# Neural Network Predictive Analysis of Echoic Behavior for Children with Autism

Chris Ninness[1] and Lee Mason[2,3]

[1]Behavioral Software Systems & Human Interventions Institute

[2]Child Study Center, Cook Children's Health Care System

[3]Burnett School of Medicine, Texas Christian University

**Author Note**

Chris Ninness https://orcid.org/0000-0003-3884-5327

Lee Mason https://orcid.org/0000-0001-9431-9132

Correspondence concerning this article should be addressed to Chris Ninness, Behavioral Software Systems & Human Interventions Institute, 2207 Pinecrest Dr., Nacogdoches, TX 75965. Email: cninness@suddenlink.net

The authors would like to thank Christopher Allmon and Aurora Bradford for their assistance with data aggregation and interobserver agreement. Additionally, we thank Jeff Weatherly, Mark Mattaini, and

Gordon Bourland for critiquing an earlier draft of this paper.

## Abstract

As a primer on machine learning strategies applied to behavior analytic investigations, we sought to predict the likelihood that particular children with autism spectrum disorder (ASD) would develop echoic verbal behavior. After retrieving 143 case records of children assessed with the VB-MAPP from a data repository, we reserved 22 case records for holdout analysis. Through iterative development, the model achieved 100% accuracy in predicting the presence or absence of an echoic repertoire. While the functional value of predicting a known outcome might be questioned, it is only possible to gauge the accuracy of any prediction technique when known outcomes are available. Thus, rather than trying to predict a completely unknown outcome, our experimental preparations were directed at gauging the multilayer perceptron's (MLP) ability to predict at high levels of precision, sensitivity, specificity, and overall prediction accuracy while accessing a limited dataset from one children's healthcare setting. We discuss procedures for employing this machine learning strategy in conjunction with a receiver operator characteristics analysis of our findings. We developed this study as a general guide for using one of several quickly evolving neural network methodologies for behavior analytic researchers, and we examine specifics regarding MLP functions and operations as they relate to providing more efficacious treatment for children with ASD.

*Keywords:* children with autism, echoic skills assessment, multilayer perceptron’s, machine learning, cross-validation, behavior analysis

# Neural Network Predictive Analysis of Echoic Behavior for Children with Autism

During recent years, a growing body of research points to the advantages of using echoic prompts when training complex verbal behavior in children with autism spectrum disorder (ASD), a neurodevelopmental disorder characterized by behavioral deficits (i.e., social communication and interaction) and excesses (i.e., restricted, repetitive behaviors; Kodak et al., 2020; Muharib et al., 2021; Roncati et al., 2019; Shillingsburg et al., 2022). While early intensive behavioral intervention has been shown helpful to remediate language deficits (Rivard & Forget, 2012), as many as 1/3 of children with ASD will fail to develop vocal language (Rose et al., 2016). Consequently, these and related studies (e.g., Vladescu et al., 2021) speak to the potential benefits of being able to predict echoic repertoires.

To highlight these possibilities, consider the following example. A practitioner working with a child with ASD may be interested in determining whether the child will be able to echo the language of others. Predictive analytic models might help address this issue by identifying the factors associated with this particular ability. By looking for commonalities among an extensive database of various characteristics of children with ASD who do and do not demonstrate echoic repertoires, applied researchers might construct a model of relevant variables specific to children who demonstrate the ability to echo. Then, by comparing the child's skills against the model, the practitioner can determine whether the conditions are right for their child to begin echoing other people's language. However, until very recently, no existing technology has been able to predict such outcomes with high levels of accuracy.

The heterogeneity of ASD makes it challenging to draw broad generalities about the disorder. The multiple causes of ASD have yet to be identified, and symptoms of the disorder range broadly across a variety of different behavioral excesses and deficits. Consequently, predicting treatment effects for this group has been problematic (Bent & Hendren, 2010; Trembath & Vivanti, 2014; Vivanti et al, 2014). Modern advances in machine learning (ML) may prove beneficial to enhancing the predictive outcomes for children with ASD.

Behavior analysts have recently examined the use of ML for predicting treatment outcomes of children with ASD who received behavior-analytic intervention. These researchers have compared various types of statistical and machine learning algorithms, such as linear regression (Linstead et al., 2017; Maharjan et al., 2023; Préfontaine et al., 2022), decision trees (Maharjan et al.; Préfontaine et al.), k-nearest neighbors, Gaussian process, and support vector machines (Préfontaine et al.), cosine similarity and collaborative filtering (Kohli et al., 2022), as well as artificial neural networks (Linstead et al.).

Using different neural network techniques, researchers have examined a variety of predictive outcomes for children diagnosed with ASD. These variables include treatment intensity (Linstead et al., 2017; Maharjan et al., 2023), the selection of treatment goals (Kohli et al., 2022); mastery of learning outcomes (Linstead et al.), and treatment prognosis (Préfontaine et al., 2022). Across all studies machine learning was effective, with the predictive accuracy positively correlated with the complexity of the machine learning algorithm. That is, for the three studies that compared complex (i.e., artificial neural networks, decision trees) against simple (i.e., linear regression) algorithms (Linstead et al., Maharjan et al.), the more complex machine learning algorithms always outperformed linear regression. Préfontaine et al. (2022), who compared five complex algorithms against one another, concluded, “None of the algorithms produced systematically worst predictions than all the others” (p. 4). Indeed, other approaches to machine learning are having an impact across the academic spectrum. In particular, Turgeon & Lanovaz (2020) demonstrated that ML algorithms such as random forest, support vector classifier, stochastic gradient descent, and k-nearest neighbors are powerful prognostic systems when classifying a variety of datasets in applied behavior analytic research.

With recent advances in machine learning, neural network techniques have become increasingly accurate, flexible, and inexplicable (Dafoe, 2018; Kissinger et al., 2021). As will be described, neural networks are frequently capable of solving the most intractable of problems. When operating at peak performance, these systems can disentangle and predict outcomes for the most evasive social (e.g., Ninness et al., 2021), physiological (e.g., Almeida & Azkune, 2018; Ninness et al., 2012), and behavior issues (e.g., Phan et al., 2017; Dufour, 2020; Ninness & Ninness, 2020a). See Turgeon & Lanovaz (2020) for a well-organized machine learning tutorial focusing on behavior-analytic investigations.

An irrefutable fact is that neural networks can be remarkably accurate but retroactively inscrutable. Unlike logistic regression or multiple regression procedures, these algorithms generate an extraordinarily large number of weights and bias values. Identifying the precise functions of these values is an active area of investigation but is far from being accomplished (Ninness & Ninness, 2020b). Notwithstanding, because of their high levels of precision, sensitivity, and specificity, neural networks (and a wide range of related machine learning systems) are in a state of exponential evolution and expansion throughout the academic (e.g., Lyddy & Barnes-Holmes 2007; Ninness et al., 2018; Smith & Hayes, 2022; Tovar & Westermann, 2017), medical (e.g., Chicco & Jurman, 2020), and corporate communities (e.g., Özener et al., 2013; Tegmark, 2017).

## MLP Overview

We used the Python programming language to train our multilayer perceptron (MLP) neural network model. MLPs are a type of feedforward neural network architecture employed extensively throughout the area of artificial intelligence (see McCaffrey, 2023, for examples). MLPs are comprised of layers of nodes that have activation functions applied to their output (i.e., see Table 1) that learn unique data patterns as they process input values. During network training, the weights of the connections between the nodes within the network layers adjust given their input and output values. These weights are continually updated as the data are passed forward (feedforward/forward pass) and backward (backpropagation/backward pass) across the layers of the network throughout training (Nielsen, 2015; Ninness et al., 2019).

