Comparative Analysis of Sequential and Vector-Based Machine Learning Approaches for ECG Pattern Recognition

Fatema Fidvi: 230024865
Fatema.Fidvi@city.ac.uk

Abstract: This report explores the application of Long Short-Term Memory (LSTM) networks and Support Vector Machines (SVM) to analyze electrocardiogram (ECG) signals. It assesses the effectiveness of these models in addressing the complexities of ECG data for cardiac abnormality diagnosis. LSTMs are tested for their sequential data processing capabilities, while SVMs are evaluated for handling high-dimensional spaces. The study highlights the importance of choosing suitable modeling techniques for medical datasets characterized by class imbalances and complex patterns. The findings aim to enhance diagnostic tools, contributing to more precise and reliable patient care, and paving the way for future advancements in machine learning applications in healthcare diagnostics.

I. Introduction

Electrocardiograms (ECGs) are an essential diagnostic tool that cardiologists rely on to assess the heart's rhythmic patterns. A standard ECG waveform, characterized by its PQRST complex, encapsulates the rhythm of a healthy heart. Amidst the rise of cardiovascular diseases (CVDs) as the foremost cause of global mortality, claiming approximately 17.3 million lives annually [1], the accurate classification of arrhythmias, reflected in the subtle deviations from the typical ECG pattern, is of critical significance. Each segment of the PQRST pattern (fig 1) corresponds to distinct electrical phases of a heartbeat, from atrial contraction to ventricular repolarization. Disparities within this rhythm can indicate different types of arrhythmias, with potential implications ranging from the benign to potentially fatal.

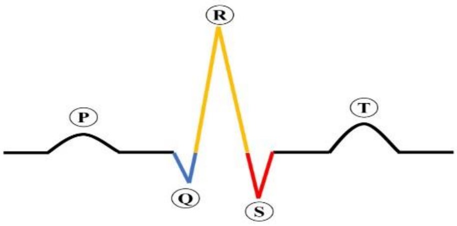


Fig 1. Construction of an ECG signal [2]

This paper aims to present a comparative analysis of two machine learning algorithms—Support Vector Machines (SVM), a vector-based approach, and Long Short-Term Memory (LSTM) networks, a method attuned to sequential data - in their task to classify five distinct types of ECG signals. These models will be compared with a baseline Multilayer Perceptron (MLP) with a single hidden layer, previously established in group coursework.

II. Brief Description of LSTM and SVM

a) LSTM-RNN (Long Short Term Memory Recurrent Neural Networks)

Long Short-Term Memory (LSTM) networks are an advanced type of Recurrent Neural Network (RNN) designed to address and overcome the limitations of traditional RNNs in learning long-term dependencies [3]. The key difference of LSTMs lies in their cell structure, which is equipped with three distinct gates: the forget gate, input gate, and output gate as shown in fig 2. Unlike a standard RNN cell, which processes inputs through a single gate, leading to challenges like the vanishing or exploding gradient problem, an LSTM cell's architecture allows for more sophisticated data handling [4].

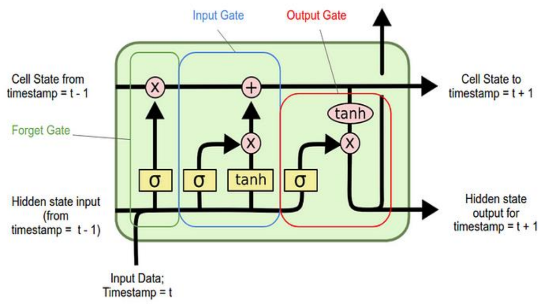


Fig 2. Single LSTM cell [3]

Pros:

- LSTMs are inherently designed to process sequential data, making them ideal choice for capturing temporal dynamics of ECG signals.
- LSTMs can learn dependencies across long sequences, crucial for ECG analysis where arrhythmias might be characterized by subtle patterns distributed over time.

Cons:

- LSTMs generally require more computational power and time to train compared to SVMs.
- The architecture of LSTMs introduces numerous hyperparameters (like the number of LSTM units, layers, hidden neurons, etc), making them harder to tune for optimal performance.
- They are prone to overfitting.
- LSTMs excel with large amounts of data, and if the data is highly imbalanced, as the ECG dataset when not balanced, the chances of LSTM performing well and generalizing decreases.

b) SVM (Support Vector Machine)

Known for their robustness, good generalization ability, and unique global optimum solutions, SVMs are probably the most popular machine learning approach for supervised learning, primarily known for handling classification tasks, although is also capable of performing regression tasks [5]. The main aim of this algorithm is to construct a hyper-plane, or a set of hyper-planes that best separates two classes based on statistical approaches. SVM chooses extreme points/vectors called support vectors that help in creating the hyperplanes [6,7].

