding new groups of data to the current ones. If the knowledge of previous data collected by the classifier is used with the new group of data to construct the updated classifier, it may be done without learning the same classifier for all of the data. In the paper, for the selection of significant features, the principles of graph based adaptive feature selection approach are used. Because the features of rice illnesses fluctuate over time owing to changes in climatic, biological, and geographical variables, the deep learning-based classifier is ideal for use on the rice disease dataset for forecasting of disease. The suggested technique has been tested on synthetic rice disease datasets as well as benchmark datasets, and the classification accuracy has been assessed and compared to other state-of-the-art classification methods. The approach is also assessed based on the algorithm's performance in order to determine its importance and efficacy.

Keyword: Rice disease prediction, deep learning, feature selection, classification, neural network and graph.

1. INTRODUCTION

In 2017, the agriculture sector contributed 6.4% to world Gross Domestic Product (GDP) growth as per the United Nations report. The agriculture sector employed an approximate 50% workforce and contributed 17-18% of India's GDP in 2017-18. The agriculture industry is extremely important in India's economy. According to estimates, the overall employment in agriculture will be decreased to 25.7 percent by 2050. As a result, farm automation and process automation need to be improved. In 2014-15, 105482.1 thousand tonnes of rice produced in India. Rice production is vital for India's GDP [1, 2].
The rice or paddy is the staple food of the people in the eastern, southern and southeastern parts of India. The estimated production of rice in India is 81 million tons, grown on

approximate 43 million hectares. Over 3000 variants of the crop are cultivated in subtropical and tropical countries under a broad variety of agronomic and climatic circumstances, ranging from puddle wet soils in deltaic coastal areas to dry soils in the tropics, coastal sandy regions at the mean sea level and up to an elevation of about 2200 meters in some of hilly tracts [3]. These varieties differ in their duration, grain quality and other plant characters, water requirement, response to fertilizers, resistance to drought, alkalinity and salinity [4, 5].

Cultivation procedures vary greatly based on environmental conditions, soil, availability of water, and crop diversity in different sections of the nation [6]. All of these factors, including production techniques, have a significant impact on a variety's susceptibility to one or more diseases. Several illnesses, both pathogenic and non-pathogenic, emerge as a result of these significantly varied conditions impacting rice growth in different regions of the nation and at different times of the year, resulting in substantial harm to grain and straw production [8].

Plant disease is a complex process triggered by an incitant (pathogen/abiotic factor) that disrupts the plant's energy-using mechanism. This may be seen at both the micro (physiological, biochemical, and cytological) and macro levels (symptoms) [9]. The ability of the sick plant to generate or live is harmed. In the widest sense, plant diseases might be thought of as anything that isn't normal. In plants, disease is defined as any deviation from the norm, shown by physiological interruptions or structural alterations that are sufficiently persistent to halt growth, induce aberrant forms, or result in the premature death of a portion of the plant or the complete individual [10]. The symptoms, or how the plant seems after the pathogens have established themselves in the host, alert us to the fact that it is infected [11].

A disease symptom or sign is evidence of response to a diseased condition in a plant, which may be expressed in a variety of ways. All of these together form the basis for clinical diagnosis and, in sum total, are referred to as diseases syndrome [12]. Discolouration is one of the most common symptoms observed in a diseased plant, whether due to parasitic or non-parasitic agents. It may reflect a change in the colour of the whole plant or one or more of its parts. Yellowing, browning, silvering, and chlorosis are common in many diseased plants. Depending upon the intensities of the disease the grades of discolouration vary. Such symptoms are more pronounced in shoots and buds than on older parts.

In some cases the stem, roots and fruits are discoloured. Spots and shot-holes are common necrotic symptoms of plant diseases. Spots vary in size, shape and color, and the margins of the spots also vary in color and pattern [13]. The spots, together with their margins, display distinct characteristics in some diseases. Spot symptoms are commonly produced by fungal and bacterial infections, sometime by virus infection and occasional they are due to nutritional deficiencies and physiologic disorder. Streaks and stripes reflect the impact of disease development in some cases. In diseases, minute linear lesions may appear on the leaf blade, leaf sheath, stem and other plant parts. Blight and blast refer to diseases that kill suddenly. The tissues of a sick plant are swiftly damaged owing to the intensity of infection, resulting in the death of leaves, blossoms, and other above-ground plant components.

