Article

Net-Net AutoML Selection of Artificial Neural Network Topology for Brain Connectome Prediction

Enrique Barreiro [1,2,3], **Cristian R. Munteanu** [1,4,5], **Marcos Gestal** [1,4,5,*],
Juan Ramón Rabuñal [1], **Alejandro Pazos** [1,4,5], **Humberto González-Díaz** [6,7] **and**
Julián Dorado [1,4,5]

1   RNASA-IMEDIR Group, Computer Science Faculty, University of A Coruña, Elviña, 150171 A Coruña,
    Spain; enrique.barreirov@udc.es (E.B.); c.munteanu@udc.es (C.R.M.); juanra@udc.es (J.R.R.);
    apazos@udc.es (A.P.); julian@udc.es (J.D.)
2   Center for Computational Science (CCS), University of Miami, Miami, FL 33136, USA
3   Computer Engineering, West Coast University, Miami Campus, Doral, FL 33178, USA
4   Centre for Information and Communications Technology Research (CITIC), Campus de Elviña s/n,
    15071 A Coruña, Spain
5   Biomedical Research Institute of A Coruña (INIBIC), University Hospital Complex of A Coruña (CHUAC),
    15006 A Coruña, Spain
6   Department of Organic Chemistry II, University of the Basque Country UPV/EHU, 48940 Leioa, Spain;
    humberto.gonzalezdiaz@ehu.eus
7   IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain
*   Correspondence: mgestal@udc.es; Tel.: +34-981167000 (ext. 1379)

check for
updates

Abstract: Brain Connectome Networks (BCNs) are defined by brain cortex regions (nodes) interacting with others by electrophysiological co-activation (edges). The experimental prediction of new interactions in BCNs represents a difficult task due to the large number of edges and the complex connectivity patterns. Fortunately, we can use another special type of networks to achieve this goal—Artificial Neural Networks (ANNs). Thus, ANNs could use node descriptors such as Shannon Entropies (Sh) to predict node connectivity for large datasets including complex systems such as BCN. However, the training of a high number of ANNs for BCNs is a time-consuming task. In this work, we propose the use of a method to automatically determine which ANN topology is more efficient for the BCN prediction. Since a network (ANN) is used to predict the connectivity in another network (BCN), this method was entitled Net-Net AutoML. The algorithm uses Sh descriptors for pairs of nodes in BCNs and for ANN predictors of BCNs. Therefore, it is able to predict the efficiency of new ANN topologies to predict BCNs. The current study used a set of 500,470 examples from 10 different ANNs to predict node connectivity in BCNs and 20 features. After testing five Machine Learning classifiers, the best classification model to predict the ability of an ANN to evaluate node interactions in BCNs was provided by Random Forest (mean test AUROC of 0.9991 ± 0.0001, 10-fold cross-validation). Net-Net AutoML algorithms based on entropy descriptors may become a useful tool in the design of automatic expert systems to select ANN topologies for complex biological systems. The scripts and dataset for this project are available in an open GitHub repository.

Keywords: artificial neural networks; brain connectome networks; machine learning; Net-Net AutoML

1. Introduction

Any system may be represented as a complex network of nodes connected by edges, which can be any property or node interaction (L_{ij} = connection between nodes i and j) [1–7]. The diversity of

systems susceptible to be studied with complex networks is very high. An important example is the representation of the human brain. Brain Connectome Networks (BCNs) are defined by anatomical connections and/or functional co-activations (L_{ij}) between brain regions (large collections of neurons). BCNs are so large and it is impossible for a person to remember and rationalize all possible connections and assign/predict correct connections in different situations [8]. Machine Learning (ML) algorithms may be useful to create classifiers for the prediction of properties of complex biological systems [9–12].

The first step in this process requires the quantification of relevant structural information of the biological system. Shannon's entropy, introduced by Claude E. Shannon in his 1948 seminar paper "A Mathematical Theory of Communication" [13] could be used to quantify the information related to network nodes or the entire network. In fact, Shannon entropy information indices have become one of the universal indices used to quantify information [14–30]. Related indices, called Shannon–Markov information-theoretic entropy measures (Sh_k) of order (k) have been used to codify structural information of complex biological systems in ML studies [31]. We can calculate 1D, 2D, or 3D Sh_k descriptors for chemical structure of drugs [32], DNA/protein sequences [33–35], RNA secondary structures [36], protein spatial structures [37], protein–protein complexes [38], and complex networks [31]. Riera-Fernandez et al. [31] published ML prediction models of a BCN using information measures.

Artificial Neural Networks (ANNs) are powerful bio-inspired predictors able to learn/infer large datasets. There are many examples of ANN applications used to build predictive models [39–42]. ANNs are also able to learn large datasets related to the connectivity (structure) of bio-systems and other complex networks [43]. In the second step, an ML classifier should be constructed to evaluate the ability of any ANN to predict BCN node connectivity. Thus, the best ANN topology could be selected for the optimal BCN prediction. For nonexperts in ML, this is a complex task. In this context, the Automated Machine Learning (AutoML) method can help automatically select ANN topologies for complex network connectivity prediction, in the development of practical ML applications by non-experts [44,45]. In order to obtain the current Net-Net AutoML classifier, we need to quantify the information about the BCN nodes and ANN topologies.

In our previous work, we have introduced the idea of the Net-Net AutoML methodology for Biological Ecosystem Networks (BENs) [46]. BENs are webs of biological species (nodes) that can establish trophic relationships (interactions/edges). Considering that the experimental confirmation of all possible interactions is difficult and generates a huge volume of information, computational prediction becomes an important goal. We used ANNs to predict BENs with inputs as node Sh_k descriptors from known ecosystems. The complex task of a priori selecting the best ANN topology for the BEN node connectivity prediction, Net-Net AutoML method was able to help us. Twelve types of classifiers have been tested as Net-Net models for BENs. The best model used 338,050 examples from 10 ANN topologies for node pairs in 69 BENs and it was obtained with a deep fully connected neural network (test AUROC (Area Under Receiver Operating Characteristic) of 0.935).

In the current work, we apply the same Net-Net AutoML methodology to predict the ability of ANNs to predict BCNs. The dataset is based on predictions of 10 different ANN topologies for 52,690 pairs of brain region co-activations in the BCN (reported in the CoCoMac experiment) [47]. Therefore, we propose Net-Net AutoML models for the selection of ANNs for the study of BCNs, with the subsequent reduction of time and computational resources for the design of new expert systems.

