Deep autoencoder-based behavioral pattern recognition outperforms standard statistical methods in high-dimensional zebrafish studies

Adrian J. Green[1,3*], Lisa Truong[4], Preethi Thunga[2,3], Connor Leong[4], Melody Hancock[1,3], Robyn L. Tanguay[4], David M. Reif [1,3]

[1]Department of Biological Sciences, [2]Department of Statistics, and [3]the Bioinformatics Research Center, NC State University, Raleigh, North Carolina, United States of America; [4]Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, Oregon, United States of America

Corresponding Author:

E-mail Address: ajgreen4@ncsu.edu (AJG)

Contents

Abstract

Zebrafish have become an essential tool in screening for developmental neurotoxic chemicals and their molecular targets. The success of zebrafish as a screening model is partially due to their physical characteristics including their relatively simple nervous system, rapid development, experimental tractability, and genetic diversity combined with technical advantages that allow for the generation of large amounts of high-dimensional behavioral data. These data are complex and require advanced machine learning and statistical techniques to comprehensively analyze and capture spatiotemporal responses. To accomplish this goal, we have trained semi-supervised deep autoencoders using behavior data from unexposed larval zebrafish to extract quintessential "normal" behavior. Following training, our network was evaluated using data from larvae shown to have significant changes in behavior (using a traditional statistical framework) following exposure to toxicants that include nanomaterials, aromatics, per- and polyfluoroalkyl substances (PFAS), and other environmental contaminants. Further, our model identified new chemicals (Perfluoro-n-octadecanoic acid, 8-Chloroperfluorooctylphosphonic acid, and Nonafluoropentanamide) as capable of inducing abnormal behavior at multiple chemical-concentrations pairs not captured using distance moved alone. Leveraging this deep learning model will allow for better characterization of the different exposure-induced behavioral phenotypes, facilitate improved genetic and neurobehavioral analysis in mechanistic determination studies and provide a robust framework for analyzing complex behaviors found in higher-order model systems.

Author Summary

We demonstrate that a deep autoencoder using raw behavioral tracking data from zebrafish toxicity screens outperforms conventional statistical methods, resulting in a comprehensive evaluation of behavioral data. Our models can accurately distinguish between normal and abnormal behavior with near-complete overlap with existing statistical approaches, with many chemicals detectable at lower concentrations than with conventional statistical tests; this is a crucial finding for the protection of public health. Our deep learning models enable the identification of new substances capable of inducing aberrant behavior, and we generated new data to demonstrate the reproducibility of these results. Thus, neurodevelopmentally active chemicals identified by our deep autoencoder models may represent previously undetectable signals of subtle individual response differences. Our method elegantly accounts for the high degree of behavioral variability associated with the genetic diversity found in a highly outbred population, as is typical for zebrafish research, thereby making it applicable to multiple laboratories. Utilizing the vast quantities of control data generated during high-throughput screening is one of the most innovative aspects of this study and to our knowledge is the first study to explicitly develop a deep autoencoder model for anomaly detection in large-scale toxicological behavior studies.

Introduction

Significant progress continues to be made in our understanding of neurodevelopmental disorders such as autism spectrum disorder, attention deficit hyperactivity disorder (ADHD), developmental delay, learning disabilities, and other neurodevelopmental problems. As incidences continue to rise globally and affect 10-15% of all births, more work must be done to improve our understanding of these disorders (Boyle et al., 2011; *Neurodevelopmental Diseases*, 2021; US EPA, 2015b). Meta-analyses suggest strong and consistent epidemiological evidence that the developing nervous system is particularly vulnerable to low-level exposure to widespread environmental contaminants, as the anatomical and functional architecture of the human brain is mainly determined by developmental transcriptional processes during the prenatal period (Grandjean & Landrigan, 2014; Green & Planchart, 2018; Miller et al., 2014; Rock & Patisaul, 2018; US EPA, 2015b). Therefore, identifying associations between developmental exposures and neurological effects is a core objective to improve public health by informing disease and disability prevention (*A Blueprint for Brain Development*, 2014; *Neurodevelopmental Diseases*, 2021).

As the number of environmental contaminants grows to nearly one million, comprehensive data on the neurodevelopmental toxicity of these contaminants remain sparse or nonexistent (Krewski et al., 2020; US EPA, 2015a, 2015b; Wambaugh et al., 2013). In response, high-throughput screening (HTS) assays have been developed to expedite chemical toxicity testing using *in vitro* and *in vivo* systems (Judson et al., 2010; Richard et al., 2016; Truong et al., 2014). However, *in vitro* cell and cell-free assays cannot fully capture systemic organismal responses in terms of anatomy, physiology, or behavior (Thomas et al., 2012). Zebrafish (*Danio rerio*) have emerged as an ideal model for studying low-level chemical exposure because of their high fecundity, rapid development, genetic tractability, and amenability to high-throughput data generation (Bugel et al., 2014; Planchart et al., 2018; Truong et al., 2014). The zebrafish brain's structural organization, cellular morphology, and neurotransmitter systems are very similar to other vertebrates, including chickens, rats, and humans (Horzmann & Freeman, 2016; Kalueff et al., 2014; Lowery & Sive, 2004; Tropepe & Sive, 2003). Furthermore, zebrafish have behavioral patterns highly similar to mammals, and genetic homologs for 70% of human genes and 82% of human disease genes, making them a powerful tool for revealing the neuronal developmental pathways underlying behavior (Basnet et al., 2019; Howe et al., 2013; Postlethwait et al., 1998).

Zebrafish larvae show mature swimming patterns following swim bladder development at four to five days post-fertilization (dpf), which can be assessed using various locomotor behavioral assays (Hernandez et al., 2018; Tegelenbosch et al., 2012). One of these assays, the larval photomotor response (LPR), utilizes a sudden transition from light to dark, eliciting a stereotyped large-angle O-bend, followed by several minutes of increased movement, which gradually reduces (Burgess & Granato, 2007c; Emran et al., 2008). Exposure to toxicants has been shown to alter this stereotypical behavioral response (Basnet et al., 2019; Truong et al., 2016). Current HTS for behavioral neurotoxicity focuses heavily on analyzing locomotor behavior using distance moved and population-based statistical methods (Basnet et al., 2019; G. Zhang et al., 2017). However, while the behavior repertoire of larval zebrafish is less sophisticated when compared to that of adult zebrafish and other higher-order vertebrates, they are capable of numerous distinct behaviors (Basnet et al., 2019; Kalueff et al., 2013; Mirat et al., 2013). These behaviors, such as thigmotaxis, and light avoidance cannot always be captured when using distance moved as a sole indicator of neurobehavioral toxicity in analyses of this data. Moreover, as most laboratory zebrafish populations feature significant genetic heterogeneity, individual responses to exotic toxicants cannot be expected to be homogeneous for simplistic measures such as distance moved (Balik-Meisner et al., 2018).

Improved accessibility to computing resources and application interfaces, together with recent advances in deep-learning makes it possible to analyze complex behavioral data in novel ways and predict neurodevelopmental toxicity (Arifoglu & Bouchachia, 2017; Pereira et al., 2020; Xia et al., 2018). The volume and diversity of data generated during HTS experiments, combined with the variety in toxicological response within populations, present an opportunity that is well-suited for machine learning (ML). In particular, analysis of zebrafish HTS data from five dpf larvae exposed to 1,060 unique chemicals reveals that only 8% of chemical-concentration pairs (a unique combination of chemical and concentration, e.g. 6.4 µM Nicotine) exhibit changes in distance moved (G. Zhang et al., 2017), which is alarmingly low given the known toxicity profiles of the chemical set. This challenge provides an opportunity to apply methods developed for anomaly detection from areas such as financial fraud (Awoyemi et al., 2017), medical application faults (Pachauri & Sharma, 2015), security systems intrusion (Sargolzaei et al., 2016), system faults (Warriach & Tei, 2013), and others (Fazai et al., 2019; Jaiswal & Ruskin,

2019). In anomaly detection, we learn the pattern of a normal process, and anything that does not follow this pattern is classified as an anomaly. This learning model is particularly applicable, as many HTS data sets have large amounts of control data to analyze (G. Zhang et al., 2017). One intriguing approach to achieving this is by applying an autoencoder (Feng et al., 2021; Frassek et al., 2021; Goodfellow et al., 2016; Le Borgne et al., 2022; Nicholaus et al., 2021; Ranjan et al., 2019). An autoencoder is a neural network of two modules, an encoder and a decoder (Goodfellow et al., 2016; Gupta & Singh, 2019). The encoder learns the underlying features of a process, and these features are typically in a reduced dimension. The decoder then uses this reduced dimension to recreate the original data from these underlying features.

