

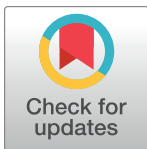
RESEARCH ARTICLE

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

Adrian J. Green^{1,2*}, Lisa Truong³, Preethi Thunga¹, Connor Leong³, Melody Hancock¹, Robyn L. Tanguay³, David M. Reif^{1,4}

1 Bioinformatics Research Center, Department of Biological Sciences, NC State University, Raleigh, North Carolina, United States of America, **2** Sciome LLC, Research Triangle Park, North Carolina, United States of America, **3** Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, Oregon, United States of America, **4** Predictive Toxicology Branch, Division of Translational Toxicology, National Institute of Environmental Health Sciences, Durham, North Carolina, United States of America

* ajgreen4@ncsu.edu



OPEN ACCESS

Citation: Green AJ, Truong L, Thunga P, Leong C, Hancock M, Tanguay RL, et al. (2024) Deep autoencoder-based behavioral pattern recognition outperforms standard statistical methods in high-dimensional zebrafish studies. *PLoS Comput Biol* 20(9): e1012423. <https://doi.org/10.1371/journal.pcbi.1012423>

Editor: Samuel V. Scarpino, Northeastern University, UNITED STATES OF AMERICA

Received: April 12, 2023

Accepted: August 15, 2024

Published: September 10, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1012423>

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: The code and data required to replicate findings reported in the article are available at <https://github.com/Tanguay-Lab/>

Abstract

Zebrafish have become an essential model organism 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 Nonafluoropentamide) 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.

[Manuscripts/tree/main/Green_et_al_\(2024\)_Manuscript.](#)

Funding: This research was supported by the National Institutes of Health (NIH) grant awards ES030287 (RLT, LT), ES030007 (AJG, DMR), ES025128 (DMR), ES033243 (DMR), and CA161608 (AJG, DMR). This research was supported [in part] by the Intramural Research Program of the NIH, ZIAES103385 (DMR). The funders played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

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 as exposure can lead to a range of neurodevelopmental disorders, including cognitive and other behavioral deficits. 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 generating similar data. 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 [1–3]. 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 [3–7]. Therefore, identifying associations between developmental exposures and neurological effects is a core objective to improve public health by informing disease and disability prevention [1,8].

As the number of environmental contaminants grows to nearly one million, comprehensive data on the neurodevelopmental toxicity of these contaminants remain sparse or nonexistent [3,9–11]. In response, high-throughput screening (HTS) assays have been developed to expedite chemical toxicity testing using *in vitro* and *in vivo* systems [12–14]. However, *in vitro* cell and cell-free assays cannot fully capture systemic organismal responses in terms of anatomy, physiology, or behavior [15]. 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 [12,16,17]. The zebrafish brain's structural organization, cellular morphology, and neurotransmitter systems are very similar to other vertebrates, including chickens, rats, and humans [18–21]. 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 model organism for revealing the neuronal developmental pathways underlying behavior [22–24].

Zebrafish larvae show swimming patterns essential for their survival following swim bladder development at four to five days post-fertilization (dpf), including exploration, foraging, and escape response which can be assessed using various locomotor behavioral assays [25,26] while more advanced continuous swimming, schooling, and reproductive behavior is still developing. 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 [27,28]. Exposure to toxicants has been shown to alter this stereotypical behavioral response [24,29]. Current HTS for behavioral neurotoxicity focuses heavily on analyzing locomotor behavior using distance moved and population-based statistical methods [24,30]. 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 [24,31,32]. 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 [33].

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 [34–36]. 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 μ M Nicotine) exhibit changes in distance moved [30], which is low given the known toxicity profiles of the chemical set. The traditional methods for analyzing zebrafish behavior data are primarily based on measurement of distance moved and instances of variations in the movement patterns, velocity changes and spatial preference is lost due to the sheer volume of data and complexity. Additionally, the traditional analysis methods is unable to identify meaningful patterns due to the noise and variability. This challenge provides an opportunity to apply methods developed for anomaly detection from areas such as financial fraud [37], medical application faults [38], security systems intrusion [39], system faults [40], and others [41,42]. Such ML techniques would allow for a more holistic evaluation of zebrafish behavior by learning complex features such as movement patterns, velocity changes and spatial preferences associated with “normal” behavior and flagging subtle deviations. These intricate nuances could be indicative of chemical toxicity and can often be missed by traditional assays relying solely on measuring distance moved as a metric. 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 [30]. One intriguing approach to achieving this is by applying an autoencoder [43–48]. An autoencoder is a neural network of two modules, an encoder and a decoder [47,49]. 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 a two sample Kolmogorov–Smirnov test (K-S test), a 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

A two sample Kolmogorov–Smirnov test (K-S test) was used to compare treated vs control distance moved and angular velocity in light/dark cycling in zebrafish larvae at five dpf. We identified 40 chemical-concentration combinations from nine chemicals and 28 chemical-concentration combinations from nine chemicals capable of inducing a significantly different ($p < 0.05$) behavioral response using both distance moved and angular velocity, respectively (S2 Table). While 10 chemicals were identified using both methods, nine chemicals were similar, with distance moved finding a significant difference in multi-walled carbon nanotubes at 75 and 100 μM and angular velocity finding a difference in sodium 2-(N-ethylperfluorooctane-1-sulfonamido)ethyl phosphate at 0.25 μM . Considering that distance moved revealed more chemical-concentration combinations in this screening application, we used this metric to identify abnormal larvae to ensure a sufficient number for training the autoencoder models. Using the 30th and 70th 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 (S1 Fig 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.

