Interpretable Artificial Neural Networks incorporating Bayesian Alphabet Models for Genome-wide Prediction and Association Studies

Tianjing Zhao, Rohan Fernando and Hao Cheng

ABSTRACT In conventional linear models for whole-genome prediction and genome-wide association studies (GWAS), it is usually assumed that the relationship between genotypes and phenotypes is linear. Bayesian neural networks have been used to account for non-linearity such as complex genetic architectures. Here, we introduce a method named NN-Bayes, where “NN” stands for neural networks, and “Bayes” stands for Bayesian Alphabet models, including a collection of Bayesian regression models such as BayesA, BayesB, BayesC, Bayesian LASSO, and BayesR. NN-Bayes incorporates Bayesian Alphabet models into non-linear neural networks via hidden layers between SNPs and observed traits. Thus, NN-Bayes attempts to improve the performance of genome-wide prediction and GWAS by accommodating non-linear relationships between the hidden nodes and the observed trait, while maintaining genomic interpretability through the Bayesian regression models that connect the SNPs to the hidden nodes. For genomic interpretability, the posterior distribution of marker effects in NN-Bayes is inferred by Markov chain Monte Carlo (MCMC) approaches and used for inference of association through posterior inclusion probabilities (PIPs) and window posterior probability of association (WPPA). In simulation studies with dominance and epistatic effects, performance of NN-Bayes was significantly better than conventional linear models for both GWAS and whole-genome prediction, and the differences on prediction accuracy were substantial in magnitude. In real data analyses, for the soy dataset, NN-Bayes achieved significantly higher prediction accuracies than conventional linear models, and results from other four different species showed that NN-Bayes had similar prediction performance to linear models, which is potentially due to the small sample size. Our NN-Bayes is optimized for high-dimensional genomic data and implemented in an open-source package called ”JWAS”. NN-Bayes can lead to greater use of Bayesian neural networks to account for non-linear relationships due to its interpretability and computational performance.

KEYWORDS Genomic prediction, GWAS, neural networks, Bayesian regression models

Introduction

Genomic prediction plays an important role in animal and plant breeding by using genomic information to estimate genotypic values or breeding values of complex traits. The adoption of genomic information, such as single-nucleotide polymorphisms (SNPs), has greatly shortened the generation interval and improved the prediction accuracy (Meuwissen et al. 2001; Hayes et al. 2009a; Heffner et al. 2009; Hickey et al. 2012, 2017). Genome-wide association studies (GWAS) are used to detect associations between SNPs and traits. It has been applied in human diseases (e.g., Ozaki et al. 2002; Hirschhorn and Daly 2005; Klein et al. 2005; Visscher et al. 2012, 2017; Buniello et al. (2019)), as well as for traits of interest in animals and plants (e.g., Atwell et al. (2010); Korte and Farlow (2013); Sharma et al. (2015); Freebern et al. (2020)).

Based on the work of Fisher (1918), we typically assume that complex traits are affected by many genes with small additive effects, and that the relationship between genotypes and phenotypes is linear. Based on this assumption, the landmark paper of Meuwissen et al. (2001) introduced Bayesian regression models for whole-genome prediction. In Meuwissen et al. (2001), they proposed a model referred to as BayesA, where the marker effects were assigned a student-t prior distribution, and a model BayesB, where a-priori only a proportion of markers had non-zero effects. Since Meuwissen et al. (2001), several variations to these regression models have been proposed, and, hereafter, they will be collectively referred as Bayesian Alphabet models. Most of the widely-used Bayesian regression models, including BayesA, BayesB, BayesC (Kizilkaya et al. 2010; Habier et al. 2011), Bayesian LASSO (Park and Casella 2008; Gianola and Fernando 2019), and BayesR (Erbe et al. 2012; Moser et al. 2015), differ only in the prior used for the marker effects. Another well-known linear model for genomic prediction is genomic best linear unbiased prediction (GBLUP) (Habier et al. 2007; VanRaden 2008; Hayes et al. 2009b), where a genomic relationship matrix is used
to account for the covariances among genetic values. It has been shown that GBLUP is equivalent to a Bayesian regression model with a normal prior for the marker effects (Fernando 1998; Habier et al. 2007; Strandén and Garrick 2009). These models, including their variations, have been widely used in GWAS (Fernando et al. 2017; Wang et al. 2012, 2016; Moser et al. 2015; Legarra et al. 2018). The assumption of linearity in these analyses, however, may affect their performance on genome-wide prediction and association studies (Nelson et al. 2013). Thus, machine learning models have been suggested due to their ability to capture the intricate non-linear relationship between high-dimensional inputs (genotypes) and outputs (phenotypes) (Szymczak et al. 2009; Azodi et al. 2019).

Artificial neural networks, inspired by information processing of the brain, are a subset of machine learning models. Neural networks utilize multi-layer architectures to learn the representation of data (LeCun et al. 2015), where the processing layers are composed of neurons (i.e., nodes) converting the input data into a more abstract representation. Neural networks have demonstrated their predictive ability in many fields such as speech recognition, object recognition and detection (LeCun et al. 2015). It has also been applied in biological fields such as protein structure prediction, protein classification, and brain decoding (Min et al. 2017).

In some genomic prediction studies, neural networks with different architectures were compared to conventional linear models, such as GBLUP and Bayesian regression models. In genomic prediction applications, single hidden layer neural networks are typically applied, but they vary in the number of nodes in the hidden layer (Gianola et al. 2011; Okut et al. 2011, 2013; Ehret et al. 2015). Neural networks achieved higher prediction accuracies in some studies (e.g., Gianola et al. (2011); Ma et al. (2018)), but a gain in prediction performance was not consistently observed (e.g., Okut et al. (2013); Bellot et al. (2018); Azodi et al. (2019); Abdollahi-Arpanahi et al. (2020)), potentially due to the sample sizes in these studies or genetic architecture of the traits (Abdollahi-Arpanahi et al. 2020).

Besides the absence of a substantial improvement in prediction accuracy in genomic prediction, a major criticism of neural network models is the lack of interpretability. For example, deep learning (“deep” typically refers to neural networks with more than two hidden layers (Schmidhuber 2015)) is applied in a black-box manner, i.e., it is hard to interpret results biologically for genome-wide analyses, although the neural network architecture may capture the complex relationships between inputs (genotypes) and outputs (phenotypes). Unlike neural networks, conventional linear models provide an explicit interpretation to link phenotypes to genotypes under some assumptions, and thus, they are used in many genome-wide analyses, e.g., GWAS. This lack of interpretability of results from neural networks impedes their application for association studies (Ehret et al. 2015; Waldmann 2018). In some studies (e.g., Okut et al. (2011); Glória et al. (2016); Waldmann (2018)), weights in neural networks were interpreted as regression coefficients to obtain the effect of each SNP, and these effects without significance measures (e.g., p-value) were used to rank all SNPs.

