Check for updates

OPEN ACCESS

EDITED BY

Hassene Seddik, University of Tunis/RIFTSI Laboratory (Smart Robotic, Friability and Signal and Image Processing Research Laboratory), Tunisia

REVIEWED BY Giorgos Livanos, Technical University of Crete, Greece Mohammad Khubeh Siddiqui

Mohammad Khubeb Siddiqui, Saudi Arabia Basic Industries, Saudi Arabia *CORRESPONDENCE

Xiaowei Wu ⊠ xwwu@vt.edu Hehuang Xie ⊠ davidxie@vt.edu

RECEIVED 18 November 2024 ACCEPTED 29 April 2025 PUBLISHED 13 May 2025

CITATION

Yin L, VanderGiessen M, Kumar V, Conacher B, Chao P-CH, Theus M, Johnson E, Kehn-Hall K, Wu X and Xie H (2025) Machine learning identifies genes linked to neurological disorders induced by equine encephalitis viruses, traumatic brain injuries, and organophosphorus nerve agents. *Front. Comput. Neurosci.* 19:1529902. doi: 10.3389/fncom.2025.1529902

COPYRIGHT

© 2025 Yin, VanderGiessen, Kumar, Conacher, Chao, Theus, Johnson, Kehn-Hall, Wu and Xie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms. Machine learning identifies genes linked to neurological disorders induced by equine encephalitis viruses, traumatic brain injuries, and organophosphorus nerve agents

Liduo Yin¹, Morgen VanderGiessen^{1,2}, Vinoth Kumar¹, Benjamin Conacher¹, Po-Chien Haku Chao¹, Michelle Theus¹, Erik Johnson³, Kylene Kehn-Hall^{1,2}, Xiaowei Wu^{4*} and Hehuang Xie^{1*}

¹Department of Biomedical Sciences and Pathobiology, Virginia-Maryland College of Veterinary Medicine, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, ²Center for Emerging, Zoonotic, and Arthropod-borne Pathogens, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, ³Neuroscience Department, Medical Toxicology Division, U.S. Army Medical Research Institute of Chemical Defense, Aberdeen, MD, United States, ⁴Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Venezuelan, eastern, and western equine encephalitis viruses (collectively referred to as equine encephalitis viruses---EEV) cause serious neurological diseases and pose a significant threat to the civilian population and the warfighter. Likewise, organophosphorus nerve agents (OPNA) are highly toxic chemicals that pose serious health threats of neurological deficits to both military and civilian personnel around the world. Consequently, only a select few approved research groups are permitted to study these dangerous chemical and biological warfare agents. This has created a significant gap in our scientific understanding of the mechanisms underlying neurological diseases. Valuable insights may be gleaned by drawing parallels to other extensively researched neuropathologies, such as traumatic brain injuries (TBI). By examining combined gene expression profiles, common and unique molecular characteristics may be discovered, providing new insights into medical countermeasures (MCMs) for TBI, EEV infection and OPNA neuropathologies and sequelae. In this study, we collected transcriptomic datasets for neurological disorders caused by TBI, EEV, and OPNA injury, and implemented a framework to normalize and integrate gene expression datasets derived from various platforms. Effective machine learning approaches were developed to identify critical genes that are either shared by or distinctive among the three neuropathologies. With the aid of deep neural networks, we were able to extract important association signals for accurate prediction of different neurological disorders by using integrated gene expression datasets of VEEV, OPNA, and TBI samples. Gene ontology and pathway analyses further identified neuropathologic features with specific gene product attributes and functions, shedding light on the fundamental biology of these neurological disorders. Collectively, we highlight a workflow to analyze published transcriptomic data using machine learning, which can be used for both identification of gene biomarkers that are unique to specific neurological conditions, as well as genes shared across multiple neuropathologies. These shared genes could serve as potential neuroprotective drug targets for conditions like EEV, TBI, and OPNA.

KEYWORDS

machine learning, neurological disorder, EEV, TBI, OPNA exposure

Introduction

Neurological disorders caused by infectious diseases, chemical exposures, and physical trauma pose significant public health challenges and are critical concerns in military medicine worldwide. Venezuelan, eastern, and western equine encephalitis viruses (collectively referred to as EEV in this manuscript), organophosphorus nerve agents (OPNA), and traumatic brain injuries (TBI) are particularly significant and interrelated threats to both civilian populations and military personnel. Despite their distinct etiologies, these conditions share common features in their neurological manifestations and potential for severe long-term consequences (Ronca et al., 2017; Mouzon et al., 2012; VanderGiessen et al., 2024) suggesting possible overlapping molecular mechanisms that could be leveraged for therapeutic development.