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 30th and 70th 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,

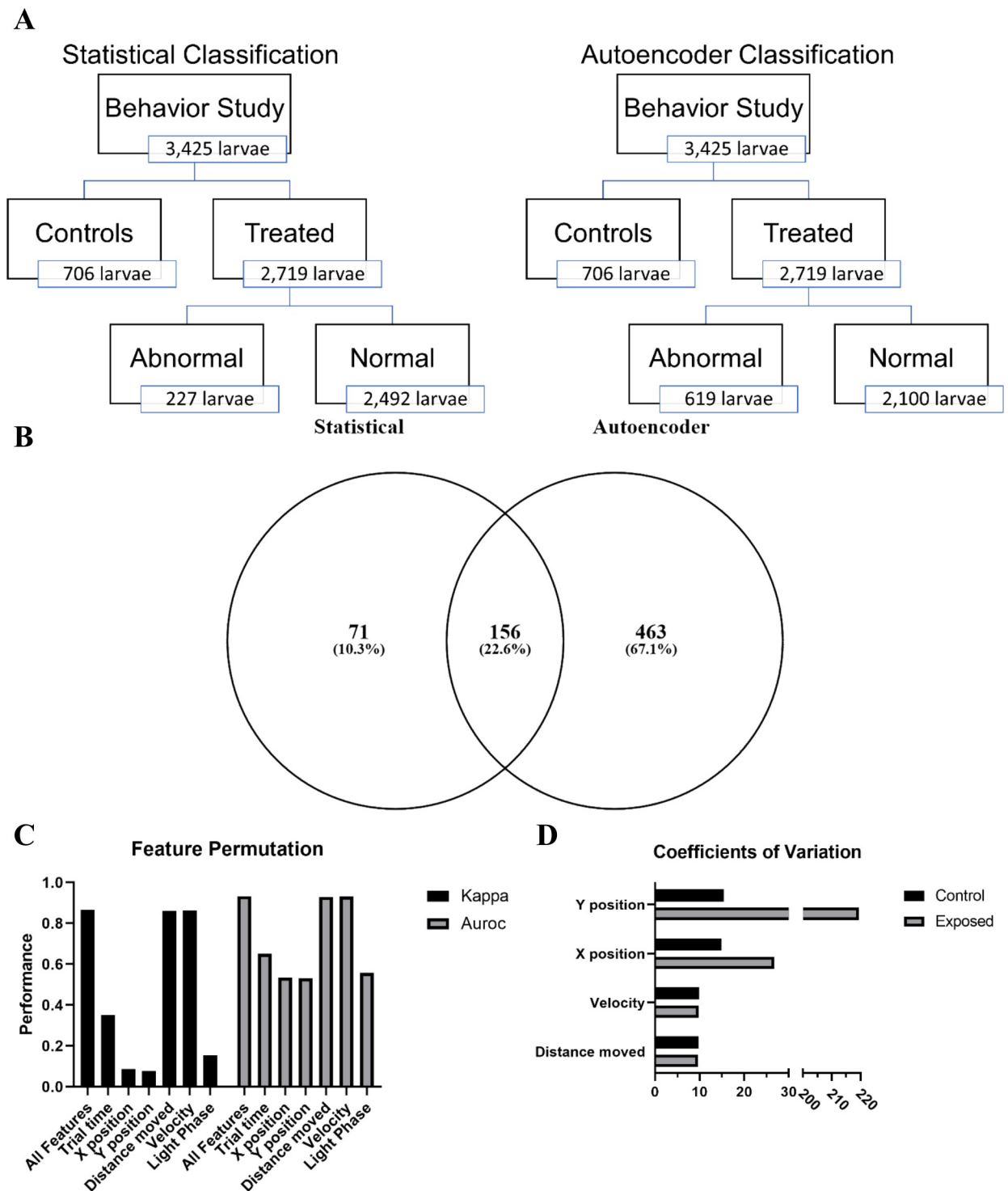


Fig 1. Assessment autoencoder performance. (A) Schematic representation of the differences in statistical and autoencoder based classification of behavioral response in larval zebrafish. (B) Venn diagram showing overlap between statistical and autoencoder classified abnormal zebrafish. (C) Evaluating 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 $\pm$ SEM. (D) Coefficients of variation for each of the main numerical features.

<https://doi.org/10.1371/journal.pcbi.1012423.g001>

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

<https://doi.org/10.1371/journal.pcbi.1012423.t001>

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.

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).

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

Table 2. Autoencoders identified chemicals. Table showing chemicals and concentrations flagged for displaying abnormal behavioral effects 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