Bayesian regularization has been applied to the neural networks to reduce their complexity by treating the network weights as random variables with prior densities (e.g., Gianola et al. (2011); Okut et al. (2011, 2013); Glória et al. (2016); van Bergen et al. (2020)). However, in most Bayesian neural network studies on genomic prediction, in part due to computational reasons, inferences of unknowns were only based on the posterior mode via the maximum a posteriori (MAP) estimator (e.g., Gianola et al. (2011); Okut et al. (2011, 2013); Glória et al. (2016)). Chen et al. (2017a) showed that in association studies, the performance of MAP inferences may be inferior to Markov chain Monte Carlo (MCMC) approaches, which exactly estimate the posterior distribution. In Demetci et al. (2020), variational inference was used to approximate the posterior distribution of posterior inclusion probabilities (PIPs) for all SNPs. However, simulations in Demetci et al. (2020) showed that such approximation may mis-estimate the network weights.

Although Bayesian methods had been proposed for analyses of genetic populations (Dempfle 1977; Gianola and Fernando 1986) in the 1980’s, these methods were not widely adopted until MCMC approaches were introduced to draw inferences from posterior distributions (Wang et al. 1994). In the method proposed here, the posterior distribution of marker effects is inferred using MCMC approaches, and it is used to produce significance measures such as posterior inclusion probability (PIP) (Guan and Stephens 2011) and window posterior probability of association (WPPA) (Fernando et al. 2017). The method is called NN-Bayes, where “NN” stands for neural networks, and “Bayes” stands for Bayesian Alphabet models, a collection of Bayesian regression models such as BayesA, BayesB, BayesC, Bayesian LASSO, and BayesR. As showed in Figure 1, the framework of NN-Bayes starts with SNPs that are linearly connected to hidden nodes, through a multi-trait, multiple regression model, that are then non-linearly connected to the observed trait. Multi-trait Bayesian Alphabet models (Cheng et al. 2018b) will be employed to fit the regression model that connects the SNPs to the hidden nodes. The hidden nodes, sampled via Hamiltonian Monte Carlo, represent unobserved intermediate traits that are non-linearly connected to the observed trait to accommodate complex genetic architecture including non-additive genetic effects. In summary, NN-Bayes attempts to improve the performance of genome-wide prediction and GWAS by accommodating non-linear relationships between the hidden nodes and the observed trait, while maintaining genomic interpretability through the Bayesian regression models that connect the SNPs to the hidden nodes.

Compared to other neural network applications of genome-wide prediction and association studies, our NN-Bayes using MCMC approaches is optimized for high-dimensional genomic data to be computationally efficient. In most neural network applications of genomic prediction, neural networks were implemented using existing general purpose software tools which are not optimized for genomic analyses, and it would be time consuming to estimate effects of a large number of genome-wide molecular markers (i.e., weights between input layer and hidden nodes). For example, in van Bergen et al. (2020), it took about 4 hours to run 1,000 iterations for a single-layer neural network with 20 hidden nodes using a dataset of 500 individuals and 5,000 SNPs. However, the computing time of our NN-Bayes, implemented in an open-source package ‘JWAS’ (Cheng et al. 2018a), is about 2 minutes for such an analysis on a laptop (with 2.6 GHz Intel Core i7 processor), and it scales nearly linearly by the number of observations and the number of markers.

Materials and Methods
Bayesian Neural Network Alphabet
Bayesian analyses using NN-Bayes will be demonstrated for a single observed trait. The hierarchical Bayesian model of NN-
Bayesian is showed in Figure 2. For individual \( i \), the observed trait is modeled as

\[
y_i = \mu^{(i)} + \sum_{j=1}^{l_1} w^{(i)}_j g(z_{ij}) + \epsilon_i,
\]

where \( y_i \) is the observed phenotype for individual \( i \), \( \mu^{(i)} \) is the overall mean for individual \( i \), \( g(.) \) is an element-wise non-linear function, \( w^{(i)}_j \) is the effect of \( j \)-th hidden node for individual \( i \), \( \epsilon_i \) is a vector of random residuals for the \( i \)-th individual, and \( z_{ij} \) is the vector of random residuals for the \( i \)-th individual. The prior for \( \epsilon_i \) is a normal distribution with null mean and unknown variance \( \sigma^2_{\epsilon} \), i.e., \( \epsilon_i \sim N(0, \sigma^2_{\epsilon}) \). A scaled inverse chi-squared distribution is assigned as the prior for \( \sigma^2_{\epsilon} \), i.e., \( \sigma^2_{\epsilon} \sim \chi^2_{\nu_{\epsilon}} / \nu_{\epsilon} \).

For \( i \)-th individual, for the prior for the hidden nodes, \( z_i = [z_{i,1}, \ldots, z_{i,l_1}]^T \), can be presented as a multi-trait Bayesian regression model (Cheng et al. 2018b):

\[
z_i = \mu^{(0)} + \sum_{m=1}^{l_0} x_{i,m} w^{(0)}_m + \epsilon_i,
\]

where \( z_i \) is the vector of \( l_1 \) hidden nodes, which can be thought of as unobserved intermediate traits, for individual \( i \), \( \mu^{(0)} = [\mu_1^{(0)}, \ldots, \mu_{l_1}^{(0)}]^T \) is a vector of overall means for the \( l_1 \) hidden nodes, \( x_{i,m} \) is the observed genotype covariate at locus \( m \) for individual \( i \) (coded as 0,1,2), \( w^{(0)}_m = [w_{i,0}^{(0)}, \ldots, w_{i,l_1}^{(0)}]^T \) are the marker effects of locus \( m \) on the \( l_1 \) unobserved intermediate traits (i.e., weights between input and all hidden nodes in the neural network), and \( \epsilon_i \) is a vector of random residuals for the \( i \)-th hidden nodes.

The overall means, \( \mu^{(0)} \), are assigned flat priors. Conditional on \( \Sigma \), the residuals, \( \epsilon_i \), have independently and identically distributed multivariate normal priors with null means and covariance matrix \( \Sigma \), which itself is assumed to have an inverse Wishart prior distribution, \( \Sigma^{-1}(\nu, \Sigma_0) \). As showed in Cheng et al. (2018b), by employing the data augmentation, \( w^{(0)}_m \) can be written as \( w^{(0)}_m = D_m \beta_m \), where \( D_m \) is a diagonal matrix, where the \( k \)-th diagonal entry indicates whether the marker effects of locus \( m \) for the unobserved intermediate trait \( k \) is zero or not. \( \beta_m \) follows some multivariate distribution, e.g., the multivariate normal distribution in multi-trait BayesCII or the multivariate t distribution in multi-trait BayesB. Note that multi-trait Bayesian regression models may be computationally intensive if a large number of hidden nodes are used. In this case, we assume that hidden nodes are independent such that multiple single-trait Bayesian regression models can be used in parallel at each iteration.

In Figure 1, we also present the hierarchical Bayesian model described above as a single hidden layer feed-forward Bayesian neural network. This architecture is typical for neural networks used in genomic prediction (Gianola et al. 2011; Okut et al. 2011, 2013; Ehret et al. 2015).

