scientific reports

Check for updates

OPEN

Automated classification of benign paroxysmal positional vertigo from video-nystagmography using a delay-aware neural network

Kunal Chaturvedi[1], Nicholas Yang[2], Imelda Hannigan[2], Nicole Reid[3], Andrew Bradshaw[3], Karthick Thiyagarajan[4], Tony Jan[5], Ali Braytee[1], Miriam S. Welgampola[2,3] & Mukesh Prasad[1]✉

Nystagmus is a key indicator of vestibular disorders, including benign paroxysmal positional vertigo (BPPV). Accurate diagnosis of BPPV is essential, as it is treatable with specific bedside maneuvers that lead to rapid symptom resolution, thereby improving patient outcomes and reducing unnecessary treatments. In clinical practice, identification of positional nystagmus relies on eliciting and interpreting eye movements during provocative maneuvers, with or without video nystagmography (VNG). This process can be subjective and difficult to standardize when signals are subtle, noisy, or temporally variable. We present DSF-BPPVNet, a delay-aware neural architecture for BPPV classification from VNG traces. The model combines temporal convolution, delayed-state feedback, and residual refinement to support classification from temporally structured eye-movement signals. The model was evaluated on 3,111 VNG traces from 705 patients using 5-fold cross-validation and compared with established deep-learning baselines. In the patient-independent setting, DSF-BPPVNet achieved the strongest overall performance among the evaluated models, with an F1-score of 0.819 ± 0.020. Explainability analyses were also performed to characterize model attribution patterns and temporal weighting behavior.

Keywords Video-nystagmography, Benign paroxysmal positional vertigo, Time-series classification, Explainable AI

Vertigo is a common, disabling, and treatable symptom that arises from dysfunction of the peripheral vestibular organs or their central connections[1]. It has substantial effects on physical and mental health, quality of life, and daily functioning, and is associated with an increased risk of falls and injury[2]. Benign paroxysmal positional vertigo (BPPV) is the most prevalent form of recurrent vertigo, accounting for nearly 20–30% of all cases[3]. BPPV arises when displaced otoconia abnormally stimulate the semicircular canals during head movement, provoking transient vertigo together with characteristic positional nystagmus[4]. In clinical practice, diagnosis depends on eliciting and interpreting these eye movements during provocative maneuvers such as the Dix–Hallpike and roll tests, because the direction and temporal evolution of the elicited eye movements contribute importantly to clinical interpretation.

This dependence on eye-movement interpretation creates an important clinical bottleneck. Recognition of positional nystagmus remains operator-dependent, and diagnostic accuracy can vary with experience, recording quality, and testing conditions. These limitations have motivated growing interest in technology-assisted vestibular assessment, including video-nystagmography (VNG) and AI-supported interpretation. However, the value of such approaches still rests on one central requirement: reliable automated analysis of noisy VNG recordings while preserving the temporal structure used in clinical decision-making.

Clinical interpretation of positional nystagmus relies not only on movement direction but also on temporal descriptors such as latency, duration, and progression over time. These temporal patterns are clinically

[1]School of Computer Science, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia. [2]Central Clinical School, University of Sydney, Sydney, Australia. [3]Institute of Clinical Neurosciences, Royal Prince Alfred Hospital, Sydney, Australia. [4]School of Engineering, Design and Built Environment, Western Sydney University, Sydney, Australia. [5]Centre for Artificial Intelligence and Optimization, Torrens University, Sydney, Australia. ✉email: mukesh.prasad@uts.edu.au

informative but may vary across canal involvement, debris location, and the biomechanics of head position relative to gravity[5,6]. Quantitative analysis can further summarize these dynamics using measures such as slow-phase velocity (SPV), from which features including onset delay, peak response, decay behavior, and episode duration may be derived[7–10]. Traditional automated approaches have used such handcrafted descriptors for classification[11,12], but their performance depends on accurate feature extraction and expert-defined representations. In this work, rather than using explicitly computed SPV signals as input, we learn directly from the recorded horizontal and vertical eye-position traces so that temporal features can be inferred from the trajectories themselves.

Recent learning-based approaches have demonstrated that nystagmus analysis from recorded eye movements is feasible. Prior work has explored convolutional models[13–15], recurrent and hybrid CNN–RNN architectures[16,17], and more recent neural ODE-based[18] and Transformer-based approaches[19] for nystagmus detection and BPPV-related classification tasks. At the same time, the field has begun to expand beyond isolated backbone comparisons toward broader AI-assisted diagnostic workflows[20,21]. Despite these advances, important challenges remain. Positional nystagmus remains a temporally structured signal in which diagnostically relevant evidence may be distributed across separated segments of the trace. Standard temporal convolutions are effective at extracting local patterns, but they do not retain prior latent states as explicitly reusable representations. Recurrent formulations can aggregate prior context, yet that history is typically compressed into a hidden representation rather than maintained as a structured set of temporally separated states that can be revisited during processing. Continuous-time formulations offer a different view of temporal evolution, although their smooth latent dynamics may be less directly matched to signals containing short-lived, phase-dependent transitions and abrupt resets. In addition, once intermediate features are formed, later evidence may still help disambiguate weak responses, onset patterns, or artifact-contaminated segments.

Motivated by this problem, we propose DSF-BPPVNet, a neural architecture for BPPV classification from VNG traces that integrates delayed-state feedback, temporal convolution, and residual refinement. The central idea is to preserve prior latent states in a structured buffer and reintroduce them into the current computation through learnable weighting, allowing the model to reuse temporally separated information when it becomes diagnostically relevant. Residual refinement complements this mechanism by progressively updating intermediate representations rather than passing them forward unchanged. This is intended to improve representation quality in the presence of variability in onset, amplitude, duration, and waveform morphology. We evaluate this approach against established deep-learning baselines and further examine its behavior using ablation and attribution analyses.