EEVs represent a significant threat to both human and animal health throughout the Americas. Venezuelan equine encephalitis virus (VEEV), eastern equine encephalitis virus (EEEV), and western equine encephalitis virus (WEEV) belong to the genus Alphavirus in the family Togaviridae (Aguilar et al., 2011; Luethy, 2023) and are transmitted primarily through mosquitos, but have also been weaponized for use as potential bioweapons by both the US and Soviet Union. These viruses can cause severe neurological diseases, with mortality rates ranging from 1% (VEEV) to 70% (EEEV) (Ronca et al., 2016; Guzman-Teran et al., 2020; Zacks and Paessler, 2010). Previous research has demonstrated that these viruses cause systemic infection which is either asymptomatic, or presents as a mild flu-like illness in the acute phase of infection (Cain et al., 2024). Neuroinvasion occurs around day 4 post infection through the olfactory epithelium, but can also increase blood-brain barrier (BBB) permeability to enter the brain via transcytosis, but this is less thoroughly understood for EEEV and WEEV (Salimi et al., 2020; Honnold et al., 2015). Recent studies using animal models have revealed that VEEV infection triggers a cascade of inflammatory responses, including the activation of pro-inflammatory cytokines such as IL-1 β , TNF- α , and IFN- γ , which contribute to neuronal damage and subsequent neurological symptoms (Dahal et al., 2020). Despite significant advances in understanding their pathogenesis, current therapeutic options are limited to supportive care, with no specific antiviral treatments available.

Organophosphorus nerve agents (OPNAs) are highly toxic chemicals that interfere with the normal functioning of the nervous system. These compounds, including G-series agents (tabun, sarin, soman) and V-series nerve agents (VE, VG, VM, VR, VX), irreversibly inhibit acetylcholinesterase, leading to excessive accumulation of acetylcholine at synapses (Wang et al., 2022). Research over the past decades has revealed that OPNA toxicity extends beyond acute cholinergic crisis. Studies have shown that OPNA exposure initiates complex cellular cascades involving oxidative stress, neuroinflammation, and excitotoxicity (Aroniadou-Anderjaska et al., 2023). Long-term studies in animal models and human survivors have documented persistent neurological deficits, including cognitive impairment, anxiety, and depression (Francois et al., 2022; Vasanthi et al., 2023). Current treatment protocols rely

primarily on a combination of anticholinergic drugs (such as atropine), oximes for enzyme reactivation, and anticonvulsants (Worek et al., 2020; Newmark, 2019; Saalbach, 2023; Reddy, 2024). However, these treatments must be administered rapidly after exposure and may not prevent long-term neurological consequences. Recent research has focused on understanding the molecular mechanisms of delayed neurotoxicity and developing more effective neuroprotective strategies.

Traumatic brain injury (TBI) is a major global health concern, affecting an estimated 69 million people worldwide each year (Dewan et al., 2019). The spectrum of TBI ranges from mild concussions to severe injuries with devastating consequences. The pathophysiology of TBI involves both primary injury mechanisms (direct mechanical damage) and secondary injury (inflammatory cascades, BBB breakdown, hemorrhage) that can persist for months or years after the initial trauma (Howlett et al., 2022). Extensive research has identified key molecular pathways involved in TBI pathogenesis, including neuroinflammation, oxidative stress, excitotoxicity, and disruption of the blood-brain barrier. Recent studies have revealed the complexity of TBI's molecular signature, with altered expression of numerous genes involved in inflammation (e.g., IL-1 β , TNF- α), cell death pathways (e.g., caspase-3, BAX), and neuroplasticity (e.g., BDNF, NGF) (Ng and Lee, 2019; Freire et al., 2023). Advanced neuroimaging techniques combined with molecular studies have demonstrated that TBI triggers both acute and chronic changes in brain structure and function (Hu et al., 2022). Despite this growing understanding, therapeutic options remain limited, with most treatments focusing on symptom management rather than addressing the underlying pathological mechanisms.

Research on the pathogenesis of these conditions faces numerous challenges. For EEVs, the requirement for high-containment facilities and the complexity of virus-host interactions have limited comprehensive studies (Hughes et al., 2021). OPNA research faces similar chemical safety concerns, along with ethical considerations that restrict human studies. While TBI research has progressed more rapidly due to greater accessibility and established animal models (Bellotti et al., 2024), many aspects of its molecular pathology remain poorly understood. However, recent advances in high-throughput genomic technologies and bioinformatics approaches have opened new avenues for investigating these conditions through comparative analysis of gene expression profiles (Siddiqui et al., 2020; Siddiqui et al., 2020; Siddiqui et al., 2017; Siddiqui and Islam, 2016; Siddiqui et al., 2019). This allows us to leverage the extensive body of research in one area to inform the understanding of related conditions, potentially identifying common pathways and novel therapeutic targets. By implementing machine learning techniques, we can now integrate and analyze complex transcriptomic datasets from various experimental platforms, identifying both shared and condition-specific molecular signatures (Long et al., 2024; Martinez et al., 2022). Such an integrated approach not only provides insights into the fundamental biology of these neurological disorders but also has the potential to guide the development of medical countermeasures that could be effective across multiple conditions. Understanding the commonalities and differences in gene expression patterns among these disorders may reveal new therapeutic targets, ultimately leading to more effective interventions for affected individuals.

