ISSN: 2834-4928

Research Article

Journal of Electrical Electronics Engineering

Machine Learning Models to Help Classification of Cardiovascular Disease

Khaoula Oueldji[1*] and Nadir Farah[2]

[1]*Badji Mokhtar University Annaba, Algeria*

[2]*Department of Computer Science, LABGED Laboratory, 23000, Annaba, Algeria*

[*]Corresponding Author
Khaoula Oueldji, Badji Mokhtar University Annaba, Algeria.

Submitted: 2024, Nov 01; **Accepted**: 2024, Nov 25; **Published**: 2024, Dec 16

Citation: Oueldji, K., Farah, N. (2024). Machine Learning Models to Help Classification of Cardiovascular Disease. *J Electrical Electron Eng, 3*(6), 01-10.

Abstract
Electrocardiography (ECG) has become a widely used noninvasive diagnostic tool, increasingly supported by algorithmic analysis. However, progress in automated ECG interpretation faces challenges due to the lack of adequate training datasets and standardized evaluation procedures, which are crucial to ensure comparability of algorithms. In this study, ECG classification models are proposed using the recently published PTB-XL dataset of 12 clinical leads. This research aims to overcome existing limitations by thoroughly investigating the performance of different deep learning-based classification algorithms. Specifically, we investigate the effectiveness of convolutional neural networks (CNNs), deep neural networks (DNNs), long short-term memory (LSTM) and U-net architectures in accurately classifying ECG signals. In addition, we explore the potential of reinforcement learning techniques using classifiers pre-trained on PTB-XL to further improve classification accuracy and robustness. This comprehensive analysis not only underscores the significant potential of deep learning algorithms in ECG analysis, but also highlights the importance of standardized datasets such as PTB-XL in advancing the field. By establishing PTB-XL as a key resource, this study aims to foster collaboration among researchers and encourage further contributions aimed at refining and extending the dataset to better serve the ECG analysis community.

Keywords: Electrocardiography (ECG), Deep Learning, Classification Algorithms, PTB-XL Dataset, Convolutional Neural Networks (CNNs), Reinforcement Learning

1. Introduction
Cardiovascular diseases (CVDs) rank among the leading causes of death worldwide second only to cancer in high-income countries [22,6]. Electrocardiography (ECG) is a non-invasive tool used to assess a patient's overall heart health and serves as a crucial diagnostic tool for cardiovascular diseases. In the United States, approximately 5% of medical visits involve an ECG [5]. Despite its importance, interpreting ECGs is a challenging task, even for cardiologists, let alone residents, general practitioners or emergency physicians who must interpret ECGs in urgent situations [16,8]. Furthermore, the growing field of telemedicine, particularly ECG Holter monitoring, will further emphasize the need for advanced decision support systems based on automated ECG interpretation algorithms. Recent years have seen significant advancements in automated ECG interpretation algorithms, particularly those based on deep learning. These approaches have not only matched or surpassed cardiologist-level performance for certain tasks but have also enabled difficult-to-make statements at the cardiologist level, such as determining ECG diagnoses [3,9,4, 11 and 14].

The ease of data collection and reduced dimensionality compared to imaging data have also drawn the broader machine learning community towards ECG classification, as evidenced by numerous research articles each year, see for a recent review [10]. Significant advancements in computer vision, such as object recognition, were largely due to the availability of large datasets and the competitive nature of classification challenges with clear evaluation procedures. However, two major issues hinder the progression of ECG analysis: firstly, publicly accessible ECG datasets are generally small and the large datasets currently available remain inaccessible to the public [17]. This issue has been partially addressed by the release of the PTB-XL dataset hosted by PhysioNet which currently stands as the largest freely accessible ECG dataset [19,20]. Furthermore, many freely accessible databases contain only single-lead recordings, making comprehensive diagnosis and clinical validation challenging. However, large and comprehensive databases with 12-lead recordings are rather an exception, hence the importance of the underlying dataset for algorithmic solution development. Secondly, most existing datasets provide raw data, but there are no clearly defined tasks with corresponding evaluation procedures. This limits the comparability of different algorithms, as experimental details such as data selection, train-test splits, evaluation metrics, and score estimation can significantly impact the final outcome. We propose a selection of models dedicated to ECG classification while also exploring the application of reinforcement learning on the PTB-XL dataset [19]. In conclusion, our main contribution in this article is to implement and adapt various deep learning and reinforcement learning-based time series classification algorithms in ECG classification. This addresses current challenges in ECG classification by providing effective models and methods for ECG classification.

2. Materials and Methods

2.1 PTB-XL Dataset

This section briefly introduces the PTB-XL dataset which underpins most of the experiments presented below [19]. The PTB-XL dataset comprises 21,837 clinical ECG recordings with 12 leads, each lasting 10 seconds, from 18,885 patients, of whom 52% were male and 48% were female. The ECG statements used for annotation adhere to the SCP-ECG standard and have been assigned to three non-mutually exclusive categories [11]. In total, there are 71 different statements, which break down into 44 diagnostic statements, 12 rhythm statements, and 19 shape statements, with 4 also used as ECG diagnostic statements. For diagnostic statements, a hierarchical organization into five coarse super classes and 24 subclasses is also provided; refer to Figure 1 for a graphical summary in terms of diagnostic super classes. Further details on the dataset, annotation scheme, and other ECG datasets can be found in [19]. In summary, PTB-XL stands out not only for its size as the largest publicly accessible clinical ECG dataset to date but also for its rich set of ECG annotations and other metadata, which make the dataset an ideal resource for training and evaluating machine learning algorithms. Throughout this article, the recommended train-test splits provided by PTB-XL is used, which take into account patient assignments and use input data sampled at a frequency of 100 Hz. Refer to Figure 1 for a visual overview of the dataset, which provides varying levels of detail. Alongside annotations such as ECG statements and likelihood information for diagnostic statements, additional metadata, such as manually annotated signal quality statements, is also available.

