



Article A Method for Improving the Performance of Ensemble Neural Networks by Introducing Randomization into Their Training Data

Bryn Richards * and Nwabueze Emekwuru 🗈

School of Mechanical, Aerospace and Automotive Engineering, Faculty of Engineering, Environment and Computing, Coventry University, Priory Street, Coventry CV1 5FB, UK * Correspondence: richa377@uni.coventry.ac.uk

Abstract: We propose a methodology for training neural networks in which ensembles of undertrained neural networks are used to obtain broadly repeatable predictions, and we augment their performance by disrupting their training, with each neural network in the ensemble being trained on a potentially different data set generated from the base data by a method that we call randomization with full range sampling. Sleep habits in animals are a function of innate and environmental factors that determine the species' place in the ecosystem and, thus, its requirement for sleep and opportunity to sleep. We apply the proposed methodology to train neural networks to predict hours of sleep from only seven correlated observations in only 39 species (one set of observations per species). The result was an ensemble of neural networks making more accurate predictions (lower mean squared error) and predictions that are more robust against variations in any one input parameter. The methodology presented here can be extended to other problems in which the data available for training are limited, or the neural network is to be applied, post-training, on a problem with substantial variation in the values of inputs (independent variables).

Keywords: neural network; ensemble; bootstrap; full range sampling; sleep patterns; animal; species comparison; factor analysis; short-wave sleep; paradoxical sleep

1. Introduction

The present work investigates the application of simple neural networks to problems with limited data sets. We argue that a simple neural network, such as a feed-forward perceptron neural network with a single hidden layer, can be extremely useful as a predictive tool in applications that require (1) quick or limited measurement of predictor variables (neural network inputs) or (2) quantifiable error (or sensitivity to variations in neural network inputs). Neural networks have been applied abundantly to the analysis and prediction of animal behavior [1], especially where data are abundant without clear emergent patterns, such as the problem of tracking species movement [2,3] or analyzing posture via body position data or images [4,5]. We apply our methodology to animal sleep patterns.

Factor analysis (or principal component analysis) involves reducing the dimensionality of problems to make them more readily comprehensible or solvable. Factor analysis could be contrasted with neural networks, where a wide range of inputs (factors) are typically desirable, but there is a parallel to be drawn in the fact that neural networks generally translate many variables (the values and weights of numerous intermediate layer nodes) into fewer (the values of output nodes) using linear mathematical operations.

Factor analysis involves defining a collection of real-world observations *X* as a linear function of a smaller number of factors *F*.

$$X = LF + M + \varepsilon \tag{1}$$



Citation: Richards, B.; Emekwuru, N. A Method for Improving the Performance of Ensemble Neural Networks by Introducing Randomization into Their Training Data. *Knowledge* **2023**, *3*, 307–319. https://doi.org/10.3390/ knowledge3030021

Academic Editor: Jose María Merigo

Received: 1 November 2022 Revised: 8 April 2023 Accepted: 26 June 2023 Published: 28 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). where *L* is a 'loading' matrix that defines the relative importance of each factor in *F* on each observation in *X*. *M* is the mean of *X*, and ε is random error. In other words, the hypothesis is that deviations in observations *X* from the mean *M* are understandable as random errors plus the systematic influence of factors *F*. The factors *F* operate linearly on *X* according to the proportions *L*.

Equation (1) is useful as an insight into the data if *F* contains fewer elements than *X*. If we measure *p* observations that describe each of *n* data points, then *X* is a matrix of size *p* by *n*. The general problem of predicting a value $x_{i,m}$ involves defining it as a function of all the other measurements in *X*. Factor analysis can reduce this to a simpler problem by defining any value $x_{i,m}$ as a linear function of a smaller set of factors. In Equation (1), *F* is a matrix *k* by *n*, where *k* is less than *p*. We hope to identify *k* underlying factors that explain all of the *p* attributes of an individual *n*.

Allison and Cicchetti [6] performed factor analysis on nine observations of 39 species of animals in an effort to investigate sleep patterns. *X* is a matrix of dimension nine by 39. Two of the observations were average daily hours of two kinds of sleep: short-wave sleep (SWS), which is sometimes called 'non-dreaming' sleep, and paradoxical sleep (PS), which is sometimes called 'dreaming' sleep. The researchers wanted to discover whether the average daily hours of these two kinds of sleep could be predicted from the other seven observations. The other seven observations were brain weight, body weight, lifespan, gestation time, degree of exposure when sleeping (an arbitrary index), the propensity to be preyed upon (an arbitrary index), and overall danger from predators (an arbitrary index).

