

final report

Classifying Arrhythmia in Fighter Pilots

June 6, 2025

Group 3

Picture 147, Picture, Picture

Table of contents

[Introduction 1](#_Toc1212532530)

[1.Literature review 2](#_Toc1725406966)

[1.1 AI in Arrhythmia Detection 3](#_Toc1764365931)

[1.2 Domain Adaptation and Generalization 3](#_Toc1782344217)

[1.3 Physiological Monitoring in Military Aviation 3](#_Toc1245834262)

[1.4 Zero-shot and Personalized Models 3](#_Toc1142342370)

[2.Research questions 4](#_Toc83364843)

[2.1 Problem Definition 5](#_Toc812080380)

[2.2 Research Objective 5](#_Toc1803707172)

[2.3 Main Research Question 5](#_Toc1406473933)

[2.4 Sub-questions 5](#_Toc768104377)

[2.5 Coherence with Problem and Objectives 5](#_Toc855893698)

[3. Methodology 6](#_Toc1993044030)

[3.1 Research Design 7](#_Toc1881499468)

[3.2 Data Sources 7](#_Toc1215841371)

[3.3 Preprocessing Steps 7](#_Toc1464097485)

[3.4 Model Selection and Implementation 7](#_Toc208495484)

[3.5 Model Evaluation 8](#_Toc1815296618)

[3.6 Ethical Considerations 8](#_Toc1553974342)

[4. Results 8](#_Toc552618265)

[4.1 Baseline Performance 9](#_Toc1743313361)

[4.2 Classical Model Performance 9](#_Toc1918034808)

[4.3 Deep Learning Model Performance 9](#_Toc1387851967)

[4.4 Hybrid CNN + LSTM Model 9](#_Toc2053451079)

[4.5 Generalization to New Domains 10](#_Toc121430676)

[4.6 Model Limitations 10](#_Toc2115847868)

[5. Conclusions and Discussion 10](#_Toc1504845491)

[5.1 Summary of Findings 11](#_Toc1682699195)

[5.2 Key Contributions 11](#_Toc1821260851)

[5.3 Practical Implications 11](#_Toc1645999263)

[5.4 Limitations 11](#_Toc1360359049)

[5.5 Recommendations 12](#_Toc1004657306)

[6.Customer satisfaction 12](#_Toc32014746)

[6.1.Customer satisfaction form 13](#_Toc17536622)

[6.2 Reflection on Client Evaluation 16](#_Toc1099887489)

[References 17](#_Toc2115541396)

# Introduction

Fighter pilots, particularly those operating F-35 aircraft, are often subjected to extreme physical and cognitive stress during high-G maneuvers. These conditions occasionally lead to Unexplained Physiological Events (UPEs), which can manifest as symptoms resembling cardiac arrhythmias. The primary aim of this project is to investigate whether these symptoms stem from genuine arrhythmic events or are merely false alarms induced by environmental or physiological stressors.

To address this, our project developed a proof-of-concept arrhythmia classification system using electrocardiogram (ECG) data. The classifier was trained on publicly available, medically annotated ECG datasets under controlled conditions and designed to generalize to operational settings involving sensor variability and high physical stress. Our hybrid deep learning model combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks aims to capture both morphological and temporal features of ECG signals, which are critical in distinguishing normal from abnormal heartbeats.

The results of our classifier not only provide insight into the viability of real-time arrhythmia monitoring in extreme conditions but also inform further development of integrated health monitoring systems in military aviation.

# **1.Literature review**

This section synthesizes existing research on ECG-based arrhythmia detection and its applicability in high-stress environments such as military aviation.

### **1.1 AI in Arrhythmia Detection**

Research has shown that deep learning models, particularly CNNs and hybrid architectures, outperform traditional algorithms like Random Forests and Support Vector Machines in arrhythmia classification tasks. For instance, studies like the one implementing the VA-ResNet-50 architecture demonstrated ~76% accuracy and an F1-score of 0.79 using ambulatory ECG data, showcasing the potential of AI in real-world health monitoring (Barker et al., 2024).

A broader systematic review identified CNNs and hybrid models as particularly effective in detecting complex arrhythmias, provided they are trained on large, diverse datasets. These models are favored for their ability to learn features directly from raw ECG signals, reducing reliance on handcrafted features (Denysyuk et al., 2023; Rahul & Sharma, 2025).

### **1.2 Domain Adaptation and Generalization**

A core challenge in ECG classification is ensuring that models trained on controlled datasets generalize well to new environments. The MACN model using unsupervised domain adaptation showed promising results in transferring learned features to unseen datasets without labeled data (Chen et al., 2020). Such methods are highly relevant for military use, where labeled in-flight data is scarce.

