# A multimodal deep learning tool for detection of junctional ectopic tachycardia in children with congenital heart disease

Yilong Ju, PhD,[*] Jamie L.S. Waugh, MASc,[†] Satpreet Singh, PhD,[‡§] Craig G. Rusin, PhD,[||¶] Ankit B. Patel, PhD,[‡**1] Parag N. Jain, MD[††1]

*From the *Department of Computer Science, Rice University, Houston, Texas, †Division of Research, Medical Informatics Corp., Houston, Texas, ‡Department of Neuroscience, Baylor College of Medicine, Houston, Texas, §Department of Neurobiology, Harvard Medical School, Boston, Massachusetts, ||Division of Cardiology, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, ¶Department of Pediatrics, Texas Children's Hospital, Houston, Texas, **Department of Electrical and Computer Engineering, Rice University, Houston, Texas, and ††Division of Critical Care Medicine, Department of Pediatrics, Baylor College of Medicine, Texas Children's Hospital, Houston, Texas.*

**BACKGROUND** Junctional ectopic tachycardia (JET) is a prevalent life-threatening arrhythmia in children with congenital heart disease. It has a marked resemblance to normal sinus rhythm, often leading to delay in diagnosis and management.

**OBJECTIVE** The study sought to develop a novel multimodal automated arrhythmia detection tool that outperforms existing JET detection tools.

**METHODS** This is a cohort study performed on 40 patients with congenital heart disease at Texas Children's Hospital. Electrocardiogram and central venous pressure waveform data produced by bedside monitors are captured by the Sickbay platform. Convolutional neural networks (CNNs) were trained to classify each heartbeat as either normal sinus rhythm or JET based only on raw electrocardiogram signals.

**RESULTS** Our best model improved the area under the curve from 0.948 to 0.952 and the true positive rate at 5% false positive rate from 71.8% to 80.6%. Using a 3-model ensemble further improved the area under the curve to 0.953 and the true positive rate at 5% false positive rate to 85.2%. Results on a subset of data show that adding central venous pressure can significantly improve area under the receiver-operating characteristic curve from 0.646 to 0.825.

**CONCLUSION** This study validates the efficacy of deep neural networks to notably improve JET detection accuracy. We have built a performant and reliable model that can be used to create a bedside alarm that diagnoses JET, allowing for precise diagnosis of this life-threatening postoperative arrhythmia and prompt intervention. Future validation of the model in a larger cohort is needed.

**KEYWORDS** Arrhythmia detection; Deep learning; Ensembles; Explainable AI; Junctional ectopic tachycardia

(Heart Rhythm O$^2$ 2024;5:452–459) 

## Introduction

Patients with congenital heart disease (CHD) routinely experience life-threatening arrhythmias, especially during the early postoperative period, with junctional ectopic tachycardia (JET) shown to be the most common.[1,2] The occurrence of JET not only extends intensive care unit stay, but also increases postoperative morbidity and mortality.[3] JET is a narrow QRS complex tachyarrhythmia, with electrical activity originating around the atrioventricular node.[4] The distinctive electrocardiographic (ECG) feature of JET is the disappearance of P waves or retrograde P waves.[5] JET often mimics sinus tachycardia, which results in delay in diagnosis and intervention. Hence, there is a need for an accurate, real-time monitoring system dedicated to JET detection.

Waugh and colleagues[6] developed a logistic regression model that takes ECG features as input and predicts the likelihood of JET on a per-cardiac-cycle basis. However, this linear algorithm exhibited a high false positive rate, leading to false alarms and consequent potential alarm fatigue for healthcare providers. Additionally, extracting ECG features required significant time and expertise from medical professionals. Neural networks, on the other hand, are known for their ability to approximate a wide variety of functions and to learn characteristic features automatically from data.[7] Therefore, in this work, we hypothesized an automated approach that leverages recent advances of deep learning (DL) could yield improved performance in JET detection.

[1]Drs Patel and Jain contributed equally to this work. **Address reprint requests and correspondence:** Dr Parag N. Jain, Division of Pediatric Critical Care, Texas Children's Hospital, MC E1420, 6651 Main Street, Houston, TX 77030. E-mail address: drparagjain@gmail.com.


https://doi.org/10.1016/j.hroo.2024.04.014

## KEY FINDINGS

- Deep learning models can be used to build a highly sensitive and specific algorithm for automated detection of postoperative junctional ectopic tachycardia in children with congenital heart disease.
- Use of an ensemble technique allows convoluted neural networks to diagnose and predict postoperative junctional ectopic tachycardia using single heartbeat.
- Explainable artificial intelligence techniques, like LIME (local interpretable model-agnostic explanations), can enable clinicians to understand the decisions made by deep learning models, thus increasing reliability in the model.