The main contributions of the paper are as follows:

- DSF-BPPVNet is proposed for BPPV classification from video nystagmography, combining delayed-state feedback, temporal convolution, and residual refinement.
- The proposed model is evaluated against representative deep-learning baselines for BPPV classification.
- We use attribution-based analyses together with examination of the learned delay-weight distributions to characterize the model's input sensitivity and internal temporal weighting patterns

Methods

Problem formulation

Let X denote a video-nystagmography (VNG) sequence comprising horizontal and vertical eye-position traces, and let y denote the corresponding class label. Depending on the task, y may represent either a binary BPPV label or a canal subtype category. The objective is to learn a mapping from the multichannel temporal input X to the predicted label $\hat{y}$.

Positional nystagmus is a temporally structured signal in which diagnostically relevant evidence may be distributed across separated portions of the trace rather than concentrated within a single local segment. This motivates a representation in which the current latent update can depend not only on the present feature state but also on delayed latent states retained from earlier stages of processing.

Delayed-state formulation

A standard state-update formulation computes the next representation primarily from the current state. More generally, a delayed-state formulation allows the current update to depend on both the present representation and a set of delayed states:

$$z^{(l+1)} = \Phi\left(z^{(l)}, z^{(l-\tau_1)}, z^{(l-\tau_2)}, \ldots, z^{(l-\tau_M)}; \theta\right), \tag{1}$$

where $z^{(l)}$ denotes the latent representation at processing stage l, $\tau_1, \ldots, \tau_M$ are discrete delays, M is the number of delay terms, and $\Phi(\cdot)$ is a learnable nonlinear transformation parameterized by θ. This formulation is useful when informative evidence is distributed across separated parts of the signal and delayed context may help refine the current representation.

In DSF-BPPVNet, this idea is implemented as a discrete delayed-state feedback mechanism operating on latent feature representations within stacked temporal processing blocks, rather than as an explicit continuous-time delay differential equation solver.

DSF-BPPVNet architecture

We propose DSF-BPPVNet, a delay-aware neural architecture for BPPV classification from VNG sequences. The model combines temporal convolution, delayed-state feedback, and residual refinement within a stacked block design. Temporal convolution extracts local sequence patterns from the eye-movement traces, delayed-state

feedback reintroduces selected prior latent representations through learned weighting, and residual refinement supports progressive updating of intermediate features across blocks. Through this design, the network integrates local temporal feature extraction with delayed latent-state reuse and stage-wise representation refinement. The overall framework is illustrated in Figure 1.

Each block first applies temporal convolutional transformations with nonlinear activation and regularization to produce an intermediate hidden representation from the incoming latent signal. In parallel, a residual pathway is maintained so that each block updates the incoming representation rather than replacing it entirely. Let $r^{(b)}$ denote the residual stream associated with block b, and let $u^{(b)}$ denote the hidden representation produced after temporal feature extraction.

To make delayed latent information available during feature extraction, each block maintains a fixed-length buffer of prior latent representations:

$$S^{(b)} = \left\{ s_1^{(b)}, s_2^{(b)}, \ldots, s_M^{(b)} \right\}, \tag{2}$$

where M denotes the number of delay slots. A block-specific delay parameter determines the lag spacing used to select buffered states. Buffered representations chosen at successive delay multiples are combined using learned normalized weights:

$$\alpha_i^{(b)} = \frac{\exp\left(\omega_i^{(b)}\right)}{\sum_{j=1}^{M} \exp\left(\omega_j^{(b)}\right)}, \quad i = 1, \ldots, M, \tag{3}$$

where $\omega_i^{(b)}$ are trainable parameters and $\alpha_i^{(b)}$ are the corresponding normalized coefficients. The delayed-state contribution is then computed as

$$d^{(b)} = \sum_{i=1}^{M} \alpha_i^{(b)} \, \tilde{s}_i^{(b)}, \tag{4}$$

where $\tilde{s}_i^{(b)}$ denotes the buffered representation selected at the i-th delay multiple determined by the block-specific delay parameter. This delayed contribution is added to the current hidden representation:

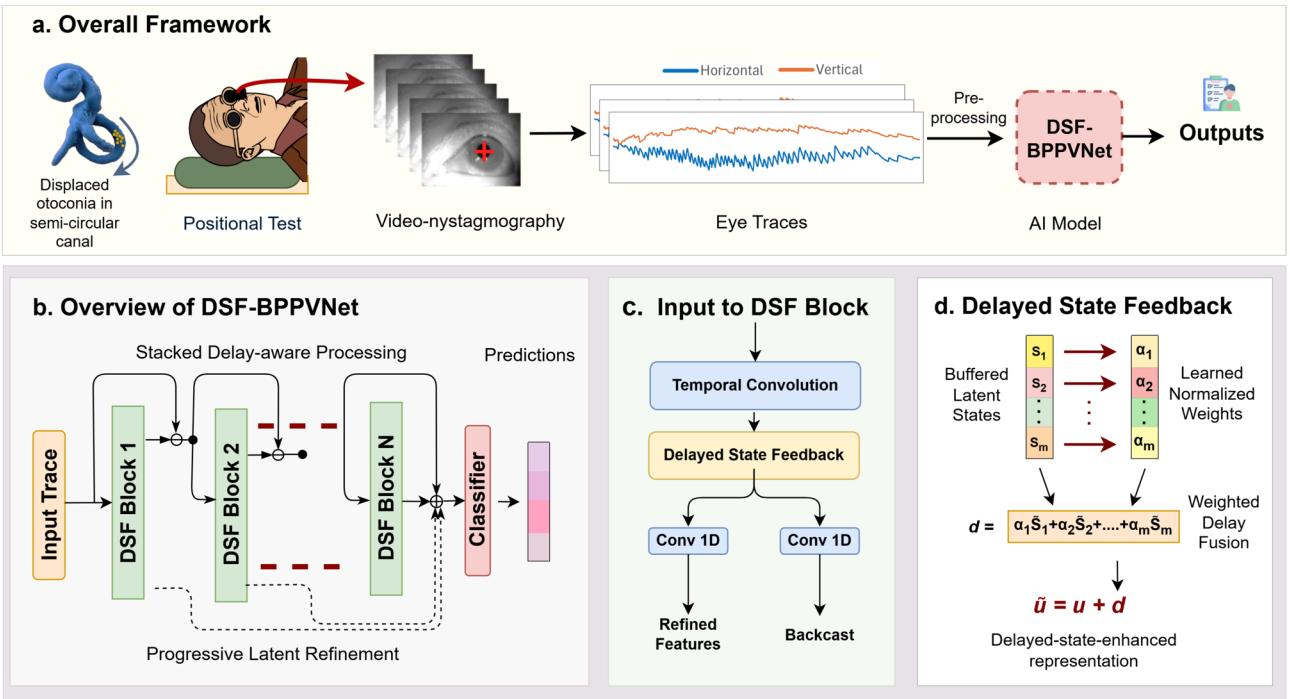


Fig. 1. Overview of the proposed DSF-BPPVNet framework. (**a**) Pipeline for classification from VNG, in which traces extracted are provided to DSF-BPPVNet to produce the output class label. (**b**) Overall DSF-BPPVNet architecture, comprising stacked delayed-state feedback blocks. (**c**) Structure of an individual block, combining temporal convolution, delayed-state feedback, and residual refinement. (**d**) Delayed-state feedback mechanism, where buffered latent representations are combined using learned normalized weights to reintroduce delayed information into the current representation.

$$\tilde{u}^{(b)} = u^{(b)} + d^{(b)}, \tag{5}$$

This operation reintroduces delayed latent information into the current block output using a learned mixture over buffered states, rather than relying only on the present convolutional features. After fusion, the current latent representation is inserted into the buffer and the oldest buffered representation is discarded, enabling delayed information to be carried forward across successive stages of processing.

From the delayed-state-enhanced representation $\tilde{u}^{(b)}$, each block produces two transformed outputs. The first is a backcast term,

$$b^{(b)} = \rho_{\text{back}}\left(\tilde{u}^{(b)}\right), \tag{6}$$

which is subtracted from the residual pathway to generate the input to the next block:

$$z^{(b+1)} = r^{(b)} - b^{(b)}, \tag{7}$$

The second is a refined feature representation,

$$f^{(b)} = \rho_{\text{feat}}\left(\tilde{u}^{(b)}\right), \tag{8}$$

which represents the transformed output of the current block. This design separates the role of residual updating from feature refinement. The residual stream provides a pathway for information to be passed forward across blocks, while the refined feature stream captures the transformed representation produced after delayed-state fusion. Across the stacked architecture, this backcast-style update is intended to support progressive refinement of intermediate representations. This residual updating strategy is conceptually related to the backcast refinement used in N-BEATS[22], although here it is adapted to sequential representation learning for VNG classification rather than forecasting.

After the final temporal processing block, the refined feature representation is aggregated over time to obtain a sequence-level summary, which is then mapped to the output by the final classification layer. The dimensionality of the output layer is task-dependent and is set according to the classification target.

Experimental protocol

Dataset

The dataset consisted of 3,111 2D eye-position traces from 705 patients tested using the Epley Omniax System (EOS) (Figure 2, left), a motorized two-axis rotator chair integrated with real-time VOG. Of these, 1,027 traces were labeled as BPPV and 2,084 as normal/non-BPPV under the study annotation protocol. The BPPV subset comprised 856 posterior-canal cases, 142 horizontal-canal cases, and 29 anterior-canal cases. The normal class comprised traces in which no positional nystagmus was identified during the recorded maneuvers. This group included non-diagnostic eye movements observed during testing, such as movements commonly seen when visual fixation is removed; we elected to retain such recordings when expert review did not indicate BPPV-related positional nystagmus. Recordings were acquired at a public outpatient facility using infrared monocular video-oculography in darkness to eliminate visual fixation. Eye movements were recorded at 30 Hz, and horizontal calibration was performed prior to each test. The dataset included recordings from four standardized diagnostic maneuvers: left Hallpike, right Hallpike, left roll, and right roll. Both horizontal and vertical eye-movement traces were used for analysis; torsional eye-movement information was not available. Raw horizontal and vertical eye-position traces were normalized using z-score normalization to reduce amplitude variability across recordings and mitigate the effect of transient artifacts. The dataset was annotated by a neuro-otologist clinician and a neuro-otologist nurse using the OmniVNG software (Figure 2, right).



Fig. 2. Data acquisition and annotation using the Epley Omniax System (EOS) (left) and OmniVNG software (right).

Model	Accuracy (%)	Precision	Recall	F1
BiLSTM	79.68 ± 5.24	0.717 ± 0.105	0.660 ± 0.089	0.686 ± 0.085
1D-CNN	86.23 ± 6.01	**0.855 ± 0.097**	0.723 ± 0.132	0.779 ± 0.097
Transformer	78.58 ± 1.36	0.707 ± 0.081	0.647 ± 0.133	0.667 ± 0.051
DSF-BPPVNet	**87.88 ± 0.84**	0.834 ± 0.036	**0.808 ± 0.044**	**0.819 ± 0.020**

Table 1. Overall comparison of patient-independent performance for BPPV classification. Values are mean ± SD across 5 folds.

Model	Accuracy (%)	Precision	Recall	F1
BiLSTM	81.71 ± 3.51	0.748 ± 0.064	0.693 ± 0.071	0.718 ± 0.058
1D-CNN	88.55 ± 1.60	**0.886 ± 0.050**	0.760 ± 0.053	0.817 ± 0.028
Transformer	78.91 ± 2.07	0.706 ± 0.047	0.649 ± 0.108	0.673 ± 0.053
DSF-BPPVNet	**89.38 ± 1.20**	0.880 ± 0.025	**0.795 ± 0.057**	**0.834 ± 0.024**

Table 2. Overall comparison of non-patient-disjoint performance for BPPV classification. Values are mean ± SD across 5 folds.

Experiment setup

We considered two evaluation settings: a non-patient-disjoint setting and a patient-independent setting. In the non-patient-disjoint setting, traces were randomly partitioned without enforcing patient-level separation. Within each training fold, 10% of the training data was set aside as a validation set using stratified sampling for model selection and early stopping. In the patient-independent setting, patients were partitioned into disjoint training, validation, and test sets such that no patient contributed traces to more than one split. Class distribution was preserved across folds by stratifying at the patient level. As in the previous setting, 10% of the training patients were allocated to the validation split.

