Workshops at 18th International Conference on Intelligent Environments (IE2022)
H.H. Alvarez Valera and M. Luštrek (Eds.)

doi:10.3233/AISE220031

Machine Learning to Predict Rapeseed Traits from RNA-seq Data

Md Ayshik RAHMAN KHAN [a,1], Khandaker Asif AHMED [b] and Md Zakir HOSSAIN [b,c]

[a] *La Trobe University, Melbourne, VIC 3083, Australia*
[b] *Commonwealth Scientific and Industrial Research Organisation, ACT 2601, Australia*
[c] *Biological Data Science Institute, College of Science, Australian National University, ACT 2601, Australia*

Abstract. Recent advances in the field of genomic trait prediction has paved the way for developing futuristic plant breeding programs. The objective of our study is to predict a single or multiple traits of rapeseed (*Brassica napus*) based on the RNA sequence data. We analyzed 12 different traits of rapeseed and evaluated how their pair-wise correlation impact on the yield production. Further, for predicting single or multi-traits of rapeseed, four state-of-art machine learning (ML) models, namely - Lasso Regression (Lasso), Random Forest (RF), Support Vector Machine (SVM) and Multi-layer Perceptron (MLP) were evaluated. For both single and multi-trait predictions, our RF and SVM models performed most consistently, where the lowest mean squared error was achieved by RF (0.045 and 0.016 for the single and multi-trait prediction respectively). A comparative analysis with related works showed the potentiality of our model for future multi-modal model development. Future study in this context could comprise of evaluating our models with other transcriptome dataset from related crops or deep learning-based methods for better outcomes.

Keywords. Trait Prediction, RNA-seq, Machine Learning, Rapeseed

1. Introduction

Rapeseed (*Brassica napus*), aka Canola, is the second most important oilseed crop around the world, widely used for oil production and livestock feed [1, 2]. Different usages of Rapeseed include biodiesel production, maintaining soil fertility, producing compound animal feed, forage for cattle, etc. In Europe, more than 80% of biodiesel production is done from rapeseed [3]. Canada, producing approximately one-fourth of the total rapeseed worldwide, is the largest rapeseed producer [4]. Australia is widely known for it's high quality exports of rapeseed. In the year 2020/21, Australia exported 3.132 million tonnes of rapeseed. [5]. Therefore, rapeseed significantly contributes not only as a food element, but also as a patron in the economy of many countries [6].

[1]Corresponding Author: Khandaker Asif Ahmed; E-mail: khandakerasif.ahmed@csiro.au, Md Zakir Hossain; E-mail: zakir.hossain@{anu.edu.au; csiro.au}

Due to its diverse usages, it has caught the eye of many researchers and multiple studies have been conducted on different phenotypic traits of rapeseed and how these traits impact yield production. A recent study by Xu et al. [7] showed that multilocular trait is exemplary for rapeseed breeding. According to Zhao et al. [8], seeds per pod played a vital role in rapeseed yield. It is also evident that high yield can be achieved through high-density planting. Experiments by Kuai et al. [9] showed that even though high density planting initially resulted in lower yields per plant, combined with increased row spacing the yield increased as well. They achieved the highest yield With the increased row spacing of 15 cm and a plant density of 45×104 plants ha^{-1}. Ali et al. [10] found positive and significant correlation between seed yield and harvest index, seed weight and flower duration in winter rapeseed varieties.

Transcriptome analysis has been proved to be very efficient for experiments such as trait prediction. Transcriptome is the snapshot of the gene expression of a cell or tissue in a specific moment [11]. Transcriptome sequencing or RNA sequencing method has an edge over the traditional approaches because of the low throughput and poor cost efficiency [12]. Canales et al. [13] performed a transcriptome analysis to examine the trade-off between seed weight and seed number and showed the impact of those traits on rapeseed yield production. Luo et al. [14] used transcriptome analysis to enumerate plant immunity trait against fungal infection, in respect to growth and defence. Azodi et al. [15] used transcriptome data for gemonic prediction of 3 important traits in maize - flowering time, height, and grain yield and in many cases they found transcriptome data to be performing similarly to the genotype-based models. Even though there are numerous transcriptomic studies focused of different economically important traits, there is indeed scarcity of machine learning (ML) model for single or multi-trait prediction tasks.