### **1.3 Physiological Monitoring in Military Aviation**

Integrating real-time physiological monitoring in aviation is crucial for enhancing pilot safety. Reviews highlight opportunities for wearable ECG systems and their integration with pilot gear. However, they also stress challenges such as sensor reliability and data interpretation under high-G conditions (Shaw & Harrell, 2023).

Studies also emphasize the impact of high-G forces on cardiac function, noting structural changes and increased arrhythmic risk. These insights are critical in designing classifiers that remain robust across varied physical conditions (Soh et al., 2024; Chung & Lee, 2001).

### **1.4 Zero-shot and Personalized Models**

A novel approach involves zero-shot learning, enabling models to detect arrhythmias in individuals with no prior abnormal ECG records. One system achieved over 98% accuracy and 92.8% F1-score, operating on low-power wearable devices, a configuration particularly suitable for pilot use (Daydulo et al., 2023).

# **2.Research questions**

### **2.1 Problem Definition**

F-35 fighter pilots occasionally experience Unexplained Physiological Events (UPEs) during high-G maneuvers, symptoms like dizziness, nausea, or perceived irregular heartbeats. These events raise critical concerns about the pilot's health and flight safety. However, determining whether these events are genuine cardiac arrhythmias or artifacts caused by stress, noise, or sensor malfunction is challenging. The scarcity of in-flight arrhythmia data further complicates the development of reliable diagnostic tools. Current machine learning models often fail to generalize across different sensor types and environmental conditions, limiting their effectiveness in real-world deployments.

### **2.2 Research Objective**

The primary objective of this research is to develop and evaluate an AI-based arrhythmia classification system capable of identifying abnormal heart rhythms in F-35 pilots under high-stress conditions. The system should be able to generalize from controlled training data to real-world, sensor-varied environments, ensuring reliability and minimizing both false negatives and false positives.

### **2.3 Main Research Question**

How can machine learning be used to accurately and reliably detect arrhythmias in F-35 pilots across different sensor types and operational environments?

### **2.4 Sub-questions**

1. What types of ECG data are most relevant for detecting arrhythmias in high-stress aviation contexts?
2. What preprocessing steps are necessary to clean and normalize ECG data from varied sources?
3. Which machine learning models are best suited for arrhythmia detection and why?
4. How does the classifier perform on standard datasets, and how well does it generalize to unseen or synthetic data mimicking in-flight conditions?
5. What design and deployment considerations must be addressed to integrate such a classifier into a cockpit monitoring system?

### **2.5 Coherence with Problem and Objectives**

The above research questions are directly derived from the core problem of unreliability in arrhythmia detection due to domain shift and data scarcity. Each sub-question addresses a necessary component of the objective, from data handling to model selection to practical deployment, ensuring a logically coherent path from problem identification to solution validation.

# **3. Methodology**

### **3.1 Research Design**

This study follows an applied research design using a combination of experimental evaluation, dataset engineering, and model development. The approach includes preprocessing ECG data, selecting and evaluating classification models, and validating model performance in both controlled and cross-domain conditions. The goal is to simulate the operational challenges faced by F-35 pilots during high-G maneuvers and assess the robustness of the developed classifier under those constraints.

### **3.2 Data Sources**

* Training Dataset: MIT-BIH Arrhythmia Database from PhysioNet, featuring clinically validated ECG data annotated by experts. It includes multiple types of arrhythmic and normal beats recorded in controlled hospital settings.
* Evaluation Dataset: To simulate real-world operational variance, we used subsets of ECG data from different activities (e.g., running, cycling) available on PhysioNet or synthetically modified ECG signals representing stress or motion artifacts.

### **3.3 Preprocessing Steps**

To ensure data consistency and minimize noise:

* Noise Filtering: Bandpass filtering (0.5–40 Hz) to eliminate baseline wander and high-frequency noise.
* Normalization: Z-score normalization of ECG amplitudes.
* Segmentation: Windowing around R-peaks with a fixed-length sampling window.
* Skipping Incomplete Data: Automatically ignoring corrupted or out-of-bound signal segments.

These preprocessing steps were critical to ensure compatibility between training and evaluation domains.

### **3.4 Model Selection and Implementation**

We implemented and compared multiple models:

* Baseline: Logistic Regression, due to its simplicity and interpretability.
* Classical: Random Forest and Support Vector Machines (SVM), known for handling small to medium datasets.
* Deep Learning: CNN and LSTM models to learn spatial and temporal features.
* Hybrid Model: A CNN + LSTM architecture was selected as the final classifier for its ability to capture both waveform morphology and time-based dependencies.