<https://doi.org/10.1371/journal.pcbi.1012423.t002>

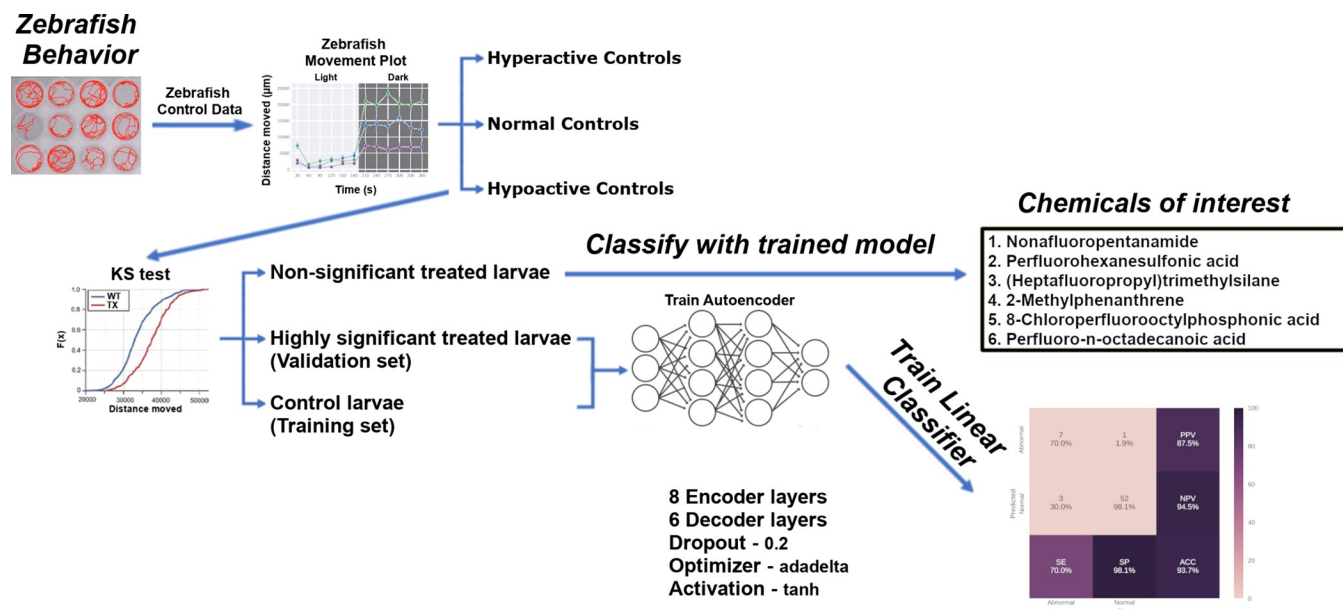


Fig 2. Summary of behavioral analysis pipeline and results. Utilizing our analysis pipeline produced six deep autoencoder models (three for the light phase and three for the dark phase) capable of classifying larval zebrafish behavior with high Kappa and AUROC values. The trained models were then used 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.

<https://doi.org/10.1371/journal.pcbi.1012423.g002>

models accurately classified all controls as normal while detecting similar levels of abnormal behavior response across the concentration range (Fig 4) ($p > 0.15$). These results show that the trained model is capable of producing similar results across experimental replicates.

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

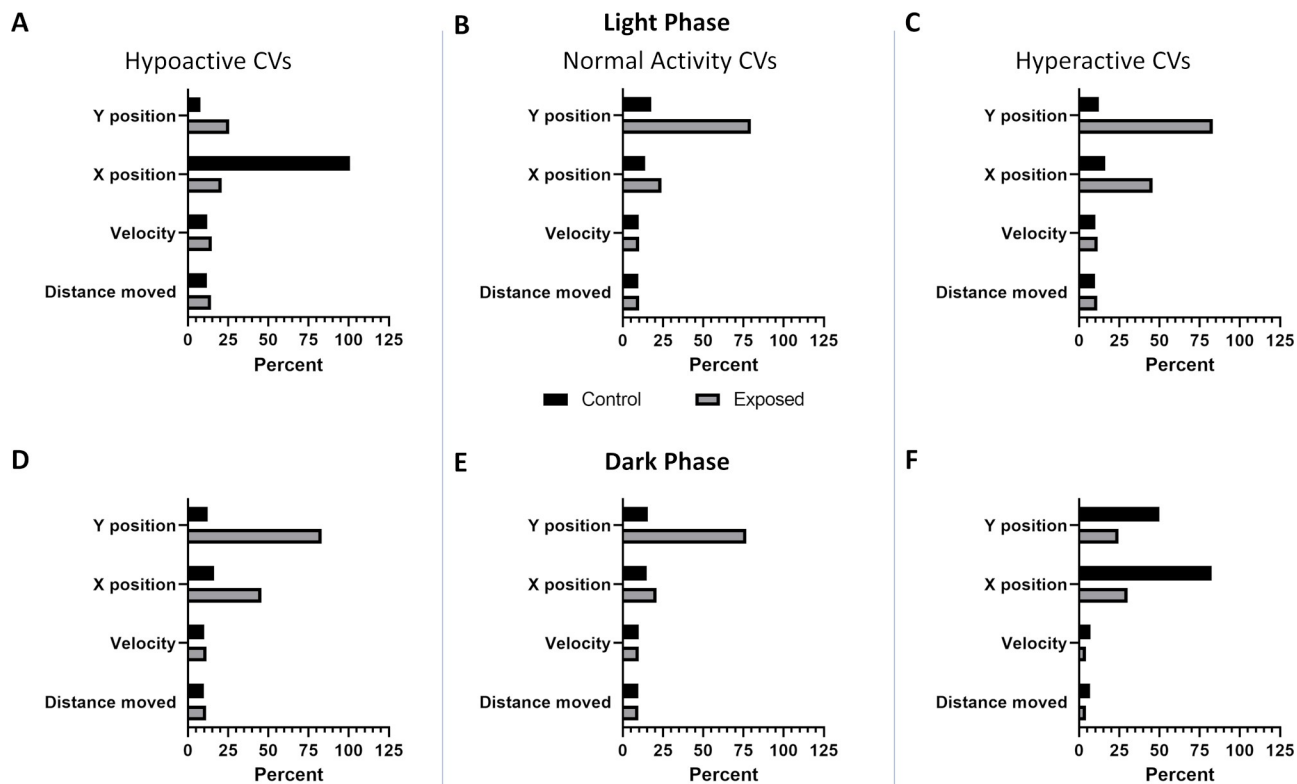


Fig 3. Coefficients of variation per larval activity state. Coefficients of variation (CVs) for each of the main numerical features (A–C) in the light (D–F) and in the dark. Columns show CVs of larval zebrafish significantly ($p < 0.05$) (A, D) hypoactive, (B, E) normal activity, or (C, F) hyperactive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