In recent years, ML techniques are very popular among researchers for predictive analysis or works in the similar domain. Kong et al. [16] used hyper-spectral imaging to detect Sclerotinia stem rot, which is one of the major diseases of rapeseed crops, causing great loss in yield production [17]. Przybyl et al. [18] used numerous deep-learning models to detect fungal contamination in rapeseed with a classification error ranging from 14% to 21%. According to Wei et al. [19], sophisticated ML approaches with a large ensemble of markers can provide improved disease risk assessment.Recent advances of statistical methods and machine learning has enabled determination of complex patterns in high dimensional settings [20] which helps with the analysis of traits and making predictions. Kurtulmus and Unal [4] used computer vision and ML to classify seven different varieties of rapeseed. Nitze et al. [21] conducted an experiment on crop type classification, an application on remote sensing, and did a comparative analysis on the used ML methods. Even though their classification accuracy varied greatly, they achieved the best results for the classification of rapeseed (over 90% for all the methods). ML techniques have been proved to be exceptionally efficient for problems such as trait prediction and classification. Traditional methods for genome based predictions such as shrinkage or regularization caused major over-fitting problems and dimensionality issues. ML models can counter these issues in effective way [22]. Azodi et al. [23] conducted an experiment which involved 18 traits of 6 different plant species. They used 6 linear and 6 non-linear ML algorithms for genomic prediction. Their experiments showed that different algo-

rithms performed better on different species and different traits. However, a generalized prediction model is yet to be validated [24].

The objective of our study was to predict different traits of rapeseed from RNA-seq data, and transform the knowledge towards high yield production. Beside correlation study among different traits, we implemented and evaluated four different ML algorithms to predict the traits. We also used weighted averaging for feature combination to perceive the impact of other features on the yield production. Mean squared error(MSE) and standard error were calculated for each trait and each method and the most promising results were discussed compared to other existing methods. Current study is a baseline for future multi-modal model for rapeseed trait prediction or models for other crops.

2. Methods

2.1. Dataset

We adopted the dataset from [13], which is a detailed transcriptome study of different rapeseed plants. The RNAseq RAW count matrix and sample trait metadata are collected from online[2]. The trait dataset included 12 different phenotypic traits including - Plant height (cm), Branch per plant, Siliques per plant, Yield (mg/ silique), Yield (kg/ ha), Seed number (n/ silique), Seed number (n/ m^2), Thousand Seed Weight (g), Seed oil (%), Protein (%), Total Biomass (kg/ m^2) and Harvest index (%). Total biomass refers to the volume of habitat, which is the weight of living plant material that are above or below the ground surface at a specific point of time. Harvest index is the weight of a harvested product, measured as a percentage of the total plant weight. The seed oil and the protein percentage refers to their concentration which can be obtained from the seed or the plant. Even though the raw count matrix had 48 bio-samples, sample metadata was available only for 24 bio-samples. Therefore, we subset the count matrix and match with the metadata.

2.2. Statistical Analysis of the Traits

A comprehensive statistical analysis was performed on the trait dataset. Firstly, all trait data were individually plotted to visualize the data distribution. Normality of individual traits were measured using Shapiro-Wilk test. Then a correlation matrix was generated for all traits. Correlation determine the relationship between two variables and how strongly they are related [25]. As we were particularly interested on rapeseed yield production, correlation coefficient was subset for each traits in respect to Yield (mg/ silique). Furthermore, a linear regression model was fitted, by considering Yield (mg/ silique) as dependent variable, and all other traits as independent variable to measure how the Yield trait can be impacted by other traits.

[2]Dataset: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE169511

2.3. RNASeq Data Preprocessing and Evaluating ML Models

For the trait prediction, a Min-Max normalization was applied column-wise to the RAW count data to ensure that the data is scaled to a specific range (0 to 1) and to maintain the relation among the original data [26]. Principal Component Analysis (PCA) was performed for dimension reduction as PCA provides a simpler view of the data while preserving most of the information [27].