For consistency, all baseline models were re-implemented within the same PyTorch framework and evaluated using the same 5-fold cross-validation splits as DSF-BPPVNet. All models were trained for up to 400 epochs using the Adam optimizer with a learning rate of 1×10^{-3}. For binary classification, training used binary cross-entropy with logits and early stopping based on validation accuracy with a patience of 100 epochs. For the 4-class subtype experiment, training used multiclass cross-entropy loss and early stopping based on validation macro F1 with the same patience. Evaluation was based on accuracy, precision, recall, and F1-score.

Hyperparameter selection

To determine the DSF-BPPVNet configuration, we performed a randomized hyperparameter search across 50 settings using 5-fold stratified cross-validation in the non-patient-disjoint setting. The search covered architectural, temporal, and optimization parameters, including the number of temporal convolutional blocks, number of delay states, initial delay stride, convolutional kernel width, pooling strategy, dropout, and learning rate. The delay stride was treated as a learnable parameter in all configurations. The final configuration selected for subsequent experiments used four temporal convolutional blocks, six delay states, a kernel size of 5, an initial delay stride of 6.0, average pooling, dropout of 0.5, and a learning rate of 0.001.

For the baseline models, hyperparameters were selected by randomized search: the 1D-CNN used 64 filters with kernel size 15, the BiLSTM used two layers with a hidden size of 64 per direction, and the Transformer used two encoder layers, four attention heads, and a model dimension of 64.

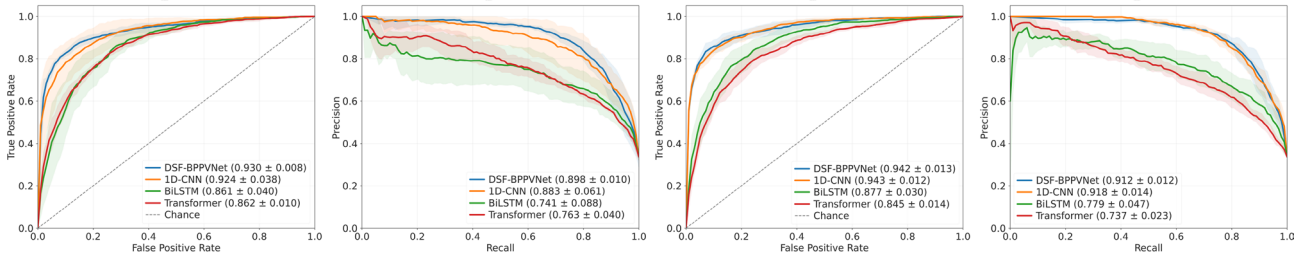
Results

Overall performance

To evaluate the proposed DSF-BPPVNet, we benchmarked its performance against three established deep learning baselines: a 1D-CNN, a Bidirectional LSTM (BiLSTM), and a Transformer encoder. Table 1 summarizes model performance in the patient-independent setting. Across the four models evaluated, DSF-BPPVNet achieved the strongest overall performance, with the highest mean F1-score of 0.819 ± 0.020 and the highest mean recall of 0.808 ± 0.044, together with the highest mean accuracy of 87.88 ± 0.84%. The strongest baseline in this setting was the 1D-CNN, which achieved an F1-score of 0.779 ± 0.097 and a recall of 0.723 ± 0.132. Although the 1D-CNN attained the highest mean precision, 0.855 ± 0.097, its lower recall indicates a less balanced performance than DSF-BPPVNet. The BiLSTM and Transformer baselines yielded lower mean F1-scores of 0.686 ± 0.085 and 0.667 ± 0.051, respectively.

Table 2 reports results in the non-patient-disjoint setting. DSF-BPPVNet again achieved the highest mean F1-score, 0.834 ± 0.024, and the highest mean recall, 0.795 ± 0.057, while also attaining the highest mean accuracy, 89.38 ± 1.20%. The 1D-CNN remained the strongest baseline, with an F1-score of 0.817 ± 0.028. As in the patient-independent setting, the performance of DSF-BPPVNet was characterized by a more favorable balance between precision and recall than the comparator models.

The discriminative behavior of the models is further illustrated by the mean Precision–Recall and Receiver Operating Characteristic curves in Figure 3. In the patient-independent setting shown in Figure 3a, DSF-



(a) Patient-independent setting (b) Non-patient-disjoint setting

Fig. 3. Receiver Operating Characteristic (ROC) and Precision–Recall (PR) curves across 5-fold cross-validation for DSF-BPPVNet and baseline models. The two left panels correspond to the patient-independent setting, and the two right panels correspond to the non-patient-disjoint setting.

DSF	Refinement	Accuracy (%)	Precision	Recall	F1
×	×	87.79 ± 1.39	0.871 ± 0.023	0.750 ± 0.053	0.805 ± 0.028
✓	×	88.55 ± 1.16	0.856 ± 0.052	0.801 ± 0.042	0.825 ± 0.015
×	✓	88.30 ± 1.36	0.851 ± 0.039	0.794 ± 0.042	0.820 ± 0.022
✓	✓	89.38 ± 1.20	0.880 ± 0.025	0.795 ± 0.057	0.834 ± 0.024

Table 3. Effect of delayed-state feedback and residual refinement in DSF-BPPVNet in the non-patient-disjoint setting. Values are mean ± SD across 5 folds.

BPPVNet achieved the highest mean ROC-AUC, 0.930 ± 0.008, and the highest mean average precision, 0.898 ± 0.010, exceeding all baselines, including the 1D-CNN with ROC-AUC of 0.924 ± 0.038 and mAP of 0.883 ± 0.061. In addition to its higher mean performance, DSF-BPPVNet showed narrower variability bands across folds, indicating more stable behavior under cross-validation. A similar pattern was observed in the non-patient-disjoint setting shown in Figure 3b, where DSF-BPPVNet again showed the strongest overall Precision–Recall and ROC characteristics.