Using factor analysis, Allison and Cicchetti [6] hypothesized that the propensity toward SWS and PS could be put in terms of two factors: a factor related to an animal's size and another factor related to an animal's typical degree of subjection to danger in its natural habitat. They showed that the two factors were substantially independent by showing that brain weight, body weight, lifespan, and gestation time were closely correlated to the 'size' factor (>0.8 correlation), while the degree of exposure when sleeping, the propensity to be preyed upon, and overall danger from predators were closely correlated to the 'danger' factor (>0.7 correlation). Furthermore, they found that the tendency to short-wave 'non-dreaming' sleep corresponded to small size (-0.630 correlation to 'size'), and paradoxical 'dreaming' sleep corresponded to an absence of danger (-0.689 correlation to 'danger').

The factor analysis of Allison and Cicchetti [6] produces

$$\dot{X}_{9\times39} = L_{9\times2}F_{2\times39} + M_{9\times39} \tag{2}$$

Since a neural network is a system for calculating outputs as linear combinations of inputs, Equation (2) is similar to a neural network with two hidden nodes (the factors) being trained on 39 data points. Since Allison and Cicchetti [6] were ultimately interested in predicting two observations (SWS and PS) from the other seven, the overall structure of the neural network is seven input neurons, two hidden neurons, and two output neurons. We noted that factor analysis has been used in combination with neural networks in other studies, mainly to reduce the dimensionality of problems to be solved by neural networks [7–9]. We also found that neural networks have been used to model sleep [10] and to classify sleep [11], but they have not been applied widely to the problem of correlating species SWS and PS requirements to other observable factors in the manner of Allison and Cicchetti [6].

Allison and Cicchetti [6] found that their two factors accounted for over 80% of the total variance observed in their data. When we trained neural networks with 7 input nodes, 2 hidden nodes, and 2 output nodes, we achieved a mean squared error of 51% of the average for SWS and 50% of the average for PS. These numbers represent the average performance of 100 neural networks of two hidden nodes each trained on the same input data.

We noted that mean squared error, *MSE*, and variance, *V*, are equivalent when bias *B* is assumed to be zero.

$$B = \sum \hat{X} - M, \tag{3}$$

$$V = \sum \left(\hat{X}^2 \right) - \left(\sum \hat{X} \right)^2,\tag{4}$$

$$MSE = \sum \left(\left(\hat{X} - M \right)^2 \right) = \sum \left(\hat{X}^2 \right) - 2M \sum \hat{X} + M^2 = \sum \left(\hat{X}^2 \right) - \left(\sum \hat{X} \right)^2 + \left(\sum \hat{X} - M \right)^2 = V + B^2.$$
(5)

Therefore, we found that a neural network of two hidden nodes does not perform as well as the factor analysis of Allison and Cicchetti [6]. We sought to train a more complex neural network to compare with Allison and Cicchetti [6]. The limited available training data makes the training of complex neural networks difficult, so we used ensemble neural networks [12,13] to address the danger of over-training. Finally, we introduced variations into the training data to speculate as to how measurement error and potentially even variations between individuals could impact the usefulness of particular observations in predicting SWS and PS.

2. Methodology Introduction

Figure 1 is a flow chart introducing the methodology of this study. First, an ensemble of neural networks is created so that we are better able to cope with the risk of under-training any one network on the extremely limited data available. Next, a hyperparameter study is conducted to select the number of hidden nodes (using a single-hidden-layer network with hidden-layer node numbers in the range 2–20). Having selected 8 hidden layer nodes, we perform an "impacts study", being somewhat analogous to the correlation coefficients study conducted by Allison and Cicchetti [6] and informing their factor analysis study. Finally, the impacts study is repeated using the same ensemble of 100 neural networks but with a different training approach, in which each of the neural networks is trained on a different randomized permutation of a training data set augmented to simulate resampling with sample error. The two impact studies, one based on the original data of Allison and Cicchetti [6] and one based on the augmented data, are compared to assess the impact upon the robustness of the ensemble neural network training state of incorporating (simulated) sample error into the training set.



Figure 1. An overview of the methodology of this study.

2.1. Creating a Neural Network Solution to Predict Sleep

We train an ensemble of 100 neural networks. Each neural network is expected to be undertrained with a maximum of 20 training cycles through the 39 available data points (species) in the study of Allison and Cicchetti [6]. The data are obtained from StatLib [14]. MATLAB's Neural Network Toolbox [15] is used with the Levenberg–Marquardt training algorithm.