Recent work has shown that deep neural networks (DNNs) can be effectively applied to various cardiac dysfunction and arrhythmia detection problems. Attia and colleagues[8] trained a convolutional neural network (CNN) to identify asymptomatic left ventricular dysfunction using paired 12-lead ECG and echocardiogram data.[9] Hannun and colleagues[10] applied their CNN to classify 12 rhythm classes using a 30-second single-lead ECG signal sampled at 200 Hz.[10] Hughes and colleagues[11] trained a CNN to detect the presence of 39 diagnostic classes spanning 5 categories from 10 seconds of 12-lead ECG data sampled at 250 Hz. More related work can be found in the review article by Siontis and colleagues.[12] However, despite the growing body of literature on this topic, the black-box nature of neural networks remains a significant barrier to their clinical adoption.[13–15] The decision-making process of a neural network model needs to be elucidated using explainable artificial intelligence (XAI) techniques, not only to enhance its trustworthiness, but also to verify and/or expand our understanding of the diagnosis. For example, local interpretable model-agnostic explanations (LIME) provides local, interpretable explanations for a complex model by approximating it locally with a simpler model, while SHAP (SHapley Additive exPlanations) explains the model outputs by computing the feature contribution based on game theory.[16,17]

To evaluate our hypothesis, we developed CNNs that take a single heartbeat as input and predict the probability that it is JET. Then, we examined whether the predictions were correct and whether the CNNs had used human-interpretable ECG features for their predictions. Both evaluations are essential to achieving an accurate and trustworthy JET detection model.

In addition, an important feature of JET is the increase in atrial pressures due to the discordant simultaneous contraction of atria and ventricles, resulting in distinctive changes in central venous pressure (CVP) waveforms.[18] Xin and colleagues[19] previously extracted distinctive features of CVP to diagnose JET. In our study, we also investigated whether a multimodal approach that combines ECG and CVP will result in improved performance of JET detection.

# Methods

## Design

We performed a retrospective single-center cohort study of all postoperative patients with CHD admitted to the cardiac intensive care unit at Texas Children's Hospital, a large tertiary care children's hospital. The Institutional Review Board of Baylor College of Medicine approved the study and waived the need for informed consent, as this was an observational study performed on aggregated de-identified patient information.

Patient cohort selection and data collection are described in Waugh and colleagues.[6] In brief, patients in the cardiac intensive care unit at Texas Children's Hospital are continuously monitored using standard bedside monitoring equipment. The Sickbay platform (Medical Informatics Corp.) is enabled on all beds at Texas Children's Hospital to capture and time-synchronize all physiologic data produced by all patient monitoring devices at their native resolution.

The ECG dataset comprising 40 patients was partitioned into 3 subsets. Specifically, 15 patients were randomly selected for the training set, while the remaining 25 patients were assigned to the held-out test set. To fine-tune the model hyperparameters, 3 patients were randomly selected from the training set to create a validation set. These partitions yielded training, validation, and test sets composed of 317,545, 44,723, and 114,412 ECG heartbeat cycles, derived from 12, 3, and 25 patients, respectively. The train-validation split ratio in terms of the number of heartbeats ($\sim$88/12) was slightly different from the typical 80/20, as we performed an 80/20 on the patients, not on the heartbeats. Because we observed that the heartbeats within a patient are very similar to each other, we argue that the ratio in terms of the number of patients matters more. One supporting evidence was that during our preliminary experiments, subsampling the training data resulted in near zero performance degradation.

To probe the value of CVP data in enhancing the JET detection performance, a specialized subdataset was curated, which only contained those segments of ECG data that were accompanied by temporally aligned CVP signals. We call this dataset the CVP subset. Using the same patient-based partition, the CVP subset was divided into training, validation, and test sets comprising 299,201, 29,988, and 82,583 ECG-CVP cycles, respectively. Similarly, the split ratio in terms of the number of patients was 80/20 here.

## ECG signal processing

We followed Waugh and colleagues'[6] approach for ECG signal processing but skipped the feature extraction steps because CNNs require only raw ECG signals. Specifically, we focused on lead II and excluded segments exhibiting movement artifacts or nonphysiological data. The retained ECG segments were then filtered to remove frequencies outside the range of 5 to 50 Hz. Following a normalization based on the segment median and interquartile range, R-wave peaks were identified using MATLAB's *findpeaks* function with adaptively adjusted parameters.