2. Materials and Methods

2.1. Brain Connectome Dataset

This dataset is represented by a version of the CoCoMac network with 383 hierarchically organized regions covering cortex, thalamus, and basal ganglia. It contains 6602 directed long-distance connections and encloses different subnetworks such as cortico-cortical, cortico-subcortical, and subcortico-subcortical fiber systems [47].

2.2. ANN Datasets and General Workflow

Figure 1 shows the workflow for the Net-Net AutoML method applied to BCNs. Our goal was to find a tool that would be able to predict BCN node connectivity (communications between brain regions). This task can be solved by training an ANN classifier using as dataset the previous known node connections (output) with any type of node descriptors (inputs). This solution involves intensive training of ANNs. Net-Net AutoML proposed a shortcut: The creation of another classifier with any ML method that can evaluate whether an ANN is able to predict with accuracy the node connectivity in BCNs, without any ANN training. In order to obtain this classifier, one should previously train different ANNs to predict BCN connectivity. Thus, this Net-Net AutoML classifier is a relationship between the ANN topologies trained for BCN prediction and the ability of these ANNs to provide good predictions.
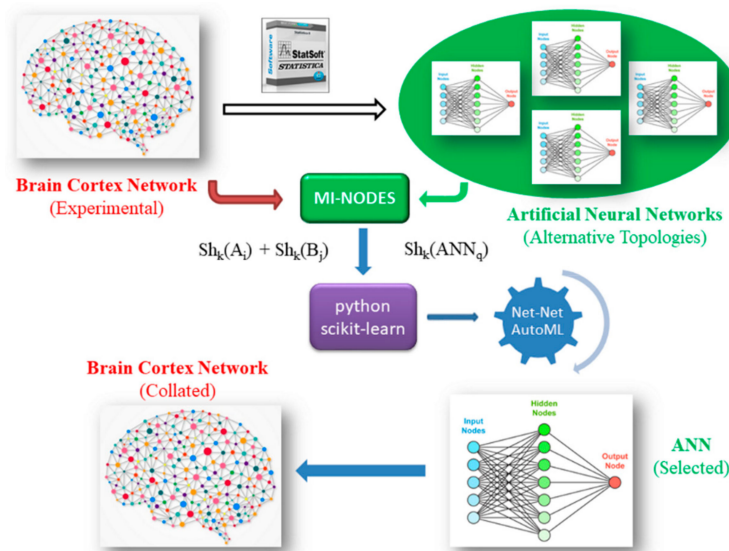

Figure 1. Proposed workflow.

In order to build ANNs for BCN prediction and Net-Net AutoML classifiers for the best ANN, both types of networks (BCNs and ANNs) should provide numerical descriptors. Therefore, for each BCN pair of nodes (brain regions), we knew the experimental connectivity (connected or not) and we could calculate the node descriptors such as Sh_k. This dataset was used to train different ANN topologies for the BCN prediction. In order to include information about ANNs to build a Net-Net AutoML classifier, we needed to include ANN information such as Sh_k of the entire ANN. Thus, we combined two types of descriptors in order to mix information about the BCN connectivity and ANN topology: We mixed Sh_k of BCN nodes (A_i–B_j pairs) with Sh_k for an ANN ($S_{hk}(ANN)$). We took into account the Shannon entropy from both types of networks but for different quantification: For nodes in BCN and for the entire network in ANN. The output of the Net-Net AutoML best classifier is the ability of $S_{hk}(ANN)$ to predict connectivity between BCN nodes A_i and B_j.

We calculated the values of entropy $Sh_k(A_i)$ and $Sh_k(B_i)$ for a large number of pair of nodes in the BCN of the CoCoMac experiment [47]. The values $Sh_k(A_i)$ are for the activated brain regions and the values $Sh_k(B_i)$ for the co-activated brain regions. With these values we created a large dataset with output values $L_{ij} = 1$ (BCN links) or $L_{ij} = 0$ (absence of region co-activation), and input values $Sh_k(A_i)$ and $Sh_k(B_i)$. The dataset was used to train 10 different ANNs with the STATISTICA software [48]. This work is a proof-of-concept method, so at this point we applied the default behavior of STATISTICA that used an automatic algorithm to calculate the best ANN by means of feature selection and topology modification. Future studies should include more ANNs. In the second step, we created the final dataset to find a Net-Net AutoML classifier: We used the outputs of 10 previously trained ANNs to predict BCN connections for 52,690 pairs of BCN nodes (500,740 examples after pre-processing).

Thus, the final dataset had the output as the ability of a specific ANN to predict the type of connection between BCN nodes, and the inputs as different entropies for BCN nodes and entire ANNs. In the next step, linear and nonlinear ML methods were tested to find the best Net-Net AutoML classifier to evaluate the ANN topology for BCN predictions.

Net-Net AutoML algorithm (Figure 1) has the following steps:

(1) For each BCN:
(1.1) Extract the connectivity matrix.
(1.2) Add weights for the BCN connections (if present).
(1.3) For each node A_i:
(1.3.1) Calculate Shannon entropies for nodes using MI-NODES: $Sh_k(A_i)$.
(1.3.2) Create pairs of entropies for all the other nodes B: $Sh_k(A_i)$–$Sh_k(B_j)$.
(2) Build ANNs to predict BCN connectivity for nodes A_j–B_j:
(2.1) For each ANN_q classifier:
(2.1.1) Calculate network Shannon entropy: $Sh_k(ANN)$.
(3) Mix the BCN node descriptors with the ANN descriptors in the Net-Net AutoML dataset: $Sh_k(A_i)$, $Sh_k(B_j)$, $Sh_k(ANN)$.
(4) Split the dataset into training and testing subsets (n-folds).
(5) Search for the best Net-Net AutoML classifier to evaluate whether a specific ANN can predict the BCN connectivity:
(5.1) For each ML method:
(5.1.1) Use a subset to train a classifier.
(5.1.2) Evaluate the model with a testing subset, calculating AUROC accuracy.
(6) Choose the best Net-Net AutoML classifier using AUROC metric.