2.1.1. Ensemble MSE Calculation

- 1. **For** each neural network $NN_g = NN_1 NN_{100}$ **Do**
 - a. **Train** NN_g through 20 epochs.
- 2. EndFor
- 3. For each species $S_n = S_1 S_m$ Do
 - a. For each neural network $NN_g = NN_1 NN_{100}$ Do
 - i. **Run** NN_g with input data from S_n .
 - ii. **Calculate** the squared error in SWS and in PS (using output from NN_g).
 - b. EndFor
 - c. **Calculate** (for the species S_n) the mean (across all $NN_1 NN_{100}$) of the squared error in SWS and the mean (across all $NN_1 NN_{100}$) of the squared error in PS.
- 4. EndFor
- 5. **Calculate** the mean (across all species $S_1 S_m$) of the MSE in SWS predictions.
- 6. **Calculate** the mean (across all species S_1 – S_m) of the MSE in PS predictions.

2.1.2. Node Sensitivity Study

- 1. **For** number of hidden layer nodes = 2–20 **Do**
 - a. **Call** the ensemble MSE calculation described in Section 2.1.1.
- 2. EndFor
- 3. **Compare** the MSE in SWS predictions and PS predictions across all node numbers.

A node sensitivity study supports us in selecting 8 hidden nodes. Alternative choices of the number of neural networks in the ensemble and the number of training iterations are also assessed in a hyperparameter sensitivity study to confirm that the results were reasonably independent of these choices.

2.2. Assessing the Impact of Each Independent Variable on the SWS and PS Predictions of an Ensemble of 100 Neural Networks Trained on the Same Data Set

The sensitivity of each dependent variable (SWS and PS) is assessed by perturbing inputs using the method of Sobhanifard [16]. Each input is perturbed by 5% to assess its relative impact on the outputs.

Impacts Study Based on 5% Perturbations of Inputs Acting on An Ensemble of 100 Neural Networks Trained All on the Same Data Set

- 1. **Call** the ensemble MSE calculation described in Section 2.1.1.
- 2. **For** each input variable i = 1-7 **Do**
 - a. **For** each species $S_n = S_1 S_m$ **Do**
 - i. **Perturb** the variable $x_{i,n}$ by 5%
 - ii. **Calculate** the MSE in the SWS predictions and the MSE in the PS predictions when feeding the perturbed variable as an input to the ensemble of NN_{1-100} already-trained neural networks.
 - b. EndFor
- 3. EndFor
- 4. **Calculate** the proportional MSE impact of each perturbation in the input variable averaged across all species S_1 – S_m .

5. **Normalize** the MSE impacts separately for each output variable (SWS and PS), with 100 being the greatest impact.

2.3. Assessing the Impact of Each Independent Variable on the SWS and PS Predictions of an Ensemble of 100 Neural Networks Trained on a Diversified Training Set

We wish to assess the change in the robustness of neural network predictions if we incorporate sample error into the training data set (using a method derived from the bootstrap method of Efron [17]) rather than averaging the sampled data prior to training. Specifically, as a proxy for robustness, we assess the change in the sensitivity of neural network predictions to variations in any one input variable using (1) training data that has been averaged prior to training versus (2) training data that includes sample error. Since we do not have access to the original raw data set, including sample error, we simulate it in the following way. We use Excel's NORMINV function [18] to generate a new data set of 10 samples per species in which each sample is randomized about a mean value (the value of the original 1 sample for that species) with a standard deviation equal to 5% of the mean. For reference, assuming a normal distribution, this means that 95% of generated sample values will fall within 10% of the corresponding mean value for that attribute in that species. The assumption of normal distribution may not be valid for all of the seven input attributes, and the further assumption that 95% of samples should fall within 10% of the mean may also be flawed. However, we lack the data to inform a better choice.

Our goal is to test whether incorporating variability in the input increases the robustness of predictions of the neural network ensemble, and in the absence of published variability in the data, we introduce random variability into the data. We do not know, from the published data of Allison and Cicchetti [6], the actual distribution of repeat measurements or the distribution of measurements taken across individuals within a given species. We hypothesize that variability if taken into account in training a neural network ensemble as described in this study, would improve the robustness of the predictions of that ensemble. We would like to test this hypothesis using a data set with simulated variability. Ideally, we would prefer to introduce a degree of variability for each species and each trait that is typical of the variability from individual to individual for that trait across that species (assuming that individual-to-individual variability probably dominates over sample-to-sample repeat sampling variability). We do not have a source of data on the individual-to-individual variability per species per trait, and it is beyond the scope of this study to produce these data. Therefore, we make the assumption of 5% across the board, recognizing that this limits the absolute accuracy of any results. However, we do maintain that the qualitative findings of this study (including demonstrating the viability of a method of training an ensemble neural network) remain valid despite this arbitrary assumption.