### ECG data augmentation

Data augmentation is a technique widely used in DL to improve model performance. By transforming the original data in various ways that simulate natural variations, DNNs are able to learn how to deal with slightly distorted data and become robust to these variations. We found 11 transformations for data augmentation described in the literature.[20,21] However, these were proposed for different arrhythmia detection problems. To determine the most effective transformations for JET vs normal sinus rhythm (NSR) classification, we performed preliminary experiments in which only 1 transformation was applied at a time. Based on the most common false positives and false negatives that we have seen, we selected 4 of them that simulate the errors during training and improve the results: adding Gaussian noise, baseline wandering, temporal warping, and temporal displacement (see Supplemental Figure A.1 for details). We also found that the best number of transformations to apply equaled 1.25 (see Supplemental Figure A.2 for details).

### CVP signal processing

Example CVP (and ECG) morphology is shown in Figure 1, in which distinctive patterns in CVP can be seen to correspond to specific rhythms. In NSR heartbeats, the CVP has a trough at the location of the P-wave in the ECG, while for JET heartbeats, the CVP instead has a peak at the typical location of the P-wave, consistent with electromechanical dissociation observed in JET. This observation indicates that CVP signals contain valuable information that may contribute to improving classification performance.

### Main outcome measures

Given that our goal was to develop a clinical bedside JET alarm, reducing alarm fatigue was of utmost importance. We used the area under the receiver-operating characteristic curve (AUROC) and the true positive rate (TPR) at the 5% false positive rate (FPR) level (TPR @ 5% FPR), on a held-out test dataset as the 2 main outcome measures of this study. AUROC is the overall classification performance, for each threshold used to distinguish between the 2 classes.[22]

### Classification models

We used a 5-layer CNN adopted from Kiyasseh and colleagues (see Supplemental Table A.1 for architecture).[21] Each input to the CNN is a single heartbeat of ECG samples (240 Hz), segmented between 2 consecutive QRS peaks and resampled to a 1-dimensional vector of length 300. When CVP data is available, we resample it to the same shape and concatenate it to the ECG vector, forming a $2 \times 300$ input. See Supplemental Appendix A for our hyperparameter choices.

### Explaining model decisions

We employ LIME to explain the decisions that the CNN makes.[16] LIME works as follows: Suppose that we would like to understand why the CNN classifies a given heartbeat $\boldsymbol{x}$ as JET. We perturb $\boldsymbol{x}$ by replacing a segment of the heartbeat with the mean signal strength, repeat this action multiple times, and collect these perturbed heartbeats as $\boldsymbol{X}^{\dagger}$. We then feed $\boldsymbol{X}^{\dagger}$ into the CNN and obtain the corresponding predictions $\boldsymbol{Y}^{\dagger}$. Then, a linear model predicting $\boldsymbol{Y}^{\dagger}$ on $\boldsymbol{X}^{\dagger}$ is fitted, which well approximates the CNN in the vicinity of $\boldsymbol{x}$. By interpreting the coefficients of the linear model, we can then understand how important each segment is to predicting $\boldsymbol{x}$ as JET.

### Ensembles

One major source of error is that the CNN is sometimes not confident in its predictions. One way to mitigate this issue

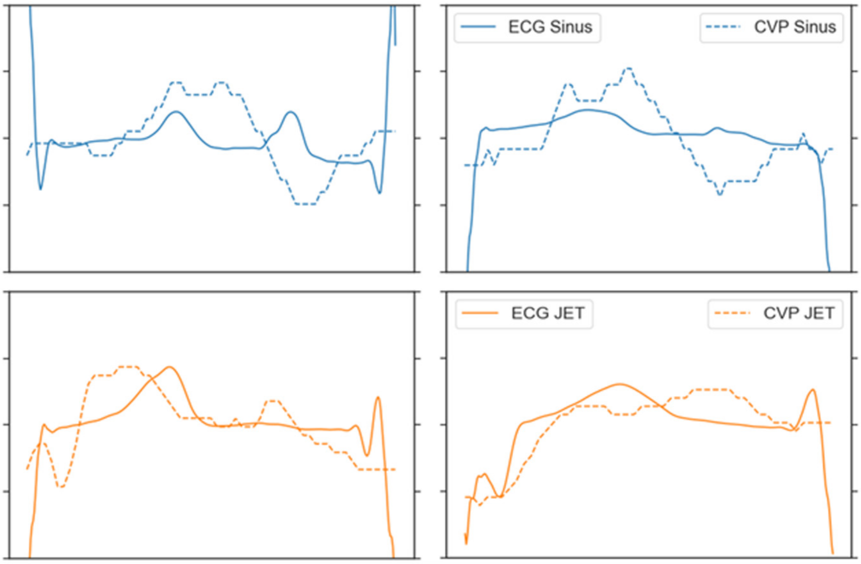


**Figure 1** Electrocardiogram (ECG) and corresponding CVP morphology. We randomly sample 2 sinus (top) and 2 junctional ectopic tachycardia (JET) (bottom) beats from our data for visualization. ECG signals are shown in solid lines and central venous pressure (CVP) signals are in dashed lines. The R-R intervals are shown for ECG and important ECG features are labeled.

is to use ensembles, in which instead of using a single CNN to make predictions, we average the predicted probability made by multiple CNNs for the final decision.[23] This can improve performance and, more importantly, provide us a reliable uncertainty metric. When the uncertainty is high, the CNN may abstain from making risky predictions and ask for human assistance.

