International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-10 Issue-1, November 2020

Developing Classification Model for Chickpea Types using Machine Learning Algorithms

Nigus Asres Ayele, Hailemichael Kefie Tamiru

***Abstract:** Ethiopia is the leading producer of chickpea in Africa and among the top ten most important producers of chickpea in the world. Debre Zeit Agriculture Research Center is a research center in Ethiopia which is mandated for the improvement of chickpea and other crops. Genome enabled prediction technologies trying to transform the classification of chickpea types and upgrading the existing identification paradigm.Current state of the identification of chickpea types in Ethiopia still sticks to a manual. Domain experts tried to recognize every chickpea type, the way and efficiency of identifying each chickpea types mainly depend on the skills and experience of experts in the domain area and this frequently causes error and sometimes inaccurate. Most of the classification and identification of crops researches were done outside Ethiopia; for local and emerging varieties, there is a need to design classification model that assists selection mechanisms of chickpea and even accuracy of an existing algorithm should be verified and optimized. The main aim of this study is to design chickpea type classification model using machine learning algorithm that classify chickpea types. This research work has a total of 8303 records with 8 features and 80% for training and 20% for testing were used. Data preprocessing were done to prepare the dataset for experiments. ANN, SVM and DT were used to build the model. For evaluating the performance of the model confusion matrix with Accuracy, Recall and Precision were used. The experimental results show that the best-performed algorithms were decision tree and achieve 97.5% accuracy. After the evaluation of results found in this research work, agriculture research centers and companies have benefited. The model of chickpea type classification will be applied in Debre Zeit agriculture research center in Ethiopia as a base to support the experts during chickpea type identification process. In addition it enables the expertise to save time, effort and cost with the support of the identification model. Moreover, this research can also be used as a corner stone in the area and will be referred by future researchers in the domain area.*

***Keywords:** Chickpea, phenotype, varieties, Identification, classification, Selection*

I. INTRODUCTION

Chickpea was known in various parts of the country and it is the second most important pulse crop in the world, after dry beans [1]. Chickpea constituting the major component of the poor people's 'diet and it plays an important role in human nutrition in Ethiopia and the arid and semiarid regions of the world. Chickpea is among important commodities and more than fifteen percent of Ethiopian legumes with an area of some 239000ha and about one million households participated in chickpea production [2] Ethiopia is the leading producer, consumer, and seller of chickpea in Africa, and is among the top ten most important producers in the world [1].

Nowadays, researchers attempt to develop genomic selection (GS) aims to improve the identification of agronomic traits by using data from high throughput genotyping platforms and phenotypic information [4]. It is originally proposed by [5] for animal breeding, and also it is promising breeding mechanism to identify crop phenotypes [6, 7, 8, 9]. GS has advantages of identifying phenotypic trait values before planting, estimating the breeding values before crosses are made, and, notably, of reducing the time of the breeding cycle [7, 8, 10, 11].

With advances in phenotype identification, volumes of data have simultaneously increased; new research efforts such as machine learning were needed.

Machine learning have emerged as new data-driven science and it focuses on estimating more accurate predictive values of unobserved individuals.

In this study, the researcher is conducting machine-learning algorithms for chickpea in order to build a model and to classifying chickpea types. In Ethiopia identifications mechanism of chickpea type is still manual and domain expert's in the domain area tries to identify the chickpea. Even if, experts tried to recognize every chickpea genome and types, the way and efficiency of identifying each chickpea varieties are mainly depending on the skills and experience of experts in the domain area. Sometimes the identification of chickpea verities and type may not well identified and even if expertise is highly-skilled they may face challenges, exposed for bias and sometimes leads to inefficiencies, errors, inconsistencies, and inaccurate. Most of the identification researches were done globally; for local and emerging varieties of Ethiopia, there is a need to design identification model that assists the selection and identification mechanisms of chickpea.

Attempts are made to design phenotype identification model globally to facilitate the selection, identification and classification of different crops phenotype such as wheat, maize, rice [12, 13, 14, 15].

However, the accuracy of an existing algorithm should be verified and optimized. The application of such kinds of models for Ethiopian crop breeding is still required highly because lack of efficiencies, consistencies, accurate selection, identification and predictions in high dimensional datasets. Finally the main aim and contribution of this study for domain experts and research community is design chickpea phenotype identification model using machine-learning algorithm. Another important contribution is collecting and preparing chickpea genotype dataset in order to help other researchers in conducting related studies to handle data problems.

Revised Manuscript Received on November 30, 2020.