Ablation study

To evaluate the contributions of delayed-state feedback and residual refinement, we compared four model variants in the non-patient-disjoint setting, as summarized in Table 3.

The baseline configuration without delayed-state feedback or residual refinement achieved the lowest F1-score, 0.805 ± 0.028. Adding residual refinement alone increased the F1-score to 0.820 ± 0.022, while adding delayed-state feedback alone increased it to 0.825 ± 0.015. The full DSF-BPPVNet model, which includes both components, achieved the highest F1-score, 0.834 ± 0.024. These results suggest that both delayed-state feedback and residual refinement contribute positively to performance. Each component improved upon the baseline when introduced individually, and the best overall result was obtained when both were used together. This pattern is consistent with the two components providing complementary benefits within the full architecture.

Subtype analysis

We further evaluated DSF-BPPVNet on a more granular 4-class subtype classification task in the patient-independent setting, with classes defined as Normal, PC-BPPV, HC-BPPV, and AC-BPPV. This task is substantially more challenging because of severe class imbalance, with HC-BPPV and AC-BPPV accounting for only 4.6% and 0.9% of the total traces, respectively. DSF-BPPVNet was first trained directly on the 4-class task. We used a Weighted Random Sampler together with a standard cross-entropy loss, as this combination provided the most stable performance during model selection. The aggregated 5-fold confusion matrix for the selected model is shown in Table 4.

To further examine the low AC-BPPV F1-score, we performed a post-hoc one-vs-rest analysis using the saved 4-class model. For each subtype, a probability threshold maximizing the F1-score was selected on the validation set and then applied to the corresponding test-set probabilities. The resulting subtype-specific performance is summarized in Table 5.

The OvR results provide additional insight into class-specific behavior. PC-BPPV was the best-performing subtype, with an F1-score of 0.704 ± 0.066, indicating that the model was able to identify this class with comparatively good precision and recall. HC-BPPV showed lower and more variable performance, with an F1-score of 0.324 ± 0.086. AC-BPPV remained the most difficult class, with a low F1-score of 0.056 ± 0.048 and PR-AUC of 0.048 ± 0.040, despite a ROC-AUC of 0.727 ± 0.038. This combination suggests that, although the model captures some ranking information for AC-BPPV, its subtype discrimination remains weak at practically useful thresholds under the present degree of class imbalance. Overall, these findings indicate that subtype-level performance is strongly influenced by class prevalence and separability, particularly for the rare AC-BPPV class.

True Label	Predicted Label			
	Normal	PC-BPPV	HC-BPPV	AC-BPPV
Normal	1745	200	67	72
PC-BPPV	154	619	62	21
HC-BPPV	25	45	68	4
AC-BPPV	16	9	1	3

Table 4. Aggregated 5-fold confusion matrix of the proposed approach.

Class	Precision	Recall	F1-score	PR-AUC	ROC-AUC
PC-BPPV	0.705 ± 0.060	0.709 ± 0.094	0.704 ± 0.066	0.775 ± 0.080	0.881 ± 0.028
HC-BPPV	0.512 ± 0.346	0.435 ± 0.266	0.324 ± 0.086	0.385 ± 0.097	0.777 ± 0.027
AC-BPPV	0.045 ± 0.046	0.247 ± 0.150	0.056 ± 0.048	0.048 ± 0.040	0.727 ± 0.038

Table 5. Post-hoc one-vs-rest (OvR) subtype classification performance. Values are mean ± SD across 5 folds.

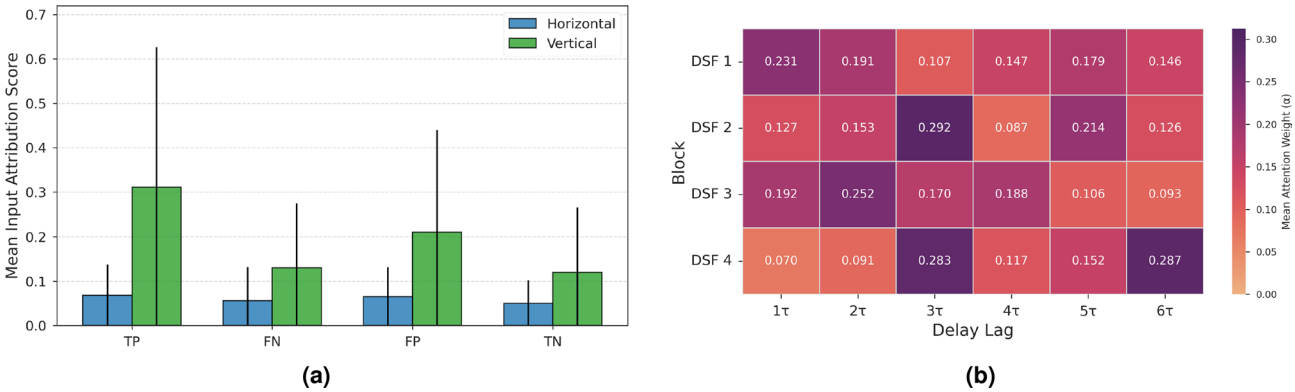


Fig. 4. Analysis of DSF-BPPVNet's learned behavior. (**a**) Aggregated Input × Gradient attribution scores for horizontal and vertical channels, computed across all samples for each prediction outcome. (**b**) Mean learned delay weights (α_i) across blocks and delays, averaged over true-positive cases.

Explainable AI

To better understand the prediction behavior of DSF-BPPVNet, we conducted an explainable AI analysis focusing on both input attribution patterns and the internal behavior of the delayed-state feedback component. Two aspects were examined: input attribution behavior, assessed quantitatively using aggregated Input × Gradient scores and qualitatively using multiple attribution methods, and the behavior of the delayed-state feedback mechanism, assessed through analysis of the learned delay-weight distributions.