Normalised count matrix was used as the input, and from the input, we made predictions for individual or multiple traits using the ML methods. We used four different ML methods namely, Lasso Regression (Lasso), Random Forest (RF), Support Vector Machine (SVM) and Multi-layer Perceptron (MLP). The experiment was conducted in two phases. In the first phase, our model evaluated each trait individually. On the second phase of the experiment, a multi-trait evaluation was implemented on all the traits combined. The feature combination was carried out by weighted averaging.We selected MSE as the evaluation metrics. The reason for choosing MSE as the evaluation metrics is its simplicity of implementation and small computational complexity [28]. After calculating MSE, we also cross-validated our result. Cross-validation is a type of data resampling method that is used for the estimation of prediction error of the ML models [29]. It is also used for preventing and detecting overfitting of data [30]. However, a rigorous use of cross-validation may also be the reason of overfitting [31]. Amongst different cross-validation techniques, we have used 10-fold cross-validation. Further, Standard error of measurement shows the approximate standard deviation of a test score for a specific group of takers [32]. Standard error allows the comparison between the estimated populations to be intuitive via graphs or tables [33].

For the multi-trait prediction, we have used weighted averaging method for the combination of the features and calculated a single MSE for that combined feature. The weight used for averaging were approximately equal (0.083 for all except Yield(mg/silique), which had a weight of 0.087). For weighted averaging, the sum of the weights of the traits requires to be 1. Therefore, keeping the weight of Yield (mg/ silique) slightly higher than the others, while the weight for the other traits were divided equally[3].

3. Results and Discussion

3.1. Data Distribution and Variances

Data distribution is used to characterize quantitative variation of original data [34]. Our statistical tests showed, most of the traits data (except for Yield (mg/ silique) and SN (n/ silique)) were normally distributed. Figure 1 shows the data distribution of all twelve traits. A correlation matrix was also generated to understand the pair-wise correlation among different traits. A correlation heatmap is a two-dimensional graphical representation of correlation matrix. Figure 2 shows the correlation heatmap of the traits. The light colored boxes indicate stronger correlation and the dark colored boxes indicate weaker correlation. Correlation coefficient of each traits against Yield (mg/ silique), showed -

[3]Github repository: https://github.com/Ayshikrk/Trait_Prediction_1.git

Branch per plant, Siliques per plant, and TSW (g) were negatively correlated to yield while the rest of the traits were positively correlated.

