

<https://doi.org/10.1038/s44328-025-00035-0>

AI-QuIC machine learning for automated detection of misfolded proteins in seed amplification assays

Kyle D. Howey^{1,2}, Manc Li^{1,2}, Peter R. Christenson^{1,2}, Peter A. Larsen^{2,3} & Sang-Hyun Oh^{1,2} ✉

Seed Amplification Assays (SAAs) detect misfolded proteins associated with neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, ALS, and prion diseases. However, current data analysis methods rely on manual, time-consuming, and potentially inconsistent processes. We introduce AI-QuIC, an artificial intelligence platform that automates analyzing data from Real-Time Quaking-Induced Conversion (RT-QuIC) assays. Using a well-labeled RT-QuIC dataset comprising over 8000 wells, the largest curated dataset of its kind for chronic wasting disease prion seeding activity detection, we applied various AI models to distinguish true positive, false positive, and negative reactions. Notably, the deep learning-based¹ Multilayer Perceptrons (MLP) model achieved a classification sensitivity of over 98% and specificity of over 97%. By learning directly from raw fluorescence data, the MLP approach simplifies the data analytic workflow for SAAs. By automating and standardizing the interpretation of SAA data, AI-QuIC holds the potential to offer robust, scalable, and consistent diagnostic solutions for neurodegenerative diseases.

Advancements in AI, in particular deep learning¹, have revolutionized numerous scientific fields, including life sciences and disease diagnostics. Pioneering developments in protein structure prediction algorithms, such as AlphaFold² and RosettaFold^{3,4} have demonstrated the impact of AI on predicting protein folding, with significant potential implications for accelerated drug discovery. These breakthroughs have set the stage for applying AI to tackle challenges in neurodegenerative diseases characterized by protein misfolding.

Early detection of misfolded proteins is crucial for diagnosing and managing neurodegenerative diseases such as Alzheimer's, Parkinson's, and amyotrophic lateral sclerosis (ALS). Seed amplification assays (SAAs), such as protein misfolding cyclic amplification (PMCA)⁵, quaking-induced conversion (QuIC)^{6–8}, and real-time QuIC (RT-QuIC)⁹, have emerged as powerful tools for detecting pathogenic proteins in various neurodegenerative diseases^{10–16}. SAAs cyclically amplify protein misfolding to enable the rapid conversion of a large excess of the monomeric substrate into prion-like amyloid fibrils with minimal quantities of seeds formed by misfolded proteins^{8,17}. These assays recapitulate the seeding-nucleation mechanisms of protein misfolding and help determine whether samples contain detectable (positive) or undetectable (negative) levels of misfolded proteins⁵. Examples of misfolded proteins detected by SAAs include prions in prion diseases (e.g., Creutzfeldt-Jakob Disease)^{6,9,14,18–20}, amyloid-beta²¹ and tau^{10,11,22} in

Alzheimer's disease, alpha-synuclein in synucleinopathies^{12,15,23,24}, and TAR DNA-binding protein 43 in ALS¹³. By achieving ultrasensitive detection of misfolded proteins, SAAs hold great potential for early diagnosis and prognosis of such disorders¹⁴.

While SAAs have broad applications across multiple neurodegenerative diseases in humans, their development has benefited significantly from prion diseases in animals, especially Chronic Wasting Disease (CWD) that naturally occurs in cervids (e.g., deer, elk, and moose)^{25–32}. Therefore, CWD is considered a robust model for studying protein misfolding in RT-QuIC.

Sharing the fundamental principles of SAAs, RT-QuIC assays are performed in multi-well plates with 4–8 technical replicates/reactions per sample⁹. The plate undergoes cyclic shaking and incubation while heated to facilitate fibril fragmentation and promote seeding-nucleation, resulting in the exponential amplification of misfolded proteins⁹ (Fig. 1a). ThT, a rotor dye, exhibits enhanced fluorescence upon binding to growing amyloid fibrils in the reaction mixture due to restricted rotational motion, enabling real-time monitoring of amyloid formation^{9,33}. Several methods have been applied to RT-QuIC data to establish predictable values using various metrics^{28,34,35} (Fig. 1a), but it is unclear whether there is unutilized information in the raw fluorescence measurements. The complexity of interpreting amplification signals is compounded by factors such as the concentration of misfolded proteins, strain conformation, and cross-

¹Department of Electrical and Computer Engineering, College of Science and Engineering, University of Minnesota, Minneapolis, MN, USA. ²Minnesota Center for Prion Research and Outreach, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, USA. ³Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, USA. ✉e-mail: sang@umn.edu

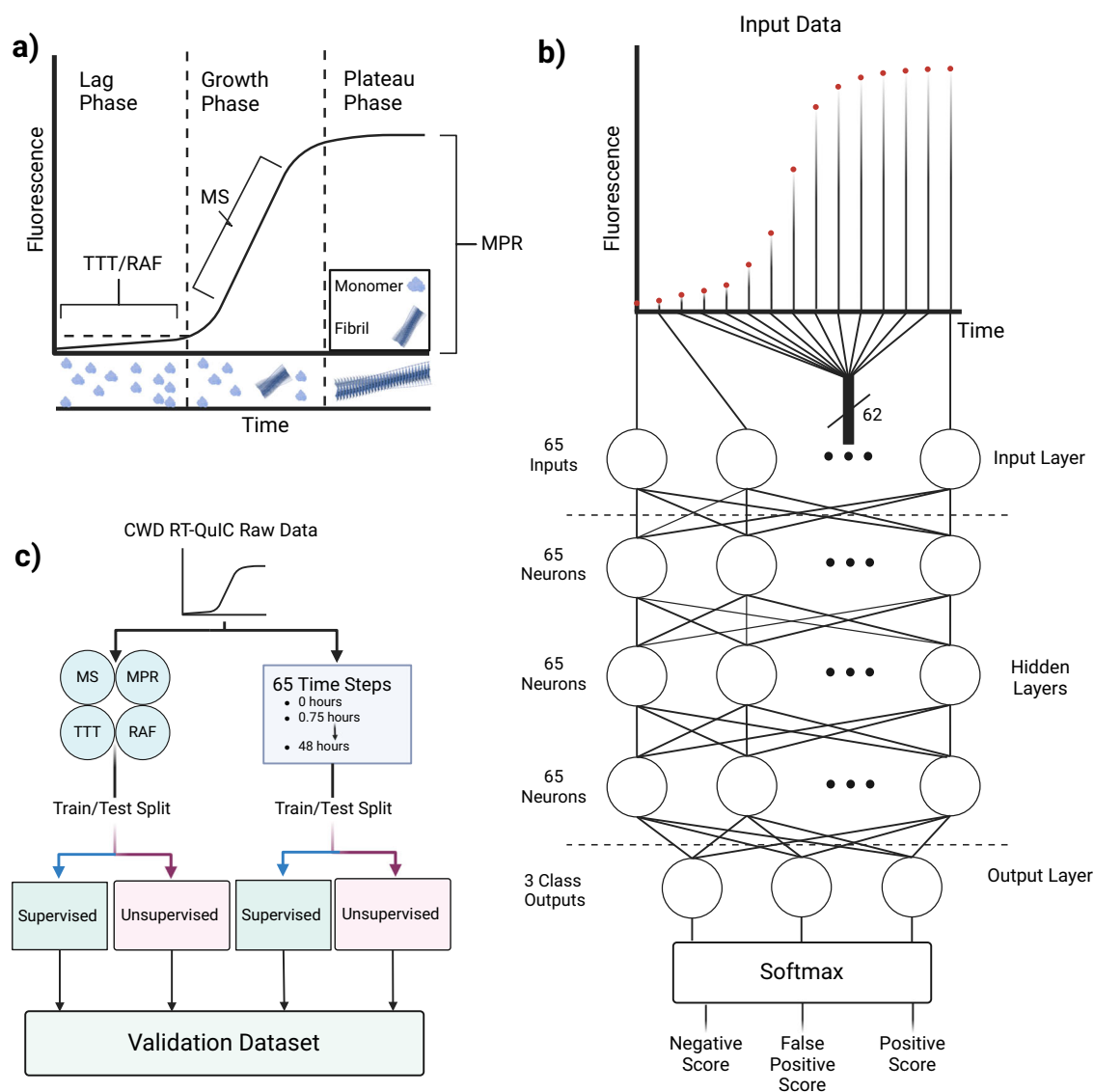


Fig. 1 | Visualization of key concepts. **a** Example graph of the RT-QuIC data from a single reaction with the source of Time to Threshold (TTT), Rate of Amyloid Formation (RAF), Max Slope (MS), and Max Point Ratio (MPR) highlighted. The different phases of the reaction are identified with a visual of how the monomers

form a fibril, increasing the fluorescence. **b** Diagram of the multilayer perceptron used in this study with 65 input features. The network also includes 3 hidden layers, each with 65 neurons and a single output layer with 3 classes. **c** Flowchart of how the data is processed in the study.

seeding^{34,36,37}. In short, processes for the optimization of data interpretation and standardization across different assay platforms and disease-associated proteins remain unexplored.

Drawing conclusions (positive vs. negative for seeding activity) from RT-QuIC reactions requires expertise by trained personnel, often requiring nuanced and subjective knowledge that is hard to formalize and articulate in human language or as a written manual³⁸. AI is uniquely suited for such challenges as it can replicate and capture expert knowledge, identify complex patterns, and make accurate predictions, thereby enhancing the reliability and efficiency of RT-QuIC assays. Integrating AI into these processes bridges gaps in human expertise and enables robust, scalable testing operations.

