



Article A Study on a Probabilistic Method for Designing Artificial Neural Networks for the Formation of Intelligent Technology Assemblies with High Variability

Vladimir V. Bukhtoyarov, Vadim S. Tynchenko *^(D), Vladimir A. Nelyub, Igor S. Masich, Aleksey S. Borodulin and Andrei P. Gantimurov

Artificial Intelligence Technology Scientific and Education Center, Bauman Moscow State Technical University, 105005 Moscow, Russia

* Correspondence: vadimond@mail.ru; Tel.: +7-95-0973-0264

Abstract: Currently, ensemble approaches based, among other things, on the use of non-network models are powerful tools for solving data analysis problems in various practical applications. An important problem in the formation of ensembles of models is ensuring the synergy of solutions by using the properties of a variety of basic individual solutions; therefore, the problem of developing an approach that ensures the maintenance of diversity in a preliminary pool of models for an ensemble is relevant for development and research. This article is devoted to the study of the possibility of using a method for the probabilistic formation of neural network structures developed by the authors. In order to form ensembles of neural networks, the influence of parameters of neural network structure generation on the quality of solving regression problems is considered. To improve the quality of the overall ensemble solution, using a flexible adjustment of the probabilistic procedure for choosing the type of activation function when filling in the layers of a neural network is proposed. In order to determine the effectiveness of this approach, a number of numerical studies on the effectiveness of using neural network ensembles on a set of generated test tasks and real datasets were conducted. The procedure of forming a common solution in ensembles of neural networks based on the application of an evolutionary method of genetic programming is also considered. This article presents the results of a numerical study that demonstrate a higher efficiency of the approach with a modified structure formation procedure compared to a basic approach of selecting the best individual neural networks from a preformed pool. These numerical studies were carried out on a set of test problems and several problems with real datasets that, in particular, describe the process of ore-thermal melting.

Keywords: artificial neural networks; ensembles; modeling; probabilistic approach; regression; simulation; structure formation

1. Introduction

Artificial neural networks are one of the most effective approaches of data mining used to solve various types of problems—modeling, classification, and clustering. This study considers the modeling problem in regression formulation, which has been successfully solved in many applications of artificial neural networks, including metallurgy, mechanical engineering, and medicine [1–3].

Nevertheless, the increasing complexity of tasks, due to the increasing technological complexity of the simulated processes and the multifactorial search of effective models, as well as increasing volumes of data processed in the construction of regression models and requirements in regard to the accuracy of such models, are the reasons for the search for approaches to improve the efficiency of regression neural network models and models of other types. In this regard, combinations of models and classifiers, which are called ensembles, are widely used [4–6]. At present, the problem of constructing effective en-



Citation: Bukhtoyarov, V.V.; Tynchenko, V.S.; Nelyub, V.A.; Masich, I.S.; Borodulin, A.S.; Gantimurov, A.P. A Study on a Probabilistic Method for Designing Artificial Neural Networks for the Formation of Intelligent Technology Assemblies with High Variability. *Electronics* **2023**, *12*, 215. https:// doi.org/10.3390/electronics12010215

Academic Editor: Maciej Ławryńczuk

Received: 9 November 2022 Revised: 21 December 2022 Accepted: 26 December 2022 Published: 1 January 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/). semble schemes that provide a high quality of the overall group solution at an increased complexity of the problem is topical.

An ensemble of neural networks is a set of individual neural networks used together to solve a single problem. The idea of combining individual neural networks into an ensemble was first proposed by Hansen and Salamon [7]. In this work, it was shown that the generalization ability of an ANN-based system can be significantly increased by replacing a single neural network with an ensemble. This approach for artificial neural networks has subsequently been significantly developed and successfully used to solve a wide range of practical problems: recognition problems [8,9], medical diagnostics [10,11], the classification of seismic signals [12], and many others [13,14].

There are various directions in the formation of ensemble schemes, among which, in the most general typing, we can distinguish schemes that involve either the monolithic use of a dataset for model construction or its division into "responsibility sets" of individual regressors from an ensemble, as well as schemes of cascade applications of regressors. In general, the design of an ensemble of neural networks includes two stages: The first stage consists of shaping the structure and training of individual neural networks. The second stage includes the selection of those neural networks that will be used in the formation of the overall solution of the ensemble in addition to the determination of the methods and parameters with which to effectively calculate the overall ensemble solution based on the solutions of individual neural networks [6,15–17].

Maintaining diversity in the formation of teams of regressors or classifiers is provided in several ways. Firstly, various basic methods can be used; that is, a heterogeneous ensemble is formed, including models obtained using various approaches. These include neural networks, support vector machines, and classification as well as regression trees [18]. This approach is focused on the fact that different individual solvers will adapt and process the original datasets differently; however, in practice, this is not guaranteed to provide a variety of models and respective solutions. It also increases the overhead costs associated with performing different types of training procedures and balancing the distribution of computing power during the parallel training of models. Therefore, within the framework of this study, oriented, among other things, to obtaining a prototype of a practice-oriented solution for the metallurgical field, we focused on the technology of artificial neural networks.

