



# NeuralLasso: Neural Networks Meet Lasso in Genomic Prediction

Boby Mathew<sup>1,2</sup>, Andreas Hauptmann<sup>3,4</sup>, Jens Léon<sup>2</sup> and Mikko J. Sillanpää<sup>3\*</sup>

<sup>1</sup> Bayer CropScience, Monheim am Rhein, Germany, <sup>2</sup> Institute of Crop Science and Resource Conservation, University of Bonn, Bonn, Germany, <sup>3</sup> Research Unit of Mathematical Sciences, University of Oulu, Oulu, Finland, <sup>4</sup> Department of Computer Science, University College London, London, United Kingdom

## OPEN ACCESS

### Edited by:

Rodomiño Ortiz,  
Swedish University of Agricultural  
Sciences, Sweden

### Reviewed by:

Hao Cheng,  
University of California, Davis,  
United States  
Osva Antonio Montesinos-López,  
Universidad de Colima, Mexico

### \*Correspondence:

Mikko J. Sillanpää  
mikko.sillanpaa@oulu.fi

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 22 October 2021

**Accepted:** 18 March 2022

**Published:** 29 April 2022

### Citation:

Mathew B, Hauptmann A, Léon J and  
Sillanpää MJ (2022) NeuralLasso:  
Neural Networks Meet Lasso in  
Genomic Prediction.  
Front. Plant Sci. 13:800161.  
doi: 10.3389/fpls.2022.800161

Prediction of complex traits based on genome-wide marker information is of central importance for both animal and plant breeding. Numerous models have been proposed for the prediction of complex traits and still considerable effort has been given to improve the prediction accuracy of these models, because various genetics factors like additive, dominance and epistasis effects can influence of the prediction accuracy of such models. Recently machine learning (ML) methods have been widely applied for prediction in both animal and plant breeding programs. In this study, we propose a new algorithm for genomic prediction which is based on neural networks, but incorporates classical elements of LASSO. Our new method is able to account for the local epistasis (higher order interaction between the neighboring markers) in the prediction. We compare the prediction accuracy of our new method with the most commonly used prediction methods, such as BayesA, BayesB, Bayesian Lasso (BL), genomic BLUP and Elastic Net (EN) using the heterogenous stock mouse and rice field data sets.

**Keywords:** neural networks, LASSO, local epistasis, genomic selection, whole genome prediction

## INTRODUCTION

The introduction of Genomic Selection (GS) (Meuwissen et al., 2001) along with the availability of low cost genotyping platforms has resulted in a major paradigm shift in both animal and plant breeding. Since then, GS has been successfully applied for efficient selection and accelerating the breeding process in various breeding programs (Spindel et al., 2015; Garner et al., 2016; Hickey et al., 2017; Voss-Fels et al., 2019). Even though GS has now been widely implemented in practice, still considerable effort has been given to improve the prediction accuracy in GS beyond the current limits. Various factors can affect the prediction accuracy in GS including marker density, heritability of the trait, population size, constitution of the learning population and the statistical model used to predict the genomic breeding values (Meuwissen, 2009; Liu et al., 2018; Norman et al., 2018). Recently many studies tried to incorporate the transcriptome data (Li et al., 2019; Azodi et al., 2020) into genomic prediction models, in order to improve the prediction accuracy in GS.

The genomic prediction models can be divided roughly into two classes: (1) genomic best linear unbiased prediction (GBLUP) based on linear mixed models and (2) the whole-genome regression (WGR) based on multilocus regression models. In the first approach, the genetic background of the trait is assumed to be polygenic while in the latter, more oligogenic genetic background is assumed. Again in the first, molecular markers are used to construct the genomic relationship matrix while in the latter, molecular markers represent considered set of regression variables in the model. However, note that WGR model can be written also as the GBLUP model with a

trait-specific relationship matrix having own variance component for each SNP in the diagonal (Zhang et al., 2010; Piepho et al., 2012; Resende et al., 2012; Shen et al., 2013).

Epistasis (genetic interaction) is one of the major reason for the non-linearity in the genotype-phenotype relationship and considerable efforts have been given to model epistasis in genomic prediction models (Hu et al., 2011; Wittenburg et al., 2011; Wang et al., 2012; Jiang and Reif, 2015). Recently, many studies even pointed out the importance of local epistasis (interactions that span short segments of the genome) (Wei et al., 2014; Akdemir and Jannink, 2015; Akdemir et al., 2017; He et al., 2017; Liang et al., 2020). Although it is well known that epistasis (both local and global) interactions contribute to many complex traits (Taylor and Ehrenreich, 2014, 2015; Albert and Kruglyak, 2015), most of the genomic prediction models account for the pair-wise interactions due to the computational complexity of screening through all possible combinations.

Most of the WGR models used in GS are based on linear regression procedure and have been successfully adopted to predict complex phenotype in plant and animal breeding programs (Meuwissen et al., 2001; Park and Casella, 2008; Mathew et al., 2019). Nonlinear extensions of these methods with dominance and epistasis has been also considered (Nishio and Satoh, 2014; Jiang and Reif, 2015; Varona et al., 2018; Olatoye et al., 2019). However, recent development in the field of machine learning enable us to use profound nonlinear methods for the prediction of complex traits in breeding. Among the machine learning methods, deep learning (DL) methods received much attention due to their outstanding prediction properties (LeCun et al., 2015). Although improved accuracy can be questioned, many recent studies successfully applied deep learning for various genomic problems (Uppu et al., 2016; Bellot et al., 2018; Montesinos-López et al., 2018, 2019; Crossa et al., 2019; Liu et al., 2019; Pérez-Enciso and Zingaretti, 2019).