***** Correspondence Author

Nigus Asres Ayele*, Information Technology, Wolkite University, Wolkite, Ethiopia.

Hailemichael Kefie Tamiru, Software Engineering, Wolkite University, Wolkite, Ethiopia

II. RELATED WORKS

W. Ma et al. [16] attempt to develop a deep learning-based neural network model (DeepGS) to predict phenotypes from genotypes, in addition, they used 10 Cross-validation test to measure the performance the developed models. The data set used for the experiment was obtained from the wheat gene bank of CIMMYT (International Maize and Wheat Improvement Center), which consists of 2,000 Iranian bread wheat landrace accessions. The accessions were phenotype for grain length, test weight, grain hardness, grain protein, and thousand-kernel weight, and grain width, plant height. They compare DeepGS with conventional neural network and RR-BLUP. The result shows DeepGS gives a relative improvement ranging from 27.70% to 46.34% over CNN and 1.44% to 65.24% over RR-BLUP. Finally, they illustrated the combination of DeepGS and RR-BLUP with an ensemble learning approach give better performances.

Roorkiwal et al. [17], the data set for the experiment was a collection of 320 elite breeding lines from the International Chickpea Screening Nursery (ICSN) and ICRISAT in the crop seasons of 2011– 12 and 2012–13. The aim of the paper is to predict breeding values of lines and selection of lines prior to field phenotyping using a genome-wide marker, which was tested for traits such as Days to flowering, 100 seed weight, Days to maturity, and Seed yield at different seasons and traits. They used five-fold cross-validation test models 80% of the lines for training 20% of the line were for testing. They used statistical GS models such us Ridge Regression Best Linear Unbiased Predictor (RR-BLUP), Random Forest, Bayes Cπ, B, Kinship Gauss model and Bayes LASSO (machine learning algorithm). The models were tested and the performance of models was analyzed across various seasons and traits. The result shows that the accuracy using RR-BLUP, Bayes Cπ, Bayes B, RF, Kinship Gauss model were 0.663, 0.823, 0.673, 0.851, 0.707, and 0.673 respectively.

González-Camacho et al. (2016) [12] compare the classification accuracy of Multi-layer perception versus probabilistic neural network (PNN). In this study, they were used 16 maize and 17 wheat genomic and phenotypic datasets with sample sizes ranged from 290 to 300 individuals using high-throughput molecular marker information. The performance criteria used to evaluate the predictive accuracy the area under the receiver operating characteristic curve, and the area under the precision-recall curve. The result shows that PNN performed well and give better accuracy than MLP in most of the datasets. Finally, the accuracy of PNN and MLP were 0.746 and 0.724 respectively. Rachmatia H, Kusuma WA, and Hasibuan LS [13] develop a prediction model that used to predict phenotype of maize based on SNPs using deep belief network. The dataset used in this experiment was maize dataset from CIMMYT's. They compared the precision of DBN with kernel Hilbert space regression, best linear unbiased predictor and Bayesian LASSO. The result shows DBN outperforms than other methods and it achieves a correlation of 0.579. Hoffstetter et al. [14] use phenotype and genotype data from a training population of 470 soft winter wheat lines or population size and 4,858 markers to assess the accuracy of genomic selection for traits such as flour yield, grain yield, softness equivalence, and Fusarium Head Blight (FHB) resistance. They compare and evaluate RRBLUP, BL, and RF, in addition, the researchers used 10-fold cross-validation test.The result shows that the accuracy using RRBLUP, BL, and RF were 0.62, 0.57, and 0.63 respectively.

III. METHODOLOGY

A. Datasets

The data set used for this study was collected from Debre Zeit Agricultural Research center (Ethiopia) and nine years chickpea data of 8303 record and 8 features were used to build the model. For detail analysis the researchers have selected 8 features based on feature selection tools and techniques. Preprocessing was done though data cleaning techniques such as data reduction, noise removal and handling missing values. After preprocessing was done the data was prepared in the form of comma separated values (CSV). In order to build chickpea identification model, the dataset is divided in to training and testing set, using percentage split technique, 80% (6656) of data used for training and 20% (1647) of the data used for testing.

B. Feature selection

In this paper feature selection were conducted by using Boruta algorithm and it is an all relevant feature selection algorithm that gives feature importance measure and it uses random forest as default [18, 19, 20]. The researchers have selected 8 features shown in figure 1 below. The features were selected using feature selection tools based on the importance of features on chickpea type classification and in this research Boruta feature selection algorithm was implemented with R package "boruta".