In the present study, we trained deep autoencoder models to recognize the pattern of quintessential larval zebrafish behavior and identify abnormal behavior following developmental chemical exposure. The performance of our deep autoencoders was compared against traditional statistical methodologies, the gold standard for behavioral assessment. In addition to model development, we assessed the features driving performance through feature permutation and generated new confirmatory data to assess model reproducibility and confirm novel findings.

Results

Statistical classification of behavior

After classifying each of the 96-well plates by differences in the movement of controls into hyperactive, normal, or hypoactive, we compared treated vs control behavioral response to light/dark cycling in zebrafish larvae at five dpf. We identified 39 chemical-concentration combinations from ten chemicals capable of inducing a significantly different ($p < 0.05$) behavioral response (Supp. Table 2). Using the 30^{th} and 70^{th} percentiles, we defined 227 individual larvae as abnormal (Fig 1a). These 227 larvae formed the validation set used to test the performance of our models.

Training performance

Autoencoder models were trained using only control data for each of the activity states (hypoactive, normal, and hyperactive) per phase of the second light cycle. This resulted in six trained models (Supp Fig 1 the training loss plots for the models). Table 1 shows the results for the six deep autoencoder models trained using control data and validated using data from zebrafish defined as abnormal using the K-S test. All the models performed well with values ranging from 0.615 – 0.867 and 0.740 – 0.922 for the Kappa and AUROC, respectively. As expected, the models consistently produced high specificity (SP) levels as this value indicated how well the models classify control data. There was greater variability in the sensitivity (SE) with the dark phase models matching or outperforming the light phase models for each activity state. Further, we observed a noteworthy trend across all models producing high positive predictive value (PPV). Overall, these results show that deep autoencoders trained using control data is capable of distinguishing between normal and abnormal larval zebrafish behavior with a high degree of accuracy.

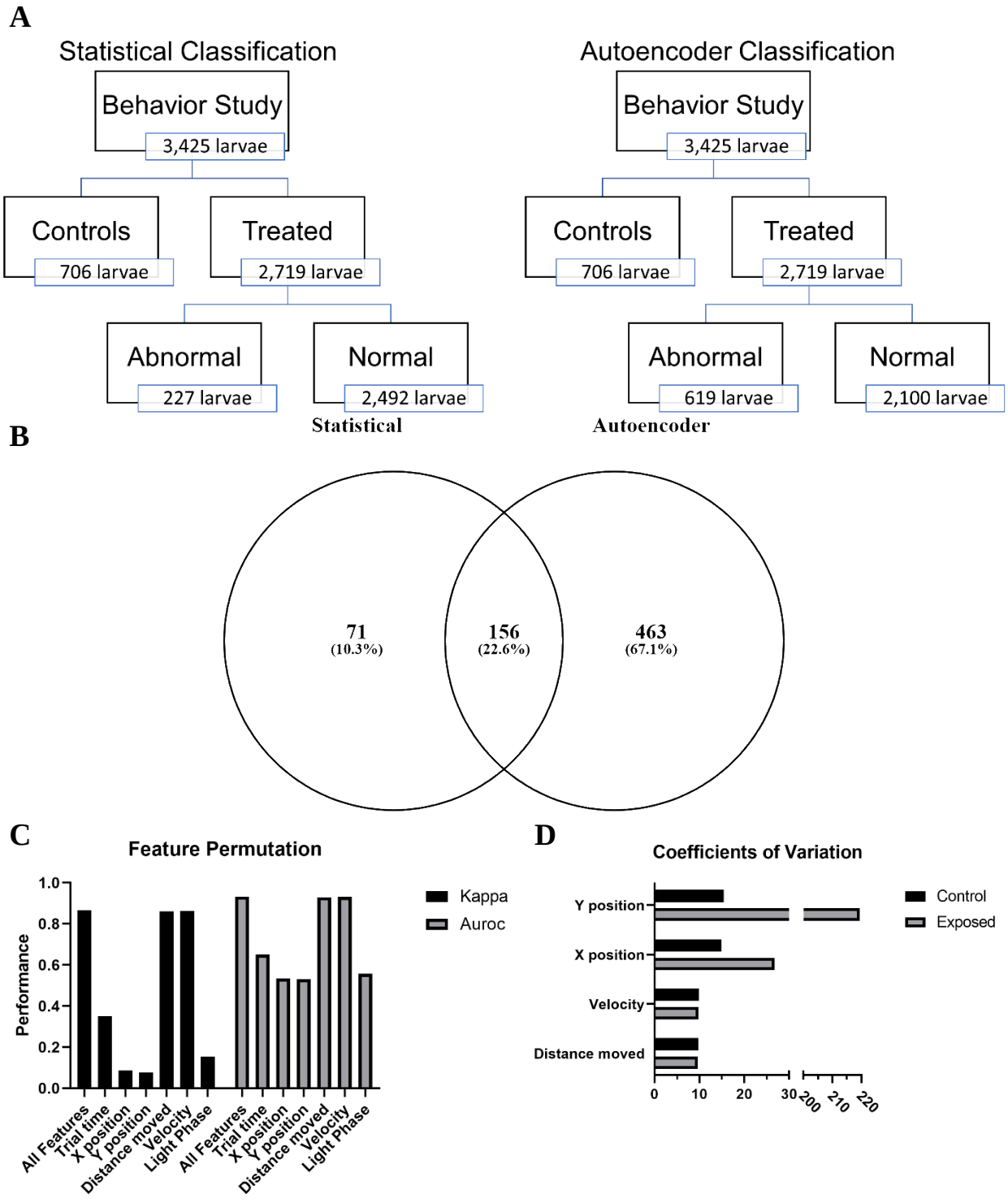

Figure 1: Assessment autoencoder performance. (A) Schematic representation of the differences in statistical and autoencoder based classification of behavioral response in larval zebrafish. (B) Venn diagram showing overlap between statistical and autoencoder classified abnormal zebrafish. (C) Evaluating the change in model performance when the values of a single feature are randomly shuffled. Kappa – Cohen's Kappa statistic, AUROC - area under the receiver operating characteristic. Figure depicts means ± SEM. (D) Coefficients of variation for each of the main numerical features.

185

Table 1. Deep autoencoder model performance in behavioral classification. Table showing performance of model trained using different activity states of the control data in both light and dark phases.

Model		Performance Metrics				
Baseline Control Activity Level	**Light Phase**	**SE**	**SP**	**PPV**	**Kappa**	**AUROC**
Hypoactive	Light	78.5	100	99.7	0.867	0.892
	Dark	78.3	98.0	88.4	0.800	0.882
Normal	Light	48.3	99.7	93.1	0.615	0.740
	Dark	73.3	94.8	77.6	0.695	0.840
Hyperactive	Light	79.2	97.5	85.5	0.790	0.883
	Dark	86.9	97.5	90.2	0.855	0.922

Evaluation of unknowns

Using the six trained models, we evaluated the 2,719 treated zebrafish larvae (Fig 1). The autoencoders correctly classified 156 of the 227 larvae that fell below or above the 30^{th} and 70^{th} percentiles, respectively. In addition, our deep autoencoders identified 463 larvae as abnormal from the 2,492 larvae defined as normal using the K-S test (Fig 1b). The majority (422) of these 619 larvae were from one of 66 chemical-concentration combinations from 13 chemicals (Table 2). The deep autoencoders successfully identified nine of the ten statistically abnormal chemicals and identified these chemicals at or below the lowest concentration shown to be statistically significant. While the deep autoencoders did not identify Perfluorodecylphosphonic acid as capable of inducing abnormal behavior, but they did identify 3-Perfluoropentyl propanoic acid (5:3), Perfluoro-n-octadecanoic acid, 8-Chloroperfluorooctylphosphonic acid, and Nonafluoropentanamide, which were missed in the statistical testing framework. These results, summarized in fig 2, show that deep autoencoders can match the performance of the K-S test and are more sensitive at detecting abnormal behavior.