The fluorescence measurements at each time point in an RT-QuIC reaction form an RT-QuIC curve, which can be used directly or summarized into derived metrics to serve as features or input parameters for training ML models. Many AI models exist, each with a unique approach to carrying out such tasks. For example, clustering classifiers such as K-Means are a subset of unsupervised AI models that identify patterns in data and assign labels associated with those patterns^{39,40}. These can be used while considering

whether an RT-QuIC reaction has seeding activity (positive) or not (negative), making them invaluable for use with large, unlabeled RT-QuIC datasets, as verifying labels can be time-consuming. More complex and supervised models, namely Support Vector Machines (SVMs) and Multi-layer Perceptrons (MLPs), can be trained to identify positive RT-QuIC reactions on labeled data, potentially with better performance^{41,42}. Investigating AI-driven automation and standardization techniques for interpreting SAA data, particularly from RT-QuIC assays, holds great promise in further enhancing the already significant impact of SAAs on the detection and diagnosis of neurodegenerative diseases.

The objective of this proof-of-concept study is to demonstrate the potential of AI in detecting the seeding activity of misfolded proteins and to explore the effectiveness of various AI models including deep learning models, such as MLPs (Fig. 1b), in analyzing data from SAAs, using RT-QuIC data from CWD prions as a representative example. In particular, we focused on the automated determination of whether seeding activity induced by CWD prions exists in individual RT-QuIC reactions. We first analyzed a well-labeled, previously generated dataset using various AI models to assess the potential of AI in distinguishing true positive, false

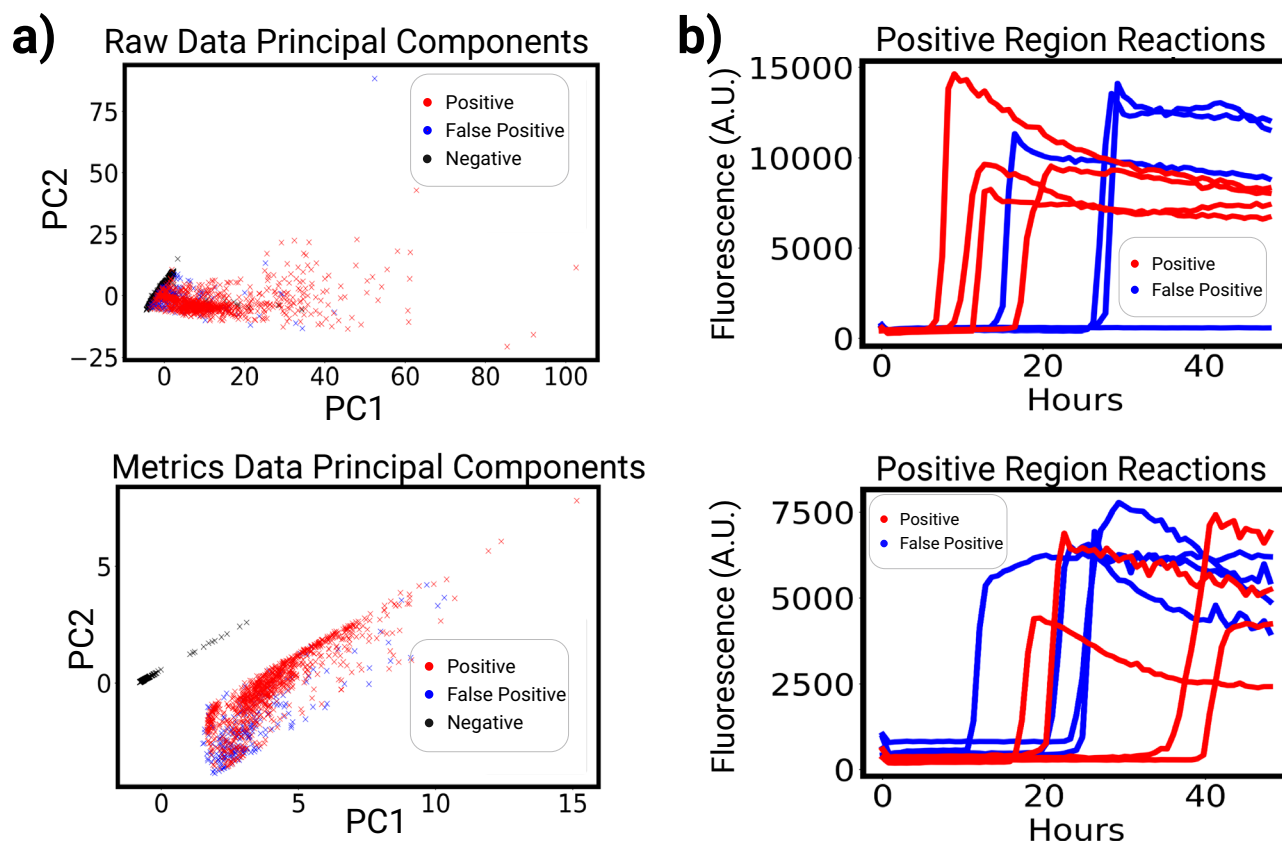


Fig. 2 | Demonstration of useful attributes captured by PCA. **a** Class Separation Achieved by PCA. A PCA was run on the raw data and metrics data and generated principal components (PCs) corresponding to linear combinations of the input features. The figures demonstrate how these PCs create distinctions in the dataset

which can be used with AI. **b** Edge-case Examples. The PCA plots from a) were used to select the most positive-like false positives for each feature extraction method and compare with clearly defined positive samples. This demonstrates any likenesses which may make it difficult for PCA to create meaningful distinctions.

positive, and negative RT-QuIC reactions for CWD prions. We then evaluated and compared the performance of various AI models, including a widely used unsupervised algorithm (K-Means) and supervised models (SVM and MLP) (Fig. 1c), on both summarized metrics and raw fluorescence data from RT-QuIC assays. Next, the best performing models were applied to a more complex subset of the data not used during training, assessing whether AI can identify patterns in RT-QuIC data that are not readily discernible using current methods. The final evaluation consisted of testing the models on an independent validation dataset to determine the generalizability of the models.

Results

Extraction of early reaction features and validation of existing metrics by PCA

The PCA algorithm provided key insights into how the data can be distilled into only the most vital components. This information can then be used to generate insights into how the data is structured and, in the case that the identified variance relates to classification, can provide visuals of how clusters form.

This visual can be demonstrated in the form of a scatter plot (Fig. 2) for both the raw data and the metrics data. In the case of the raw data, these two PCs account for nearly 80% of the variance in themselves, representing two of the four PCs used to train the raw K-Means and SVM models. In the case of the metrics data, these two PCs account for over 90% of the variance in the dataset, and both represent linear combinations of the calculated metrics. In this case, it becomes clear that using all four of these metrics was unnecessary as TTT and RAF contain much of the same information since they are simply reciprocals of one another. As such, only 3 PCs are required to represent the dataset in most cases. In addition to these insights, the two scatter plots also

demonstrate the capability both datasets possess to meaningfully separate the data. Although there is overlap (particularly between the false positive and positive reactions), some distinction can be inferred between data points with different class identities. This is particularly interesting in the case of the metrics data, as this clear distinction indicates that the metrics (a significantly simpler dataset than its raw data counterpart) are sufficient to provide meaningful distinctions between classes. This validates the use of these existing metrics as a viable way of classifying RT-QuIC data. It should be noted that all the negative samples appear to be closer to the positive samples than the false positives. This arises from the decision to set TTT for samples that never reached the threshold to a hard 0. PC1 and PC2 both depend on TTT, and positive reactions will, in general, take less time to reach the threshold than false positive reactions resulting from false seeding. Combining these PCs, then, causes this distinction to become apparent.

A few edge-case examples of positive and false positive reactions can also be plotted (Fig. 2), providing a visualization of any reactions that may not be straightforward to classify. These edge-cases were identified from examination of the scatter plots and selection of data from each of these two classes, which were furthest in the apparent positive region. Ultimately, these visualizations show the difficulty inherent to this dataset in the false positive reactions. Many of these reactions are difficult to distinguish from true positive reactions. Models with decent degrees of accuracy on these reactions, then, represent a powerful, standardized tool for handling edge cases without needing human annotation.

K-Means metrics performed well for unsupervised RT-QuIC classification

Among classification algorithms, perhaps the most interesting and challenging to apply are unsupervised learning algorithms. Able to pick out

Table 1 | K-Means Performance Metrics (Higher is Better)

Train Type	Sensitivity	Specificity	Precision	F1-Score	Negative Support	Positive Support
Raw	0.89	0.99	0.38	0.53	1444	160
Metrics	1.00	0.97	0.78	0.87	1444	160

Table 2 | SVM Performance Metrics (Higher is Better)

Train Type	Sensitivity	Specificity	Precision	F1-Score	Negative Support	Positive Support
Raw	0.94	0.93	0.59	0.72	1444	160
Metrics	0.93	0.98	0.84	0.88	1444	160

Table 3 | MLP on Raw Data Metrics (Higher is Better)

Train Type	Sensitivity	Specificity	Precision	F1-Score	Negative Support	Positive Support
Raw	0.98	0.97	0.80	0.88	1444	160

patterns in unlabeled data, unsupervised algorithms provide a critical foundation in understanding the dynamics of a dataset, allowing us to determine if the variances PCA and metrics-based feature extraction identifies are meaningful to classifying RT-QuIC reactions. The patterns the algorithms select are not directed as in supervised learning, so good performance of these models indicates a dataset in which the predominant patterns are the ones pertaining to the classification of RT-QuIC reactions as positive or negative.

K-Means. The model performance metrics for the PCA feature-extracted dataset and for the manually feature-extracted dataset were calculated (Table 1). For the raw data, the results represent an accuracy of 84%. For the feature-extracted data, the results represent an accuracy of 97%.

The K-Means model trained on the raw data was consistently poor and unstable, indicating data that was not cleanly separable into distinct clusters. The K-Means model trained on metrics, however, achieved high accuracy and excellent sensitivity. This suggests that the metrics provided a more useful basis for extracting salient features for the K-Means model to use.

Supervised learning achieved high accuracy for RT-QuIC classification

Supervised algorithms, having access to labels during training, provide an important point of comparison against unsupervised models for learning patterns in datasets. These algorithms excel at finding patterns that support class identities specified by the operator, making them more adept at classifying more complex datasets.

Support Vector Machine. The SVM results represent 93% accuracy on the raw data as well as 98% accuracy on the metrics data (Table 2).