The aggregated mean input attribution scores for the horizontal and vertical channels across the four prediction outcomes are shown in Figure 4a. These results reveal distinct attribution patterns for each outcome and provide insight into the model's channel-level emphasis. In true positive cases, the model placed substantially greater attribution on the vertical component, with a mean attribution of 0.311 ± 0.315, compared with 0.068 ± 0.069 for the horizontal component. This pattern suggests that the model places greater emphasis on the vertical channel in correctly identified positive cases. It is also consistent with the predominance of PC-BPPV in the dataset, for which vertical nystagmus is clinically prominent. The analyses of false negatives, false positives, and true negatives show a broadly similar channel-level pattern, suggesting that variation in vertical-channel emphasis is associated with differences in prediction outcome.

To examine the internal behavior of the delayed-state feedback mechanism across processing stages, we analyzed the learned delay weights α_i assigned to each of the six delay indices. Figure 4b presents a heatmap of these weights, averaged across true positive samples from the 5-fold cross-validation, for each of the four convolutional blocks. The weights are relatively distributed across multiple delay lags in the earlier blocks, whereas later blocks assign greater relative emphasis to a smaller subset of delays. In particular, Blocks 2 and 4 assign the highest mean weights to the third delay lag, with average values of 0.292 and 0.283, respectively, while Block 3 places greater weight on the second delay lag, 0.252, and Block 4 also assigns substantial weight to the longest delay, 0.287. These results indicate that the relative contribution of delayed states varies across blocks and suggest that later processing stages place greater emphasis on selected delayed representations.

These aggregated findings are further illustrated in the qualitative analysis of a representative true positive sample shown in Figure 5. The figure presents attribution maps generated using four XAI methods: Guided

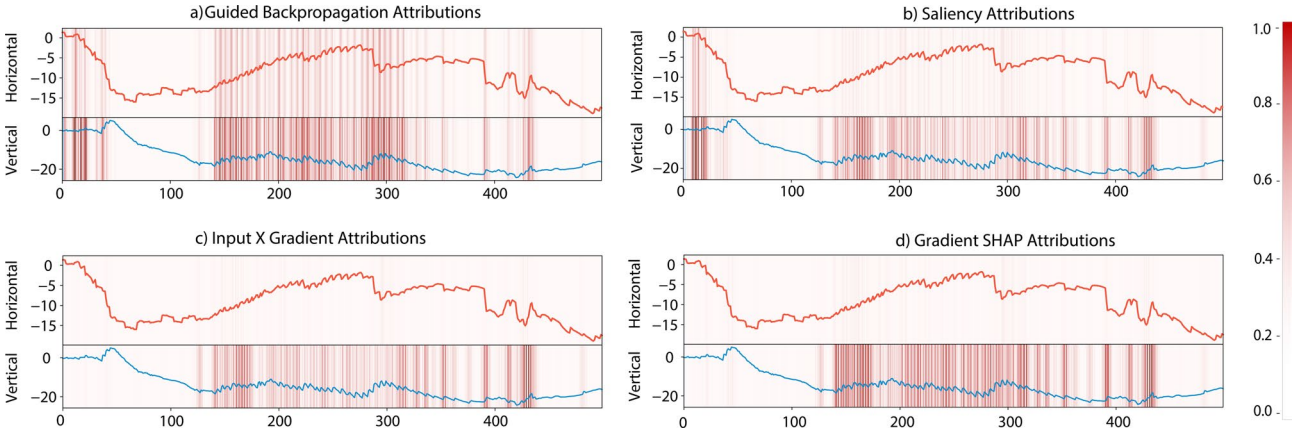


Fig. 5. Feature attributions for the horizontal and vertical components of the input traces, visualized using (**a**) Guided Backpropagation, (**b**) Saliency, (**c**) Input × Gradient, and (**d**) Gradient SHAP methods. Here, the red region represents the magnitude of attribution.

Backpropagation, Saliency, Input × Gradient, and Gradient SHAP. Across methods, the vertical component shows stronger and more spatially concentrated attribution, particularly during periods of rapid change in the sequence. This qualitative pattern is consistent with the vertical-channel dominance observed in the aggregated analysis in Figure 4a. The attributed regions are also concentrated around portions of the trace corresponding to pronounced temporal transitions, including the apparent onset and decay of nystagmus activity. Taken together, these findings suggest that DSF-BPPVNet is influenced by temporally localized regions of the trace, particularly in the vertical channel.

Discussion

We developed DSF-BPPVNet, a neural architecture that combines temporal convolution, delayed-state feedback, and residual refinement for classification of BPPV from eye-movement traces. In internal evaluation, DSF-BPPVNet outperformed the evaluated deep-learning baselines in both the patient-independent and non-patient-disjoint settings. The principal improvement was a more favorable precision–recall balance, driven primarily by higher recall while maintaining competitive precision. Within this cohort, these findings indicate that preservation of temporal structure is important for BPPV classification from VNG recordings.

A plausible explanation for this performance is the model's ability to carry forward and reuse temporally separated information. By incorporating delayed-state feedback, the model reuses information from earlier latent states rather than relying solely on local receptive fields or fully global interactions. This design preserves temporally separated information in eye-movement traces that may be relevant for BPPV classification. In contrast to feature-based methods[11,23,24] and feature-augmented models[25,26] that depend on manually derived descriptors such as slow-phase velocity, onset latency, or nystagmus duration, DSF-BPPVNet learns directly from recorded 2D eye-position trajectories. This provides a complementary approach that reduces dependence on hand-engineered features while maintaining sensitivity to temporal structure in the input signal.

The temporal structure of positional nystagmus is clinically informative, but it should not be interpreted as following a single canonical waveform. In clinical practice, its observed time course may vary across patients and BPPV variants and is influenced by interacting factors including endolymph–cupula mechanics, otoconial sedimentation, canal orientation relative to gravity, and the position of debris within the canal[6,27]. These biomechanical determinants help explain clinically observed variation in onset latency, peak intensity, and decay behavior[5]. Accordingly, DSF-BPPVNet is intended to preserve and integrate temporally separated information in the recorded signal rather than to represent a single canonical waveform. Because the network operates directly on eye-position traces rather than explicitly derived slow-phase velocity signals, the learned representation may reflect a combination of slow-phase dynamics, fast-phase reset structure, and broader waveform morphology.