In this study, we used advanced computational methods to systematically analyze and compare the transcriptomic profiles associated with TBI, EEV infection, and OPNA exposure. Our analytical framework is able to (1) integrate gene expression data from diverse experimental platforms and conditions, overcoming the limitations posed by data variability among these brain disorders, and (2) identify shared and condition-specific molecular signatures that are associated with the three neuropathologies. Our findings in this study not only highlight potential common therapeutic targets but also reveal unique pathways that could guide the development of targeted interventions, offering new insights into the treatment of these complex neurological disorders.

Results

Workflow implemented in this study

In this study, we aimed to extract salient association signals from integrated gene expression datasets derived from VEEV, OPNA, and TBI samples, with the objective of enabling accurate prediction of diverse neurological disorders. To this end, we acquired and reanalyzed a collection of 6 datasets encompassing 395 samples related to VEEV, OPNA, and TBI, across various experimental conditions and organismal models. These datasets underwent a rigorous normalization and integration process to facilitate downstream analysis. Differential expression analysis was performed simultaneously for each condition to identify informative key genes exhibiting substantial expression variations relative to control samples. Subsequently, the expression matrix of these selected key genes was extracted from the integrated dataset and utilized as input for machine learning algorithms, enabling the identification of distinctive features associated with different neurological diseases (Figure 1).

Collection, normalization and integration of gene expression datasets for OPNA, EEV, and TBI derived from various platforms

To investigate the transcriptional characteristics of rodent brains injured by EEV, TBI or OPNA, we systematically searched and downloaded multiple datasets from NCBI's Gene Expression Omnibus (GEO) database, ensuring representation of each condition. The acquired data encompassed gene expression profiles for these disorders and their corresponding control samples, consisting of 395 samples across two mammalian species, mouse (Mus musculus) and rat (Rattus norvegicus) (Figure 2). As a prerequisite to downstream analysis, data normalization was conducted by identifying unique genes across all platforms used in the study. To create a comprehensive gene set, we compiled a list of all genes involved in each dataset and then obtained the intersection set. This process allowed us to identify orthologous genes and determine which ones were similar across species. To account for batch effects caused by diverse experimental designs across the different GEO datasets, we applied the ComBat tool from the pycombat package (Zhang et al., 2020), which corrects for artificial differences in the overall expression distribution of each sample by using Location and Scale (L/S) adjustments (Figure 2B). We then integrated samples across different diseases or species into a single expression matrix by using the unique gene symbol as the key, with genes in rows and samples in columns. The standardized dataset was ready for downstream analysis of differentially expressed genes (DEGs) and for training machine learning models.

Differential expression analysis between OPNA/EEV/TBI and control samples in rodent models

To investigate the characteristics of gene expression changes under different conditions, we implemented differential expression analysis between disease and control samples for each dataset separately. For the comparison between EEV/TBI/OPNA and their corresponding normal controls, we identified 2,542, 5,211, and 766 DEGs, respectively. Interestingly, in all three conditions, there appear to be more up-regulated DEGs than down-regulated DEGs (Figure 3A). To further explore the intrinsic connections of the gene expression changes among these three brain diseases, we compared the three up-and down-regulated DEGs to identify the common and distinct gene expression changes among different diseases. We found that most DEGs are disease-specific, in other words, the DEGs in all three diseases do not overlap significantly. For up-and down-regulated DEGs, only 45 and 1 DEGs were found to have simultaneous changes in all three diseases, respectively (Figure 3B).





FIGURE 2

Data collection and normalization in this study. (A) Publicly available datasets used in this study across species and diseases. (B) Gene expression levels across multiple datasets after ComBat normalization.



Differential expression analysis between disease and control samples. (A) Transformed volcano plot depicting differentially expressed genes (DEGs) between disease and control samples. (B) Overlap of DEGs among the three types of brain diseases studied. The top panel shows the number of upregulated DEGs, while the bottom panel displays the number of down-regulated DEGs. (C) GO ontology enrichment for the 45 shared up-regulated DEGs.

To analyze the functional consequences of these gene expression changes, gene functional enrichment analysis was conducted on both shared and disease-specific gene sets using the DAVID website (Sherman et al., 2022), by which an over-representation analysis was performed to identify the enriched biological processes of input gene set. As expected, the 45 shared up-regulated DEGs are significantly enriched in immune-and neuron-related functions, including "immune system process," "innate immune response," "inflammatory response," "synapse disassembly," and "positive regulation of angiogenesis" (Figure 3C). Moreover, we found a list of immunerelated and neuron-related processes enriched in disease-specific DEGs, suggesting that specific brain functions may be affected by different brain diseases (Supplementary Figure S1). Above all, our findings provide valuable insights into the molecular mechanisms underlying various brain pathologies and identify potential therapeutic targets for further investigation. The shared gene signature across multiple brain diseases suggests common pathways that could be targeted for broad-spectrum treatments, while the disease-specific signatures offer opportunities for developing targeted therapies for individual conditions.