The SVM performed very well on the testing data and with minimal training time, with the metrics-trained model outperforming the raw data/PCA approach in most metrics. The high accuracy of this model, compounded with the highest precision of any of the models evaluated in this study, makes this a compelling choice. The sensitivity is also the lowest out of all the models (with the exception of the raw data K-Means model), however, incurring a cost for the high precision.

Multilayer Perceptron. The metrics produced by the MLP were calculated and represent an accuracy of 97% (Table 3).

Requiring less preprocessing than the SVMs and K-Means, the MLP benefited from a simpler implementation (feature-engineering-free) while still achieving high accuracy and specificity. This positions the MLP as an effective and adaptable model that is not significantly affected by specific implementation. The MLP appeared to be an excellent all-around candidate

for classification, maintaining comparatively high precision without a major sacrifice to sensitivity. Leveraging its high performance, compounded with its versatile implementation, the MLP represents the capability of deep learning to be broadly applied without the need for expert-dependent feature engineering. This makes the MLP and other deep learning techniques prime candidates for the classification of SAA data generated using other neurodegenerative diseases without the need for application-specific feature extraction.

Summary. The supervised learning methods used in this study represent a useful alternative to the unsupervised learning methods presented previously. Both the SVM trained on metrics and the MLP achieved relatively higher precision than the unsupervised models, making them more useful in applications where negative wells greatly outnumber positive wells. Particularly in the case of the MLP, deep learning was able to identify useful relationships in the dataset with a relatively simple implementation and achieved excellent all-around results. This highlights the capability of DNNs to extract salient features from data without the need for expert feature engineering, making these networks powerful choices for datasets that have not been well characterized.

A comparative analysis of unsupervised and supervised learning for RT-QuIC classification revealed key performance differences

Unsupervised Overview. The confusion matrices (Fig. 3a) and ROC curves (Fig. 3b) provide greater insight into how the performance of the models varies under different conditions. The K-Means trained on the metrics data is the best performer of the two, never missing a positive sample in the entire testing dataset. Additionally, the 3.6% of negative samples it classified as positive are made up entirely of false positives. This model never successfully classified a false positive, and the false positives are the only reactions the model misclassified. The example curves (Fig. 3c) also support the finding in the PCA analysis that these false positive curves are not always trivial to classify.

Supervised Overview. Examining the supervised model results, the first important detail is how the different training approaches for the SVMs affected the outcome. The SVM trained on raw data with the PCA-based feature extraction was more balanced between specificity and sensitivity, but faced an overall performance decrease that the metrics-trained SVM did not exhibit (Fig. 4). Comparing this metrics-trained SVM to the MLP, there is no clear better performer as each performed well with different strengths and weaknesses. In the case of the MLP for example, while the results were more balanced than the metrics-based SVM approach, the slightly worse performance on correctly identifying negative reactions

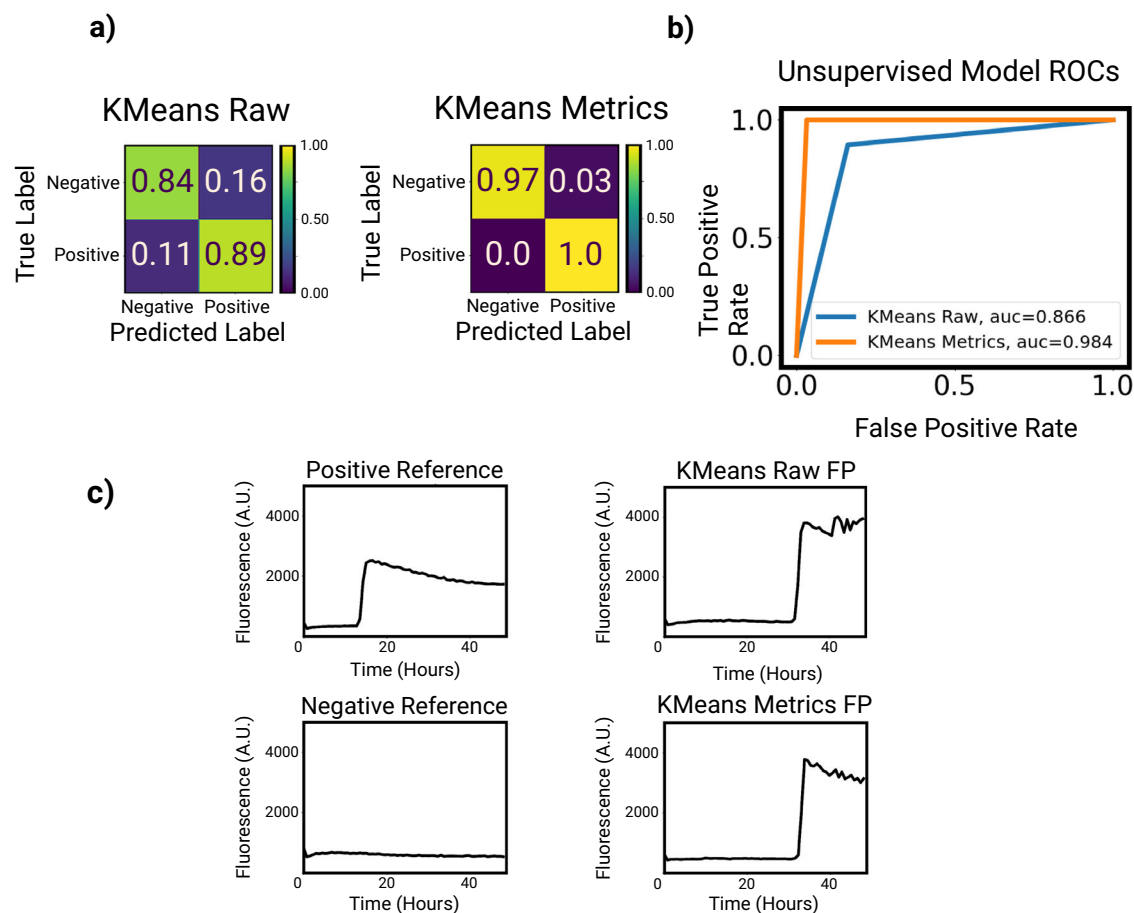


Fig. 3 | Comparison of K-Means model results. **a** Model confusion matrices represent the ratio of how many samples from each true class were classified in each predicted class. **b** The Receiver Operator Characteristic (ROC) illustrates the sensitivity-specificity tradeoff of each model, with an ideal model having an area

under the curve (AUC) of 1. **c** The examples include a single false positive a given model misclassified along with correctly classified reference samples, highlighting how distinguishable a misclassified false positive was.

represents a much larger portion of the dataset as negative reactions are more common in general (Figs. 4, 5). This higher prioritization of correct classification of positive samples, however, favors the MLP in the ROC evaluation (Figs. 4b, 5b). As was the case with the unsupervised models, the example plots demonstrate the challenge of classifying some of the false positive samples (Figs. 4c, 5c).

Performance on Dataset with Label Ambiguity. In addition to the 554 samples that made up the training and testing dataset for the models, 19 samples with ambiguous labels were tested separately. This subset of the dataset consists of swab samples where human annotators, using a plot of the replicates overlayed for each swab, classified the swab sample differently than what was suggested by statistics. In particular, each of these samples was identified as positive by a human annotator, but determined to be negative by statistical analysis in the original study⁴³. As a point of comparison, the three best-performing models were selected (the K-Means and SVM models trained on metrics as well as the MLP) and tested on these samples.

Each sample dilution was annotated by the K-Means and SVM models trained on metrics, the MLP, and a human annotator (ML) for comparison (Table 4). Examining these reactions, it becomes clear that machine learning represents a far superior method of classification to other statistical methods of evaluation. The excellent agreement between the models and human annotation represents a major improvement and demonstrates the capability of AI to limit the need for human annotation, providing a standardized, automated approach with an accuracy approaching that of humans.

Models performance on external validation data shows generalizability

In addition to the testing and evaluation performed on a subset of the original dataset, each model was applied to an external validation dataset^{27,44,45} to highlight generalizability. The results of this evaluation (Fig. 6) demonstrate the capabilities of the different models to generalize to this new dataset. With the exception of the KMeans model trained on raw data, each model performed generally well on this dataset. One important distinction arises when comparing the supervised and unsupervised models, however. Each of the unsupervised models generalized poorly, losing significant performance on this validation dataset. The supervised models all performed similarly, however, with the raw data SVM achieving the highest specificity while the MLP and SVM trained on metrics both achieving similar performance with near perfect sensitivity. With the smaller size of this dataset, the results are insufficient to pick a single, statistically best-generalized model. The supervised models, however, were far superior at generalizing than the unsupervised models.