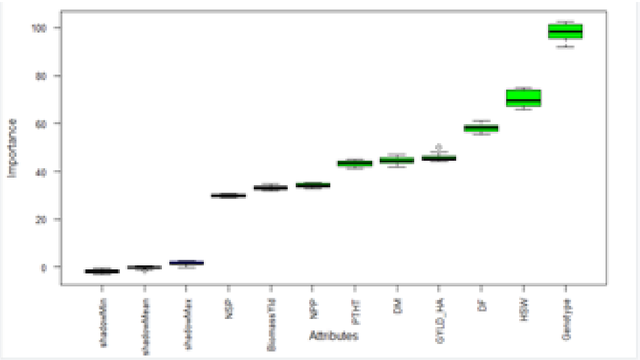

Figure 1:All importance features/attributes

The selected features and their description are listed below.

- Genotype: Name of genotypes
- Days of flowering (DF): Number of days to achieve 50% flowering in the whole plot.
- Days of maturity (DM): Number of days to achieve maturity in the whole plot.
- Number of Pods per plant (NPP): Plants from each plot were harvested and weighed to measure the number of pods.
- Plant Height (PTH): The height of the plant to achieve maturity in the whole pilot
- Number of pod damaged(NPD): the number of pod damaged per plant Grain Yield per pilot (GYLDP): Plants from each plot were harvested and weighed to measure the grain yield.

Grain Yield per hectare (GYLDHA): Plants from each hectare were harvested and weighed to measure the grain yield

C. Classification Algorithms

Three different algorithm artificial neural network, support vector machines and decision tree were used for classification of chickpea type. Description of those algorithm describe below.

Multilayer Perception: It is a class of feed forward ANN and it is arranged in a layer of the neuron. Mostly it has three layers. The first layer is known as input layer, the second layer is known as hidden layer and the third layer is output layer. We used MLP algorithm to generate the network. The network is also possible to alter and monitored at training phase. The multilayer perception neural network use back propagation algorithm for learning and back propagates the error from output layer to hidden layer and input layer.

In this research work, we used MLP method in order to classify the data set to Desi and Kabuli because it has been widely applied in classification and pattern identification, it can approximate any continuous function to desired accuracy. MLP has a simple architecture and relatively easy to implement. The choice of the network architecture and the choice of the parameters are the crucial factor to build an accurate MLP identification model. The performance of neural networks depends on the network type, numbers of input neurons, number of hidden layer, number of hidden layer neuron, training algorithm, and activation function used at each layer of the network [21]. This study uses the selected features of chickpea and genotypes as input neurons, one hidden layer, three hidden layer neurons, and sigmoid activation as activation function and back propagation algorithm as training algorithm. Finally in this research MLP was implemented with R package "neural net".

Support Vector Machines: It is classification algorithm for both liner and non- liner data. Additionally, SVMs is an algorithm that uses a nonlinear mapping to transform the original training data into a higher dimension [22]. The working principles of SVM are based on the concept of decision planes that defines the decision boundaries. A decision plane is one that separates between a set of objects having different class memberships [22]. An appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane [22]. Support vector machine has various characteristics such as the ability to handle large feature space, ability to prevent over fitting and information-dense in a given data set [22]

SVM operating to minimize the mathematical classification error and maximize the geometric margin; it is also called Maximum Margin Classifiers. SVM map input vector to a higher dimensional space where a maximal separating hyperplane is constructed. Two parallel hyper-planes are constructed on each side of the hyper-plane that separates the data. The separating Hyper-plane is the hyper-plane that maximizes the distance between the two parallel hyperplanes [23]. SVMs are one of an excellent algorithm used for classification. In this study, we have used this algorithm to split the data into Desi and Kabuli because it has the ability to handle large feature space, the ability to prevent over fitting and information dense in a given data set. SVM was implemented using R package "Kernellab".

Decision Tree: Researches related to statistics, machine learning, pattern recognition, and data mining have explained with a decision tree for available data to be classified. According to Han and Kamber, explained decision tree is a flow chart tree structure where each internal node represents a test on an attribute, each branch represents an outcome of the test and the leaf nodes represent class or class distribution [24]. Decision tree involves both nominal and numerical attributes. According to Two Crows Corporation (1999) decision tree is the ways that represent the available data into classes or labels. "Classification trees label records and assign them to the appropriate class" [25]. We have a various reason to select decision tree, which is easy to understand, it is easily converted to a set of production rules, and it is very efficient, which is desirable for a large amount of data. This is in case of the partitioning nature of the algorithm and it is the most powerful algorithm for binary classification. Finally in this research Decision tree was implemented using R package "rpart".