<https://doi.org/10.1371/journal.pcbi.1012423.g003>

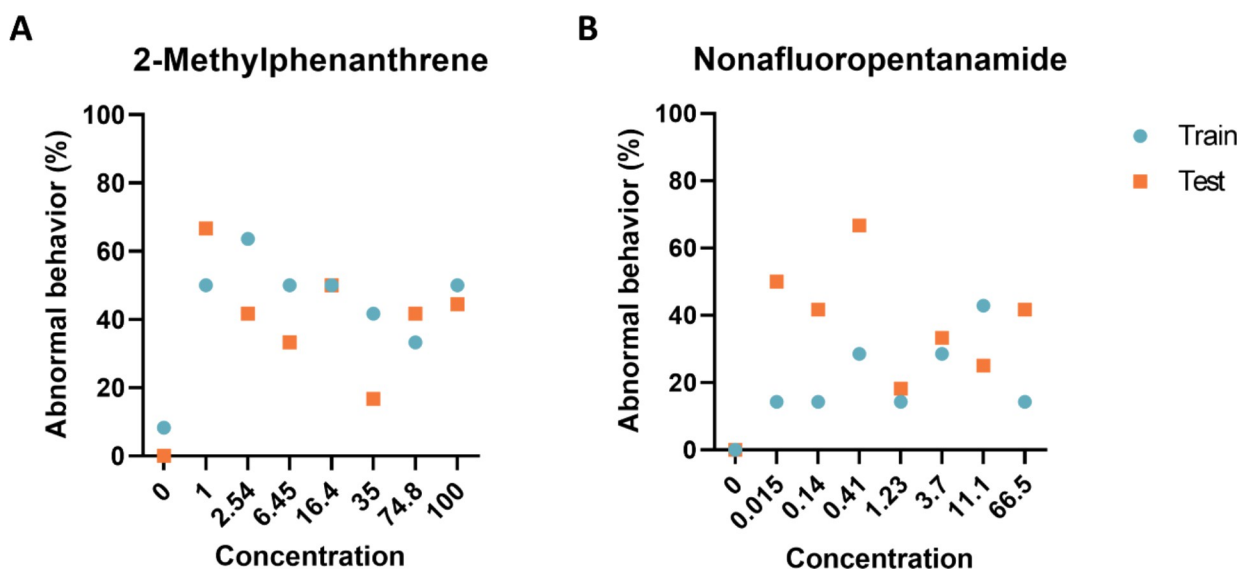


Fig 4. Experimental model evaluation. Comparison of the performance of deep autoencoder models between the training set and two chemicals identified by the models to elicit abnormal larval zebrafish behavior. Percent of larval zebrafish classified as abnormal based on their behavioral response to developmental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

<https://doi.org/10.1371/journal.pcbi.1012423.g004>

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). While making a definitive claim will require further experimentation, it does appear that the autoencoder method is particularly sensitive at detecting changes due to PFAS exposure. PFAS are associated with increased glutamate levels in the hippocampus and catecholamine levels in the hypothalamus, decreased dopamine in the whole brain after PFAS exposure, and increased extracellular glutamate has been observed in the hippocampus epileptic rats [50,51]. Thus, it is reasonable to infer that these neurochemical changes are capable of altering autoencoder-detectable patterns without changing locomotor magnitude or direction.

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 [31,52] or through the application of unsupervised ML [27,32,53–57]. These studies have focused on the analysis and categorization of behavioral patterns in wild-type strains [27,57], mutant strains [32,53], or larvae exposed to neuroactive chemicals [32] 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 [41–43,58], information systems [38,40], security systems [39,45], and financial fraud [37] they have only very recently been applied to biological data [44,59,60]. 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 such as the two sample Kolmogorov–Smirnov test (K-S test). 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. Our deep autoencoders only needed seven hundred controls and a three-minute light and three-minute dark cycle to identify differences. The majority of zebrafish labs have historical or the ability to generate similar data that can be used to train their own deep autoencoder models. 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 [32], 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 control data generated as part of any high-throughput screening (HTS)—setting the stage for predictive toxicological applications and safety assessments for the enormous backlog of as-yet untested chemicals.

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 2. 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 (S1 Table).

Ethics statement

This study was conducted in accordance with the guidelines and regulations set forth by the Institutional Animal Care and Use Committee (IACUC) at Oregon State University. The protocol was reviewed and approved by the IACUC under the approval number 2021–0227. All procedures involving animals were performed in compliance with the ethical standards of the institution and adhered to the principles of humane animal treatment.

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 Oregon State University Animal Use Care and Protocol: 2021–0227 [61]. 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 [62,63]. 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 MgSO₄, 0.15 mM KH₂PO₄, 0.05 mM Na₂HPO₄, and 0.7 mM NaHCO₃ [63].

Developmental chemical exposure

The empirical data used to develop our model were gathered as described in Truong et al. and Noyes et al. [12,64,65]. The experimental design consisted of the 30 unique chemicals tested (S1 Table) with at least 7 replicates (an individual embryo in singular wells of a 96-well plate) at each concentration for each chemical. The concentrations evaluated were based on preliminary studies within the authors' lab to span lethal and sub-lethal concentration range were possible based on physical chemicals properties including solubility.

Developmental toxicity assessments

Mortality and morphology. At 24 hours post-fertilization (hpf), embryos were screened for mortality, developmental delay, and spontaneous movement [12]. At 120 hpf, mortality, craniofacial abnormalities (eye, snout and jaw), body axis abnormalities, edema (yolk sac and pericardial edema), upright body abnormalities (swim bladder, somite and circulation), touch response brain abnormalities (brain, otic vesicle and pectoral fin), pigment, notochord, and trunk abnormalities (trunk and caudal fin) [12,66,67]. The incidence of abnormality across all 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-round well plates of larvae were placed into a Zebrafish box (Viewpoint Life-Sciences) 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 with an initial 6 minute dark acclimation period. 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 (S2 Fig). 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 [29].

Data preprocessing and statistical analysis pipeline

Preprocessing. All data processing, statistical analysis and ML were implemented in Python using the open source libraries Tensorflow [68], Keras [69], Scikit-learn [70], Pandas [71], and Numpy [72] within a purpose build Singularity container environment [73]. 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 its 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 [53,74]. This resulted in 5,445 of the 30,825,000 frames being normalized.

Statistical analysis

A two sample Kolmogorov–Smirnov test (K-S test), a non-parametric two-sided test with no adjustments for normality or multiple comparisons, was used to compare each chemical-concentration combination with their respective same plate controls ($p < 0.05$). Interexperimental zebrafish larval response to light/dark cycling is highly variable (S2 Fig). Therefore, it was essential to group the unexposed controls based on the mean from individual 96-well plates compared to mean movement for unexposed controls across all plates. 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 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 adadelat 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 adadelat optimizer (learning_rate = 0.001, rho = 0.95, and epsilon = 1e-07) and mean squared error for the loss function [75–77]. 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 dataset are available at GitHub [[https://github.com/Tanguay-Lab/Manuscripts/tree/main/Green_et_al_\(2024\)_Manuscript](https://github.com/Tanguay-Lab/Manuscripts/tree/main/Green_et_al_(2024)_Manuscript)].