Discussion

This proof-of-concept study demonstrated the promising potential of AI in enhancing the interpretation and automation of RT-QuIC data, using Chronic Wasting Disease (CWD) as a robust model for protein misfolding disorders. The results addressed two primary objectives: first, our comparative analysis of unsupervised (K-Means) and supervised (SVMs, MLPs) AI models unveiled their unique strengths in processing both summarized metrics and raw fluorescence data; second, we established that AI models can effectively identify seeding activity in RT-QuIC reactions. Our findings

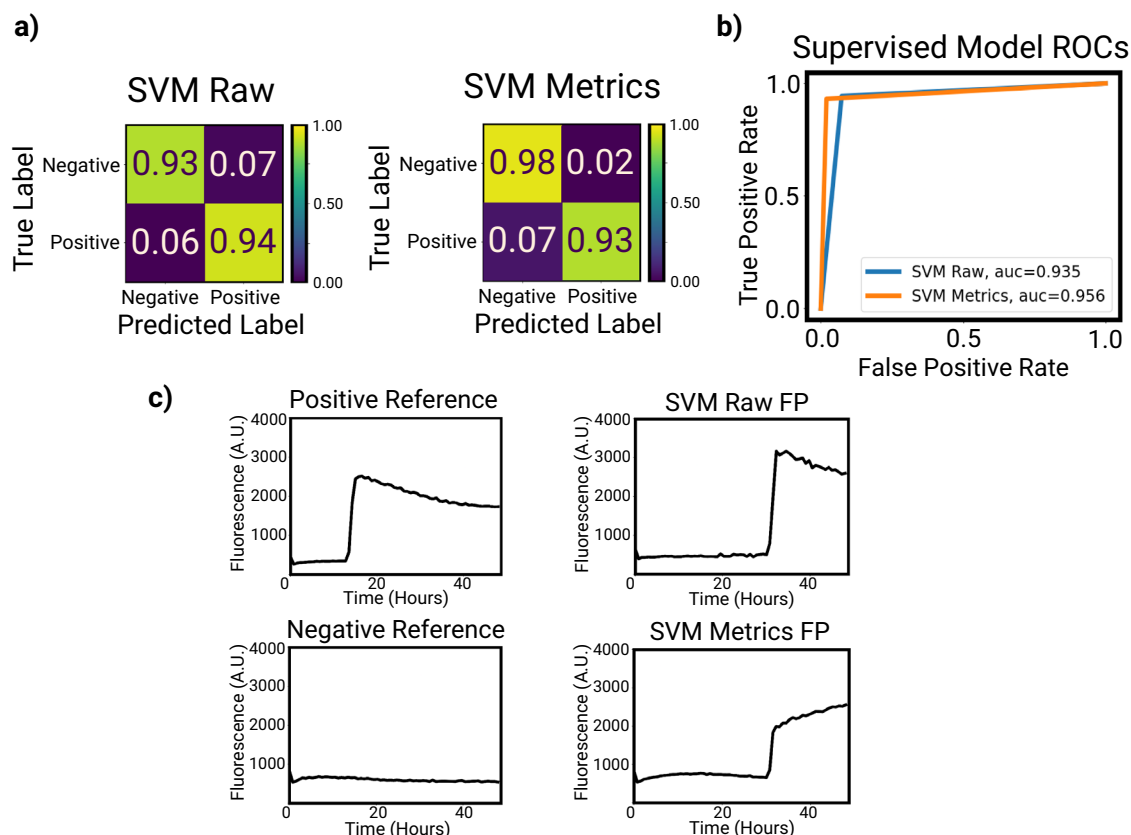


Fig. 4 | Comparison of Support Vector Machine (SVM) model results. a Model confusion matrices represent the ratio of how many samples from each true class were classified in each predicted class. **b** The Receiver Operator Characteristic (ROC) illustrates the sensitivity-specificity tradeoff of each model, with an ideal

model having an area under the curve (AUC) of 1. **c** The examples include a single false positive a given model misclassified along with correctly classified reference samples, highlighting how distinguishable a misclassified false positive was.

Table 4 | Replicate Reactions in Agreement with Human Annotation

	SVM Metric	MLP	K-Means Metrics
Reactions Agreed with Human annotation	304/304	292/304	304/304

demonstrate that AI models can effectively automate the detection of misfolded protein seeding activity in RT-QuIC assays. Notably, the deep learning-based MLP approach performed similarly to or better than other models by leveraging its inherent ability to extract relevant features automatically, directly learning from raw data without requiring dimensionality reduction techniques like PCA. This eliminates the need for expert-dependent and time-consuming feature engineering, streamlining the analysis process. The MLP model possesses adaptability to diverse datasets and assay conditions. By reducing reliance on data preprocessing, the MLP approach offers a more efficient and scalable solution for standardizing SAA data interpretation across various neurodegenerative diseases. Collectively, our study lays the groundwork for AI-driven enhancements in RT-QuIC data analysis and the development of AI-assisted diagnostic tools for a spectrum of neurodegenerative disorders characterized by protein misfolding.

In examining the capabilities of the various models and comparing performance between the two datasets, a pattern of support emerges for the metrics selected to describe the curves. Firstly, it is worth acknowledging the apparent similarities between the PCA-selected features and the chosen metrics. The analysis of the PCA (Fig. 2) demonstrates evidence that machine learning and human intuition for curve description are both

capable techniques for clustering data. As is the case with nearly all machine learning applications, the exact patterns identified by the algorithm do not perfectly line up with human-identified patterns and concepts. The other primary goal of this study was to provide a method of reliably automating curve identification. Identifying a curve as positive or negative by the human eye can be time-consuming and inconsistent among labs and technicians. As such, AI represents a powerful solution, being highly adaptable and capable of identifying key patterns without significant technician time. While a few models stood apart as particularly effective at classifying the data in this study, nearly all of them demonstrate the viability of this approach, obtaining high accuracy similar to that of a technician. In addition to this high accuracy, the models are all highly reproducible and consistent. Once the dataset has been generated and the code written, these models can each classify tens of thousands of samples in seconds.

Artificial intelligence (AI), particularly machine learning (ML), has shown remarkable effectiveness in identifying patterns that hold predictive value. Our study utilized the dataset of over 8000 RT-QuIC reactions, the largest of its kind to contain both human annotations and metrics commonly used on RT-QuIC data, providing unprecedented depth for analysis⁴³. With research efforts involving more comprehensive datasets, various misfolded proteins, and the development of AI models specifically tailored for SAA data, these models could potentially match or surpass human experts in distinguishing positive from negative samples, overcoming current challenges such as strain differences, cross-seeding, and concentration effects that could hinder accurate diagnosis in clinical settings. Future integration of AI with SAAs holds the potential to enhance diagnostic accuracy, streamline workflows, and ensure consistent interpretation across various assays and laboratory conditions, mitigating discrepancies in interpretation among personnel. The implications of our

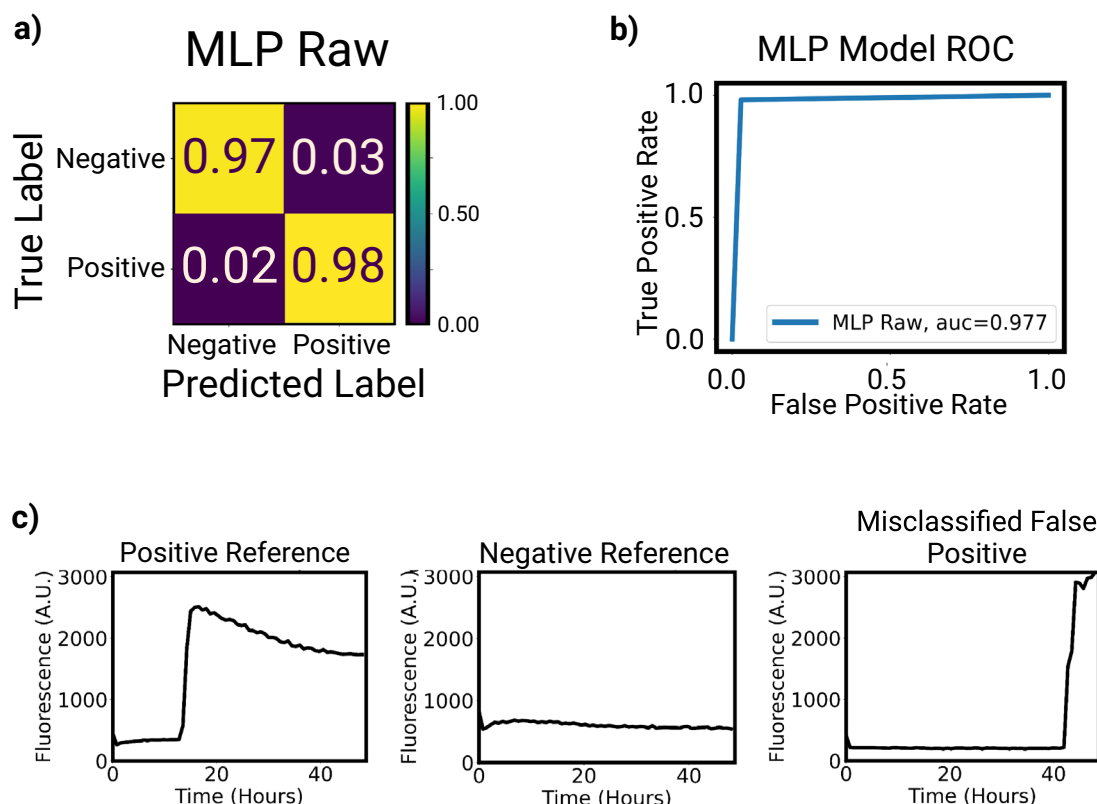


Fig. 5 | Multilayer Perceptron (MLP) model results. **a** Model confusion matrix represents the ratio of how many samples from each true class were classified in each predicted class. **b** The Receiver Operator Characteristic (ROC) illustrates the sensitivity-specificity tradeoff of the model, with an ideal model having an area

under the curve (AUC) of 1. **c** The examples include a single false positive given model misclassified along with correctly classified reference samples, highlighting how distinguishable a misclassified false positive was.

findings extend beyond CWD and RT-QuIC to other SAAs, such as PMCA^{21,23}, HANABI⁴⁶, Nano-QuIC^{31,47}, Cap-QuIC⁴⁸, MN-QuIC³², which face similar challenges in data interpretation. AI-driven approaches offer the potential for standardization and automation across a spectrum of neurodegenerative diseases that have already shown seeding activity detectable by SAAs, including Alzheimer's^{10,21,49,50}, Parkinson's^{23,51}, as well as ALS and frontotemporal dementia^{13,52}.