2.3. Computational Methods

2.3.1. Markov–Shannon Entropy Centralities for Nodes

In the present work, we used the MI-NODES application [49]. In the first step, the connectivity matrix L for the BCN was obtained using the CoCoMac experiment [47]: $n \times n$ matrix with n vertices/nodes with values of 1 for the connected nodes and 0 for not connected ones. In the next step, a Markov matrix Π was derived from L by calculating the probabilities of vertices p_{ij} as matrix **P**. The power k ($k = 1$–5) of the probability matrix Π noted with $(^1\Pi)^k$ was multiplied with a vector containing the initial probabilities (^0p_j). The results are vectors with the absolute probabilities to reach nodes by walking throughout k nodes n_i (^kp_j). These vectors for each k value were used to calculate the entropy centrality (Sh_k) (see Equations (1) and (2)). For more mathematical details about combining the Markov chain theory and Shannon entropy see Reference [49] for MI-NODES desktop application and Reference [46] for the previous Net-Net AutoML application on Biological Ecosystem Networks. G represents any graph/network with j nodes. Thus, the input descriptors for BCN nodes or entire networks as ANNs are modified Shannon entropies that include the data about the path throughout k nodes based on probabilities.

$$^k\mathbf{P} = {}^0\mathbf{P} \times \left({}^1\mathbf{\Pi}\right)^k = \left[{}^k\mathbf{p}_1, \ {}^k\mathbf{p}_2, \ldots, \ {}^k\mathbf{p}_j\right] \tag{1}$$

$$_k(\mathbf{G}) = \sum_{j \in \mathbf{G}} \mathbf{Sh_k}(\mathbf{j}) = -\sum_{j \in \mathbf{G}} {}^k\mathbf{p_j} \log {}^k\mathbf{p_j} \tag{2}$$

2.3.2. Net-Net AutoML Models

Once the values of the Markov–Shannon entropies were obtained for the BCN nodes and ANNs, twelve Machine Learning classifiers from scikit-learn (python) were tested to find the best Net-Net

AutoML classifier able to determine the optimal ANN topology to predict BNCs (co-activations between brain regions):

- KNeighborsClassifier = KNN—k-nearest neighbors: A nonparametric classifier that assigns an unclassified sample to the same class as the nearest of k samples in the training set [50].
- LinearDiscriminantAnalysis = LDA—linear discriminant analysis [51]: A statistical supervised method that projects the input data to a lower dimension in order to maximize the scatter between classes versus the scatter within each class.
- GaussianNB = GBN—Gaussian Naive Bayes, a simple "probabilistic classifier" [52].
- SVC(kernel = 'rbf') = SVM_RBF—support-vector machines with nonlinear radial basis functions [53].
- LogisticRegression = LogR—Logistic regression [54] is a linear model that estimates the probability of a binary response using different factors.
- MLPClassifier = MLP—multilayer perceptron (artificial neural network) using 20 neurons in a hidden layer [55].
- DecisionTreeClassifier = DT—Decision Tree (DT) represents a set of decision rules inferred from the features as a tree of rules (the paths from root to leaf represent classification rules) [56].
- RandomForestClassifier = RF—Random Forest [57] aggregates several decision trees (parallel trees). Each tree is generated using a bootstrap sample randomly drawn from the original dataset.
- XGBClassifier = XGB—an optimized distributed gradient boosting library based on serial trees [58].
- GradientBoostingClassifier = GB—gradient boosting library [59].
- AdaBoostClassifier = Ada—is a meta-estimator that starts the fitting with a classifier based on the original dataset and then adds additional copies of the original classifier to the adjusted weights for the incorrectly classified instances [60].
- BaggingClassifier = Bagging—similar with Ada but the additional classifiers are based on subsets of the original dataset [61].

LDA could represent the simplest model equation. Let $^qS(L_{ij})$ be the output variable of a Net-Net AutoML classifier used to score the ability of a given q ANN to correctly predict the link or brain region co-activation L_{ij} between two nodes/brain regions, A_i–B_j ($L_{ij} = 1$). Equation (3) describes the general formula for the LDA model. *k* means Sh_k codifies information for nodes placed at least at a topological distance *k* (from the reference node).

$$^qS\left(L_{ij}\right) = \sum_{k=1}^{5} a_{ki} \cdot Sh_k(A_i) + \sum_{k=1}^{5} b_{ki} \cdot Sh_k\left(B_j\right) + \sum_{k=1}^{5} c_{ki} \cdot Sh_k\left(A_i - B_j\right) + \sum_{k=1}^{5} d_{ki} \cdot Sh_k(\text{ANN}) \tag{3}$$

Therefore, 20 features were obtained for the final dataset: $Sh_1(A)$, $Sh_1(B)$, $Sh_1(A\text{-}B)$, $Sh_1(\text{ANN})$, $Sh_2(A)$, $Sh_2(B)$, $Sh_2(A\text{-}B)$, $Sh_2(\text{ANN})$, $Sh_3(A)$, $Sh_3(B)$, $Sh_3(A\text{-}B)$, $Sh_3(\text{ANN})$, $Sh_4(A)$, $Sh_4(B)$, $Sh_4(A\text{-}B)$, $Sh_4(\text{ANN})$, $Sh_5(A)$, $Sh_5(B)$, $Sh_5(A\text{-}B)$, and $Sh_5(\text{ANN})$. In order to avoid overfitting, a 10-fold cross-validation was performed. The performance of the models was measured using the Area Under the Receiver Operating Characteristics (AUROC) [62]. The scripts and dataset are available at an open repository in GitHub from one of the authors, Cristian R. Munteanu [63].

3. Results and Discussion

This study used the CoCoMac BCN dataset in order to build a computational tool that is able to select the best ANN topologies for BCN connectivity prediction. A recent study conducted by Van Essen et al. [64] has reviewed The Human Connectome Project, an amazing five-year enterprise devoted to characterize brain connectivity and function in healthy adult human beings. Lang [65] discussed the need for computational methods to disentangle the relationships between anatomical and functional connections. However, we have to deal with the estimation of data reliability and the presence of contradictory reports to develop new BCN models. Consequently, there is a need for

computational tools for data mining using integrated large sets of partially redundant or inconsistent data in brain maps [66]. More often, data on BCN have to be systematically re-evaluated (collated) [67].

ANNs can discriminate between the correct connectivity for nodes (n_j) in complex systems (L_{ij}) from the incorrect and/or randomly distributed links. In order to be able to select the best ANN topology for BCN node connectivity prediction, we propose the Net-Net AutoML method. This tool represents a classifier based on Shannon entropies (Sh_k) of BCN pairs of nodes and ANN topology. In the first step, the BCN was turned into numerical input parameters ($Sh_k(ANN)$ values) for all pairs of nodes to feed alternative ANN classifiers. Next, we trained different ANNs to predict the BCN connectivity. In the final step, we joined all the data (Sh_k values of BCNs and ANNs) of selected pre-trained cases to search the best AutoML classifier using twelve Machine Learning methods. Table 1 shows the values of $Sh_k(ANN)$ for 10 different topologies. As previously explained, $Sh_k(A_i)$ and $Sh_k(B_i)$ were used as inputs and $L_{ij} = 1$ (BCN links) or $L_{ij} = 0$ as outputs. At that point, we used Multilayer Perceptron (MLP) and Liner Neural Network (LNN) as topologies with the default values provided by STATISTICA after its automatic algorithm based on feature selection and topology modification to improve them. In order to obtain a better generalization, the future studies should use more ANN topologies and/or different configurations thereof. We may need to use a High Performance Computing (HPC) service if we want to test a high number of ANNs for many complex systems [68–73].