2.3.1. Simulating Sample Error in the Training Data with Full Range Sampling

- 1. Set standard deviations $\sigma_{i,m} = \{\sigma_{1,1}, \sigma_{2,1}, \dots, \sigma_{p,1}, \sigma_{1,2}, \dots, \sigma_{p,2}, \dots, \sigma_{p,n}\}$ for p = 9 observations of n = 39 species. This study uses $\sigma_{i,m} = 0.05\mu_{i,m}$ for i = 1-p and n = 1-n. This could be improved with attention to each attribute and species, but 5% of the mean is considered reasonable. This is at the lower end of standard deviations among brain sizes and lifespans in mammals (commensurate with cows) but at the upper end of standard deviations of gestation time (commensurate with gorillas and humans). In total, 5% of the mean represents approximately 15–45 min of short-wave sleep or up to 30 min of paradoxical sleep for most of the animals in this study.
- 2. **For** j = 1-10 **Do**
 - a. **Perturb** every variable $x_{i,m}$ (independent and dependent variables ranging from I = 1-p and species m = 1-n) in the training set by a randomly generated amount whereby the mean value produced by the random number generator is 1 and the standard deviation is σ_i . Save the perturbed value in an expanded input set of 390 training data points (n = 39 species by $j_{max} = 10$ perturbations).

3. EndFor

- 4. For each neural network $NN_g = NN_1 NN_{100}$ Do
 - a. For each species $S_n = S_1 S_m$ Do
 - i. Set j = a random integer $j \in \mathbb{N}, j \in \{1, 2, \dots, 10\}$
 - ii. **Select** data set *j* for species S_n in the training set of neural network NN_g .
 - b. EndFor
 - **Train** NN_g through 20 epochs.
- c. Tr 5. EndFor

2.3.2. Impacts Study Based on 5% Perturbations of Inputs Acting on an Ensemble of 100 Neural Networks Each Trained on a Data Set Randomized with Full Range Sampling (FRS)

- 1. **Call** the function of Section 2.3.1 to simulate sample error in the training data with FRS.
- 2. **For** each independent variable i = 1-7 **Do**
 - a. For each species $S_n = S_1 S_m$ Do
 - i. **Perturb** the variable $x_{i,n}$ by 5%
 - ii. **Calculate** the MSE in the SWS predictions and the MSE in the PS predictions when feeding the perturbed variable as an input to the ensemble of NN_{1-100} already-trained neural networks.
 - b. EndFor
- 3. EndFor
- 4. **Calculate** the proportional MSE impact of each perturbation in the input variable averaged across all species.
- 5. **Normalize** the MSE impacts separately for each output variable (SWS and PS), with 100 being the greatest impact.

This method is adapted from the bootstrap confidence interval of Efron [17] applied to neural networks by Trichakis et al. [19] with full range sampling developed by Richards and Emekwuru [20] for problems with a limited or periodic range from which to draw input data. The method ensures that every species S_{1-39} is represented in every neural network NN_{1-100} , but the set of NN_{1-100} is more diverse than in Section 2.1.1 because each is trained on a potentially different randomized subset of a new, larger data set created by adding random perturbations to the original data set.

3. Results

As noted above, we conducted a hyperparameter selection exercise at the outset of this work. The mean squared error of the ensemble neural network approach is presented in Table 1 against the number of nodes used in the hidden layers of the networks. The hyperparameter selection study informs the complexity of neural networks used later in the study of the robustness of ensemble neural network predictions. We also evaluated the number of neural networks to be used in the neural network ensemble (a value of 100 was selected), the learning rate, and the number of training epochs to be allowed in the training of each of the ensemble neural networks (a value of 20 was selected). Only the node number hyperparameter sensitivity results are presented here since node number is typical of particular interest to AI researchers, and an increasing node number has an important, non-linear impact on computation time (which means that node number is to be minimized where possible).