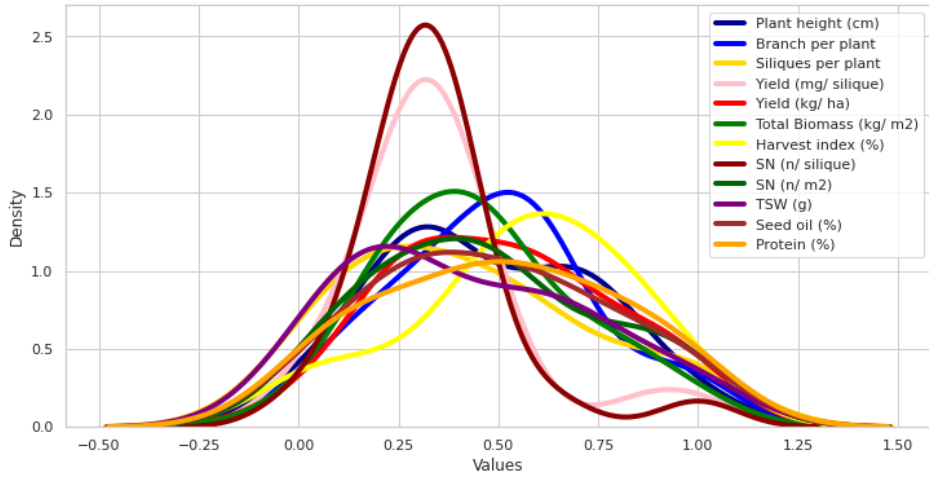

Figure 1. Data Distribution of Traits

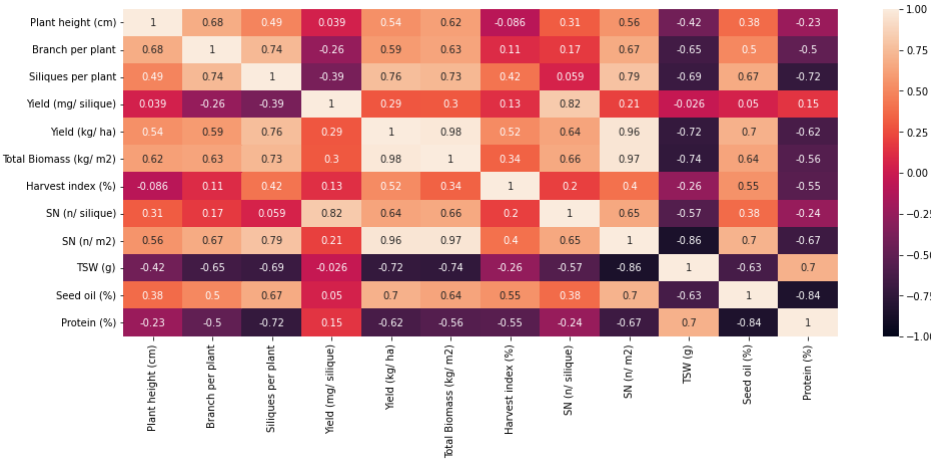

Figure 2. Correlation Heatmap

Further, we performed a regression analysis on the trait dataset and generated an Ordinary Least Squares (OLS) regression table as shown in Table 1. In this analysis, Yield (mg/ silique) was the dependent variable, while the rest of the traits were the independent variable. The intercept of Yield (mg/ silique) had a coefficient value of -0.3126 with a standard error of 0.233. The R-squared value of the OLS regression was 0.991 and F-statistics value was 116.5. We found SN(n/silique) independently have a statistically significant relationship (p=0.007) with Yield (mg/ silique), followed by TSW (p=0.05).

Table 1. OLS Regression Results

Traits	Coefficient	Standard Error	t	P-Value
Plant height (cm)	-0.066	0.44	-1.473	0.166
Branch per plant	-0.0233	0.062	-0.360	0.725
Siliques per plant	0.297	0.422	0.703	0.495
Yield (kg/ ha)	-0.827	1.443	-0.573	0.577
Total Biomass (kg/ m^2)	1.346	1.196	1.125	0.283
Harvest index (%)	0.244	0.261	0.937	0.367
SN (n/ silique)	1.707	0.529	3.229	0.007
SN (n/ m^2)	-0.943	0.338	-2.789	0.16
TSW (g)	0.302	0.143	2.113	0.056
Seed oil (%)	-0.022	0.055	-0.402	0.694
Protein (%)	0.0009	0.051	0.018	0.986

Table 2. MSE for all the traits with Standard Error

Traits	Lasso	RF	SVM	MLP
Plant height (cm)	0.131 ±0.02	0.102 ±0.025	0.089 ±0.021	0.132 ±0.029
Branch per plant	0.091 ±0.023	0.078 ±0.022	0.071 ±0.021	0.07 ±0.029
Siliques per plant	0.16 ±0.047	0.09 ±0.032	0.096 ±0.031	0.095 ±0.036
Yield (mg/ silique)	0.075 ±0.044	0.074 ±0.044	0.074 ±0.044	0.095 ±0.057
Yield (kg/ ha)	0.105 ±0.023	0.081 ±0.017	0.082 ±0.018	0.09 ±0.018
Total Biomass (kg/ m^2)	0.106 ±0.03	0.83 ±0.023	0.01 ±0.022	0.08 ±0.023
Harvest index (%)	0.109 ±0.021	0.101 ±0.023	0.093 ±0.022	0.157 ±0.057
SN (n/ silique)	0.057 ±0.031	0.045 ±0.026	0.049 ±0.031	0.067 ±0.03
SN (n/ m^2)	0.08 ±0.021	0.098 ±0.023	0.104 ±0.024	0.1 ±0.022
TSW (g)	0.112 ±0.023	0.112 ±0.03	0.119 ±0.027	0.131 ±0.026
Seed oil (%)	0.075 ±0.028	0.091 ±0.027	0.084 ±0.026	0.104 ±0.033
Protein (%)	0.15 ±0.052	0.112 ±0.046	0.12 ±0.039	0.151 ±0.044

3.2. Performance of ML Models

Firstly, we calculated MSE for individual traits using RNA-seq data. The results were cross validated using 10-fold cross-validation. We have also calculated the standard error for the cross-validated results. Table 2 shows the cross-validated MSE for all the traits, along with the standard error of them for the four different methods.