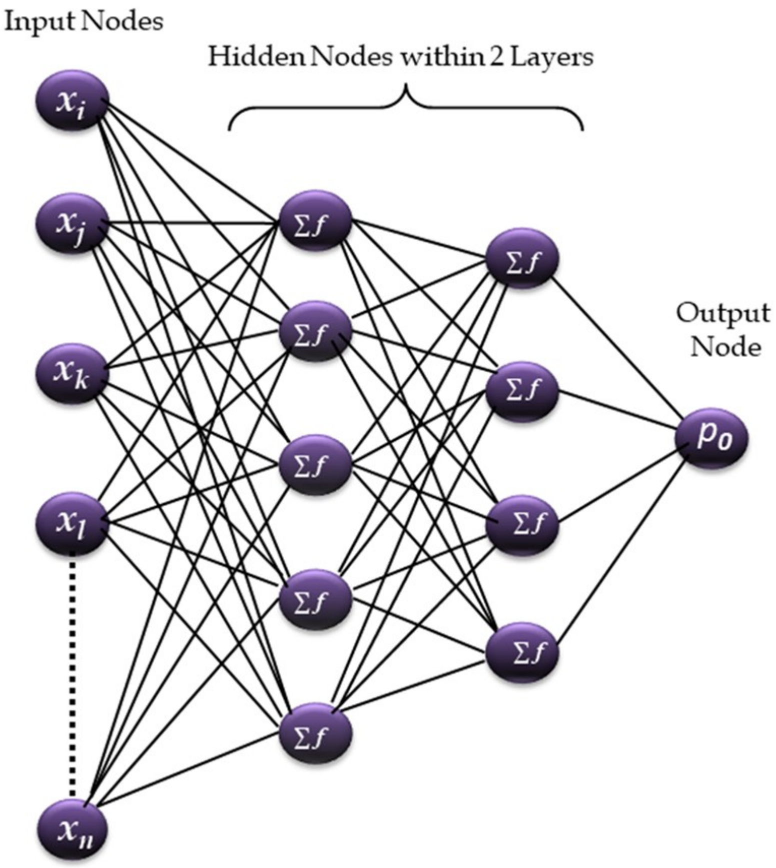
One of the fascinating features of MLPs is the way in which training is initiated using a series of entirely randomized values. During the feedforward pass, connections between individual nodes (neurons) within layers (node layers) are initialized with small random weights (e.g., between -0. 10 and +0. 10). Data are presented to the network via the input layer, and this layer contains one node for each independent variable within the dataset. Dependent variables are represented on an output layer.

For example, the current architecture (network structure) includes 10 input nodes and one output node. When training begins, the input values are passed to the first hidden layer and multiplied by small random weights. As illustrated in Figure 1, the sums of all weighted input values are then passed to the hyperbolic tangent (*tanh*) activation function within each hidden node (see ∑**ƒ** in Figure 1). This activation function (see Figure 2 and Table 1) is referred to as a squashing function because it systematically compresses input values of any size into outputs ranging between -1 and +1. Note that alternative activation functions (e.g., ReLu) accept values from negative to positive infinity and return values within a specific range (see Islam et al., 2021, for a discussion). Upon passing through all of the hidden layers, the data (signals) reach the node in the output layer, where the sigmoid function compresses outputs ranging between 0 and 1(see $p_0$ in Figure 1).

**Bias Nodes**. Although not displayed in Figure 1, bias nodes are constants usually initialized with values of 1. Bias node values are used during training to offset predicted outcomes away from the center of the coordinate axis. These values are trained much the same as other inputs in that they are multiplied by small random weights and ultimately become part of the trained model (see Table 1 for related details).

**Figure 1**

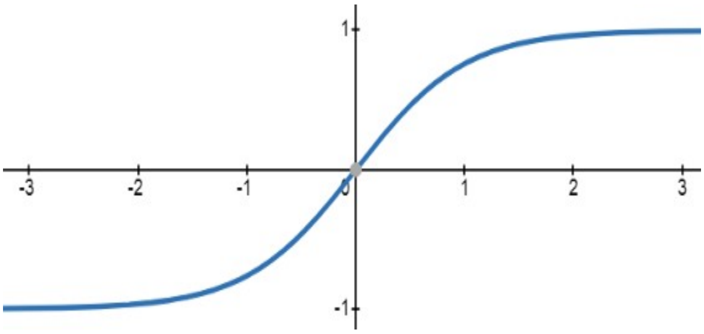
*Diagram of the MLP Architecture Employed in the Current Study*




Outputs from the *tanh* function in the first layer are again multiplied by randomized weights, summed, and passed downstream to the nodes within the succeeding layers. Throughout the feedforward pass, the data (signals) are repeatedly multiplied by weights, summed, and fed to the next *tanh* transformation function within each hidden layer. The hidden nodes operate as feature detectors, playing an essential role in the MLP learning process.

**Figure 2**

*Illustration of the Hyperbolic Tangent Function*




Haykin (2009) notes that as the network training process advances, "…the hidden neurons begin to gradually 'discover' the salient features that characterize the training data. They do so by performing a nonlinear transformation on the input data into a new space called the feature space" (p. 126). During the early training iterations, there are large differences between the network's calculated output values and the known target values. These differences are forms of error. During backpropagation, the algorithm attempts to gradually minimize these errors by updating its weights so that the calculated outputs become better approximations of the known target values.

**Backpropagation**

Backpropagation is a network training process that entails a few specialized terms and computations. In conjunction with the feedforward process, backpropagation performs a series of corrections to the error values that occur during training. As described earlier, over the course of many forward and backward passes these operations shrink the error value, creating a neural network model with increasingly accurate predictions. Error is never completely eliminated, but is instead brought into an acceptable level of error by minimizing a loss function which minimizes the difference between the *predicted* outputs and the *actual* outputs. If the error level is minimized too much, then the model may be subject to overfitting which is liable to cause poor out-of-sample performance.

Depending on the size and diversity of the input values, network training may entail several thousand iterations before the MLP learns the relations among input values and known target values. In the current study, this algorithm required 576 iterations to learn the relations among the variables employed. Table 1 provides a glossary of common MLP terms.

**Stochastic Gradient Descent**

Analogous to conventional statistics, a neural network determines predicted outcomes in terms of probability and error. Just as the standard deviation is one of the more frequently employed error values used in conventional statistics, the mean square error (MSE) is one of the more commonly used error values in the neural network literature. Loss and cost are two more important terms in the neural network lexicon. Loss refers to the error value computed for a training instance, while cost refers to the average loss value when all instances have been computed. Within the current illustrative study, we will emphasize the loss function.

In equation 1, MSE is the loss function for the neural network. Throughout training, it measures how well the algorithm reduces the differences between the known target value ($\hat{Y}$) and its calculated ($Y$)

value. As its name suggests, the MSE is obtained by repeatedly squaring the differences between the target value and the MLP's most recently computed output, summing (∑) this squared difference, and then dividing the sum by the number (*n*) of data points employed for the number of completed iterations. More concisely, the MSE is the mean of the squared differences between the MLP's predicted and known target outcomes (see Equation 1).

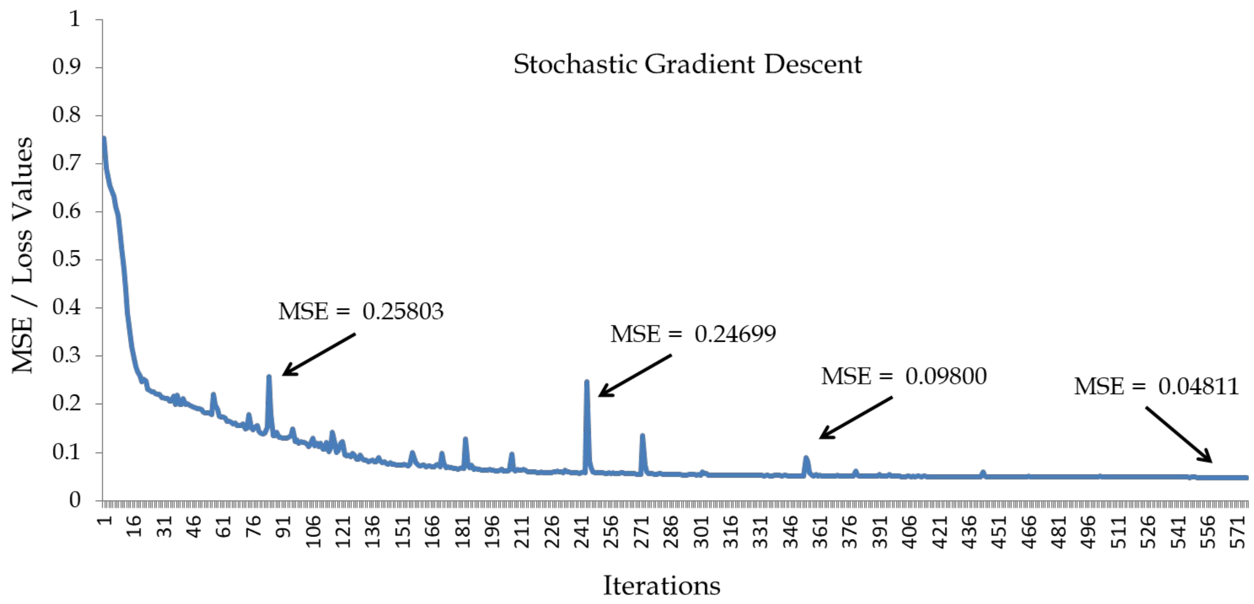
$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$