Pros:

- SVMs excel in high dimension spaces, making them suitable for ECG datasets that, after feature extraction, become highly informative yet manageable in size.
- They're robust against overfitting, crucial for accurate arrhythmia classification.
- The “Kernel Trick” allows SVM to handle non-linear relationships within the ECG data by mapping inputs into higher-dimensional spaces, enabling the identification of a linear decision boundary in this new space [8].
- Once trained, SVMs require only the calculation of dot products between the support vectors and the new data points for classification.

Cons:

- Training and optimizing SVMs on very large dataset such as this can be computationally intensive and time-consuming if not accompanied with feature extraction and PCA.
- The performance of SVM is heavily dependent on the choice of kernel and regularization parameters. Finding the optimal set of parameters like C, gamma for RBF kernel requires intensive search and cross-validation which can be meticulous and resource-intensive process.

Hypothesis Statement: The hypothesis is that Long Short-Term Memory (LSTM) networks are likely to significantly outperform Support Vector Machines (SVMs) in the classification of Electrocardiogram (ECG) signals, due to LSTMs' enhanced ability to process sequential data and capture temporal dependencies.

III. About Dataset and Initial Data Analysis

This paper uses PhysioNet MIT-BIH Arrythmia ECG dataset consisting of ECG recordings from 47 different subjects recorded at the sampling rate of 360Hz [9]. With 109,446 records and 187 numerical features, pre-divided into training and testing sets, it provides a rich basis for developing advanced predictive models aimed at classifying 5 distinct types of ECG signals. Fig 3 showcases the 5 distinct classes in the dataset: Normal (0), Supraventricular ectopic beats (1), Ventricular ectopic beats (2), Fusion beats (3) and unclassified beats (4). Although complete and standardized across all 187 features, it presented a significant challenge of class imbalance, as highlighted in Fig 4. This was addressed by blending under-sampling and SMOTE, achieving class balance and preserving data integrity, where under-sampling lowers computational load by reducing the number of majority class samples and SMOTE carefully increases the minority class representation by producing augmented data.

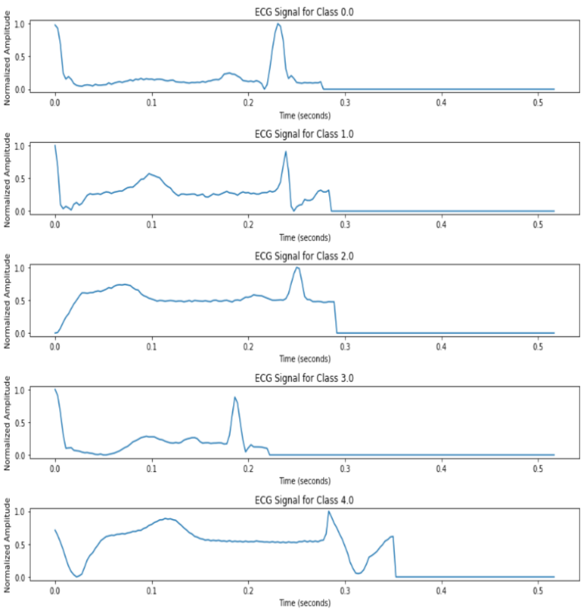


Fig 3. 5 classes of ECG Signals in the dataset

IV. Training and Evaluation Methodology

In developing the best models for this study, a strategic approach was adopted involving splitting the training set further into training and validation segments using the 80:20 split. The methodology for each model was designed to leverage their respective strengths effectively.