In the neural network, for each observation, the input layer is composed of a vector of \( l_0 \) molecular markers (coded as 0, 1, 2), and the output layer is the observed trait. Nodes in the hidden layer, between input and output layer, represent unob-
served intermediate traits, and the relationship between hidden
nodes and observed trait is non-linear. The weights between
the input layer and the hidden layer represent marker effects
on unobserved intermediate traits, the “biases” of hidden nodes
can be regarded as the overall means. The weights and bias
between the hidden layer and the output layer help to define the
non-linear relationship between unobserved intermediate traits
and observed trait. Note that extra input nodes can be included
in the neural network if additional random or fixed effects are
fitted in the model.

Inference using Markov chain Monte Carlo
From a Bayesian perspective, inferences of unknowns are based
on their posterior distributions. However, closed-form expres-
sions of those posteriors are usually unavailable. Thus, in prac-
tice, samples are obtained using MCMC techniques, where statis-
tics computed from the resulting Markov chain converge to
those from the posterior as chain length increases (Norris 1998;
Sorensen and Gianola 2007). Here, to construct the Markov
chain, we will use Gibbs sampling, where each unknown vari-
able or block of variables is sampled from its full conditional
distribution, conditioned on the observed data and the latest
samples of all the other unknowns. Given the sampled values of hidden nodes $z_i$ in equation (2),
unknown parameters between the input layer and the hidden
layer do not depend on the observed trait. Sampling of those
unknowns, including the overall means $\mu^{(0)}$, the marker effects
$w^{(0)}_m$, and the residual covariance matrix $\Sigma$, given the sampled
values of $z_i$, is based on the multi-trait Bayesian regression mod-
els described in Cheng et al. (2018b). Given the sampled values
for the hidden nodes (after activation) $g(z_i)$ and output node $y_i$,
the sampling of weights and bias between the hidden layer and
the output layer is straightforward (derivations are given in the
Appendix). As showed above, we consider Hamiltonian Monte
Carlo to draw samples for the hidden nodes. The sampler is
implemented in the open-source package "JWAS" (Cheng et al.
2018a).

Sampling hidden nodes using Hamiltonian Monte Carlo
Hamiltonian Monte Carlo (HMC) is used to sample hidden nodes, i.e.,
$z_i$. In HMC, each unknown parameter is paired with a “mo-
momentum” variable $\phi_i$. The HMC constructs the Markov chain
by a series of iterations. A useful introduction to the principles
and concepts underlying HMC is given by Betancourt (2017).
Following notations in Gelman et al. (2013), there are three steps
in each iteration of the HMC:
1. updating the momentum variable independently of the cur-
rent values of the paired parameter, i.e., $\phi_i \sim MVN(0, M)$.
2. updating $(z_i, \phi_i)$ via L “leapfrog steps”. In each leapfrog
step, $z_i$ and $\phi_i$ are updated dependently and scaled by $\epsilon$.
The leapfrog step below is repeated $L$ times:
   (a) $\phi_i \sim \phi_i + \frac{1}{2} \epsilon \frac{d log p(z_i | y_i, ELSE)}{dz}$;
   (b) $z_i \sim z_i + \epsilon M^{-1} \phi_i$;
   (c) $\phi_i \sim \phi_i + \frac{1}{2} \epsilon \frac{d log p(z_i | y_i, ELSE)}{d\phi}$.

   The resulting state at the end of $L$ repetitions will be denoted
   as $(z_i^*, \phi_i^*)$.
3. calculating the acceptance rate, $r$, and above resulting state
will be accepted with probability $\min(1, r)$.

In our method, the gradient of log full conditional posterior distribution of $z_i$ is:

$$
\frac{d \log f(z_i | y_i, ELSE)}{dz_i} = -\Sigma^{-1}(z_i - \mu^{(0)}) - \frac{1}{n} \sum_{m=1}^{n} x_{i,m} w^{(0)}_m
$$

where $w^{(1)} = [w^{(1)}_1, ..., w^{(1)}_n]^T$ is the vector of weights between
hidden nodes and observed trait, ELSE denotes the sampled
values of remaining unknowns, and the symbol $\circ \phi$ denotes the element-wise production.

In the round $t$ of HMC, the acceptance rate, $r$ is:

$$
r = \epsilon \frac{\log p(z_i^* | y_i, ELSE) \circ \phi_i^*) - \log \phi_i^{(t-1)} | y_i, ELSE) \circ \phi_i^{(t-1)}}{\Sigma_i}
$$

A detailed derivation can be found in the Appendix. In our analyses, 10 leapfrog steps was applied in each iteration of HMC, i.e., $L = 10$, the scale parameter $\epsilon$ was 0.1, and $M$ was set as an
identity matrix.

NN-Bayes for Genomic Prediction
The MCMC sample from the posterior distribution of genotypic
values for individuals of interest is obtained as

$$
\hat{g} = g(1 \mu^{(0)} + X w^{(0)}) w^{(1)},
$$

where $X$ is the genotype covariate matrix for individuals of interest, $w^{(0)}$ is a matrix of size $p$-by-$l_1$ of samples of marker
effects on the $l_1$ hidden nodes, $\mu^{(0)}$ is a vector of samples of
overall means for the $l_1$ hidden nodes, and $w^{(1)}$ is a vector of samples of $l_1$ weights between hidden nodes and the observed
trait, and $g(.)$ is a non-linear function. The prediction accuracy is
calculated as the Pearson correlation between the posterior mean
of genotypic values and the phenotypic values in the validation
dataset.

NN-Bayes for Genome-wide Association Studies
In NN-Bayes, the PIP for each single marker is computed as the
frequency that its effect is non-zero on at least one of hidden
nodes. Here we prefer the use of the posterior probability of
association of a genomic window (WPPA) (Fernando et al. 2017)
to account for the fact that highly correlated SNPs within a
genomic window jointly affect the phenotype, and it is difficult
to identify the effect of a single marker (Hayes et al. 2010).

Given that an investigator is interested in identifying genomic
segments that explain more than a proportion $T$ of the genetic
variance, WPPA is defined as the posterior probability of this
event (Fernando et al. 2017). To estimate WPPA, first, MCMC
samples of genotypic values for all individuals are obtained
from their posterior distribution, using equation (5). Next, using
the notation in equation (5), the genotypic values that are
attributed to genomic window $t$ are sampled from their posterior
distribution as

$$
\hat{g}_t = g(X_t w^{(0)}_t) w^{(1)},
$$

where $X_t$ is the genotype covariate matrix of markers in window
t, and $w^{(0)}_t$ represents the samples of marker effects in
window $t$ on hidden nodes. The proportion of the genetic variance explained by the genomic window $t$, $q_t$, is now sampled as $c_t^2$, where $c_t^2$ and $c_t^2$ are the variances of a random sample from $g$ and $g_t$, respectively. Then, the posterior probability that window $t$ accounts for more than a proportion $T$ of the genetic variance (i.e., WPPA) can be estimated from the MCMC samples by counting the number of samples where $q_t > T$ (Fernando and Garrick 2013; Chen et al. 2017b; Lloyd-Jones et al. 2017). In this paper, non-overlapping windows of 1 Mb were used to identify genomic windows that explain over 1% of the total genetic variance (i.e., $T = 0.01$).