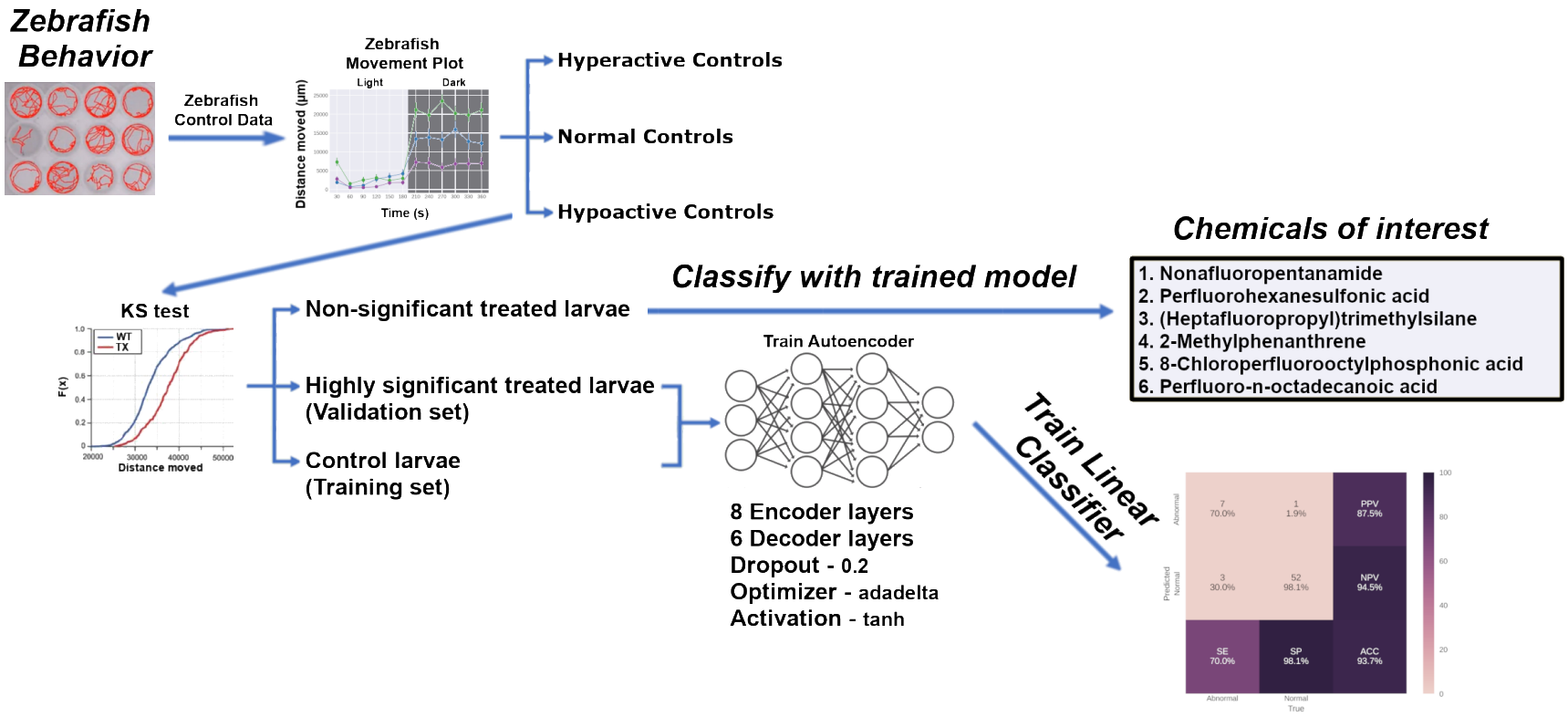

Figure 2: Summary of behavioral analysis pipeline and results. Utilizing our analysis pipeline produced six deep autoencoder models (three for the light phase and three for the dark phase) capable of classifying larval zebrafish behavior with high Kappa and AUROC values. The trained models were then used to classify the non-significant exposed larvae and identified Nonafluoropentanamide, Perfluorohexanesulfonic acid, (Heptafluoropropyl)trimethylsilane, 2-Methylphenanthrene, 8-Chloroperfluorooctylphosphonic acid, Perfluoro-n-octadecanoic acid, and others as capable of inducing abnormal behavior.

Table 2. Autoencoders identified chemicals. Table showing chemicals and concentrations flagged for displaying abnormal behavioral when evaluated using Autoencoder. Compounds that were picked up by Autoencoder, but not KS test are highlighted in red.

CASRN	Chemical Name	Concentration (µM)
71751-41-2	Abamectin	0.1, 0.2, 0.4, 0.6
308068-56-6	Multi-Walled Carbon Nanotube	10, 23.2, 50, 75, 100
2531-84-2	2-Methylphenanthrene	1, 2.54, 6.45, 16.4, 35, 74.8, 100
832-69-9	1-Methylphenanthrene	1, 2.54, 6.45, 16.4, 35, 74.8, 100
914637-49-3	3-Perfluoropentyl propanoic acid (5:3)	0.25
192-51-8	Dibenzo[e-l]pyrene	0.01, 0.025, 0.065, 0.164, 0.35, 0.75, 1, 2.54, 16.4, 35, 100
16517-11-6	Perfluoro-n-octadecanoic acid	0.25
355-46-4	Perfluorohexanesulfonic acid	0.015, 0.14, 0.41, 3.7, 11.1, 33.3, 66.5, 100
3834-42-2	(Heptafluoropropyl)trimethylsilane	0.015, 0.046, 0.41, 1.23, 11.1, 33.3
	8-Chloroperfluorooctylphosphonic acid	0.167
31253-34-6	2-Aminohexafluoropropan-2-ol	0.015, 0.046, 0.41, 1.23, 3.7, 11.1, 33.3, 66.5, 100
13485-61-5	Nonafluoropentanamide	0.41, 3.7, 11.1
439-14-5	Diazepam	1, 3, 5, 8, 12

Features driving improved autoencoder performance

To determine the features in the model that were most important in driving classification performance, we employed permutation feature importance. This technique is a model agnostic inspection technique used for any fitted estimator to determine the importance of each feature in the model. Larger the decrease in model performance (Kappa or AUROC) when a single feature value is randomly shuffled, the more important the feature. Our results, shown in fig 1c, indicate that phase, trial time, x position, and y position are the largest drivers of model performance, while distance moved and velocity contribute very little. Coefficients of variation show greater variability in the x and y positional data between control and exposed groups compared to either velocity or distance moved (fig 1d). This trend is consistent irrespective of the larval activity state (hypoactive, normal activity, or hyperactive) relative to their respective controls (Fig 3).