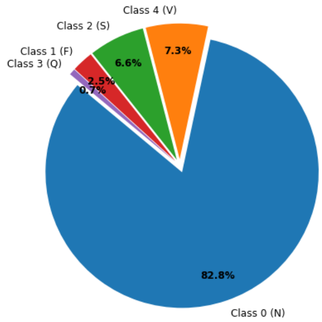


Fig 4. Class imbalance

a. LSTM Model:

Architecture & Training approach: The LSTM model's development began with the conversion of data into tensors, a format compatible with the PyTorch framework, essential for processing sequential data. The model's architecture featured a straightforward configuration of hidden units, layers, and dropout rates. A distinctive feature of this setup was the implementation of learnable initial states, instead of using traditional zero initialization method. This approach aimed to enhance the model's starting conditions, improving its efficiency in minimizing loss and capturing the complex temporal patterns of the ECG data. To refine the model's performance, an iterative tuning process was selected over traditional optimization methods like GridSearch and Optuna, which were found to be impractical due to the dataset's size and computational limits. The model underwent adjustments in architecture and hyperparameters during each iteration, involving layer count, hidden neurons, learning rate and dropout rate. The iterative training approach was chosen for its ability to facilitate personalized, step-by-step refinements, enhancing model efficiency and precision. This method forms the basis of the detailed improvements detailed in subsequent sections.

Parameters & Evaluation Strategy: In the LSTM model development for ECG signal classification, key parameters were carefully selected to optimize performance. The Adam optimizer was chosen for its flexible and efficient weight updates, crucial for managing time-series data. The use of cross-entropy loss, incorporating softmax functionality, streamlined the model for better efficiency and accuracy for classification task. Critical parameters such as hidden neuron count, hidden layers, dropout rate, and learning rate were iteratively adjusted to grasp the complexity of the data without risking overfitting. The model underwent iterative refinements, guided by real-time feedback from the validation set, which ensured that each parameter adjustment enhanced the accuracy and generalizability of the outcomes. The optimal LSTM iteration was determined through a dual-criterion approach focusing on accuracy and loss on the validation set. This strategy confirmed that the selected model configuration not only achieved high classification accuracy but also minimized errors, demonstrating its strong generalization capabilities.

b. SVM Model:

Architecture & Training Approach: The SVM model underwent initial enhancement through Discrete Wavelet Transform (DWT) [10], which expanded the feature set from 188 to 207 dimensions, thereby capturing a broader array of arrhythmia patterns essential for accurate classification. This was followed by scaling these additional features for maintaining performance consistency across different scales. To further enhance computational efficiency and manage the expanded feature set, Principal Component Analysis (PCA) was applied to reduce dimensionality, which is particularly effective in vector-based models like SVM. This contrasts with LSTM's handling of sequential data. A linear kernel was initially chosen for baseline training, setting the stage for subsequent hyperparameter optimization.

Parameters & Evaluation Strategy: The SVM model underwent a strategic hyperparameter tuning phase to refine its accuracy and generalization capabilities This optimization was carried out through a grid search on a 20% subset of the training data, chosen via stratified sampling to maintain class distribution and using accuracy as the evaluation metric. This method efficiently leveraged computational resources while yielding detailed insights into the model's performance under different configurations. The parameter grid included variations in the regularization constant (C), gamma values, and kernel types (linear and RBF), with linear kernels favored for their simplicity and RBF kernels for their capacity to handle non-linear patterns found in ECG signals. Following the grid search, the best hyperparameters were applied to retrain the SVM on the full training dataset, with settings adjusted to enable probability estimates essential for ROC curve analysis later. The model's effectiveness was then assessed on the validation set, emphasizing both classification accuracy and a detailed classification report. This evaluation strategy provided a robust measure of the model's ability to generalize, effectively confirming its improved accuracy and reliability in detecting arrhythmia patterns.

V. Results, Analysis and Critical Evaluation

a. Model Development and Analysis:

The LSTM model development followed a carefully structured iterative approach, informed by the detailed metrics of each phase as documented in Table 1. The initial setup for iteration 1 utilized a foundational baseline configuration of 128 hidden units across two layers, with a high learning rate of 0.01 and no dropout. This configuration led to poor convergence, as evidenced by significantly higher and fluctuating validation and training losses, alongside poor and inconsistent accuracy on both training and validation sets. The elevated learning rate and the absence of dropout likely contributed to this instability, lacking sufficient regularization needed to manage overfitting effectively. To address these challenges, Iteration 2 introduced an additional layer to increase model complexity and incorporated a 0.2 dropout rate to combat overfitting. Additionally, the learning rate was reduced to 0.001 to stabilize training. These changes significantly improved generalization, evidenced by lower validation loss and higher validation accuracy, resulting in a more stable and balanced model performance. Iteration 3 built on the previous adjustments by increasing the hidden units to 256 and simplifying back to two layers, maintaining the dropout rate and learning rate. This iteration resulted in the most effective model configuration, achieving the highest validation accuracy and demonstrating higher generalization on validation data. This progressive fine-tuning, guided by continuous monitoring of training and validation metrics, led to the selection of the best model configuration, highlighted in Table 1.