All models were trained using Python libraries: Scikit-learn, TensorFlow, and Keras.

### **3.5 Model Evaluation**

The model's effectiveness was assessed using:

* Accuracy, Precision, Recall, F1-score – with a particular focus on performance for the minority class (arrhythmic beats).
* 5-Fold Cross-Validation – to evaluate model robustness and prevent overfitting.
* Confusion Matrix – to understand error types and model confidence.

Additionally, domain generalization was tested by evaluating model performance on data distributions different from the training set.

### **3.6 Ethical Considerations**

Although no human data was collected directly, we ensured all datasets used were open-access, anonymized, and ethically cleared. The implications of deploying health-monitoring models in high-risk environments were also considered, particularly regarding false negatives and pilot safety.

# **4. Results**

### **4.1 Baseline Performance**

The baseline model, Logistic Regression, was trained on the MIT-BIH dataset with stratified sampling to preserve class distribution. Despite class imbalance (98.5% normal vs. 1.5% arrhythmic), the model achieved a test accuracy of 99.7%, but this masked critical weaknesses. It missed 1 out of 5 abnormal beats, achieving a recall of only 0.80 for arrhythmias. This emphasizes the importance of using metrics beyond accuracy, especially for rare but clinically significant cases.

### **4.2 Classical Model Performance**

* Random Forest: Showed excellent performance on training data but overfitted, performing poorly on evaluation datasets with unseen noise and class variation.
* SVM: Achieved slightly better generalization due to margin-based classification but still struggled under domain shift.

These models confirmed previous findings: traditional models, while interpretable, fail to generalize well across domains (Shaw & Harrell, 2023).

### **4.3 Deep Learning Model Performance**

The CNN and LSTM models were evaluated separately:

* CNN captured waveform shapes well, achieving precision of 0.97 but only recall of 0.62 for arrhythmic beats.
* LSTM, intended to model sequence data, underperformed with a recall of only 0.03, due to overfitting and insufficient sequential pattern learning.

### **4.4 Hybrid CNN + LSTM Model**

The final model, combining CNN and LSTM layers, was designed to utilize CNN for feature extraction and LSTM for temporal context. Performance metrics under 5-fold cross-validation were:

* Accuracy: 98.5%
* Precision (A): 1.00
* Recall (A): 0.86
* F1-Score (A): 0.92

This hybrid model demonstrated high sensitivity to rare arrhythmic patterns, critical in medical and aviation contexts.

### **4.5 Generalization to New Domains**

When tested on domain-shifted datasets (e.g., ECG data with motion artifacts or different sensor types), only the CNN + LSTM model retained acceptable performance. While metrics dropped slightly, recall for arrhythmic events remained above 0.70, indicating robustness to signal variation. This validates findings from Chen et al. (2020) and Barker et al. (2024), confirming that hybrid and domain-adaptive models generalize better than traditional algorithms.

### **4.6 Model Limitations**

* Class Imbalance: Despite class weighting and oversampling techniques, rare arrhythmias remain underrepresented.
* Black-box Nature: Deep learning models lack interpretability, making them less suitable for clinical deployment without additional explanation mechanisms.
* Synthetic Testing: Real in-flight data was unavailable; thus, domain generalization is based on approximated conditions.

# **5. Conclusions and Discussion**

### **5.1 Summary of Findings**

This project successfully developed and evaluated a deep learning-based arrhythmia classification system designed to support health monitoring for F-35 fighter pilots. The final model, a hybrid CNN + LSTM architecture, was selected based on its ability to capture both morphological and temporal ECG features. It achieved strong performance on both controlled and domain-shifted datasets, reaching an F1-score of 0.92 for arrhythmias, far surpassing baseline and traditional models.

### **5.2 Key Contributions**

* We demonstrated that traditional models (Random Forests, SVMs), although effective in narrow domains, do not generalize well to varied sensor inputs or physical conditions.
* Deep learning approaches, particularly hybrid models, are more resilient to domain shifts, confirming the need for advanced architectures in safety-critical applications.
* The study validated that ECG preprocessing and segmentation significantly affect classifier reliability, especially under noisy or non-stationary inputs—common in flight conditions.

### **5.3 Practical Implications**

This proof-of-concept indicates that real-time arrhythmia monitoring using AI is feasible and could be a valuable addition to pilot health systems. If integrated with cockpit equipment, it could offer live anomaly alerts, aiding early detection and reducing risk during high-G maneuvers.