Neural Network	Mean Squa	ared Error *	
Node Number	SWS	PS	
2	48.59%	50.74%	
3	46.87%	42.06%	
4	39.13%	36.34%	
5	39.19%	29.87%	
6	38.92%	29.75%	
7	33.98%	32.54%	
8	33.13%	26.41%	
9	35.11%	25.37%	
10	31.33%	27.04%	
11	31.16%	23.27%	
12	32.73%	22.35%	
13	32.56%	21.38%	
14	23.60%	17.56%	
15	31.50%	19.45%	
16	28.51%	19.15%	
17	28.55%	16.78%	
18	30.02%	18.53%	
19	29.23%	15.32%	
20	25.20%	17.78%	

Table 1. MSE of slow-wave sleep (SWS) and paradoxical sleep (PS) predictions with node number.

* MSE is presented here as a percentage of the mean value of the prediction (hours slept per 24-h cycle in SWS or PS). It is also an average MSE over an ensemble of 100 neural networks (each neural network trained on the same input data, without perturbations).

Having selected our hyperparameters, we move on to the study proper, in which we assessed the sensitivity of the neural network predictions to each of the inputs. The sensitivity of neural network outputs to inputs is presented in Table 2. These are the results for an ensemble of neural networks, all trained on the same data by the methodology of Section 2.2. We did not have a means of quantitatively comparing our normalized MSE impacts with the correlation coefficients of Allison and Cicchetti [6], but we noted that this analysis was qualitatively similar. Therefore, we expected a similar ranking in the relative importance of input variables (but not agreement in the quantitative degree of correlation or impact of any particular input variable).

Table 2. Relative influence of inputs on short-wave sleep (SWS) and paradoxical sleep (PS) with neural network training data in which randomization has not been introduced (each of the 100 neural networks in the ensemble trained from an identical data set).

	Present Study ¹ Normalized MSE Impacts		Allison and Cicchetti ² Correlation Coefficients	
Input Variable	SWS	PS	SWS	PS
Body weight	10	2	-0.712	-0.370
Brain weight	15	4	-0.679	-0.435
Lifespan	33	14	-0.377	-0.342
Gestation time	33	24	-0.589	-0.651
Predation index	88	57	-0.369	-0.536
Sleep exposure index	58	31	-0.580	-0.591
Overall danger index	100	100	-0.542	-0.686

¹ The impact on mean squared error (normalized) is assessed with perturbations of 5% to input variables. Inputs are perturbed to assess their impact, but inputs are not perturbed during training. All 100 neural networks are trained on the same data without perturbations. MSE impacts are normalized whereby 100 represents the largest impact. ² Correlation coefficients of SWS and PS with input variables, originally published by Allison and Cicchetti [6].

Furthermore, we wished to assess how the sensitivity of neural network outputs to inputs changes when the ensemble neural network is trained on a randomized data set

(simulating sample error). The sensitivity study is repeated in Table 3, but in this case, the ensemble of neural networks was trained according to the methodology of Section 2.3.2 using randomized data sets prepared according to the methodology of Section 2.3.1.

Table 3. The impacts of input variable variations on mean squared error of sleep predictions (short-wave sleep and paradoxical sleep) from a neural network ensemble trained on diversified data randomized with 5% variations and full range sampling into the ensemble training sets.

	Normalized MSE Impacts on Ensembles with FRS Diversified Training Data ¹		Change in Absolute MSE Impact When Using FRS Diversified Training Data ²	
Input Variable	SWS	PS	SWS	PS
Body weight	6	10	-0.61×10^{-2}	0.39×10^{-2}
Brain weight	16	24	$-0.46 imes10^{-2}$	$1.07 imes 10^{-2}$
Lifespan	42	58	$-0.61 imes10^{-2}$	$1.97 imes 10^{-2}$
Gestation time	35	41	$-1.07 imes10^{-2}$	$-0.59 imes10^{-2}$
Predation index	100	95	$-2.47 imes10^{-2}$	$-1.63 imes 10^{-2}$
Sleep exposure index	50	98	-2.62×10^{-2}	2.32×10^{-2}
Overall danger index	83	100	$-4.73 imes 10^{-2}$	$-7.44 imes 10^{-2}$
Totals			$-12.5 imes 10^{-2}$	$-3.91 imes 10^{-2}$

¹ The impact on mean squared error is assessed with perturbations of 5% to input variables. Results are normalized, whereby 100 represents the largest impact. ² Changes in impact are calculated relative to the impacts of the same input variable variations on ensemble predictions created without diversified training sets. (See "Normalized MSE impacts" from Table 2) The published value is (a - b), where *a* is the MSE for a prediction using the ensemble with diversified training data, and *b* is the MSE for a prediction using an ensemble with uniform training data.