IV. EXPERIMENTS

In order to build chickpea type classification model the dataset is divided in to training and testing set, using percentage split technique, 80% of the total data (6656 data which contains about 8303) are used for training and the remaining 20% (1647 data which contains about 8303) for testing. The diagram Figure 2 below illustrated that the general approaches of designing classification model for chickpea.

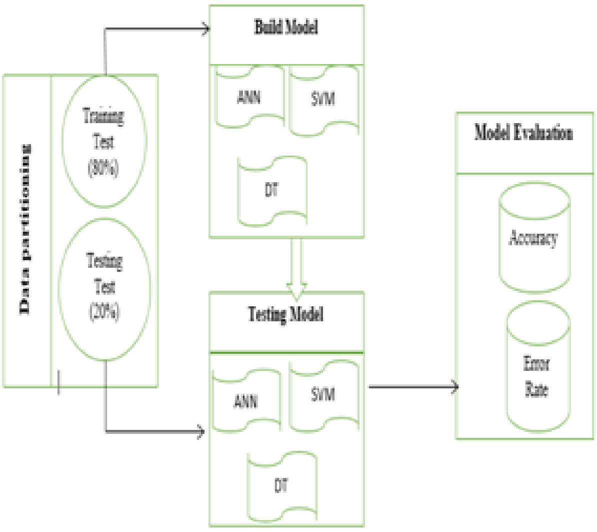

Figure 2: General approaches for designing chickpea type classification model

A. Experiment 1: Multi-Layer Perception

In this experiment, our neural network model used 8 features. Hence, the neural network model has 8 inputs, 3 hidden unit and 2 output units (8:3:2) where input units are the number of features and the outputs are class and we are used sigmoid activation function that used in the hidden and output layers.

Table 1 below illustrated that the parameters such as network layer, input neurons, hidden layer, number of neuron in hidden layer and output neurons used in MLP experiment.

Table 1: MLP parameters used

Parameter	Values
Input Neurons	9
Number of hidden layers	1
Number of hidden layer neurons	3
Transfer Function	Sigmoid
Output neuron	1
Maximum epochs (iterations)	16034

B. Experiment2: Support Vector Machines

The second classifier used in this paper is Support Vector Machines. The researcher's design the models using the selected features were applied to the data sets and the parameters that we applied for the classifier with correspondence value shown in Table 2 below.

Table 2: Parameters used for SVM classifiers

Parameter	Values used
Epsilon	0.1
kernel	rbfdot
Tolerance parameter	0.001

C. Experiment 3: Decision Tree

The third classifier used in this paper was decision tree classifier in identifying chickpea type using testing datasets was evaluated. A model was built using library rpart with the selected features. The parameters used in decision tree experiment illustrated in Table 3 below.

Table 3: Decision tree parameters used

Parameter	Values used
Algorithm	Ctree
Mincriterion	0.99
Minsplit	1000
maxdepth	8

V. RESULTS

A. Experiment 1 Result

The diagram illustrated in Figure 3 below shows that the network diagram generated from multilayer perception neural network and it shows there are 8 input neurons, one hidden layer, three hidden layer neurons and output class.

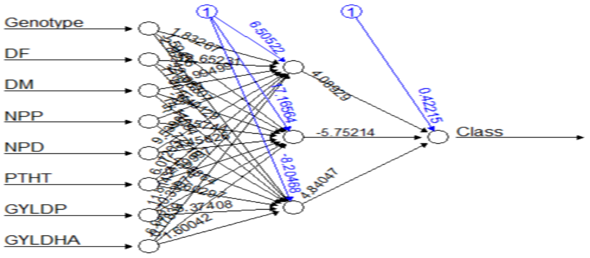

Figure 3: MLP network diagram on result

Confusion Matrix of the MLP model with chickpea dataset illustrated in Figure 4 below , the brown parts represent the value of the TP=765 and TN=108 for the decision model whereas green parts represent the value of the FP=133 and FN=108.

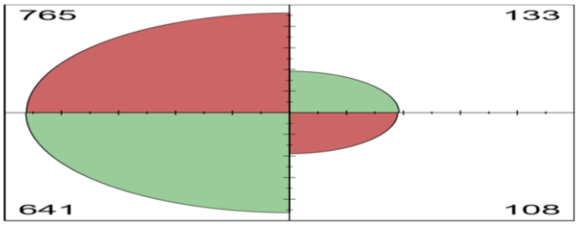

Figure 4: confusion matrix MLP model

Figure 5, above shows a confusion matrix for chickpea type classification model using MLP in experiment 1. In this experiment we have used our testing dataset 1647 instances and applied MLP with correctly classified instances were 873 which means 53% and incorrectly classified instances were 774 which means 47%. The Precision, Sensitivity (Recall), specificity of the model was 0.5441, 0.8518, and 0.4481 respectively. Lastly the diagram Figure 5 below illustrated that the Receiver Operating Character curve and Area under Curve of MLP experiment.