Table 1. Information indices $Sh_k(ANN)$ of the ANNs used as inputs to find the best Net-Net Automated Machine Learning (AutoML) model.

ANN No.	ANN Profile (inputs:hidden layers EPs:outputs)	$Sh_k(ANN)$					
		k = 0	k = 1	k = 2	k = 3	k = 4	k = 5
1	MLP15:15-14-1:1	0.05	0.054	0.054	0.054	0.054	0.054
2	MLP4:4-6-11-1:1	0.06	0.067	0.069	0.072	0.072	0.072
3	MLP5:5-8-1:1	0.078	0.088	0.097	0.097	0.097	0.097
4	MLP7:7-11-1:1	0.067	0.074	0.081	0.081	0.081	0.081
5	MLP9:9-12-1:1	0.061	0.067	0.071	0.071	0.071	0.071
6	MLP10:10-12-1:1	0.061	0.067	0.071	0.071	0.071	0.071
7	MLP4:4-8-11-1:1	0.057	0.059	0.06	0.061	0.061	0.061
8	MLP10:10-11-12-1:1	0.046	0.048	0.044	0.046	0.046	0.046
9	LNN14:14-1:1	0.056	0.146	0.146	0.146	0.146	0.146
10	LNN15:15-1:1	0.053	0.146	0.146	0.146	0.146	0.146

Thus, our final classifier is based on entropies for different nodes of the BCN—$Sh_k(A_i)$ and $Sh_k(B_j)$—and entropies of different ANNs topologies—$Sh_k(ANN)$. The best Net-Net AutoML model determined the scores $^qS(L_{ij})$ for a given ANN topology to predict the connectivity of BCN pair of nodes. This study presents a proof-of-concept model that fits very well, 500,740 outcomes obtained with 10 different ANNs.

Twelve ML classifiers were tested and the AUROC values were calculated using a 10-fold cross-validation (see Table 2). The best model was obtained with RF10 (RF based on 10 trees) and AUROC was 0.9983 ± 0.0001. Figure 2 shows the box-plot for the AUROC values of 12 ML methods (10-fold CV). The boxplot suggests that the AUROC values for all ML methods were stable within each fold. In addition, the difference between the RF and the other methods (box-plots are far from overlapping) proved to be statistically significant. All the Net-Net classifiers obtained good performance with an AUROC value greater than 0.90. An interesting result was provided by the simple KNN method with a mean AUROC value of 0.9958, still less than RF10 with a mean AUROC value of 0.9983. The linear methods (LDA and LogR) can provide models with a mean AUROC value > 0.92 and 0.95. The change to a tree-based method provided a better AUROC with values greater than 0.99, except for Ada. Bagging had a performance very similar to the RF classifier with a mean AUROC value of 0.9980.

Table 2. AUROC for Net-Net AutoML classification models.

Fold	KNN	LDA	GNB	SVM_RBF	LogR	MLP	DT	RF10 [1]	XGB	GB	Ada	Bagging
1	0.9956	0.9195	0.9164	0.9625	0.9510	0.9876	0.9926	0.9983	0.9923	0.9937	0.9652	0.9983
2	0.9955	0.9217	0.9198	0.9642	0.9529	0.9879	0.9924	0.9982	0.9923	0.9933	0.9662	0.9980
3	0.9957	0.9203	0.9193	0.9637	0.9524	0.9880	0.9920	0.9983	0.9919	0.9934	0.9659	0.9980
4	0.9962	0.9207	0.9192	0.9637	0.9525	0.9851	0.9927	0.9981	0.9926	0.9938	0.9660	0.9982
5	0.9957	0.9204	0.9204	0.9645	0.9538	0.9882	0.9924	0.9983	0.9926	0.9934	0.9662	0.9980
6	0.9960	0.9192	0.9190	0.9633	0.9518	0.9885	0.9925	0.9982	0.9925	0.9936	0.9656	0.9979
7	0.9958	0.9219	0.9205	0.9645	0.9535	0.9888	0.9925	0.9983	0.9926	0.9940	0.9669	0.9982
8	0.9955	0.9215	0.9196	0.9642	0.9529	0.9884	0.9927	0.9981	0.9925	0.9940	0.9664	0.9978
9	0.9959	0.9176	0.9183	0.9625	0.9514	0.9875	0.9922	0.9982	0.9923	0.9933	0.9646	0.9979
10	0.9964	0.9182	0.9168	0.9623	0.9508	0.9879	0.9933	0.9984	0.9922	0.9933	0.9651	0.9982
Mean	**0.9958**	**0.9201**	**0.9189**	**0.9635**	**0.9523**	**0.9878**	**0.9925**	**0.9983**	**0.9924**	**0.9936**	**0.9658**	**0.9980**
SD	**0.0003**	**0.0015**	**0.0014**	**0.0009**	**0.0010**	**0.0010**	**0.0003**	**0.0001**	**0.0002**	**0.0003**	**0.0007**	**0.0002**