4. Discussion

4.1. Node Number Sensitivity Study

The node number sensitivity study (see Table 1) did not produce a local minimum, but diminishing returns were observed beyond four nodes. Eight nodes were selected as a conservative choice, and processing times with ensembles of 100 neural networks with eight nodes were reasonable on MATLAB's cloud computing platform.

The node number sensitivity study did not produce any result at or below 20 nodes that is competitive with the factor analysis exercise of Allison and Cicchetti [6], where variation was 80% predictable from two factors. This may be explained by the use of a varimax rotation [21] by Allison and Cicchetti [6] to ensure that their factors were uncorrelated. The perceptron neural network used here did not incorporate this complexity.

The node number sensitivity study (see Table 1) did not produce a local minimum, but diminishing returns were observed beyond four nodes. Based on Table 1, the choice of eight nodes was made for the neural networks trained for impact studies. This choice seemed conservative (being large enough that MSE is relatively stable as a function of node number and nevertheless small enough to allow quick training of ensembles of 100 neural networks using MATLAB's cloud computing service).

4.2. Neural Network Ensembles

Without neural network ensembles, this study would not have produced comprehensible results. A single neural network of the complexities studied here (seven inputs and two outputs, with as few as two hidden nodes) cannot be adequately trained on the limited available data, over-training or under-training will result. This study followed a methodology of erring toward under-training (generalizability) and refining model accuracy by creating ensembles of neural networks. Although each network within the ensemble is expected to be under-trained, the use of an ensemble brings the ensemble prediction closer to the observed values of SWS and PS. In the sensitivity study (results not tabulated here), higher numbers of neural networks within the ensemble tended toward lower MSE on ensemble predictions, although at 100 neural networks, the ensemble was not substantially sensitive to small changes in the number of networks (hence the choice of 100 neural networks). Likewise, a sensitivity study on the number of training epochs revealed that, at 20 epochs, although the networks tended strongly toward under-training, small changes in the epoch number did not produce significant changes in the MSE on ensemble predictions.

The neural network ensembles trained in this study related both types of sleep strongly to predation and danger. In contrast, Allison and Cicchetti [6] found that paradoxical sleep (PS) correlated well with danger measures, and short-wave sleep (SWS) correlated well with animal size measures. Later studies [22] have tended to look at total sleep and the proportion of total sleep time that is spent in a state of paradoxical sleep or dream sleep. Thus, later studies tend to examine PS + SWS and PS/(PS + SWS) rather than PS and SWS as independent observations. Allison and Cicchetti [6] pointed out that PS and SWS are not mutually exclusive and that one species may have more or less total sleep (of all kinds) than another and therefore, a species achieving high levels of SWS does not exclude high levels of PS or vice versa. However, Zepelin et al. [22] characterized total sleep (PS + SWS) as a function of an animal's opportunity to sleep safely and PS (dream sleep) as a function of an animal's need (driven by the complexity of its brain). Thus, SWS itself is a residual term, according to Zepelin et al. [22]; it is the excess opportunity for safe sleep (in excess of the need for dream sleep), and as such, it may (we hypothesize) be less predictable than PS. Therefore, we conclude that the relatively good agreement in the results of this study with the results of Allison and Cicchetti [6] in identifying predictors for PS is more significant than the relatively poor agreement in identifying predictors for SWS. This study agrees with Allision and Cicchetti [6] on the top predictor of PS (overall danger) and the top 3 predictors overall (overall danger, predation index, and sleep exposure).

Ultimately, the goal of this study was not to improve upon the conclusions of Allison and Cicchetti [6] in respect of animal sleep. Rather, we sought to show that neural networks can be trained on the limited data of Allison and Cicchetti [6] and that the resulting training state can be reproducible and robust (especially with ensemble neural networks and fullrange sampling with raw data, including sample error). The analogy between neural network analysis and factor analysis was of interest, and we sought to demonstrate how robustly-trained ensembles of neural networks can be used to predict the significance of real inputs on outputs in the real systems that they model.

4.3. Randomized Training Sets with Full Range Sampling

The method of training each of the ensemble neural networks on a different randomized training data set with full range sampling has made predictions (1) more accurate and (2) more robust. Claim 1, that the predictions are more accurate, is supported by Table 3, where columns 4 and 5 show overall improvements in MSE for both SWS predictions and PS predictions (reductions in the absolute MSE impact of inputs on outputs in 10 out of 14 permutations of inputs with outputs). Claim 2, that the predictions are more robust, is also supported by Table 3, where the dependency of both SWS predictions and PS predictions on the "overall danger index" (the strongest predictor of both SWS and PS in Table 2) dropped substantially. This means that neither prediction is as heavily dependent on this one input. The introduction of randomized training data with FRS has made both predictions (SWS and PS) less dependent on any input than both had been dependent on the "overall danger index" when predicted by a homogeneously trained ensemble of neural networks.