Secondly, the focus on maintaining diversity in an ensemble can be realized through the formation of solvers of the same type, but also by using different hyperparameters of the corresponding model building method [19]. The result is the construction of models with different data processing structures. In terms of artificial neural networks, this is determined by a different structure, namely the number of layers, neurons on layers, the presence of connections between them, and the types of activation functions in specific neurons. It is this type of ensuring diversity in the ensemble that is considered in this study. Thus, taking into account the choice of artificial neural networks as the basic technology for data analysis, the aim of the present work is to develop and study a method for generating various structures of artificial neural networks for the formation of ensemble regressors. At the same time, along with maintaining diversity, it also seems rational to form a structure based on the requirement of an acceptable basic efficiency of each of the solvers. It was within the framework of this paradigm that the study was carried out, the results of which are described in this article.

Additionally, as a completion of the review of methods for maintaining the diversity of individual solvers for ensemble approaches, we should mention methods based on manipulating the original data samples, particularly bagging, based on the formation of bootstrap subsamples, as well as various boosting schemes aimed at associating with the elements of the samples of the corresponding probabilities to be chosen as learning patterns for each next individual solver [20–23]. Boosting schemes have been studied quite well, with various variants of adaptive and gradient boosting among the most effective. In our study, we decided to encapsulate the boosting scheme at the level of the structures of individual solvers, providing variability due to the structural reconfiguration of models based on the

probabilistic approach to designing artificial neural networks that we developed earlier. To do this, we accomplish the following steps during our study:

- We propose a scheme for redistributing the probabilities of the appearance of neurons of a certain type in specific positions of the neural network. The basic method and the proposed adaptive probabilistic procedure for generating neural network structures as individual solvers are discussed in Section 2.
- We execute a number of numerical experiments to investigate the efficiency of the proposed scheme for the generation of neural network structures in the scope of an ensemble technique, comparing it with other ensemble and single-model techniques. A description of the scheme for performing numerical experiments and their results are given in Section 3.

A possible approach is also the formation of various options for describing objects due, for example, to a randomized selection of a set of attributes, as is done explicitly in the random forest approach. In part, such an approach can also be considered within the framework of the implementation of the chosen direction of research covered in this article, since when forming the structures of neural networks, the structure of the first hidden layer also varies significantly. This, together with the procedure for randomly initializing the weights of the neural network during training, makes it possible to ensure the variability in the use of attributes of training patterns.

2. Materials and Methods

2.1. A Probabilistic Method for Designing the Structure of Artificial Neural Networks

Within the ensemble scheme of use, a relevant problem is the formation of individual solvers, possibly including the use of creating a redundant preselection pool. Let us consider a scheme for forming a preselection for selecting ensemble members and selecting individual regression models in the next section in view of the need to automate the procedure for constructing individual regressors for the efficient use of an ensemble scheme on a wide range of problems. In this section, we consider the proposed approach to forming artificial neural network structures for subsequent training. This problem of the optimal design of artificial neural network structures is relevant for many applications of artificial neural networks, involving their use both individually and as part of ensemble circuits.

The necessity of the formation of rational procedures of artificial neural network structure selection is connected with the fact that, while forming a model for a real dataset, a case, characterizing a particular practical problem that is almost impossible to rationally determine in advance of the structure of a neural network, is supposed to be used. The productivity of the further training of neural network models and, eventually, the quality of model problem solving largely depend on the efficiency of artificial neural network structure selection.

Additionally, there are no universal recommendations for the choice of artificial neural network structure even within one class of neural network architectures. The optimizable parameters of artificial neural network structures, such as the number of hidden layers, the number of neurons on layers, and the type of activation functions, require rational choice within the solution of specific applied problems. The expansion of the field of artificial neural networks requires the formation of effective automated approaches to a choice of their structures, instead of either in addition to possible "manual" step-by-step or arbitrary choices, which require deep expert expertise and are often not possible due to the high dimensions of the search space of optimal architectures of the high degree of task nonlinearity. At the same time, overly complex architectures of neural network models are also bearers of negative effects, associated with overlearning and high computational costs, sensitive even with modern specialized neuroaccelerators.

Therefore, the formalized problem of neural network architecture selection is determined by the following basic factors: it is necessary to determine the number of layers of artificial neural networks, the number of neurons on each layer, and their type. In regard to the latter aspect, in this research we consider the type of activation function used in each neuron. The construction of automated procedures for solving such a problem in the general case is reduced to the solution of the optimization problem, as a criterion in which the quality of the solution of the original regression problem with the trained neural network of the appropriate structure is considered. Additional criteria, which can be composed for a multicriteria optimization problem, can be a criterion of network complexity, a criterion of the average time of calculation by the network by the appropriate structure, or special components, defining the quality in terms of an ensemble scheme of using the current neural network solver.

Taking into account the optimization formulation of the problem of designing artificial neural network structures, one of the widest classes of applied approaches is evolutionary optimization algorithms, which have proven their effectiveness in solving a wide range of optimization problems. Of this class of methods, genetic algorithms and their probabilistic modification are used quite effectively. However, as in other applications of genetic algorithms, it is important that their efficiency should also be ensured by selecting parameters, which are quite numerous. Moreover, these parameters also cannot be determined in advance for a wide class of problems; experimentation and the selection of specific application conditions, as well as the topography of the target function, are required. In this regard, the authors propose a probabilistic method, the description of which is given below.

Taking the genetic algorithm of neural network structure design as a starting point, and taking into account the premises described above, we can highlight the following requirements for new methods for the automated design of neural network structures:

- The application of the method should require a smaller number of parameters to be adjusted;

- The representation of solutions in the method should, if possible, be devoid of disadvantages typical for the representation of solutions in the form of binary strings;

- The problems of sensitivity to permutations and destructive crossing should be solved in the method.

