Human-centered Explainable Psychiatric Disorder Diagnosis System using Wearable ECG Monitors

Hung Nguyen[1,2,⋆][0000−0002−6750−9536],
Alireza Rahimi[1][0009−0006−6512−7303], Veronica Whitford[1][0000−0001−5681−9929],
Hélène Fournier[2][0000−0002−8470−3226], Irina Kondratova[2][0000−0002−5406−2309],
René Richard[2,1][0000−0002−1342−6225], and Hung Cao[1][0000−0002−0788−4377]

[1] University of New Brunswick, Fredericton, Canada
{hung.ntt,alireza.rahimi,veronica.whitford,hcao3}@unb.ca
[2] National Research Council Canada, Fredericton, Canada
{helene.fournier,irina.kondratova,rene.richard}@nrc-cnrc.gc.ca

Abstract. Psychiatric disorders (PDs), such as schizophrenia (SZ) and bipolar disorder (BP), significantly impact global populations, yet their diagnosis remains heavily reliant on subjective clinical methods. This paper presents the PD diagnosis system that integrates wearable electrocardiogram (ECG) monitoring, Time Series Convolutional Attention Network, dual visual and textual explanations with Explainable AI (XAI) and Large Visual Language Models (LVLMs), and user-centered interfaces to ensure model transparency, increase applicability in clinical settings. By investigating the relationship between PDs and heart rate variability (HRV), this work paves the way for more objective and accessible clinical assessments with wide-ranging applications in mental health diagnostics. We evaluate our system on the detection of PDs, demonstrating superior performance compared to recent literature models and providing interpretable explanations for model decisions. Additionally, we showcase the system's applicability to a related use case, highlighting its scalability and potential for widespread adoption.

1 Introduction

Psychiatric disorders (PDs), such as schizophrenia (SZ) and bipolar disorder (BP), affect 1–3% of the global population [19]. SZ is characterized by psychotic symptoms (e.g., delusions, hallucinations) and negative symptoms (e.g., alogia, blunted affect), whereas BP involves extreme mood swings with manic and depressive episodes, sometimes accompanied by psychotic features [21]. Despite their differences, SZ and BP share genetic overlap and cognitive deficits, impacting quality of life. Current diagnostic methods, based on self-reports and clinical interviews, are subjective, time-consuming, and reliant on clinician expertise. Efforts to improve objectivity, such as standardized scales (e.g., Positive and Negative Syndrome Scale/PANSS) and blood biomarkers, face challenges

⋆ Corresponding author

like accessibility and cost [8]. Wearable technology offers a promising alternative, with heart rate variability (HRV)—an autonomic dysregulation marker, emerging as a non-invasive biomarker for SZ and BP [2]. HRV can be measured by consumer-grade devices and integrated with artificial intelligence (AI) to enhance diagnostics. However, prior approaches often relied on less precise sensors or focused on heart rate rather than HRV [9]. Additionally, AI's "black-box" nature poses interpretability challenges, prompting calls for transparency under regulations. In response to these challenges, our research makes the following key contributions:

1. **Psychiatric Disorder Diagnosis System:** We introduce a comprehensive system integrating single-lead electrocardiogram (ECG) data with the Time Series Convolutional Attention Network (TSCAN) to capture local and global temporal patterns for detecting PDs. By utilizing consumer-grade wearable ECG monitors and user-friendly interfaces, our system promotes widespread adoption in clinical settings for more accessible mental health diagnostics.
2. **Dual Visual and Textual Explanations:** The system combines dual Explainable AI (XAI) methods, i.e. attention-based and gradient-based visual explanations, with Large Vision Language Models (LVLMs) to generate textual insights, enhancing model interpretability and clinician trust.
3. **Evaluation and Applicability:** We evaluate the system on the PD detection task, achieving superior performance with TSCAN compared to recent literature models and providing interpretable decisions through dual explanations. Our work explores the relationship between HRV and PDs, paving the way for easier assessments and wider clinical applications. We also demonstrate the system's adaptability to a related use case, showcasing its generalizability across mental health applications.