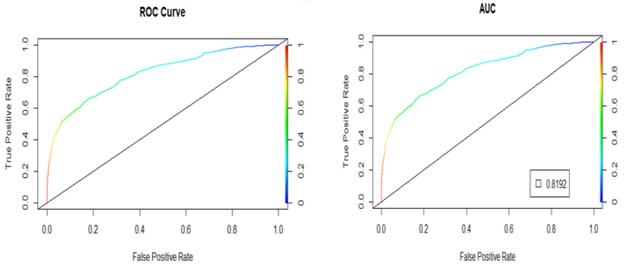

Figure 5: ANN ROC curve and AUC scores

Table 4: MLP results

Model	Features used	Recall	Precision	Specificity	AUC Scores	Accuracy
MLP	8	0.8518	0.5441	0.4481	0.8192	53%

B. Experiment 2 Result

Confusion Matrix of the SVMs model with chickpea dataset illustrated in Figure 6 below and the green parts represent the value of the TP=850 and TN=612 whereas green parts represent the value of the FP=65 and FN=120 for SVMs.

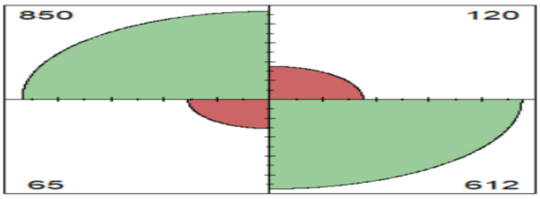

Figure 6: Confusion matrix SVMs model

Figure 6 above shows confusion matrix for chickpea type classification model using SVMs in experiment 2. In this experiment, we have used our testing dataset 1647 instances and applied SVM with correctly classified instances were 1462 which means 88.8% and incorrectly classified instances were 185 which means 11.2%.

The Precision of the model was 0.9289, Sensitivity of the model was 0.8762 and lastly the specificity of the model was 0.8360. Lastly the diagram Figure 7 below illustrated that the Receiver Operating Character curve and Area under Curve of SVMs experiment

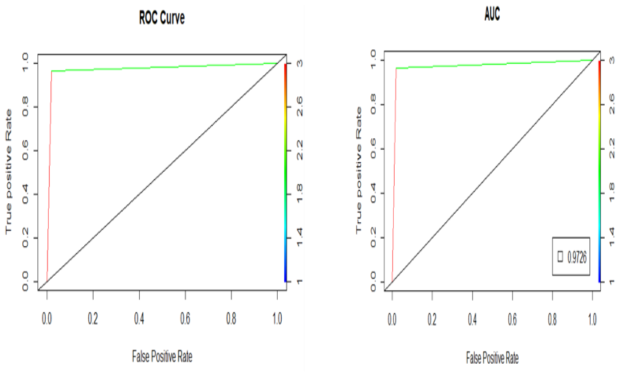

Figure 7: SVMs ROC curve and AUC scores

Table 5: SVMs results

Model	Features used	Recall	Precision	Specificity	AUC Scores	Accuracy (%)
SVMs	8	0.8762	0.9289	0.8360	0.9726	88.8%

C. Experiment 3 Result

The tree generated in the decision tree experiment illustrated in Figure 8 below.

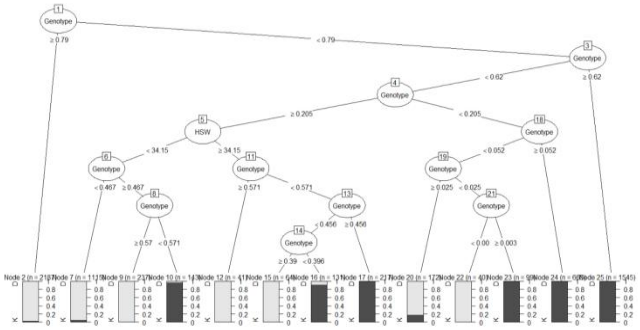

Figure 8: Decision tree structure

Figure 9: below illustrate the Confusion Matrix of the Decision Tree model with chickpea dataset and from the Figure, the blue parts represent the value of the TP=888 and TN=717 for the decision model whereas white parts represent the value of the FP=10 and FN=32.