**Data Analysis**

**Simulated data** Real pig genotypes in Duarte et al. (2014) were used to simulate phenotypes with dominance and epistatic effects. All SNP markers on the chromosome 1 were used, resulting in a genotypic data set with 5,023 markers for 928 individuals. A random sample of five percent of these 5,023 markers were selected as quantitative trait loci (QTL). Following the simulation in van Bergen et al. (2020), the additive effects ($a$) of QTL were sampled from a univariate normal distribution with null mean and variance one. Dominance effects ($d$) were sampled from a univariate normal distribution with mean 1.2 and standard deviation 0.3. Dominance effects ($d$) were computed as $\delta[a]$. The epistatic factors ($\gamma$) of all pairwise combinations of QTL were sampled from a univariate normal distribution with null mean and variance one. The epistatic effect ($e$) between $i$th and $j$th QTL was computed as $\gamma \sqrt{|a_i| a_j}$. The complementary epistasis scenario in van Bergen et al. (2020) was used to simulate the total genetic value. Phenotypes were simulated with a broad-sense heritability of 0.5 and were scaled to have a phenotypic variance equal to one. Based on the simulation processes described above, 20 different datasets were simulated.

**Real data** Publicly available genotypic and phenotypic data of multiple species were used to compare NN-Bayes and conventional linear models. These datasets include the pig dataset from Duarte et al. (2014), and data for soy, maize, switchgrass, and spruce from Azodi et al. (2019). Traits used in our analyses were the 13 week tenth rib backfat (BF), yield (YLD), wood density (DE) and flowering time (FT). The description of each dataset, such as the number of SNP markers and the number of observations, are showed in Table 1.

**Genomic Prediction** 20 and 50 replicates were applied for each real and simulated dataset, respectively. In each replicate, a random subset of 80% of all observations were used for training, and the remaining were used for validation. The prediction accuracy is calculated as the Pearson correlation between the posterior mean of genotypic values and the phenotypic values in the validation dataset. In NN-Bayes, different priors for marker effects (i.e., weights between inputs and hidden nodes) in Bayesian Alphabet models (i.e., RR-BLUP, BayesA, BayesB, BayesC$\pi$, and Bayesian LASSO) were used. Linear models with priors in Bayesian Alphabet models were also performed. Different number of hidden nodes (i.e., 2, 3, 5 and 10) were tested in the analyses. Chains of length 10,000 and 20,000 were applied to simulated datasets and real datasets, respectively, to ensure the convergence.

**Genome-wide association studies** For each dataset, PIP and WPPA were used for association inference. In the simulated study, non-overlapping windows of 1 Mb were tested. Genomic windows that explain over 1% of the total genetic variance were assumed to be of potential interest (i.e., $T = 0.01$). The area under the receiver operating characteristic curve (AUC) was calculated using the R package pROC (Robin et al. 2011) to assess the performance of NN-Bayes in GWAS.

**Data availability** The genotypic and phenotypic data used in the real data analyses are publicly available in Duarte et al. (2014) and Azodi et al. (2019). The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

**Results**

**Genomic Prediction**

**Simulated data** Overall, the prediction accuracies of NN-Bayes were significantly higher than those of conventional linear models, and the differences were substantial in magnitude. The number of hidden nodes did not significantly affect the performance of NN-Bayes. No significant differences were found with different priors for marker effects used in NN-Bayes. Here we only present the comparison of prediction accuracy between NN-Bayes composed of 10 hidden nodes with BayesCII priors (named as NN-BayesCII-10) and the linear model with BayesC$\pi$ prior (LM-BayesC$\pi$). In Figure 3, the 20 simulated datasets were distinguished by color. Prediction accuracies of NN-BayesCII-10 were higher than LM-BayesC$\pi$ in 949 out of 1,000 validation sets (i.e., above the diagonal black line), and the prediction accuracy for NN-BayesCII-10 was significantly higher than for LM-BayesC$\pi$ in 16 out of 20 simulated datasets under the t-test with a significance level of 0.05. The results show that our NN-Bayes has the potential to improve the prediction accuracy for a trait with non-additive genetic effects.

**Real data** In each of the five species, the prediction accuracies of NN-Bayes with different numbers of hidden nodes were not significantly different under the t-test with a significance level of 0.05. Significant differences were found among different priors.

---

**Figure 3** The prediction accuracy of NN-Bayes composed of 10 hidden nodes with BayesCII priors (NN-BayesCII-10) versus a linear model with BayesC$\pi$ prior (LM-BayesC$\pi$) on scenarios with both additive and non-additive effects (i.e., dominance and epistasis). 20 simulated datasets were distinguished by color. For each dataset, 50 replicates for validation were applied. The diagonal line is used for reference such that a dot above the line represents a validation with higher accuracy for NN-Bayes.
for marker effects in the pig data. Comparison of prediction accuracies among all methods is showed in Table 2 in the Appendix. In detail, NN-Bayes using BayesA and BayesB priors performed better than methods using other priors in the pig data. The prediction accuracies of NN-Bayes composed of 10 hidden nodes with BayesC\(\pi\) priors (named as NN-BayesC\(\pi\)-10) and the linear model with BayesC\(\pi\) prior (LM-BayesC\(\pi\)) are showed in the last column of Table 1. Among the five species, the prediction accuracies for NN-BayesC\(\pi\)-10 were higher than those for LM-BayesC\(\pi\) when the ratio between number of observations and number of markers was relatively high (e.g., soy and spruce). However, the prediction accuracies of NN-BayesC\(\pi\)-10 were not significantly different from those for LM-BayesC\(\pi\) under the t-test with a significance level of 0.05. This may be because the sample size is not large enough for accurate parameter estimates, or the architecture of the neural network may not be sufficiently complex to represent the intricate hidden relationships between genotypes and phenotype. Alternatively, the additive genetic effects may have already accounted for a majority proportion of the variation, so the improvement based on non-additive effects is limited. When we considered a significance level of 0.1, the prediction accuracies for NN-Bayes composed of 10 hidden nodes with priors for marker effects in BayesC\(\pi\) or RR-BLUP were significantly higher than conventional linear models for the soy dataset. As showed in Figure 4, for the soy dataset, in 15 out of 20 validation sets, the NN-BayesC\(\pi\)-10 had a higher prediction accuracy than the LM-BayesC\(\pi\) (i.e., above the diagonal black line).

**Genome-wide association studies**

**Simulated Data** GWAS was conducted on 20 simulated datasets. NN-Bayes composed of 2 hidden nodes with priors for marker effects in BayesC\(\pi\) (named as NN-BayesC\(\pi\)-2) was tested and compared to results from a linear model with BayesC\(\pi\) prior (LM-BayesC\(\pi\)). Chain of length 500,000 with 200,000 burn-in was used to ensure the convergence, and samples were saved for every 50 iterations.

For both NN-BayesC\(\pi\)-2 and LM-BayesC\(\pi\), when PIP was used for association inference, the AUC was around 0.5, indicating the performance of PIP was similar to a random classifier. Thus, we will only present the GWAS results using WPPA in all analyses. The AUC of NN-BayesC\(\pi\)-2 was significantly higher than that of LM-BayesC\(\pi\), under the t-test with a significance level of 0.15. As showed in Figure 5, the AUC for NN-BayesC\(\pi\)-2 was higher than that for LM-BayesC\(\pi\) in 16 out of the 20 simulated datasets (i.e., above the diagonal black line).