Table 1: LSTM best iteration

Iteration	Architecture (Hidden Units, Layers)	Dropout Rate	Learning Rate	Avg Training Loss	Avg Validation Loss	Avg Training Accuracy	Avg Validation Accuracy
1	128, 2	0.0	0.010	1.610	1.617	20.05%	23.02%
2	128, 3	0.2	0.001	0.260	0.383	85.47%	82.28%
3	**256, 2**	**0.2**	**0.001**	**0.368**	**0.414**	**85.64%**	**84.13%**

In the SVM model development, the initial setup with a linear kernel and C = 1 achieved a modest overall accuracy of 72.7% but performed very poorly on underrepresented classes. This prompted a strategic hyperparameter optimization using grid search to enhance model complexity and performance. The search, detailed in Table 2, assessed various combinations of kernels, C values, and gamma settings, calculating mean accuracies on a portion of the training data reserved for validation. The optimal parameters—C = 10, gamma = 0.01, and kernel = 'rbf'—were then applied to retrain the SVM on the full training dataset. This adjustment significantly enhanced the model's performance, achieving a 96% accuracy on the validation set, a substantial increase from the initial 72%. This improvement highlights the grid search's effectiveness in optimizing the SVM, significantly enhancing its robustness and generalization capabilities on complex data.

C	Gamma	Kernel	Avg Test Accuracy
0.1	scale	linear	0.7975
0.1	scale	rbf	0.8167
0.1	0.01	linear	0.7975
0.1	0.01	rbf	0.8130
0.1	0.1	linear	0.7975
0.1	0.1	rbf	0.4111
0.1	1	linear	0.7975
0.1	1	rbf	0.2338
1.0	scale	linear	0.8007
1.0	scale	rbf	0.9046
1.0	0.01	linear	0.8007
1.0	0.01	rbf	0.9281
1.0	0.1	linear	0.8007
1.0	0.1	rbf	0.8168
1.0	1	linear	0.8007
1.0	1	rbf	0.4957
10	scale	linear	0.8031
10	scale	rbf	0.9492
10	0.01	linear	0.8031
10	**0.01**	**rbf**	**0.9524**
10	0.1	linear	0.8031
10	0.1	rbf	0.8312
10	1	linear	0.8031
10	1	rbf	0.5153

Table 2 SVM GridSearch

b. Performance Comparison and Critical Insights:

The analysis and comparison of the LSTM and SVM best models on the unseen dataset, as illustrated in the classification reports in Table 3 and visualized through confusion matrices in Fig.5, offers a window into their distinct capabilities in handling complex sequential dataset, providing vital insights for their application in medical diagnostics.

Table 3: Classification Reports of LSTM and SVM models

LSTM	precision	recall	f1-score	support
Class 0	0.99	0.97	0.98	18118
Class 1	0.55	0.82	0.66	556
Class 2	0.91	0.94	0.93	1448
Class 3	0.62	0.83	0.71	162
Class 4	0.96	0.98	0.97	1608
Overall Acc			0.96	21892
Macro Avg.	0.81	0.91	0.85	21892
Weighted Avg	0.97	0.96	0.97	21892

SVM	precision	recall	f1-score	support
Class 0	0.98	0.93	0.96	18118
Class 1	0.41	0.65	0.50	556
Class 2	0.71	0.86	0.78	1448
Class 3	0.42	0.78	0.55	162
Class 4	0.94	0.95	0.94	1608
Overall Acc			0.92	21892
Macro Avg.	0.69	0.83	0.75	21892
Weighted Avg	0.94	0.92	0.93	21892

For the LSTM model, the classification report reveals commendable performance across the board, with outstanding precision and recall in underrepresented classes, such as Class 1, which exhibits a recall of 0.82. The LSTM's overall accuracy stands at an impressive 96.30%, signifying its adeptness at generalizing across a variety of ECG signal patterns.