Often these learning methods are applied in a black-box manner and standard architectures that worked well in disciplines like natural language processing and computer vision are transferred to genomic prediction. Even though results are encouraging, interpretability remains an issue (Waldmann, 2018). However, as an exception, there is a study presenting an interpretable neural network model (see Zhao et al., 2021). Also, in this study we propose to design a domain specific learning system that is motivated by neural networks, but incorporates classical elements of lasso. The resulting algorithm is termed NeuralLasso, that is capable of incorporating higher order nonlinear interactions between contributing markers in the local neighborhood. Unlike the method of Zhao et al. (2021), our non-Bayesian approach is focusing on modeling high-order local interactions. In the terminology of neural networks, predictions are performed in a single layer and  $\ell_1$  sparsity on the learned parameters is incorporated, hence the relation to classical lasso models. We compare the prediction accuracy of NeuralLasso with the most commonly used GP methods such as BayesA, BayesB, BL, GBLUP, and EN using the mouse and rice data sets.

## MODELS AND METHODS

### Whole Genome Regression Model

Let us consider a standard genomic prediction model

$$y = X\beta + Zw + \epsilon. \quad (1)$$

Here,  $y$  is a vector of observed phenotypes for  $n$  lines,  $\beta$  contains the fixed effects,  $X$  represents the incidence matrix for the fixed effects,  $Z = Z_{i1}, Z_{i2}, \dots, Z_{ip}$  is the  $n \times p$  ( $p$  is the number of markers) matrix for the genotypes coded as 0,1,2,  $w = (w_1, w_2, \dots, w_p)$  is a column vector of marker effects and  $\epsilon$  corresponds to the residual, following a normal distribution as  $\epsilon \sim N(0, I\sigma_\epsilon^2)$ . For simplicity, here we assume no fixed effects other than overall mean (note that it is possible to pre-correct fixed effects away from the phenotype before neural network analysis).

The number of markers usually exceeds the number of observations in genomic prediction problems and regularization is applied in order to obtain solution to Equation (1). A regularized regression function can be formulated as

$$\hat{\beta}, \hat{w} = \operatorname{argmin}_{\beta, w} \left[ \sum_{i=1}^n (y_i - X\beta - \sum_{j=1}^p Z_{ij}w_j)^2 + P(\lambda, w) \right]. \quad (2)$$

Here, the function  $P(\lambda, w)$  is the penalty function with regularization parameter  $\lambda \geq 0$ . Least absolute shrinkage and selection operator (lasso) (Tibshirani, 1996) based on the penalty term called  $\ell_1$ -norm, which is the sum of the absolute coefficients and Ridge Regression based on the  $\ell_2$ -norm penalty which the sum of squared coefficients are the most commonly used regularized regression methods. The EN method which is a compromise between lasso and ridge regression penalties can be represented as:

$$\hat{\beta}, \hat{w} = \operatorname{argmin}_{\beta, w} \left\{ \sum_{i=1}^n (y_i - X\beta - \sum_{j=1}^p Z_{ij}w_j)^2 + \lambda \sum_{j=1}^p \left[ \frac{1}{2}(1 - \alpha)w_j^2 + \alpha|w_j| \right] \right\} \quad (3)$$

where  $0 \leq \alpha \leq 1$  is the penalty weight. The EN penalty is controlled by  $\alpha$  and when  $\alpha = 1$  EN is identical to lasso, whereas EN is equivalent to ridge regression when  $\alpha = 0$ .

### NeuralLasso

We design our model based on the underlying Equation (1). That is, given the genotypes in  $Z$ , instead of finding the matrix  $w$  we seek to find a parametrizable (nonlinear) mapping  $\Lambda_\theta$ , with parameters  $\theta$ , such that

$$\Lambda_\theta(Z) = y. \quad (4)$$

The question is how to construct such a mapping  $\Lambda_\theta$  and what do the parameters of  $\theta$  represent. In the following we aim to derive a model, that is motivated by neural networks, but follows the classical architecture of lasso as presented in Equation (3). For this purpose, we will shortly review classic neural network architectures.