Machine learning framework to identify key expression features for OPNA, EEV, and TBI

We combined the DEGs identified in each dataset, and from which selected orthologous genes across species to build an informative gene set for machine learning. A total of 2,525 genes were retained for downstream classifier training and prediction. Different machine learning models, including k-nearest neighbor (KNN), random forest (RF), linear discriminant analysis (LDA), support vector machine (SVM), and artificial neural network (ANN), were then employed for prediction purpose. The performance of these classifiers was evaluated by the misclassification rate (MCR), which is defined as the proportion of incorrect predictions, i.e.,

$MCR = \frac{\# \text{incorrect predictions}}{\# \text{total predictions}}.$

Under a training to test ratio of 4:1, we obtained the misclassification rate of the five classifiers (Table 1), which shows that the ANN classifier achieves the best performance for both ternary (across the three disorders to identify distinct features for each disorder) and binary (between disease and control to identify common features of the three disorders) classification (Figures 4A,B). We therefore chose ANN as a suitable machine learning model in this study for its outperformance among others.

We next used the ANN model to select a subset of feature genes that are most important for classification. For ternary and binary

TABLE 1 Misclassification rates of different machine learning models.

Classification	KNN	RF	LDA	SVM	ANN
Ternary (OPNA/ EEV/TBI)	0.083	0.125	0.486	0.097	0.024
Binary (Disease/ Control)	0.158	0.096	0.465	0.123	0.038

classifications, 238 and 252 genes were selected, respectively, and by using these selected feature genes, the ANN classifier was able to achieve the same misclassification rate as by using all genes. To ensure that the selected features are representative for disease in binary classification and for each disorder in ternary classification, we set between-group standard variation over 0.15 as an additional threshold to filter the selected features. For binary classification and ternary classification, 50 and 81 features were retained as the final expression features, respectively. These retained features showed strong expression representation in both binary and ternary classifications (Figures 4C,D). This provides strong evidence that the selected feature genes play a critical role in separating the classes, either across OPNA/ EEV/TBI or between disease and control.

Lastly, we examined the function of the selected feature genes. Functional enrichment analysis showed that both common and distinct features of the three brain disorders are significantly enriched in immune-related terms, such as "immune system process" and "innate immune response" (Figures 4E,F). Among these immunerelated genes, we identified a core set of upregulated genes across all three disorders compared with mock samples (Figure 4G). These genes included B2m (beta-2-microglobulin, crucial for MHC class I antigen presentation), Fosl2 (a key transcription factor in inflammatory responses), Map3k8 (a central regulator of inflammatory cytokine production), Rsad2 (an interferon-stimulated gene with antiviral properties), Tap1 (involved in antigen processing), Trim25 (a critical regulator of innate immunity), and Ube2l6 (involved in ISGylation). All these genes were consistently upregulated, suggesting a common inflammatory signature across these conditions. Among the features identified in ternary classification, we identified disorderspecific immune signatures (Figure 4H). OPNA samples showed elevated expression of C1qb (complement cascade initiator), Bst2 (type I interferon-induced antiviral protein), and Irf1 (interferon regulatory factor). EEV showed specifically upregulated Il7r (lymphocyte development regulator), Rnasel (viral RNA degradation), and Arg1 (immunosuppressive mediator in myeloid cells). TBI samples distinctively expressed Morc3 (nuclear protein involved in immune response) and Lgals9 (immunomodulatory galectin). In summary, through machine learning algorithms, we have successfully extracted the common and distinctive expression features underlying EEV, TBI, and OPNA.

Conclusions and discussion

In this study, we developed and implemented a comprehensive framework for analyzing and comparing transcriptomic profiles across three distinct neurological conditions: TBI, EEV, and OPNA. By leveraging machine learning approaches, particularly using artificial neural networks, we identified both shared and condition-specific gene signatures that provide valuable insights into the underlying molecular mechanisms of these neurological disorders. Our comparative analysis revealed several key findings. First, the integration of diverse transcriptomic datasets demonstrated the feasibility of cross-platform data normalization and analysis. Second, the machine learning models enabled the identification of critical genes associated with each condition, suggesting potential therapeutic targets for medical countermeasures. Third, the pathway and gene ontology analyses highlighted specific biological processes and



molecular functions that may play crucial roles in the pathogenesis of these neurological conditions.