The proposed method for the automatic design of neural network structures is based on the calculation and use of probability estimates, Equation (1):

$$p_{i,i}^k, i = \overline{1, N_l}, j = \overline{1, N_{neuron}}, k = \overline{0, N_F}$$
(1)

where *i* is the number of the hidden layer of the neural network; *j* is the number of neurons on the hidden layer of the network; N_l is the maximum number of hidden layers; N_{neuron} is the maximum number of neurons on the hidden layer; and *k* is an identifier, the value of which is interpreted as follows:

- If *k* = 0, then *p*⁰_{*i*,*j*} is an estimate of the probability that the neuron on the *i*-th layer of the network is absent;
- If k ∈ [1, N_F], then p^k_{i,j} k is an estimate of the probability that the neuron on the *i*-th network layer exists and its activation function is an activation function with the number k from the set of activation functions available to the algorithm. N_F is a power of the set of activation functions, which can be used when designing the structures of neural networks.

The numbering of layers and neurons is carried out according to the number of layers and number of neurons on layers in a full network with parameters N_l and N_{neuron} . As the most complete in the sense of the number of layers and the number of neurons on layers of the neural network architecture, a full multilayer perceptron with the number of hidden layers equal to N_l and the number of neurons on each hidden layer equal to N_{neuron} is used. This architecture allows using the number of hidden layers and the number of neurons on layers to explicitly indicate all possible positions of a neuron in the network.

$$p_{i,j}^0 + \sum_{k=1}^{N_F} p_{i,j}^k = 1$$
⁽²⁾

There is no focus on the types of input variables and their descriptions since we used the standard approach to input data processing, where the numerical parameters of the objects from the sample of the problem, scaled in the range from 0 to 1, are fed to the input of the neural network. For categorical data, if available, the standard one-hot encoding approach is also assumed. Thus, the approach considered in this article does not involve additional manipulations with input data, and is aimed at the rational generation of the structure of neural networks as members of an ensemble.

2.2. Proposed Procedure for Forming Elements of a Neural Network Ensemble Based on the Adaptation of the Probability Distribution of Neuron Activation Functions

One of the key features of problem formation in the form of ensembles is to maintain the diversity of individual solutions in order to ensure the efficiency and generalizability of the overall solution. There are various ways to maintain diversity in ensemble models; these are the structure formation and training of individual predictors on slightly different sets of training data generated from the original set, including so-called boosting schemes. An alternative option is to diversify the structures of the models assumed to be included in an ensemble at the stage of their design. Such a strategy is based on the assumption that basically structurally different models imply the segregation of solution quality in the initial space of regression model construction.

This approach was considered in this research, with the focus on the formation of ensembles on the basis of artificial neural networks. In relation to artificial neural networks, the differences in structure are determined by the differences in their basic parameters. In this study, we consider a set of three structural parameters. This is the number of hidden layers of a neural network, the structure of each layer, which is determined by the number of neurons per layer, and their type (type of activation function). In order to provide the diversification of the structures of generated models, the probabilistic method of artificial neural network structure design, described in the previous section, is supplemented by the following modification scheme: When calculating the probabilities of the presence of a neuron of a certain type on a hidden layer, a special multiplier is used. Its purpose is to decrease the probability of neuron type proportionally to the number of already formed neural networks which have a neuron of this type on this layer.

Equation (3):

$$\widetilde{p}_{i,i}^k, i = d_i^k \cdot p_{i,i}^k \tag{3}$$

where $p_{i,j}^k$, *i* and iterators are defined by Equation (1). The value of the coefficient is calculated by Equation (4):

$$d_{i,j}^{k} = 1 - \frac{n_{i,j}^{k}}{Ensemble \ Size}$$

$$\tag{4}$$

where *n* is the number of neural networks with k-type neurons on the *i*-th layer in the *j*-th position already in the ensemble. Thus, the probability correction factor, d_i^k , is determined for each neuron type (activation function type) with respect to each individual possible neuron position on the hidden layers of the network. The interval of change in the value of such a coefficient is (0;1]. Value 1 corresponds to a situation when there are no neuron networks in the ensemble or preliminary pool, for which on the *i*-th layer there is a neuron with the activation function of *k*-type neurons in Equation (4).

For the first network of the preliminary pool the coefficient is equal to one for all types of neurons for all possible positions on the layers, after which nets are recruited into the preliminary pool or immediately into the ensemble.

2.3. Evolutionary Method for Forming Ensembles of Artificial Neural Networks

In the general case, the formation of an ensemble of solvers, in particular artificial neural networks, involves solving two problems. First, this is the formation of a set of solvers whose solutions will be taken into account in the formation of a general solution. Second, to determine how to compute the total solution for the ensemble. In regard to the formation of a set of solvers, approaches involving the direct formation of individual solvers with their inclusion in the ensemble without additional selection procedures can be used. Various boosting mechanisms are implemented in this case, for example, to ensure the realization of the ensemble efficiency increment requirements by maintaining the diversity of individual solvers based on different ways of varying the datasets used to construct the individual solvers [10,24,25]. Another option is to form a preliminary pool of individual solvers, artificial neural networks in the study under consideration, with the further selection of solvers for inclusion in the ensemble [26–28].

One of the problems of the effective application of ensemble methods is the choice of a way to form a general solution. In the basic version there are two possible approaches: The first assumes, for the regression problem, the use of some static formula with respect to the input features. The most commonly used methods are averaging or weighted averaging, where weighting coefficients can be calculated by taking into account the accuracy estimates of individual regressors. An alternative option implies forming cascade schemes that involve the sequential connection of solvers from a set according to certain rules, for example, from simpler to more complex ones when crossing the boundary values of the solvers' confidence in correctness.