### Background on Neural Networks

The underlying premise of a neural network is to combine affine linear mappings and pointwise nonlinearities to construct a nonlinear mapping in a repeating multi-layered fashion (LeCun et al., 2015; Schmidhuber, 2015; Goodfellow et al., 2016). In its most general form we can write the main building block of a neural network for an input  $z \in \mathbb{R}^n$  (genotypes) and output  $y \in \mathbb{R}^m$  (phenotype) as

$$y = \varphi(Cz + b), \tag{5}$$

where  $C \in \mathbb{R}^{m \times n}$  is a linear transformation matrix,  $b \in \mathbb{R}^m$  an additive affine component, the so-called bias, and finally  $\varphi(\cdot)$  a point-wise acting nonlinear function. A popular choice for this nonlinear function is given by the rectified linear unit  $\text{ReLU}(x) := \max(x, 0)$ . A multi-layered neural network would be now given as a repeated composition the blocks in Equation (5), where each block is called one layer. Nevertheless, in this work we concentrate on so-called shallow networks that consist only of one layer. The specific network architecture is now defined by the structure of the affine linear transformation  $C$ . The obvious choice of a dense matrix  $C$  is called a fully connected layer, as each data point in the input vector is related with each point in the output vector. Such a fully connected layer learns specific weights for each location in the input and hence is locally varying. Another option for the choice of linear mapping would be given by convolutions, if represented as matrices this would result in a sparse representation. Sparse representations are desirable, as they can be implemented efficiently and reduce the amount of parameters significantly. Nevertheless, the choice of convolutions as linear transformation is not optimal in our setting, as these are translationally invariant and hence do not encode any locality. In the following, we aim to design a transformation that is sparse, but does also encode locality to combine the strength of both.

### Formulating NeuralLasso

The first important part is to define the underlying transformation given as the matrix  $C$  for our proposed model is based on the requirement to encode locality, while taking neighborhood relationships into account. For this purpose, we follow (Arridge and Hauptmann, 2019) and define a sparse subdiagonal matrix  $C \in \mathbb{R}^{p \times p}$ , where  $p$  is the number of markers, and a neighborhood of size  $N$ , such that the main diagonal and the  $N$  subdiagonals below and above are non-zero. That is for  $N = 0$  we simply have a diagonal matrix and for  $N = 1$  we have a tridiagonal matrix such as

$$C = \begin{pmatrix} c_{0,1} & c_{1,1} & & & \\ c_{-1,2} & c_{0,2} & c_{1,2} & & \\ & \ddots & \ddots & \ddots & \\ & & c_{-1,n-1} & c_{0,n-1} & c_{1,n-1} \\ & & & c_{-1,n} & c_{0,n} \end{pmatrix}. \tag{6}$$

Given the matrix  $C$  we could formulate a lasso problem that takes interactions in the local neighborhood into account

by minimizing

$$\hat{C} = \arg \min_C \sum_{i=1}^n \left( y_i - \sum_{j=1}^p (CZ_i^T)_j \right)^2 + \lambda \sum_{j=-N}^N \sum_{i=1}^p |c_{j,i}|. \tag{7}$$

Note, that for  $N = 0$  no neighborhood relation is taken into account and the model reduces to the basic lasso scheme similar to Equation (3). As the above model in Equation (7) only considers linear interactions in the local neighborhood, we want to combine this sparse subdiagonal matrix with classical elements of neural networks, *i.e.*, nonlinear activation functions and additional bias vectors to allow for nonlinear interactions, as outlined previously.

### The Proposed Model for Local Epistatic Interactions

We will now consider the building block of a neural network as in Equation (5) for one layer, but consider multiplication with the subdiagonals of  $C$  separately to introduce nonlinear effects between neighboring loci. In the following, we will fix the neighborhood to  $N = 2$ , that is a neighborhood window of 5 loci. We will model the nonlinear interaction by a maximum thresholding using ReLU for the 3 central loci and no nonlinearity for the outer two loci. This way we enforce an interaction effect of the 5-neighborhood. Given the (sub)diagonal vectors  $c_i \in \mathbb{R}^p$  for  $i = -2, \dots, 2$  the non-linear parametrized model can be formulated as

$$\Lambda_C(Z_j) = \sum_{n=-2}^2 \sum_{i=1}^p \varphi_i(c_{n,i}z_{i+n} + b_{n,i}), \tag{8}$$

where  $z_i = 0$  for  $i < 1$  or  $i > p$ , and  $\varphi_i(x) = \text{ReLU}(x) = \max(x, 0)$  for  $i = -1, 0, 1$  and  $\varphi_i(x) = x$  otherwise. That is, if we write all terms down we get

$$\begin{aligned} \Lambda_C(Z_j) = & \sum_{i=1}^p (c_{-2,i}z_{i-2} + b_{-2,i}) + (c_{2,i}z_{i+2} + b_{2,i}) \\ & + \text{ReLU}(c_{-1,i}z_{i-1} + b_{-1,i}) + \text{ReLU}(c_{0,i}z_i + b_{0,i}) \\ & + \text{ReLU}(c_{1,i}z_{i+1} + b_{1,i}). \end{aligned} \tag{9}$$

The resulting NeuralLasso then formulates as

$$\{\hat{C}, \hat{b}\} = \arg \min_{\{C, b\}} \sum_{j=1}^n [y_j - \Lambda_C(Z_j)]^2 + \lambda \sum_{n=-2}^2 \sum_{i=1}^p (|c_{n,i}| + |b_{n,i}|). \tag{10}$$

The parameters  $\hat{C}$  and  $\hat{b}$  can then be found by any suitable optimisation algorithm. ReLU functions were chosen, here, because of their ability to keep some of the linearity and introducing nonlinearity only by thresholding. Note that if all the ReLU activations are changed to linear functions then the model reduces to a sparse perceptron with biases (*i.e.*, a single-layer neural network), which will be an overparametrized version of the lasso approach. We will shortly discuss our implementation in the next section.