Several limitations exist in this study. First, OPNA and EEV studies are highly limited, which limits our ability to have consistent tissue types in the data selected. Future studies would benefit from incorporating additional neurological conditions with more comprehensive data from the same species, tissue type, and background. Next, our analysis relies on microarray data, which provides only tissue-level expression profiles, and may overlook cell type-specific responses that are crucial for understanding the complex pathophysiology of neurological disorders. Additionally, microarray technology has inherent limitations in detecting novel transcripts and may identify fewer transcripts compared to more recent sequencing approaches. Future studies could address these limitations by incorporating single-cell RNA sequencing (scRNA-seq) data to identify cell type-specific responses to TBI, EEV infection and OPNA exposure and reveal cellular heterogeneity within affected tissues.

Regarding the specific findings of this study, the overwhelming similarities between EEVs, TBIs, and OPNAs are associated with the upregulation of the immune response. While this is an expected finding, it is challenging to utilize broad immune-related markers for biomarker analysis or therapeutics. Binary analysis of neurological disease phenotypes could be broadened to include other neuropathologies to identify signatures of gene expression that are indicators of damage in cases where the injury is unknown. Especially in the context of the warfighter, early markers of inflammation could be beneficial in distinguishing a healthy phenotype where immune-associated genes are lowly expressed from a recent injury or exposure to chemical or biological agents where these genes are upregulated. Further research is required to assess whether these biomarkers of disease in the brain also correlate with differences in the blood to make sample collection and testing realistic in the field. For the ternary analysis (EEV vs. TBI vs. OPNA), these results could be utilized to characterize types of injury. For example, in this analysis the gene Morc3 and Lgals9 appear to be upregulated in OPNA exposure and downregulated in EEVs. Therefore, these genes could be a feasible biomarker in clinical setting to distinguish whether an individual posing general malaise-like symptoms could have been exposed to a viral or chemical threat. This method of biomarker analysis has previously been used in the clinical context to

assess whether patient inflammation is associated with infection where supportive antimicrobial therapeutics are necessary, or an underlying disease state such as cancer, ischemia, or pulmonary embolism where administration of antimicrobial agents could worsen disease and in some cases be fatal (Alexandra Binnie and Dos Santos, 2019). As there are few comparable transcriptomic studies for EEVs and OPNA, further validations and incorporation of additional datasets are crucial for assessing the feasibility of transcriptional biomarker identification.

Our findings provide a prospect for future investigations into these neurological conditions and demonstrate the value of machine learning in understanding complex disease mechanisms. Looking forward, we expect to expand our research in the following two directions. First, we would like to incorporate single-cell RNA sequencing data to explore cellular heterogeneity and identify specific cell-type contributions to the pathogenesis of TBI, EEV infection, and OPNA exposure. This expansion will provide a more granular understanding of the molecular and cellular mechanisms underlying these disorders. Second, we will focus more on the transparency, accountability, and fairness of the machine learning models used in this study. Advanced explainable AI methods, such as LIME, SHAP, and Saliency Maps and Attention Mechanisms, will be compared and incorporated to enhance the interpretability and precision of our machine learning framework. These advancements will enable us to uncover deeper insights into disease-specific pathways and refine the identification of potential therapeutic targets. Together, these efforts will contribute to the development of more effective medical countermeasures and advance our understanding of these complex neurological conditions.

Methods

Data collection, normalization, and integration

The publicly available gene expression data were obtained from the National Center for Biotechnology Information (NCBI), under the Gene Expression Omnibus (GEO) database. A total of 395 samples were used in this study for investigating the transcriptional responses of different diseases such as EEV, TBI, and OPNA exposure. These datasets were obtained from various experiments across mammalian species, primarily rats and mice, as summarized in Table 2. Samples for TBI, OPNA, and EEV were selected based on data availability, and limited to the most consistent brain regions across different experimental conditions. Each dataset was filtered to remove knockouts, treatments, or additional treatments to ensure these data did not impact the results (Supplementary Table S1).

All the samples were integrated into one expression matrix with genes listed in rows and samples in columns, by merging genes with the same symbol in different samples or species. ComBat (Zhang et al., 2020) was adopted to normalize the data across platforms and experiments and adjust for batch effects. The normalized expression matrix, with dimension $6,289 \times 395$, was used for downstream analysis.

Differential expression analysis

The Limma package (Ritchie et al., 2015) in R was adopted to identify differentially expressed genes (DEGs) between the disease samples and control samples for EEV, OPNA, and TBI, respectively. An empirical-Bayes based method was used to determine expression difference and statistical significance, so that genes with fold-change over 1.5 and *p*-value less than 0.05 were identified as DEGs.

Gene functional enrichment analysis

Gene functional enrichment analysis was conducted using the DAVID website (Sherman et al., 2022), which implemented a hypergeometric test to evaluate the enrichment score, and the significance of the input gene sets in certain gene ontology (GO) terms. The GO terms with p-value less than 0.05 were determined as the over-represented biological processes.