The addition of randomized training data sets notably brought weight, lifespan, and sleep exposure factors more prominently into play as predictors of PS, whereas PS was predicted strongly by only the "predation index" and the "overall danger index" before the introduction of randomized training data. Referring to the factor analysis work of Allision and Cicchetti [6], we noted that "sleep exposure" was considered as a size factor rather than a danger factor because an animal's size is one of the most important factors in determining whether it is generally able to find a sheltered setting for sleep. Taken together, we found that the addition of randomized training data with FRS allowed factors related

to animal size to become more significant in predicting paradoxical sleep (sleep typically associated with predators).

The ensemble neural networks trained on randomized input data with FRS have the potential to be more useful in practice than neural networks and other correlation strategies that do not account for variability in inputs because they are potentially more robust against small changes in input values. For example, while an average species parameter (such as species average bring weight) is unlikely to change quickly, researchers could elect to use a local population average brain weight instead of a global species average. If a model is highly sensitive to brain weight, then this choice will be of paramount importance, whereas a model less sensitive to brain weight might be less prone to error in practice. A neural network trained on a data set that includes a slightly lower than true mean brain weight and a slightly higher than true mean brain weight will tend to be more robust against variations in brain weight than one trained only on true mean brain weight. We observed this effect in the impact scores for predicting PS in Table 3, where the scores became more evenly distributed (closer to 100) after introducing variations into the training set. The same general trend can be observed for SWS, where Table 3 shows a more even set of scores than Table 2. The effect was less pronounced than for PS but still qualitatively similar. The difference in degree could simply mean that danger factors are simply more significant for predicting SWS than PS.

Again, our focus in this work was to demonstrate a methodology for the application of neural networks to factor analysis studies in instances with limited available training data. We argued that the viability of the method is qualitatively supported by the results. We were less interested in quantitative predictions in respect of specific animal sleep habits.

4.4. Interpretations of Randomised Data and Variations in Data

Ensemble neural networks with heterogenous, randomized training data selected with full range sampling using the methodologies presented here were expected to be more robust against random error in the input variables, meaning that their predictions were expected to vary less in response to the presence of errors or other variations in the values of input data. This is because they were trained on examples of variations in input data and achieved better MSE by learning to be more responsive to a broad combination of inputs and less responsive to any one input. This had implications for applications of such ensemble neural networks to noisy or uncertain data, where the neural networks produced more consistent predictions when fed noisy data or offered tighter confidence intervals (smaller standard deviation on output values) when making predictions on inputs that varied over a confidence interval range.

Whether the benefit of the ensemble neural network can be extended to making predictions for individual members of a species is not clear. The data used here contain species averages, and it is not necessarily valid to compute an SWS prediction or a PS prediction from an individual's specific metrics using a neural network trained on cross-species comparative data. A neural network trained on collections of data from many individuals across species (rather than from species average data) could have more applications, and the addition of full range sampling ensemble neural networks with randomized inputs (or simply raw per-individual data containing measurement error) could potentially be used for predicting individual sleep patterns from group training data within or across species.

5. Conclusions

We presented methodologies for coping with very limited training data (only 39 data points) when training a relatively complex neural network (seven inputs and two outputs). This was based on the ensemble neural network method with under-training (stopping training at a pre-defined epoch number). We applied randomized data selected for the ensemble of neural networks using full range sampling to re-train the ensemble on a more diverse set of points.

We showed that the mean squared error on predictions of both short-wave sleep and paradoxical sleep improved (were reduced) when the ensemble neural networks were trained on a heterogeneous training data set that had been randomized using full range sampling. Additionally, the predicted tendency toward each type of sleep was less dependent on any particular input to the neural network, instead making predictions based on a more evenly distributed set of weighting factor values drawing from a broader set of inputs. This suggests that the addition of randomized training data with full range sampling to ensemble neural networks can improve their accuracy in general and robustness in the face of variations (including errors) in input values. This could extend their usefulness in field applications to make predictions based on noisy or uncertain data.