For the final estimation, we are only left with estimating the penalty weight  $\lambda$  as in the classic lasso model. This can be

achieved in a similar manner as used by Waldmann et al. (2019), here, we use a slightly modified bisection method and a single set of training data *i.e.*, one realization of training and validation split. We then initialize a starting interval  $[a, b]$  for  $\lambda$ , chosen based on prior knowledge for the range of  $\lambda$ . We then compute the correlation coefficient for  $\lambda = a, b$ , *i.e.*, the end points of the interval  $[a, b]$ , and the mid point  $\lambda = a + b/2$ . Then, we identify the value for  $\lambda$  with the largest correlation coefficient. If it is one of the end points, we shift the interval around the end point, which becomes the new mid point. If, otherwise, the mid point has the highest correlation value, we keep the mid point, but halve the interval size. We then repeat the process for the new subinterval and compute the correlation for either one new point, if shifted, or two, if halved.

We note that for simplicity we have made here certain fixed choices and formulated NeuralLasso only for univariate continuous outcomes using fixed neighborhood size of 5, with ReLU as activation function and one layer. However, note that these are not arbitrary choices. As was stated earlier, ReLU was employed for its beneficial property of including linear functions as special case, if appropriate biases are learned. Some choices were found based on experimenting (*e.g.*, neighborhood size of 5 provided good predictive performance) and some of the choices (use of ReLU and linear activation functions) are discussed more in the discussion section. We refer to the **Appendix A** for more a general formulation of NeuralLasso using variable neighborhood size and activation functions.

## Example Analysis

In order to compare the prediction accuracy of different methods, we analyzed the rice field data which is publicly available at <http://www.ricediversity.org/data/> and a heterogeneous stock mouse population [see Valdar et al. (2006) for more details]. We selected traits in these data sets, which cover many levels of heritabilities (ranging from 0.25 to 0.75) and arguably many different genetic architectures.

**Rice field data:** The rice data set consists of 413 diverse accessions of *O. sativa* collected from 82 different countries (Zhao et al., 2011). The accessions were genotyped with single nucleotide polymorphism (SNP) markers and 33,569 SNPs were available for the analysis after excluding markers with minor allele frequency (MAF) > 0.05, duplicated markers and missing values > 20%. In this study, we analyzed the traits flowering time (FT) (in three different locations) and amylose content (AMY). The trait FT was measured in three different locations, the first location (ARK) was in Stuttgart, Arkansas, USA, the second one in Aberdeen (ABR) and the third location was Faridpur (FAD), Bangladesh (see Zhao et al., 2011 for more details). Out of the 413 lines, phenotypic informations were available for 371 lines in all three environments with the trait FT and 393 lines for the trait AMY. Genetic architecture underlying the trait (some traits are affected by many genes and some are by only few number of genes) is often play an important role in the prediction accuracy of different statistical methods. Thus we decided to consider two traits (FT and AMY) with different genetic architecture in this study. The narrow-sense SNP-heritabilities ( $h^2$ ) of the traits

were 0.50, 0.70, 0.50, and 0.26 for the phenotypes AMY, ARK, ABR, and FAD, respectively. Here,  $h^2$  were estimated as:  $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$ , where  $\sigma_g^2$  and  $\sigma_e^2$  are the genomic and residual variances, respectively. The variance components were estimated using GBLUP method.

**Heterogeneous stock mouse data:** The mice data (see Valdar et al., 2006) consists of 1940 individuals with 10345 biallelic SNP markers after excluding markers with minor allele frequency (MAF) = 0.05 and missing values = 20%. In this study, we analyzed the trait “body weight,” which was measured at the age of 6 weeks. The narrow-sense SNP-heritability ( $h^2$ ) of the trait “body weight” was 0.58.

**Results:** To demonstrate the superiority of our new approach, we compared the prediction accuracy (Pearson correlation coefficient between the observed and predicted phenotypes) of NeuralLasso with the most commonly used GP methods using the rice data set. The GP methods we are considering here are the GBLUP (Meuwissen et al., 2001), least absolute shrinkage and selection operator (lasso) (Tibshirani, 1996) and elastic net (EN) (Hoerl and Kennard, 1970). Also, Bayesian WGR models we choosed to consider here are the BL (Park and Casella, 2008), BayesA and BayesB (Meuwissen et al., 2001). Predictive abilities of BayesA, BayesB and BL were estimated using the R-package BGLR (Pérez and de los Campos, 2014). Whereas the predictive abilities of GBLUP and EN were estimated using the R-packages rrBLUP (Endelman, 2011) and glmnet (Simon et al., 2011), respectively. To estimate the predictive accuracy of NeuralLasso, the model was implemented in Python with TensorFlow and the scripts used in this study will be publicly available at: <https://github.com/asHauptmann/NeuralLasso>. Optimization was performed with the Adam algorithm and a cosine decay from  $10^{-3}$  to  $10^{-5}$  with 3,000 iterations, as batch size we used the full sample size.