221

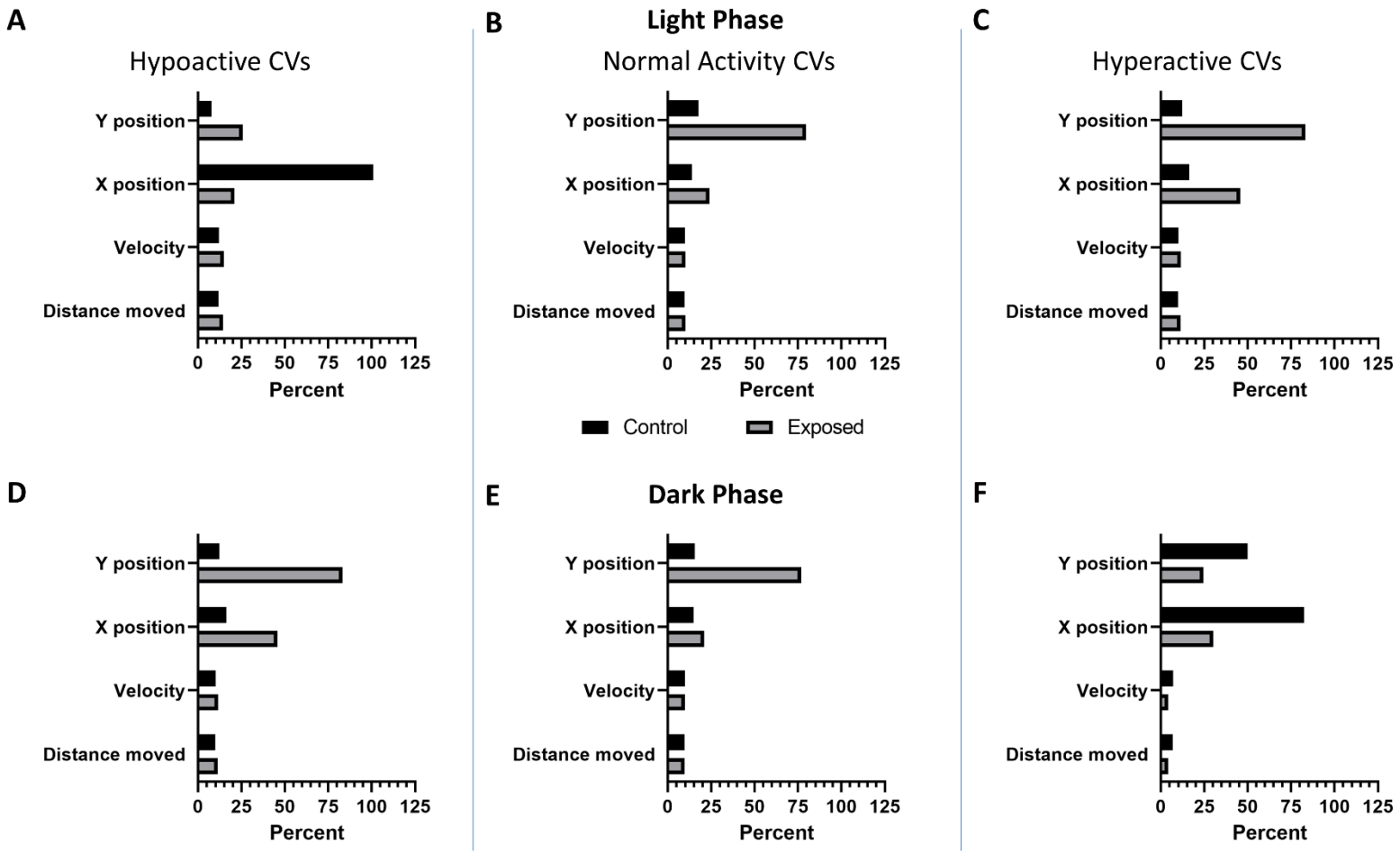

Figure 3: Coefficients of variation per larval activity state. Coefficients of variation (CVs) for each of the main numerical features (A – C) in the light (D – F) and in the dark. Columns show CVs of larval zebrafish significantly ($p < 0.05$) (A, D) hypoactive, (B, E) normal activity, or (C, F) hyperactive relative to their respective controls.

222

Experimental confirmation of autoencoder findings

To provide an unbiased evaluation of the final model fits, we generated new data using 2-Methylphenanthrene, and Nonafluoropentanamide. The data collected confirmed that our models accurately classified all controls as normal while detecting similar levels of abnormal behavior response across the concentration range (Fig 4). These results show that the trained model is capable of producing similar results across experimental replicates.

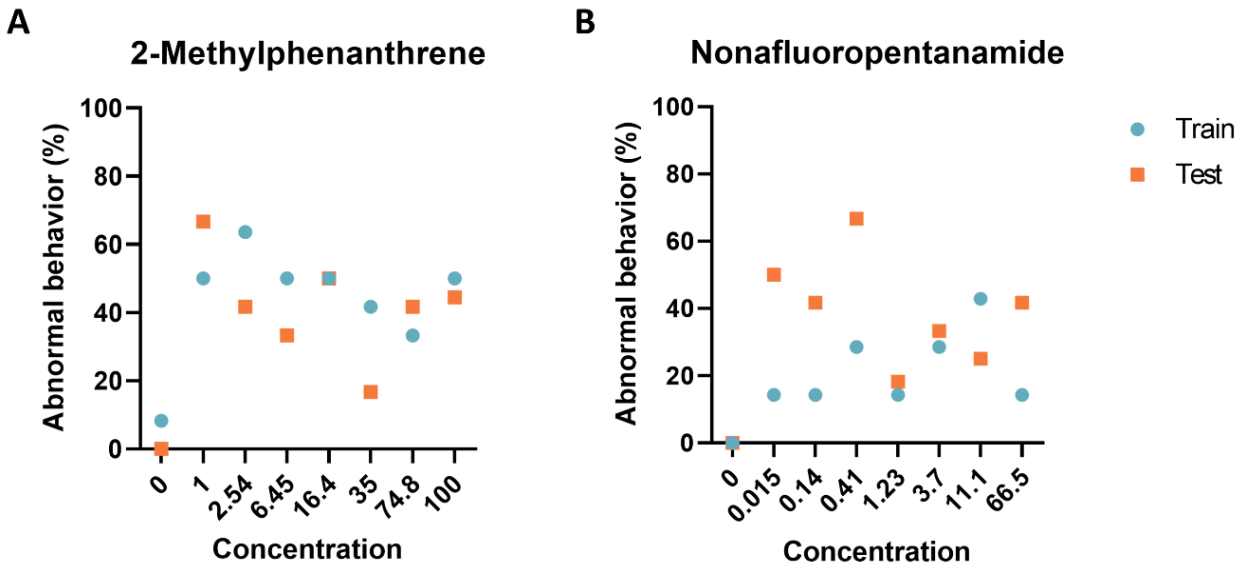

Figure 4: Experimental model evaluation. Comparison of the performance of deep autoencoder models between the training set and two chemicals identified by the models to elicit abnormal larval zebrafish behavior. Percent of larval zebrafish classified as abnormal based on their behavioral response to developmental exposure to (A) 2-Methylphenanthrene and (B) Nonafluoropentanamide

Discussion

Statistical analysis identified 39 chemical-concentration combinations from ten chemicals capable of inducing a significantly different ($p < 0.05$) behavioral response. Utilizing the 227 abnormal individuals identified by the statistical test as our validation set, we trained six deep autoencoder models using control data for each of the activity states (hypoactive, normal, and hyperactive). All of the resulting models performed well with values ranging from 0.615 – 0.867 and 0.740 – 0.922 for the Kappa and AUROC, respectively. All models achieved SP values above 94.8% and PPV values above 77.6% while SE values for all dark phase models outperformed the light phase models for each activity state (Table 1). Assessment of permutation feature importance indicates that phase, trial time, x-position, and y-position are the largest drivers of model performance (fig 1c). The calculated coefficients of variation shed some light on this surprising finding (fig 1d). They show that variation in the x and y positional data is greater than observed for velocity or distance moved between control and exposed groups. These differences in variation likely make it easier for the models to distinguish between treated and exposed groups.

When we examined exposed larvae defined as normal using the K-S test (Fig 1), our deep autoencoders identified 66 chemical-concentration combinations from 12 chemicals (Table 2) with Perfluoro-n-octadecanoic acid, 8-Chloroperfluorooctylphosphonic acid, and Nonafluoropentanamide only identified by our autoencoders. These results show that a deep autoencoder-based model can classify larval zebrafish behavior as normal or abnormal with very good efficacy and often identified abnormal behaviors at lower concentrations than current statistical methods. Further, the models identified three novel chemicals, Perfluoro-n-octadecanoic acid, 8-Chloroperfluorooctylphosphonic acid, and Nonafluoropentanamide as capable of inducing abnormal behavior (Fig 3).