The explainability and ablation analyses support the internal consistency of this design. Delay-weight distributions and attribution maps indicate that deeper blocks make use of non-adjacent temporal context, consistent with the intended function of delayed-state feedback. The ablation results further show that both delayed-state feedback and residual refinement contributed to its performance, with the full model yielding the strongest results. Together, these findings suggest that the observed gains are linked to the proposed architectural components rather than to model scale alone.

Several considerations are important when interpreting these findings. All experiments were conducted on a single-center dataset. Although patient-independent evaluation provides a more rigorous estimate of internal performance, external validation across other devices and acquisition settings will still be necessary to establish broader robustness. In addition, because the present dataset was acquired using the Epley Omniax system, the extent to which these findings transfer to other VNG platforms or bedside VOG recordings remains to be established through external validation. A related consideration is that the non-patient-disjoint setting permits traces from the same patient to appear across splits. Although the difference between the two evaluation settings

was modest in the present study, the patient-independent setting provides the more rigorous estimate of internal generalization because it removes subject overlap between training and test data.

The present model operates on 2D eye-movement traces and does not incorporate torsional information. Although vertical eye-movement information can still be informative for distinguishing some vertical-canal patterns, including posterior-canal versus anterior-canal BPPV, the absence of torsional information reduces the completeness of the motion representation and may limit lateralization and finer subtype characterization. This is particularly relevant because canal-specific diagnosis in clinical practice depends on the observed pattern of nystagmus across maneuvers and its relation to canal orientation[6]. The current results therefore support the feasibility of the model for classification from 2D VNG traces, while highlighting the potential value of richer motion representations in future work.

Beyond these representational limitations, the binary task considered here distinguishes BPPV from recordings labeled as normal, but does not encompass the full range of clinically relevant positional disorders. In clinical practice, an important diagnostic challenge is distinguishing benign positional nystagmus from central positional nystagmus and other non-BPPV causes of positional vertigo, including vestibular migraine and cerebellar disorders[4,12]. Extension to such settings, together with finer-grained outputs such as canal- and side-localization and differentiation between canalolithiasis and cupulolithiasis[28], will be an important next step in assessing clinical applicability more broadly.

Subtype performance should also be interpreted in the context of substantial class imbalance. The dataset was dominated by posterior-canal BPPV, whereas horizontal-canal and especially anterior-canal cases were much less frequent. The attribution analysis suggested predominant reliance on the vertical channel, which may be consistent with stronger separation of posterior-canal cases but could also be associated with weaker performance for anterior-canal BPPV. More balanced datasets, ideally incorporating 3D eye-movement recordings, will be valuable for refining subtype-level classification and for establishing whether these findings generalize more reliably across less prevalent canal subtypes.

Conclusion

This study presented DSF-BPPVNet, a delay-aware neural architecture for BPPV classification from VNG traces. The model combines temporal convolution, delayed-state feedback, and residual refinement to support classification from temporally structured eye-movement traces. Across the evaluated baselines, DSF-BPPVNet achieved the strongest overall performance in the patient-independent setting and showed similar behavior in the non-patient-disjoint setting. Additional ablation, subtype, and explainability analyses helped characterize the contribution of the proposed components and the model's prediction behavior. Future work should focus on external multi-site validation, more robust evaluation of rare subtype performance, and extension of the framework to broader vestibular classification tasks.

Data availability

Data used in the study is not publicly available however access to the data and code will be made available on request.

Received: 19 January 2026; Accepted: 8 May 2026
Published online: 18 May 2026

References

1. Young, A. S., Rosengren, S. M. & Welgampola, M. S. Disorders of the inner-ear balance organs and their pathways. In *Handbook of Clinical Neurology* Vol. 159 385–401 (Elsevier, 2018). https://doi.org/10.1016/B978-0-444-63916-5.00025-2.
2. Kovacs, E., Wang, X. & Grill, E. Economic burden of vertigo: A systematic review. *Health Econ. Rev.* **9**37. https://doi.org/10.1186/s13561-019-0258-2 (2019).
3. Von Brevern, M. et al. Epidemiology of benign paroxysmal positional vertigo: A population based study. *J. Neurol. Neurosurg. Psychiatry* **78**, 710–715. https://doi.org/10.1136/jnnp.2006.100420 (2006).
4. Argaet, E., Bradshaw, A. & Welgampola, M. Benign positional vertigo, its diagnosis, treatment and mimics. *Clin. Neurophysiol. Pract.* **4**, 97–111. https://doi.org/10.1016/j.cnp.2019.03.001 (2019).
5. Obrist, D. & Hegemann, S. Fluidmechanics of benign paroxysmal positional vertigo (BPPV). In Proceeding of Fifth International Symposium on Turbulence and Shear Flow Phenomena, 1029–1034, (2007), https://doi.org/10.1615/TSFP5.1600 (Begellhouse, Technische Universität München, Munich, Germany).
6. Bhandari, A., Kingma, H. & Bhandari, R. BPPV simulation: A powerful tool to understand and optimize the diagnostics and treatment of all possible variants of BPPV. *Front. Neurol.* **12**, 632286. https://doi.org/10.3389/fneur.2021.632286 (2021).
7. Bisdorff, A. R., Staab, J. P. & Newman-Toker, D. E. Overview of the International Classification of Vestibular Disorders. *Neurol. Clin.* **33**, 541–550. https://doi.org/10.1016/j.ncl.2015.04.010 (2015).
8. Aw, S. T., Todd, M. J., Aw, G. E., McGarvie, L. A. & Halmagyi, G. M. Benign positional nystagmus: A study of its three-dimensional spatio-temporal characteristics. *Neurology* **64**, 1897–1905. https://doi.org/10.1212/01.WNL.0000163545.57134.3D (2005).
9. Levo, H., Aalto, H. & Hirvonen, T. P. Nystagmus measured with video-oculography. *ORL. J. Otorhinolaryngol. Relat. Spec.* **66**, 101–104. https://doi.org/10.1159/000079327 (2004).
10. Wang, H., Yao, Q., Li, Z., Yu, D. & Shi, H. Characteristics of positional nystagmus in patients with horizontal canal canalolithiasis or cupulopathy. *J. Neurol.* **266**, 2475–2480. https://doi.org/10.1007/s00415-019-09435-5 (2019).
11. Winnick, A. A. et al. Automated nystagmus detection: Accuracy of slow-phase and quick-phase algorithms to determine the presence of nystagmus. *J. Neurol. Sci.* **442**, 120392. https://doi.org/10.1016/j.jns.2022.120392 (2022).
12. Ahmadi, S.-A. et al. Modern machine-learning can support diagnostic differentiation of central and peripheral acute vestibular disorders. *J. Neurol.* **267**, 143–152. https://doi.org/10.1007/s00415-020-09931-z (2020).
13. Lim, E.-C. et al. Developing a diagnostic decision support system for benign paroxysmal positional vertigo using a deep-learning model. *J. Clin. Med.* **8**, 633. https://doi.org/10.3390/jcm8050633 (2019).
14. Newman, J. L., Phillips, J. S. & Cox, S. J. 1D convolutional neural networks for detecting nystagmus. *IEEE J. Biomed. Health Inform.* **25**, 1814–1823. https://doi.org/10.1109/JBHI.2020.3025381 (2021).