In order to compare the prediction accuracies we used five-fold cross-validation (CV), for that we used 80% of the data as the training set and the remaining 20% as the validation set. To remove the influence of random partitions on the accuracy, we repeated the cross-validation procedure 50 times and took the mean value. Additionally, we also used the same training and validation sets with the different GP methods. In the analysis using BGLR, we used the default priors and considered 10,000 Markov Chain Monte Carlo iterations with a burn-in period of 3,000 iterations. For the EN estimation using glmnet, we set  $\alpha$  to 0.33 in Equation (3) based on cross validation.

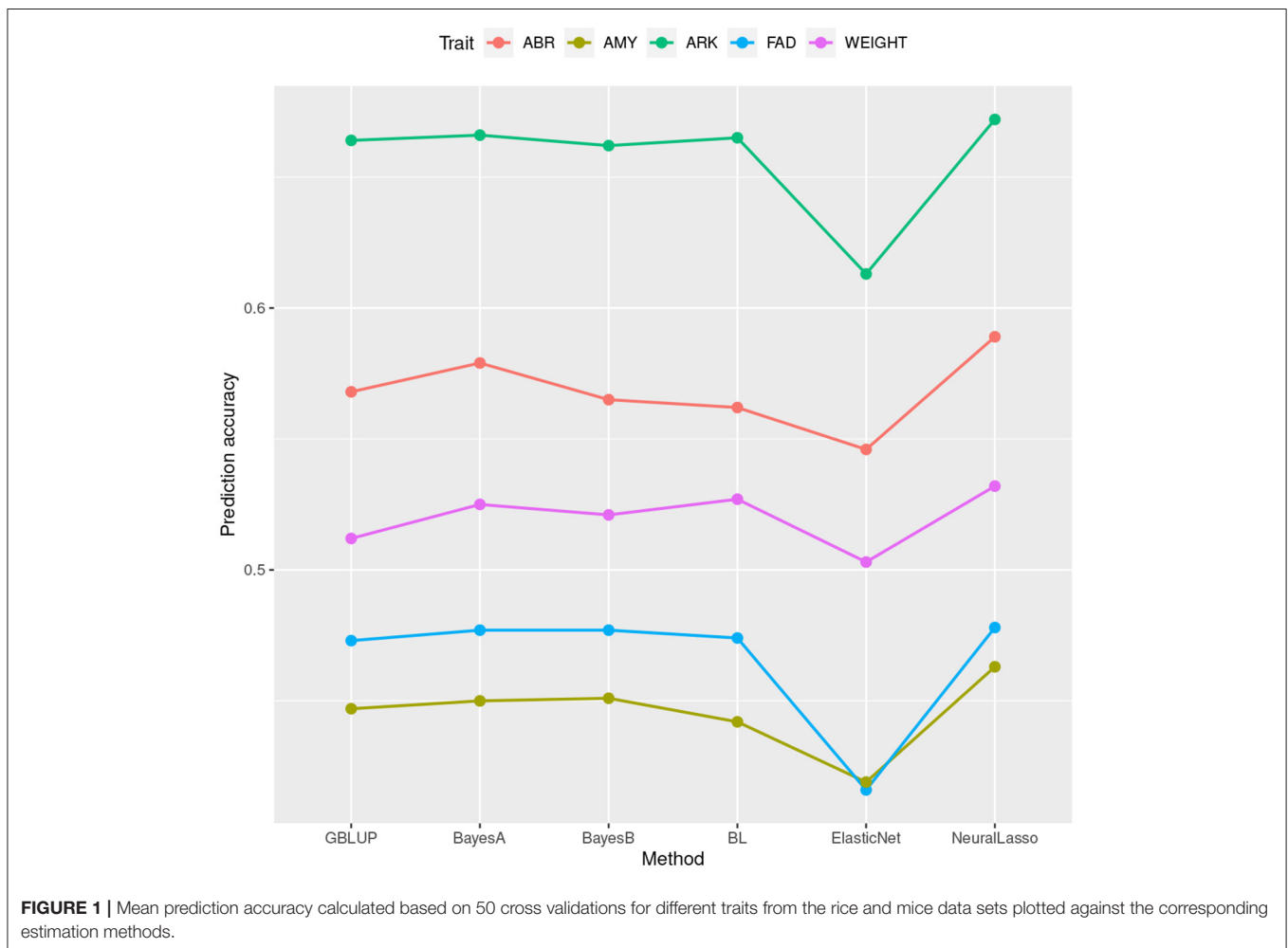
In **Table 1**, we can see the prediction accuracies of different methods in four traits (ARK, ABR, FAD, AMY) of rice data set, as well as trait body weight of mice data set. These traits together cover many levels of SNP-heritabilities. In traits ARK, ABR, and AMY of rice and trait body weight of mice, NeuralLasso seems to slightly outperform all the other methods, suggesting some role of local interactions in the genetic architecture of the trait. In trait FAD, the superior performance is much smaller and performance is practically the same with the GBLUP. This is likely due to much smaller SNP-heritability of the FAD than the other traits (see also **Figure 1** which shows ordering of the methods in their prediction



**TABLE 1** | Mean prediction accuracy based on 50 CV replicates using different approaches for the traits with rice (ARK, ABR, FAD, AMY) and mice (WEIGHT) data sets are shown along with the corresponding heritability ( $h^2$ ) estimate for the trait.