Recognition and categorization of swimming patterns in larvae is a challenging task and a number of approaches have been used. These can range from subjective analysis based on experienced observations (Fero et al., 2011; Kalueff et al., 2013, p. 0) or more recently through the application of unsupervised ML (Budick & O'Malley, 2000; Burgess & Granato, 2007a, 2007b, 2007c; Kimmel et al., 1974; Mirat et al., 2013; H. Zhang et al., 2013). These studies have focused on the analysis and categorization of behavioral patterns in wild-type strains (Burgess &

Granato, 2007c; H. Zhang et al., 2013), mutant strains (Burgess & Granato, 2007b; Mirat et al., 2013), or larvae exposed to neuroactive chemicals (Mirat et al., 2013) but do not classify behavior as normal or abnormal. In addition, these unsupervised approaches have utilized highspeed camera systems which are medium to low throughput and have limited potential in the screening of tens of thousands of chemicals for behavioral effects. As introduced above, classification of behavior is one of the primary goals of toxicological screening and tends to result in highly imbalanced datasets and lend themselves to anomaly detection methodologies. While these methods are common in manufacturing (Fan et al., 2018; Fazai et al., 2019; Jaiswal & Ruskin, 2019; Nicholaus et al., 2021), information systems (Pachauri & Sharma, 2015; Warriach & Tei, 2013), security systems (Feng et al., 2021; Sargolzaei et al., 2016), and financial fraud (Awoyemi et al., 2017) they have only very recently been applied to biological data (Frassek et al., 2021; Homayouni et al., 2021; Nwokedi et al., 2021). To the best of our knowledge, this is the first study to explicitly develop a deep autoencoder model for anomaly detection in toxicological behavior studies.

Overall, our results show that a deep autoencoder utilizing raw behavioral tracking data from five dpf zebrafish larvae can accurately distinguish between normal and abnormal behavior. We show that these results are reproducible and allow for the identification of new compounds capable of eliciting abnormal behavior. Further, our models were able to identify abnormal behavior following chemical exposure at lower concentrations than with traditional statistical tests. Our approach accounts for the high degree of behavioral variability associated with the genetic diversity found within a highly outbred population typical of zebrafish studies, thereby making it extensible to use across labs. Looking to the future, neurodevelopmentally active chemicals identified using our deep autoencoder models may represent heretofore undetectable signals of subtle differences in individual responses, suggesting chemicals that should be investigated further as eliciting differential population responses (i.e. interindividual susceptibility differences).

These findings will facilitate the application of behavioral characterization methods discussed above, such as Zebrazoom (Mirat et al., 2013), using highspeed cameras to identify the behavioral traits most perturbed by the chemical exposure and allow for more mechanistic discovery. One of the key innovations presented in this study is leveraging vast amounts of

290 control data generated as part of any high-throughput screening (HTS) – setting the stage for
291 predictive toxicological applications and safety assessments for the enormous backlog of as-yet
292 untested chemicals.

293

Materials and methods

This section describes the autoencoder models utilizing a semi-supervised ML algorithm and logistic regression (LR) to discriminate between normal and abnormal behavior in chemically exposed five dpf zebrafish. An overview of our approach is shown in Fig 3. Briefly, we created and trained six autoencoder models for each phase of the assay; namely, hyperactive, normal, and hypoactive depending on the control movement in the light or dark phases of the assay. Finally, treated plates were tested on one of these, depending on which category, its controls fell under. We used experimental data collected on a large and diverse compound set of 30 chemicals including an insecticide, nanomaterial, perfluorinated chemicals, and aromatic pollutants at a range of concentrations (133 chemical-concentration pairs) to assess the neurotoxic effects of these chemicals following developmental exposure (Supp. Table 1).

Zebrafish husbandry

Tropical 5D wild-type zebrafish were housed at Oregon State University's Sinnhuber Aquatic Research Laboratory (SARL, Corvallis, OR) in densities of 1000 fish per 100-gallon tank according to the Institutional Animal Care and Use Committee protocols (Barton et al., 2016). Fish were maintained at 28 °C on a 14:10 h light/dark cycle in recirculating filtered water, supplemented with Instant Ocean salts. Adult, larval and juvenile fish were fed with size-appropriate GEMMA Micro food 2–3 times a day (Skretting). Spawning funnels were placed in the tanks the night prior, and the following morning, embryos were collected and staged (Kimmel et al., 1995; Westerfield, 2007). Embryos were maintained in embryo medium (EM) in an incubator at 28 °C until further processing. EM consisted of 15 mM NaCl, 0.5 mM KCl, 1 mM $MgSO_4$, 0.15 mM KH_2PO_4, 0.05 mM Na_2HPO_4, and 0.7 mM $NaHCO_3$ (Westerfield, 2007).

Developmental chemical exposure

The empirical data used to develop our model were gathered as described in Truong et al. and Noyes et al.(Noyes et al., 2015; Truong et al., 2014, 2022). The experimental design consisted of the 30 unique chemicals tested (Supp Table 1) with at least 7 replicates (an individual embryo in singular wells of a 96-well plate) at each concentration for each chemical.

Developmental toxicity assessments

Mortality and morphology

At 24 hours post-fertilization (hpf), embryos were screened for mortality and 2 developmental endpoints. At 120 hpf, mortality and incidence of abnormality in 9 morphology endpoints were evaluated as binary outcomes. Any individuals identified with a physical abnormality were excluded from any behavioral analysis as these abnormalities might confound the results.

Photomotor responses

The larval photomotor response (LPR) assay was conducted at 120 hpf when the 96-well plates of larvae were placed into a Zebrabox (Viewpoint LifeSciences) and larval movement was recorded. The recorded videos were then tracked with Ethovision XT v.11 analysis software for 24 min across 3 cycles of 3 min light: 3 min dark. The trial time(s), x-position, y-position, distance moved (µm), and velocity (mm/s) by each larva in the 2nd light/dark cycle were the features used for behavioral assessment (Supp Fig 2). The 2nd light/dark cycle was chosen as it exhibited less noise than the 1st cycle and was less influenced by any learning that might have occurred in the 3rd cycle. For all assessments, data were collected from embryos exposed to nominal concentrations of chemical and uploaded under a unique well-plate identifier into a custom LIMS (Zebrafish Acquisition and Analysis Program [ZAAP]) – a MySQL database and analyzed using custom R scripts that were executed in the LIMS background (Truong et al., 2016).

Data preprocessing and statistical analysis pipeline

Preprocessing

All data processing, statistical analysis and ML were implemented in Python using the open source libraries Tensorflow (Martín Abadi et al., 2015), Keras (U.S. Environmental Protection Agency, 2021), Scikit-learn (Pedregosa et al., 2011), Pandas (McKinney, 2010), and Numpy (Harris et al., 2020) within a purpose build Singularity container environment (Sylabs.io, 2019). The x-position and y-position data was standardized relative to the center of each well and

forward filled if datapoints were missing. Outliers were normalized to the maximum likely distance a zebrafish larva could move in 1/25th of a second. Considering that the average length of a 5 dpf larval zebrafish is 3.9 mm and can move about 2.5 times it's body length during a startle response (120 frames at 1000 frames/second) the threshold for distance moved in our system was set at 3.25 mm per frame (Burgess & Granato, 2007b; *ZFIN Zebrafish Developmental Stages*, n.d.). This resulted in 5,445 of the 30,825,000 frames being normalized.

Statistical analysis

Interexperimental zebrafish larval response to light/dark cycling is highly variable (Supp Fig 2). Therefore, a two sample Kolmogorov–Smirnov test (K-S test) was used to compare mean of controls from individual 96-well plates to mean control movement across all plates. The K-S test is a non-parametric two-sided test and no adjustments were made for normality or multiple comparisions. Controls from individual plates with statistically significant ($p < 0.01$) differences in movement compared to the average of all controls were grouped together as hyperactive, normal, or hypoactive. Following grouping the K-S test was used to compare each chemical-concentration combination with their respective same plate control ($p < 0.05$). Individuals in the 30th and 70th percentiles of each chemical-concentration combination were defined as abnormal.