Author Contributions: Conceptualization, B.R. and N.E.; methodology, software, data curation, and investigation, B.R.; writing—original draft preparation, B.R.; writing—review and editing, N.E.; supervision, N.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Tadeusiewicz, R. Neural networks as a tool for modeling of biological systems. *Bio-Algorithms Med. Syst.* 2015, 11, 135–144. [CrossRef]
- 2. Dalzeil, B.D.; Morales, J.M.; Fryxell, J.M. Fitting probability distributions to animal movement trajectories: Using artificial neural networks to link distance, resources, and memory. *Am. Nat.* 2008, *172*, 248–258. [CrossRef] [PubMed]
- Tracey, J.A.; Zhu, J.; Crooks, K.R. Modeling and inference of animal movement using artificial neural networks. *Env. Ecol. Stats* 2011, 18, 393–410. [CrossRef]
- Jeantet, L.; Vigon, V.; Geiger, S.; Chevallier, D. Fully convolutional neural network: A solution to infer animal behaviours from multi-sensor data. *Ecol. Model.* 2021, 450, 109555. [CrossRef]
- Fang, C.; Zhang, T.M.; Zeng, H.K.; Huang, J.D.; Cuan, K.X. Pose estimation and behaviour classification of broiler chickens based on deep neural networks. *Comp. Electron. Agric.* 2021, 180, 105863. [CrossRef]
- Allison, T.; Cicchetti, D.V. Sleep in Mammals: Ecological and Constitutional Correlates. *Science* 1976, 194, 732–734. [CrossRef] [PubMed]
- 7. Pan, Z.X.; Pan, D.J.; Sun, P.Y.; Zhang, M.S.; Zuberbuhler, A.D.; Jung, B. Spectroscopic quantitation of amino acids by using artificial neural networks combined with factor analysis. *Spectrochim. Acta Part A* **1997**, *53*, 1629–1632.
- 8. Zhang, Y.X. Artificial neural networks based on principal component analysis input selection for clinical pattern recognition analysis. *Talanta* **2007**, *73*, 68–75. [CrossRef] [PubMed]
- Ding, S.F.; Jia, W.K.; Su, C.Y.; Zhang, L.W.; Liu, L.L. Research of neural network algorithm based on factor analysis and cluster analysis. *Neural Comput. Appl.* 2011, 20, 297–302. [CrossRef]
- 10. Crick, Francis, and Graeme Mitchison. REM sleep and neural nets. Behav. Brain Res. 1995, 69, 147–155. [CrossRef] [PubMed]
- 11. Tagluk, M.E.; Akin, M.; Sezgin, N. Classification of sleep apnea by using wavelet transform and artificial neural networks. *Expert. Syst. Appl.* **2010**, *37*, 1600–1607. [CrossRef]
- 12. Opitz, D.W.; Shavlik, J.W. Actively Searching for an Effective Neural-Network Ensemble. Connect. Sci. 1996, 8, 3–4. [CrossRef]
- 13. Sagi, O.; Rokach, L. Ensemble learning: A survey. WIREs Data Min. Knowl. Discov. 2018, 8, 255–272. [CrossRef]
- 14. StatLib, Carnegie Mellon University. Available online: https://lib.stat.cmu.edu/datasets/ (accessed on 29 October 2022).
- 15. MATLAB and Statistics Toolbox; ver. R2022a; The Mathworks, Inc.: Natick, MA, USA, 2022.
- 16. Sobhanifard, Y. Hybrid modelling of the consumption of organic foods in Iran using exploratory factor analysis and an artificial neural network. *BJF* **2017**, *120*, 44–58. [CrossRef]
- 17. Efron, B. Bootstrap methods: Another look at the jackknife. Ann. Stat. 1979, 7, 1–26. [CrossRef]
- 18. Microsoft Excel; ver. 16.62; Microsoft Corp.: Redmond, WA, USA, 2022.
- Trichakis, I.; Nikolos, I.; Karatzas, G.P. Comparison of bootstrap confidence intervals for an ANN model of a karstic aquifer response. *Hydrol. Process* 2011, 25, 2827–2836. [CrossRef]
- Richards, B.; Emekwuru, N. Using machine learning to predict synthetic fuel spray penetration from limited experimental data without computational fluid dynamics. In Proceedings of the ICESF International Conference on Energy and Sustainable Futures, Coventry, UK, 7–8 September 2022.

- 21. NIST/SEMATECH e-Handbook of Statistical Methods; NIST: Gaithersburg, MD, USA, 2012. [CrossRef]
- 22. Zepelin, H.; Siegel, J.M.; Tobler, I. Mammalian sleep. Chapter 8. In *Principles and Practice of Sleep Medicine*, 4th ed.; Roth, K.M.H.T., Dement, W.C., Eds.; Elsevier/Saunders: Philadelphia, PA, USA, 2005; pp. 91–100.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.