Blight is a term used to describe such a condition. Blast occurs when the entire leaf blade, bud, or other plant components are involved, leading in the rapid death of the portion or plant as a whole [14].The process of identification of rice plant disease with the images is a complicated process and the size of the data also necessitated the computational approaches. The accuracy of the classification can be enriched with the feature selection scheme. In this article, graph based feature selection scheme is used and the selected features are classified with the deep learning process.

The remainder of the article is organised as follows, the related works in the plant disease identification is given in Section 2, the proposed adaptive feature selection and deep learning is given in Section 3, the acquired results are discussed in Section 4 and the article is concluded in Section 5.

Related Work

The most important problem in computer vision-based precision farming is detecting and identifying things of interest (such as crops, weeds, or fruits) from the rest of the view. A well-designed hardware system with a robust mining technique on explicit picture datasets is required for the creation of such a model. The number and quality of datasets are critical for the created model's high performance. The review work [15] describes the specifics of certain well-known public picture databases. Crop and weed field picture collection is one of the first publicly available weed control datasets. At a carrot farm, a multispectral camera mounted on a field robot for picture capture [16]. The grassBroadleaf dataset was constructed from RGB photos collected by an unpiloted aerial aircraft (UAV) flying in a soybean field at a height of roughly 4 m above ground level [17].

The Orchard fruit dataset was collected in the orchard fields for three fruit varieties (i.e., apple, mango, and almond). The images for the apple and mango trees were acquired using an autonomous ground vehicle, while the almond data was acquired with a hand-held camera [18]. The Minne-Apple data was generated by a research team at the University of Minnesota for apple detection. This dataset includes diverse images from multiple fruit varieties over two growth seasons [19]. The Fruit Flowers dataset was created for the evaluation of semantic segmentation networks for flower detection of tree fruits. The image data was collected for the flowers of three species, apple, peach, and pear, using color cameras in natural orchard conditions [20].

The Maize Disease dataset was dedicated to automated, field-based detection of North leaf blight fungal foliar disease of maize. This data set was generated using a hand-held camera, a camera mounted on a boom, and a camera mounted on a small UAV [21]. Image and video datasets used in this investigation is not available in the benchmark datasets; therefore it is generated under the supervision of experts using a digital camera and mobile handset.

Adaptive Feature Selection Approach using Neural Network (AFSNN)

Effective feature selection using graph is accomplished by the wrapper method in combination with incremental training method to identify a subset of a feature from the available set of features. To attain the best generalization of learning process and the AFSNN automatically decides the count of the neurons during the process of feature selection where the hidden neuron determined by the incremental approach. The AFSNN is initialized by the minimum count of neuron and feature. During the incremental process, neurons and features are added whereas simple criteria is used to determine the addition of neuron and features.

Grouping the features

Based on the similarity, features are classified into groups with similar objects and the process is termed as clustering and it huge group of original clusters are needed. For grouping process, threshold value taken from user and handling the overall process is tedious. To overcome the shortcomings, AFSNN aims at identifying the relationship among the features. Through, the identification informative and distinct features for developing the robust feature selection model. The relationship among two variable is determined by the general statistics approach called correlation. In AFSNN approach, Pearson product moment correlation is applied to estimate the correlation measure among diverse features of training set. The coefficient of correlation p$_{ij}$ among two features i and j is equated as,