Looking forward, we propose the continued development of AI-QuIC to integrate diverse data types, such as image and spectral data and accommodate a wider range of assay parameters, enhancing its applicability to numerous neurodegenerative diseases. In addition, forecasting techniques and in-depth time-series analysis could lay the foundation for the reduction of runtimes of the RT-QuIC assay, increasing speed and accuracy in an all-inclusive diagnostic framework. While the utilization of AI-QuIC on RT-QuIC data in this study is in its early stages, it represents a significant advancement toward the early detection and diagnosis of neurodegenerative diseases. By harnessing the pattern-recognition capabilities of AI - particularly the streamlined, feature-engineering-free approach of the deep learning MLP model - AI-QuIC effectively captures the dynamics of protein misfolding during SAA assays, enabling automated and standardized data interpretation. Just as AI has revolutionized protein folding modeling, the integration of AI into SAA analysis shows the potential to enhance diagnostic workflows and contribute to drug discovery pipelines. Recent breakthroughs, such as the cryo-electron microscopy determination of the atomic structure of misfolded prions⁵³ have unveiled the structures of misfolded protein aggregates, providing new insights into their pathogenic mechanisms. These structural data also offer training data for AI to better predict not only normal protein folding but also misfolding processes. As AI continues to advance in modeling both protein folding and misfolding, it

may facilitate the identification of new drug targets for protein misfolding diseases. By combining AI-enhanced diagnostics, such as AI-QuIC, with AI-based modeling of protein misfolding, we can accelerate our understanding of neurodegenerative disorders and our ability to diagnose them at an early stage.

Methods

Data Generation and Processing

The dataset was sourced from Milstein, Gresch, et. al.⁴³, a study that tested different disinfectants on CWD and consisted of the RT-QuIC analysis of swab samples after disinfection of various surfaces. The portion of the dataset used for this study comprised 8028 individual wells, representing 573 samples, each prepared with 2-3 dilutions and assayed in 4-8 technical replicates/reactions. Data from a total of 7331 sample reactions and 984 control reactions were used to evaluate the ML models. Among the 573 samples, 19 were withheld due to ambiguous labels, being classified as negative by statistical tests but positive by a "blind" human evaluation. These samples were further reviewed to exclude those with complex identities for additional characterization. The final dataset used for training and initial testing of the models included 7027 sample wells and 984 control wells. Each of these wells was characterized by taking a fluorescence reading 65 times at 45-minute intervals over a 48-hour run. Considering this sourcing, the ground-truth CWD positive/negative status of each sample was not directly known due to the introduction of disinfectant and the indirect nature of the swabbing method used to generate the RT-QuIC data (Fig. 7). The "true" sample identities were instead determined by experienced researchers using statistical assessments, providing an excellent opportunity to compare the performance of AI solutions against existing methods for identifying CWD prions in RT-QuIC data.

Fig. 6 | Model results on external validation data. Model confusion matrices represent the ratio of how many samples from each true class were classified in each predicted class. The data used to generate these confusion matrices was generated from studies^{28,45,46} independent from the rest of the swab dataset analyzed in Figs. 3–5.

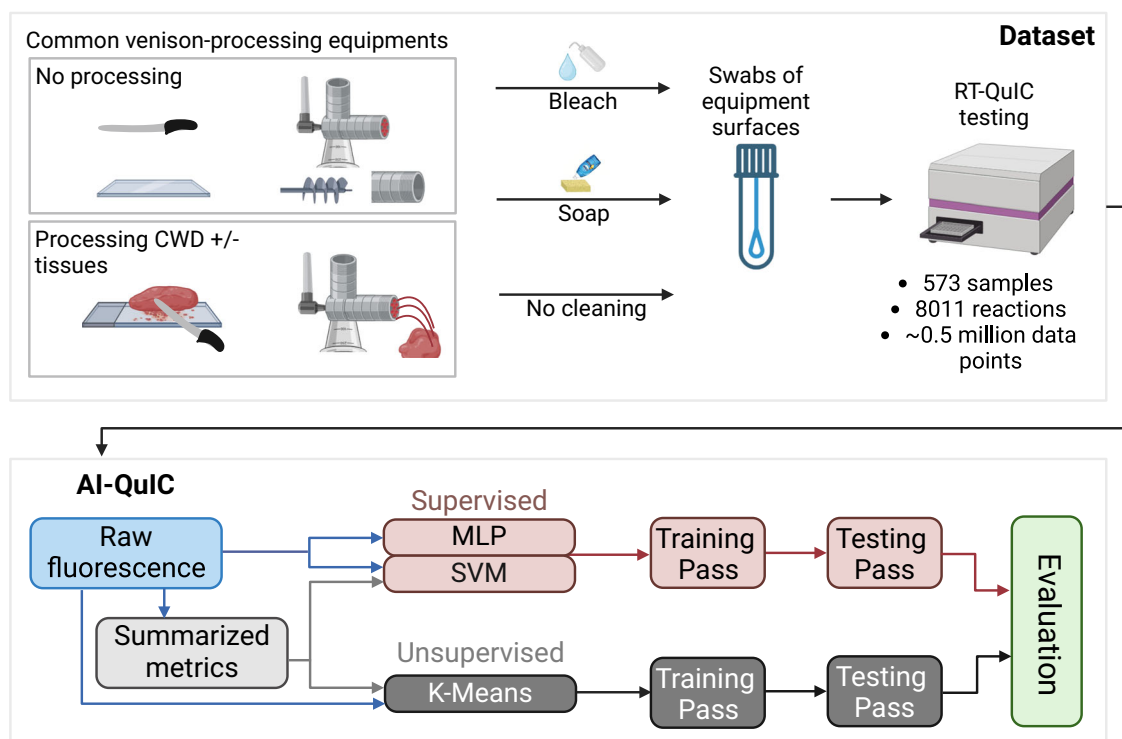
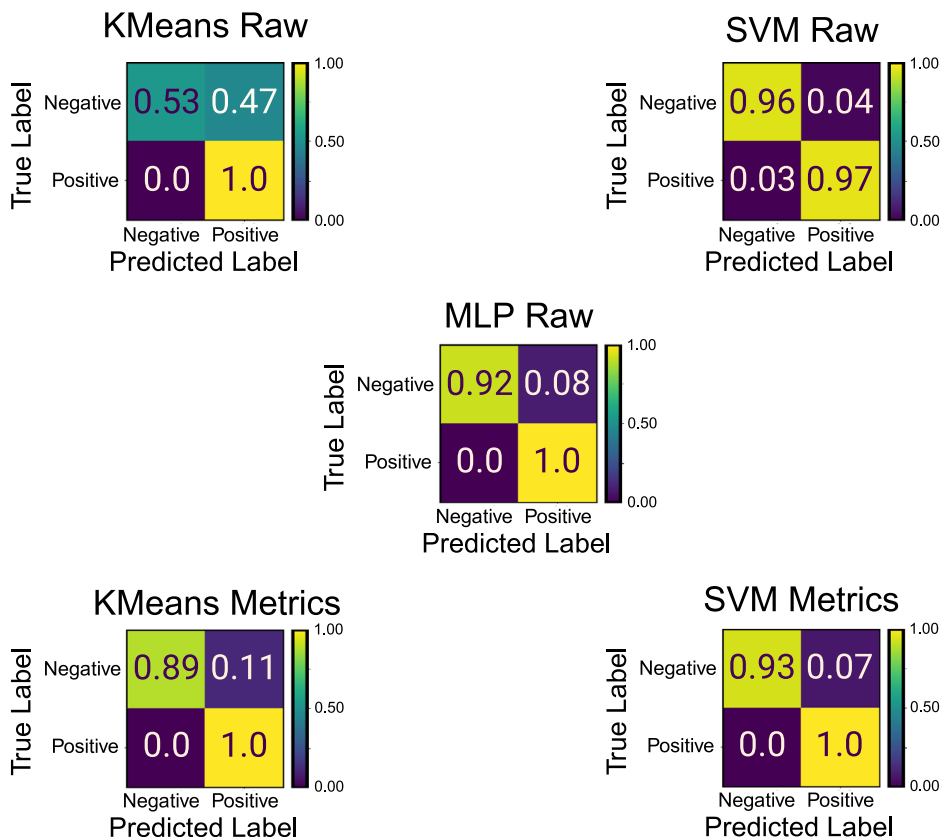


Fig. 7 | Visualization of data generation and processing. Each reaction is obtained from the RT-QuIC analysis of CWD positive/negative samples obtained through the experimental design process outlined in the Milstein, Gresch, et. al. paper⁴³. The fluorescence readings from this analysis are then processed into summarized metrics

to be used as a comparative dataset. Both datasets are split into a training and testing set. These datasets are used respectively to train and test the models, with the testing data providing the basis for evaluation.

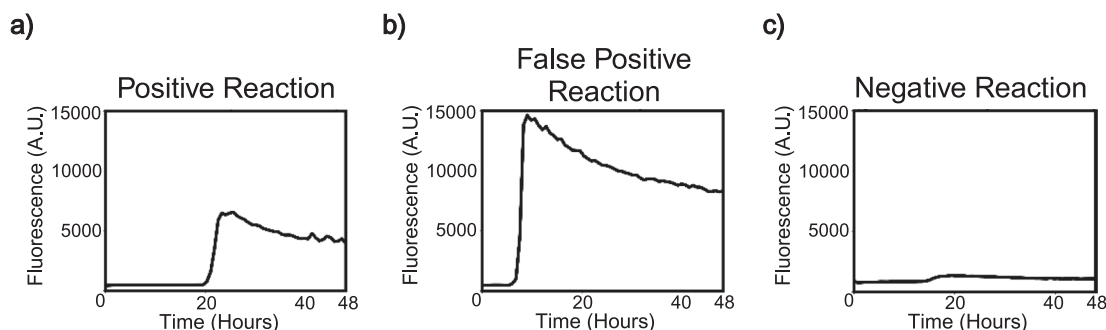


Fig. 8 | Example reaction plots. **a** A single positive annotated reaction plot, **b** a single false positive annotated reaction plot, and **c** a single negative annotated reaction is shown in the three plots. The plots demonstrate an example visualization of the raw

data used in this study and allow for an intuitive comparison of reactions with different class identities.

Three labels were introduced for this study based on the results of each RT-QuIC reaction: negative, false positive, and true positive. Each of these labels is related to sample identity and the threshold defined in the original study⁴³. A negative label represents a reaction involving a negative or a positive sample that never crossed the threshold (i.e. was identified as negative by analysis of the RT-QuIC results). A false positive label indicates a reaction with a CWD negative sample that crossed the threshold after being identified as a negative sample in the Milstein, Gresh, et. al. study⁴³. A true positive label was identified as a reaction involving a CWD positive sample that crossed the positive threshold after being considered as positive in the original study⁴³. In this dataset, there were 6781 negative reactions, 239 false positive reactions, and 991 positive reactions.