2 Related Work

2.1 Detection of Psychiatric Disorders

The detection of PDs has increasingly benefited from AI techniques. While various AI and ML techniques have been applied to predict SZ, BP, or their episodes, relatively few studies have specifically utilized HRV or ECG data for this purpose. The use of wearable devices capable of recording HRV/ECG data has opened new avenues for PD detection. Recently, Buza et al. [4] used Convolutional Nearest Neighbor for detecting PDs using R-R intervals (RRI) sequences recorded from the wearable Polar H10 device. Further exploring the relationship between HRV and PDs, Corponi et al. [6] also employed Bayesian analysis to examine physiological changes during acute episodes of BP, focusing on how HRV patterns change during different disorder phases.

2.2 Human-centered Explainable Systems

The XAI field has grown rapidly, driven by the demand for transparency in AI, particularly in sensitive areas like healthcare [18]. In PD detection, XAI

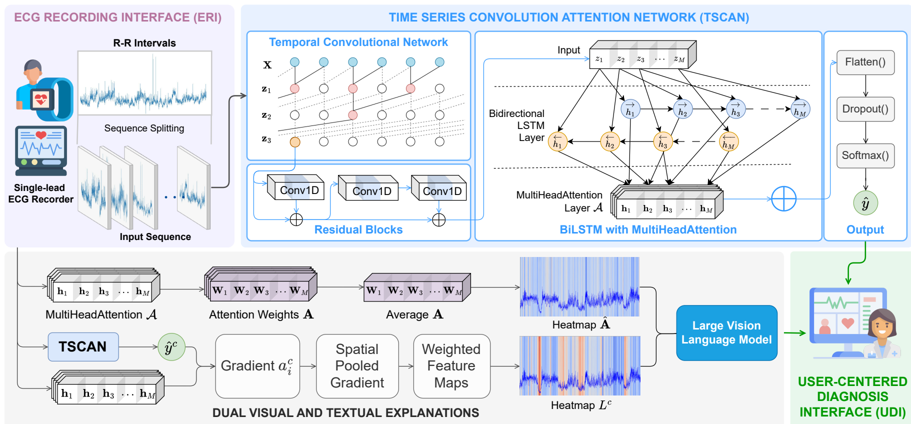


Fig. 1: The Psychiatric Disorder Diagnosis System includes (1) ECG Recording Interface, (2) Time Series Convolutional Attention Network, (3) Dual Visual and Textual Explanations, and (4) User-centered Diagnosis Interface.

applications enhance trust and ethical AI use by offering clinicians insights into diagnostic decisions. Techniques like SHAP [14] and GradCAM [20] have been applied to interpret models analyzing SZ classification [1,10] and motor activity data [15]. These advancements promote human-centered, transparent systems that present insights via visual and textual explanations, accessible even to those without AI expertise.

Efforts to make AI systems more transparent and user-friendly focus on aligning explanations with human psychology [22], providing interactive interfaces [3], and generating textual explanations [16]. LVLMs have emerged as powerful tools for blending language understanding with visual reasoning, excelling in tasks like visual question answering and multi-modal learning. These capabilities present new opportunities for enhancing explainability in visual perception tasks through textual explanations, advancing human-centered XAI systems [16,17].

3 Psychiatric Disorder Diagnosis System

In this section, we present our PD diagnosis system (Fig. 1) integrating an ECG Recording Interface (ERI), the Time Series Convolutional Attention Network (TSCAN), Dual Visual and Textual Explanations, and a User-centered Diagnosis Interface (UDI) to enhance accessibility and interpretability for clinicians.

3.1 ECG Recording Interface (ERI)

The ERI is designed to capture and process the ECG data from users. This interface integrates seamlessly with wearable devices to record real-time ECG data, calculate the RRI sequence and store data in the database. Users can initiate the ECG recording only by activating the wearable devices for connectivity, and then the ECG signals are streamed to the interface. After the ECG is recorded for a set amount of time (at least 70 minutes), when the interface prompts users to stop, the ECG is saved. After acquiring the ECG signals, incorporated algorithms detect R-peaks and then calculate the RRI. The raw ECG data and the calculated RRI sequence data are stored in a database shared with clinicians.

3.2 Time Series Convolutional Attention Network (TSCAN)

TSCAN integrates a Temporal Convolutional Network, residual blocks, and bidirectional LSTMs (BiLSTMs) with multi-head attention to capture both local and global temporal dependencies in RRI sequences, which adapts Temporal Convolutional Attention Network (TCAN) [7] to the time-series input.