A more complex approach, as compared to the basic ones, is the formation of an adaptive skipping procedure, which forms a rule for calculating the overall solution in accordance with the implicit partitioning of the model construction space [29]. The evolutionary method of genetic programming [30] is proposed to be used as such a procedure within the holistic evolutionary approach of creating high-performance methods for the complex design of ensembles of neural networks. The genetic programming method allows to form in an explicit form, as a formula, a way to calculate a general solution on the basis of individual solutions of neural networks. The outputs of individual neural networks are used as the inputs of such a procedure, which is a formula for calculating the solution of an ensemble of models in a symbolic expression. In this case, the procedure implies the possibility of using a terminal set, which is incomplete in relation to the preformed pool of individual artificial neural networks. In addition, for binding to the features of the factor space of regression model building, the attributes of the initial task of regression modeling can also be used as inputs. Taking into account the automation of the construction of such a rule for the calculation of a general solution in an ensemble on the basis of the evolutionary optimization procedure, various hybridized schemes of calculation can be constructed, combining in a terminal set of such procedures the outputs of individual neural networks as well as the values of some of their inputs. Within the framework of this study, just such a complete scheme has been implemented. A comparison with the original version, using only the outputs of individual solvers, is a subject for a separate study.

The scheme of using ensembles of neural networks assumes that the calculation of an overall ensemble solution is based on the solutions obtained by individual neural networks; that is, the overall ensemble solution is some function that depends on the solutions of individual neural networks, Equation (5):

$$o = f(o_1, o_2, \dots, o_n) \tag{5}$$

where *o* - is the general solution, *o_i* is the individual solution of the *i*-th network, and *n* is the number of networks in the ensemble. One of the effective methods for the reconstruction of

functional dependencies in explicit (symbolic) form is the method of genetic programming. This is the basis for the adaptation and use of the hybrid genetic programming method proposed in this paper for choosing the method of forming a general solution in ensembles of neural networks. Since the method of genetic programming refers to evolutionary methods, we can talk about the development of an evolutionary method for forming a common solution in ensembles of neural networks:

- 1. In order for the method of genetic programming, originally developed for solving symbolic regression problems, to be used to form a way to calculate the total solution in ensembles of neural networks, it is necessary to adapt it for solving this problem. For this purpose, the following changes and additions were made to the method:
- 2. The set $T = \{o_1, o_2, \dots, o_n, C\}$, where o_i is the individual solution of the *i*-th network, n is the number of networks in the ensemble, and C is the set of constants (numerical coefficients of the model describing the way of common solution formation), is used as a terminal set of the genetic programming method used to form a common solution. Thus, individual solutions of networks are used as input variables, not input variables of the problem;
- 3. The hybrid genetic programming method includes evolutionary algorithms to tune numerical parameters (genetic, hybrid, and probabilistic genetic algorithms);
- 4. The method provides a mechanism for limiting the number of input variables used; that is, the number of networks whose solutions are used to form the overall solution. There are also no obstacles that limit the total complexity of the ensemble (for example, the total number of computational nodes in all of the networks in the ensemble). Such constraints can be introduced through the use of appropriate estimates in the calculation of fitness during the evolutionary process. If such mechanisms are not activated, the number of input parameters is arbitrarily determined by the genetic programming procedure itself.

3. Numerical Studies

To evaluate the effectiveness of the proposed approach, numerical studies were performed on the generated sets of test problems and datasets of real problems placed in the Machine Learning Repository [30,31]. A description of the generated and test sets is given below. As a unique task, in this study a dataset obtained from metallurgical production at the stage of final products in plums was used to assess the quality of regression modeling. The name of the company that provided the data and the characteristics of the dataset are not disclosed to ensure the nondisclosure of commercial information.

3.1. Description of Test Tasks

For the primary evaluation of the performance of the methods under consideration, we used datasets generated on the basis of the functions given in Table 1. The table also indicates the ranges of variation in the variables and the volume of the generated datasets.

| Task No. | Modeled Function | Input Variable Range | Sample Volume |
|----------|--|---|---------------|
| 1 | 3-d Mexican-Hat: $y = \frac{\sin \sqrt{x_1^2 + x_2^2}}{\sqrt{x_1^2 + x_2^2}}$ | $x_{1,2}\in [-2\pi,2\pi]$ | 1000 |
| 2 | Friedman 1: $y = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5$ | $x_{15} \in [0, 1]$ | 1000 |
| 3 | Friedman 2: $y = \sqrt{x_1^2 + (x_2x_3 - \frac{1}{x_2x_4})^2}$ | $x_{1} \in [0, 100] x_{2} \in [40\pi, 560\pi] x_{3} \in [0, 1] x_{3} \in [1, 11]$ | 1000 |
| 4 | Gabor: $y = \frac{\pi}{2} exp \left[-2(x_1^2 + x_2^2) \right] cos[2\pi(x_1 + x_2)]$ | $x_4 \in [1, 11]$ $x_i \in [0, 1]$ | 1000 |

Table 1. Functions and information for generating test data.

The choice of these functions is due to the experience of their use in similar studies and the obtained acceptable level of consistency of the results of such datasets and real datasets.