**(1)**

If network training progresses as intended, the MLP learns through gradual but slightly erratic approximations. Figure 3 illustrates that as the number of iterations increases, the MSE shrinks while the MLP's calculated output values become better estimates of the known target values. While the gradient descent illustrates a transition from 1 to a stabilized value 0. 0026, there are conspicuous spikes of higher error across iterations.

**Figure 3**

*Display of Stochastic Gradient Descent across 576 Iterations*




Gradients represent the magnitude and direction of change of the output signals (Haykin, 2009). That is, gradients measure weight variations (error changes) that occur while training a model. Particularly at the beginning of training there is oscillation in the error values (see Figure 3). When the gradient changes are consistent with improved learning, the error shrinks toward some minimal threshold. These error values tell us the distance between the network's current calculated values and the known target values. As the gradient gradually descends, it moves toward a stable series of minimal errors. In accordance with the values shown in Figure 3, when these values stabilize within ± 0. 0001 over 100 consecutive iterations, training is discontinued.

**Hyperparameters**

Before network training begins, there are a series of external settings that the researcher specifies. *Hyperparameters* (often referred to as parameters) are the settings by the researcher that initiate and continue to adjust the neural network learning process. The *learning* *rate* hyperparameter specifies each updated weight value's speed (step size) within the network's hidden layers. The *momentum* factor identifies the fraction of the previous weight change induced by the learning rate. It regulates accurate training and helps sustain the network's most functional learning speed. Likewise, the number of hidden layers and nodes within each layer are set, and the number of iterations to be performed and when to stop performing training (stopping rules) are designated by the researcher. Batch size refers to the number of sample records employed during training iterations, and iterations refers to the number of times the network's parameters are updated while training is in progress (See Table 1 for a list of commonly used neural network terms).

There are a growing number of approaches for selecting hyperparameters when training different types of neural networks, including a range of automated algorithms. To name only a few of these approaches, grid search (Jiménez et al., 2007), randomized search (Probst et al., 2019), Bayesian optimization (Victoria & Maragatham, 2020), and gradient-based optimization (Bengio, 2000) are popular techniques for selecting optimal hyperparameters. Providing details on these techniques is beyond the scope of this paper, but see Yang & Shami (2020) for an examination of these and related systems.

As an alternative to automated algorithms, a manual search allows the developer to systematically explore a combination of hyperparameters and develop a model that maximizes network performance. Manual hyperparameter tuning is an interface between human logic and machine learning. Analogous to how we adjust various audio settings on a stereo system (e.g., bass, treble, balance, etc.), a network algorithm can be tuned to improve its performance during training. In the current study, our hyperparameters were tuned by conducting training sequences wherein we observed the network's accuracy levels and adjusted the hyperparameters accordingly. Note that raw scores that were not originally expressed as binary values were normalized to numbers falling between 0 and 1using the Min-Max normalization function.

We set the maximum number of iterations ('max_iter') at 5,000 while specifying that if the model did not reduce error during a series of 100 consecutive iterations, training would stop ('n_iter_no_change': 100). This strategy allowed us to verify that the MSE had reached a threshold of diminished and stabilized loss. Several of these strategies and parameters follow those recommended by McCaffrey, 2023. Rather than list all parameter settings here, we provide the complete MLP neural network model and all parameters employed in this study at [https://www.behavioralsoftwaresystems.com](https://www.behavioralsoftwaresystems.com)

**Table 1**

*General Deep Neural Network Terms and Explanations*

| **Term** | **Explanation** |
|---|---|
| *Iterations* | Iterations refer to the number of times the network's parameters are updated during training, |
| *Momentum* | During backpropagation, the momentum factor sustains accurate training and improves the network's ability to maintain the most functional learning speed. This is accomplished when the |

| | |
|---|---|
| | momentum factor identifies the portion of the previous weight change to be added to the next weight change. |
| $\sum \boldsymbol{f}$ | The summation symbol (∑) refers to the cumulating sum of weighted input values. The symbol $f$ within (∑$f$) refers to a particular activation function that accepts input values. The hyperbolic tangent function (tanh) is the activation function employed within the current architecture. |
| *Sigmoid* | A function that receives inputs and produces an s-shaped curve consisting of output values ranging between 0 and 1. |
| *Tanh* | The hyperbolic tangent function (tanh) is one of several commonly employed neural network activation functions. Tanh is a squashing function that accepts the sum of weighted inputs and returns a value lying within -1. 0 and +1. 0. |
| *Training data* | A randomly selected set of case records presented to a neural network algorithm allowing the model to identify and learn patterns during cross-validation trials. |
| *Validation data* | A randomly selected subset used to assess the model’s progress during training. |
| *Activation Function* | Calculates the manner in which weighted sums of input values are transformed into outputs within the nodal layers of the network |
| *Holdout data* | Refers to a segregated dataset used to assess the generalization accuracy of the model on data that were not employed during any part of the training procedures. |
| *Bias* | During gradient descent bias values (typically 1s) help keep the outputs from the activation function (tanh) away from the center of the y-axis. |
| *Loss Function* | Throughout training, the loss function continually contrasts the differences between the known target values and the MLP’s predicted values. In doing so, it measures how accurately the MLP is learning these relations. |
| *Stochastic* | A network algorithm that gradually approximates |

| | |
|---|---|
| *Gradient Descent* | the least squared regression line by reducing the loss function (MSE) during a series of training iterations. |

The goal in developing a new dedicated MLP architecture is to increase the accuracy of predictions for a specific type of investigation. Unlike conventional statistical prediction methods (e.g., binary logistic regression), this goal can only be reached by training and systematically evaluating the network's predictive accuracy "as it is being developed" (Wenliang & Seitz, 2018). If the training process is completed successfully, the researcher can employ the newly developed model to predict outputs that match or closely resemble the target outputs. This technique allows the investigator to employ the newly-developed architecture to forecast entirely new and previously unseen data. The finalized architecture is identified as a specific type (or version) of neural network model. As described by Ninness et al. (2018), such a model is designed to forecast performance outcomes for new data. That is, the trained model is likely to generalize its learning and make accurate predictions when presented with unseen or previously nonexistent data – data that were completely unknown when the model was being trained. Haykin (2009) describes the network training process as follows:

> The hope is that the network becomes well trained so that it learns enough about the past to generalize to the future. From such a perspective, the learning process amounts to a choice of network parameterization for a given set of data. More specifically, we may view the network selection problem as choosing, within a set of