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) [78]. 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 [79]. 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 [80–82]. 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 [83]. 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).

Experimental confirmation of autoencoder findings

Following model development two chemicals were identified for follow-up laboratory testing. We generated new data using 2-Methylphenanthrene, and Nonafluoropentanamide. 2-Methylphenanthrene was chosen as the autoencoder identified it was different from controls at a much lower concentration than a K-S test of distance moved and angular velocity while Nonafluoropentanamide was selected as it was not identified using either a K-S test of distance moved and angular velocity. Similarity between the results was determined by comparing fourth order polynomial curve fits with and a significance threshold of $p < 0.05$.

Supporting information

S1 Table. Study chemicals and their common use.
(XLSX)

S2 Table. Statistical results for behavioral response analysis.
(XLSX)

S1 Fig. Loss function results during training. Changes of loss functions during the training of (A) light-hypoactive controls, (B) light-normal controls, (C) light-hyperactive controls, (D) dark-hypoactive controls, (E) dark-normal controls, (F) dark-hyperactive controls. Blue line—training data (controls-only), orange line—test data (abnormal-only). (TIF)

S2 Fig. Interexperimental behavioral response to light/dark cycling in control larval zebrafish. Zebrafish larvae were statically exposed to a chemical from six hpf until five dpf. At five dpf, behavior was measured under environmental conditions of continuous light for three minutes (0–180) followed by three minutes of dark (180–360). This plot shows representative control behavior data ($n = 7$ per line) classified as hyperactive (blue line), normal (green line) or hypoactive (purple line). The insert shows an example of larval behavioral tracks produced by Ethovision XT software. Figure depicts means $\pm$ SEM. (TIF)

Acknowledgments

We would like to thank the staff at Sinnhuber Aquatic Research Laboratory, and John Lam for his contribution to reprocessing videos.

Author Contributions

Conceptualization: Adrian J. Green, David M. Reif.

Data curation: Lisa Truong.

Formal analysis: Adrian J. Green, Preethi Thunga, Melody Hancock.

Funding acquisition: Robyn L. Tanguay, David M. Reif.

Investigation: Lisa Truong, Connor Leong.

Methodology: Adrian J. Green.

Project administration: David M. Reif.

Resources: Robyn L. Tanguay, David M. Reif.

Supervision: Robyn L. Tanguay, David M. Reif.

Validation: Adrian J. Green, Lisa Truong, Connor Leong.

Writing – original draft: Adrian J. Green.

Writing – review & editing: Adrian J. Green, Lisa Truong, Preethi Thunga, Connor Leong, Melody Hancock, Robyn L. Tanguay, David M. Reif.

References

1. Neurodevelopmental Diseases. In: National Institute of Environmental Health Sciences [Internet]. 12 Jan 2021 [cited 12 Jan 2021]. Available: <https://www.niehs.nih.gov/research/supported/health/neurodevelopmental/index.cfm>.
2. Boyle CA, Boulet S, Schieve LA, Cohen RA, Blumberg SJ, Yeargin-Allsopp M, et al. Trends in the prevalence of developmental disabilities in US children, 1997–2008. *Pediatrics*. 2011; 127: 1034–1042. <https://doi.org/10.1542/peds.2010-2989> PMID: 21606152
3. US EPA. Health: Neurodevelopmental Disorders—Report Contents. In: Health: Neurodevelopmental Disorders—Report Contents [Internet]. 10 Jun 2015 [cited 12 Jan 2021]. Available: <https://www.epa.gov/americaschildrenenvironment/health-neurodevelopmental-disorders-report-contents/>.