**Real data** GWAS was conducted on the real pig dataset. NN-Bayes composed of 2 hidden nodes with priors for marker effects in BayesC\(\pi\) (named as NN-BayesC\(\pi\)-2) was tested. Chain of length 1 million with 500,000 burn-in was used to ensure the convergence, and samples were saved for every 10 iterations.

For NN-BayesC\(\pi\)-2, a WPPA > 0.95 was used to identify significant associations because it results in controlling the proportion of false positives to ≤ 0.05 (Fernando et al. 2017). The GWAS results from NN-BayesC\(\pi\)-2 had a similar pattern as those in Duarte et al. (2014). The significant window on chromosome 6 was identified, which is consistent with the GWAS results in Duarte et al. (2014).

**Discussion**

In conventional linear models used for genomic prediction, it is usually assumed that complex traits are affected by many genes with small additive effects, and that the relationship between genotypes and phenotypes is linear. To account for non-linear relationships such as non-additive genetic effects, we proposed a method named NN-Bayes, where "NN" stands for neural networks, and "Bayes" stands for Bayesian Alphabet models, a collection of Bayesian regression models such as BayesA, BayesB,

| Reference       | Species | Trait | #SNPs | #Observations | Prediction accuracy |
|-----------------|---------|-------|-------|---------------|---------------------|
| Duarte et al. (2014) | pig     | BF    | 42,246| 928           | 0.530 (0.05)        |
| Azodi et al. (2019) | soy     | YLD   | 5,014 | 4,234         | 0.663 (0.02)        |
| Azodi et al. (2019) | spruce  | DE    | 6,785 | 1,722         | 0.418 (0.05)        |
| Azodi et al. (2019) | switchgrass | FT  | 217,150| 514          | 0.860 (0.03)        |
| Azodi et al. (2019) | maize   | FT    | 244,771| 391          | 0.754 (0.10)        |

a BF: 13 week tenth rib backfat; YLD: yield; DE: wood density; FT: flowering time; #SNP: number of SNP markers; #Observations, number of observations; Prediction accuracy: the average prediction accuracy from 20 replicates with the standard deviation in the bracket.

![Figure 4](image-url) The prediction accuracy of NN-Bayes composed of 10 hidden nodes with BayesC\(\pi\) priors (NN-BayesC\(\pi\)-10) versus the linear model with BayesC\(\pi\) prior (LM-BayesC\(\pi\)) for the soy dataset. 20 replicates for validation were applied. The diagonal line is used for reference such that a dot above the line represents a validation with higher accuracy for NN-Bayes.
BayesC, BayesR, and Bayesian LASSO. NN-Bayes incorporates Bayesian Alphabet models into non-linear neural networks via hidden layers between SNPs and observed traits. Priors in multi trait Bayesian Alphabet models are assumed for marker effects on hidden nodes, and flexible non-linear relationships between hidden nodes and the observed trait are assumed through activation functions, e.g., the hyperbolic tangent function. Thus, NN-Bayes attempts to improve the accuracy of prediction by accommodating non-linear relationships between the hidden nodes and the observed trait, while maintaining genomic interpretability through the Bayesian regression methods that connect the SNPs to the hidden nodes. Compared to other neural network applications on genomic prediction, NN-Bayes has been optimized for MCMC based inference with high-dimensional genomic data, and it is thus more computationally efficient. Our NN-Bayes is implemented in an open-source package called “JWAS” (Cheng et al. 2018a).

In the analyses of the simulated data, where the phenotypes were simulated with both additive and non-additive effects (i.e., dominance and epistasis), performance of NN-Bayes was significantly better than conventional linear models for both GWAS and whole-genome prediction, and the differences on prediction accuracy were substantial in magnitude. In real data analyses for the soy dataset, NN-Bayes was able to achieve significantly higher prediction accuracies than conventional linear models, and results from other four different species showed that NN-Bayes had similar prediction performance to linear models. Results in Azodi et al. (2019) showed that artificial neural networks did not tend to outperform linear models for the multiple plant datasets used in our analyses. This may be because the sample size is not large enough for accurate parameter estimates, or the architecture of the neural network may not be sufficiently complex to represent the intricate hidden relationships between genotypes and phenotype. Alternatively, the additive genetic effects may have already accounted for a majority proportion of the variation, so the improvement based on non-additive effects is limited.

In most neural network applications on genomic prediction, single hidden layer, feed-forward neural networks have been used (e.g., Gianola et al. (2011); Okut et al. (2011, 2013); Ehret et al. (2015)). These neural networks are often trained by gradient-based back-propagation, and the training process stops when it reaches the maximum number of training iterations or stops early when a criterion is met, for example, the optimal mean squared error (MSE) on a validation dataset. To reduce the overfitting in such neural networks with large numbers of molecular markers as input, regularization technology, such as Bayesian regularization (Gianola et al. 2011) and L1-norm penalty on unknown parameters (Wang et al. 2018), are usually applied. In Bayesian regularized neural networks, the effective number of parameters were similar when different numbers of neurons were used in the middle layer (Gianola et al. 2011; Okut et al. 2013; Ehret et al. 2015) and were usually much smaller than the sample size. Moreover, dropout may be applied to reduce the complexity of neural networks by setting a random proportion of weights to zero in each training iteration (Waldmann 2018; Azodi et al. 2019). In van Bergen et al. (2020), variable selection was used in a Bayesian neural network by fitting a “marker selection vector” into the Bayesian neural networks.

Most Bayesian neural network studies on genomic prediction heavily rely on approximations in part because they were performed using general algorithms implemented in existing software tools that were not optimized for genomic data analyses. For example, in Gianola et al. (2011) and Okut et al. (2013), the rationale of using relationship matrix as input of neural networks is from the representation of the infinitesimal model as a regression on relationship matrix, so weights connecting the input and hidden layers should follow the prior distribution of $N(0, A^{-1}c_p^2)$, where $A$ is the relationship matrix. However, due to the use of MATLAB (Demuth and Beale 2009) in Gianola et al. (2011) and Okut et al. (2013), this prior was restricted in the form of $N(0, Ic_p^2)$. In addition to approximations, it is usually computationally intensive or infeasible to analyze large datasets on genomic prediction using general algorithms implemented in existing software tools. Okut et al. (2013) reported that MATLAB required about 2 hours to train a neural network with ~2500 SNPs as input, and van Bergen et al. (2020) reported that PyMC3 (Salvatier et al. 2016) took about 4 hours with 500 individuals and 5000 SNPs. For a dense genotype, such as the Jersey cow data with 35,798 SNPs, the relationship matrix of size 297 was used as input to make computations feasible in MATLAB (Gianola et al. 2011). These methods (Gianola et al. 2011; Okut et al. 2011, 2013; Ehret et al. 2015; Bellot et al. 2018; Azodi et al. 2019) were not used for GWAS due to the lack of interpretability. Our NN-Bayes using MCMC approaches is computationally efficient and can be further sped up by a recently developed parallel computing strategy (Zhao et al. 2020).