Temporal Convolutional Network employs stacked 1D convolutional layers with filter sizes f_1 and f_2, capturing temporal patterns at varying scales. Each layer applies batch normalization (BN) and ReLU activation:

$$\mathbf{z}_1 = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv1D}_{f_1}(\mathbf{X}))), \quad \mathbf{z}_2 = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv1D}_{f_2}(\mathbf{z}_1))) \tag{1}$$

followed by a max pooling $P = 2$ to downsample the feature maps to achieve $\mathbf{z}_3$.

Residual Blocks Two residual blocks capture complex patterns and mitigate vanishing gradients. The first block creates a shortcut connection via 1×1 convolution and max pooling:

$$\mathbf{z}_{\varphi_1} = \mathrm{MaxPooling1D}(\mathrm{Conv1D}_{f_2}(\mathbf{X})), \quad \mathbf{z}_4 = \mathrm{ReLU}(\mathbf{z}_{\varphi_1} + \mathbf{z}_3) \tag{2}$$

followed by the second block stacking two convolutional layers with filter sizes f_3 and f_4 and adding a shortcut connection.

BiLSTM with Multi-Head Attention The BiLSTM layer processes residual block outputs in both directions, generating concatenated hidden states $\mathbf{h}_m = [\overrightarrow{\mathbf{h}_m}; \overleftarrow{\mathbf{h}_m}]$. Multi-head attention ($\mathcal{A}$) computes queries $\mathbf{Q}$, keys $\mathbf{K}$, values $\mathbf{V}$ and attention scores $\mathbf{A}$: Attention outputs are concatenated and passed through a linear layer for final aggregation:

$$\mathbf{O} = \mathrm{LayerNorm}(\mathbf{H} + \mathrm{Concat}(\mathbf{A}_1, \ldots, \mathbf{A}_h)\mathbf{W}^O) \tag{3}$$

then $\mathbf{O}$ is flattened and passed through dense layers with ReLU activation and dropout regularization. The dense layer has u_d units, and the dropout rate is set to r. The final dense layer with softmax activation outputs the predicted probabilities for each class: $\hat{y} = \mathrm{softmax}(\mathbf{W}\mathbf{O}_\Theta + \mathbf{b})$, where $\mathbf{W}$ and $\mathbf{b}$ are the weights and biases of the dense layer.

3.3 Dual Visual and Textual Explanations

To gain insights into the decision-making process of TSCAN and enhance its interpretability, we integrate two XAI approaches (Attention-based and Gradient-based) and LVLMs to generate dual visual and textual explanations.

Visual Attention-based Explanation leverages the attention mechanism employed in the TSCAN. Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ be the input sequence of length T, where $\mathbf{x}_t \in \mathbb{R}^d$ is the feature vector at time step t. The attention mechanism computes a set of attention weights $\mathbf{A} = (\alpha_1, \alpha_2, \ldots, \alpha_T)$, where $\alpha_t \in [0, 1]$ represents the importance of time step t in the model's detection. The attention weights are obtained from the multi-head attention layer in the TSCAN. Given $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$ matrices, and $\mathcal{A}$, to interpret the attention weights, they are normalized to the range $[0, 1]$ using min-max scaling: $\hat{\alpha}_t = \frac{\alpha_t - \min(\mathbf{A})}{\max(\mathbf{A}) - \min(\mathbf{A})}$. The normalized attention weights $\hat{\mathbf{A}} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_T)$ are visualized as a heatmap overlay on the input sequence, highlighting the time steps the model attends to most when making decisions and providing insights into the temporal dependencies and patterns relevant for the classification task.

Visual Gradient-based Explanation The gradient-based explanation technique highlights the regions of the input sequence that have the highest impact on the model's output $\hat{y}$. Let $\hat{y}^c$ be the output of the model for class c, and $A^l \in \mathbb{R}^{T \times C}$ be the activations of layer l, where T is the sequence length and C is the number of channels. In our experiment, we choose the multi-head attention $\mathcal{A}$ as the target layer. The gradient-based heatmap for class c is computed as follows:

$$L^c = \mathrm{ReLU}\left(\sum_i \alpha_i^c A_i^l\right), \quad \alpha_i^c = \frac{1}{T} \sum_t \frac{\partial \hat{y}^c}{\partial A_{i,t}^l} \tag{4}$$

where α_i^c is the gradient of the output $\hat{y}^c$ with respect to the i-th channel of the activations A^l. The normalized heatmap L^c is visualized as an overlay on the input sequence $\mathbf{X}$.