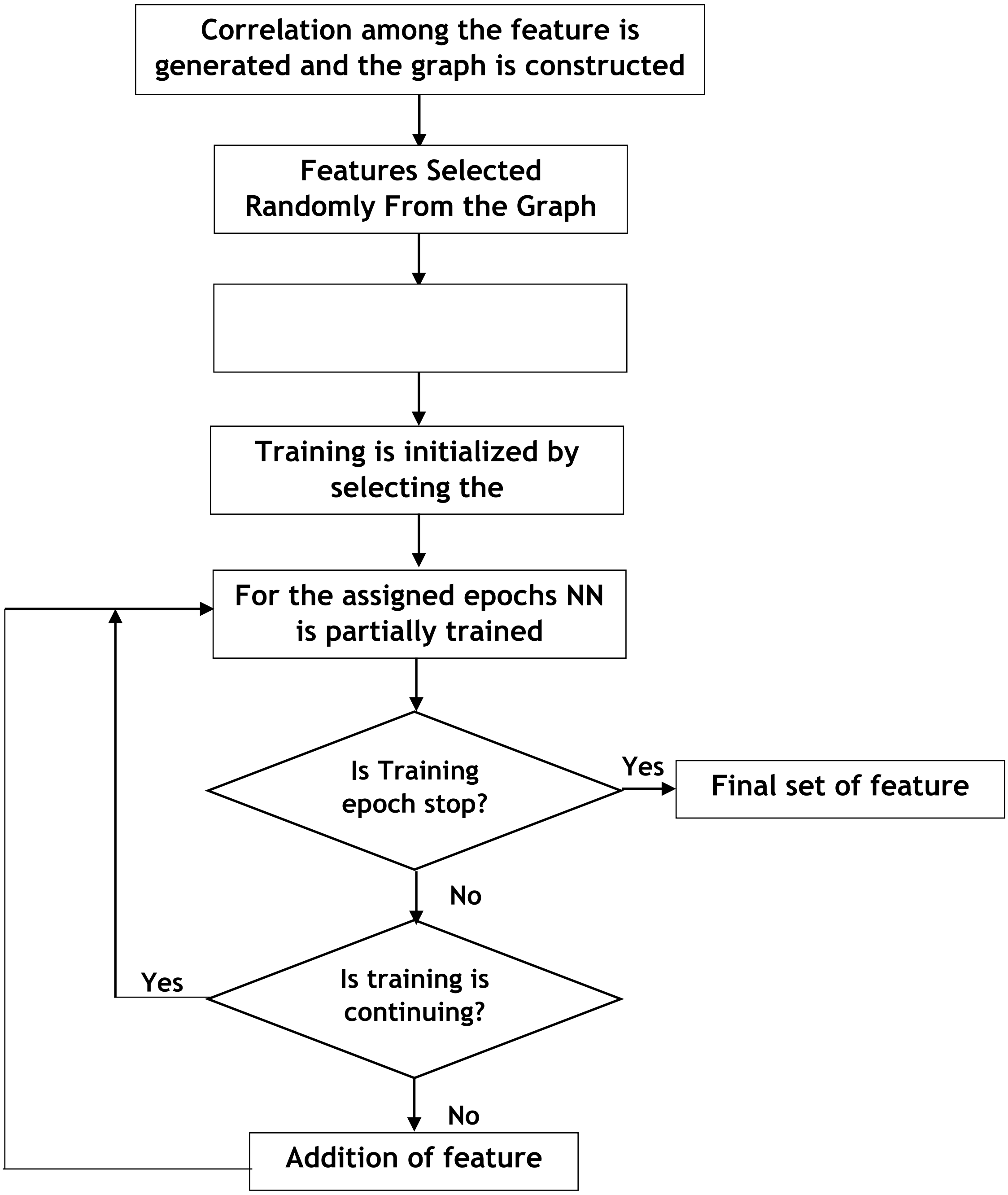
$$p_{ij} = \frac{\sum_r (y_i - \overline{y_\imath})(y_j - \overline{y_\jmath})}{\sqrt{(y_i - \overline{y_\imath})^2 \sqrt{(y_j - \overline{y_\jmath})^2}}}$$

where the values of the features i and j is signified by y$_i$ and y$_j$ respectively. The mean values of y$_i$ and y$_j$is signified by the $\overline{y_\imath}$ and $\overline{y_\jmath}$ which is averaged over the r value. Existence of exact linear dependency is determined by the complete correlation of values i and j whereby p$_{ij}$ would be 1 or -1. If the values are entirely uncorrelated i and j is assigned with 0. After estimating the all combinations of correlation coefficient, every feature is sorted as descending order. The correlation value of every feature i is estimated by,

$$crln_i = \frac{\sum_{j=1}^{NF} |p_{ij}|}{NF - 1} \quad if\ i\ \neq j$$

where NF is the count of feature and it is incorporated in signifying a given dataset. The overall process of the proposed AFSNN is described in Figure 1.