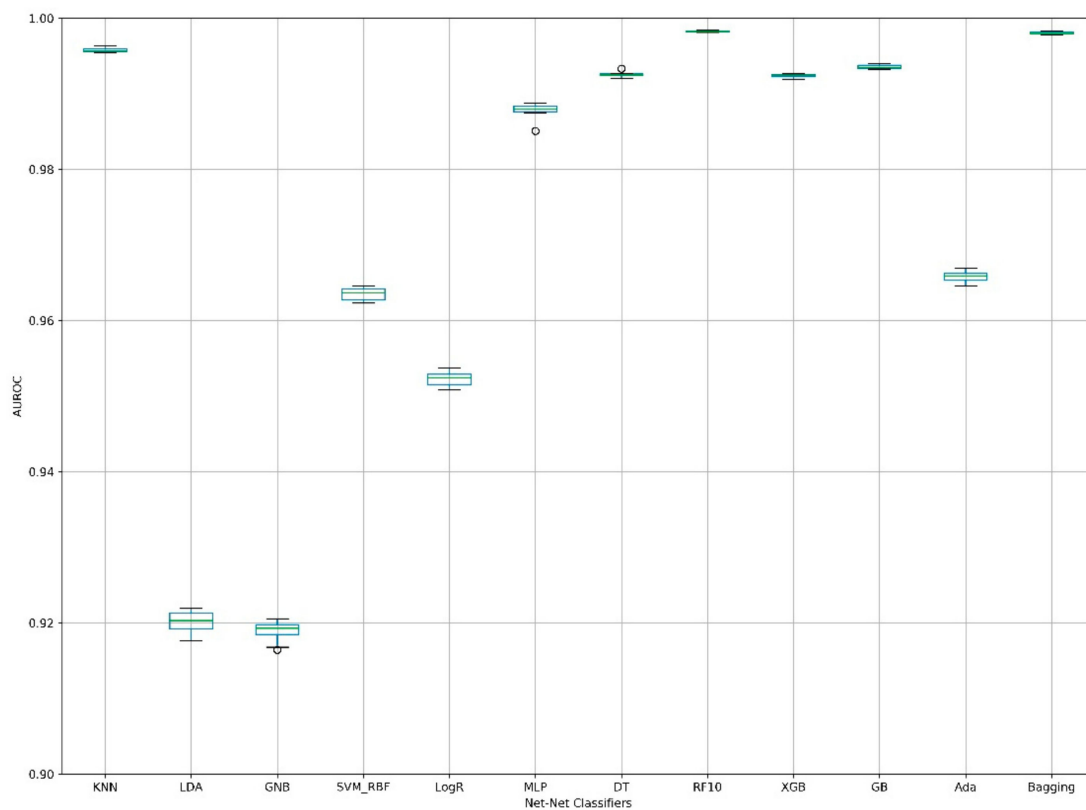[1] RF10 = Random Forest with 10 trees.

Figure 2. Box plot for the AUROC values of 12 Net-Net AutoML classification models.

In the next step, we performed a grid search for the RF number of trees: 1 tree (DT), and 2, 5, 10, 50, 100 trees (RF2, RF5, RF10, RF50, RF100). The results show that by using two trees with Random Forest, it is possible to increase the mean AUROC value from 0.9925 to 0.9944 (DT) (see Table 3 and Figure 3). By doubling the number of trees from 50 to 100, the difference between RF50 and RF100 for the mean AUROC value was just 0.0001. Therefore, we chose the best Net-Net AutoML model as RF50 with a mean AUROC value of 0.9991 (SD = 0.0001). The dataset, scripts, and results are available as an open GitHub repository [63].

Table 3. AUROC for Net-Net AutoML classification models.

Fold	DT	RF2 [1]	RF5 [2]	RF10 [3]	RF50 [4]	RF100 [5]
1	0.9926	0.9948	0.9976	0.9983	0.9990	0.9992
2	0.9924	0.9944	0.9971	0.9982	0.9990	0.9992
3	0.9920	0.9943	0.9971	0.9983	0.9992	0.9993
4	0.9927	0.9945	0.9974	0.9981	0.9992	0.9993
5	0.9924	0.9942	0.9972	0.9983	0.9991	0.9993
6	0.9925	0.9945	0.9974	0.9982	0.9991	0.9992
7	0.9925	0.9944	0.9974	0.9983	0.9992	0.9992
8	0.9927	0.9943	0.9973	0.9981	0.9990	0.9992
9	0.9922	0.9942	0.9972	0.9982	0.9991	0.9992
10	0.9933	0.9948	0.9975	0.9984	0.9991	0.9992
Mean	**0.9925**	**0.9944**	**0.9973**	**0.9983**	**0.9991**	**0.9992**
SD	**0.0003**	**0.0002**	**0.0002**	**0.0001**	**0.0001**	**0.0000**

[1] RF with two trees; [2] RF with five trees; [3] RF with 10 trees; [4] RF with 50 trees; [5] RF with 100 trees.
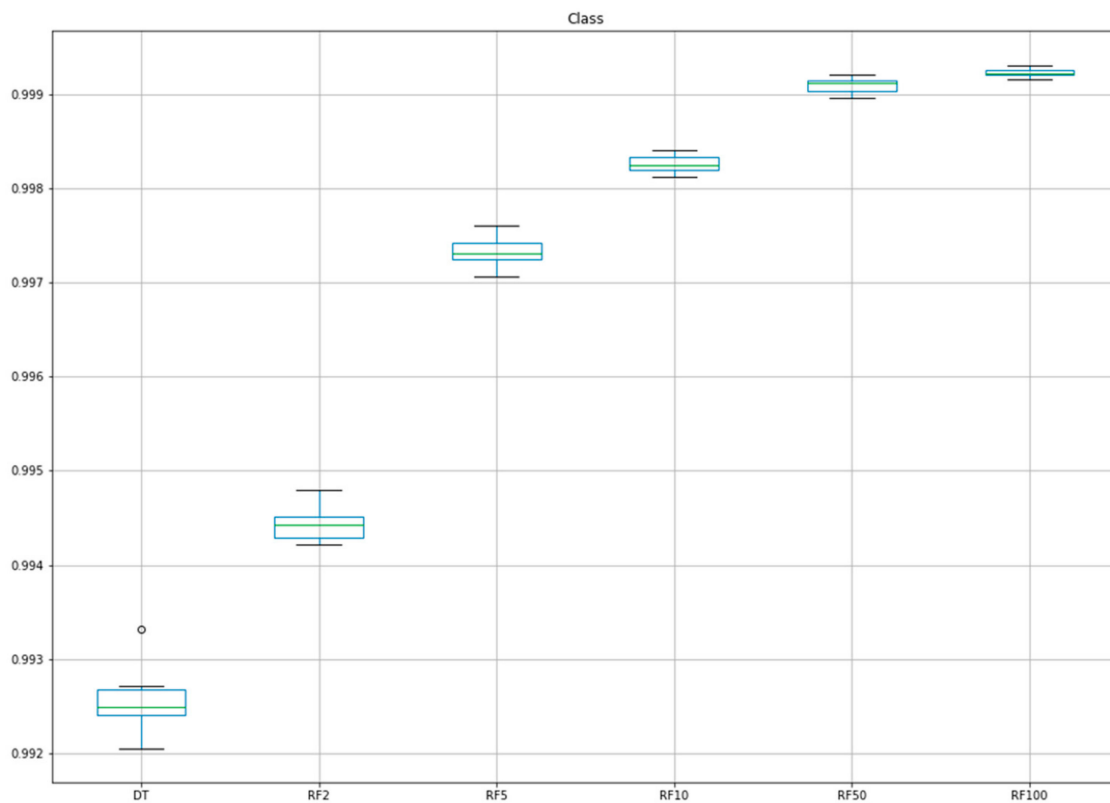
Figure 3. Box plot for the AUROC values of the NET-NET AutoML classification models.