Textual Explanations via LVLMs offer a human-readable interpretation of the model's detection of PDs, enhancing the reliability and trustworthiness of the model via UDI for clinicians. We employ a recent member of the LVLMs family, GPT-4o Vision, as the core vision language model. This LVLM processes a designed prompt (Fig. 2), the input RRI, and the dual visual explanation maps to generate textual explanations. The model's responses are based on its understanding of the visual content from explanation maps and RRI-related metrics (i.e. SDNN, RMSSD, and pNN50) to associate relevant textual descriptions. The generated explanations provide a concise and intuitive summary of PD detection, allowing clinicians to understand the model's behaviour and the rationale behind its decisions.

```
SYSTEM: You are a mental health professional analyzing
AI model explanations for detecting mental health
conditions (healthy or bipolar disorder/schizophrenia)
from ECG-derived RR intervals.

INPUT PARAMETERS:
{
    "classification": str,
    "files": {
        "attention_map": {path_to_attention_png},
        "gradient_map": {path_to_gradient_png},
        "rri_data": {path_to_RRI_csv}
    }
}

You will analyze the provided files where:
1. Attention-based saliency map shows feature
importance:
- Red = high importance, Blue = low importance
- Shows timestamped regions of interest
2. Gradient-based saliency map shows feature importance:
- Red = high importance, Blue = low importance
- Shows timestamped regions of interest
3. RRI data CSV contains values for calculating:
- SDNN (Standard Deviation of NN intervals)
- RMSSD (Root Mean Square of Successive Differences)
- pNN50 (Percentage of NN intervals > 50ms)

TASK: Provide a single paragraph analyzing:
- Key RRI features highlighted for mental health classification
- Agreement between the two explanation methods
- Model confidence and potential limitations
- Temporal patterns in the explanations
- Cause-effect relationship between identified features and
classification
- Whether the explanations support the model's diagnosis of
{classification}

Use clear, clinical language. Reference specific timesteps and
metrics that support your assessment. Focus on how the identified
patterns relate to known cardiovascular manifestations of mental
health conditions.
```

Fig. 2: Prompt template for generating textual explanations in PD detection.

3.4 User-centered Diagnosis Interface (UDI)

The UDI serves as the interface for clinicians to monitor, diagnose, and verify the model's decisions with dual visual and textual explanations. The clinicians can choose an RRI sequence from the database, then the signal is plotted, and the detection from TSCAN is delivered. After that, the clinicians can observe two visual attention-based and gradient-based explanations and read the textual explanation to further investigate the reliability of the model's decisions.

Our PD diagnosis system ensures a straightforward workflow: ECG collection from the users via the ERI, detection by an interpretable AI system, and investigation by clinicians via the UDI. By integrating wearable ECG monitors and providing intuitive data collection and analysis interfaces, our system aims to make advanced PD diagnosis more comprehensible and widely available in clinical settings.

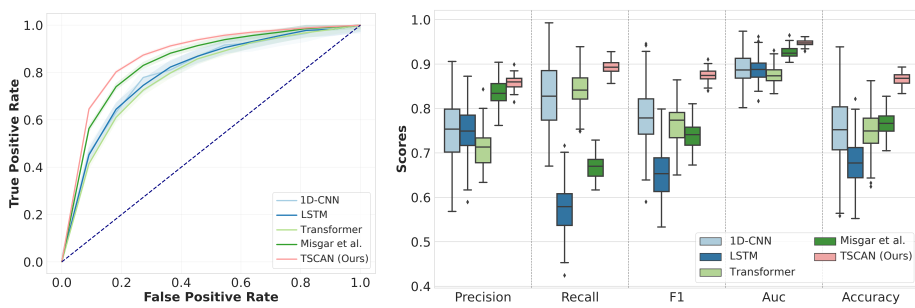
4 Experimental Setup and Dataset

Wearable ECG Monitor Our PD diagnosis system uses the Polar H10 as a wearable ECG monitor due to its high precision in single-lead ECG signal and RRI measurements [13]. The ECG signal is sampled at 130 Hz and recorded in microvolts (µV).