In contrast, the SVM model achieves an overall accuracy of 92.26% but exhibits disparities across different classes. It demonstrates high precision and recall for Class 0, yet it encounters challenges with Class 1, showing a precision of only 0.41. While it achieves a recall of 0.65 for Class 1, the model also exhibits a high rate of false positives. In the medical field, such inaccuracies are particularly critical and can lead to detrimental outcomes. Classes 2 and 3 also reflect a decent recall but lower precision, pointing to potential confusion in class differentiation for SVM.

The macro-average scores paint a comprehensive picture of the LSTM's superiority in precision, recall, and f1-score, suggesting a more evenly balanced performance despite class imbalances, which is detailed in the comparative table.

In Figure 5, the SVM's confusion matrix on the right displays a significant challenge: it misclassifies supraventricular ectopic beats (Class 1) as normal (Class 0) approximately 28.96% of the time. This misclassification is particularly concerning because failing to detect supraventricular ectopic beats can have severe health implications, potentially leading to critical conditions if left untreated. While the LSTM also shows some misclassification, it predicts Class 1 as Class 0 only 16% of the time, offering a more reliable detection rate. Similarly, for Class 3, another critical but underrepresented category the SVM often misclassifies these potentially fatal abnormal beats as normal, occurring in approximately 11% of instances. In contrast, the LSTM model reduces this error rate to just 2.6%.

The critical insights from this comparative analysis boil down to recognizing the LSTM's superior handling of temporal dynamics, making it particularly suited for intricate, time-series medical data as hypothesized earlier. Conversely, while the SVM is robust in feature-rich environments, it needs further refinement to boost precision and handle minority classes better.

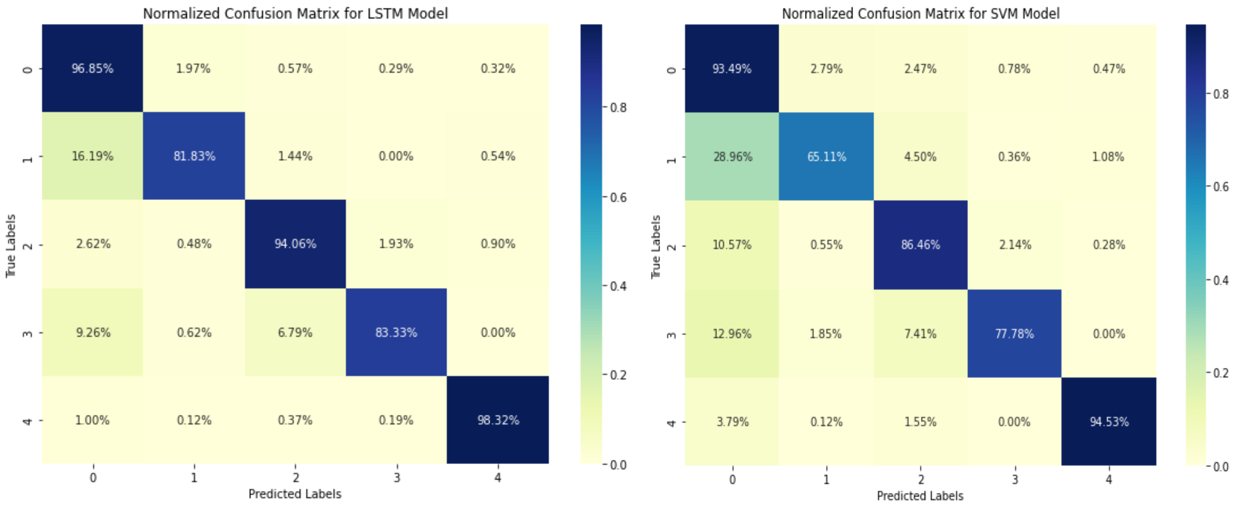


Fig 5: Normalized Confusion Matrix for LSTM & SVM

c. Conclusion

The final assessment between the LSTM and SVM models, as highlighted in the comparative ROC curve of Figure 6, reveals the LSTM's slightly superior performance, with its continuous lines consistently above the SVM's dashed lines for all classes. This performance highlights the LSTM's proficiency in handling the complexities of sequential ECG data, crucial for accurate medical diagnostics.