Amongst the four ML methods used for the single trait prediction, Lasso and MLP had comparatively higher MSE than RF and SVM. On the other hand, SVM and RF maintained a consistency in their result for most of the traits. From Table 2, it can be discerned that all the methods performed the best for the trait SN (n/ silique). For this trait, Lasso, RF, SVM, and MLP had MSE of 0.057, 0.045, 0.049, and 0.067 respectively. However, on the other hand, the worst performing traits were different for different methods. Lasso and SVM exhibited highest MSE for protein (%). They had a MSE of 0.15 and 0.12 respectively. Highest MSE for RF was 0.112, which was seen for two traits -

protein (%) and TSW (g), where MLP showed its worst MSE (0.157) for harvest index (%).

Finally, we focused to predict multi-traits using RNA seq dataset. Weighted averaing was used for feature combination. The result was cross validated and standard error was also calculated. Both the outcomes of MSE and standard error of the multi-trait prediction had shown a notably better output than our single trait prediction. The MSE reduced significantly for all the methods in the case of multi-trait prediction (both RF and SVM had MSE of 0.016). We also tried to adjust the weight of the traits based on the p-values, showed in Table 1. However, the approach didn't improve the outcomes much. The overall outcome for multi-trait prediction is illustrated in Figure 3.

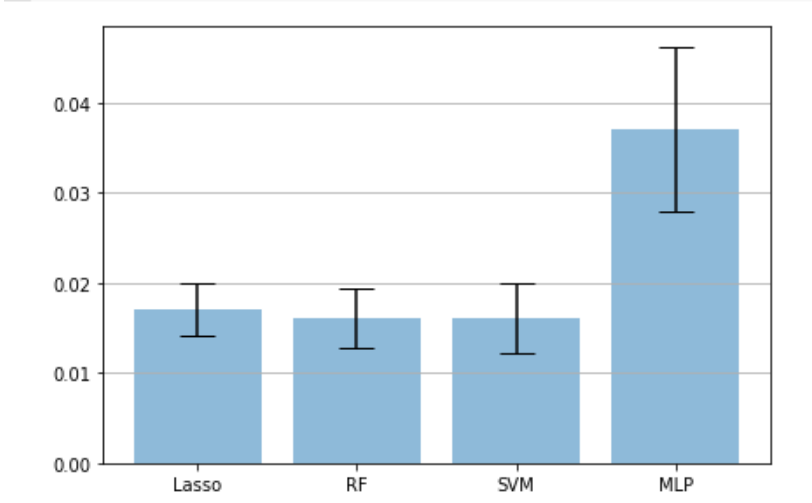

Figure 3. MSE for multi-trait prediction

3.3. Comparative Analysis

A few research works have been seen in recent years where different ML or deep learning methods were used for trait prediction. Zingaretti et al. [35] conducted an experiment similar to ours where they proposed a generalized deep learning approach for single trait prediction from wheat data. They used MLP model on non-normalised data and got a MSE of 1.518 which is quite higher than ours. Another notable experiment was conducted by Riley Mcdowell [36] on multiple traits of three different datasets (Arabidopsis, maize and wheat). Even though the study utilised numerous ML and neural network-based models, most of the models performed poorly (accuracy as low as 15%) and none of the models showed consistent performance. However, the regularized neural network methods performed slightly better than the unregularized methods. Therefore, regularization was recommended. It is also observable that similar to our experiment, Lasso did not perform too well in this experiment. A similar eventuality is observed in the experiments of liu et al [37]. Their experiment on the grain yield trait on soybean and stem height trait on loblolly pine dataset showed that Convolutional Neural Network (CNN)

outperformed lasso models. According to Washburn et al. [38], deep neural networks show promise in the field of trait predictions. However, in our experiment, MLP did not perform as good as RF or SVM. Even though the study by Bellot et al. [39] was done genetic traits of humans, their study also showed that the results obtained from MLP were not consistent as they largely depended on SNPs and phenotypes. So, for future model development, a well-annotated and trustworthy dataset is a pre-requisite.