Dataset We utilized the HRV-ACC dataset [12], which contains physiological data from 60 participants. The dataset comprises participants: 30 diagnosed with SZ or BP (labelled as treatment/positive) and 30 controls (labelled as control/negative). This dataset is considered balanced for training the model. Each participant contributed 1.5–2 hours of ECG recordings using a wearable Polar H10 device. To prepare the data for our model, we create input sequences $\mathbf{X}_i = (\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_T})$ of length T for each i-th sequence. We employ a sliding window approach with a fixed sequence length of $T = 50$. This method generates overlapping sequences where $\mathbf{X}_i \cap \mathbf{X}_{i+1} = \mathbf{x}_{i_2}, \mathbf{x}_{i_3}, \ldots, \mathbf{x}_{i_T}$, meaning consecutive

Table 1: The performance results are presented with the highest values indicated in **bold**, and the second-highest values underlined. "-" metrics are not available in the original work.

Model	Precision	Recall	F1 Score	AUC	Accuracy
(a) 5-fold cross-validation					
1D-CNN	0.747 ± 0.070	0.826 ± 0.070	0.784 ± 0.070	0.895 ± 0.035	0.750 ± 0.070
LSTM	0.760 ± 0.056	0.583 ± 0.056	0.659 ± 0.056	0.889 ± 0.028	0.667 ± 0.056
Transformer	0.710 ± 0.042	0.833 ± 0.042	0.767 ± 0.042	0.875 ± 0.021	0.750 ± 0.042
Misgar et al. [15]	0.833 ± 0.028	0.667 ± 0.028	0.741 ± 0.028	0.928 ± 0.014	0.766 ± 0.028
TSCAN (Ours)	**0.858 ± 0.014**	**0.896 ± 0.014**	**0.876 ± 0.014**	**0.948 ± 0.007**	**0.866 ± 0.014**
(b) Leave-one-out cross-validation					
1D-CNN	0.781	0.833	0.806	0.826	0.8
LSTM	0.727	0.8	0.762	0.799	0.75
Transformer	0.92	0.767	0.836	0.85	0.85
Misgar et al. [15]	0.884	0.767	0.821	0.906	0.833
Buza et al. [4]	-	-	-	0.910	0.833
TSCAN (Ours)	**0.962**	**0.833**	**0.893**	**0.933**	**0.900**



(a) AUC-ROC curves

(b) Metrics

Fig. 3: Performance results with 5-fold cross-validation.

sequences share $T-1$ elements. For a dataset of length N, we generate $N-T+1$ sequences, where each RRI $\mathbf{x}_k$ (except for the first and last $T-1$ intervals) appears in T different sequences: $\mathbf{X}_{k-T+1}, \mathbf{X}_{k-T+2}, \ldots, \mathbf{X}_k$. This approach effectively captures temporal dependencies by allowing the model to observe each RRI in various contexts within the surrounding data. To ensure consistent scale across all inputs, we apply normalization to the sequences.

5 Experimental Results

5.1 Model Performance

We evaluated the TSCAN model against several baselines from state-of-the-art such as Misgar et al. [15] and Buza et al. [4] with 5-fold cross-validation and leave-one-out cross-validation. We also included components of the TSCAN model,