3.2. Real Datasets Used for Approach Investigation

To test the methods of solving regression problems using ensembles of artificial neural networks considered and proposed in the study, we also used publicly available datasets from the Machine Learning Repository's Concrete Slump Data Set. This dataset includes 103 records that describe the relationships between compositions of concrete and measured indices that characterize the ultimate strengths of test specimens made from concrete with appropriate ingredient compositions.

We also used private datasets that describe the parameters of the implementation of metallurgical production processes. As one of such datasets, a set describing the static dependence of the implementation of the process of ore-thermal smelting at different input parameters was used. The initial data of the problem are as follows: there are samples of data characterizing the efficiency of ore-thermal smelting furnaces. As control parameters (input influences) are used—electrical parameters and charge loading on the individual components-these input parameters have a significant impact on the processes in the furnace and also provide an opportunity to obtain continuous and reliable information about them. The main control parameters for the furnace of ore-thermal melting include the following: the amount of sinter loaded into the furnace; the amount of silica loaded into the furnace; the amount of coke loaded into the furnace; the amount of converter slag loaded into the furnace; electrical power input; electrode burial; voltage; current intensity; and specific power consumption. These parameters allow for the estimation of technological, energy, and economic aspects of the smelting process, characterizing the efficiency of furnace operation. The nickel content in the spent slag in percent was selected as output parameter y.

Additionally, for an extended evaluation of the effectiveness of the approaches under consideration, a set of metallurgical production data, characterizing the content of target metals in the drains of metallurgical production, subjectively evaluated as complex for regression modeling, was used. The set is characterized by 20 input and 4 output parameters, considered in lag space with a delay in measurement periods of minus one and minus two periods.

The results of the study of the examined methods for this dataset and the previously considered datasets are given in the results section of the numerical study.

3.3. Alternative Approaches to Regression Model Construction and Artificial Neural Network Ensemble Formation

To investigate the effectiveness of the proposed approach a pool of methods was formed, which also obtained results on the considered problems. The following methods were included in the research as basic nonensemble methods: individual artificial neural networks, a regression method of reference vectors, and a method of multivariate adaptive splines [32,33].

As alternative methods of ensembling we considered GASEN methods, methods based on the use of a genetic algorithm to select networks from the pool and a boosting gradient method [34–36]. In order to perform numerical investigations, the considered methods were implemented in the IT-Pegas software system. The verification of the implementation correctness and preliminary tuning of the parameters for the considered approaches have been performed using this software system during the preliminary study on the basic set of test problems (Table 1).

3.4. Numerical Study and Alignment Conditions of Functioning

A five-fold cross-validation scheme with five-fold sampling re-partitioning (the socalled five-on-five scheme) was used to generate a statistically robust estimate of the results. Thus, for each method, sets of 25 estimates of the accuracy of the regression model were obtained and then averaged. The scatter of the quality of the obtained solutions was also considered by calculating the estimates of the standard deviation.

Due to the different structures of the models and methods considered during the study, in order to provide standard conditions for the construction of regression models during

the preliminary study with the help of software processor time counters, the alignment settings of the computational complexity of the approaches considered were determined.

3.5. Results and Discussion

The results of the experiments are shown in Table 2. Test tasks are presented in Table 1. To assess the statistical significance of the results, we used ANOVA methods. Tables 2–4 present the results of calculating the R2 index for the models formed on the basis of the methods described above.

Concrete **Ore-Thermal** The Method Test Task 1 Test Task 2 Test Task 3 Test Task 4 Slump Data Melting Data Stochastic Gradient 0.815 0.976 0.934 0.898 0.984 0.785 Boosting Multidimensional regression 0.785 0.981 0.931 0.826 0.985 0.813 splines Single neural 0.928 0.968 0.965 0.947 0.987 0.806 networks 0.996 0.997 GASEN 0.987 0.995 0.983 0.934 GA-based 0.981 0.997 0.995 0.981 0.992 0.941 Proposed 0.986 0.996 0.996 0.984 0.995 0.985 Approach

Table 2. Data on the results of the method study on test problems with additive noise 0%.

Table 3. Data on the results of the study of methods on test problems with additive noise 5%.

| The Method | Test Task 1 | Test Task 2 | Test Task 3 | Test Task 4 |
|------------------|-------------|-------------|-------------|-------------|
| Stochastic | | | | |
| Gradient | 0.701 | 0.911 | 0.867 | 0.746 |
| Boosting | | | | |
| Multidimensional | | | | |
| regression | 0.678 | 0.924 | 0.854 | 0.682 |
| splines | | | | |
| Single neural | 0.764 | 0.017 | 0.883 | 0.785 |
| networks | 0.764 | 0.917 | 0.005 | 0.785 |
| GASEN | 0.931 | 0.969 | 0.935 | 0.923 |
| GA-based | 0.930 | 0.958 | 0.942 | 0.937 |
| Proposed | 0.052 | 0.087 | 0.045 | 0.046 |
| Approach | 0.932 | 0.907 | 0.903 | 0.940 |

Table 4. Data on the results of the study of methods on test problems with additive noise 20%.

| The Method | Test Task 1 | Test Task 2 | Test Task 3 | Test Task 4 |
|------------------|-------------|-------------|-------------|-------------|
| Stochastic | | | | |
| Gradient | 0.637 | 0.860 | 0.821 | 0.639 |
| Boosting | | | | |
| Multidimensional | | | | |
| regression | 0.628 | 0.851 | 0.752 | 0.641 |
| splines | | | | |
| Single neural | 0.681 | 0.845 | 0.819 | 0.695 |
| networks | | | | |
| GASEN | 0.865 | 0.889 | 0.891 | 0.882 |
| GA-based | 0.850 | 0.905 | 0.885 | 0.920 |
| Proposed | 0.927 | 0.917 | 0.933 | 0.936 |
| Approach | | | | |