This formation of the dataset imposes a few limitations to the utility of the models evaluated in this study. Firstly, without knowing the identity or “ground truth” of the samples, these models can only be compared against the labels of annotators and, in the case of the supervised models, will be trained with any implicit biases held by the annotators. Additionally, the ambiguity in sample identity also limits the results; rather than classifying a reaction or sample as a whole, this method is limited to the classification of each replicate individually based on its annotation. While this individual replicate identity, or well-level identity, can inform sample identity, this limitation prevents the model from being an all-inclusive diagnostic tool. Next, there is additional classification difficulty in the extreme similarity between some false positive samples and true positive samples. A false positive RT-QuIC reaction is characterized as a reaction that was identified as positive but is known to be CWD negative, resulting from false seeding. As such, many of the false positive reactions were difficult to distinguish from true positive reactions (Fig. 8). Despite these limitations, the models still learned meaningful patterns which could be invaluable to future sample-level identification and serve as a proof of concept for models that could potentially surpass current annotator/metrics-based classification methods.

In addition to this dataset, a smaller, external validation dataset was used to assert the validity of these methods by testing them on data generated completely independently from the training dataset. This data featured 122 negative reactions, 15 false positive reactions, and 31 positive reactions, with 156 wells being sourced from samples and 12 being control reactions. For this dataset, the samples were derived directly from tissue samples (including ear, muscle, and blood), making the labels more reliable^{27,44,45}.

Every method chosen for this study attempts to contribute to two goals: (1) to evaluate classification methods based on prevailing knowledge on the analysis of RT-QuIC data, and (2) to create a new method with the automation capabilities that machine learning represents. These two goals were supported by two datasets—a dataset of precalculated metrics and a raw fluorescence dataset (Fig. 1c). The first of these datasets, in line with the first goal of the study, consisted of four features extracted from the raw fluorescence values of each well. The feature set included the time to a

precalculated threshold (TTT), the rate of amyloid formation (RAF), the maximum slope of the fluorescence curve (MS), and the maximum point ratio (MPR) between the smallest and largest fluorescence readings. These features were selected for their ability to characterize the curves RT-QuIC generates and for their relevance to the current method of using a precalculated threshold for classifying a test result as positive or negative. RAF and TTT are reciprocals of one another, with TTT measuring the “time required for fluorescence to reach twice the background fluorescence”⁴³ and RAF being the rate at which this threshold is met. Both were included in the metrics dataset to give the feature selection algorithms used in this study the freedom to select whatever combination of these it deems best. The inclusion of RAF also gave models an indication of the dilution factor of the sample in the reaction as the two features are linearly related³⁴. Additionally, reactions that did not reach the threshold were set to a TTT of 0 as this value is not physically realizable and works better for scaling than choosing some other large value that would need to depend on the runtime of the assay. MS was calculated with a sliding window with a width of 3.75 hours or 6 cycles. By finding the difference between the fluorescence reading at a current point and one taken 3.75 hours previously, the entire curve can be characterized into slopes. The maximum of these is selected. MPR was determined by dividing the maximum fluorescence reading by the background fluorescence reading⁴³. In contrast to this first, distilled dataset, the second dataset consisted of the raw fluorescence readouts, preserving all the information obtained from a particular run for the algorithms to extract. Once the two datasets were generated in the aforementioned formats, a few universal preprocessing steps were applied to ensure the data was ready for machine learning.

Firstly, RT-QuIC data in its raw form is not well suited to machine learning due to the wide variabilities in the absolute fluorescence. Different machines and concentrations, for example, can cause fluorescence curves to reach higher in some plates than in others, making it difficult to derive meaningful patterns in the larger dataset. The Scikit-Learn (sklearn) Standard Scaler⁵⁴ was used to remedy this, centering the means of each well at 0 and scaling the dataset to unit variance. This was done separately to the training and testing partitions of the data. Secondly, as this study focused on the behavior of individual wells, it was crucial to remove any patterns in the data that could be derived from the ordering of the dataset. To account for this, the samples were shuffled into a random order using tools in the Numpy package⁵⁵.

Creation of training/testing sets

All of the models utilized distinct training and testing sets to ensure the models could generalize beyond the training data. To accomplish this, 20% of the dataset was selected at random and withheld from training to be used in the evaluation of the models. The separation was done at the sample level to ensure different reactions from the same sample were not present in both the training and testing sets. The model predictions on this testing dataset are the source of the evaluation metrics and plots generated for each model.

All of the models were configured to use the same data for training and testing to allow for direct comparison.

Principal component analysis

Before training two of the model types, SVM and K-Means, a Principal Component Analysis (PCA) was applied to the raw dataset. The PCA implementation used in this study was a part of the sklearn package⁵⁴. PCA is a linear dimensionality reduction method that extracts relationships in a dataset and creates a new set of features based on these relationships. The algorithm works by identifying linear combinations of the different features in the dataset, matching each variable with every other variable and assigning a score based on how related they are to one another. This score is called the covariance. These scores are stored in a matrix that can be used to calculate eigenvectors, which transform the original features into a set of linear combinations of features. These linear combinations are called Principal Components (PCs). The eigenvalues of these eigenvectors, then, represent the overall variance/feature importance of each individual feature. PCA can then select the features with the largest eigenvalues to represent the dataset⁵⁶. Typically, the number of PCs to incorporate into the final dataset is identified by using a scree plot and identifying a point at which adding additional features has greatly reduced improvement in overall variance (also called finding the “elbow”). Using this method on the PCs obtained from the raw data, it was determined that 4 or 5 PCs were all that was necessary to represent the dataset. 4 PCs were selected from this set to be used in training some of the models as this aligned with the 4 features of the metrics dataset and captured nearly 90% of the variance in the dataset. Both K-Means and SVMs train poorly on data with large numbers of features (such as the raw data with 65 timesteps used in this study), making PCA an effective way to give these algorithms an assist. Additionally, applying PCA enhances the variance in the data, making the training of K-Means and SVM more efficient. In contrast, deep learning models, like the MLP, inherently handle high-dimensional data through their robust feature extraction capabilities¹. Therefore, we did not apply PCA when training the MLP model, allowing it to learn directly from the raw data.

Unsupervised Learning Approaches

K-Means Clustering. A K-Means model was trained on both the raw dataset and on the dataset of metrics, yielding a total of two models. KMeans is a powerful clustering utility and one of the most famous unsupervised learning algorithms. The algorithm works by identifying cluster centers through various methods and adjusting these locations/scales to optimally separate the dataset into distinct sets. The KMeans algorithm evaluated in this study was developed using the Scikit-Learn package in Python and used Lloyd’s KMeans Clustering Algorithm (LKCA)⁵⁷. This algorithm uses an iterative optimization approach to approximate cluster centers for data with arbitrary numbers of features, making it more practical for multidimensional datasets like RT-QuIC reactions than manually calculated approaches to clustering. However, like many machine learning algorithms, LKCA frequently finds solutions that are only locally optimal. Practically, this means that the solution KMeans will find to a clustering task is highly influenced by the initial conditions of the algorithm. In order to generate these initial conditions, the model was set to use a random initialization, selecting the features of two random samples as cluster centers, which were iteratively adjusted by LKCA to find an optimal solution. LKCA was allowed to run up to 500 iterations or until the algorithm had converged on a solution - whichever required less time. Due to the high initialization sensitivity of KMeans, the algorithm was configured to select 150 random samples to use as initial cluster centers, selecting the best version. The goal was to greatly increase the odds that KMeans selected two samples, one which nearly epitomizes a negative sample and one which nearly epitomizes a positive sample. Then, with the optimization of LKCA, this increases the likelihood of finding a globally optimal solution. Considering the high efficiency of the KMeans algorithm, this process takes little time. The raw

data was passed through a PCA algorithm to highlight the variance in the data and reduce dimensionality.

The model was configured to find three clusters, with the idea that one cluster would represent positive RT-QuIC reactions, another would represent false positive reactions, and the final cluster would represent negative reactions. As most of the analysis used to evaluate the models was not compatible with multiclass output, the model output was converted to yield a 0 or a 1, corresponding to negative or positive. False positive classifications were treated as negative except when evaluating performance on false positives. Since the model was not given labels, it randomly selected whether the positive label was a 0 or a 1. The rest of the evaluation required a uniform system of outputs, so a standard was selected (1 represented positive, and 0 represented negative). Any model that underperformed below what was predicted by random guessing had its labels flipped to match this standard.

Supervised learning approaches

Support vector machines. While SVMs are elegant in their simplicity, they often suffer from poor performance on datasets with many features (as in our case with 65 time steps). Considering this, the SVM trained on raw data was considered a prime candidate for the PCA preprocessing step, which was applied to the raw dataset via the Scikit-Learn package⁵⁴ prior to training and testing. This created a much smaller feature space of just the relationships for the SVM to learn. The SVM was constructed using the Scikit-Learn implementation⁵⁴. The model itself uses a radial basis function (RBF)⁵⁸ kernel. RBFs are a family of functions that are symmetric about a mean, such as a Gaussian distribution. These functions allow SVMs to classify data that is not separable by a polynomial function (such as concentric circles), making the SVM more versatile but also more complex⁵⁸.

SVMs implemented with Scikit-Learn are inherently binary classifiers and do not support multiple classification boundaries. While it is possible to implement multiple classes by training multiple models, the implementation in this study did not use that approach in order to simplify the training and testing of the model. The PCA/SVM combo was run on the raw data and the SVM alone was run on the feature-extracted data.

Deep learning: multilayer perceptron. Multilayer Perceptrons (MLPs) are one of the simpler types of Deep Neural Networks¹ (DNNs) developed for classification tasks. MLPs consist of a set of 1-dimensional layers of neurons, featuring an input layer, an output layer, and at least one hidden layer in between. The weights between each layer are used to identify patterns that correspond to the training task⁴².