Figure 1. Flowchart for overall training process


Correlation among the feature is generated and the graph is constructed
Features Selected Randomly From the Graph
Training is initialized by selecting the
For the assigned epochs NN is partially trained
Is Training epoch stop?
Yes
Final set of feature
No
Is training is continuing?
Yes
No
Addition of feature

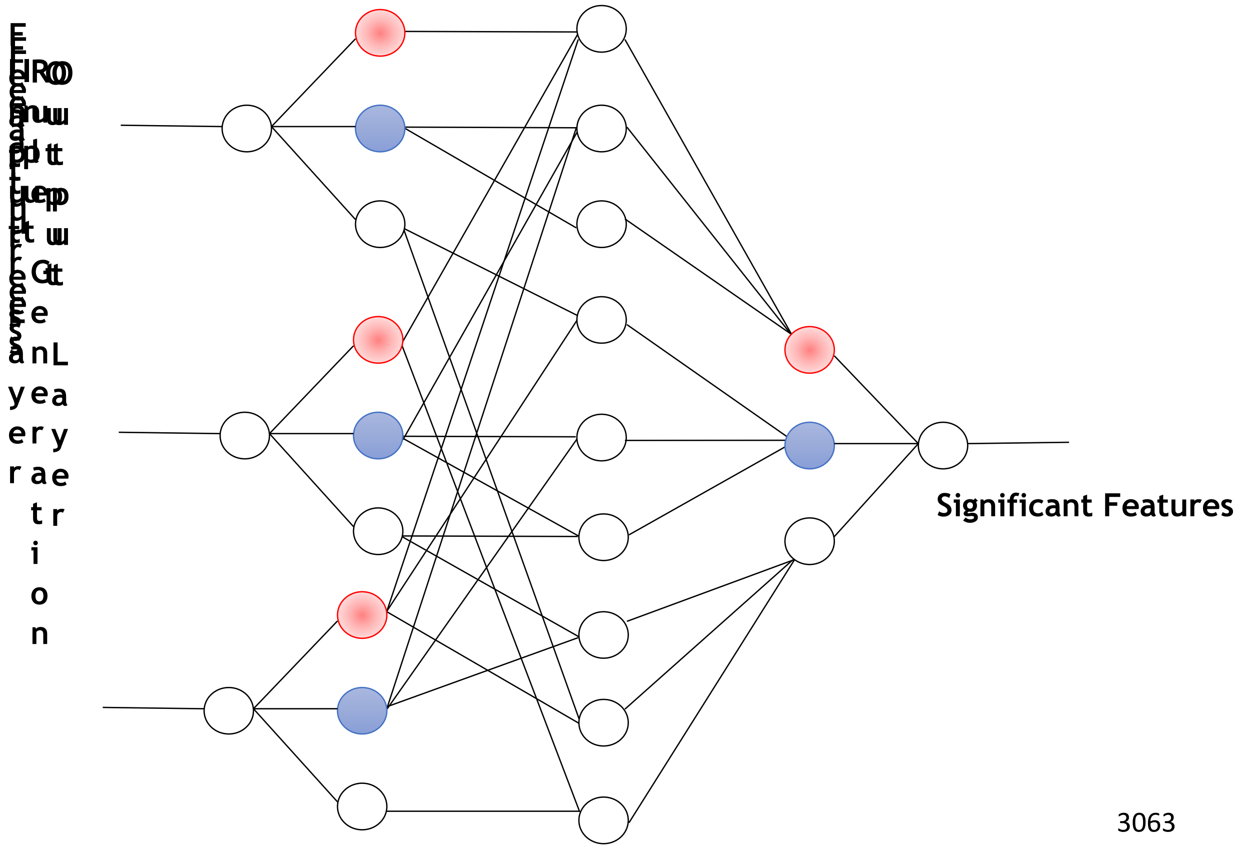
At the end of the process, AFSNN is categorized into two groups namely NF/2 features of similar (SI) group and NF/2 features of dissimilar (DS) group. The first feature in the group SI is the most correlated and last feature in the group DS is the least correlated in the dataset. Highly correlated features are identified and neighborhood information's are considered for generating a graph. The correlated features involved in the formation of graph is passed to the training phase of the neural network.