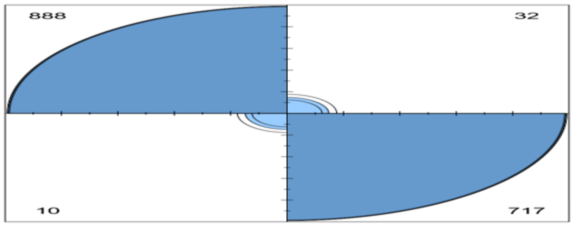

Figure 9: Confusion matrix of decision tree model

Figure 9 above shows confusion matrix for chickpea type classification using Decision tree with testing data and the results of experiment 3. With our testing dataset, out of 1647 instances, 1605 (97.5%) were correctly classified and the remaining 42 (2.5%) instances were incorrectly classified and, The Precision, Sensitivity, specificity of the model was 0.9885, 0.9655 and 0.9575 respectively. Lastly the diagram Figure 10 below illustrated that the Receiver Operating Character curve and Area under Curve of Decision Tree experiment.

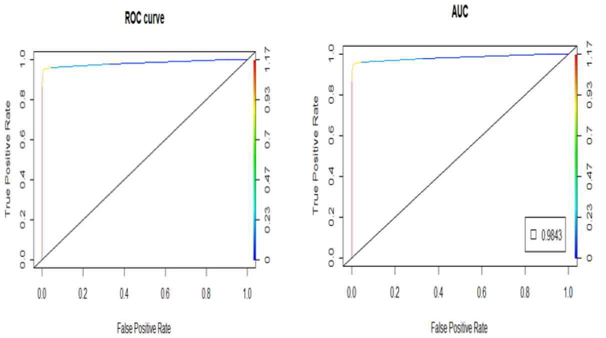

Figure 10: Decision Tree ROC curve and AUC scores

Table 6: Decision tree results

Model	Features used	Recall	Precision	Specificity	AUC Scores	Accuracy (%)
Decision Tree	8	0.9655	0.9885	0.9575.	0.9848	97.5%

Table 7: Summary of experimental results

Model	Recall	Precision	Specificity	AUC Score	Accuracy (%)
MLP	0.8518	0.5441	0.4481	0.8192	53%
SVM	0.8762	0.9289	0.8360	0.9726	88.8%
Decision tree	0.9655	0.9885	0.9575.	0.9848	97.5%

As shown in Table 7 above the comparison between the models focus on which algorithms perform better outcomes. The highest accuracy of the model was generated by decision tree algorithm that is Ctree classifier and having an accuracy of 97.5%, precision of 98.85%, recall of 96.55%, specificity of 95.75% and higher AUC score of 98.48% in experiment 3. The second well-performed model was SVM and having an accuracy of 88.8%, precision of 92.89%, recall of 87.62%, and specificity of 83.6% and AUC score of 97.26% in experiment 2. The lowest accuracy was performed by MLP with having accuracy of 53%, precision of 54.41%, recall of 85.18%, specificity of 44.81% and AUC of 81.92% in experiment 1.The highest sensitivity (recall) and specificity was registered by Ctree classifier, whereas lowest sensitivity was generated by MLP.

VI. VI. Model Comparison with Related Works

As the best of researchers‟ knowledge, no research work was done in the area of Ethiopian chickpea type classification; this study is the first attempt in the area. So, we compare the results of this research work with other researchers work done globally. In this study, during experimentation 8 features in the decision tree model have been found to work very well in classification of Ethiopian chickpea type. The results of the current classification model and the previous researches [17][13][12] [16][14] are depicted in Table 9 below.

Researchers	Data size	Features used	Classifiers	Performance %
Roorkiwal et al (2016) [17]	320 elite breeding lines from ICSN and ICRISAT.	4 features	RR-BLUP, BayesCπ, Bayes B, RF, BL.	The maximum accuracy was 85.1% using Random Forest.
Rachmatia H, Kusuma WA, and Hasibuan LS (2017) [13]	2283 genotype of maize dataset from CIMMYT’s	8 features	DBN, RKHS, BL, BLUP	The highest accuracy was 57.9% using DBN
González Camacho et al. (2016) [12]	16 maize and 17 wheat genomic and phenotypic datasets.	15 features	MLP, PNN	The highest accuracy was 0.746 using PNN
W. Ma et al. (2017) [16]	2000 Iranian bread wheat genotype data fromCIMMYT gene bank.	8 features	DeepGS, CNN, RR-BLUP	DeepGS achieved the highest accuracy of 65.24%
Hoffstetter et al. (2016) [14]	470 soft winter wheat lines and 4,858 genotype data.	4 features	RRBLUP, BL, RF	The highest accuracy was 63% using RF
Current Study	**8303 chickpea genotype data from DZARC.**	**8 features**	**ANN, SVMs, Decision Tree**	**Decision Tree gives the highest accuracy of 97.5%**