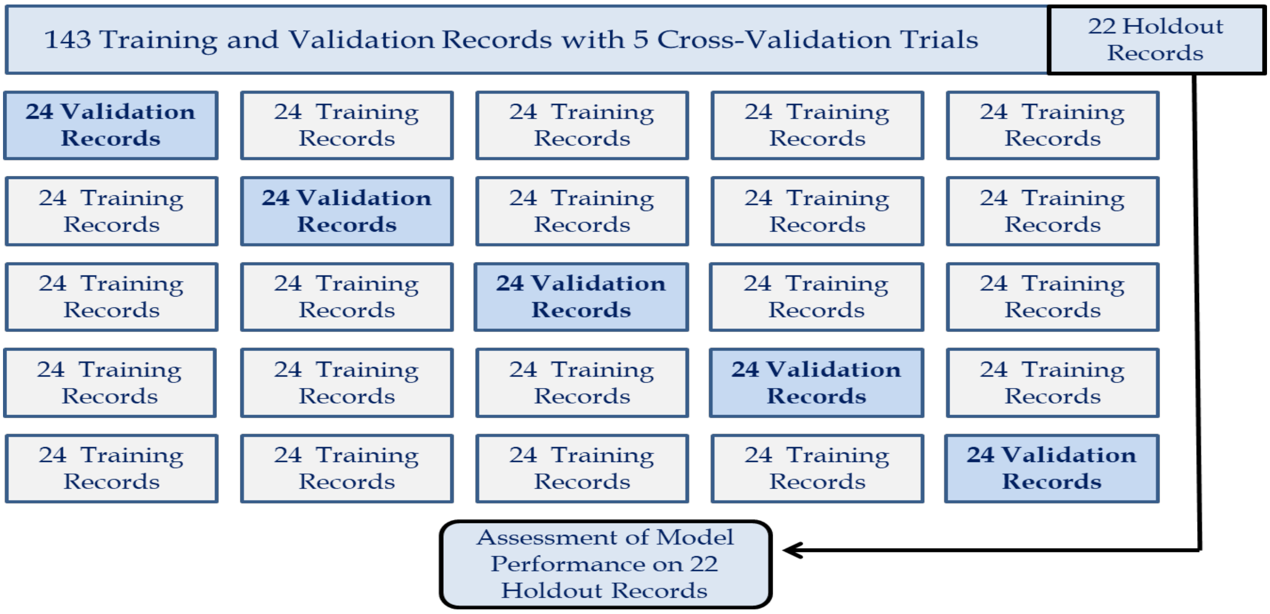
> candidate model structures (parameterizations), the “best” one according to a certain criterion. (p. 171)

**Cross-Validation**

Nothing provides predictive confirmation like empirical evidence. Cross-validation (CV) techniques are employed to confirm, or negate, a statistical or neural network algorithm’s ability to forecast performance outcomes accurately when presented with previously unseen input values. Stevens (2001) describes the advantages of performing CVs when conducting any type of logistic-regression analysis. In particular, Stevens emphasizes, “…although there are many procedures for selecting a ‘good’ model, the acid test is the generalizability of the model. In other words, how well will the model chosen fit on an independent sample? This leads us once again into cross-validation…” (p. 586). There are several strategies for conducting CVs (e.g., Haykin, 2009; Plis et al., 2014; Wong, 2015). As an overview, the technique entails conducting a series of assessments wherein subsets (folds) of randomly selected training and validation data are extracted from an available database. Typically, CV procedures divide large datasets into separate training sets, validation sets, and a single holdout (test) set. As described by Hagan (2002), “When we divide the data, approximately 70% is typically used for training, with 15% for validation and 15% for testing. These are only approximate numbers.” (Chapter 13, p. 8) Outcomes from CV procedures allow us to compare and contrast the accuracy, precision, sensitivity (also termed recall), and specificity of the predictive models. Similar procedures have been used to simulate and predict human behavior (cf. Greene, 2017; Ninness & Ninness, 2020a). Figure 4 illustrates our CV configuration in which training and validation is conducted across 120 participant records before assessing the model's performance on an unseen dataset. To allow a balanced and consistent set of 24 training and holdout records as shown in Fig, 4, one participant record was not included during our CV analysis. However, as will be discussed, this record was included in the final 121 training records:

**Figure 4**

*Display of the Training and Validation Sequences during Cross-Validation*




As shown in Fig 4, CV trials were rotated across five-folds, and each trial employed a different subset for validation. For each trial, accuracy was assessed by measuring how well the trained model predicted outcomes for its particular validation fold. The model's performance was obtained by averaging its predictive accuracy across all CV trials. Irrespective of CV accuracy, a network model remains an abstraction until it accurately predicts outcomes for the completely unseen holdout dataset.

**Predicting Verbal Behavior.** Estimates indicate 1 in 44 children has been identified with ASD (Maenner et al., 2021). Autism is a neurodevelopmental disability characterized by language and social skills deficits, in addition to excesses of circumscribed behavior. While symptoms can vary broadly, approximately one-half of individuals with ASD show profound language deficits, of which one-fourth are functionally non-verbal (Baghdadli et al., 2012; Rose et al., 2016). Some children with ASD may acquire language skills through early intensive behavioral intervention, while others will remain non- or minimally verbal. For practitioners, it may be helpful to identify the children who show the prerequisite skills for developing language.

Prior research on early intensive behavioral intervention has suggested that children with ASD who demonstrate an echoic repertoire, in conjunction with higher IQ and lower age, are more likely to develop spoken language (Goldstein, 2002; Sallows & Graupner, 2005). Echoic behavior is distinctive in that it shares point-to-point correspondence with a prior verbal stimulus (Ingvarsson, 2016). Also known as verbal imitation, echoic behavior is frequently considered the mechanism by which other forms of verbal behavior are learned (Palmer, 2012), a cornerstone for conditioning the behavior of the listener (Schlinger, 2008a), and the basis for remembering (Schlinger, 2008b). The acquisition of an echoic repertoire has been shown to increase independence and reduce problem behavior (Cividini-Motta et al., 2017). Consequently, the ability to predict an echoic repertoire may be of special value to applied researchers and clinicians.

In the current demonstration study, our ambition in conducting predictive-analytic procedures is to cultivate interventions for a child's current verbal skills and their potential for advancing these skills. To fulfill this ambition, we examined the intake evaluations of children with ASD on the *Verbal Behavior Milestones Assessment and Placement Program, Second Edition* (VB-MAPP; Sundberg, 2014).

## Experimental Objectives

We explored the extent to which scores on the various domains of the VB-MAPP are predictive of a child’s echoic repertoire as measured by the *Early Echoic Skills Assessment* (EESA; Esch, 2008). The ability to reliably and accurately predict the presence and absence of an echoic repertoire is an indication of the interrelatedness of the domains on the VB-MAPP. Based on these outcomes, we may be able to determine the existence of behavioral profiles indicative of vocal imitation.

To assess these potentials, we employed procedures that allowed us to identify the degree of accuracy, precision, sensitivity, and specificity of these techniques. In doing so, we examine MLP operations as they relate to more efficacious treatment for children with ASD.

## Research Issues and Methods

**Participants and Settings.** Archival data were collected from 143 young children with ASD who received early intensive behavioral intervention (EIBI) across a seven-year period between 2015 and 2022. As part of the comprehensive EIBI services provided by a not-for-profit children’s healthcare center, a *Verbal Behavior Milestones Assessment and Placement Program, Second Edition* (VB-MAPP; Sundberg, 2014) was administered to each patient every six months. Patients were 80% male, and ranged in age from 1 year, 10 months to 8 years, 11 months at the time of their initial assessment. Patients had been independently diagnosed with ASD by a medical doctor prior to their enrollment in EIBI services.

The VB-MAPP is a criterion-referenced assessment of basic language and language-related skills that is commonly used to assess the communication and social skills deficits of children with ASD. The VB-MAPP consists of five different components: Milestones Assessment, Barriers Assessment, Transition Assessment, Task Analysis and Supporting Skills, and Placement and IEP Goals. The Milestones Assessment is the core feature of the VB-MAPP, from which the other four components are derived.

The Milestones Assessment contains 170 language and collateral milestones that are sequenced across three developmental levels. Level 1 consists of nine domains balanced across a scale of 1 to 5, which correspond with the language skills acquired by typically developing children from birth to 18 months of age. Level 2 consists of 12 domains balanced across a scale of 6 to 10, which correspond with skills acquired from 18 to 30 months of age. Embedded within Levels 1 and 2 is a subtest of verbal imitation called the Early Echoic Skills Assessment (EESA; Esch, 2008) that is used to rate the speaker's ability to echo a variety of speech phonemes, syllable combinations, and intonation patterns. Level 3 consists of 13 domains balanced across a scale of 11 to 15, which correspond with skills acquired from 30 to 48 months of age. For the present study, we were particularly interested in the data from Level 1 of the initial VB-MAPP administration that was collected at the onset of EIBI for each child.