Figure 4 shows the feature importance for the best Random Forest model based on 50 trees. We can observe that the entropies of the BCN nodes are the most important features followed by the ANN entropies. The differences between the BCN node entropies are less important for this RF classifier. The most important feature was the BCN node entropy containing information about the nodes placed at a minimum distance of 2—$Sh_2(B)$.
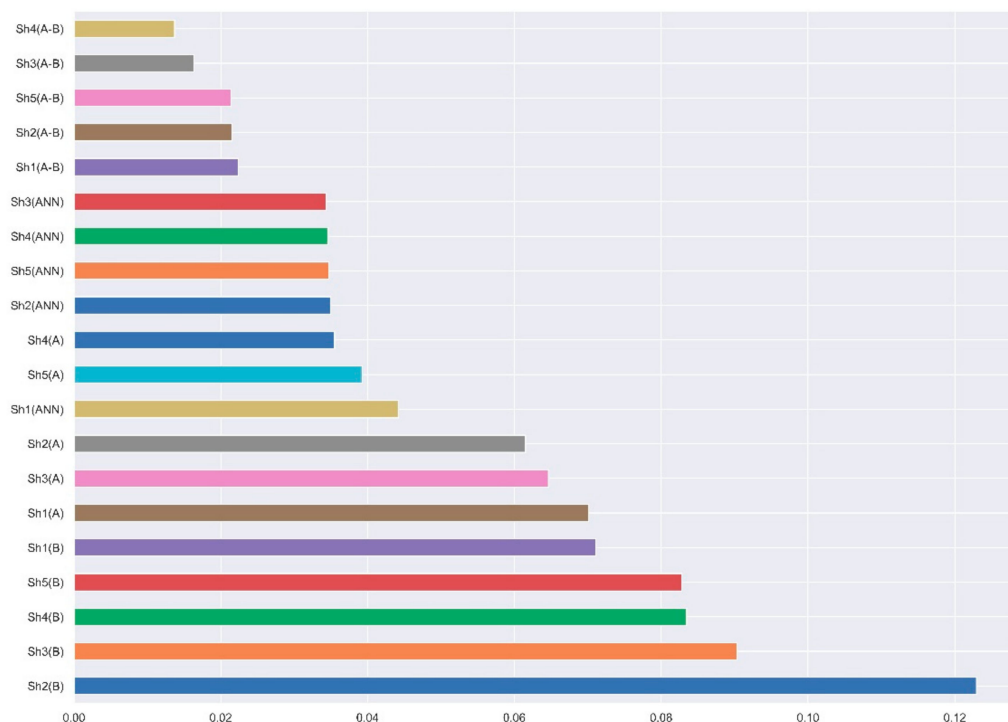
Figure 4. Feature importance for the best Random Forest model.

4. Conclusions

This work confirms that Markov chains are useful to calculate Shannon entropy information indices Sh_k that quantify the connectivity patterns on both BCNs and ANNs. We demonstrated how to develop Net-Net AutoML models based on Sh_k values of both networks. The dataset contains 500,470 examples and 20 features. Twelve linear and non-linear Machine Learning classifiers were tested and the best classification model was provided by Random Forest (50 trees) with the mean test AUROC of 0.9991 ± 0.0001 (10-fold cross-validation). The Net-Net AutoML models are useful to determine the ANN topology, which is better to predict the connectivity in the BCN system. Consequently, we can use this methodology to predict the ANN topologies, with better performance before training them. This may lead to an optimization of computing resources. The scripts and dataset for this project are available in an open GitHub repository from one of the authors, Cristian R. Munteanu [63].

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

References

1. Sandhu, K.S.; Li, G.; Poh, H.M.; Quek, Y.L.; Sia, Y.Y.; Peh, S.Q.; Mulawadi, F.H.; Lim, J.; Sikic, M.; Menghi, F.; et al. Large-scale functional organization of long-range chromatin interaction networks. *Cell Rep.* **2012**, *2*, 1207–1219. [CrossRef] [PubMed]

2. Gaspar, M.E.; Csermely, P. Rigidity and flexibility of biological networks. *Brief. Funct. Genom.* **2012**, *11*, 443–456. [CrossRef] [PubMed]

3. Csermely, P.; Korcsmaros, T.; Kiss, H.J.; London, G.; Nussinov, R. Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. *Pharm. Ther.* **2013**, *138*, 333–408. [CrossRef] [PubMed]

4. Vidal, M.; Cusick, M.E.; Barabasi, A.L. Interactome networks and human disease. *Cell* **2011**, *144*, 986–998. [CrossRef]

5. Barabasi, A.L.; Gulbahce, N.; Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **2011**, *12*, 56–68. [CrossRef]

6. Barabasi, A.L.; Oltvai, Z.N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **2004**, *5*, 101–113. [CrossRef]

7. Strogatz, S.H. Exploring complex networks. *Nature* **2001**, *410*, 268–276. [CrossRef]

8. Reijneveld, J.C.; Ponten, S.C.; Berendse, H.W.; Stam, C.J. The application of graph theoretical analysis to complex networks in the brain. *Clin. Neurophysiol.* **2007**, *118*, 2317–2331. [CrossRef]

9. Guo, L.; Rivero, D.; Dorado, J.; Munteanu, C.R.; Pazos, A. Automatic feature extraction using genetic programming: An application to epileptic EEG classification. *Expert Syst. Appl.* **2011**, *38*, 10425–10436. [CrossRef]

10. Liu, Y.; Tang, S.; Fernandez-Lozano, C.; Munteanu, C.R.; Pazos, A.; Yu, Y.Z.; Tan, Z.; González-Díaz, H. Experimental study and Random Forest prediction model of microbiome cell surface hydrophobicity. *Expert Syst. Appl.* **2017**, *72*, 306–316. [CrossRef]

11. Aguiar-Pulido, V.; Seoane, J.A.; Gestal, M.; Dorado, J. Exploring patterns of epigenetic information with data mining techniques. *Curr. Pharm. Des.* **2013**, *19*, 779–789. [CrossRef] [PubMed]

12. Fernandez-Blanco, E.; Rivero, D.; Gestal, M.; Dorado, J. Classification of signals by means of genetic programming. *Soft Comput.* **2013**, *17*, 1929–1937. [CrossRef]

13. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]

14. Dehmer, M.; Emmert-Streib, F. *Analysis of Complex Networks. From Biology to Linguistics*; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2009.