The graph is traversed for the random selection of features and it is passed to the training process. The training model is generated from training process, which is attained from the random features and it is called as learning process. The testing data is classified with the assistance of generated model. The acquired accuracy from the testing process is retained and random features are selected again for the classification process. The process is accomplished iteratively and the best accuracy attained feature is considered as best feature. The best classification accuracy, precision, recall, f-measure, and error rate is included in the comparison of results.

NN training process termination

Hidden neurons and characteristics are introduced to the proposed model one by one throughout the training phase. The training progression reduces the occurrence of mistake in the training process. During the training phase, a fraction of the dataset utilised in the training process is used for testing, which determines the method' exact correctness. The AFSNN procedure, on the other hand, enhances the NN's ability to generalise, therefore training is not a good way to end the NN training process. Figure 2 depicts the overall scheme of a neural network and the training method.

Figure 2. Overall structure of neural network



For the end of the training phase, a separate validation dataset is used. The validation error is utilised to obtain unbiased estimate, however it is not used to change the weights of the NN. Validation error ends the training process, which provides the best generalisation and is clear and uncomplicated. Strips are used to measure the inaccuracy in the validation process after each training period. When the validation error exceeds a certain threshold, the training is halted. For each TM subsequent time, the validation error is computed for each successive strip TM. The termination conditions are as follows:

$T(\tau+i)-T(\tau)> \lambda$, i=1,2,3….TM

where TM and τ are positive integers, and the value is determined by the user. Once the criterion specified in the preceding equation is met, the training process is ended. If the termination requirement is not met, the NN may have certain hidden neurons and significant features that are added, and the accuracy is then confirmed over a certain number of epochs, which is equivalent to

$$ACY = 100(\frac{R_{ua}}{R_u})$$

where the count of the patterns correctly classified R_{ua} and R_u denotes the whole set of data. If the value is not intensified or raised then the process will terminated automatically.

Selection and Addition of Feature

In the existing neural network, the features are added by the straightforward criteria and it determines the classification accuracy in the validation set. The best generalization is attained by the selection of salient features from the graph. The classification accuracy is described as,

$$ACY(T+\tau) > ACY(T), T = \tau, 2\tau, 3\tau, \dots \dots \dots$$

The proposed approach test criteria for selecting the feature for every τ epoch and add one feature to the subset of feature if the criteria is fulfilled. The classification accuracy is enriched by the feature addition process, which is the significant strategy in the proposed learning approach. The AFSNN improves the network processing power by the feature addition process. The feature selection process is accomplished completely by the selection and addition process until the group reaches the empty set.

2. RESULT AND DISCUSSION

Classifiers' effectiveness is assessed using various performance metrics like accuracy, recall, precision, f measure, Jaccard index, and MCC. The description of each metric is given below.

Recall: Recall quantifies the number of positive class predictions made out of all positive examples in the dataset as indicated as follows:

$$Recall = \frac{TP}{TP + FN}$$

Precision: Precision quantifies the number of positive class predictions that actually belong to the positive class, and it is estimated as follows:

$$Precision = \frac{TP}{TP + FP}$$

F-measure: F-Measure provides a single score that balances both the concerns of precision and recall in one number, and it is estimated as follows:

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

Accuracy: It is one of the most commonly used measures for classification performance, and it is defined as a ratio between the correctly segmented samples to the total number of samples as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Jaccard Index: It is one of the measures which only focuses on the likelihood of an object being positive and avoids the true negatives. It also relates the objects in same class or same cluster as True positives.

$$Jaccard\ index = \frac{TP}{(TP + FP + FN)}$$

F-measure: F-Measure provides a single score that balances both the concerns of precision and recall in one number, and it is estimated as follows:

$$F\ measure = \frac{(2TP)}{(2TP + FP + FN)}$$

Mathew's correlation coefficient: Mathew's correlation coefficient is one of the important measures which involves the balance between all the measures in the confusion matrix. MCC is given as:

$$MCC = \frac{FP * FN}{\sqrt{(TP + FP) * (TP * FN) * (TN + FP) * (TN + FN)}}$$

Table 1. Comparison of Classifier Performance

Metrics	Iteration	Decision tree	CNN	SVM	AFSNN
Accuracy	100	90.93	95.10	97.71	93.99
	200	89.79	95.07	97.59	93.19
	300	90.89	95.97	98.09	95.41
F1-measure	100	87.49	91.44	97.94	89.49
	200	87.15	90.77	97.03	89.37
	300	87.97	91.73	97.37	89.94
Precision	100	87.54	90.49	98.98	89.39
	200	87.03	90.37	98.01	89.90
	300	87.39	90.79	99.13	89.97
Recall	100	88.31	99.48	94.99	89.77
	200	88.30	91.17	94.13	89.53
	300	88.57	99.80	95.77	89.99
MCC	100	87.31	99.13	97.37	89.11
	200	87.94	99.05	97.97	89.07
	300	87.77	99.79	97.85	89.99
Jaccard Index	100	89.00	85.45	89.48	87.39
	200	89.00	85.31	89.18	87.18
	300	89.71	85.75	89.79	87.84

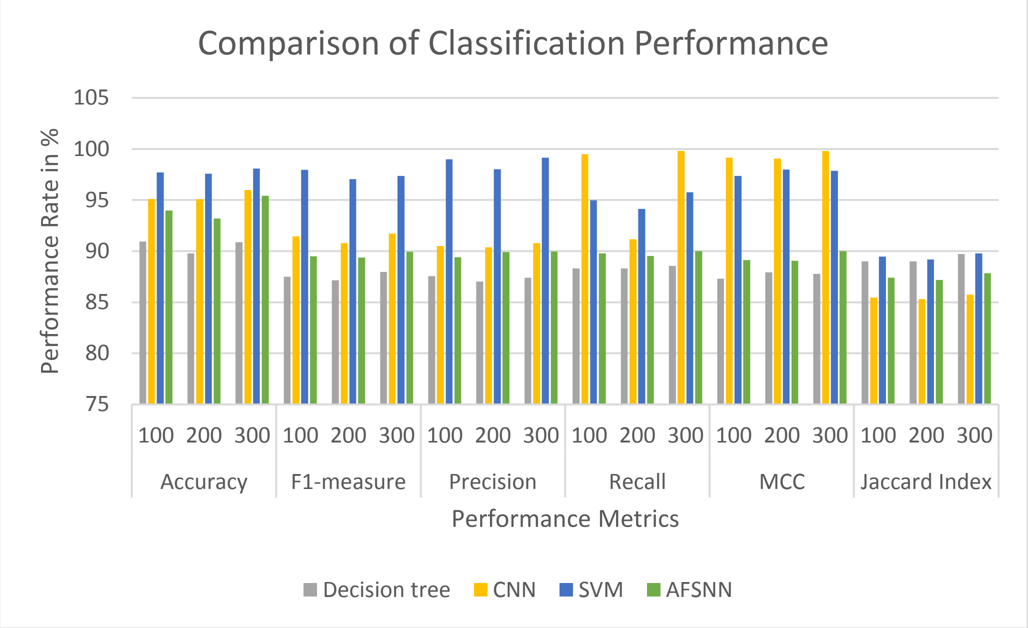


Figure 3. Comparison of Classifier Performance

3. CONCLUSION

The amount of data sets changes over time, running a static knowledge acquisition process for the full dataset is exceedingly time intensive. Machine intelligence is being developed by researchers to handle this type of challenge in a variety of real-world applications. One such data exploration task in a dynamic setting is designing a deep learning-based classifier for decision-making and, as a consequence, upgrading the entire system's knowledge base. The process of creating a classifier entails retrieving interesting patterns from a large collection of data and predicting future trends using those patterns. The classification accuracy of the adaptive feature selection based neural network has been analysed and compared to various state-of-the-art classification methods using simulated rice disease datasets as well as benchmark datasets. In order to identify the relevance of the technique, it is also evaluated based on the algorithm's performance.

4. REFERENCE

[1] Yue, W., Gao, J., & Yang, X. (2014). Estimation of gross domestic product using multi-sensor remote sensing data: A case study in Zhejiang province, East China. *Remote Sensing*, *6*(8), 7260-7275.

[2] Pawlak, K., & Kołodziejczak, M. (2020). The role of agriculture in ensuring food security in developing countries: Considerations in the context of the problem of sustainable food production. *Sustainability*, *12*(13), 5488.

[3] Agustina, E., Pratomo, I., Wibawa, A.D. and Rahayu, S. 2017. Expert System for Diagnosis Pests and Diseases of the Rice Plant using Forward Chaining and Certainty Factor Method. In International Seminar on Intelligent Technology and Its Applications (ISITIA).IEEE:266-270.