All VB-MAPP assessments were completed in a clinical setting under the supervision of a Board Certified Behavior Analyst. Patients were either directly observed or tested for their ability to complete each of the skills included on the Milestones Assessment. On their initial intake assessment, patient scores ranged from Level 1 to Level 3 across the various domains of the VB-MAPP. Patient records were maintained in an online database that - subsequent to receiving IRB approval - was accessed to create a de-identified dataset for our analyses.

**Dataset**

Subsequent to retrieving and compiling the case records within this dataset from the initial administrations of the VB-MAPP Milestones Assessment for each patient, we formatted the data into 121 rows that consisted of the patient's age (at the time of assessment), sex, and the score recorded for each of the nine domains from Level 1: Mand, Tact, Listener Responding, Visual Perceptual Skills and Matching to Sample (VP-MTS), Independent Play, Social Behavior and Social Play, Motor Imitation, Spontaneous Vocal Behavior, and Echoic, as demonstrated by the EESA (Mason, 2022). The score for each domain ranged from 0 to 5 in half-point increments. According to Sundberg (2014), the scores across domains are approximately balanced, such that a Mand score of 5 is developmentally equivalent to a score of 5 in each of the remaining eight domains of Level 1. A typically developing child, therefore, would show similar scores across each domain. However, prior research on individuals with ASD has shown circumscribed learning patterns that may result in isolated behavioral excesses or deficits (Mayes et al., 2011; Ploog, 2010). In contrast to their typically developing peers, a high score in one domain of the VB-MAPP is not necessarily representative of high scores in the other domains.

At the start of the study, 121 patient records were available for review. Twenty-two additional case records became available over the course of the archival review, and these records were allocated to the holdout dataset. Thus, the entire dataset included 143 case records, wherein 22 of these were allocated to the holdout dataset.

**Variables**

Along with each patient's age and sex, we examined their scores on Level 1 of the VB-MAPP Milestones Assessment. The nine domains assessed within Level 1 are designed to correspond with the skills displayed by typically developing children up to 18 months of age. Five of these skills are directly related to functional language development. The Mand domain examines a person's ability to request access to preferred items and activities. The Tact domain examines a person's ability to label their surroundings. The Listener domain examines a person's ability to follow instructions. The Vocal domain examines a person's physical ability to produce speech sounds. Finally, the Echoic domain examines a person's ability to repeat what they hear. Echoing is frequently considered the earliest form of language; a prerequisite to other skills like requesting, labeling, and conversing (Sundberg, 2014).

Notably, the language skills assessed within Level 1 of the VB-MAPP are not mutually exclusive. For example, the first Mand and Tact milestones allow the use of echoic prompts. Similarly, the second Mand milestone allows the presence of the desired item (i.e., Tact). The interaction of these basic language skills has been well documented within the verbal behavior literature (Gamba et al., 2015; Mason & Andrews, 2019; Michael et al., 2011).

Sundberg (2014) describes the remaining four domains as language-related skills. The VP-MTS domain examines a person's visual ability to discriminate between similar features of the environment. The Play domain examines a person's ability to manipulate their environment independently, while the Social domain examines a person's ability to engage with other people. Finally, the Imitation domain examines a person's ability to imitate the actions of others.

The EESA subtest was used as the binary dependent variable for this investigation. Verbal imitation skills typically emerge by 11 months of age, and are frequently considered a prerequisite to more complex verbal behavior (Esch, 2008). The EESA is designed to evaluate a person's ability to echo the language models provided by another person, and is completed in full with each administration of the VB-MAPP.

**Experimental Preparations**

For the current explorative study, the presence or absence of an echoic repertoire served as the dependent variable, while the remaining eight domains of Level 1 of the VB-MAPP served as independent variables. Prior to conducting the analyses, we determined a threshold of four points on the EESA as the cut point for distinguishing between participants with and without an echoic repertoire. That is, participants who scored less than or equal to four on the EESA subtest were considered to have no echoic repertoire (0). Participants who scored 5 or above on the EESA subtest were classified as demonstrating an echoic repertoire (1).

Given that all of the domains of the VB-MAPP, including the EESA, are part of the same instrument, we were well-positioned to identify the degree to which a MLP can make such predictions even when employing a dataset composed of only 143 participant records. While the functional value of predicting a known outcome might be questioned, it is only possible to gauge the accuracy of any prediction technique when known outcomes are available. Thus, rather than trying to predict an unknown outcome, our experimental preparations were directed at gauging the MLP's ability to predict at high levels of precision, sensitivity, specificity, and overall prediction accuracy while accessing a limited dataset from one children's healthcare setting.

Interobserver agreement (IOA) was assessed by asking a second rater to independently re-code 49% (n = 59) of the records used in our analysis. The IOA sample was randomly selected by assigning each record a random number. Next, the dataset was sorted by random number in ascending order. We then renumbered the dataset sequentially from 1 to 118, and extracted the odd-numbered records for IOA. Of the 59 records that were re-coded, 55 (93%) were an exact match. For the remaining four records, point-by-point agreement on the nine VB-MAPP domains averaged 86% (range, 78% to 89%).

**Procedures**

We randomly selected and segregated 22 participant case records from the existing database to be analyzed as holdout data, and we employed the remaining 121 records to conduct and MLP followed by a ROC analyses.

**MLP Analysis.** To summarize several of our earlier details, we employed the five-layer MLP architecture shown in Figure 1. Extracting five separate training and validation sets from our sample, we conducted a series of five CV folds from our database (cf. Smith, 2005). During these trials, it became apparent that employing two hidden layers produced more accurate outcomes than could be obtained using three hidden layers. In particular, we found that two hidden layers where the number of hidden nodes per layer was set to 5 and 4 generated the highest level of predictive accuracy. In a similar way, we found using a Learning Rate set to 0. 01 produced the best performance outcomes.

This strategy allowed us to save a model consistent with the most functional weights and bias values and the number of layers and nodes within layers used during training. Then, during the holdout test, using all 121 training case records, we employed the sklearn.neural_network MLPClassifier to ascertain the effects of age and sex in conjunction with the ability to mand, tact, listen, demonstrate VP/MTS, play, interact socially, imitate, and vocalize on the probability that participants are able to acquire echoic behaviors.

Figure 5 shows the classification table depicting performance outcomes for the MLP analysis of the 22 participant records in the holdout dataset. The diagonal cells show the number of correctly identified cases. Predicted positives and predicted negatives are classified as 11 and 11, respectively. Off-diagonal cells show the number of cases classified incorrectly as 0 and 0. Precision is displayed at 1. 00, and sensitivity is 1. 00. Specificity is shown as 1. 00. Test Accuracy, 1. 00, is the overall proportion of correct class identifications when the holdout dataset was presented to the trained MLP model.