4. Grandjean P, Landrigan PJ. Neurobehavioural effects of developmental toxicity. *The Lancet Neurology*. 2014; 13: 330–338. [https://doi.org/10.1016/S1474-4422\(13\)70278-3](https://doi.org/10.1016/S1474-4422(13)70278-3) PMID: 24556010
5. Rock KD, Patisaul HB. Environmental Mechanisms of Neurodevelopmental Toxicity. *Curr Environ Health Rep*. 2018; 5: 145–157. <https://doi.org/10.1007/s40572-018-0185-0> PMID: 29536388
6. Green AJ, Planchart A. The neurological toxicity of heavy metals: A fish perspective. *Comp Biochem Physiol C Toxicol Pharmacol*. 2018; 208: 12–19. <https://doi.org/10.1016/j.cbpc.2017.11.008> PMID: 29199130
7. Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature*. 2014; 508: 199–206. <https://doi.org/10.1038/nature13185> PMID: 24695229
8. A Blueprint for Brain Development. In: NIH Director's Blog [Internet]. 8 Apr 2014 [cited 12 Jan 2021]. Available: <https://directorsblog.nih.gov/2014/04/08/a-blueprint-for-brain-development/>.
9. US EPA. About the TSCA Chemical Substance Inventory. In: About the TSCA Chemical Substance Inventory [Internet]. 2 Mar 2015 [cited 23 Aug 2019]. Available: <https://www.epa.gov/tsca-inventory/about-tsca-chemical-substance-inventory>.
10. Krewski D, Andersen ME, Tyshenko MG, Krishnan K, Hartung T, Boekelheide K, et al. Toxicity testing in the 21st century: progress in the past decade and future perspectives. *Arch Toxicol*. 2020; 94: 1–58. <https://doi.org/10.1007/s00204-019-02613-4> PMID: 31848664
11. Wambaugh JF, Setzer RW, Reif DM, Gangwal S, Mitchell-Blackwood J, Arnot JA, et al. High-Throughput Models for Exposure-Based Chemical Prioritization in the ExpoCast Project. *Environ Sci Technol*. 2013; 130711145716006. <https://doi.org/10.1021/es400482g> PMID: 23758710
12. Truong L, Reif DM, St Mary L, Geier MC, Truong HD, Tanguay RL. Multidimensional In Vivo Hazard Assessment Using Zebrafish. *Toxicol Sci*. 2014; 137: 212–233. <https://doi.org/10.1093/toxsci/kft235> PMID: 24136191
13. Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, et al. ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. *Chem Res Toxicol*. 2016; 29: 1225–1251. <https://doi.org/10.1021/acs.chemrestox.6b00135> PMID: 27367298
14. Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. *Environ Health Perspect*. 2010; 118: 485–492. <https://doi.org/10.1289/ehp.0901392> PMID: 20368123
15. Thomas RS, Black MB, Li L, Healy E, Chu T-M, Bao W, et al. A Comprehensive Statistical Analysis of Predicting In Vivo Hazard Using High-Throughput In Vitro Screening. *Toxicol Sci*. 2012; 128: 398–417. <https://doi.org/10.1093/toxsci/kfs159> PMID: 22543276
16. Bugel SM, Tanguay RL, Planchart A. Zebrafish: A Marvel of High-Throughput Biology for 21st Century Toxicology. *Curr Environ Health Rpt*. 2014; 1: 341–352. <https://doi.org/10.1007/s40572-014-0029-5> PMID: 25678986
17. Planchart A, Green AJ, Hoyo C, Mattingly CJ. Heavy Metal Exposure and Metabolic Syndrome: Evidence from Human and Model System Studies. *Curr Environ Health Rep*. 2018; 5: 110–124. <https://doi.org/10.1007/s40572-018-0182-3> PMID: 29460222
18. Kalueff AV, Stewart AM, Gerlai R. Zebrafish as an emerging model for studying complex brain disorders. *Trends in Pharmacological Sciences*. 2014; 35: 63–75. <https://doi.org/10.1016/j.tips.2013.12.002> PMID: 24412421
19. Lowery LA, Sive H. Strategies of vertebrate neurulation and a re-evaluation of teleost neural tube formation. *Mechanisms of Development*. 2004; 121: 1189–1197. <https://doi.org/10.1016/j.mod.2004.04.022> PMID: 15327780
20. Tropepe V, Sive HL. Can zebrafish be used as a model to study the neurodevelopmental causes of autism? *Genes, Brain and Behavior*. 2003; 2: 268–281. <https://doi.org/10.1034/j.1601-183x.2003.00038.x> PMID: 14606692
21. Horzmann KA, Freeman JL. Zebrafish Get Connected: Investigating Neurotransmission Targets and Alterations in Chemical Toxicity. *Toxics*. 2016; 4: 19. <https://doi.org/10.3390/toxics4030019> PMID: 28730152
22. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*. 2013; 496: 498–503. <https://doi.org/10.1038/nature12111> PMID: 23594743
23. Postlethwait JH, Yan YL, Gates MA, Horne S, Amores A, Brownlie A, et al. Vertebrate genome evolution and the zebrafish gene map. *Nat Genet*. 1998; 18: 345–349. <https://doi.org/10.1038/ng0498-345> PMID: 9537416
24. Basnet RM, Zizioli D, Taweedet S, Finazzi D, Memo M. Zebrafish Larvae as a Behavioral Model in Neuropharmacology. *Biomedicines*. 2019; 7: 23. <https://doi.org/10.3390/biomedicines7010023> PMID: 30917585