Autoencoder architecture

Deep autoencoders were developed using zebrafish control data to distinguish between normal and abnormal zebrafish behavior. The model was trained on a Dell R740 containing two Intel Xeon processors with 18 cores per processor, 512 GB RAM, and a Tesla-V100-PCIE (31.7 GB). The autoencoders consisted of an input and output layer of fixed-size based on the size of a single phase (25 frames per 180s) of the second light cycle (4500 frames by 5 features). The encoder network was composed of eight fully connected hidden layers using a normal kernel initialization, tanh activation, a dropout value of 0.2, L1 and L2 regularization values of 1e^{-05}, and an adadelta optimizer. The size of each hidden layer was reduced by increasing multiples of 15 and resulted in a compressed representation (bottleneck) size of 250. The decoder network was composed of six fully connected hidden layers using tanh activation, and a dropout value of 0.2. All hidden layers used an adadelta optimizer (learning_rate=0.001, rho=0.95, and

epsilon=1e-07) and mean squared error for the loss function (He et al., 2015; Osl et al., 2012; Ramachandran et al., 2017). For each model, we optimized the hyperparameters (i.e., the number of hidden layers, the number of nodes in the layers, loss functions, optimizers, regularization rates, and dropout rates) by grid search technique trained on all control data over 500 epochs using Cohens Kappa statistic as the objective metric. The final encoder models were trained over the course of 125000 epochs. The resulting compressed representation was used as input into a logistic regression layer trained using a 100 fold cross-validation with each fold consisting of 4000 epochs using a limited-memory BFGS solver. The code and sample training data that implements the models are available at GitHub [https://github.com/Tanguay-Lab/Manuscripts/tree/main/Green_et_al_(2023)_Manuscript]. A complete dataset is available apon request.

Network performance and evaluation

The data showed strong normal vs abnormal class imbalance (Fig 1). Classifiers may be biased towards the major class (normal) and therefore, show poor performance accuracy for the minor class (abnormal) (Lemaître et al., 2017). Normal vs abnormal classification accuracy was evaluated using a confusion matrix, Cohen's Kappa statistic, and area under the receiver operating characteristic (AUROC) as Kappa and AUROC measure model accuracy, while compensating for simple chance (Ben-David, 2008). The primary metrics we used from the confusion matrix included sensitivity (SE), specificity (SP), and positive predictive value (PPV) as these parameters give us the true positive rate, true negative rate, and the proportion of true positives amongst all positive calls (Parikh et al., 2008; Pearson, 1904; Townsend, 1971). Chemical-concentration combinations were defined as abnormal if the autoencoders identified more individual as abnormal in the exposed than their respective controls and at least 25% of the individuals were abnormal. Permutation feature importance was used to evaluate which features are the most important for model performance. In brief, one feature (variable) is shuffled randomly and all features are fed into the model the resulting Kappa and AUROC values are calculated. This is repeated 1000 times per feature and average Kappa and AUROC are calculated across each shuffle (Breiman, 2001). To determine why one feature might be more

important than another a coefficient of variation was calculated for each of the features in the control and exposed groups (variation() in the Scipy package).

Acknowledgments

This research was supported by the National Institutes of Health, through the National Institute of Environmental and Health Sciences (P30 ES030287, R56 ES030007, P30 ES025128) and the National Cancer Institute (R01 CA161608). We would like to thank the staff at Sinnhuber Aquatic Research Laboratory, and John Lam for his contribution to reprocessing videos.

References

A Blueprint for Brain Development. (2014, April 8). NIH Director's Blog. https://directorsblog.nih.gov/2014/04/08/a-blueprint-for-brain-development/

Arifoglu, D., & Bouchachia, A. (2017). Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks. *Procedia Computer Science*, *110*, 86–93. https://doi.org/10.1016/j.procs.2017.06.121

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. *2017 International Conference on Computing Networking and Informatics (ICCNI)*, 1–9. https://doi.org/10.1109/ICCNI.2017.8123782

Balik-Meisner, M., Truong, L., Scholl, E. H., La Du, J. K., Tanguay, R. L., & Reif, D. M. (2018). Elucidating Gene-by-Environment Interactions Associated with Differential Susceptibility to Chemical Exposure. *Environmental Health Perspectives*, *126*(06). https://doi.org/10.1289/EHP2662

Barton, C. L., Johnson, E. W., & Tanguay, R. L. (2016). Facility Design and Health Management Program at the Sinnhuber Aquatic Research Laboratory. *Zebrafish*, *13*(S1), S-39-S-43. https://doi.org/10.1089/zeb.2015.1232

Basnet, R. M., Zizioli, D., Taweedet, S., Finazzi, D., & Memo, M. (2019). Zebrafish Larvae as a Behavioral Model in Neuropharmacology. *Biomedicines*, *7*(1), 23. https://doi.org/10.3390/biomedicines7010023

Ben-David, A. (2008). About the relationship between ROC curves and Cohen's kappa. *Engineering Applications of Artificial Intelligence*, *21*(6), 874–882. https://doi.org/10.1016/j.engappai.2007.09.009

436 Boyle, C. A., Boulet, S., Schieve, L. A., Cohen, R. A., Blumberg, S. J., Yeargin-Allsopp, M.,
437 Visser, S., & Kogan, M. D. (2011). Trends in the prevalence of developmental disabilities
438 in US children, 1997-2008. *Pediatrics*, *127*(6), 1034–1042.
439 https://doi.org/10.1542/peds.2010-2989
440 Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.
441 https://doi.org/10.1023/A:1010933404324
442 Budick, S. A., & O'Malley, D. M. (2000). Locomotor repertoire of the larval zebrafish:
443 Swimming, turning and prey capture. *Journal of Experimental Biology*, *203*(17), 2565–
444 2579. https://doi.org/10.1242/jeb.203.17.2565
445 Bugel, S. M., Tanguay, R. L., & Planchart, A. (2014). Zebrafish: A Marvel of High-Throughput
446 Biology for 21st Century Toxicology. *Current Environmental Health Reports*, *1*(4), 341–
447 352. https://doi.org/10.1007/s40572-014-0029-5
448 Burgess, H. A., & Granato, M. (2007a). *Flote v2.1: Biological Tracking Software*.
449 Burgess, H. A., & Granato, M. (2007b). Sensorimotor Gating in Larval Zebrafish. *Journal of*
450 *Neuroscience*, *27*(18), 4984–4994. https://doi.org/10.1523/JNEUROSCI.0615-07.2007
451 Burgess, H. A., & Granato, M. (2007c). Modulation of locomotor activity in larval zebrafish
452 during light adaptation. *Journal of Experimental Biology*, *210*(14), 2526–2539.
453 https://doi.org/10.1242/jeb.003939
454 Emran, F., Rihel, J., & Dowling, J. E. (2008). A behavioral assay to measure responsiveness of
455 zebrafish to changes in light intensities. *Journal of Visualized Experiments: JoVE*, *20*.
456 https://doi.org/10.3791/923

457 Fan, C., Xiao, F., Zhao, Y., & Wang, J. (2018). Analytical investigation of autoencoder-based
458 methods for unsupervised anomaly detection in building energy data. *Applied Energy*,
459 *211*, 1123–1135. https://doi.org/10.1016/j.apenergy.2017.12.005
460 Fazai, R., Abodayeh, K., Mansouri, M., Trabelsi, M., Nounou, H., Nounou, M., & Georghiou, G.
461 E. (2019). Machine learning-based statistical testing hypothesis for fault detection in
462 photovoltaic systems. *Solar Energy*, *190*, 405–413.
463 https://doi.org/10.1016/j.solener.2019.08.032
464 Feng, J., Liang, Y., & Li, L. (2021). Anomaly Detection in Videos Using Two-Stream
465 Autoencoder with Post Hoc Interpretability. *Computational Intelligence and*
466 *Neuroscience*, *2021*, 7367870. https://doi.org/10.1155/2021/7367870
467 Fero, K., Yokogawa, T., & Burgess, H. A. (2011). The Behavioral Repertoire of Larval
468 Zebrafish. In A. V. Kalueff & J. M. Cachat (Eds.), *Zebrafish Models in Neurobehavioral*
469 *Research* (pp. 249–291). Humana Press. https://doi.org/10.1007/978-1-60761-922-2_12
470 Frassek, M., Arjun, A., & Bolhuis, P. G. (2021). An extended autoencoder model for reaction
471 coordinate discovery in rare event molecular dynamics datasets. *The Journal of Chemical*
472 *Physics*, *155*(6), 064103. https://doi.org/10.1063/5.0058639
473 Goodfellow, I., Bengio, Y., & Courville, A. (2016). Chapter 14—Autoencoders. In *Deep*
474 *Learning* (pp. 499–523). MIT Press.
475 Grandjean, P., & Landrigan, P. J. (2014). Neurobehavioural effects of developmental toxicity.
476 *The Lancet Neurology*, *13*(3), 330–338. https://doi.org/10.1016/S1474-4422(13)70278-3
477 Green, A. J., & Planchart, A. (2018). The neurological toxicity of heavy metals: A fish
478 perspective. *Comparative Biochemistry and Physiology. Toxicology & Pharmacology:*
479 *CBP*, *208*, 12–19. https://doi.org/10.1016/j.cbpc.2017.11.008