Appendix

**MCMC in NN-Bayes**

**Sampling weights between hidden layers and output layers**

The full conditional posterior distribution of $\mu^{(1)}$ and $\mathbf{w}^{(1)}$,

$$
\begin{bmatrix}
\mu^{(1)} \\
\mathbf{w}^{(1)}
\end{bmatrix}
$$

is a multivariate normal distribution with mean:

$$
\left(\mathbf{1}^\prime \mathbf{g(Z)}\right)^{-1}\mathbf{y}
$$

and covariance matrix:

$$
\left(\mathbf{1}^\prime \mathbf{g(Z)}\right)^{-1}\mathbf{c_p^2}
$$

Bayesian Neural Network Alphabet 7
where \( \mathbf{y} \) is a vector of \( n \) phenotypes for the observed trait, and \( \mathbf{Z} \) is a matrix of size \( n \times l \) for hidden nodes.

**Sampling hidden nodes using Hamiltonian Monte Carlo** In the round \( t \) of HMC, the acceptance rate, \( r \), can be expressed as:

\[
    r = \frac{p(z_1^t, \phi^t|y, \text{ELSE})}{p(z_1^{t-1}, \phi^{t-1}|y, \text{ELSE})}
    \]

\[
    = \frac{p(z_1^t|y, \text{ELSE})p(\phi^t)}{p(z_1^{t-1}|y, \text{ELSE})p(\phi^{t-1})}
    \]

\[
    \approx e^{\log(p(z_1^t|y, \text{ELSE})p(\phi^t)) - \log(p(z_1^{t-1}|y, \text{ELSE})p(\phi^{t-1}))}
    \]  

As shown in equation (3) and (9), the log full conditional posterior distribution of the hidden nodes and their gradients are required in HMC. Following equation (3), the full conditional posterior distribution of the hidden nodes for individual \( i \), i.e., \( z_i \), can be expressed as:

\[
    f(z_i|y, \text{ELSE}) \propto f(z_i|\mu^{(0)}, w^{(0)}_1, \ldots, w^{(0)}_m, x_{i,1}, \ldots, x_{i,n}, \Sigma) 
    \]

\[
    f(y_i|\mu^{(1)}, w^{(1)}_1, \ldots, w^{(1)}_m, z_i, c^2_i) 
    \]

\[
    \propto |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{(z_i - \mu^{(0)} - \sum x_{i,m}w^{(0)}_m)^T \Sigma^{-1} (z_i - \mu^{(0)} - \sum x_{i,m}w^{(0)}_m)}{2}\right\} 
    \]

\[
    \cdot \left(c^2_i\right)^{-\frac{1}{2}} \exp\left\{-\frac{(y_i - \mu^{(1)} - (w^{(1)}_m)^T g(z_i))^2}{2c^2_i}\right\} 
    \] (10)

Thus, the log full conditional posterior distribution of \( z_i \) is:

\[
    \log f(z_i|y, \text{ELSE}) 
    \]

\[
    \propto -\frac{1}{2} \log(|\Sigma|) 
    \]

\[
    -\frac{(z_i - \mu^{(0)} - \sum x_{i,m}w^{(0)}_m)^T \Sigma^{-1} (z_i - \mu^{(0)} - \sum x_{i,m}w^{(0)}_m)}{2} 
    \]

\[
    -\frac{1}{2} \log(c^2_i) - \left|y_i - \mu^{(1)} - (w^{(1)}_m)^T g(z_i)\right|^2 
    \]

Thus, the gradient of the log-full conditional posterior distribution of \( z_i \) is:

\[
    \frac{d\log f(z_i|y, \text{ELSE})}{dz_i} 
    \]

\[
    \propto -\Sigma^{-1} (z_i - \mu^{(0)} - \sum_{m=1}^l x_{i,m}w^{(0)}_m) 
    \]

\[
    + \frac{y_i - \mu^{(1)} - (w^{(1)}_m)^T g(z_i)}{c^2_i} 
    \]

where the symbol “\( \circ \)” denotes the element-wise production.

**Sampling residual variance of the observed trait** The full conditional posterior distribution of \( c^2 \) is a scaled inverse chi-squared distribution with \( n + v_t \) degrees of freedom and scale parameter \( e' e + S^2_e \), where the \( e \) is \( y - \hat{\mu}^{(1)} - g(Z)w^{(1)} \).

**Genomic Prediction using real datasets**

Prediction accuracies for each real dataset are shown in table 2. In each of the five species, the prediction accuracies of NN-Bayes with different numbers of hidden nodes were not significantly different under the t-test with a significance level of 0.05. Significant differences were found among different priors for marker effects in the pig data.

---

**References**

Abdollahi-Arpanahi, R., D. Gianola, and F. Peñaagaricano, 2020 Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. Genetics Selection Evolution 52: 1–15.

Atwell, S., Y. S. Huang, B. J. Vilhjálmsson, G. Willems, M. Horton, et al., 2010 Genome-wide association study of 107 phenotypes in arabidopsis thaliana inbred lines. Nature 465: 627–631.

Azodi, C. B., E. Bolger, A. McCarron, M. Roantree, G. de Los Campos, et al., 2019 Benchmarking parametric and machine learning models for genomic prediction of complex traits. G3: Genes, Genomes, Genetics 9: 3691–3702.

Bellot, P., G. de Los Campos, and M. Pérez-Enciso, 2018 Can deep learning improve genomic prediction of complex human traits? Genetics 210: 809–819.

Betancourt, M., 2017 A conceptual introduction to hamiltonian monte carlo. arXiv preprint arXiv:1701.02434.

Buniello, A., J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, et al., 2019 The ngdhri-ebi gwas catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic acids research 47: D1005–D1012.

Chen, C., J. P. Steibel, and R. J. Tempelman, 2017a Genome-Wide Association Analyses Based on Broadly Different Specifications for Prior Distributions, Genomic Windows, and Estimation Methods. Genetics 206: 1791 – 1806. Learn how to use estimation from GBLUP to get solutions to EMMAX The use of ROC.

Chen, C., J. P. Steibel, and R. J. Tempelman, 2017b Genome-Wide Association Analyses Based on Broadly Different Specifications for Prior Distributions, Genomic Windows, and Estimation Methods. Genetics 206: 1791–1806.

Cheng, H., R. Fernando, and D. Garrick, 2018a JwAs: Julia implementation of whole-genome analysis software. In Proceedings of the world congress on genetics applied to livestock production, volume 11, p. 859.

Cheng, H., K. Kizilkaya, J. Zeng, D. Garrick, and R. Fernando, 2018b Genomic prediction from multiple-trait bayesian regression methods using mixture priors. Genetics 209: 89–103.

Demetci, P., W. Cheng, G. Darnell, X. Zhou, S. Ramachandran, et al., 2020 Multi-scale genomic inference using biologically annotated neural networks. bioRxiv.

Dempfle, L., 1977 Relation entre BLUP (Best Linear Unbiased Prediction) et estimateurs bayésiens. Genet Sel Evol 9.