25. Hernandez RE, Galitan L, Cameron J, Goodwin N, Ramakrishnan L. Delay of Initial Feeding of Zebrafish Larvae Until 8 Days Postfertilization Has No Impact on Survival or Growth Through the Juvenile Stage. *Zebrafish*. 2018; 15: 515–518. <https://doi.org/10.1089/zeb.2018.1579> PMID: 30089231
26. Tegelenbosch RAJ, Noldus LPJJ, Richardson MK, Ahmad F. Zebrafish embryos and larvae in behavioural assays. *Behaviour*. 2012; 149: 1241–1281. <https://doi.org/10.1163/1568539X-00003020>
27. Burgess HA, Granato M. Modulation of locomotor activity in larval zebrafish during light adaptation. *Journal of Experimental Biology*. 2007; 210: 2526–2539. <https://doi.org/10.1242/jeb.003939> PMID: 17601957
28. Emran F, Rihel J, Dowling JE. A behavioral assay to measure responsiveness of zebrafish to changes in light intensities. *J Vis Exp*. 2008. <https://doi.org/10.3791/923> PMID: 19078942
29. Truong L, Bugel SM, Chlebowski A, Usenko CY, Simonich MT, Simonich SLM, et al. Optimizing multi-dimensional high throughput screening using zebrafish. *Reproductive Toxicology*. 2016; 65: 139–147. <https://doi.org/10.1016/j.reprotox.2016.05.015> PMID: 27453428
30. Zhang G, Truong L, Tanguay RL, Reif DM. A New Statistical Approach to Characterize Chemical-Elicited Behavioral Effects in High-Throughput Studies Using Zebrafish. *PLoS ONE*. 2017; 12: e0169408. <https://doi.org/10.1371/journal.pone.0169408> PMID: 28099482
31. Kalueff AV, Gebhardt M, Stewart AM, Cachat JM, Brimmer M, Chawla JS, et al. Towards a Comprehensive Catalog of Zebrafish Behavior 1.0 and Beyond. *Zebrafish*. 2013; 10: 70–86. <https://doi.org/10.1089/zeb.2012.0861> PMID: 23590400
32. Mirat O, Sternberg JR, Severi KE, Wyart C. ZebraZoom: an automated program for high-throughput behavioral analysis and categorization. *Front Neural Circuits*. 2013;7. <https://doi.org/10.3389/fncir.2013.00107> PMID: 23781175
33. Balik-Meisner M, Truong L, Scholl EH, La Du JK, Tanguay RL, Reif DM. Elucidating Gene-by-Environment Interactions Associated with Differential Susceptibility to Chemical Exposure. *Environmental Health Perspectives*. 2018;126. <https://doi.org/10.1289/EHP2662> PMID: 29968567
34. Arifoglu D, Bouchachia A. Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks. *Procedia Computer Science*. 2017; 110: 86–93. <https://doi.org/10.1016/j.procs.2017.06.121>
35. Xia C, Fu L, Liu Z, Liu H, Chen L, Liu Y. Aquatic Toxic Analysis by Monitoring Fish Behavior Using Computer Vision: A Recent Progress. *Journal of Toxicology*. 2018; 2018: e2591924. <https://doi.org/10.1155/2018/2591924> PMID: 29849612
36. Pereira TD, Shaevitz JW, Murthy M. Quantifying behavior to understand the brain. *Nat Neurosci*. 2020; 23: 1537–1549. <https://doi.org/10.1038/s41593-020-00734-z> PMID: 33169033
37. Awoyemi JO, Adetunmbi AO, Oluwadare SA. Credit card fraud detection using machine learning techniques: A comparative analysis. 2017 International Conference on Computing Networking and Informatics (ICCNi). 2017. pp. 1–9. <https://doi.org/10.1109/ICCNi.2017.8123782>
38. Pachauri G, Sharma S. Anomaly Detection in Medical Wireless Sensor Networks using Machine Learning Algorithms. *Procedia Computer Science*. 2015; 70: 325–333. <https://doi.org/10.1016/j.procs.2015.10.026>
39. Sargolzaei A, Crane CD, Abbaspour A, Noei S. A Machine Learning Approach for Fault Detection in Vehicular Cyber-Physical Systems. 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA). 2016. pp. 636–640. <https://doi.org/10.1109/ICMLA.2016.0112>
40. Warriach EU, Tei K. Fault Detection in Wireless Sensor Networks: A Machine Learning Approach. 2013 IEEE 16th International Conference on Computational Science and Engineering. 2013. pp. 758–765. <https://doi.org/10.1109/CSE.2013.116>
41. Fazai R, Abodayeh K, Mansouri M, Trabelsi M, Nounou H, Nounou M, et al. Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems. *Solar Energy*. 2019; 190: 405–413. <https://doi.org/10.1016/j.solener.2019.08.032>
42. Jaiswal V, Ruskin A. Mooring Line Failure Detection Using Machine Learning. *OnePetro*; 2019. <https://doi.org/10.4043/29511-MS>
43. Nicholas IT, Park JR, Jung K, Lee JS, Kang D-K. Anomaly Detection of Water Level Using Deep Autoencoder. *Sensors (Basel)*. 2021; 21: 6679. <https://doi.org/10.3390/s21196679> PMID: 34640997
44. Frassek M, Arjun A, Bolhuis PG. An extended autoencoder model for reaction coordinate discovery in rare event molecular dynamics datasets. *J Chem Phys*. 2021; 155: 064103. <https://doi.org/10.1063/5.0058639> PMID: 34391359
45. Feng J, Liang Y, Li L. Anomaly Detection in Videos Using Two-Stream Autoencoder with Post Hoc Interpretability. *Comput Intell Neurosci*. 2021; 2021: 7367870. <https://doi.org/10.1155/2021/7367870> PMID: 34354745