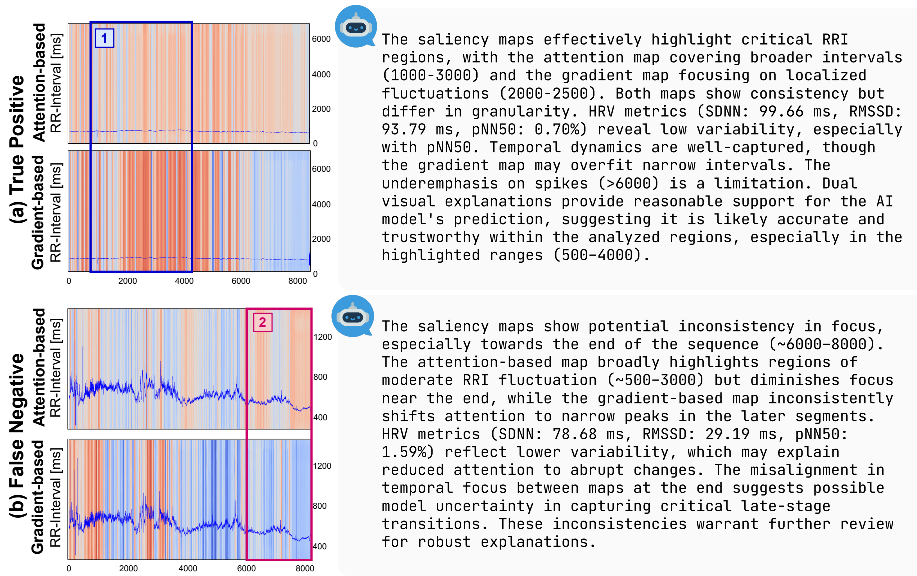


Fig. 4: The dual visual and textual explanations of model's detection of PDs.

such as 1D-Convolutional Neural Network (1D-CNN), LSTM, and Transformer, as part of an ablation study.

Setup To optimize the performance of all models, we applied the hyperparameter search conducted on the training set, with 20% of this set reserved as a validation set for hyperparameter selection. For TSCAN, the best hyperparameters are as follows: the number of filters $f_1 = 256$, $f_2 = 1024$, $f_3 = 256$, $f_4 = 1024$, kernel sizes $k = 7$, LSTM units $u_b = 128$, attention heads $h = 2$, attention key vectors dimension $d = 32$, dense units $u_d = 64$, and drop out rate $r = 0.2$.

5-fold cross-validation partitions the HRV-ACC dataset into 80% training (48 people) and 20% test (12 people) set for each fold. For a fair comparison, we adapted the output layer architecture of the Misgar et al. [15] to align with the specific requirements of our dataset. Table 1a summarizes the performance metrics obtained for all models under 5-fold cross-validation. Figure 3a illustrates the AUC-ROC curves, highlighting TSCAN's consistent performance across folds. TSCAN demonstrates superior performance across all metrics, achieving an average AUC of 0.948 and consistently high precision, recall, and F1 scores. While Misgar et al.'s model achieves competitive AUC, the Transformer exhibits strong recall but slightly lower precision, indicating a tendency to overpredict positive cases. The 1D-CNN balances precision and recall effectively, whereas LSTM shows limitations in recall, missing some positive cases.

Leave-one-out cross-validation further evaluates TSCAN and its components, where the model is trained on all participants except one and tested on the held-out individual. As shown in Table 1b, TSCAN outperforms all baseline models across metrics. The complementary strengths of TSCAN's components contribute to its robust performance: Transformer captures long-range dependencies, 1D-CNN enhances spatial feature extraction, and LSTM contributes effective temporal modelling. Misgar et al's model and Buza et al.'s baselines show competitive AUC and Accuracy.

Overall, the results highlight the effectiveness of TSCAN in accurately and robustly classifying instances. The strong performance of its components offers local and global information into the factors contributing to TSCAN's high performance across all metrics.

5.2 Visual and Textual Explanations

Dual visual and textual explanations provide insights into the decision-making process of TSCAN, as shown in Figure 4. Explanations are analyzed for both a correctly classified case (true positive) and a misclassified case (false negative):

1. **True positive** (Figure 4a): The visual explanations highlighted consistent regions of importance in the RRI sequence $\boxed{1}$, indicating reliable model attention. The LVLM-generated textual explanation further supports the decision, enhancing the model's trustworthiness.
2. **False negative** (Figure 4b): Disagreements between the attention-based and gradient-based visual explanations are observed in the latter part of the RRI sequence $\boxed{2}$. The LVLM textual explanation identified this inconsistency, suggesting potential areas where the model may misinterpret low variability as indicative of a healthy individual.

These explanations offer valuable insights into feature importance, agreement between visual mechanisms, and potential areas for error analysis; thus, improving model interpretability and potential clinical relevance.