**Figure 5**

*Display of Predictive Classification Findings*

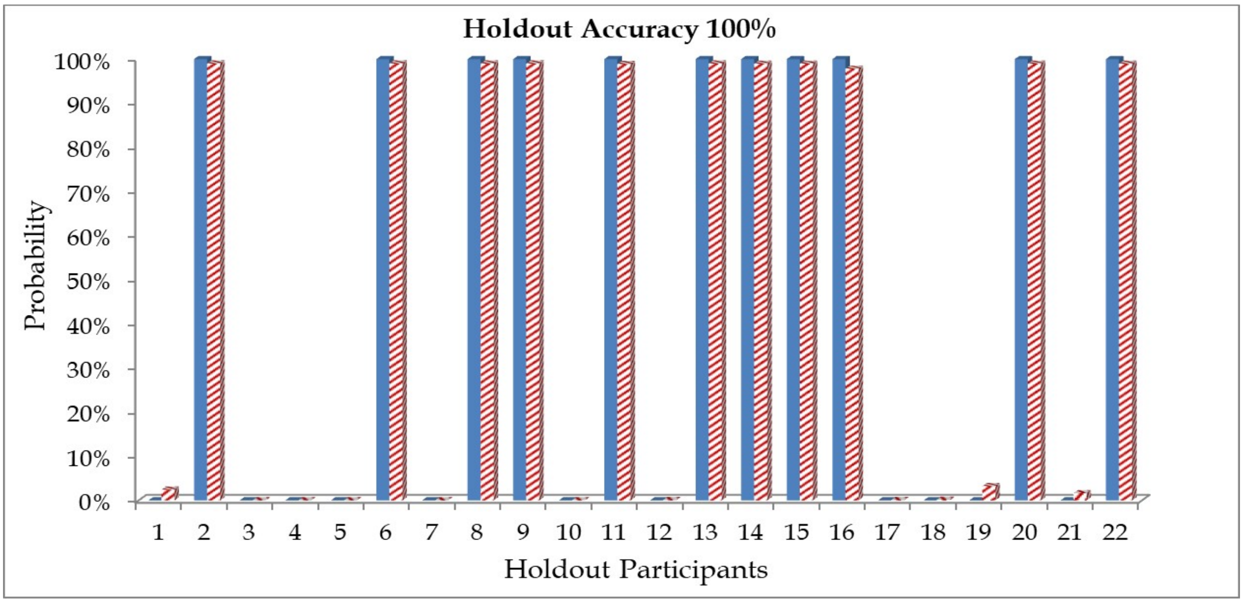
| | | | | |
|---|---|---|---|---|
| Precision: 1.00 | | | | F1 Score: 1.00 |
| | Actual Positives | Actural Negatives | Classification | Sensitivity |
| Predicted Positives | 11 | 0 | 11 | 1.00 |
| Predicted Negatives | 0 | 11 | 11 | Specificity 1.00 |
| | 11 | 11 | 22 | |
| Test Accuracy: 1.00 | | | | Kappa 1.00 |

*Note.* As viewed from our application C# Windows panel, the center of the figure shows a 2x2 contingency table displaying the predicted and actual positive and negative values. The sklearn.neural_network MLPClassifier shows the same outcomes.

Figure 6 shows predictive findings where the blue bars display the known participants' echoic performance levels. Hatched bars show the MLPs prediction of each participant's echoic performance. The trained MLP model accurately predicted the presence or absence of echoic behavior for 22 of 22 cases.

**Figure 6**

*Display of the MLP Model’s Predictive Findings*



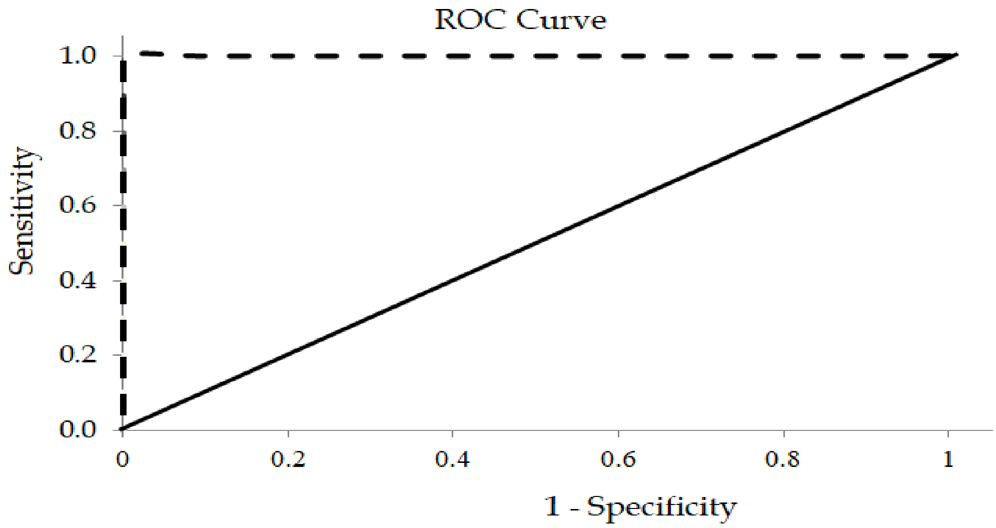

*Note.* Solid bars display the presence (1) or absence (0) of echoic verbal behavior for each participant. Hatched bars display MLP's prediction of the participant's ability to echo on a continuous scale of 0-1.

**Receiver Operator Characteristics (ROC) Analysis**

To further address binary classification thresholds, we provide an accuracy threshold continuum by way of the ROC curve. The ROC curve conveys binary classifiers' social validity and diagnostic ability in terms of the thresholds achieved at consecutive levels of sensitivity and specificity. Figure 7 illustrates a performance scale on a range of accuracy thresholds attained by the MLP.

**Figure 7**

*Display of the MLP Model's Predictive Findings*




Again, sensitivity identifies the proportion of accurately predicted positive cases or true positive rate, while specificity denotes the proportion of accurately predicted negative cases or true negative rate. By transposing specificity to 1 – specificity along the *x*-axis, the ROC curve illustrates the convergence of sensitivity and specificity as a single metric. The diagonal 45-degree line demarcates random classifiers.

**Utility and Social Validity.** A Type I error is an inferential statistical determination that occurs when the null hypothesis is erroneously rejected. That is, an experimental outcome is inconsistent with performances that would occur within the broader parent population. Such inaccurate statistical inferences fall under the more general heading of false positive errors. The ROC curve represents the true positive rate in conjunction with the false positive rate by means of a pooled accuracy analysis. Models with the fewest Type I and Type II error rates (the best-performing models) lay adjacent to the upper left corner of the graph, indicating that the model can accurately differentiate and predict class membership. Figure 7 illustrates that the

MLP model aligns with the *y*-axis and continues this constant alignment at 1. 0 along the path of the upper *x*-axis.

## Discussion

As it relates to the natural science of behavior, we concern ourselves with the extent to which ML can describe, predict and control various environmental relations. A primary purpose of the current study was to demonstrate the predictive ability of ML for identifying whether or not children with autism would develop echoic verbal behavior.

For the purposes of the current analysis, we included the information available to us from the initial intake results of the VB-MAPP, which included sex, age, and the nine domains of Level 1. Consequently, it is important to delineate the Mand, Tact, Listener, VP/MTS, Play, Social, Imitation, Echoic, and Vocal skills as measured by the VB-MAPP from the more general description for each of these operants. For example, many children with ASD demonstrate echolalia, which is emblematic of an echoic repertoire. Yet, the echoic ability of these same children may not register on the EESA subtest of the VB-MAPP. Therefore, we make it a point to distinguish the domains of the VB-MAPP, denoted with a capital letter (e.g., Mand), from the more general operant class, denoted with a lowercase letter (e.g., mand).

The results of the current study serve as one version of a phenotypic marker for vocal speech. By identifying the relevant collateral behavior associated with a functional Echoic repertoire, our findings provide a better understanding of human language and point to future research opportunities. Specifically, our results help to paint a broad picture of the environmental determinants of language development, and we urge caution when interpreting these outcomes within the clinical case context. Sidman's (1960) warning that "we have no way of evaluating whether or not a given example of group data actually does provide a true picture of individual behavioral processes" (p. 274) was reiterated by Skinner (1966) who cautioned, "A curve which enables us to predict the performance of another organism does not therefore represent a basic process" (p. 20).

First, we should acknowledge the limitations of our findings by reiterating that the participants in this study were 143 children diagnosed with ASD. Our findings say nothing about neurotypical language development, or the language development of individuals with other verbal behavior disorders. Additionally, the heterogeneity of ASD may be associated with differing basic repertoires that serve as either prerequisite or collateral supports for vocal language. Future researchers should consider replicating our findings with similar and diverse groups of participants.