	GBLUP	BayesA	BayesB	BL	ElasticNet	NeuralLasso	$h^2$
<b>Rice</b>							
ARK	0.664	0.666 (+0.30)	0.662 (−0.30)	0.665 (+0.15)	0.613 (−7.68)	0.672 (+1.20)	0.70
ABR	0.568	0.579 (+1.93)	0.565 (−0.52)	0.562 (−1.05)	0.546 (−3.87)	0.589 (+3.69)	0.50
FAD	0.473	0.477 (+0.84)	0.477 (+0.84)	0.474 (+0.21)	0.416 (−12.05)	0.478 (+1.05)	0.26
AMY	0.447	0.45 (+0.67)	0.451 (+0.89)	0.442 (−1.11)	0.419 (−6.26)	0.463 (+3.58)	0.50
<b>Mice</b>							
WEIGHT	0.512	0.525 (+2.53)	0.521 (+1.75)	0.527 (+2.92)	0.503 (−1.75)	0.532 (+3.90)	0.58

Additionally, the percentage difference in prediction accuracy compared to the commonly used GBLUP estimation method is provided in the bracket.



accuracies). Superiority of NeuralLasso method becomes clear also from here.

## DISCUSSION

In this study, we have presented a shallow neural network method which takes into account higher order local epistatic interactions in each marker's neighborhood. In recent years, machine learning methods including deep learning (DL) methods have been widely considered for GP, however neural network methods perform

similarly or worst to the classical linear methods (Azodi et al., 2019; Zingaretti et al., 2020; Montesinos-López et al., 2021). In this study with the tested cases, our proposed method seems to improve the prediction accuracy slightly over traditional methods. We believe that the accuracy of NeuralLasso will depend on the complexity of the trait. Unlike the traditional genomic prediction models which are able to account for the two-loci genome-wide interactions, NeuralLasso account for only the additive and higher order local epistatic genetic effects. Thus there is a reduced chance that the local epistatic genetic effect will

disappear due to recombination and will be passed on to several generations (Akdemir and Jannink, 2015).

Even though, deep neural networks have been popular so far, size of learning data need to be large in many cases. We believe that more shallow networks like the one presented here may turn out to be useful and important in the future due to their more limited learning data size requirements.

In fact, our proposed model is not a single-layer neural network, *i.e.*, a perceptron. In the classic perceptron the nonlinearity is applied after summation, in our case the nonlinearity is applied before to allow for nonlinear interactions. On the other hand, one could say it is only one layer, but with several channels, for each member of the neighborhood, that are combined nonlinearly. In summary, this is why we say the model is motivated by neural networks, but does not clearly fit in the classic notion of a neural network. Finally, that is why we also do not describe our model as a neural network, but as NeuralLasso, motivated by the design of neural networks.

We also tested the performance of NeuralLasso when changing all non-linear ReLU functions to linear ones (results not shown). In those experiments, the prediction accuracies of NeuralLasso method clearly dropped down in the rice data set but stayed at about the same level in the mice data [when ReLU functions in Equation (9) were replaced by linear functions]. This is well in line with what one expects to see in rice data (high level of epistasis) and in mice data (small or no level of epistasis). Therefore, the latter experiment arguably means that our NeuralLasso may also be capable of taking into account some other context-specific effects than only epistasis, because its predictive performance was so high in mice data set.

In this study, we only considered small genomic region, however, NeuralLasso can be adjusted to account for higher order genetic interactions in larger genomic region of interest, chromosome-wise or whole genome scale. Although this might be computationally challenging, it will be interesting to see if this turns out to be important in the future. In order to reduce

the computational burden, one can also first perform a genome-wide association study (GWAS) and only account the regions of interest (*e.g.*, candidate gene regions) in NeuralLasso.

As in all genomic predictions, not any single statistical method is clearly superior in their prediction accuracy for all traits, but their performance depends on factors such as genetic architecture and heritability of the trait. However, NeuralLasso performance was found here to be promising and it is worth of considering in the future.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: details are provided in the article.

## AUTHOR CONTRIBUTIONS

BM, AH, JL, and MS: writing—review and editing and conceptualization. BM and AH: writing—original draft and formal analysis. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We are grateful for the editor and two reviewers for their comments which significantly helped us to improve presentation of this work. AH acknowledges funding from the Academy of Finland projects 338408 and 336796. The codes used in this study are made available at: <https://github.com/asHauptmann/NeuralLasso>.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.800161/full#supplementary-material>