4. Conclusion

Trait prediction models can play a vital role in the field of plants and breeding. Use of ML or deep learning methods for trait prediction has manifested satisfactory results and they are already outperforming the traditional methods. In our study, we evaluated different traits of Rapeseed and the relationship among different traits in respect to yield. Our study showed seed number per silique is a significant trait for yield production. Therefore, considering its positive correlation with yield, increasing the number of seed per silique can increase the yield of rapeseed. Our Prediction models, namely Lasso, RF, SVM and MLP, showed low MSE scores for predicting single or multi-traits of rapeseed. While we got low MSE score from RF and SVM, the results from Lasso and MLP were a little inferior in comparison. These models will be beneficial in terms of real-life return of experience, where yield or seed characteristics can be predicted easily from their molecular-based data. Even though we took an unique approach of multi-trait prediction, scope of further research still remains. Beside the four ML models, different ensemble ML methods can be applied in future studies. To fine-tune our models for other crops, different transcriptome datasets could be taken into consideration. Trait heritability was not considered in our experiments, which may have some level of impact on the outcome. Therefore, the experiments could be designed while taking the trait heritability into consideration for better outcomes.

Acknowledgment

We express our gratitude to Biological Data Science Institute (BDSI) at Australian National University for their generous support.

References

[1] Stewart D, Shepherd LVT. 9 - Metabolomics for the safety assessment of genetically modified (GM) crops. In: Weimer BC, Slupsky C, editors. Metabolomics in Food and Nutrition. Woodhead Publishing Series in Food Science, Technology and Nutrition. Woodhead Publishing; 2013. p. 192-216.

[2] Friedt W, Tu J, Fu T. In: Liu S, Snowdon R, Chalhoub B, editors. Academic and Economic Importance of *Brassica napus* Rapeseed. Cham: Springer International Publishing; 2018. p. 1-20.

[3] Bajpai D, Tyagi V. Biodiesel: source, production, composition, properties and its benefits. Journal of OLEo science. 2006;55(10):487-502.

[4] Kurtulmuş F, Ünal H. Discriminating rapeseed varieties using computer vision and machine learning. Expert Systems with Applications. 2015;42(4):1880-91.

[5] Australian Oilseeds Federation;. Available from: http://www.australianoilseeds.com/oilseeds_industry/industry_facts_and_figures.