VII. CONCLUSION AND FUTURE WORK

Chickpea was known in various parts of the country and most important pulse crop in the world, and adapted in cooler agro-ecological environments, which are located in the central highlands of Ethiopian. It plays an important role in human nutrition and consumed in Ethiopia in different preparations like snacks, curry, green pea and blend to bread/Enjera powder and known for its high market values in Ethiopia. Ethiopia is the leading producer, consumer, and seller of chickpea in Africa, and is among the top ten most important producers in the world. The development of chickpea type classification model can support the selection process of chickpea varieties, which is the main procedure of chickpea production. Ethiopian chickpea type classification was done manually and it spends more resource and time. Domain experts tried to recognize every chickpea type, the way and efficiency of identifying each chickpea varieties are mainly depending on the skills and experience of experts in the domain area. Sometimes the identification of chickpea verities and type may not well identify and even if expertise is highly-skilled they may face challenges and exposed for bias. Most of the identification researches were done globally; for local and emerging varieties of Ethiopia, there is a need to design classification model that assists the selection and identification mechanisms of chickpea. In this study, an effort has been made to design an optimal model for the classification of chickpea types. The primary objective of this paper is to design a model for classifying chickpea type using machine learning algorithm. To this end, this study follows an experimental type, which involves data collection, preparation for training and evaluating chickpea type prediction model. A total of 8303 records and 8 features were used to predict chickpeatype. In addition, data partitioning were employed that is training and testing set, using percentage split technique, 80% (6656) of the data are used for training and 20% (1647) of data used for testing. To building the model, we were used machine learning algorithms such as artificial neural network, support vector machine and decision tree and for evaluating the models we are used confusion matrix rule with Accuracy, Sensitivity (Recall), precision and specificity. In this study, an experiment was conducted using an artificial neural network, support vector machine and decision tree.

The experimental result shows that accuracy achieved for artificial neural network, support vector machine and decision tree is 53%, 88.8%, and 97.5% respectively. In addition, this research work has got promising results and based on performance evaluator the best classification algorithms were decision tree classifier which is 97.5% accuracy. Finally, the main contribution of this study for domain experts, research community and the whole population is design chickpea type classification model from chickpea using machine learning algorithm to assist chickpea type selection and identification mechanism. Another important contribution of this study is collecting and preparing chickpea genotype dataset based on the feedback from domain experts on agricultural research institutes in order to help other researchers in conducting related studies to handle data problems.In the future works, the researchers will cooperation with agriculture research institutes to develop an expert system for genotyping and phenotyping of practical crops. In addition, the researchers will cooperate with agriculture research centers to carryout practical real applications of the model in the genomic selection and phenotyping of chickpea and similar crops.

REFERENCES

1. L. Korbu, T. Damte and A. Fikre, "Harnessing Chickpea Value Chain for Nutrition Security and Commercialization of Smallholder Agriculture in Africa," TLIII Project, ICRISAT, 2016.
2. Authority, Central Statistics, "Agricultural Sample Survey Report on Area and Production of Crops Private Peasant Holdings, Meher Season. Addis Ababa, Ethiopia," 2004.
3. R. Gonzlez, R. Oscar, G. JM and D. Gianola, "Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits," Livestock Science, vol. 166, pp. 217-231, 2014.
4. L. Ornella, P. Perez, E. Tapia, J. Gonzalez-Camacho, J. Burgueno, X. Zhang, S. Singh, F. Vicente, D. Bonnett and S. Dreisigacker, "Genomic-enabled prediction with classification algorithms," Heredity, vol. 112, p. 616, 2014.