## Results

In Figure 2, the proposed model development process with human feedback is illustrated. Sickbay is used to continuously and passively collect all bedside data. After preprocessing, heartbeat by heartbeat of labeled ECG data are fed into a DNN, which predicts whether each beat is NSR or JET. In addition to evaluating the model performance, a root cause error analysis is performed by applying an XAI method to understand how the CNN makes its predictions. Errors are then categorized based on the LIME visualization, by electrophysiologists. Finally, the model is updated to improve performance using tools such as data augmentation focusing on variations in the morphology of the identified error examples, and applying ensembles. This process is repeated until no significant gain is obtained.

### Classification performance

Table 1 reports the classification performance of the 4 selected trained models. AUROC and TPR @ 5% FPR were calculated on the held-out test dataset. The logistic regression model (LRM) from Waugh and colleagues was used as a baseline. It is evident that the DNNs significantly improve TPR @ 5% FPR. The best CNN ensemble model achieved a TPR of 85.2%, outperforming the 71.8% TPR

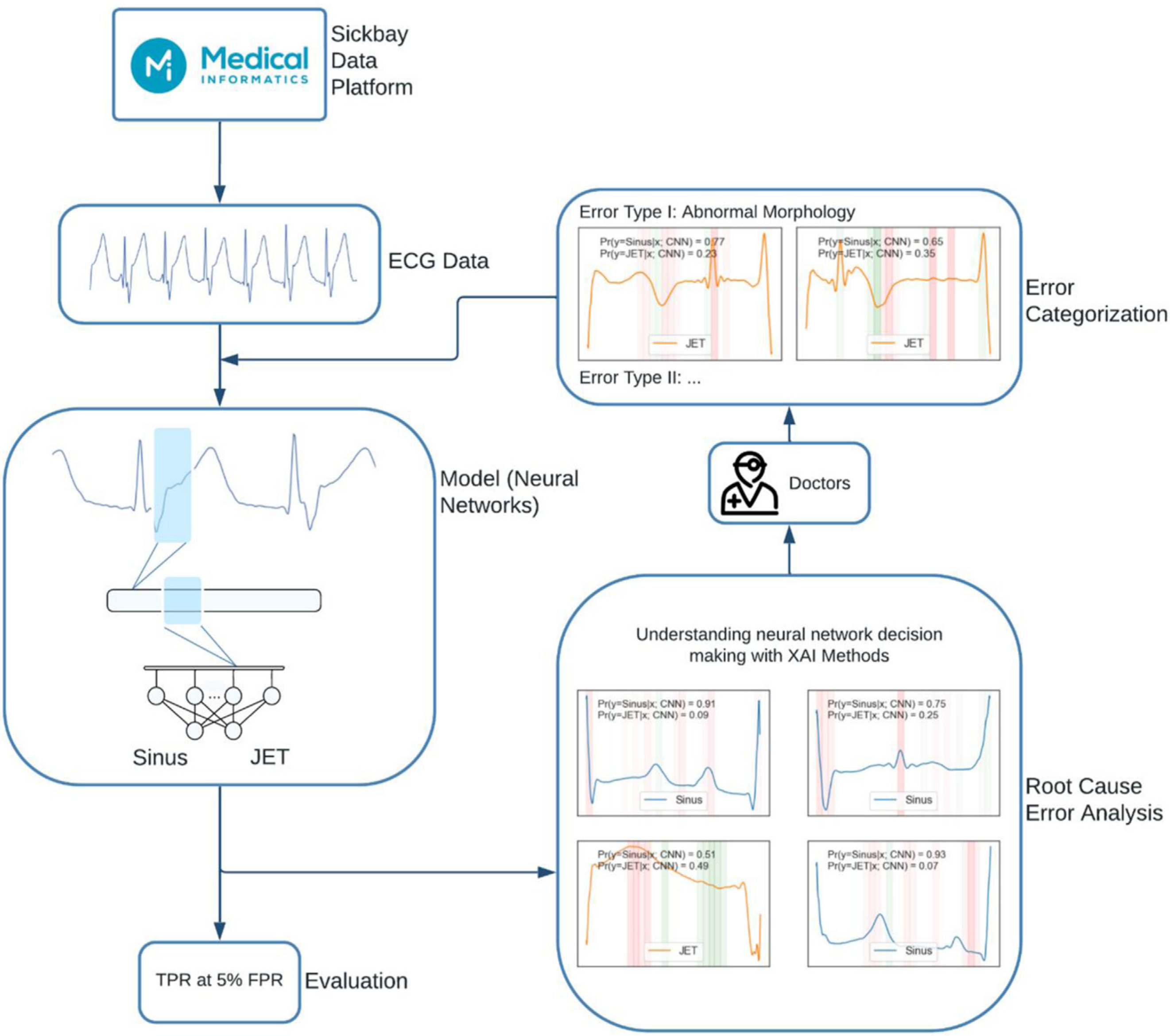


**Figure 2** An overview of the model development process with human feedback. CNN = convolutional neural network; ECG = electrocardiogram; FPR = false positive rate; JET = junctional ectopic tachycardia; TPR = true positive rate.

**Table 1** Classification performance of the models on the held-out test dataset.

| Model | AUROC | TPR at 5% FPR |
|---|---|---|
| Logistic regression[6] | 0.948 | 71.8% |
| CNN | 0.952 | 80.6% |
| CNN (3-model ensemble) | 0.953 | 85.2% |
| CNN (5-model ensemble) | 0.950 | 84.0% |

AUROC = area under the receiver-operating characteristic curve; CNN = convolutional neural network; FPR = false positive rate; TPR = true positive rate.

of the LRM. It is worth noting that during development, we did not aim for an improvement in the TPR at a specific FPR, but rather we focused only on improving the AUROC, as, in general, a higher AUROC leads to a higher TPR at any FPR level. On the CVP subset, we also found that using CVP in addition to ECG improves the test AUROC from 0.645 to 0.825, compared with using just ECG.