15. Dehmer, M.; Grabner, M.; Varmuza, K. Information indices with high discriminative power for graphs. *PLoS ONE* **2012**, *7*, e31214. [CrossRef] [PubMed]

16. Dehmer, M.; Varmuza, K.; Borgert, S.; Emmert-Streib, F. On entropy-based molecular descriptors: Statistical analysis of real and synthetic chemical structures. *J. Chem. Inf. Model.* **2009**, *49*, 1655–1663. [CrossRef] [PubMed]

17. Estrada, E.; Avnir, D. Continuous symmetry numbers and entropy. *J. Am. Chem. Soc.* **2003**, *125*, 4368–4375. [CrossRef]

18. Graham, D.J.; Grzetic, S.; May, D.; Zumpf, J. Information properties of naturally-occurring proteins: Fourier analysis and complexity phase plots. *Protein J.* **2012**, *31*, 550–563. [CrossRef]

19. Graham, D.J.; Greminger, J.L. On the information expressed in enzyme structure: More lessons from ribonuclease A. *Mol. Divers.* **2011**, *15*, 769–779. [CrossRef]

20. Graham, D.J.; Greminger, J.L. On the information expressed in enzyme primary structure: Lessons from Ribonuclease A. *Mol. Divers.* **2010**, *14*, 673–686. [CrossRef]

21. Graham, D.J.; Kim, M. Information and classical thermodynamic transformations. *J. Phys. Chem. B* **2008**, *112*, 10585–10593. [CrossRef]

22. Graham, D.J.; Malarkey, C.; Sevchuk, W. Experimental investigation of information processing under irreversible Brownian conditions: Work/time analysis of paper chromatograms. *J. Phys. Chem. B* **2008**, *112*, 10594–10602. [CrossRef] [PubMed]

23. Graham, D.J. Information content in organic molecules: Brownian processing at low levels. *J. Chem. Inf. Model.* **2007**, *47*, 376–389. [CrossRef] [PubMed]

24. Graham, D.J. Information content in organic molecules: Aggregation states and solvent effects. *J. Chem. Inf. Model.* **2005**, *45*, 1223–1236. [CrossRef] [PubMed]

25. Graham, D.J.; Schulmerich, M.V. Information content in organic molecules: Reaction pathway analysis via Brownian processing. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1612–1622. [CrossRef]

26. Graham, D.J.; Malarkey, C.; Schulmerich, M.V. Information content in organic molecules: Quantification and statistical structure via Brownian processing. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1601–1611. [CrossRef]

27. Graham, D.J. Information and organic molecules: Structure considerations via integer statistics. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 215–221. [CrossRef]

28. Graham, D.J.; Schacht, D.V. Base information content in organic formulas. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 942–946. [CrossRef]

29. Barigye, S.J.; Marrero-Ponce, Y.; Santiago, O.M.; Lopez, Y.M.; Perez-Gimenez, F.; Torrens, F. Shannon's, Mutual, conditional and joint entropy information indices. Generalization of global indices defined from local vertex invariants. *Curr. Comput. Aided Drug Des.* **2013**, *9*, 164–183. [CrossRef]

30. Aguiar-Pulido, V.; Munteanu, C.R.; Seoane, J.A.; Fernández-Blanco, E.; Pérez-Montoto, L.G.; González-Díaz, H.; Dorado, J. Naïve Bayes QSDR classification based on spiral-graph Shannon entropies for protein biomarkers in human colon cancer. *Mol. Biosyst.* **2012**, *8*, 1716–1722. [CrossRef]

31. Riera-Fernandez, P.; Munteanu, C.R.; Escobar, M.; Prado-Prado, F.; Martin-Romalde, R.; Pereira, D.; Villalba, K.; Duardo-Sanchez, A.; Gonzalez-Diaz, H. New Markov-Shannon Entropy models to assess connectivity quality in complex networks: From molecular to cellular pathway, Parasite-Host, Neural, Industry, and Legal-Social networks. *J. Theor. Biol.* **2012**, *293*, 174–188. [CrossRef] [PubMed]

32. Prado-Prado, F.J.; Garcia, I.; Garcia-Mera, X.; Gonzalez-Diaz, H. Entropy multi-target QSAR model for prediction of antiviral drug complex networks. *Chemom. Intellig. Lab. Syst.* **2011**, *107*, 227–233. [CrossRef]

33. Munteanu, C.R.; Magalhaes, A.L.; Uriarte, E.; Gonzalez-Diaz, H. Multi-target QPDR classification model for human breast and colon cancer-related proteins using star graph topological indices. *J. Theor. Biol.* **2009**, *257*, 303–311. [CrossRef] [PubMed]

34. Munteanu, C.R.; Gonzalez-Diaz, H.; Borges, F.; de Magalhaes, A.L. Natural/random protein classification models based on star network topological indices. *J. Theor. Biol.* **2008**, *254*, 775–783. [CrossRef] [PubMed]

35. Munteanu, C.R.; Gonzalez-Diaz, H.; Magalhaes, A.L. Enzymes/non-enzymes classification model complexity based on composition, sequence, 3D and topological indices. *J. Theor. Biol.* **2008**, *254*, 476–482. [CrossRef]

36. González-Díaz, H.; Pérez-Bello, A.; Cruz-Monteagudo, M.; González-Díaz, Y.; Santana, L.; Uriarte, E. Chemometrics for QSAR with low sequence homology: Mycobacterial promoter sequences recognition with 2D-RNA entropies. *Chemom. Intell. Lab. Syst.* **2007**, *85*, 20–26. [CrossRef]

37. González-Díaz, H.; Saíz-Urra, L.; Molina, R.; Uriarte, E. Stochastic molecular descriptors for polymers. 2. Spherical truncation of electrostatic interactions on entropy based polymers 3D-QSAR. *Polymer* **2005**, *46*, 2791–2798. [CrossRef]

38. Rodriguez-Soca, Y.; Munteanu, C.R.; Dorado, J.; Rabunal, J.; Pazos, A.; Gonzalez-Diaz, H. Plasmod-PPI: A web-server predicting complex biopolymer targets in plasmodium with entropy measures of protein-protein interactions. *Polymer* **2010**, *51*, 264–273. [CrossRef]