Second, our findings are premised upon the structural hierarchy of skills outlined by the VB-MAPP. The VB-MAPP purports to be matched across domains within each level. That is, typically developing speakers who show Level 1 Echoic skills also tend to show Level 1 skills within the Mand, Imitation, Vocal, and other domains. To the extent that the developmental profiles within and across each of these domains remains valid, we can be confident in the accuracy of our predictions. Padilla and Akers (2021) found strong support for developmental age appropriateness of all but two of the Level 1 domains; the two exceptions, Independent Play and Social Behavior & Social Play, were found to have moderate support. This evidence suggests that the scores of the VB-MAPP provide information relevant to the target behaviors of interest. Nevertheless, our findings are limited to the use of the VB-MAPP, and are not inevitably representative of other verbal behavior measures. Future researchers should seek to replicate our findings with other language assessments.

Clinicians who use the VB-MAPP to monitor progress and as a curriculum guide may identify results from individual clients with ASD that align with the findings of the current study to varying degrees. A client that demonstrates Mand, Imitation, and Vocal skills, who also displays Echoics as measured by the EESA fits the hybrid predictive model perfectly. The same is true for a client that does not demonstrate Mand, Imitation, or Vocal skills, who also does not display an Echoic repertoire. Given that our study examined only the initial VB-MAPP scores gathered at intake, we cannot say whether the relationship between Mand/Imitation/Vocal skills and Echoics is correlational or causal. In the case of the latter client, future researchers could investigate the extent to which remediation of the Mand, Imitation, or Vocal domains leads to the development of an Echoic repertoire.

**Prediction with Small Samples**

This demonstration study sought to predict the likelihood that children with ASD would develop echoic verbal behavior. Using 143 case records of the VB-MAPP, we built predictive models based on MLP strategies. We then assessed the predictive ability using 22 holdout case records, and the techniques described in this study may contribute to a broader body of current investigations focusing on the development of improved network accuracy (Dubey et al., 2022). It should be noted that other machine learning algorithms may produce similar outcomes and require less computing power for small datasets (see Turgeon & Lanovaz, 2020, for a tutorial). Moreover, other popular ML techniques such as random forest, support vector classifier, stochastic gradient descent, and k-nearest neighbors are increasingly valuable and functional techniques particularly when classifying small-*n* datasets.

**Test Accuracy**

The test accuracy level (100%) obtained by this model is noteworthy, given the small sample employed to train our model. The potential for predicting skills measured by the VB-MAPP to children who will benefit from early intensive behavioral intervention is supported by this investigation, and this is an important step toward broader treatment efficacy. In this predictive analysis, the MLP model's high level of accuracy may help identify children with the repertoires necessary for emitting echoic behavior and guide treatment decisions for children who do not already demonstrate this skill. As a practical matter, these findings suggest that predictions based on MLP models can support decisions for more intensive training aimed at students with particular skill sets.

**Funding:** No type of funding or other compensation was provided for conducting any part of this investigation.

**Ethical Approval**: All procedures performed with respect to human participants were conducted in accordance with the ethical standards of the institutional research committee and in conformance with the 1964 Helsinki declaration and all of its subsequent amendments or similar ethical standards.

**Conflict of Interest**: On behalf of both authors, the corresponding author affirms that there are no conflicts of interest with regard to any aspect of this investigation.

**Data and Software Availability:** The archival dataset analyzed and described within this manuscript was retrieved from, and is downloadable to interested researchers at, the Texas Christian University Digital Repository:

https://repository.tcu.edu/handle/116099117/56144

Copies of the Python and C# software employed in this study can be downloaded from Behavioral Software Systems:

https://www.behavioralsoftwaresystems.com

**References**

Almeida, A., & Azkune, G. (2018). Predicting human behaviour with recurrent neural networks. *Applied Sciences*, *8*(2), 305. https://doi.org/10.3390/app8020305

Baghdadli, A., Assouline, B., Sonié, S., Pernon, E., Darrou, C., Michelon, C., Picot, M. C., Aussilloux, C., & Pry, R. (2011). Developmental trajectories of adaptive behaviors from early childhood to adolescence in a cohort of 152 children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*(7), 1314–1325. https://doi.org/10.1007/s10803-011-1357-z

Bengio, Y. (2000). Gradient-Based Optimization of Hyperparameters. *Neural Computation*, *12*(8), 1889–1900. https://doi.org/10.1162/089976600300015187

Bent, S., & Hendren, R. L. (2010). Improving the prediction of response to therapy in autism. *Neurotherapeutics, 7,* 232-240. https://doi.org/10.1016/j.nurt.2010.05.011

Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*, *20*(1). https://doi.org/10.1186/s12911-020-1023-5

Cividini-Motta, C., Scharrer, N., & Ahearn, W. H. (2016). An assessment of three procedures to teach echoic responding. *The Analysis of Verbal Behavior, 33*(1), 41–63. https://doi.org/10.1007/s40616-016-0069-z

Dafoe, A. (2018). *AI governance: A research agenda.* University of Oxford. Retrieved from https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf.

Dufour, M., Lanovaz, M. J., & Cardinal, P. (2020). Artificial intelligence for the measurement of vocal stereotypy. *Journal of the Experimental Analysis of Behavior*, *114*(3), 368–380. Portico. https://doi.org/10.1002/jeab.636

Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 503, 92–108. https://doi.org/10.1016/j.neucom.2022.06.111

Esch, B. E. (2008). Early echoic skills assessment. In M. L. Sundberg, *The verbal behavior milestones assessment and placement program* (pp. 62-63). Concord, CA: AVB Press.

Gamba, J., Goyos, C., & Petursdottir, A. I. (2015). The functional independence of mands and tacts: Has it been demonstrated empirically? *The Analysis of Verbal Behavior, 31*(1), 10-38. https://doi.org/10.1007%2Fs40616-014-0026-7

Goldstein, H. (2002).communication intervention for children with autism: A review of treatment efficacy. *Journal of Autism and Developmental Disorders, 32*(5), 373-396. https://doi.org/10.1023/a:1020589821992

Greene, M. N., Morgan, P. H., & Foxall, G. R. (2017). NEURAL Networks and consumer behavior: NEURAL models, logistic regression, and the behavioral perspective model. *The Behavior Analyst*, *40*, 393–418. https://doi.org/10.1007/s40614-017-0105-x.

Hagan, M. T., Demuth, H. B., Beale, M. H., & De Jesús, O. (2002). *Neural network design.* PWS Publishing Company. https://hagan.okstate.edu/NNDesign.pdf

Haykin, S. (2009). *Neural network and learning machines* (3rd ed.). Prentice Hall. https://doi.org/10.1007/s10278-012-9556-5

Ingvarsson, E. T. (2016). Tutorial: Teaching verbal behavior to children with ASD. International Electronic Journal of Elementary Education, 9(2), 433-450. https://files.eric.ed.gov/fulltext/EJ1126669.pdf

Islam, M. A., Wimmer, H., & Rebman, C. M. (2021). Examining Sigmoid vs ReLu Activation Functions in Deep Learning. *Interdisciplinary Research in Technology and Management*, 432–437. https://doi.org/10.1201/9781003202240-68

Jiménez, Á. B., Lázaro, J. L., & Dorronsoro, J. R. (2007). Finding Optimal Model Parameters by Discrete Grid Search. Innovations in Hybrid Intelligent Systems, 120–127. https://doi.org/10.1007/978-3-540-74972-117

Kissinger, H. A., Schmidt, E., & Huttenlocher, D. (2021). *The age of AI: And our human future.* Hachette UK.

Kohli, M., Kar, A. K., Bangalore, A., & A P, P. (2022). Machine learning-based ABA treatment recommendation and personalization for autism spectrum disorder: An exploratory study. *Brain Informatics, 9*(1): 16. https://doi.org/10.1186/s40708-022-00164-6

Linstead, E., Dixon, D. R., French, R., Granpeesheh, D., Adams, H., German, R., Powell, A., Stevens, E., Tarbox, J., & Kornack, J. (2016). Intensity and learning outcomes in the treatment of children with autism spectrum disorder. *Behavior Modification, 41*(2), 229–252. https://doi.org/10.1177/0145445516667059

Lyddy, F., & Barnes-Holmes, D. (2007). Stimulus equivalence as a function of training protocol in a connectionist network. *Journal of Speech & Language Pathology & Applied Behavior Analysis*, *2*, 14–24. https://doi.org/10.1037/h0100204.