## Error analysis

To understand where the LRM fails, we examined heartbeats predicted incorrectly and categorized the major types of error. We show an example of each type in Figure 3. In the first row, for false positives (NSR predicted as JET), the top 3 error types are (1) P-wave falsely detected as T-wave, in which P waves are falsely detected as T waves by the peak feature extraction (PFE) algorithm used in Waugh and colleagues,[6] which leads to no P waves detected; (2) slanted baseline of P-wave, in which the slanted baseline is not identified by the PFE, resulting in an incorrect estimate of peak prominence; and (3) inverted P/T waves, which also confused the PFE and led to feature values that do not make sense. The second row in Figure 3 shows the top error types for false negatives (JET predicted as NSR): (1) other peaks detected as P-wave, which again caused the PFE to extract features from an incorrect location; (2) mislabeled heartbeats, which can happen if, for example, a few JET heartbeats are inside a big chunk of NSR-labeled heartbeats; and (3) abnormal morphology, due to patient movement, incorrect detection of the QRS complex, etc.

We observed these error types during evaluation, even though the PFE had been iterated upon a number of times. This suggests that there are too many factors to consider for a manually designed PFE to be perfect. One solution is building an end-to-end DNN that does not require an explicit PFE. Hence, we augment the training data with transformations accounting for the majority of variation which causes the observed error (see Supplemental Figure A.3 for examples of transformed heartbeats). We applied baseline wandering, which tilts the baseline of the heartbeat, so that the CNN learns to detect peaks with slanted baselines. Adding random temporal warping and displacement enables the CNN to detect T/P waves at slightly abnormal locations. We also added Gaussian noise so that the CNN is robust to local fluctuations. We then fed the original and transformed heartbeats into the CNN and obtained the models reported in Table 1.

## Explanation of neural network predictions

After training the CNNs, we again performed an error analysis. Because CNNs are black boxes, we employed LIME to analyze which heartbeat segments contribute to the final prediction. In Figure 4, the contribution of each segment is shown in green/red masks. The darker the green/red, the