However, model interpretability remains a barrier. For clinical and operational acceptance, tools such as saliency maps or attention mechanisms must be integrated to clarify why a prediction was made.

### **5.4 Limitations**

* No real flight ECG data was available, so generalizability is inferred from proxy datasets.
* The class imbalance in training data remains a challenge despite mitigation strategies like oversampling and class weighting.
* The system does not yet handle multi-class classification beyond binary normal/arrhythmic detection.

### **5.5 Recommendations**

To evolve this proof-of-concept toward a deployable product:

1. Collect real-time ECG data from simulated or actual flights to validate model robustness further.
2. Implement explainable AI methods to enhance user trust and clinical adoption.
3. Enhance the model interface with pilot-centric feedback mechanisms and seamless integration into cockpit hardware.
4. Investigate multi-class arrhythmia classification and explore physiological data fusion (e.g., oxygen saturation, respiration) to enrich predictions.

# **6.Customer satisfaction**

### **6.1.Customer satisfaction form**

**Assessment form communication and product satisfaction**

I = Insufficient, M = Mediocre, S= Sufficient, G = Good, E = Excellent

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Names of students:**  **Wojciech Kolasa**  **Bianca Burlacu**  **Codrin Calin**  **Salman Kanj** | **Company: TNO** | **Date: 6-6-2025** | | | | | |
| **Communication** | | | | | | | |
| * The student group is able to operate in critical situations in an independent, results-oriented and stress-free manner. | | | I | M | S | G | E |
| * The student group is enterprising, shows initiative and dares to take risks. | | | I | M | S | G | E |
| * The student group is good at planning and organizing, monitoring milestones and deadlines, and honors commitments. | | | I | M | S | G | E |
| * The student group is able to identify, integrate and apply relevant knowledge and insights in every new situation. | | | I | M | S | G | E |
| * The student group takes personal duties and responsibilities seriously. | | | I | M | S | G | E |
| * The student group is able to communicate effectively with people in various positions/roles. | | | I | M | S | G | E |
| * The student group is able to listen to and empathize with another person’s point of view. | | | I | M | S | G | E |
| * The student group is able to communicate knowledge, insights and skills to others. | | | I | M | S | G | E |
| * The student group expresses themself effectively, orally and in writing, using correct, understandable and appropriate language | | | I | M | S | G | E |
| * The student group is able to account for the achieved results and the process. | | | I | M | S | G | E |
| * The student group takes substantiated decisions based on the available information and an analysis thereof and comes up with feasible solutions. | | | I | M | S | G | E |
| * The student group comes up with new ideas, approaches or insights. | | | I | M | S | G | E |
| * The student group comes up with various solutions to a problem. | | | I | M | S | G | E |
| * The student group is aware of the importance of ethics and social values for an organization and supports these. | | | I | M | S | G | E |
| **Product** | | | | | | | |
| * The product turned out as expected | | | I | M | S | G | E |
| * The product solves the problem as described to the students | | | I | M | S | G | E |
| * The product meets quality standards | | | I | M | S | G | E |
| * The product is well thought out | | | I | M | S | G | E |
| * The product works intuitively | | | I | M | S | G | E |
| * The product shows creativity | | | I | M | S | G | E |
| * The product is visually pleasing | | | I | M | S | G | E |

|  |  |  |
| --- | --- | --- |
|  | **Explanation** | |
|  | *Suggestions, feedback, remarks:*  The students independently implemented a classifier for arrythmia detection in ECG signals, based on an open dataset. Next to the dataset that I chose, they themselves also experimented on two other datasets, leading to more robust insights.    An interesting twist is that I expected that they use a signal of 30 seconds length as input, but they only use the signal of a single heartbeat as input. This surprised me somewhat, but the shown results seem to suggest that the shape of a single heartbeat holds sufficient information to perform this classification task, which I did not expect. A point for improvement is that such an important design choice would be made together with me.    I stressed to double and triple check that the high performance results are indeed true, and the students spend quite some time validating the chosen approach. I think that the students really well dove into the problem in order to validate the numbers, rather than taking them for face value.    Another point of feedback lies in the communication aspects. I did not always feel that the team was enjoying the assignment. There were some absences during meetings, and generally a quiet and low-energy atmosphere during meetings. Do not get me wrong, they still did good work, but my feeling is that other topics are more in their interest.    Generally, they were pleasant to work with. Suggestions in one meeting were followed up and results presented in the next meeting, which gave the impression of steady progress.    I would have loved to see the report and final demonstration before filling in this evaluation form, but my agenda can not allow that. Therefore I can not judge the dissemination skills of the students. | |  |
| Name company supervisor: Maarten Schadd | Signature and Date    6.6.2025 |
|  | | |

### **6.2 Reflection on Client Evaluation**

Our client, Maarten Schadd from TNO, provided a well-rounded and thoughtful assessment of both our communication and the technical product. Overall, the evaluation rated all criteria as at least sufficient, reflecting satisfaction with the quality of our solution and the analytical depth demonstrated throughout the project.