Gupta, A., & Singh, S. (2019, June 25). ML | Classifying Data using an Auto-encoder. *GeeksforGeeks*. https://www.geeksforgeeks.org/ml-classifying-data-using-an-auto-encoder/

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1026–1034. https://doi.org/10.1109/ICCV.2015.123

Hernandez, R. E., Galitan, L., Cameron, J., Goodwin, N., & Ramakrishnan, L. (2018). Delay of Initial Feeding of Zebrafish Larvae Until 8 Days Postfertilization Has No Impact on Survival or Growth Through the Juvenile Stage. *Zebrafish*, *15*(5), 515–518. https://doi.org/10.1089/zeb.2018.1579

Homayouni, H., Ray, I., Ghosh, S., Gondalia, S., & Kahn, M. G. (2021). Anomaly Detection in COVID-19 Time-Series Data. *SN Computer Science*, *2*(4), 279. https://doi.org/10.1007/s42979-021-00658-w

Horzmann, K. A., & Freeman, J. L. (2016). Zebrafish Get Connected: Investigating Neurotransmission Targets and Alterations in Chemical Toxicity. *Toxics*, *4*(3), 19. https://doi.org/10.3390/toxics4030019

Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M.,

503 Churcher, C., Scott, C., Barrett, J. C., Koch, R., Rauch, G.-J., White, S., … Stemple, D.
504 L. (2013). The zebrafish reference genome sequence and its relationship to the human
505 genome. *Nature*, *496*(7446), 498–503. https://doi.org/10.1038/nature12111
506 Jaiswal, V., & Ruskin, A. (2019, April 26). *Mooring Line Failure Detection Using Machine*
507 *Learning*. Offshore Technology Conference. https://doi.org/10.4043/29511-MS
508 Judson, R. S., Houck, K. A., Kavlock, R. J., Knudsen, T. B., Martin, M. T., Mortensen, H. M.,
509 Reif, D. M., Rotroff, D. M., Shah, I., Richard, A. M., & Dix, D. J. (2010). In vitro
510 screening of environmental chemicals for targeted testing prioritization: The ToxCast
511 project. *Environmental Health Perspectives*, *118*(4), 485–492.
512 https://doi.org/10.1289/ehp.0901392
513 Kalueff, A. V., Gebhardt, M., Stewart, A. M., Cachat, J. M., Brimmer, M., Chawla, J. S.,
514 Craddock, C., Kyzar, E. J., Roth, A., Landsman, S., Gaikwad, S., Robinson, K., Baatrup,
515 E., Tierney, K., Shamchuk, A., Norton, W., Miller, N., Nicolson, T., Braubach, O., …
516 Schneider, H. (2013). Towards a Comprehensive Catalog of Zebrafish Behavior 1.0 and
517 Beyond. *Zebrafish*, *10*(1), 70–86. https://doi.org/10.1089/zeb.2012.0861
518 Kalueff, A. V., Stewart, A. M., & Gerlai, R. (2014). Zebrafish as an emerging model for
519 studying complex brain disorders. *Trends in Pharmacological Sciences*, *35*(2), 63–75.
520 https://doi.org/10.1016/j.tips.2013.12.002
521 Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., & Schilling, T. F. (1995). Stages of
522 embryonic development of the zebrafish. *Developmental Dynamics: An Official*
523 *Publication of the American Association of Anatomists*, *203*(3), 253–310.
524 https://doi.org/10.1002/aja.1002030302

Kimmel, C. B., Patterson, J., & Kimmel, R. O. (1974). The development and behavioral characteristics of the startle response in the zebra fish. *Developmental Psychobiology*, *7*(1), 47–60. https://doi.org/10.1002/dev.420070109

Krewski, D., Andersen, M. E., Tyshenko, M. G., Krishnan, K., Hartung, T., Boekelheide, K., Wambaugh, J. F., Jones, D., Whelan, M., Thomas, R., Yauk, C., Barton-Maclaren, T., & Cote, I. (2020). Toxicity testing in the 21st century: Progress in the past decade and future perspectives. *Archives of Toxicology*, *94*(1), 1–58. https://doi.org/10.1007/s00204-019-02613-4

Le Borgne, Y.-A., Siblini, W., Lebichot, B., & Bontempi, G. (2022). Autoencoders and anomaly detection—Reproducible Machine Learning for Credit Card Fraud detection—Practical handbook. In *Reproducible Machine Learning for Credit Card Fraud Detection—Practical Handbook*. Université Libre de Bruxelles. https://github.com/Fraud-Detection-Handbook/fraud-detection-handbook

Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*, *18*(17), 1–5.

Lowery, L. A., & Sive, H. (2004). Strategies of vertebrate neurulation and a re-evaluation of teleost neural tube formation. *Mechanisms of Development*, *121*(10), 1189–1197. https://doi.org/10.1016/j.mod.2004.04.022

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz,

Lukasz Kaiser, Manjunath Kudlur, … Xiaoqiang Zheng. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. https://www.tensorflow.org/

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Miller, J. A., Ding, S.-L., Sunkin, S. M., Smith, K. A., Ng, L., Szafer, A., Ebbert, A., Riley, Z. L., Royall, J. J., Aiona, K., Arnold, J. M., Bennet, C., Bertagnolli, D., Brouner, K., Butler, S., Caldejon, S., Carey, A., Cuhaciyan, C., Dalley, R. A., … Lein, E. S. (2014). Transcriptional landscape of the prenatal human brain. *Nature, 508*(7495), 199–206. https://doi.org/10.1038/nature13185

Mirat, O., Sternberg, J. R., Severi, K. E., & Wyart, C. (2013). ZebraZoom: An automated program for high-throughput behavioral analysis and categorization. *Frontiers in Neural Circuits*, *7*. https://doi.org/10.3389/fncir.2013.00107

Neurodevelopmental Diseases. (2021, January 12). National Institute of Environmental Health Sciences. https://www.niehs.nih.gov/research/supported/health/neurodevelopmental/index.cfm

Nicholaus, I. T., Park, J. R., Jung, K., Lee, J. S., & Kang, D.-K. (2021). Anomaly Detection of Water Level Using Deep Autoencoder. *Sensors (Basel, Switzerland), 21*(19), 6679. https://doi.org/10.3390/s21196679

Noyes, P. D., Haggard, D. E., Gonnerman, G. D., & Tanguay, R. L. (2015). Advanced Morphological—Behavioral Test Platform Reveals Neurodevelopmental Defects in Embryonic Zebrafish Exposed to Comprehensive Suite of Halogenated and Organophosphate Flame Retardants. *Toxicological Sciences, 145*(1), 177–195. https://doi.org/10.1093/toxsci/kfv044

570 Nwokedi, E. I., Bains, R., Bidaut, L., Wells, S., Ye, X., & Brown, J. M. (2021). Unsupervised
571 detection of mouse behavioural anomalies using two-stream convolutional autoencoders.
572 *ArXiv*.
573 Osl, M., Netzer, M., Dreiseitl, S., & Baumgartner, C. (2012). Applied Data Mining: From
574 Biomarker Discovery to Decision Support Systems. In Z. Trajanoski (Ed.),
575 *Computational Medicine* (pp. 173–184). Springer Vienna. https://doi.org/10.1007/978-3-
576 7091-0947-2_10
577 Pachauri, G., & Sharma, S. (2015). Anomaly Detection in Medical Wireless Sensor Networks
578 using Machine Learning Algorithms. *Procedia Computer Science*, *70*, 325–333.
579 https://doi.org/10.1016/j.procs.2015.10.026
580 Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., & Thomas, R. (2008). Understanding and
581 using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*,
582 *56*(1), 45–50.
583 Pearson, K. (1904). On the theory of contingency and its relation to association and normal
584 correlation. In *Drapers Company Research Memoirs*. Dulau and Co.
585 https://archive.org/details/cu31924003064833/page/n1/mode/2up
586 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,
587 Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D.,
588 Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in
589 Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
590 Pereira, T. D., Shaevitz, J. W., & Murthy, M. (2020). Quantifying behavior to understand the
591 brain. *Nature Neuroscience*, *23*(12), 1537–1549. https://doi.org/10.1038/s41593-020-
592 00734-z