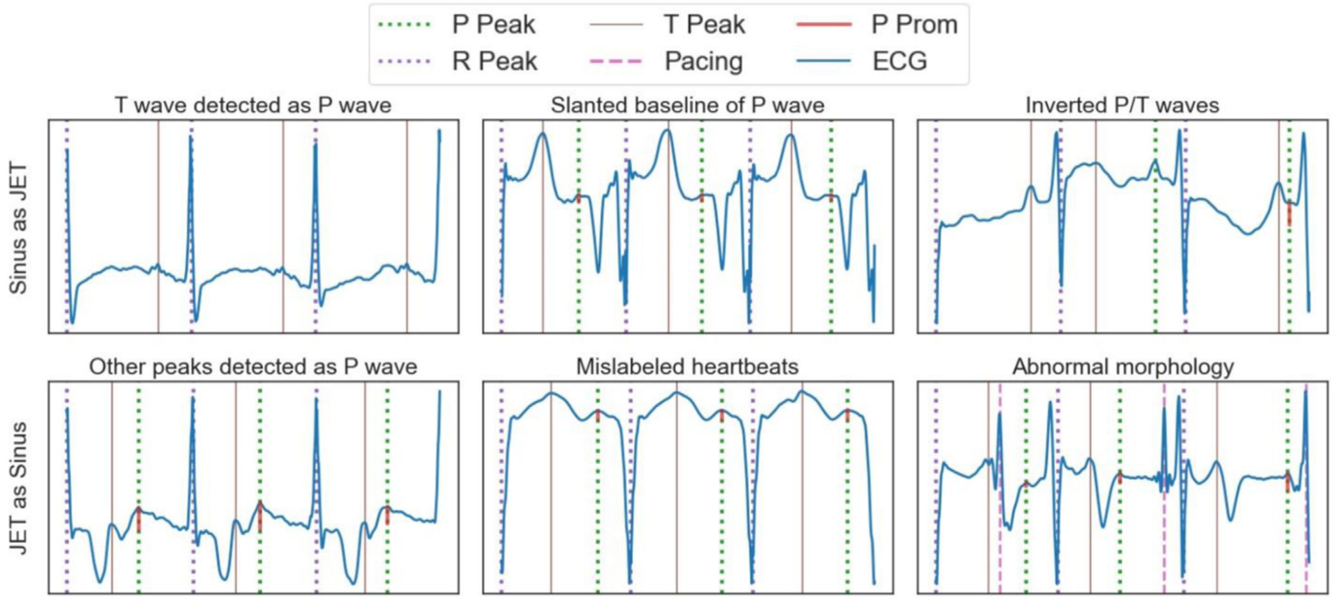


**Figure 3** Logistic regression model error types. We show one representative example for each type of error made by the logistic regression model in Waugh and colleagues.[6] We also label the peaks with the peak-finding algorithm used in the same work. The specific reason for each error is shown in the title of each subfigure. Top row: Normal sinus rhythm (NSR) predicted as junctional ectopic tachycardia (JET). Bottom row: JET predicted as NSR. ECG = electrocardiogram.

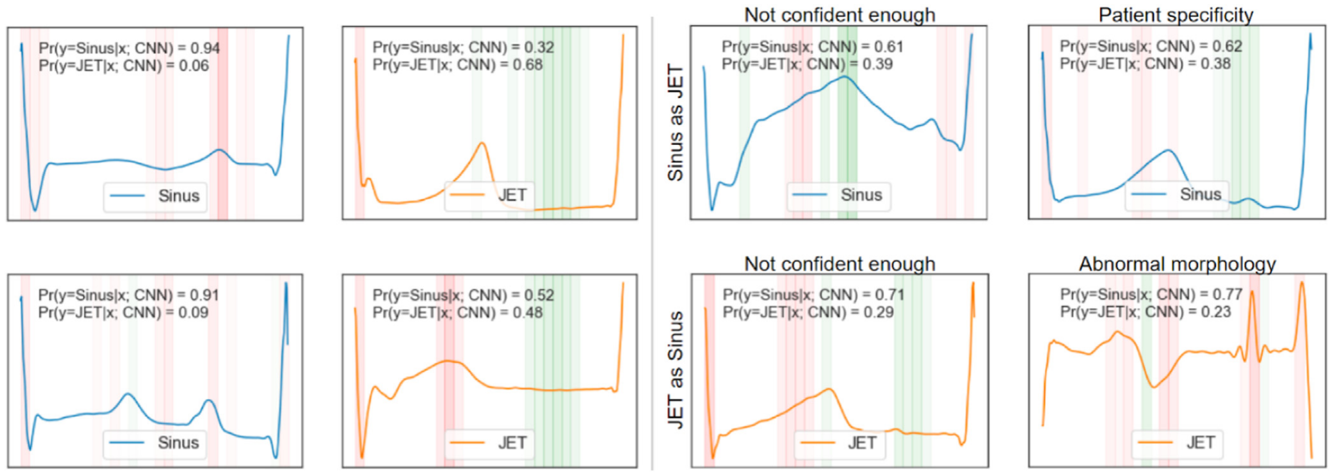


**Figure 4** Explaining convolutional neural networks' (CNNs') decision making with local interpretable model-agnostic explanations (LIME). The LIME masks show that the CNN focuses on the typical P-wave area, where the presence of a P-wave contributes negatively to junctional ectopic tachycardia (JET) prediction and its absence contributes positively. The predicted probability is show for each class. A threshold of 0.38 is used, which corresponds to 5% false positive rate (FPR) in the validation set. Left: Examples of correct predictions. Right: Examples of errors. LIME enables us to understand why these errors occur – whether the convolutional neural network (CNN) fails to identify the key P-wave feature, the patient has a unique JET/normal sinus rhythm (NSR) electrocardiogram (ECG) morphology (e.g., due to pacing spikes), or there is an abnormal beat.