15. Nguyen, T.-A.-Q. Benign Paroxysmal Positional Vertigo Disorders Classification Using Eye Tracking Data. In *Artificial Intelligence Applications and Innovations* Vol. 712 (eds Maglogiannis, I. et al.) 174–185 (Springer Nature Switzerland, 2024). https://doi.org/10.1007/978-3-031-63215-0_13.
16. Wu, P. et al. Diagnosing the benign paroxysmal positional vertigo via 1D and deep-learning composite model. *J. Neurol.* **270**, 3800–3809. https://doi.org/10.1007/s00415-023-11662-w (2023).
17. Mun, S. B. et al. Deep learning-based nystagmus detection for BPPV diagnosis. *Sensors* **24**, 3417. https://doi.org/10.3390/s24113417 (2024).
18. Qiu, X. et al. Gram-based attentive neural ordinary differential equations network for video nystagmography classification. In 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 21282–21291. https://doi.org/10.1109/ICCV51070.2023.01951 (2023).
19. Han, J.-H. et al. Transformer-based detection and clinical evaluation system for torsional nystagmus. *Sensors* **25**, 4039. https://doi.org/10.3390/s25134039 (2025).
20. Phillips, V. et al. A pilot study of smartphone eye tracking for detection of positional nystagmus. *Digit. Biomark.* **9**, 124–129. https://doi.org/10.1159/000547008 (2025).
21. Dai, S., Wu, Y., Kang, X., Shen, Z. & Zhong, P. Artificial intelligence-driven clinical auxiliary diagnosis of benign paroxysmal positional vertigo. *Front. Neurol.* **17**, 1774729. https://doi.org/10.3389/fneur.2026.1774729 (2026).
22. Oreshkin, B. N., Carpov, D., Chapados, N. & Bengio, Y. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting, https://doi.org/10.48550/ARXIV.1905.10437 (2019).
23. Charoenpong, T., Pattrapisetwong, P. & Mahasitthiwat, V. A new method to detect nystagmus for vertigo diagnosis system by eye movement velocity. In 2015 14th IAPR International Conference on Machine Vision Applications (MVA), 174–177, (2015), https://doi.org/10.1109/MVA.2015.7153161 (IEEE, Tokyo, Japan).
24. Punuganti, S. A. & PhD, J. O.-M. Detection of Saccades and Quick-Phases in Eye Movement Recordings with Nystagmus. In ACM Symposium on Eye Tracking Research and Applications, 1–5, (2020), https://doi.org/10.1145/3379156.3391353 (ACM, Stuttgart Germany).
25. Newman, J. L., Phillips, J. S., Cox, S. J., FitzGerald, J. & Bath, A. Automatic nystagmus detection and quantification in long-term continuous eye-movement data. *Comput. Biol. Med.* **114**, 103448. https://doi.org/10.1016/j.compbiomed.2019.103448 (2019).
26. Ben Slama, A. et al. DBN-DNN: Discrimination and classification of VNG sequence using deep neural network framework in the EMD domain. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **8**, 681–690. https://doi.org/10.1080/21681163.2020.1799867 (2020).
27. Rabbitt, R. D. Semicircular canal biomechanics in health and disease. *J. Neurophysiol.* **121**, 732–755. https://doi.org/10.1152/jn.00708.2018 (2019).
28. Nuti, D., Masini, M. & Mandalà, M. Chapter 18 - benign paroxysmal positional vertigo and its variants. In Furman, J. M. & Lempert, T. (eds.) Neuro-Otology, vol. 137 of Handbook of Clinical Neurology, 241–256, 2016, https://doi.org/10.1016/B978-0-444-63437-5.00018-2 (Elsevier).

Author contributions

K.C. developed the methodology, implemented the models, performed experiments, and wrote the initial manuscript draft. M.W. conceived the study and contributed to clinical interpretation, dataset curation, and domain expertise in vestibular and ocular motor assessment. N.Y. and I.H. contributed to clinical interpretation, dataset curation, and domain expertise. N.R. and A.B. (Andrew Bradshaw) supported data collection, clinical annotation, and validation of findings. M.P., A.B. (Ali Braytee), T.J. and K.T. assisted with model development, analysis, and manuscript revision. M.P., M.W., and A.B. (Ali Braytee) supervised the study, provided overall guidance, and contributed to manuscript writing and refinement. All authors reviewed and approved the final manuscript.

Declarations

Ethical approval and informed consent

This study was approved by the Sydney Local Health District Human Research Ethics Committee (X18-0087 & 2019/ETH06926). All methods were performed in accordance with the relevant guidelines and regulations. The need to obtain informed consent was waived by the Sydney Local Health District Human Research Ethics Committee due to the retrospective nature of the study.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.P.

Reprints and permissions information is available at 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 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/.

© The Author(s) 2026