Demuth, H. and M. Beale, 2009 Matlab neural network toolbox user’s guide version 6. the mathworks inc.

Duarte, J. L. G., R. J. Cantet, R. O. Bates, C. W. Ernst, N. E. Raney, et al., 2014 Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. BMC bioinformatics 15: 1–11.

Ehret, A., D. Hochstuhl, D. Gianola, and G. Thaller, 2015 Application of neural networks with back-propagation to genome-enabled prediction of complex traits in holstein-friesian and german fleckvieh cattle. Genetics Selection Evolution 47: 22.

Erbe, M., B. Hayes, L. Matukumalli, S. Goswami, P. Bowman, et al., 2012 Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. Journal of dairy science 95: 4114–4129.

Fernando, R., 1998 Genetic evaluation and selection using genotypic, phenotypic and pedigree information. In 6th Wild Cong. Genet. App.Liv. Prod., volume 26, pp. 329–336, Armidale, Australia.
Table 2 Prediction accuracy of NN-Bayes and linear models for real datasets.

| #hidden nodes | Priors | Pig        | Soy        | Spruce     | Switchgrass | Maize       |
|---------------|--------|------------|------------|------------|-------------|-------------|
| 2             | RR-BLUP| 0.534(0.04)| 0.657(0.02)| 0.413(0.04)| 0.865(0.03)| 0.755(0.09) |
|               | BayesA | 0.560(0.05)| 0.655(0.02)| 0.410(0.05)| 0.866(0.03)| 0.754(0.09) |
|               | BayesB | 0.562(0.05)| 0.655(0.02)| 0.411(0.05)| 0.866(0.03)| 0.755(0.09) |
|               | BayesC | 0.533(0.04)| 0.656(0.02)| 0.412(0.05)| 0.865(0.03)| 0.757(0.09) |
|               | BayesL | 0.527(0.04)| 0.655(0.02)| 0.409(0.04)| 0.864(0.03)| 0.755(0.09) |
| 3             | RR-BLUP| 0.529(0.04)| 0.660(0.01)| 0.418(0.05)| 0.865(0.03)| 0.756(0.10) |
|               | BayesA | 0.562(0.05)| 0.659(0.02)| 0.420(0.05)| 0.866(0.03)| 0.757(0.10) |
|               | BayesB | 0.568(0.05)| 0.658(0.02)| 0.420(0.05)| 0.865(0.03)| 0.757(0.10) |
|               | BayesC | 0.529(0.04)| 0.660(0.01)| 0.419(0.05)| 0.864(0.03)| 0.758(0.10) |
|               | BayesL | 0.523(0.05)| 0.658(0.02)| 0.420(0.05)| 0.866(0.03)| 0.756(0.10) |
| 5             | RR-BLUP| 0.530(0.04)| 0.663(0.01)| 0.419(0.05)| 0.861(0.03)| 0.757(0.10) |
|               | BayesA | 0.565(0.05)| 0.660(0.02)| 0.420(0.05)| 0.860(0.03)| 0.760(0.10) |
|               | BayesB | 0.565(0.05)| 0.660(0.02)| 0.418(0.05)| 0.864(0.03)| 0.754(0.10) |
|               | BayesC | 0.529(0.04)| 0.662(0.01)| 0.419(0.05)| 0.857(0.03)| 0.757(0.10) |
|               | BayesL | 0.523(0.05)| 0.662(0.02)| 0.418(0.05)| 0.861(0.03)| 0.757(0.10) |
| 10            | RR-BLUP| 0.531(0.04)| 0.664(0.02)| 0.419(0.05)| 0.859(0.03)| 0.754(0.10) |
|               | BayesA | 0.563(0.04)| 0.660(0.02)| 0.419(0.05)| 0.860(0.03)| 0.753(0.10) |
|               | BayesB | 0.569(0.04)| 0.660(0.02)| 0.418(0.05)| 0.859(0.03)| 0.755(0.10) |
|               | BayesC | 0.530(0.05)| 0.663(0.02)| 0.418(0.05)| 0.860(0.03)| 0.754(0.10) |
|               | BayesL | 0.524(0.05)| 0.662(0.02)| 0.418(0.05)| 0.859(0.03)| 0.756(0.10) |

* #hidden nodes, number of hidden nodes in NN-Bayes. No significant difference among NN-Bayes with difference number of hidden nodes (i.e., 2, 3, 5, 10). Priors, prior used for marker effects. Prediction accuracy, the average of prediction accuracies from 20 replications with standard deviation in the brackets.

† For pig dataset, with significance level=0.05, the priors used for marker effects in BayesA and BayesB had significantly higher prediction accuracy than other priors, as denoted by † symbol.