greater the impact the segment has on predicting the heartbeat as JET/NSR. The probabilities of NSR and JET predicted by the CNN are also shown at the top of each heartbeat. On the left side of Figure 4, interestingly, the LIME masks focus on the P waves: the presence of P waves contributes to a prediction of NSR, while their absence contributes to a prediction of JET. This indicates that the CNN may have learned to use P-wave features during its decision-making process. On the right side of Figure 4, we categorize the major error types made by the single CNN (80.6% TPR) with the help of LIME. For example, one source of FPs is that although the CNN pays attention to the P-wave, the P prominence is not large enough to be considered "present" and result in a prediction of NSR. In fact, upon inspecting other heartbeats from this specific patient, we found that all P prominences were smaller than average (compared with other patients). Thus, this is a patient-specific issue, and one possible resolution is to develop an automated pipeline to fine-tune a patient-specific model. As for the cause of FNs, abnormal morphology is still a factor, which could potentially be resolved by introducing more training data. This is left as future work. We include a number of additional LIME mask visualizations in Supplemental Figures A.4 to A.6, to more exhaustively illustrate how the CNN makes predictions.

## Discussion

Our primary objective was to enhance the diagnosis of JET in postoperative patients with CHD using CNNs. Our results demonstrate a substantial improvement in the TPR of automated JET detection. In addition, by using LIME for model interpretation, we were able to validate that the CNN correctly learned the characteristic features of JET, namely the absence of P waves, thereby substantiating the reliability and effectiveness of our DL approach in a clinical setting.

Compared with previous arrhythmia detection work, our model is designed to analyze a single heartbeat cycle resampled to only 300 samples, opposed to an extended period oof heartbeats (tens of seconds) consisting of thousands of samples.[8,10,11] This leads to a more cost-efficient model capable of a finer-grained level of JET detection. Leveraging the per-cardiac-cycle prediction, we can architect higher-order JET diagnostic strategies for improved robustness—for instance, diagnosing JET only when ***m*** out of ***n*** consecutive heartbeats are classified as JET, where ***m*** and ***n*** are calibrated based on specific performance criteria.

Also, by using the ensemble method, we are not only able to further improve the prediction performance, but also obtain a high-quality uncertainty estimation. By using the variance of the 5 predicted JET probabilities made by the 5-model ensemble as the uncertainty metric, we are able to identify (see details in Supplemental Figure A.7) that heartbeats that are the most uncertain to the ensemble are those with high volatility, many peaks, or even incorrect QRS segmentation; the heartbeats that are the most "certain" to the ensemble are those with very simple morphology and clear T/P waves. These findings not only enhance our understanding and enable automatic abstention, but also suggest pathways to improving the algorithm's performance, specifically, by making the algorithm more robust against noise, atypical peaks, and segmentation errors. Possible strategies include adding a denoising step in data processing, augmenting data with pacing spikes, and training the model using inputs that comprise 2 or more heartbeats.

## Limitations

This study's generalizability is currently circumscribed to a similar clinical setting due to the single-center data source and the specific patient demographic from postoperative pediatric CHD patients. The external validity of these findings to a broader population or to different clinical environments remains a subject of further investigation. Additionally, as we plan to deploy the algorithm prospectively, the real-time performance of the model remains to be evaluated. While 5% FPR is relatively low, it can still generate high rates of false alarms based on how frequently a classification is made. However, there is no current gold standard of clinically acceptable FPR related to critical alarms.. To give a rough estimate, there are 114,412 heartbeats from 25 patients in the test set, which leads to 114,412 $\times$ 5% / 25 $\approx$ 229 false alarms per patient. The average ECG recording duration for these patients is 39 minutes. Hence, on average, we have about 6 false alarms per minute. While this is still not perfect, we could use the aforementioned higher-order diagnostic strategies to alleviate it.

Although our approach can be generally applicable to any number of classes, we focused on a binary classification problem and did not verify the model performance in multiclass settings. As for the effectiveness of LIME, while we can verify with it that the CNN is indeed using the P-wave as an important feature, it also sometimes generates uninterpretable masks around the T-wave, potentially causing more confusions. One possible explanation is that as LIME may not capture global model behavior, it sometimes can be very sensitive to local behavior and show spurious masks around nonimportant features.[24] Therefore, future studies including multicenter data, diverse patient demographics, and varied clinical settings are essential to ascertain the broader applicability and generalizability of our CNN-based approach to the detection of JET, as well as other arrhythmias. More advanced XAI techniques also need to be applied for better understanding.