Machine learning on transcriptomic datasets

Different machine learning models, such as KNN, RF, LDA, SVM, and ANN, were used to predict disease classes based on the normalized expression matrix. These models differ in various aspects, including flexibility, training complexity, scalability, interpretability, etc. In this study, since the goal is to accurately predict samples into disease classes and identify key expression features associated with the prediction, we focus on the ANN model, particularly deep learning ANN, as it achieves lowest misclassification rate compared to the other models (Table 1). To enhance model efficiency, only DEGs obtained by contrasting the disease and control samples were included in the models, reducing the dimension from 6,289 to 2,525. Both ternary and binary classifications were implemented to compare prediction accuracy across various conditions, where the former used OPNA, EEV, and TBI as responses and the latter used disease (by combining samples from three disorders) and control (by combining the corresponding control samples) as responses. Specifically, for KNN classification, we set k = 3; for RF classification, we used 1,000 decision tree; and for SVM classification, we chose radial basis kernel. In ANN classification, the entire dataset was split into training and testing sets, according to a ratio of 4:1. The ANN model consists of three hidden layers, with the number of nodes in each layer specified according to the geometric pyramid rule (Masters, 1993). Model performance was evaluated subsequently using the misclassification rate.

Besides prediction, ANNs were also used to identify feature genes that are shared or unique among different disease classes. The basic idea is to remove or mask specific features and observe the impact on model performance (i.e., MCR) (Kavzoglu and Mather, 2002). If removing a feature significantly degrades performance, it suggests the feature plays a key role in prediction and should be selected. Specifically, such a feature selection was achieved by heuristically searching the 2,525-dimensional feature space along certain path to

Data	Data Type	Disease Type	Brain Location	Species	Title	PMID
GSE91074	Microarray	EEV	Whole Brain	Mouse	Gene expression in the brains of different strains of laboratory mice upon intranasal infection with vaccine strain (TC83) of Venezuelan equine encephalitis virus	28184218
GSE96550	Microarray	EEV	Whole Brain	Mouse	Differential host gene responses from infection with neurovirulent and partially-neurovirulent strains of Venezuelan equine encephalitis virus	28446152
GSE128543	Microarray	TBI	Cortex	Mouse	Identification of Novel Targets of RBM5 in the Healthy and Injured Brain	32335213
GSE131435	Microarray	TBI	Hippocampus	Rat	TBI weight-drop model with variable impact heights differentially perturbs hippocampus-cerebellum specific transcriptomic profile.	33172833
GSE13428	Microarray	OPNA	Hippocampus	Rat	Gene Expression Profiling of Rat Hippocampus Following Exposure to the Acetylcholinesterase Inhibitor Soman	19281266
GSE28435	Microarray	OPNA	Piriform cortex	Rat	Transcriptomic Analysis of Rat Brain Following Exposure to the Organophosphonate Anticholinesterase Sarin	21777429

TABLE 2 Resource of datasets used in this study

find the most parsimonious model that attains comparable or slightly higher MCR to the full model. The search path was determined by calculating the leave-one-out model MCR and ranking the marginal effect of each gene.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

LY: Data curation, Formal analysis, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. MV: Data curation, Resources, Writing – review & editing. VK: Data curation, Writing – review & editing, Software. BC: Data curation, Software, Writing – review & editing. P-CC: Writing – review & editing. MT: Conceptualization, Investigation, Project administration, Resources, Supervision, Writing – review & editing. EJ: Conceptualization, Investigation, Project administration, Resources, Supervision, Writing – review & editing. KK-H: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. XW: Data curation, Formal analysis, Methodology, Supervision, Writing – original draft, Writing – review & editing. HX: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. XW: Data curation, Formal

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This project was supported

References

Aguilar, P. V., Estrada-Franco, J. G., Navarro-Lopez, R., Ferro, C., Haddow, A. D., and Weaver, S. C. (2011). Endemic Venezuelan equine encephalitis in the Americas: hidden under the dengue umbrella. *Future Virol.* 6, 721–740. doi: 10.2217/fvl.11.50

Alexandra Binnie, J.L., and Dos Santos, Claudia C, (2019). How can biomarkers be used to differentiate between infection and non-infectious causes of inflammation? Evid. Based Pract. Crit. Care, doi: 10.1016/B978-0-323-64068-8.00055-9. [Epub ahead of print].