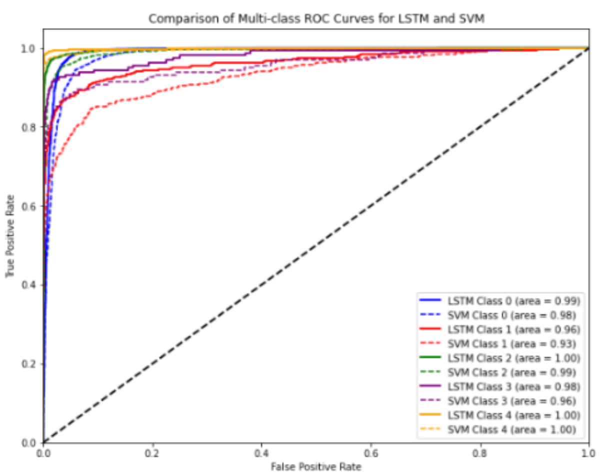


Fig 6: Comparative ROC curve

The LSTM model excels with higher accuracy and sensitivity, particularly in managing class imbalances—a critical advantage in healthcare applications where precision is vital to prevent misdiagnosis and ensure timely treatment. In contrast, the SVM, despite respectable overall accuracy, falls short in precision and shows higher false positive rates in underrepresented classes. This could lead to significant clinical risks, such as inappropriate treatment interventions.

It is important to acknowledge that this comparison may not entirely be fair due to the distinct preprocessing requirements for SVM, including feature extraction using DWT and dimensionality reduction using PCA, aimed at enhancing its data handling capabilities. These steps, necessary to maximize SVM's performance, underscore the varied approaches required to tune each model according to its operational strengths. This evaluation underlines the importance of model selection tailored to the complexities of the task at hand, reinforcing the LSTM's suitability for time-sensitive, high-accuracy medical diagnostics.

VI. Lessons Learned and Future Work

From the group coursework, we discovered that a single hidden layer MLP, even with increased hidden neurons, fell short in achieving desired precision across all classes of our dataset. Despite high overall accuracy, the model struggled with class-specific precision. Building on these insights, we recognized the need for more sophisticated models to enhance precision for individual classes. Key lessons learned include:

- Complex Models for Complex Data: The intricacies of our dataset required advanced models capable of capturing detailed patterns, leading us to explore LSTM and SVM architectures.

- Tailored Approaches Are Crucial: Customizing models to fit specific needs significantly outperforms the use of generic pretrained models. Tailoring allows for adjustments that directly address the unique challenges of the dataset.

For future works, we can focus on:

- Exploring Bidirectional LSTMs: Implementing a bidirectional architecture in the LSTM model could enhance its capability to grasp complex temporal relationships by processing data in both directions, potentially improving accuracy and precision across all classes.

- Investigating Complex SVM Kernels: Testing more sophisticated kernels like polynomial and sigmoid in the SVM could provide new ways to handle nonlinear interactions in the data, improving the model's efficacy beyond the capabilities of linear or RBF kernels.

- Advanced Feature Engineering: Delving deeper into feature engineering for the SVM could improve performance, particularly for minority classes. This would involve analyzing the feature selection post-DWT and PCA to pinpoint features that offer more discriminative power for challenging categories.

- Denoising ECG signals: Explore the impact of denoising ECG signals during preprocessing, as outlined in [11], to assess how effectively the optimized models perform on cleaner data.

VII. References

[1] S. C. Smith *et al.*, "Our Time: a call to save preventable death from cardiovascular disease (Heart disease and stroke)," *Circulation*, vol. 126, no. 23, pp. 2769–2775, Dec. 2012, doi: 10.1161/cir.0b013e318267e99f.
[2] B. T. Pham *et al.*, "Electrocardiogram heartbeat classification for arrhythmias and myocardial infarction," *Sensors (Basel)*, vol. 23, no. 6, p. 2993, Mar. 2023, doi: 10.3390/s23062993.
[3] R. T. J. J, "LSTMs Explained: A Complete, Technically Accurate, Conceptual Guide with Keras," *Medium*, Mar. 06, 2024. [Online]. Available: https://medium.com/analytics-vidhya/lstms-explained-a-complete-technically-accurate-conceptual-guide-with-keras-2a650327e8f2
[4] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks," *arXiv (Cornell University)*, Sep. 2019, doi: 10.48550/arxiv.1909.09586.
[5] M. Awad and R. Khanna, "Support vector clasification" in *Apress eBooks*, 2015, pp. 67–80. doi: 10.1007/978-1-4302-5990-9_4.
[6]"Support Vector Machine (SVM) Algorithm - JavatPoint," *www.javatpoint.com*. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm
[7] "Papers with Code - SVM Explained." https://paperswithcode.com/method/svm
[8] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.
[9] "MIT-BIH Arrhythmia Database V1.0.0," Feb. 24, 2005. https://physionet.org/content/mitdb/1.0.0/
[10] R. J. Martis *et al.*, "APPLICATION OF HIGHER ORDER CUMULANT FEATURES FOR CARDIAC HEALTH DIAGNOSIS USING ECG SIGNALS," *International Journal of Neural Systems*, vol. 23, no. 04, p. 1350014, Jun. 2013, doi: 10.1142/s0129065713500147.
[11] U. R. Acharya *et al.*, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Information Sciences*, vol. 405, pp. 81–90, Sep. 2017, doi: 10.1016/j.ins.2017.04.012.