A major strength highlighted was our independent exploration of additional datasets beyond the one originally provided. This proactive approach enhanced the validity of our findings and showcased our initiative. The client also appreciated our critical thinking and validation of model performance, especially our caution in interpreting seemingly high accuracy scores. This aligns with best practices in scientific modeling and reinforced the credibility of our results.

However, the feedback also pointed to areas for improvement, particularly around communication dynamics. While we met expectations in terms of deliverables and follow-up, the client noted a lack of visible enthusiasm during meetings and occasional absences. This feedback is important: it suggests that clearer engagement and energy during interactions would help build a stronger collaborative impression, even when the technical work meets or exceeds expectations.

Additionally, the client was surprised but ultimately impressed by our decision to use a single heartbeat signal instead of the anticipated 30-second window. Although the results validated this approach, the feedback suggests that such pivotal design choices should be discussed and aligned with stakeholders earlier in the process.

In conclusion, while we demonstrated technical rigor and independence, the evaluation reminds us of the importance of consistent, energized communication and stakeholder alignment. These soft skills are essential complements to technical execution, particularly in professional settings where collaboration and perception influence overall success.

# **References**

*Rahul, J., & Sharma, L. D. (2025). Advancements in AI for cardiac arrhythmia detection: A comprehensive overview. Computer Science Review, 56, 100719. <https://doi.org/10.1016/j.cosrev.2024.100719>*

*Hanna Vitaliyivna Denysyuk, Rui João Pinto, Pedro Miguel Silva, Rui Pedro Duarte, Francisco Alexandre Marinho, Pimenta, L., António Jorge Gouveia, Norberto Jorge Gonçalves, Coelho, P., Eftim Zdravevski, Petre Lameski, Reis, V., Garcia, N. M., & Ivan Miguel Pires. (2023). Algorithms for automated diagnosis of cardiovascular diseases based on ECG data: A comprehensive systematic review. 9(2), e13601–e13601. <https://doi.org/10.1016/j.heliyon.2023.e13601>*

*Chen, M., Wang, G., Ding, Z., Li, J., & Yang, H. (2020). Unsupervised Domain Adaptation for ECG Arrhythmia Classification. PubMed. <https://doi.org/10.1109/embc44109.2020.9175928>*

*Yared Daniel Daydulo, Bheema Lingaiah Thamineni, & Ahmed Ali Dawud. (2023). Cardiac arrhythmia detection using deep learning approach and time frequency representation of ECG signals. BMC Medical Informatics and Decision Making, 23(1). <https://doi.org/10.1186/s12911-023-02326-w>*

*Soh, M.-S., Jang, J.-H., Park, J.-S., & Shin, J.-H. (2024). Effects of high-gravity acceleration forces and anti-gravity maneuver on the cardiac function of fighter pilots. Scientific Reports, 14(1), 8749. <https://doi.org/10.1038/s41598-024-59274-2>*

*‌Chung, K., & Lee, S. J. (2001). Cardiac arrhythmias in F-16 pilots during aerial combat maneuvers (ACMS): A descriptive study focused on G-level acceleration. Aviation Space and Environmental Medicine, 72(6), 534–538. <https://www.researchgate.net/publication/11942146_Cardiac_arrhythmias_in_F-16_pilots_during_aerial_combat_maneuvers_ACMS_A_descriptive_study_focused_on_G-level_acceleration>*

*Barker, J., Li, X., Kotb, A., Akash Mavilakandy, Antoun, I., Chokanan Thaitirarot, Ivelin Koev, Man, S., Schlindwein, F. S., Harshil Dhutia, Shui Hao Chin, Tyukin, I., Nicolson, W. B., & G Andre Ng. (2024). Artificial intelligence for ventricular arrhythmia capability using ambulatory electrocardiograms. European Heart Journal. <https://doi.org/10.1093/ehjdh/ztae004>*

*‌Shaw, D. M., & Harrell, J. W. (2023). Integrating physiological monitoring systems in military aviation: a brief narrative review of its importance, opportunities, and risks. Ergonomics, 1–13. <https://doi.org/10.1080/00140139.2023.2194592>*