[6] Ma N, Yuan J, Li M, Li J, Zhang L, Liu L, et al. Ideotype population exploration: growth, photosynthesis, and yield components at different planting densities in winter oilseed rape (*Brassica napus* L.). PloS one. 2014;9(12):e114232.
[7] Xu P, Wang X, Dai S, Cui X, Cao X, Liu Z, et al. The multilocular trait of rapeseed is ideal for high-yield breeding. Plant Breeding. 2021;140(1):65-73.
[8] Zhao H, An F, Du D. New idioplasmic resource *B. napus* L. with multi-loculus founded by interspecific hybridization. In: Proceedings of the 12th International Rapeseed Congress, Wuhan. vol. 3; 2007. p. 294-5.
[9] Kuai J, Sun Y, Zuo Q, Huang H, Liao Q, Wu C, et al. The yield of mechanically harvested rapeseed (*Brassica napus* L.) can be increased by optimum plant density and row spacing. Scientific reports. 2015;5(1):1-14.
[10] Ali N, Javidfar F, Elmira JY, Mirza M. Relationship among yield components and selection criteria for yield improvement in winter rapeseed (*Brassica napus* L.). Pak J Bot. 2003;35(2):167-74.
[11] Ward JA, Ponnala L, Weber CA. Strategies for transcriptome analysis in nonmodel plants. American journal of botany. 2012;99(2):267-76.
[12] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics. 2009;10(1):57-63.
[13] Canales J, Verdejo J, Carrasco-Puga G, Castillo FM, Arenas-M A, Calderini DF. Transcriptome Analysis of Seed Weight Plasticity in *Brassica napus*. International Journal of Molecular Sciences. 2021;22(9):4449.
[14] Luo J, Xia W, Cao P, Xiao Z, Zhang Y, Liu M, et al. Integrated transcriptome analysis reveals plant hormones jasmonic acid and salicylic acid coordinate growth and defense responses upon fungal infection in poplar. Biomolecules. 2019;9(1):12.
[15] Azodi CB, Pardo J, VanBuren R, de Los Campos G, Shiu SH. Transcriptome-based prediction of complex traits in maize. The Plant Cell. 2020;32(1):139-51.
[16] Kong W, Zhang C, Cao F, Liu F, Luo S, Tang Y, et al. Detection of sclerotinia stem rot on oilseed rape (*Brassica napus* L.) leaves using hyperspectral imaging. Sensors. 2018;18(6):1764.
[17] Del Rio L, Bradley C, Henson R, Endres G, Hanson B, McKay K, et al. Impact of Sclerotinia stem rot on yield of canola. Plant disease. 2007;91(2):191-4.
[18] Przybył K, Wawrzyniak J, Koszela K, Adamski F, Gawrysiak-Witulska M. Application of deep and machine learning using image analysis to detect fungal contamination of rapeseed. Sensors. 2020;20(24):7305.
[19] Wei Z, Wang K, Qu HQ, Zhang H, Bradfield J, Kim C, et al. From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. PLoS genetics. 2009;5(10):e1000678.
[20] de Los Campos G, Vazquez AI, Hsu S, Lello L. Complex-trait prediction in the era of big data. Trends in Genetics. 2018;34(10):746-54.
[21] Nitze I, Schulthess U, Asche H. Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. Proceedings of the 4th GEOBIA, Rio de Janeiro, Brazil. 2012;79:3540.
[22] González-Recio O, Rosa GJ, Gianola D. Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. Livestock Science. 2014;166:217-31.
[23] Azodi CB, Bolger E, McCarren A, Roantree M, de Los Campos G, Shiu SH. Benchmarking parametric and machine learning models for genomic prediction of complex traits. G3: Genes, Genomes, Genetics. 2019;9(11):3691-702.
[24] Pasolli E, Truong DT, Malik F, Waldron L, Segata N. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. PLoS computational biology. 2016;12(7):e1004977.
[25] Taylor R. Interpretation of the correlation coefficient: a basic review. Journal of diagnostic medical sonography. 1990;6(1):35-9.
[26] Patro S, Sahu KK. Normalization: A preprocessing stage. arXiv preprint arXiv:150306462. 2015.
[27] Paul LC, Suman AA, Sultan N. Methodological analysis ofprincipal component analysis (PCA) method. International Journal of Computational Engineering & Management. 2013;16(2):32-8.
[28] Wang Z, Bovik AC. Mean squared error: Love it or leave it? A new look at signal fidelity measures. IEEE signal processing magazine. 2009;26(1):98-117.
[29] Berrar D. Cross-Validation. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. Encyclopedia of Bioinformatics and Computational Biology. Oxford: Academic Press; 2019. p. 542-5.

[30] Santos MS, Soares JP, Abreu PH, Araujo H, Santos J. Cross-validation for imbalanced datasets: avoiding overoptimistic and overfitting approaches [research frontier]. ieee ComputatioNal iNtelligeNCe magaziNe. 2018;13(4):59-76.
[31] Moore AW. Cross-validation for detecting and preventing overfitting. School of Computer Science Carneigie Mellon University. 2001.
[32] Harvill LM. Standard error of measurement: an NCME instructional module on. Educational Measurement: issues and practice. 1991;10(2):33-41.
[33] Lee DK, In J, Lee S. Standard deviation and standard error of the mean. Korean journal of anesthesiology. 2015;68(3):220.
[34] Limpert E, Stahel WA. Problems with using the normal distribution–and ways to improve quality and efficiency of data analysis. PloS one. 2011;6(7):e21403.
[35] Zingaretti LM, Gezan SA, Ferrão LFV, Osorio LF, Monfort A, Muñoz PR, et al. Exploring deep learning for complex trait genomic prediction in polyploid outcrossing species. Frontiers in plant science. 2020;11:25.
[36] McDowell R. Genomic selection with deep neural networks; 2016.
[37] Liu Y, Wang D. Application of deep learning in genomic selection. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2017. p. 2280-0.
[38] Washburn JD, Burch MB, Franco JAV. Predictive breeding for maize: Making use of molecular phenotypes, machine learning, and physiological crop models. Crop Science. 2020;60(2):622-38.
[39] Bellot P, de Los Campos G, Pérez-Enciso M. Can deep learning improve genomic prediction of complex human traits? Genetics. 2018;210(3):809-19.