Appendix 1: Glossary

1. **Electrocardiograms (ECG) -** A medical test that records the electrical activity of the heart over a period of time using electrodes placed on the skin, commonly used to detect heart abnormalities.
2. **Arrythmias -** Irregular heartbeats that can range from harmless to life-threatening, characterized by abnormal electrical activity in the heart.
3. **Sequential/ Temporal Data** - Data where the order of entries is significant, often involving time as a variable (e.g., time series data from sensors, stock prices, or ECG data).
4. **Sampling rate (360 Hz) -** The frequency at which an ECG signal is recorded, with 360 Hz indicating that the heart's electrical signals are sampled 360 times per second.
5. **Supraventricular Ectopic Beats (Class 1)** - Premature heartbeats originating from the atria, above the ventricles, which can disrupt the regular rhythm of the heart.
6. **Ventricular Ectopic Beats (Class 2)** - Premature heartbeats originating from the ventricles, the lower chambers of the heart.
7. **Support Vector Machine (SVM) -** A supervised machine learning model that uses classification algorithms for classification problems.
8. **Hyper-planes** - In the context of SVM, a hyper-plane is a decision boundary that separates different classes in high-dimensional space, effectively classifying the data.
9. **Kernel, C, Gamma -** These are parameters in SVM where 'Kernel' determines the type of hyper-plane used to separate data, 'C' controls the trade-off between smooth decision boundary and classifying training points correctly, and 'Gamma' defines the influence of a single training example.
10. **RBF (Radial Basis Function) -** A type of kernel used in SVM that measures the distance between a test point and every other point in the data, with a higher influence for points closer to the test point.
11. **Discrete Wavelet Transform (DWT) -** A mathematical transform that analyzes data at different frequency bands with different resolutions, often used for signal processing.
12. **Principal Component Analysis (PCA) -** A statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.
13. **Recurrent Neural Networks (RNN) -** A class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence, allowing it to exhibit temporal dynamic behavior.
14. **Long-Short Term Memory (LSTM)** - A type of recurrent neural network (RNN) used in deep learning that can learn order dependence in sequence prediction problems.
15. **Bidirectional LSTM's** - A type of LSTM that processes data in both forward and backward directions (future to past, past to future), providing additional context, improving model performance especially in tasks involving sequence data.
16. **Cross-Entropy Loss** - A loss function used typically in classification problems, which quantifies the difference between two probability distributions.
17. **Adam Optimizer** - An algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments.
18. **Learnable Initial States -** These are initial values for hidden states that are learned during the training process, which can potentially improve model performance by retaining useful information from the start of the sequence.

Appendix 2: Intermediate Results, Challenges, and Justification for Implementation Choices

A. For LSTM model

1. Challenges with Zero Initialization: The initial LSTM setup trials used zero-initialized states, which led to significant delays in learning and erratic training and validation results, as shown in Figure 7. Combined with a high learning rate and no dropout, this setup resulted in significant overfitting. To tackle these challenges, we shifted to learnable initial states, which greatly enhanced the model's stability and effectiveness.