6 Applicability to Another Use Case

In this section, we evaluated the adaptability of our proposed diagnosis system by applying it to a new use case: atrial fibrillation (AF) detection using the 2017 PhysioNet/Computing in Cardiology Challenge dataset [5]. This dataset contains single-lead ECG recordings sampled at 300 Hz, pre-processed through band-pass filtering with the AliveCor device. It has been extensively used to benchmark classification algorithms to distinguish various cardiac rhythms. For this experiment, we focused on two classes from the dataset, AF (labelled as positive) and normal sinus rhythm (labelled as negative), to retrain the TSCAN model. Fig. 5 presents the model's detection and corresponding dual visual and textual explanations generated by our diagnosis system on the correct classification (true positive) and misclassification (false negative) cases. LVLM demonstrates its effectiveness by providing additional interpretative highlights, such

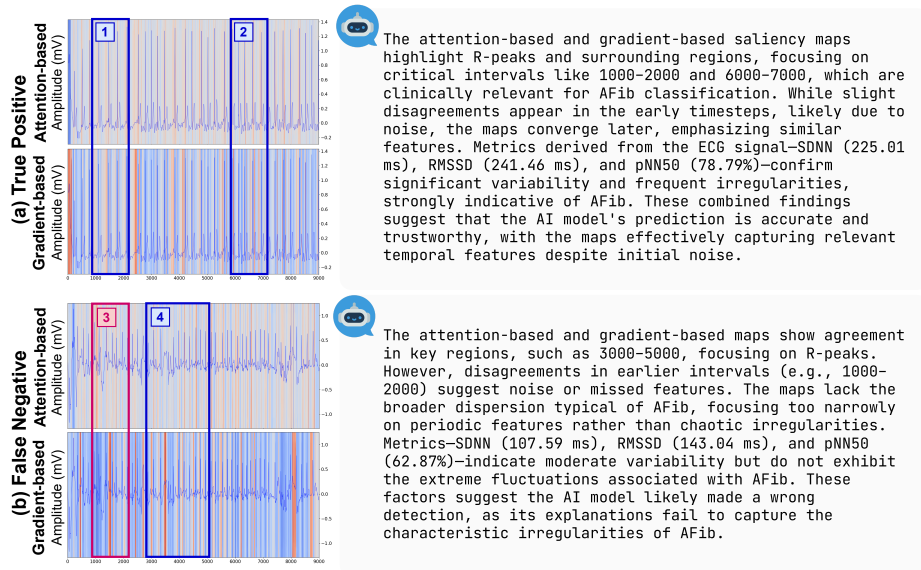


Fig. 5: The dual visual and textual explanations of model's detection of AF.

as $\boxed{1}$, $\boxed{2}$, $\boxed{3}$ and $\boxed{4}$. This emphasizes the potential of incorporating XAI and LVLM in healthcare applications such as AF detection to enhance clinicians' relevance in AI-enhanced clinical decision support systems.

7 Discussion

We found that both explanation methods consistently highlight areas with rapid changes in the RRI sequence, aligning with prior research showing lower HRV in people with PDs (e.g., SZ, BP) compared to healthy controls [2,11], although some research suggests that this is less observable in BP [11]. Hence, our future studies will include broader diagnostic labels to explore HRV differences among PDs and improve clinical insights. We will also detail the comparative predictive performance for different use cases (i.e. detection of AF and other conditions).

Our dual approach to model interpretation offers a valuable framework for enhancing clinical decision-making without substituting clinical judgment. We also underscore the importance of open-access and customizable HRV data collection in wearable devices. This highlights a potential area for development in consumer wearables, where we can balance the flexibility and security of HRV monitoring in both research and clinical applications.

Future research should integrate additional physiological signals (e.g., skin conductance, temperature, physical activity) and expand datasets to include diverse, representative populations and a broader spectrum of PDs. A multimodal approach combining various data streams (e.g., eye-tracking, electroencephalography/EEG) with XAI algorithms can offer a holistic understanding of physiological states, improving the accuracy of detecting PDs. We aim to implement this approach using our high-performance-computing infrastructure for large-scale, multi-dimensional data processing and analysis.