46. Ranjan C, Reddy M, Mustonen M, Paynabar K, Pourak K Dataset: Rare Event Classification in Multivariate Time Series. arXiv:180910717 [cs, stat]. 2019 [cited 4 Jan 2022]. Available: <http://arxiv.org/abs/1809.10717>.
47. Goodfellow I, Bengio Y, Courville A. Chapter 14—Autoencoders. Deep Learning. MIT Press; 2016. pp. 499–523.
48. Le Borgne Y-A, Siblini W, Lebichot B, Bontempi G. Autoencoders and anomaly detection—Reproducible Machine Learning for Credit Card Fraud detection—Practical handbook. Reproducible Machine Learning for Credit Card Fraud Detection—Practical Handbook. Université Libre de Bruxelles; 2022. Available: <https://github.com/Fraud-Detection-Handbook/fraud-detection-handbook>.
49. Gupta A, Singh S. ML | Classifying Data using an Auto-encoder. In: GeeksforGeeks [Internet]. 25 Jun 2019 [cited 10 Feb 2020]. Available: <https://www.geeksforgeeks.org/ml-classifying-data-using-an-auto-encoder/>.
50. Brown-Leung JM, Cannon JR. Neurotransmission Targets of Per- and Polyfluoroalkyl Substance Neurotoxicity: Mechanisms and Potential Implications for Adverse Neurological Outcomes. Chem Res Toxicol. 2022; 35: 1312–1333. <https://doi.org/10.1021/acs.chemrestox.2c00072> PMID: 35921496
51. Soukupova M, Binaschi A, Falcicchia C, Palma E, Roncon P, Zucchini S, et al. Increased extracellular levels of glutamate in the hippocampus of chronically epileptic rats. Neuroscience. 2015; 301: 246–253. <https://doi.org/10.1016/j.neuroscience.2015.06.013> PMID: 26073699
52. Fero K, Yokogawa T, Burgess HA. The Behavioral Repertoire of Larval Zebrafish. In: Kalueff AV, Cachat JM, editors. Zebrafish Models in Neurobehavioral Research. Totowa, NJ: Humana Press; 2011. pp. 249–291. https://doi.org/10.1007/978-1-60761-922-2_12
53. Burgess HA, Granato M. Sensorimotor Gating in Larval Zebrafish. J Neurosci. 2007; 27: 4984–4994. <https://doi.org/10.1523/JNEUROSCI.0615-07.2007> PMID: 17475807
54. Kimmel CB, Patterson J, Kimmel RO. The development and behavioral characteristics of the startle response in the zebra fish. Developmental Psychobiology. 1974; 7: 47–60. <https://doi.org/10.1002/dev.420070109> PMID: 4812270
55. Budick SA, O'Malley DM. Locomotor repertoire of the larval zebrafish: swimming, turning and prey capture. Journal of Experimental Biology. 2000; 203: 2565–2579. <https://doi.org/10.1242/jeb.203.17.2565> PMID: 10934000
56. Burgess HA, Granato M. Flote v2.1: Biological Tracking Software. 2007.
57. Zhang H, Lenaghan SC, Connolly MH, Parker LE. Zebrafish Larva Locomotor Activity Analysis Using Machine Learning Techniques. 2013 12th International Conference on Machine Learning and Applications. 2013. pp. 161–166. <https://doi.org/10.1109/ICMLA.2013.35>
58. Fan C, Xiao F, Zhao Y, Wang J. Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. Applied Energy. 2018; 211: 1123–1135. <https://doi.org/10.1016/j.apenergy.2017.12.005>
59. Homayouni H, Ray I, Ghosh S, Gondalia S, Kahn MG. Anomaly Detection in COVID-19 Time-Series Data. SN Comput Sci. 2021; 2: 279. <https://doi.org/10.1007/s42979-021-00658-w> PMID: 34027432
60. Nwokedi EI, Bains R, Bidaut L, Wells S, Ye X, Brown JM. Unsupervised detection of mouse behavioural anomalies using two-stream convolutional autoencoders. ArXiv. 2021.
61. Barton CL, Johnson EW, Tanguay RL. Facility Design and Health Management Program at the Sinnhuber Aquatic Research Laboratory. Zebrafish. 2016; 13: S-39-S-43. <https://doi.org/10.1089/zeb.2015.1232> PMID: 26981844
62. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. Stages of embryonic development of the zebrafish. Dev Dyn. 1995; 203: 253–310. <https://doi.org/10.1002/aja.1002030302> PMID: 8589427
63. Westerfield M. The zebrafish book: a guide for the laboratory use of zebrafish (Danio rerio). Eugene, OR: Eugene, OR: Univ. of Oregon Press, 2007.; 2007. Available: <https://catalog.lib.ncsu.edu/catalog/NCSU2481113>.
64. Noyes PD, Haggard DE, Gonnerman GD, Tanguay RL. Advanced Morphological—Behavioral Test Platform Reveals Neurodevelopmental Defects in Embryonic Zebrafish Exposed to Comprehensive Suite of Halogenated and Organophosphate Flame Retardants. Toxicol Sci. 2015; 145: 177–195. <https://doi.org/10.1093/toxsci/kfv044> PMID: 25711236
65. Truong L, Rericha Y, Thunga P, Marvel S, Wallis D, Simonich MT, et al. Systematic developmental toxicity assessment of a structurally diverse library of PFAS in zebrafish. Journal of Hazardous Materials. 2022; 431: 128615. <https://doi.org/10.1016/j.jhazmat.2022.128615> PMID: 35263707
66. Thunga P, Truong L, Tanguay RL, Reif DM. Concurrent Evaluation of Mortality and Behavioral Responses: A Fast and Efficient Testing Approach for High-Throughput Chemical Hazard Identification. Frontiers in Toxicology. 2021; 3. Available: <https://www.frontiersin.org/articles/10.3389/ftox.2021.670496>. <https://doi.org/10.3389/ftox.2021.670496> PMID: 35295121

67. Zhang G, Marvel S, Truong L, Tanguay RL, Reif DM. Aggregate entropy scoring for quantifying activity across endpoints with irregular correlation structure. *Reprod Toxicol*. 2016; 62: 92–99. <https://doi.org/10.1016/j.reprotox.2016.04.012> PMID: 27132190
68. Abadi Martín, Agarwal Ashish, Barham Paul, Brevdo Eugene, Chen Zhifeng, Citro Craig, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available: <https://www.tensorflow.org/>.
69. U.S. Environmental Protection Agency. Comptox Chemicals Dashboard: Master List of PFAS Substances (Version2). 10 Aug 2021 [cited 19 May 2022]. Available: <https://comptox.epa.gov/dashboard/chemical-lists/pfasmaster>.
70. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011; 12: 2825–2830.
71. McKinney W. Data Structures for Statistical Computing in Python. Austin, Texas; 2010. pp. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>
72. Harris CR, Millman KJ, Walt SJ van der, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020; 585: 357–362. <https://doi.org/10.1038/s41586-020-2649-2> PMID: 32939066
73. Sylabs.io. Singularity. Sylabs.io; 2019. Available: <https://sylabs.io/singularity/>.
74. ZFIN Zebrafish Developmental Stages. [cited 5 Apr 2022]. Available: https://zfin.org/zf_info/zfbook/stages/index.html.
75. Ramachandran P, Zoph B, Le QV. Searching for Activation Functions. arXiv:171005941 [cs]. 2017 [cited 4 Sep 2020]. Available: <http://arxiv.org/abs/1710.05941>.
76. Osl M, Netzer M, Dreiseitl S, Baumgartner C. Applied Data Mining: From Biomarker Discovery to Decision Support Systems. In: Trajanoski Z, editor. *Computational Medicine*. Vienna: Springer Vienna; 2012. pp. 173–184. https://doi.org/10.1007/978-3-7091-0947-2_10
77. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 IEEE International Conference on Computer Vision (ICCV). 2015. pp. 1026–1034. <https://doi.org/10.1109/ICCV.2015.123>
78. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*. 2017; 18: 1–5.
79. Ben-David A. About the relationship between ROC curves and Cohen’s kappa. *Engineering Applications of Artificial Intelligence*. 2008; 21: 874–882. <https://doi.org/10.1016/j.engappai.2007.09.009>
80. Pearson K. On the theory of contingency and its relation to association and normal correlation. *Drapers Company Research Memoirs*. Dulau and Co.; 1904. Available: <https://archive.org/details/cu31924003064833/page/n1/mode/2up>.
81. Townsend JT. Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*. 1971; 9: 40–50. <https://doi.org/10.3758/BF03213026>
82. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008; 56: 45–50. <https://doi.org/10.4103/0301-4738.37595> PMID: 18158403
83. Breiman L. Random Forests. *Machine Learning*. 2001; 45: 5–32. <https://doi.org/10.1023/A:1010933404324>