Aroniadou-Anderjaska, V., Figueiredo, T. H., de Araujo Furtado, M., Pidoplichko, V. I., and Braga, M. F. M. (2023). Mechanisms of organophosphate toxicity and the role of acetylcholinesterase inhibition. *Toxics* 11:866. doi: 10.3390/toxics11100866

Bellotti, C., Samudyata, S., Thams, S., Sellgren, C. M., and Rostami, E. (2024). Organoids and chimeras: the hopeful fusion transforming traumatic brain injury research. *Acta Neuropathol. Commun.* 12:141. doi: 10.1186/s40478-024-01845-5

Cain, M. D., Klein, N. R., Jiang, X., Salimi, H., Wu, Q., Miller, M. J., et al. (2024). Postexposure intranasal IFNalpha suppresses replication and neuroinvasion of Venezuelan equine encephalitis virus within olfactory sensory neurons. *J. Neuroinflammation* 21:24. doi: 10.1186/s12974-023-02960-1

Dahal, B., Lin, S. C., Carey, B. D., Jacobs, J. L., Dinman, J. D., van Hoek, M. L., et al. (2020). EGR1 upregulation following Venezuelan equine encephalitis virus infection is regulated by ERK and PERK pathways contributing to cell death. *Virology* 539, 121–128. doi: 10.1016/j.virol.2019.10.016

Dewan, M. C., Rattani, A., Gupta, S., Baticulon, R. E., Hung, Y. C., Punchak, M., et al. (2019). Estimating the global incidence of traumatic brain injury. *J. Neurosurg.* 130, 1080–1097. doi: 10.3171/2017.10.JNS17352

by the Defense Threat Reduction Agency (HDTRA1-23-1-0009). The content of the information does not necessarily reflect the position or the policy of the federal government, and no official endorsement should be inferred.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom.2025.1529902/ full#supplementary-material

Francois, S., Mondot, S., Gerard, Q., Bel, R., Knoertzer, J., Berriche, A., et al. (2022). Long-term anxiety-like behavior and microbiota changes induced in mice by sublethal doses of acute sarin surrogate exposure. *Biomedicines* 10:1167. doi: 10.3390/biomedicines10051167

Freire, M. A. M., Rocha, G. S., Bittencourt, L. O., Falcao, D., Lima, R. R., and Cavalcanti, J. R. L. P. (2023). Cellular and molecular pathophysiology of traumatic brain injury: what have we learned so far? *Biology (Basel)* 12:1139. doi: 10.3390/biology12081139

Guzman-Teran, C., Calderón-Rangel, A., Rodriguez-Morales, A., and Mattar, S. (2020). Venezuelan equine encephalitis virus: the problem is not over for tropical America. *Ann. Clin. Microbiol. Antimicrob.* 19:19. doi: 10.1186/s12941-020-00360-4

Honnold, S. P., Mossel, E. C., Bakken, R. R., Lind, C. M., Cohen, J. W., Eccleston, L. T., et al. (2015). Eastern equine encephalitis virus in mice II: pathogenesis is dependent on route of exposure. *Virol. J.* 12:154. doi: 10.1186/s12985-015-0385-2

Howlett, J. R., Nelson, L. D., and Stein, M. B. (2022). Mental health consequences of traumatic brain injury. *Biol. Psychiatry* 91, 413–420. doi: 10.1016/j.biopsych.2021.09.024

Hu, L., Yang, S., Jin, B., and Wang, C. (2022). Advanced neuroimaging role in traumatic brain injury: a narrative review. *Front. Neurosci.* 16:872609. doi: 10.3389/fnins.2022.872609

Hughes, H. R., Velez, J. O., Davis, E. H., Laven, J., Gould, C. V., Panella, A. J., et al. (2021). Fatal human infection with evidence of Intrahost variation of eastern equine encephalitis virus, Alabama, USA, 2019. *Emerg. Infect. Dis.* 27, 1886–1892. doi: 10.3201/eid2707.210315

Kavzoglu, T., and Mather, P. M. (2002). The role of feature selection in artificial neural network applications. *Int. J. Remote Sens.* 23, 2919–2937. doi: 10.1080/01431160110107743

Long, Q., Zhang, X., Ren, F., Wu, X., and Wang, Z. M. (2024). Identification of novel biomarkers, shared molecular signatures and immune cell infiltration in heart and kidney failure by transcriptomics. *Front. Immunol.* 15:1456083. doi: 10.3389/fimmu.2024.1456083

Luethy, D. (2023). Eastern, Western, and Venezuelan equine encephalitis and West Nile viruses. Vet. Clin. N. Am. Equine Pract. 39, 99–113. doi: 10.1016/j.cveq.2022.11.007

Martinez, B. A., Shrotri, S., Kingsmore, K. M., Bachali, P., Grammer, A. C., and Lipsky, P. E. (2022). Machine learning reveals distinct gene signature profiles in lesional and nonlesional regions of inflammatory skin diseases. *Sci. Adv.* 8:eabn4776. doi: 10.1126/sciadv.abn4776

Masters, T. (1993). Practical neural network recipies in C++.