39. Jalali-Heravi, M.; Fatemi, M.H. Prediction of thermal conductivity detection response factors using an artificial neural network. *J. Chromatogr. A* **2000**, *897*, 227–235. [CrossRef]

40. Prado-Prado, F.J.; Garcia-Mera, X.; Gonzalez-Diaz, H. Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. *Bioorg. Med. Chem.* **2010**, *18*, 2225–2231. [CrossRef]

41. Tenorio-Borroto, E.; Penuelas Rivas, C.G.; Vasquez Chagoyan, J.C.; Castanedo, N.; Prado-Prado, F.J.; Garcia-Mera, X.; Gonzalez-Diaz, H. ANN multiplexing model of drugs effect on macrophages; theoretical and flow cytometry study on the cytotoxicity of the anti-microbial drug G1 in spleen. *Bioorg. Med. Chem.* **2012**, *20*, 6181–6194. [CrossRef]

42. Gonzalez-Diaz, H.; Bonet, I.; Teran, C.; De Clercq, E.; Bello, R.; Garcia, M.M.; Santana, L.; Uriarte, E. ANN-QSAR model for selection of anticancer leads from structurally heterogeneous series of compounds. *Eur. J. Med. Chem.* **2007**, *42*, 580–585. [CrossRef] [PubMed]

43. Gonzalez-Diaz, H.; Arrasate, S.; Sotomayor, N.; Lete, E.; Munteanu, C.R.; Pazos, A.; Besada-Porto, L.; Ruso, J.M. MIANN models in medicinal, physical and organic chemistry. *Curr. Top. Med. Chem.* **2013**, *13*, 619–641. [CrossRef] [PubMed]

44. Kotthoff, L.; Thornton, C.; Hoos, H.H.; Hutter, F.; Leyton-Brown, K. Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. In *Automated Machine Learning*; The Springer Series on Challenges in Machine Learning; Frank, H., Ed.; Springer: Berlin, Germany, 2017; pp. 81–95. [CrossRef]

45. Feurer, M.; Klein, A.; Eggensperger, K.; Springenberg, J.; Blum, M. Efficient and robust automated machine learning. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2015; pp. 2962–2970.

46. Barreiro, E.; Munteanu, C.R.; Cruz-Monteagudo, M.; Pazos, A.; Gonzalez-Diaz, H. Net-Net auto machine learning (AutoML) prediction of complex ecosystems. *Sci. Rep.* **2017**, *8*, 12340. [CrossRef] [PubMed]

47. Modha, D.S.; Singh, R. Network architecture of the long-distance pathways in the macaque brain. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13485–13490. [CrossRef]

48. StatSoft. Inc. *STATISTICA (Data Analysis Software System), Version 6.0*; StatSoft. Inc.: Tulsa, OK, USA, 2002.

49. Duardo-Sanchez, A.; Gonzalez-Diaz, H.; Pazos, A. MI-NODES multiscale models of metabolic reactions, brain connectome, ecological, epidemic, world trade, and legal-social networks. *Curr. Bioinform.* **2015**, *10*, 692–713. [CrossRef]

50. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]

51. Winkel, P.; Juhl, E. Assumptions in linear discriminant analysis. *Lancet* **1971**, *2*, 435–436. [CrossRef]

52. Lowd, D.; Domingos, P. Naive Bayes models for probability estimation. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 529–536.

53. Han, S.; Qubo, C.; Meng, H. Parameter selection in SVM with RBF kernel function. In Proceedings of the World Automation Congress 2012, Puerto Vallarta, Mexico, 24–28 June 2012; pp. 1–4.

54. Hilbe, J.M. *Logistic Regression Models*; Chapman & Hall/CRC Press: Boca Raton, FL, USA, 2009.

55. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1994.

56. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]

57. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

58. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

59. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

60. Freund, Y.; Schapire, R.E. A Decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 20. [CrossRef]

61. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]

62. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]

63. Van Essen, D.C.; Ugurbil, K.; Auerbach, E.; Barch, D.; Behrens, T.E.; Bucholz, R.; Chang, A.; Chen, L.; Corbetta, M.; Curtiss, S.W.; et al. The Human Connectome Project: A data acquisition perspective. *NeuroImage* **2012**, *62*, 2222–2231. [CrossRef] [PubMed]

64. Lang, E.W.; Tome, A.M.; Keck, I.R.; Gorriz-Saez, J.M.; Puntonet, C.G. Brain connectivity analysis: A short survey. *Comput. Intell. Neurosci.* **2012**, *2012*, 412512. [CrossRef]

65. Stephan, K.E.; Kamper, L.; Bozkurt, A.; Burns, G.A.; Young, M.P.; Kotter, R. Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **2001**, *356*, 1159–1186. [CrossRef]

66. Kotter, R. Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics* **2004**, *2*, 127–144. [CrossRef]

67. Sanbonmatsu, K.Y.; Tung, C.S. High performance computing in biology: Multimillion atom simulations of nanoscale systems. *J. Struct. Biol.* **2007**, *157*, 470–480. [CrossRef]

68. Pitera, J.W. Current developments in and importance of high-performance computing in drug discovery. *Curr. Opin. Drug Discov. Dev.* **2009**, *12*, 388–396.

69. Maniatis, T.A.; Nikita, K.S.; Uzunoglu, N.K. Ultrasonic diffraction tomography: An application connecting high performance computing centers with clinical environment. *Stud. Health Technol. Inform.* **2000**, *79*, 214–243. [CrossRef]

70. Johnston, W.E.; Jacobson, V.L.; Loken, S.C.; Robertson, D.W.; Tierney, B.L. High-performance computing, high-speed networks, and configurable computing environments: Progress toward fully distributed computing. *Crit. Rev. Biomed. Eng.* **1992**, *20*, 315–354. [PubMed]

71. Fernandez, J.J. High performance computing in structural determination by electron cryomicroscopy. *J. Struct. Biol.* **2008**, *164*, 1–6. [CrossRef] [PubMed]

72. Dunning, T.H., Jr.; Harrison, R.J.; Feller, D.; Xantheas, S.S. Promise and challenge of high-performance computing, with examples from molecular modelling. *Philos. Trans. Ser. Math. Phys. Eng. Sci.* **2002**, *360*, 1079–1105. [CrossRef] [PubMed]

73. Cant, S. High-performance computing in computational fluid dynamics: Progress and challenges. *Philos. Trans. Ser. Math. Phys. Eng. Sci.* **2002**, *360*, 1211–1225. [CrossRef] [PubMed]