593 Planchart, A., Green, A. J., Hoyo, C., & Mattingly, C. J. (2018). Heavy Metal Exposure and
594 Metabolic Syndrome: Evidence from Human and Model System Studies. *Current*
595 *Environmental Health Reports*, *5*(1), 110–124. https://doi.org/10.1007/s40572-018-0182-
596 3
597 Postlethwait, J. H., Yan, Y. L., Gates, M. A., Horne, S., Amores, A., Brownlie, A., Donovan, A.,
598 Egan, E. S., Force, A., Gong, Z., Goutel, C., Fritz, A., Kelsh, R., Knapik, E., Liao, E.,
599 Paw, B., Ransom, D., Singer, A., Thomson, M., … Talbot, W. S. (1998). Vertebrate
600 genome evolution and the zebrafish gene map. *Nature Genetics*, *18*(4), 345–349.
601 https://doi.org/10.1038/ng0498-345
602 Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for Activation Functions.
603 *ArXiv:1710.05941 [Cs]*. http://arxiv.org/abs/1710.05941
604 Ranjan, C., Reddy, M., Mustonen, M., Paynabar, K., & Pourak, K. (2019). Dataset: Rare Event
605 Classification in Multivariate Time Series. *ArXiv:1809.10717 [Cs, Stat]*.
606 http://arxiv.org/abs/1809.10717
607 Richard, A. M., Judson, R. S., Houck, K. A., Grulke, C. M., Volarath, P., Thillainadarajah, I.,
608 Yang, C., Rathman, J., Martin, M. T., Wambaugh, J. F., Knudsen, T. B., Kancherla, J.,
609 Mansouri, K., Patlewicz, G., Williams, A. J., Little, S. B., Crofton, K. M., & Thomas, R.
610 S. (2016). ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology.
611 *Chemical Research in Toxicology*, *29*(8), 1225–1251.
612 https://doi.org/10.1021/acs.chemrestox.6b00135
613 Rock, K. D., & Patisaul, H. B. (2018). Environmental Mechanisms of Neurodevelopmental
614 Toxicity. *Current Environmental Health Reports*, *5*(1), 145–157.
615 https://doi.org/10.1007/s40572-018-0185-0

Sargolzaei, A., Crane, C. D., Abbaspour, A., & Noei, S. (2016). A Machine Learning Approach for Fault Detection in Vehicular Cyber-Physical Systems. *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 636–640. https://doi.org/10.1109/ICMLA.2016.0112

Sylabs.io. (2019). *Singularity* (3.5.2). Sylabs.io. https://sylabs.io/singularity/

Tegelenbosch, R. A. J., Noldus, L. P. J. J., Richardson, M. K., & Ahmad, F. (2012). Zebrafish embryos and larvae in behavioural assays. *Behaviour*, *149*(10–12), 1241–1281. https://doi.org/10.1163/1568539X-00003020

Thomas, R. S., Black, M. B., Li, L., Healy, E., Chu, T.-M., Bao, W., Andersen, M. E., & Wolfinger, R. D. (2012). A Comprehensive Statistical Analysis of Predicting In Vivo Hazard Using High-Throughput In Vitro Screening. *Toxicological Sciences*, *128*(2), 398–417. https://doi.org/10.1093/toxsci/kfs159

Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, *9*(1), 40–50. https://doi.org/10.3758/BF03213026

Tropepe, V., & Sive, H. L. (2003). Can zebrafish be used as a model to study the neurodevelopmental causes of autism? *Genes, Brain and Behavior*, *2*(5), 268–281. https://doi.org/10.1034/j.1601-183X.2003.00038.x

Truong, L., Bugel, S. M., Chlebowski, A., Usenko, C. Y., Simonich, M. T., Simonich, S. L. M., & Tanguay, R. L. (2016). Optimizing multi-dimensional high throughput screening using zebrafish. *Reproductive Toxicology*, *65*, 139–147. https://doi.org/10.1016/j.reprotox.2016.05.015

637 Truong, L., Reif, D. M., St Mary, L., Geier, M. C., Truong, H. D., & Tanguay, R. L. (2014).
638 Multidimensional In Vivo Hazard Assessment Using Zebrafish. *Toxicological Sciences*,
639 *137*(1), 212–233. https://doi.org/10.1093/toxsci/kft235
640 Truong, L., Rericha, Y., Thunga, P., Marvel, S., Wallis, D., Simonich, M. T., Field, J. A., Cao,
641 D., Reif, D. M., & Tanguay, R. L. (2022). Systematic developmental toxicity assessment
642 of a structurally diverse library of PFAS in zebrafish. *Journal of Hazardous Materials*,
643 *431*, 128615. https://doi.org/10.1016/j.jhazmat.2022.128615
644 U.S. Environmental Protection Agency. (2021, August 10). *Comptox Chemicals Dashboard:*
645 *Master List of PFAS Substances (Version2)*.
646 https://comptox.epa.gov/dashboard/chemical-lists/pfasmaster
647 US EPA, O. (2015a, March 2). *About the TSCA Chemical Substance Inventory* [Overviews and
648 Factsheets]. US EPA. https://www.epa.gov/tsca-inventory/about-tsca-chemical-
649 substance-inventory
650 US EPA, O. (2015b, June 10). *Health: Neurodevelopmental Disorders – Report Contents*
651 [Reports and Assessments]. US EPA.
652 https://www.epa.gov/americaschildrenenvironment/health-neurodevelopmental-
653 disorders-report-contents
654 Wambaugh, J. F., Setzer, R. W., Reif, D. M., Gangwal, S., Mitchell-Blackwood, J., Arnot, J. A.,
655 Joliet, O., Frame, A., Rabinowitz, J., Knudsen, T. B., Judson, R. S., Egeghy, P., Vallero,
656 D., & Cohen Hubal, E. A. (2013). High-Throughput Models for Exposure-Based
657 Chemical Prioritization in the ExpoCast Project. *Environmental Science & Technology*,
658 130711145716006. https://doi.org/10.1021/es400482g

659 Warriach, E. U., & Tei, K. (2013). Fault Detection in Wireless Sensor Networks: A Machine
660 Learning Approach. *2013 IEEE 16th International Conference on Computational Science*
661 *and Engineering*, 758–765. https://doi.org/10.1109/CSE.2013.116
662 Westerfield, M. (2007). *The zebrafish book: A guide for the laboratory use of zebrafish (Danio*
663 *rerio)* (Veterinary Medicine Library). Eugene, OR : Univ. of Oregon Press, 2007.
664 https://catalog.lib.ncsu.edu/catalog/NCSU2481113
665 Xia, C., Fu, L., Liu, Z., Liu, H., Chen, L., & Liu, Y. (2018). Aquatic Toxic Analysis by
666 Monitoring Fish Behavior Using Computer Vision: A Recent Progress. *Journal of*
667 *Toxicology*, *2018*, e2591924. https://doi.org/10.1155/2018/2591924
668 *ZFIN Zebrafish Developmental Stages*. (n.d.). Retrieved April 5, 2022, from
669 https://zfin.org/zf_info/zfbook/stages/index.html
670 Zhang, G., Truong, L., Tanguay, R. L., & Reif, D. M. (2017). A New Statistical Approach to
671 Characterize Chemical-Elicited Behavioral Effects in High-Throughput Studies Using
672 Zebrafish. *PloS One*, *12*(1), e0169408. https://doi.org/10.1371/journal.pone.0169408
673 Zhang, H., Lenaghan, S. C., Connolly, M. H., & Parker, L. E. (2013). Zebrafish Larva
674 Locomotor Activity Analysis Using Machine Learning Techniques. *2013 12th*
675 *International Conference on Machine Learning and Applications*, *1*, 161–166.
676 https://doi.org/10.1109/ICMLA.2013.35

677