Mouzon, B., Chaytow, H., Crynen, G., Bachmeier, C., Stewart, J., Mullan, M., et al. (2012). Repetitive mild traumatic brain injury in a mouse model produces learning and memory deficits accompanied by histological changes. *J. Neurotrauma* 29, 2761–2773. doi: 10.1089/neu.2012.2498

Newmark, J. (2019). Therapy for acute nerve agent poisoning: an update. *Neurol. Clin. Pract.* 9, 337–342. doi: 10.1212/CPJ.00000000000641

Ng, S. Y., and Lee, A. Y. W. (2019). Traumatic brain injuries: pathophysiology and potential therapeutic targets. *Front. Cell. Neurosci.* 13:528. doi: 10.3389/fncel.2019.00528

Reddy, D. S. (2024). Neurosteroids as novel anticonvulsants for refractory status epilepticus and medical countermeasures for nerve agents: a 15-year journey to bring ganaxolone from bench to clinic. *J. Pharmacol. Exp. Ther.* 388, 273–300. doi: 10.1124/jpet.123.001816

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007

Ronca, S. E., Dineley, K. T., and Paessler, S. (2016). Neurological sequelae resulting from encephalitic alphavirus infection. *Front. Microbiol.* 7:959. doi: 10.3389/fmicb.2016.00959

Ronca, S. E., Smith, J., Koma, T., Miller, M. M., Yun, N., Dineley, K. T., et al. (2017). Mouse model of neurological complications resulting from encephalitic alphavirus infection. *Front. Microbiol.* 8:188. doi: 10.3389/fmicb.2017.00188

Saalbach, K.P., (2023). Therapeutic treatment of nerve agent toxicity, in sensing of deadly toxic chemical warfare agents, nerve agent simulants, and their toxicological aspects.

Salimi, H., Cain, M. D., Jiang, X., Roth, R. A., Beatty, W. L., Sun, C., et al. (2020). Encephalitic alphaviruses exploit Caveola-mediated Transcytosis at the blood-brain barrier for central nervous system entry. *MBio* 11, e02731-e02719. doi: 10.1128/mBio.02731-19

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., et al. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* 50, W216–W221. doi: 10.1093/nar/gkac194

Siddiqui, M. K., Islam, M. Z., and Kabir, M. A. (2019). A novel quick seizure detection and localization through brain data mining on ECoG dataset. *Neural. Comput. & Applic.* 31, 5595–5608. doi: 10.1007/s00521-018-3381-9

Siddiqui, M. K., Huang, X., Morales-Menendez, R., Hussain, N., and Khatoon, K. (2020). Machine learning based novel cost-sensitive seizure detection classifier for imbalanced EEG data sets. *Int. J. Interact. Des. Manuf. (IJIDeM)* 14, 1491–1509. doi: 10.1007/s12008-020-00715-3

Siddiqui, M.K., and Islam, M. Z., (2016). "Data mining approach in seizure detection." in IEEE. p. 3579–3583.

Siddiqui, M. K., Islam, M. Z., and Kabir, M. A. (2017). "Analyzing performance of classification techniques in detecting epileptic seizure," in *Advanced data mining and applications. ADMA 2017. Lecture notes in computer science, vol 10604.* eds. G. Cong, W. C. Peng, W. Zhang, C. Li, and A. Sun (Cham: Springer).

Siddiqui, M. K., Morales-Menendez, R., Huang, X., and Hussain, N. (2020). A review of epileptic seizure detection using machine learning classifiers. *Brain Inform.* 7:5. doi: 10.1186/s40708-020-00105-1

VanderGiessen, M., de Jager, C., Leighton, J., Xie, H., Theus, M., Johnson, E., et al. (2024). Neurological manifestations of encephalitic alphaviruses, traumatic brain injuries, and organophosphorus nerve agent exposure. *Front. Neurosci.* 18:4940. doi: 10.3389/fnins.2024.1514940

Vasanthi, S. S., Rao, N. S., Samidurai, M., Massey, N., Meyer, C., Gage, M., et al. (2023). Disease-modifying effects of a glial-targeted inducible nitric oxide synthase inhibitor (1400W) in mixed-sex cohorts of a rat soman (GD) model of epilepsy. *J. Neuroinflammation* 20:163. doi: 10.1186/s12974-023-02847-1

Wang, J., Lu, X., Gao, R., Pei, C., and Wang, H. (2022). Current Progress for retrospective identification of nerve agent biomarkers in biological samples after exposure. *Toxics* 10:439. doi: 10.3390/toxics10080439

Worek, F., Thiermann, H., and Wille, T. (2020). Organophosphorus compounds and oximes: a critical review. Arch. Toxicol. 94, 2275–2292. doi: 10.1007/s00204-020-02797-0

Zacks, M. A., and Paessler, S. (2010). Encephalitic alphaviruses. Vet. Microbiol. 140, 281-286. doi: 10.1016/j.vetmic.2009.08.023

Zhang, Y., Parmigiani, G., and Johnson, W. E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform.* 2:lqaa078. doi: 10.1093/nargab/lqaa078