The results of a series of numerical experiments allow us to formulate the following conclusions. When using a baseline dataset without noise, the reliability on relatively simple test functions and sets—test problems 2, 3 and Concrete Slump Data set and standard methods that are not based on the procedure for building ensembles of neural networks—is quite high. Statistically, the results of the considered methods of building ensembles of neural networks and inidual models are indistinguishable, with a small exception for Problem 3. As the complexity of the problem increases and test problems 1, 4 and Ore-Thermal Melting data set can be referred to more complex ones, the potential of ensemble approaches is realized, among which the proposed approach with maintaining the diversity of networks in the ensemble is statistically significantly more effective. According to our observations, this is due to the efficiency ceiling of the structure of non ensemble models, which is overcome by forming an ensemble, and this effect is stronger when the configuration of the described set of points becomes more complex.

In order to assess the stability of the observed effects under the condition of noisy samples, we carried out investigations with additive random noise superimposed on the test functions, for the simulation of which a generator of uniformly distributed random numbers was used. The results of the study on the test data sets at a noise level of 5% are shown in Table 3.

The slump in the estimated efficiency parameter is most obvious for nonensemble methods; the drop in the quality of the regression model for a number of method-test-task intersections was even higher than the level of the superimposed noise. The results of the considered ensemble methods proved to be more stable, although a sag of the estimated index is also observed. There is also a statistically significant difference in the estimated efficiency of the proposed approach, with diversity maintenance in the ensemble and standard evolutionary methods of neural network ensemble formation.

Increasing the level of noisiness of the initial data up to 20% further widens the gap in estimating the quality of the regression model between the ensemble approaches under consideration and alternative, "traditional" approaches to regression modeling (Table 4).

It is the proposed approach, with modifications providing a variety of individual regressors in the ensemble, that allows for statistically significant results. A stable statistically significant effect is observed on the entire set of test functions used, which is especially noteworthy provided that the computational resources spent on model building are, on average, equal. Thus, in the considered series of experiments, taking into account the execution of a series of numerical studies according to the conventional cross-validation schemes, an increase in the efficiency of regression models in the conditions of noisy samples and the complex problem of modeling the ore-thermal smelting process is provided when checking the statistical significance of the results.

4. Conclusions

This paper describes a probabilistic evolutionary method developed by the authors for the automatic design of neural network structures, which differs from known evolutionary methods by a smaller number of adjustable parameters due to the use of original procedures for generating new solutions and a reduced number of coding–decoding operations of neural network structures. A new method for the formation of ensembles of neural networks is presented in which the methods of the automatic selection of a method for generating a general solution and the selection of neural networks developed by the authors are implemented. Comparative research on a developed probabilistic method for forming structures of neural networks and a widespread method based on the use of a genetic algorithm for tuning the structures of neural networks has been carried out. It is shown that the proposed method is not less effective than the competing evolutionary approach and has a smaller number of adjustable parameters, which facilitates its adaptation for solving a particular problem. In addition, the developed method makes it possible to slightly reduce the use of computational resources by eliminating the need for coding–decoding solutions into binary strings. A comparative analysis of the developed method and other common methods of ANN teams' design was carried out. It was shown that, on all of the test problems, the efficiency of the proposed method is not lower than the efficiency of other methods. On average, on most test problems, the proposed method outperforms other methods in terms of efficiency, evaluated by the average R2 value. The statistical significance of the difference in the results for test functions with noise is confirmed by processing samples of the results of numerical studies using the ANOVA method. The presented approaches have been implemented and tested in software systems. The results obtained demonstrate an improvement in the quality of ensemble approaches in comparison with a number of methods widely used and implemented in applied packages for statistical data analyses: gradient boosting variants, methods for constructing ensembles based on evolutionary algorithms, and the GASEN method. The proposed approach is focused on ensuring the diversity of individual solvers in an ensemble at the stage of their structural design. Thus, this method can be combined with other methods to ensure the efficient use of ensemble circuits. Thus, a combination with gradient boosting schemes and methods for selecting solvers and from a preliminary pool is possible, which, in our opinion, will give a synergistic effect of increasing the efficiency of solving regression problems. However, this will require an additional amount of numerical research on an extended set of problems, which is our priority for further research.

Within the framework of this study, we focused on the problem of maintaining diversity in ensembles; issues related to the justification of the ensemble approach were considered in our other works. Additionally, for comparison, we included several nonensemble approaches in the research scheme. The results of the numerical study show that the ensemble approaches are more efficient. In the future, we plan to evaluate the effectiveness of the proposed approaches to the formation of collective regressors in comparison with approaches such as transfer learning and federated learning for the same kinds of problems.