## REFERENCES

- Akdemir, D., Jannink, J.-L., and Isidro-Sánchez, J. (2017). Locally epistatic models for genome-wide prediction and association by importance sampling. *Genet. Select. Evol.* 49, 1–14. doi: 10.1186/s12711-017-0348-8
- Akdemir, D., and Jannink, J. L. (2015). Locally epistatic genomic relationship matrices for genomic association and prediction. *Genetics* 199, 857–871. doi: 10.1534/genetics.114.173658
- Albert, F. W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* 16, 197–212. doi: 10.1038/nrg3891
- Arridge, S., and Hauptmann, A. (2019). Networks for nonlinear diffusion problems in imaging. *J. Math. Imag. Vis.* 62, 1–17. doi: 10.1007/s10851-019-00901-3
- Azodi, C. B., Bolger, E., McCarren, A., Roantree, M., de Los Campos, G., and Shiu, S.-H. (2019). Benchmarking parametric and machine learning models for genomic prediction of complex traits. *G3 Gen. Genet.* 9, 3691–3702. doi: 10.1534/g3.119.400498
- Azodi, C. B., Pardo, J., VanBuren, R., de los Campos, G., and Shiu, S.-H. (2020). Transcriptome-based prediction of complex traits in maize. *Plant Cell* 32, 139–151. doi: 10.1105/tpc.19.00332
- Bellot, P., de los Campos, G., and Pérez-Enciso, M. (2018). Can deep learning improve genomic prediction of complex human traits? *Genetics* 210, 809–819. doi: 10.1534/genetics.118.301298
- Crossa, J., Martini, J. W., Gianola, D., Pérez-Rodríguez, P., Jarquin, D., Juliana, P., et al. (2019). Deep kernel and deep learning for genome-based prediction of single traits in multi-environment breeding trials. *Front. Genet.* 10, 1168. doi: 10.3389/fgene.2019.01168
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Gen.* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024
- Garner, J., Douglas, M., Williams, S. O., Wales, W., Marett, L., Nguyen, T., et al. (2016). Genomic selection improves heat tolerance in dairy cattle. *Sci. Rep.* 6, 34114. doi: 10.1038/srep34114
- Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep Learning*. Cambridge, MA: MIT Press.
- He, S., Reif, J. C., Korzun, V., Bothe, R., Ebmeyer, E., and Jiang, Y. (2017). Genome-wide mapping and prediction suggests presence of local epistasis in a vast elite winter wheat populations adapted to Central Europe. *Theor. Appl. Genet.* 130, 635–647. doi: 10.1007/s00122-016-2840-x
- Hickey, J. M., Chiurugwi, T., Mackay, I., Powell, W., Eggen, A., Kilian, A., et al. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* 49, 1297. doi: 10.1038/ng.3920
- Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67.