5. B. Hayes and M. Goddard, "Prediction of total genetic value using genome-wide dense marker maps," Genetics, vol. 157, pp. 1819-1829, 2001.
6. J. A. Bhat, S. Ali, R. K. Salgotra, Z. A. Mir, S. Dutta, V. Jadon, A. Tyagi, M. Mushtaq, N. Jain and P. K. Singh, "Genomic selection in the era of next generation sequencing for complex traits in plant breeding}," Frontiers in genetics, vol. 7, p. 221, 2016.
7. Z. Desta, Abera and R. Ortiz, "Genomic selection: genome-wide prediction in plant improvement," Trends in plant science, vol. 19, pp. 592-601, 2014.
8. E. Jonas, K. de and J. Dirk, "Does genomic selection have a future in plant breeding?}," Trends in biotechnology, vol. 31, pp. 497-504, 2013.
9. J. Poland and J. Rutkoski, "Advances and challenges in genomic selection for disease resistance," Annual review of phytopathology, vol. 54, pp. 79-98, 2016.
10. J. Jannink, A. J. Lorenz and H. Iwata, "Genomic selection in plant breeding: from theory to practice," Briefings in functional genomics, vol. 9, pp. 166-177, 2010.
11. X. Yu, X. Li, T. Guo, C. Zhu, Y. Wu, S. E. Mitchell, K. L. Roozeboom, D. Wang, M. L. Wang and G. A. Pederson, "Genomic prediction contributing to a promising global strategy to turbocharge gene banks," Nature plants, vol. 2, p. 16150, 2016.
12. C. Gonzalez, J. Manuel, J. Crossa, R. P. Perez, L. Ornella and D. Gianola, "Genome-enabled prediction using probabilistic neural network classifiers," BMC genomics, vol. 17, p. 208, 2016.
13. R. H, K. W and H. L, "Prediction of maize phenotype based on whole-genome single nucleotide polymorphisms using deep belief networks," in Journal of Physics: Conference Series, vol. 835, IOP Publishing.
14. A. Hoffstetter, A. Cabrera and M. a. S. C. Huang, "Optimizing training population data and validation of genomic selection for economic traits in soft winter wheat," G3: Genes, Genomes, Genetics, pp. 2919-2928, 2016.
15. J. Crossa, D. Jarquin, J. Franco, P. Perez-Rodriguez, J. Burgueno, C. Saint-Pierre, P. Vikram, C. Sansaloni, C. Petroli and D. Akdemir, "Genomic prediction of gene bank wheat landraces," G3: Genes, Genomes, Genetics, pp. g3-116, 2016.
16. W. Ma, Z. Qiu, J. Song, Q. Cheng and C. Ma, "DeepGS: Predicting phenotypes from genotypes using Deep Learning," bioRxiv, pp. 241-414, 2017.
17. M. Roorkiwal, A. Rathore, R. R. Das, M. K. Singh, A. Jain, S. Srinivasan, P. M. Gaur, B. Chellapilla, S. Tripathi and Y. Li, "Genome-enabled prediction models for yield related traits in chickpea," Frontiers in plant science, vol. 7, p. 1666, 2016.
18. M. B. Kursa and W. R. Rudnicki, "Feature selection with the Boruta package," J Stat Softw, vol. 36, pp. 1-13, 2010.
19. M. B. J. A. Kursa and W. R. Rudnicki, "Boruta-a system for feature selection," Fundamental Informatics, vol. 101, pp. 271-285, 2010.
20. W. R. Rudnicki, Wrzesie and M. P. W, "All relevant feature selection methods and applications," in Feature Selection for Data and Pattern Recognition, Springer, 2015, pp. 11-28.
21. I. a. B. M. Kaastra, "Designing a neural network for forecasting financial and economic time series," Neurocomputing, vol. 10, pp. 215-236, 1996.
22. S. Aruna and S. a. N. L. Rajagopalan, "An empirical comparison of supervised learning algorithms in disease detection," International Journal of Information Technology Convergence and Services, vol. 1, pp. 81-92, 2011.
23. K. Durgesh and B. SRIVASTAVA and Lekha, "Data classification using support vector machine," Journal of theoretical and applied information technology, vol. 12, pp. 1-7, 2010.
24. Han, Jiawei; Pei, Jian and Kamber, Micheline;, Data mining: concepts and techniques, Elsevier, 2011.
25. B. L, "The Role of Data Mining Technology In Electronic Transaction Expansion At Dashen Bank S.C.MSc Thesis, Addis Ababa University," 2011.

AUTHORS PROFILE

Nigus Asres Ayele: His MSc is form Addis Ababa Science and Technology University since 2020; He has BSc Degree from Wolkite University since 2015. His Research Interest areas are Artificial intelligence, Machine learning, Deep learning, Image processing, Big data Analytics, Data science, Bioinformatics, Software testing and quality assurance, Network security and Software defined network.

Hailemichael Kefie Tamiru received his BSc and MSc degrees in Software Engineering, from Adama Science and Technology University, Ethiopia in 2013and 2017, respectively. His research interest areas are, Artificial Intelligence, Machine leering, Data science and Software Testing and Quality Assurance.