Author Contributions: Conceptualization, V.V.B., V.S.T. and I.S.M.; methodology, V.V.B., V.S.T., and A.P.G.; validation, V.V.B., I.S.M., A.S.B. and A.P.G.; formal analysis, V.V.B., I.S.M., A.S.B. and A.P.G.; investigation, V.V.B., V.S.T. and V.A.N.; resources, V.V.B., V.A.N. and I.S.M.; data curation, V.S.T., V.A.N. and I.S.M.; writing—original draft preparation, V.V.B., V.S.T., V.A.N., I.S.M., A.S.B. and A.P.G.; writing—review and editing, V.V.B., V.S.T., V.A.N., I.S.M., A.S.B. and A.P.G.; using—review and editing, V.V.B., V.S.T., V.A.N., I.S.M., A.S.B. and A.P.G.; supervision, V.S.T., V.A.N., I.S.M., A.S.B. and A.P.G.; using—review and editing, V.V.B., V.S.T., V.A.N., I.S.M., A.S.B. and A.P.G.; using supervision, V.S.T., V.A.N. and I.S.M.; project administration, V.S.T., V.A.N. and A.S.B.; funding acquisition, V.A.N. and A.S.B. All authors have read and agreed to the published version of the manuscript.

Funding: The studies were carried out within the program of the Russian Federation of strategic academic leadership "Priority-2030", aimed at supporting the development programs of educational institutions of higher education, and the scientific project PRIOR/SN/NU/22/SP5/16 "Building intelligent networks, determining their structure and architecture, operation parameters in order to increase productivity systems and bandwidth of data transmission channels using trusted artificial intelligence technologies that provide self-learning, self-adaptation and optimal reconfiguration of intelligent systems for processing large heterogeneous data".

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

- Smith, J.L. Advances in Neural Networks and Potential for Their Application to Steel Metallurgy. *Mater. Sci. Technol.* 2020, 36, 1805–1819. [CrossRef]
- Guo, L.; Lei, Y.; Li, N.; Yan, T.; Li, N. Machinery Health Indicator Construction Based on Convolutional Neural Networks Considering Trend Burr. *Neurocomputing* 2018, 292, 142–150. [CrossRef]
- Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional Neural Networks: An Overview and Application in Radiology. Insights Imaging 2018, 9, 611–629. [CrossRef] [PubMed]