[4] Dhingra, G., Kumar, V. and Joshi, H.D. 2018. Study of Digital Image Processing Techniques for Leaf Disease Detection and Classification. Multimedia Tools and Applications. 77(15):19951-20000.

[5] Ferentinos, K.P. 2018. Deep Learning Models for Plant Disease Detection and Diagnosis. Computers and Electronics in Agriculture. 145:311-318.

[6] Delgado-Vera, C., Mite-Baidal, K., Gomez-Chabla, R., Solís-Avilés, E., Merchán-Benavides, S. and Rodríguez, A. 2018. Use of Technologies of Image Recognition in Agriculture: Systematic Review of Literature. In International Conference on Technologies and Innovation. Springer, Cham: 15-29

[7] Dos S. F. A., Freitas, D.M., Da S.G.G., Pistori, H. and Folhes, M.T. 2017. Weed Detection in Soybean Crops Using ConvNets. Computers and Electronics in Agriculture. 143:314-324.

[8] Haug, S. and Ostermann, J. 2014. A Crop/Weed Field Image Dataset for the Evaluation of Computer Vision Based Precision Agriculture Tasks. In European Conference on Computer Vision. Springer, Cham.: 105-116.

[9] Lu, Y. and Young, S. 2020. A Survey of Public Datasets for Computer Vision Tasks in Precision Agriculture. Computers and Electronics in Agriculture. 178: p.105760.

[10] Mahlein, A.K. 2016. Plant Disease Detection by Imaging Sensors–Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. Plant Disease. 100(2):241-251.

[11] Pantazi, X.E., Moshou, D. and Tamouridou, A.A. 2019. Automated Leaf Disease Detection in Different Crop Species Through Image Features Analysis and One Class Classifiers. Computers and Electronics in Agriculture. 156:96-104.

[12] Patrício, D.I. and Rieder, R. 2018. Computer Vision and Artificial Intelligence in Precision Agriculture for Grain Crops: A Systematic Review. Computers and Electronics in Agriculture. 153:69-81

[13] Peng, Y., Kondo, N., Fujiura, T., Suzuki, T., Yoshioka, H. and Itoyama, E. 2019. Classification of Multiple Cattle Behavior Patterns Using a Recurrent Neural Network with Long Short-Term Memory and Inertial Measurement Units. Computers and Electronics in Agriculture. 157:247-253.

[14] Phadikar, S. and Goswami, J. 2016. Vegetation Indices Based Segmentation for Automatic Classification of Brown Spot and Blast Diseases of Rice. In International Conference on Recent Advances in Information Technology (RAIT).IEEE:284-289.

[15] Lu, Y. and Young, S. 2020. A Survey of Public Datasets for Computer Vision Tasks in Precision Agriculture. Computers and Electronics in Agriculture. 178: p.105760.

[16] Haug, S. and Ostermann, J. 2014. A Crop/Weed Field Image Dataset for the Evaluation of Computer Vision Based Precision Agriculture Tasks. In European Conference on Computer Vision. Springer, Cham.: 105-116

[17] Bargoti, S. and Underwood, J. 2017. Deep Fruit Detection in Orchards. In IEEE International Conference on Robotics and Automation (ICRA). IEEE:3626-3633.

[18] Häni, N., Roy, P. and Isler, V. 2020. Minneapple: A Benchmark Dataset for Apple Detection and Segmentation. IEEE Robotics and Automation Letters. 5(2):852- 858.

[19] Dias, P.A., Tabb, A. and Medeiros, H. 2018. Multispecies Fruit Flower Detection Using a Refined Semantic Segmentation Network. IEEE Robotics and Automation Letters. 3(4):3003-3010.

[20] Wiesner-Hanks, T., Stewart, E.L., Kaczmar, N., DeChant, C., Wu, H., Nelson, R.J., Lipson, H. and Gore, M.A. 2018. Image Set for Deep Learning: Field Images of Maize Annotated with Disease Symptoms. BMC research notes. 11(1):p.440.

[21] Wilamowski, B.M. and Yu, H. 2010. Improved Computation for Levenberg– Marquardt Training. IEEE transactions on neural networks. 21(6):930-937.