The MLP was selected for this task due to its natural compatibility with the dimensions of the problem. Since the study examined each replicate individually, the data was one-dimensional, consisting of just the fluorescence reading at each of the 65 time steps. MLPs are able to compare each timestep with every other timestep, allowing them to extract a wide range of features from the dataset, even when these features are not linearly separable by class.

The MLP that produced the results in this paper used the Keras package in Python. Keras is a machine learning package that makes model development simpler and is part of the Tensorflow package⁵⁹. The model consisted of 3 hidden layers. The input layer matched the input shape (65 nodes, one for each timestep). The next three hidden layers used 65 neurons each, matching the shape of the input, which proved effective in testing. Each of these layers used a Rectified Linear Unit (ReLU) activation function, a popular activation that eliminates negative inputs completely⁶⁰. The model, while originally tested as a binary classification setup, is configured with 3 output neurons, each corresponding to a class (negative, false positive, or positive). This forces the MLP to learn distinctions between positive and false positive classes in a more meaningful way. The model uses a softmax activation function in the final layer, converting the input to this layer into a score from 0 to 1 for each of the three classes. These scores add up to 1 and can be seen as the model’s “confidence” that a reaction belongs to a

given class. With this information, the highest scoring class can be taken as the model's prediction or can be processed more carefully to intuit new insights about the data.

Before the model is trained on the data, the inputs are first processed according to the steps outlined in the Data Generation and Processing section of the main manuscript. An important note, however, is that the MLP is the only model type explored in this study that does not require additional feature extraction. Neural networks are themselves feature extractors, so attempting to train the model on PCA extracted features or metrics would limit the potential features from which the model can extract information. This represents a massive advantage DNNs have over other ML methods, as DNNs generally do not require time-consuming feature engineering to achieve good performance¹.

Training an MLP involves passing training data through the network in groups called batches. For this implementation, the number of reactions in a batch was 64. Once the network obtains predictions on this data, the prediction is compared to the true class using an error function (more commonly known as the loss function). This loss is then used to update the weights in between neurons of the network through a process called backpropagation^{1,61}. This process continues until all the training data has been passed through the network, completing one iteration or epoch of training. The MLP used in this study was set to train for a maximum of 100 epochs. While training, a portion of the training dataset is withheld (10% in this case) for validation of the model after each epoch. Should the model begin to overfit, a process in which the model loses the ability to generalize beyond the training data without learning more generally applicable patterns, the loss of the model on the validation data will begin to increase even as the training data loss decreases. The loss of the model on this validation set was monitored during training (Fig. 9) and used to limit overfitting. At each epoch, the model was saved as a checkpoint if the validation loss at that epoch was less than the best previous validation loss. At the end of training, the last checkpoint was restored to the model iteration that performed best on the validation set. For the MLP used in this study, however, the model always improved on the validation set, so the checkpoint that was restored was the final state of the model after the 100th epoch.

Once the model was trained, the outright classification was a non-optimal representation of the data in this study. The false positive samples and the positive samples were not clearly distinguished by the model, meaning any well with a curve would score highly in both the false positive and positive classes. This behavior was undesirable as many positives were

mistaken for false positives and vice versa. In order to rectify this issue, a weighted average was applied to the output to get a single, unified score. The model output used the function shown in Eq. 1 to calculate the final score used for final classification. In the equation, Y_{neg} is the score for the negative class, Y_{fp} is the score for the false positive class, Y_{pos} is the score for the positive class, and Y_{out} is the overall binary score used in the results. This smoothed out the poorly defined class boundary into a spectrum from 0 to 2 (noting that the sum of the score outputs produced by the model are normalized), with any score less than 0.5 being labeled a negative prediction, any value between 0.5 and 1.5 being labeled a false positive prediction, and anything above 1.5 being labeled a positive prediction.

$$Y_{out} = 0 \times Y_{neg} + 1 \times Y_{fp} + 2 \times Y_{pos}$$

Equation 1: Weighted Average Applied to MLP Output

Evaluation criteria

Evaluation of the models primarily considered five metrics, accuracy, sensitivity/recall, specificity, precision, and F1-score. The accuracy metric corresponds to how many wells were classified the same as the human labels. As this study used 2 classes, the worst possible accuracy score should be 50% (when adjusted for class weight), corresponding to a random guess. A significantly lower accuracy would likely be the result of a procedural issue, so for each accuracy metric, it is imperative to consider the improvement over 50% as the true success rate of the model. While accuracy evaluates the dataset holistically, sensitivity (or recall) evaluates the performance of the model on identifying positive samples correctly. Specifically, sensitivity corresponds to the fraction of human-labeled positive wells that were classified correctly. Low sensitivity corresponds to a model that frequently classifies annotator-labeled positive samples as negative. Specificity, contrary to sensitivity, identifies the fraction of human-labeled negative wells that were classified correctly. Low specificity, then, correlates to a model that frequently classifies negative samples as positive. Another metric, precision, provides additional insight into how well a model is distinguishing between classes. Precision is defined as the fraction of model-labeled positive samples that were actually positive. This metric is of particular interest in this application as the number of negative reactions in the dataset is much greater than the number of positive reactions - an environment that can allow the previous metrics to be high despite a poorly performing model. In particular, precision measures a model's false positive rate, with a low precision indicating a high rate of false positives relative to the number of true positives, making a positive prediction less meaningful. F1-score is a combination of sensitivity and precision, creating a metric that evaluates the performance of a model on positive samples considering these factors together. While less useful for identifying patterns of mistakes in models, F1-score can be used to put sensitivity and precision in context together. F1-score is calculated as the harmonic mean of precision and sensitivity, meaning a particularly low precision or sensitivity will create a low F1-score.

Data availability

Data used in this manuscript can be found at: <https://github.com/HoweyUMN/QuICSeedAI>.

Code availability

Code to reproduce the work this manuscript can be found at: <https://github.com/HoweyUMN/QuICSeedAI>.

Received: 16 October 2024; Accepted: 26 February 2025;

Published online: 08 April 2025

References

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
2. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

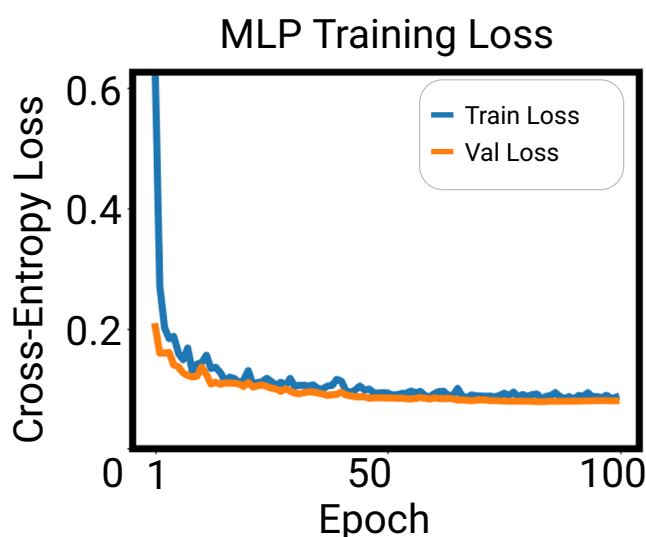


Fig. 9 | Multilayer perceptron loss. The value of the cross-entropy loss is measured after each epoch on the dataset used to train the model (Train Loss) and a small portion of the larger training set which was withheld for validation (Val Loss). A lower loss indicates less error between predicted and true labels.

3. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
4. Anishchenko, I. et al. De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).
5. Saborio, G. P., Permanne, B. & Soto, C. Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding. *Nature* **411**, 810–813 (2001).
6. Atarashi, R. et al. Simplified ultrasensitive prion detection by recombinant PrP conversion with shaking. *Nat. Methods* **5**, 211–212 (2008).
7. Orrù, C. D. et al. Human variant Creutzfeldt-Jakob disease and sheep scrapie PrP(res) detection using seeded conversion of recombinant prion protein. *Protein Eng. Des. Sel.* **22**, 515–521 (2009).
8. Wilham, J. M. et al. Rapid end-point quantitation of prion seeding activity with sensitivity comparable to bioassays. *PLoS Pathog.* **6**, e1001217 (2010).
9. Atarashi, R., Sano, K., Satoh, K. & Nishida, N. Real-time quaking-induced conversion: a highly sensitive assay for prion detection. *Prion* **5**, 150–153 (2011).
10. Tennant, J. M., Henderson, D. M., Wisniewski, T. M. & Hoover, E. A. RT-QulC detection of tauopathies using full-length tau substrates. *Prion* **14**, 249–256 (2020).
11. Metrick, M. A. 2nd et al. A single ultrasensitive assay for detection and discrimination of tau aggregates of Alzheimer and Pick diseases. *Acta Neuropathol. Commun.* **8**, 22 (2020).
12. Fellner, L., Jellinger, K. A., Wenning, G. K. & Haybaeck, J. Commentary: Discriminating α -synuclein strains in parkinson's disease and multiple system atrophy. *Front. Neurosci.* **14**, 802 (2020).
13. Scialò, C. et al. TDP-43 real-time quaking induced conversion reaction optimization and detection of seeding activity in CSF of amyotrophic lateral sclerosis and frontotemporal dementia patients. *Brain Commun* **2**, fcaa142 (2020).
14. Coysh, T. & Mead, S. The Future of Seed Amplification Assays and Clinical Trials. *Front. Aging Neurosci.* **14**, 872629 (2022).
15. Brockmann, K. et al. CSF α -synuclein seed amplification kinetic profiles are associated with cognitive decline in Parkinson's disease. *NPJ Parkinsons Dis* **10**, 24 (2024).
16. Manca, M. et al. Tau seeds occur before earliest Alzheimer's changes and are prevalent across neurodegenerative diseases. *Acta Neuropathol.* **146**, 31–50 (2023).
17. Atarashi, R. et al. Ultrasensitive human prion detection in cerebrospinal fluid by real-time quaking-induced conversion. *Nat. Med.* **17**, 175–178 (2011).
18. Rhoads, D. D. et al. Diagnosis of prion diseases by RT-QulC results in improved surveillance. *Neurology* **95**, e1017–e1026 (2020).
19. Mok, T. H. et al. Seed amplification and neurodegeneration marker trajectories in individuals at risk of prion disease. *Brain* **146**, 2570–2583 (2023).
20. Mok, T. H. & Mead, S. Preclinical biomarkers of prion infection and neurodegeneration. *Curr. Opin. Neurobiol.* **61**, 82–88 (2020).
21. Salvadores, N., Shah Nawaz, M., Scarpini, E., Tagliavini, F. & Soto, C. Detection of misfolded A β oligomers for sensitive biochemical diagnosis of Alzheimer's disease. *Cell Rep.* **7**, 261–268 (2014).
22. Coughlin, D. G. et al. Selective tau seeding assays and isoform-specific antibodies define neuroanatomic distribution of progressive supranuclear palsy pathology arising in Alzheimer's disease. *Acta Neuropathol.* **144**, 789–792 (2022).
23. Shah Nawaz, M. et al. Development of a Biochemical Diagnosis of Parkinson Disease by Detection of α -Synuclein Misfolded Aggregates in Cerebrospinal Fluid. *JAMA Neurol.* **74**, 163–172 (2017).
24. Candelise, N. et al. Effect of the micro-environment on α -synuclein conversion and implication in seeded conversion assays. *Transl. Neurodegener.* **9**, 5 (2020).
25. Escobar, L. E. et al. The ecology of chronic wasting disease in wildlife. *Biol. Rev. Camb. Philos. Soc.* **95**, 393–408 (2020).
26. Tennant, J. M. et al. Shedding and stability of CWD prion seeding activity in cervid feces. *PLoS One* **15**, e0227094 (2020).
27. Li, M. et al. RT-QulC detection of CWD prion seeding activity in white-tailed deer muscle tissues. *Sci. Rep.* **11**, 16759 (2021).
28. Haley, N. J. et al. Prion-seeding activity in cerebrospinal fluid of deer with chronic wasting disease. *PLoS One* **8**, e81488 (2013).
29. Cooper, S. K. et al. Detection of CWD in cervids by RT-QulC assay of third eyelids. *PLoS One* **14**, e0221654 (2019).
30. Ferreira, N. C. et al. Detection of chronic wasting disease in mule and white-tailed deer by RT-QulC analysis of outer ear. *Sci. Rep.* **11**, 7702 (2021).
31. Christenson, P. R., Li, M., Rowden, G., Larsen, P. A. & Oh, S.-H. Nanoparticle-Enhanced RT-QulC (Nano-QulC) Diagnostic Assay for Misfolded Proteins. *Nano Lett.* **23**, 4074–4081 (2023).
32. Christenson, P. R. et al. A field-deployable diagnostic assay for the visual detection of misfolded prions. *Sci. Rep.* **12**, 12246 (2022).
33. Housmans, J. A. J., Wu, G., Schymkowitz, J. & Rousseau, F. A guide to studying protein aggregation. *FEBS J.* **290**, 554–583 (2023).
34. Henderson, D. M. et al. Quantitative assessment of prion infectivity in tissues and body fluids by real-time quaking-induced conversion. *J. Gen. Virol.* **96**, 210–219 (2015).
35. Rowden, G. R. et al. Standardization of Data Analysis for RT-QulC-based Detection of Chronic Wasting Disease. *Pathogens* **12**, 309 (2023).
36. De Luca, C. M. G. et al. Efficient RT-QulC seeding activity for α -synuclein in olfactory mucosa samples of patients with Parkinson's disease and multiple system atrophy. *Transl. Neurodegener.* **8**, 24 (2019).
37. Standke, H. G. & Kraus, A. Seed amplification and RT-QulC assays to investigate protein seed structures and strains. *Cell Tissue Res.* **392**, 323–335 (2023).
38. Goodfellow, I., Bengio, Y. & Courville, A. Deep Learning. (MIT Press, 2016).
39. Ng, A., Jordan, M. & Weiss, Y. On Spectral Clustering: Analysis and an algorithm. in *Advances in Neural Information Processing Systems* (eds. Dietterich, T., Becker, S. & Ghahramani, Z.) vol. 14 (MIT Press, 2001).
40. Na, S., Xumin, L. & Yong, G. Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. in *2010 Third International Symposium on Intelligent Information Technology and Security Informatics* 63–67 (IEEE, 2010). <https://doi.org/10.1109/IITSI.2010.74>.
41. Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J. & Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Appl.* **13**, 18–28 (1998).
42. Taud, H. & Mas, J. F. Multilayer Perceptron (MLP). in *Geomatic Approaches for Modeling Land Change Scenarios* (eds. Camacho Olmedo, M. T., Paegelow, M., Mas, J.-F. & Escobar, F.) 451–455 (Springer International Publishing, Cham, 2018). https://doi.org/10.1007/978-3-319-60801-3_27.
43. Marissa S. et al. Detection and decontamination of chronic wasting disease prions during venison processing. *Emerging Infectious Diseases* (2024) <https://doi.org/10.1101/2024.07.23.604851>.
44. Schwabenlander, M. D. et al. Comparison of chronic wasting disease detection methods and procedures: Implications for free-ranging white-tailed deer (*Odocoileus virginianus*) surveillance and management. *J. Wildl. Dis.* **58**, 50–62 (2022).
45. Li, M. et al. QulCSeedR: an R package for analyzing fluorophore-assisted seed amplification assay data. *Bioinformatics* **41**, btac752 (2024).
46. Goto, Y. et al. Development of HANABI, an ultrasonication-forced amyloid fibril inducer. *Neurochem. Int.* **153**, 105270 (2022).
47. Christenson, P. R. et al. Blood-based Nano-QulC: Accelerated and inhibitor-resistant detection of misfolded alpha-synuclein. *Nano Lett.* **24**, 15016–15024 (2024).
48. Christenson, P. R. et al. Visual detection of misfolded alpha-synuclein and prions via capillary-based quaking-induced conversion assay (Cap-QulC). *npj Biosens.* **1**, 2 (2024).

49. Pilotto, A. Alpha-synuclein co-pathology in Alzheimer's disease identified by RT-QUIC identification. *Alzheimers. Dement.* **18**(Suppl. 6):e068145 (2022).
50. Mastrangelo, A. et al. Evaluation of the impact of CSF prion RT-QuIC and amended criteria on the clinical diagnosis of Creutzfeldt-Jakob disease: a 10-year study in Italy. *J. Neurol. Neurosurg. Psychiatry* **94**, 121–129 (2023).
51. Concha-Marambio, L. et al. Seed Amplification Assay to Diagnose Early Parkinson's and Predict Dopaminergic Deficit Progression. *Mov. Disord.* **36**, 2444–2446 (2021).
52. Saijo, E. et al. Ultrasensitive and selective detection of 3-repeat tau seeding activity in Pick disease brain and cerebrospinal fluid. *Acta Neuropathol.* **133**, 751–765 (2017).
53. Kraus, A. et al. High-resolution structure and strain comparison of infectious mammalian prions. *Mol. Cell* **81**, 4540–4551.e6 (2021).
54. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**, 2825–2830 (2011).
55. Harris & R, C. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
56. Maćkiewicz, A. & Ratajczak, W. Principal components analysis (PCA). *Comput. Geosci.* **19**, 303–342 (1993).
57. Wilkin, G. A. & Huang, X. K-Means Clustering Algorithms: Implementation and Comparison. in *Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007)* 133–136 (IEEE, 2007). <https://doi.org/10.1109/IMSCCS.2007.51>.
58. Ghosh, J. & Nag, A. An Overview of Radial Basis Function Networks. in *Radial Basis Function Networks 2: New Advances in Design* (eds. Howlett, R. J. & Jain, L. C.) 1–36 (Physica-Verlag HD, Heidelberg, 2001). https://doi.org/10.1007/978-3-7908-1826-0_1.
59. Chollet, F. & Others. Keras. <https://keras.io> (2015).
60. Agarap, A. F. Deep Learning using Rectified Linear Units (ReLU). arXiv:1803.08375 [cs.NE] (2018).
61. Plaut, D. C. & Hinton, G. E. Learning sets of filters using back-propagation. *Comput. Speech Lang.* **2**, 35–61 (1987).

Acknowledgements

A portion of this research was supported by the Minnesota State Legislature through the Minnesota Legislative-Citizen Commission on Minnesota Resources (LCCMR) and Minnesota Agricultural Experiment Station Rapid Agricultural Response Fund to K.D.H., M.L., P.R.C., P.A.L. and S.-H.O. M.L., P.A.L., and S.-H.O. acknowledge support from the Minnesota Partnership for Biotechnology and Medical Genomics. S.-H.O. further acknowledges support from the Sanford P. Bordeaux Chair and the International Institute for Biosensing (IIB) at the University of Minnesota. The figures in this manuscript were created with the assistance of Biorender <https://BioRender.com>.

Author contributions

S.-H.O. conceived the study. M.L. and K.D.H. designed the study. K.D.H. developed the code. K.D.H., M.L., and S.-H.O. wrote the manuscript based on the input from other co-authors. M.L. curated data for the study. K.D.H. conducted the analysis utilized in the study. K.D.H., M.L., and P.R.C. visualized the data. K.D.H., M.L., P.R.C., P.A.L., and S.-H.O. discussed the results. All authors reviewed the manuscript. All authors approved the manuscript.

Competing interests

Authors K.D.H., M.L., and P.R.C. declare no financial or non-financial competing interests. P.A.L. and S.-H.O. are co-founders and equity holders of Priogen Corp, a diagnostic company specializing in the detection of prions and protein-misfolding diseases. Author S.-H.O. serves as Editor-in-Chief of this journal and had no role in the peer review or decision to publish this manuscript.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44328-025-00035-0>.

Correspondence and requests for materials should be addressed to Sang-Hyun Oh.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025