- 4. AL-Qutami, T.A.; Ibrahim, R.; Ismail, I.; Ishak, M.A. Virtual Multiphase Flow Metering Using Diverse Neural Network Ensemble and Adaptive Simulated Annealing. *Expert Syst. Appl.* **2018**, *93*, 72–85. [CrossRef]
- Ribeiro, G.T.; Mariani, V.C.; Coelho, L.d.S. Enhanced Ensemble Structures Using Wavelet Neural Networks Applied to Short-Term Load Forecasting. *Eng. Appl. Artif. Intell.* 2019, 82, 272–281. [CrossRef]
- 6. Melin, P.; Monica, J.C.; Sanchez, D.; Castillo, O. Multiple Ensemble Neural Network Models with Fuzzy Response Aggregation for Predicting Covid-19 Time Series: The Case of Mexico. *Healthcare* **2020**, *8*, 181. [CrossRef]
- 7. Hansen, L.K.; Salamon, P. Neural Network Ensembles. IEEE Trans. Pattern Anal. Mach. Intell. 1990, 12, 993–1001. [CrossRef]
- 8. Irvine, N.; Nugent, C.; Zhang, S.; Wang, H.; Ng, W.W.Y. Neural Network Ensembles for Sensor-Based Human Activity Recognition within Smart Environments. *Sensors* 2020, 20, 216. [CrossRef]
- 9. Li, S.; Yao, Y.; Hu, J.; Liu, G.; Yao, X.; Hu, J. An Ensemble Stacked Convolutional Neural Network Model for Environmental Event Sound Recognition. *Appl. Sci.* 2018, *8*, 1152. [CrossRef]
- 10. ALzubi, J.A.; Bharathikannan, B.; Tanwar, S.; Manikandan, R.; Khanna, A.; Thaventhiran, C. Boosted Neural Network Ensemble Classification for Lung Cancer Disease Diagnosis. *Appl. Soft Comput. J.* **2019**, *80*, 579–591. [CrossRef]
- Khan, R.U.; Almakdi, S.; Alshehri, M.; Kumar, R.; Ali, I.; Hussain, S.M.; Haq, A.U.; Khan, I.; Ullah, A.; Uddin, M.I. Probabilistic Approach to COVID-19 Data Analysis and Forecasting Future Outbreaks Using a Multi-Layer Perceptron Neural Network. *Diagnostics* 2022, 12, 2539. [CrossRef] [PubMed]
- 12. Jia, D.W.; Wu, Z.Y. Seismic Fragility Analysis of RC Frame-Shear Wall Structure under Multidimensional Performance Limit State Based on Ensemble Neural Network. *Eng. Struct.* 2021, 246, 112975. [CrossRef]
- Masich, I.S.; Tyncheko, V.S.; Nelyub, V.A.; Bukhtoyarov, V.V.; Kurashkin, S.O.; Borodulin, A.S. Paired Patterns in Logical Analysis of Data for Decision Support in Recognition. *Computation* 2022, 10, 185. [CrossRef]
- 14. Mikhalev, A.S.; Tynchenko, V.S.; Nelyub, V.A.; Lugovaya, N.M.; Baranov, V.A.; Kukartsev, V.V.; Sergienko, R.B.; Kurashkin, S.O. The Orb-Weaving Spider Algorithm for Training of Recurrent Neural Networks. *Symmetry* **2022**, *14*, 2036. [CrossRef]
- Li, H.; Wang, X.; Ding, S. Research and Development of Neural Network Ensembles: A Survey. Artif. Intell. Rev. 2018, 49, 455–479. [CrossRef]
- 16. Shu, C.; Burn, D.H. Artificial Neural Network Ensembles and Their Application in Pooled Flood Frequency Analysis. *Water Resour. Res.* **2004**, *40*, 1–10. [CrossRef]
- Giacinto, G.; Roli, F. Design of Effective Neural Network Ensembles for Image Classification Purposes. *Image Vis. Comput.* 2001, 19, 699–707. [CrossRef]
- Mendes-Moreira, J.; Soares, C.; Jorge, A.M.; Sousa, J.F.D. Ensemble approaches for regression: A survey. ACM Comp. Surv. 2012, 45, 1–40. [CrossRef]
- 19. Shahhosseini, M.; Hu, G.; Pham, H. Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. *Mach. Learn. Appl.* **2022**, *7*, 100251. [CrossRef]
- 20. Kotsiantis, S.; Kanellopoulos, D. Combining bagging, boosting and random subspace ensembles for regression problems. *Int. J. Innov. Comput. Inf. Control* **2012**, *8*, 3953–3961.
- Nguyen, P.T.; Ha, D.H.; Avand, M.; Jaafari, A.; Nguyen, H.D.; Al-Ansari, N.; Van Phong, T.; Sharma, R.; Kumar, R.; Le, H.V.; et al. Soft Computing Ensemble Models Based on Logistic Regression for Groundwater Potential Mapping. *Appl. Sci.* 2020, 10, 2469. [CrossRef]
- 22. Phyo, P.-P.; Byun, Y.-C.; Park, N. Short-Term Energy Forecasting Using Machine-Learning-Based Ensemble Voting Regression. Symmetry 2022, 14, 160. [CrossRef]
- 23. Jozdani, S.E.; Johnson, B.A.; Chen, D. Comparing Deep Neural Networks, Ensemble Classifiers, and Support Vector Machine Algorithms for Object-Based Urban Land Use/Land Cover Classification. *Remote Sens.* **2019**, *11*, 1713. [CrossRef]
- 24. Khwaja, A.S.; Anpalagan, A.; Naeem, M.; Venkatesh, B. Joint Bagged-Boosted Artificial Neural Networks: Using Ensemble Machine Learning to Improve Short-Term Electricity Load Forecasting. *Electr. Power Syst. Res.* **2020**, *179*, 106080. [CrossRef]
- Liu, L.; Wei, W.; Chow, K.H.; Loper, M.; Gursoy, E.; Truex, S.; Wu, Y. Deep Neural Network Ensembles against Deception: Ensemble Diversity, Accuracy and Robustness. In Proceedings of the 16th International Conference on Mobile Ad Hoc and Smart Systems, MASS 2019, Monterey, CA, USA, 4–7 November 2019; Institute of Electrical and Electronics Engineers Inc.: Monterey, CA, USA, 2019; pp. 274–282.
- 26. Ai, S.; Chakravorty, A.; Rong, C. Household Power Demand Prediction Using Evolutionary Ensemble Neural Network Pool with Multiple Network Structures. *Sensors* 2019, *19*, 721. [CrossRef]
- 27. Huang, C.; Li, M.; Wang, D. Stochastic Configuration Network Ensembles with Selective Base Models. *Neural Netw.* 2021, 137, 106–118. [CrossRef]
- 28. Van Roode, S.; Ruiz-Aguilar, J.J.; González-Enrique, J.; Turias, I.J. An Artificial Neural Network Ensemble Approach to Generate Air Pollution Maps. *Environ. Monit. Assess.* **2019**, *191*, 727. [CrossRef]
- Ahvanooey, M.T.; Li, Q.; Wu, M.; Wang, S. A Survey of Genetic Programming and Its Applications. *KSII Trans. Internet Inf. Syst.* 2019, 13, 1765–1794. [CrossRef]
- 30. Chandrasekaran, K.; Karp, R. Finding a Most Biased Coin with Fewest Flips. J. Mach. Learn. Res. 2014, 35, 394–407.
- 31. Yen, I.C. Modeling Slump of Concrete with Fly Ash and Superplasticizer. Comput. Concr. 2008, 5, 559–572. [CrossRef]

- Gackowski, M.; Szewczyk-Golec, K.; Pluskota, R.; Koba, M.; Madra-Gackowska, K.; Woźniak, A. Application of Multivariate Adaptive Regression Splines (MARSplines) for Predicting Antitumor Activity of Anthrapyrazole Derivatives. *Int. J. Mol. Sci.* 2022, 23, 5132. [CrossRef] [PubMed]
- 33. Qin, W.; Wang, L.; Liu, Y.; Xu, C. Energy Consumption Estimation of the Electric Bus Based on Grey Wolf Optimization Algorithm and Support Vector Machine Regression. *Sustainability* **2021**, *13*, 4689. [CrossRef]
- 34. Friedman, J.H. Stochastic Gradient Boosting. Comput. Stat. Data Anal. 2002, 38, 367–378. [CrossRef]
- Lee, M.C.; Boroczky, L.; Sungur-Stasik, K.; Cann, A.D.; Borczuk, A.C.; Kawut, S.M.; Powell, C.A. Computer-Aided Diagnosis of Pulmonary Nodules Using a Two-Step Approach for Feature Selection and Classifier Ensemble Construction. *Artif. Intell. Med.* 2010, 50, 43–53. [CrossRef]
- Yao, C.; Dai, Q.; Song, G. Several Novel Dynamic Ensemble Selection Algorithms for Time Series Prediction. *Neural Process. Lett.* 2019, 50, 1789–1829. [CrossRef]

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.