‡ For soy dataset, with significance level=0.1, NN-Bayes composed of 10 hidden nodes with priors for marker effects in BayesC or RR-BLUP had significantly higher prediction accuracy than conventional linear models, as denoted by ‡ symbol.
Fernando, R., A. Toosi, A. Wolc, D. Garrick, and J. Dekkers, 2017
Application of whole-genome prediction methods for genome-wide
association studies: a bayesian approach. J Agric Biol
Environ Stat 22(2): 172–93.
Fernando, R. L. and D. Garrick, 2013 Bayesian Methods Applied
to GWAS. In Genome-Wide Association Studies and Genomic Predic-
tion, pp. 237–274, Humana Press, Totowa, NJ.
Fisher, R. A., 1918 Xc—the correlation between relatives on
the supposition of mendelian inheritance. Transactions of the
Royal Society of Edinburgh 52: 399–433.
Freebern, E., D. J. Santos, L. Fang, J. Jiang, K. L. P. Gaddis, et al.,
2012 Gwas and fine-mapping of livability and six disease traits
in holstein cattle. BMC genomics 21: 1–11.
Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari,
et al., 2013 Bayesian data analysis. CRC press.
Gianola, D. and R. L. Fernando, 1986 Bayesian methods in ani-
mal breeding 63: 217–244.
Gianola, D. and R. L. Fernando, 2019 A Multiple-Trait Bayesian
Lasso for Genome-Enabled Analysis and Prediction of Com-
plex Traits. Genetics 214:4: 302934.2019 – 331.
Gianola, D., H. Okut, K. A. Weigel, and G. J. Rosa, 2011 Predict-
ing complex quantitative traits with bayesian neural networks:
a case study with jersey cows and wheat. BMC genetics 12: 87.
Glória, L. S., C. D. Cruz, R. A. M. Vieira, M. D. V. de Resende, P. S.
Lopes, et al., 2016 Accessing marker effects and heritability
estimates from genome prediction by bayesian regularized
neural networks. Livestock Science 191: 91–96.
Guan, Y. and M. Stephens, 2011 Bayesian variable selection re-
gression for genome-wide association studies and other large-
scale problems. The Annals of Applied Statistics pp. 1780–
1815.
Habier, D., R. L. Fernando, and J. C. Dekkers, 2007 The impact of
genetic relationship information on genome-assisted breeding
values. Genetics 177: 2389–2397.
Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick,
2011 Extension of the bayesian alphabet for genomic selec-
tion. BMC bioinformatics 12: 186.
Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. God-
dard, 2009a Invited review: Genomic selection in dairy cattle:
Progress and challenges. Journal of dairy science 92: 433–443.
Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and
M. E. Goddard, 2010 Genetic architecture of complex traits
and accuracy of genomic prediction: Coat colour, milk-fat
percentage, and type in holstein cattle as contrasting model
traits. PLoS Genet 6(9): e1001139.
Hayes, B. J., P. M. Visscher, and M. E. Goddard, 2009b Increased
accuracy of artificial selection by using the realized relation-
ship matrix. Genetics research 91: 47–60.
Heffner, E. L., M. E. Sorrells, and J.-L. Jannink, 2009 Genomic
selection for crop improvement. Crop Science 49: 1–12.
Hickey, J. M., T. Chiurugwi, I. Mackay, W. Powell, A. Eggen, et al.,
2017 Genomic prediction unifies animal and plant breeding
programs to form platforms for biological discovery. Nature
genetics 49: 1297.
Hirschhorn, J. N. and M. J. Daly, 2005 Genome-wide associa-
tion studies for common diseases and complex traits. Nature
reviews genetics 6: 95–108.
Kizilkaya, K., R. L. Fernando, and D. J. Garrick, 2010 Genomic
prediction of simulated multibreed and purebred performance
using observed fifty thousand single nucleotide polymor-
phism genotypes. Journal of Animal Science 88: 544–551.
Klein, R. J., C. Zeiss, E. Y. Chew, J.-Y. Tsai, R. S. Sackler, et al., 2005
Complement factor h polymorphism in age-related macular
degeneration. Science 308: 385–389.
Korte, A. and A. Farlow, 2013 The advantages and limitations of
trait analysis with gwas: a review. Plant methods 9: 1–9.
LeCun, Y., Y. Bengio, and G. Hinton, 2015 Deep learning. nature
521: 436–444.
Legarra, A., A. Ricard, and L. Varona, 2018 Gwas by gblup:
single and multilocator emax and bayes factors, with an
example in detection of a major gene for horse gait. G3: Genes,
Genomes, Genetics 8: 2301–2308.
Leshno, M., V. Y. Lin, A. Pinkus, and S. Schoken, 1993 Multi-
layer feedforward networks with a nonpolynomial activation
function can approximate any function. Neural networks 6:
861–867.
Lloyd-Jones, L. R., M. R. Robinson, G. Moser, J. Zeng, S. Beleza,
et al., 2017 Inference on the Genetic Basis of Eye and Skin
Color in an Admixed Population via Bayesian Linear Mixed
Models. Genetics 206: 1113–1126.
Ma, W., Z. Qiu, J. Song, J. Li, Q. Cheng, et al., 2018 A deep convo-
lutional neural network approach for predicting phenotypes
from genotypes. Planta 248: 1307–1318.
Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Predic-
tion of total genetic value using genome-wide dense marker
maps. Genetics 157: 1819–1829.
Min, S., B. Lee, and S. Yoon, 2017 Deep learning in bioinformatics.
Briefings in bioinformatics 18: 851–869.
Moser, G., S. H. Lee, B. J. Hayes, M. E. Goddard, N. R. Wray,
et al., 2015 Simultaneous discovery, estimation and prediction
analysis of complex traits using a bayesian mixture model.
PloS Genet 11: e1004969.
Nelson, R.M., M. E. Pettersson, and O. Carlberg, 2013 A century
after fischer: time for a new paradigm in quantitative genetics.
Trends in Genetics 29: 669–676.
Norris, J. R., 1998 Markov chains. Number 2, Cambridge univer-
sity press.
Okut, H., D. Gianola, G. J. Rosa, and K. A. Weigel, 2011 Pre-
diction of body mass index in mice using dense molecular
markers and a regularized neural network. Genetics research 93:
189–201.
Okut, H., X.-L. Wu, G. J. Rosa, S. Bauck, B. W. Woodward,
et al., 2013 Predicting expected progeny difference for mar-
bling score in angus cattle using artificial neural networks and
bayesian regression models. Genetics Selection Evolution 45:
1–13.
Ozaki, K., Y. Ohnishi, A. Iida, A. Sekine, R. Yamada, et al., 2002
Functional snps in the lymphotoxin-α gene that are associated
with susceptibility to myocardial infarction. Nature genetics
32: 650–654.
Park, T. and G. Casella, 2008 The bayesian lasso. Journal of the
American Statistical Association 103: 681–686.
Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, et al., 2011
proc: an open-source package for r and s+ to analyze and
compare roc curves. BMC bioinformatics 12: 1–8.
Salvatier, J., T. V. Wiecki, and C. Fonesbeck, 2016 Probabilis-
tic programming in python using pymc3. PeerJ Computer
Science 2: e55.
Schmidhuber, J., 2015 Deep learning in neural networks: An
overview. Neural networks 61: 85–117.
Sharma, A., J. S. Lee, C. G. Dang, P. Sudrajad, H. C. Kim, et al.,
2015 Stories and challenges of genome wide association stud-
ies in livestock—a review. Asian-Australasian journal of ani-
Sorensen, D. and D. Gianola, 2007 Likelihood, Bayesian, and MCMC methods in quantitative genetics. Springer Science & Business Media.

Strandén, I. and D. J. Garrick, 2009 Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. J Dairy Sci 92: 2971–2975.

Szymczak, S., J. M. Biernacka, H. J. Cordell, O. González-Recio, I. R. König, et al., 2009 Machine learning in genome-wide association studies. Genetic epidemiology 33: S51–S57.

van Bergen, G. H., P. Duenk, C. A. Albers, P. Bijma, M. P. Calus, et al., 2020 Bayesian neural networks with variable selection for prediction of genotypic values. Genetics Selection Evolution 52: 1–14.

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. Journal of dairy science 91: 4414–4423.

Visscher, P. M., M. A. Brown, M. I. McCarthy, and J. Yang, 2012 Five years of gwas discovery. The American Journal of Human Genetics 90: 7–24.

Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, et al., 2017 10 years of gwas discovery: biology, function, and translation. The American Journal of Human Genetics 101: 5–22.

Waldmann, P., 2018 Approximate bayesian neural networks in genomic prediction. Genetics Selection Evolution 50: 70.

Wang, C. S., J. J. Rutledge, and D. Gianola, 1994 Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. Genet. Sel. Evol. 26: 91–115.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. Muir, 2012 Genome-wide association mapping including phenotypes from relatives without genotypes. Genetics Research 94: 73–83.

Wang, T., Y.-P. P. Chen, P. J. Bowman, M. E. Goddard, and B. J. Hayes, 2016 A hybrid expectation maximisation and mcmc sampling algorithm to implement bayesian mixture model based genomic prediction and qtl mapping. BMC genomics 17: 1–21.

Wang, Y., X. Mi, G. J. Rosa, Z. Chen, P. Lin, et al., 2018 an r package for fitting sparse neural networks with application in animal breeding. Journal of animal science 96: 2016–2026.

Zhao, T., R. Fernando, D. Garrick, and H. Cheng, 2020 Fast parallelized sampling of Bayesian regression models for whole-genome prediction. Genetics Selection Evolution 52: 1–11.