- Hu, Z., Li, Y., Song, X., Han, Y., Cai, X., Xu, S., et al. (2011). Genomic value prediction for quantitative traits under the epistatic model. *BMC Genet.* 12, 1–11. doi: 10.1186/1471-2156-12-15
- Jiang, Y., and Reif, J. C. (2015). Modeling epistasis in genomic selection. *Genetics* 201, 759–768. doi: 10.1534/genetics.115.177907
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Li, Z., Gao, N., Simianer, H., and Martini, J. W. (2019). Integrating gene expression data into genomic prediction. *Front. Genet.* 10, 126. doi: 10.3389/fgene.2019.00126
- Liang, Z., Tan, C., Prakapenka, D., Ma, L., and Da, Y. (2020). Haplotype analysis of genomic prediction using structural and functional genomic information for seven human phenotypes. *Front. Genet.* 11, 1461. doi: 10.3389/fgene.2020.588907
- Liu, X., Wang, H., Wang, H., Guo, Z., Xu, X., Liu, J., et al. (2018). Factors affecting genomic selection revealed by empirical evidence in maize. *Crop J.* 6, 341–352. doi: 10.1016/j.cj.2018.03.005
- Liu, Y., Wang, D., He, F., Wang, J., Joshi, T., and Xu, D. (2019). Phenotype prediction and genome-wide association study using deep convolutional neural network of soybean. *Front. Genet.* 10, 1091. doi: 10.3389/fgene.2019.01091
- Mathew, B., Sillanpää, M. J., and Léon, J. (2019). “Advances in crop breeding techniques in cereal crops,” in *Advances in Statistical Methods To Handle Large Data Sets for GWAS in Crop Breeding*, eds F. Ordon and W. Friedt (London: Burleigh Dodds Science Publishing), 437–450. doi: 10.19103/AS.2019.0051.20
- Meuwissen, T. H. (2009). Accuracy of breeding values of unrelated individuals predicted by dense SNP genotyping. *Genet. Select. Evol.* 41, 35. doi: 10.1186/1297-9686-41-35
- Meuwissen, T. H., Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819
- Montesinos-López, O. A., Martín-Vallejo, J., Crossa, J., Gianola, D., Hernández-Suárez, C. M., Montesinos-López, A., et al. (2019). New deep learning genomic-based prediction model for multiple traits with binary, ordinal, and continuous phenotypes. *G3 Gen. Gen. Genet.* 9, 1545–1556. doi: 10.1534/g3.119.300585
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Gianola, D., Hernández-Suárez, C. M., and Martín-Vallejo, J. (2018). Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. *G3 Gen. Gen. Genet.* 8, 3829–3840. doi: 10.1534/g3.118.200728
- Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W., Fajardo-Flores, S. B., Gaytan-Lugo, L. S., Santana-Mancilla, P. C., and Crossa, J. (2021). A review of deep learning applications for genomic selection. *BMC Gen.* 22, 1–23. doi: 10.1186/s12864-020-07319-x
- Nishio, M., and Satoh, M. (2014). Including Dominance Effects in the Genomic BLUP Method for Genomic Evaluation. *PLoS ONE* 9, e85792. doi: 10.1371/journal.pone.0085792
- Norman, A., Taylor, J., Edwards, J., and Kuchel, H. (2018). Optimising genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. *G3 Gen. Gen. Genet.* 8, 2889–2899. doi: 10.1534/g3.118.200311
- Olatoye, M. O., Hu, Z., and Aikpokpodion, P. O. (2019). Epistasis detection and modeling for genomic selection in cowpea (*Vigna unguiculata* L. Walp.). *Front. Genet.* 10, 677. doi: 10.3389/fgene.2019.00677
- Park, T., and Casella, G. (2008). The Bayesian Lasso. *J. Am. Stat. Assoc.* 103, 681–686. doi: 10.1198/016214508000000337
- Pérez, P., and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442
- Pérez-Enciso, M., and Zingaretti, L. M. (2019). A guide on deep learning for complex trait genomic prediction. *Genes* 10, 553. doi: 10.3390/genes10070553
- Piepho, H., Ogutu, J., Schulz-Streeck, T., Estaghvirou, B., Gordillo, A., and Technow, F. (2012). Efficient computation of ridge-regression best linear unbiased prediction in genomic selection in plant breeding. *Crop Sci.* 52, 1093–1104. doi: 10.2135/cropsci2011.11.0592
- Resende M. Jr., Muñoz, P., Resende, M. D., Garrick, D. J., Fernando, R. L., Davis, J. M., et al. (2012). Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190, 1503–1510. doi: 10.1534/genetics.111.137026
- Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003
- Shen, X., Alam, M., Fikse, F., and Rönnegård, L. (2013). A novel generalized ridge regression method for quantitative genetics. *Genetics* 193, 1255–1268. doi: 10.1534/genetics.112.146720
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization paths for cox’s proportional hazards model via coordinate descent. *J. Stat. Softw.* 39, 1–13. doi: 10.18637/jss.v039.i05
- Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redona, E., et al. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* 11, e1004982. doi: 10.1371/journal.pgen.1004982
- Taylor, M. B., and Ehrenreich, I. M. (2014). Genetic interactions involving five or more genes contribute to a complex trait in yeast. *PLoS Genet.* 10, e1004324. doi: 10.1371/journal.pgen.1004324
- Taylor, M. B., and Ehrenreich, I. M. (2015). Higher-order genetic interactions and their contribution to complex traits. *Trends Genet.* 31, 34–40. doi: 10.1016/j.tig.2014.09.001
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Series B (Methodol.)* 58, 267–288.
- Uppu, S., Krishna, A., and Gopalan, R. P. (2016). A deep learning approach to detect SNP interactions. *J. Softw.* 11, 965–975. doi: 10.17706/jsw.11.10.960-975
- Valdar, W., Solberg, L. C., Gauguier, D., Burnett, S., Klennerman, P., Cookson, W. O., et al. (2006). Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* 38, 879–887. doi: 10.1038/ng1840
- Varona, L., Legarra, A., Toro, M. A., and Vitezica, Z. G. (2018). Non-additive effects in genomic selection. *Front. Genet.* 9, 78. doi: 10.3389/fgene.2018.00078
- Voss-Fels, K. P., Cooper, M., and Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* 132, 669–686. doi: 10.1007/s00122-018-3270-8
- Waldmann, P. (2018). Approximate Bayesian neural networks in genomic prediction. *Genet. Select. Evol.* 50, 70. doi: 10.1186/s12711-018-0439-1
- Waldmann, P., Ferenčaković, M., Mészáros, G., Khayatzaeh, N., Curik, I., and Sölkner, J. (2019). AUTALASSO: an automatic adaptive LASSO for genome-wide prediction. *BMC Bioinform.* 20, 1–10. doi: 10.1186/s12859-019-2743-3
- Wang, D., El-Basyoni, I. S., Baenziger, P. S., Crossa, J., Eskridge, K., and Dweikat, I. (2012). Prediction of genetic values of quantitative traits with epistatic effects in plant breeding populations. *Heredity* 109, 313–319. doi: 10.1038/hdy.2012.44
- Wei, W.-H., Hemani, G., and Haley, C. S. (2014). Detecting epistasis in human complex traits. *Nat. Rev. Genet.* 15, 722–733. doi: 10.1038/nrg3747
- Wittenburg, D., Melzer, N., and Reinsch, N. (2011). Including non-additive genetic effects in Bayesian methods for the prediction of genetic values based on genome-wide markers. *BMC Genet.* 12, 1–14. doi: 10.1186/1471-2156-12-74
- Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D.-J., and Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS ONE* 5, e12648. doi: 10.1371/journal.pone.0012648
- Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2, 467. doi: 10.1038/ncomms1467
- Zhao, T., Fernando, R., and Cheng, H. (2021). Interpretable artificial neural networks incorporating Bayesian alphabet models for genome-wide prediction and association studies. *G3 Gen. Gen. Genet.* 11, jkab228. doi: 10.1093/g3journal/jkab228
- Zingaretti, L. M., Gezan, S. A., Ferrão, L. F. V., Osorio, L. F., Monfort, A., Muñoz, P. R., et al. (2020). Exploring deep learning for complex trait genomic prediction in polyploid outcrossing species. *Front. Plant Sci.* 11, 25. doi: 10.3389/fpls.2020.00025

**Conflict of Interest:** BM was employed by company Bayer CropScience.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Mathew, Hauptmann, Léon and Sillanpää. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC

BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.