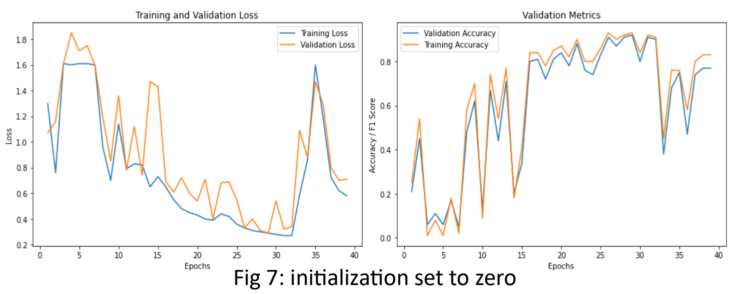


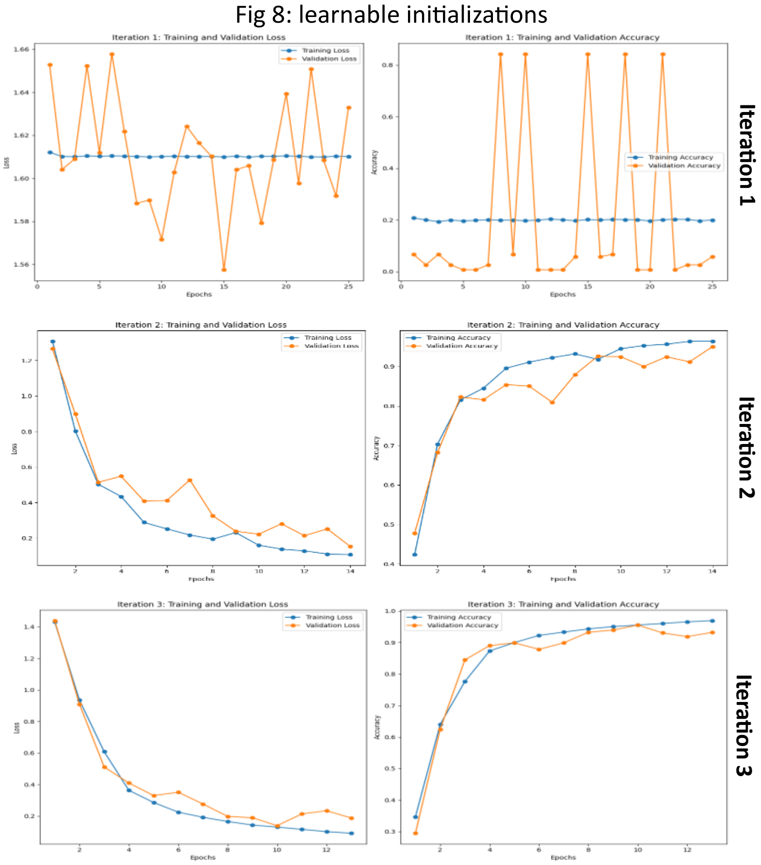
Fig 7: initialization set to zero

2. Iterative Experimentation: Attempts to increase the model's complexity with 256 hidden neurons and three layers worsened performance, suggesting overfitting due to excessive model complexity and thus, for the third iteration, the decision of balancing this complexity by reducing layers to 2 and increasing the hidden neurons to 256 was taken, which was proven to be the best model from all our iteration as shown in fig 8.

3. Computational Challenges and optimization technique difficulties: Our project faced considerable computational constraints given the extensive size of our dataset, which includes over 100,000 records. Due to unavailability of GPU on my computer, coupled with the quick depletion of resources on cloud platforms like Colab, significantly hindered our capacity for comprehensive model training and detailed hyperparameter optimization. This limitation not only reduced the scope of models we could effectively explore but also slowed down the iterative improvement process crucial for model refinement. Additionally, our initial attempts at employing advanced optimization methods such as GridSearch and Optuna were overly resource-intensive, resulting in prolonged execution times without substantial improvements in model performance. These challenges highlighted the practical difficulties in applying sophisticated machine learning techniques to large-scale, complex datasets.

The visuals of the metrics of three iterations included in the report are shown in fig 8.

Fig 8: learnable initializations



B. For svm model:

For SVM, implementing PCA was a necessary decision despite the potential for an unfair comparison with LSTM. The original dataset featured 187 features across 109,446 records, posing significant computational challenges. Without PCA, conducting a comprehensive grid search to optimize the SVM would have been impractical due to these constraints. Thus, despite the risk of an uneven comparison, PCA, combined with DWT, taken from literature [10], was employed to maximize SVM's performance by reducing dimensionality and enhancing its capability to handle the data efficiently. This approach allowed us to leverage SVM's strengths effectively, while for LSTM, which excels with sequential data, such preprocessing was not appropriate.