8 Conclusion

Our paper presents an interpretable system for PD diagnosis using wearable ECG devices and the TSCAN model. Combining visual and textual explanations enhances AI transparency and clinical usability. Experiments show superior performance, interpretability through XAI and LVLMs, and adaptability to related use cases. It underscores HRV as a key biomarker for PDs, offering a scalable approach to mental health diagnostics. Future work will integrate multimodal data, expand disorder coverage, and improve the user interface for better clinician trust and satisfaction.

Acknowledgment

This study is supported by the UNB-FCS Startup Fund (22–23 START UP/H CAO). It is also funded in part by the National Research Council of Canada's Aging in Place Challenge program.

References

1. Arias, J.T., Astudillo, C.A.: Enhancing schizophrenia prediction using class balancing and shap explainability techniques on eeg data. In: 2023 IEEE 13th International Conference on Pattern Recognition Systems. pp. 1–5. IEEE (2023)
2. Benjamin, B.R., et al.: Heart rate variability is associated with disease severity in psychosis spectrum disorders. Progress in Neuro-Psychopharmacology and Biological Psychiatry **111**, 110108 (2021)
3. Bertrand, A., et al.: On selective, mutable and dialogic xai: A review of what users say about different types of interactive explanations. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–21 (2023)
4. Buza, K., et al.: A simple and effective classifier for the detection of psychotic disorders based on heart rate variability time series. In: Workshop on Bioinformatics and Computational Biology WBCB 2023 (2023)
5. Clifford, G.D., et al.: Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017. In: 2017 Computing in Cardiology (CinC). pp. 1–4. IEEE (2017)
6. Corponi, F., et al.: Does heart rate variability change over acute episodes of bipolar disorder? a bayesian analysis. (2024)
7. Hao, H., et al.: Temporal convolutional attention-based network for sequence modeling. arXiv preprint arXiv:2002.12530 (2020)
8. Hill, M., et al.: Precision medicine for psychotic disorders: objective assessment, risk prediction, and pharmacogenomics. Molecular Psychiatry pp. 1–22 (2024)
9. Hinde, K., et al.: Wearable devices suitable for monitoring twenty four hour heart rate variability in military populations. Sensors **21**(4), 1061 (2021)

10. Jimenez-Mesa, C., et al.: Machine learning in small sample neuroimaging studies: Novel measures for schizophrenia analysis. Human Brain Mapping **45**(5), e26555 (2024)
11. Książek, K.M., et al.: The analysis of heart rate variability and accelerometer mobility data in the assessment of symptom severity in psychotic disorder patients using a wearable polar h10 sensor. medRxiv pp. 2023–08 (2023)
12. Ksikażek, K., et al.: Hrv-acc: a dataset with rr intervals and accelerometer data for the diagnosis of psychotic disorders using a polar h10 wearable sensor, 2023b https://doi.org/10.5281/zenodo
13. Li, X., thers: Evaluation of measurement accuracy of wearable devices for heart rate variability. Iscience **26**(11) (2023)
14. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. Advances in neural information processing systems **30** (2017)
15. Misgar, M.M., Bhatia, M.: Unveiling psychotic disorder patterns: A deep learning model analysing motor activity time-series data with explainable ai. Biomedical Signal Processing and Control **91**, 106000 (2024)
16. Nguyen, H., et al.: Langxai: Integrating large vision models for generating textual explanations to enhance explainability in visual perception tasks. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24. pp. 8754–8758 (8 2024)
17. Nguyen, H.T.T., et al.: Xedgeai: A human-centered industrial inspection framework with data-centric explainable edge ai approach. Information Fusion p. 102782 (2024)
18. Nguyen, T.T.H., et al.: Towards trust of explainable ai in thyroid nodule diagnosis. In: International Workshop on Health Intelligence. pp. 11–26. Springer (2023)
19. Pedersen, C.B., et al.: A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. JAMA psychiatry **71**(5), 573–581 (2014)
20. Selvaraju, R.R., et al.: Grad-cam: visual explanations from deep networks via gradient-based localization. International journal of computer vision **128**, 336–359 (2020)
21. Simonsen, U., et al.: The effect of tempol on endothelium-dependent vasodilatation and blood pressure. Pharmacology & therapeutics **122**(2), 109–124 (2009)
22. Wang, D., et al.: Designing theory-driven user-centric explainable ai. In: Proceedings of the 2019 CHI conference on human factors in computing systems. pp. 1–15 (2019)