## Conclusion

This study aimed to enhance the diagnosis of JET on a per-cardiac-cycle basis in postoperative pediatric patients using CNNs. The results show a significant improvement in the JET detection TPR, and the features learned by the CNNs were verified to be human-interpretable, underscoring the reliability and potential of the proposed DL approach for clinical applications. The added value of CVP in the diagnosis of JET is also demonstrated. This work lays the foundation for the development of a bedside alarm system for precise JET diagnosis and prompt intervention.

**Funding Sources:** This study was supported by Texas Children's Hospital Pediatric Pilot Award as well as funding from the National Institutes of Health (1R01HL142994).

**Disclosures:** Craig G. Rusin reports a conflict of interest with Medical Informatics Corp. Medical Informatics Corp. did not financially support this work. All other authors report no conflicts of interest.

**Authorship:** All authors attest they meet the current ICMJE criteria for authorship.

**Patient Consent:** The Institutional Review Board of Baylor College of Medicine waived the need for informed consent, as this was an observational study performed on aggregated de-identified patient information.

**Ethics Statement:** The Institutional Review Board of Baylor College of Medicine approved the study.

## Appendix
### Supplementary data

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.hroo.2024.04.014.

## References

1. Sahu MK, Das A, Siddharth B, et al. Arrhythmias in children in early postoperative period after cardiac surgery. World J Pediatr Congenit Heart Surg 2018;9:38–46.
2. Talwar S, Patel K, Juneja R, Choudhary SK, Airan B. Early postoperative arrhythmias after pediatric cardiac surgery. Asian Cardiovasc Thorac Ann 2015;23:795–801.
3. Dodge-Khatami A, Miller OI, Anderson RH, Gil-Jaurena JM, Goldman AP, de Leval MR. Impact of junctional ectopic tachycardia on postoperative morbidity following repair of congenital heart defects. Eur J Cardiothorac Surg 2002;21:255–259.
4. Entenmann A, Michel M, Egender F, Hessling V, Kramer HH. Impact of different diagnostic criteria on the reported prevalence of junctional ectopic tachycardia after pediatric cardiac surgery. Pediatr Crit Care Med 2016;17:845–851.
5. Mond HG, Haqqani HM. The electrocardiographic footprints of atrial ectopy. Heart Lung Circ 2019;28:1463–1471.
6. Waugh JLS, Patel R, Ju Y, Patel AB, Rusin CG, Jain PN. A novel automated junctional ectopic tachycardia detection tool for children with congenital heart disease. Heart Rhythm O2 2022;3:302–310.
7. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw 1989;2:359–366.
8. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med 2019;25:70–74.
9. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60:84–90.
10. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019;25:65–69.
11. Hughes JW, Olgin JE, Avram R, et al. Performance of a convolutional neural network and explainability technique for 12-lead electrocardiogram interpretation. JAMA Cardiol 2021;6:1285–1295.
12. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat Rev Cardiol 2021;18:465–478.
13. Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multimodal and multi-centre data fusion: a mini-review, 2 showcases and beyond. Inf Fusion 2022;77:29–52.
14. Quinn TP, Jacobs S, Senadeera M, Le V, Coghlan S. The three ghosts of medical AI: can the black-box present deliver? Artif Intell Med 2022;124:102158.
15. Zhang Z, Beck MW, Winkler DA, et al. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. Ann Transl Med 2018;6:216.
16. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2016. p. 1135–1144.
17. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates; 2017. p. 4768–4777.

18. Pittman JA, Ping JS, Mark JB. Arterial and central venous pressure monitoring. Int Anesthesiol Clin 2004;42:13–30.
19. Tan X, Dai Y, Humayun AI, Chen H, Allen GI, Jain PN. Detection of junctional ectopic tachycardia by central venous pressure. Artif Intell Med Conf Artif Intell Med (2005-) 2021;12721:258–262.
20. Raghu A, Shanmugam D, Pomerantsev E, Guttag J, Stultz CM. Data augmentation for electrocardiograms. Proc Mach Learn Res 2022;174:282–310.
21. Kiyasseh D, Zhu T, Clifton DA. Clocs: contrastive learning of cardiac signals across space, time, and patients. Proc Mach Learn Res 2021;139:5606–5615.
22. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5:1315–1316.
23. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates; 2017. p. 6405–6416.
24. Rao S, Mehta S, Kulkarni S, Dalvi H, Katre N, Narvekar M. A study of LIME and SHAP model explainers for autonomous disease predictions. In: 2022 IEEE Bombay Section Signature Conference (IBSSC). IEEE; 2022. p. 1–6.