Maharjan, J., Garikipati, A., Dinenno, F. A., Ciobanu, M., Barnes, G., Browning, E., DeCurzio, J., Mao, Q., & Das, R. (2023). Machine learning determination of applied behavioral analysis treatment plan type. *Brain Informatics, 10*(1). https://doi.org/10.1186/s40708-023-00186-8

Mason, L. (2022). *The VB-MAPP Level 1 Milestones Assessment profiles of children with autism spectrum disorder.* Texas Christian University Library Data Management. https://doi.org/10.18776/tcu/data/56144

Mason, L. L., & Andrews, A. (2019). The verbal behavior stimulus control ratio equation: A quantification of language. *Perspectives on Behavior Science, 42*(2), 323-343. https://doi.org/10.1007%2Fs40614-018-0141-1

Mayes, S. D., Calhoun, S. L., Murray, M. J., Morrow, J. D., Cothren, S., Purichia, H., Yurich, K. K. L., & Bouder, J. N. (2011). Use of Gilliam Asperger’s Disorder Scale in differentiating high and low functioning autism and ADHD. *Psychological Reports, 108*(1), 3–13. https://doi.org/10.2466/04.10.15.pr0.108.1.3-13

McCaffrey, J. (2023). *Binary Classification Using a scikit MLPClassifier Neural Network.* Retrieved from https://jamesmccaffrey.wordpress.com/2023/05/05/binary-classification-using-a-scikit-mlpclassifier-neural-network/

Michael, J., Palmer, D. C., & Sundberg, M. L. (2011). The multiple control of verbal behavior. *The Analysis of Verbal Behavior, 27*(1), 3-22. https://doi.org/10.1007%2FBF03393089

Nielsen, M. A. (2015). *Neural networks and deep learning* (Vol. 25). San Francisco, CA, USA: Determination Press.

Ninness, C., Lauter, J., Coffee, M., Clary, L., Kelly, E., Rumph, M., et al. (2012). Behavioral and biological neural network analyses: A common pathway toward pattern recognition and prediction. *The Psychological Record*, *62*, 579–598. https://doi.org/10.5210/bsi.v22i0.4450.

Ninness, C., Ninness, S. K., Rumph, M., & Lawson, D. (2018). The emergence of stimulus relations: Human and computer learning. *Perspective on Behavioral Science*, *41*, 121-154. https://doi.org/10.1007/s40614-017-0125-6

Ninness, C., Rehfeldt, R. A. & Ninness, S. (2019). Identifying accurate and inaccurate stimulus relations: Human and computer learning. *The Psychological Record*, *69*(3), 333–356. https://doi.org/10.1007/s40732-019-00337-6

Ninness, C., Newton, R., Saxon, J., Rumph, R., Bradfield, A., Harrison, C., Vasquez, E., & Ninness, S. K. (2002a). Small Group Statistics: A Monte Carlo Comparison of Parametric and Randomization Tests. *Behavior and Social Issues*, *12*(1), 53–63. https://doi.org/10.5210/bsi.v12i1.79

Ninness, C., Rumph, R., Vasquez, E., Bradfield, A., & Ninness, S. K. (2002b). Multivariate Randomization Tests for Small-N Behavioral Research: A Web-Based Application. *Behavior and Social Issues*, *12*(1), 64–74. https://doi.org/10.5210/bsi.v12i1.80

Ninness, C. & Ninness**,** S. K**. (**2020**a).** Emergent virtual analytics: Modeling contextual control of derived stimulus relations. *Behavior and Social Issues*, *29*(1), 119–137. doi: https://doi.org/10.1007/s42822-020-00032-0

Ninness, C. & Ninness, S. K. (2020b). Emergent virtual analytics: Artificial intelligence and human-computer interactions. *Behavior and Social Issues*, *29*(1), 100–118. https://doi.org/10.1007/s42822-020-00031-1

Ninness, C., Yelick, A., Ninness, S. K., & Cordova, W. (2021). Predicting heuristic decisions in child welfare: A neural network exploration. *Behavior and Social Issues*, *30*(1), 194–208. https://doi.org/10.1007/s42822-021-00047-1

Palmer, D. C. (2012). The role of atomic repertoires in complex behavior. The Behavior Analyst, 35(1), 59–73. https://doi.org/10.1007/bf03392266

Özener, O., Yüksek, L., & Özkan, M. (2013). Artificial neural network approach to predicting engine-out emissions and performance parameters of a turbo charged diesel engine. *Thermal Science*, *17*(1), 153–166. https://doi.org/10.2298/tsci120321220o

Probst, P., Wright, M. N., & Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, *9*(3). Portico. https://doi.org/10.1002/widm.1301

Rivard, M., & Forget, J. (2012). Verbal behavior in young children with autism spectrum disorders at the onset of an early behavioral intervention program. *The Psychological Record, 62,* 165-186. https://doi.org/10.1007/BF03395796

Rose, V., Trembath, D., Keen, D., & Paynter, J. (2016). The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *Journal of Intellectual Disability Research, 60*(5), 464-477. https://doi.org/10.1111/jir.12284

Smith, P., & Hayes, S. C. (2022). An open-source relational network derivation script in R for modeling and visualizing complex behavior for scientists and practitioners. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022. 914485

Smith, J. T., (2005). Neural network verification. In B. J. Taylor (Ed.), *Methods and Procedures for the Verification and Validation of Artificial Neural Networks*. Springer.

Stevens, J. P. (2001). *Applied multivariate statistics for the social sciences* (4th ed.). Psychology Press. https://doi.org/10.4324/9781410604491

Tegmark, M. (2018). *Life 3. 0*. Penguin Books.

Tovar, Á. E., & Westermann, G. (2017). A neurocomputational approach to trained and transitive relations in equivalence classes. *Frontiers in Psychology*, *8*, 1–15. https://doi.org/10.3389/fpsyg.2017.01848.

Turgeon, S., & Lanovaz, M. J. (2020). Tutorial: Applying Machine Learning in Behavioral Research. *Perspectives on Behavior Science*, *43*(4), 697–723 https://doi.org/10.1007/s40614-020-00270-y

Phan, N., Dou, D., Wang, H., Kil, D., & Piniewski, B. (2017). Ontology-based deep learning for human behavior prediction with explanations in health social networks. *Information Sciences*, *384*, 298–313. https://doi.org/10.1016/j.ins.2016.08.038

Ploog, B. O. (2010). Stimulus overselectivity four decades later: A review of the literature and its implications for current research in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 40*(11), 1332-1349. https://doi.org/10.1007/s10803-010-0990-2

Préfontaine, I., Lanovaz, M. J., & Rivard, M. (2022). Brief report: Machine learning for estimating prognosis of children with autism receiving early behavioral intervention—A proof of concept. *Journal of Autism and Developmental Disorders,* 1-6. https://doi.org/10.1007/s10803-022-05641-9

Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal on Mental Retardation, 110*(6), 417-438. https://doi.org/10.1352/0895-8017(2005)110[417:ibtfcw]2.0.co;2

Sundberg, M. L. (2014). *The verbal behavior milestones assessment and placement program: The VB-MAPP* (2nd ed.). Concord, CA: AVB Press.

Trembath, D., & Vivanti, G. (2014). Problematic but predictive: Individual differences in

children with autism spectrum disorders. *International Journal of Speech-Language Pathology, 16*(1), 57-60. https://doi.org/10.3109/17549507.2013.859300

Victoria, A. H., & Maragatham, G. (2020). Automatic tuning of hyperparameters using Bayesian optimization. *Evolving Systems*, *12*(1), 217–223. https://doi.org/10.1007/s12530-020-09345-2

Wenliang, L. K., & Seitz, A. R. (2018). Deep neural networks for modeling visual perceptual learning. *Journal of Neuroscience*, *38*, 6028–6044. https://doi.org/10.1523/jneurosci. 1620-17.2018

Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, *48*(9), 2839-2846. https://doi.org/10.1016/j.patcog.2015.03.009

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice.

*Neurocomputing*, *415*, 295–316.

https://doi.org/10.1016/j.neucom.2020.07.061