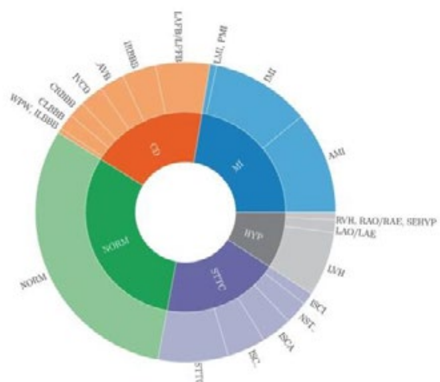

Figure 1: Graphical Summary of the PTB-XL Dataset in Terms of Diagnostic Super classes and Subclasses [21]

2.2 Preprocessing the PTB-XL Dataset

The PTB-XL dataset was preprocessed to convert diagnostic classes into a binary format compatible with machine learning models. The preprocessing process involved several steps:

Data Loading ECG signals were loaded from .npy files for both training and test sets. Corresponding class labels were loaded from .csv files. Binarization of Class Labels The sklearn. Preprocessing library and its Multi Label Binarizer tool were used to convert diagnostic classes into a format suitable for machine learning. MultiLabel Binarizer can handle cases where a sample can belong to multiple classes (multi-label). The fit transform method was applied to the training labels (y_train) to fit the binarizer to the data and convert them into a binary matrix. The transform method was then applied to the test labels (y_test) to obtain their binary representation using the binarizer fitted on the training data.

This preprocessing process transformed the PTB-XL's diagnostic classes into a binary format usable by machine learning algorithms. The conversion to a binary format facilitates model learning and improves classification accuracy.

2.3 The used Machine Learning Models

The primary objective of this article was to explore the complex field of electrocardiogram (ECG) classification, with a particular focus on the delineation of diagnostic super classes. Departing from conventional methods, which often involve complex pre-processing steps such as baseline noise detection and suppression, we opted for a more direct approach. This article sought to harness the power of cutting-edge deep learning algorithms capable of navigating the complexities inherent in raw multivariate time series data.

Convolutional Neural Networks (CNNs) These pillars of the deep learning domain are renowned for their innate ability to automatically extract relevant features from raw data. By employing a series of convolutional layers, punctuated by pooling layers, CNNs possess unparalleled prowess in discerning complex local patterns within ECG signals, positioning them as leading candidates for classification tasks.

Deep Neural Networks (DNNs) Although perhaps less inherently suited for sequence modeling than their recurrent counterparts, DNNs wield considerable influence in the realm of ECG classification. Their strength lies in their proficiency in extracting salient features from raw data, thereby significantly enhancing the efficacy of the classification pipeline [7].

Long Short-Term Memory (LSTM) Networks Stepping into the limelight, LSTMs represent a sophisticated variant of recurrent neural networks designed to excel in capturing long-term dependencies within sequential data. Leveraging intricate gate mechanisms, LSTMs adeptly navigate the ebb and flow of information over time, making them exceptionally well-suited for modeling the nuanced temporal dynamics inherent in ECG data [12].

U-net Architecture Venturing into the realm of hybrid architectures, the Unet stands as a beacon of innovation, seamlessly integrating the strengths of both CNNs and RNNs. By adeptly capturing both spatial and temporal information embedded within the data, the U-net emerges as a formidable ally in the quest for efficient processing of sequential ECG signals [15].

Reinforcement Learning with LSTM Pushing the boundaries of conventional paradigms, we delved into the realm of reinforcement learning, harnessing the formidable capabilities of LSTMs as the cornerstone of this learning framework. This pioneering approach holds immense promise for facilitating sequential decision-making and dynamic adaptation to the ever-evolving complexities of input data dynamics. More specifically, the reinforcement learning framework adopted relies on LSTMs to model the temporal dynamics of ECG signals. LSTMs serve as function approximators in a Markov decision process, enabling the agent to learn to make optimal decisions over time to correctly classify ECG signals. The reward function is designed to encourage the agent to produce accurate predictions while minimizing the number of classification errors. The key mathematical formulas of this reinforcement learning framework are as follows: The Value Function:

$$V(s_t) = E[R_t \mid s_t] \quad (1)$$

represents the expected cumulative reward R_t given the state s_t of the ECG at time t. It serves as a critical metric for evaluating the desirability of different states.

The Reward Function r_t is defined as −1 if the prediction is incorrect and 1 if it is correct. It provides immediate feedback to the agent based on its actions. The Update of the Value Function:

$$V(s_t) = V(s_t) + \alpha \bullet (r_t + \gamma \bullet V(s_{t+1}) - V(s_t)) \quad (2)$$

Outlines how the value function is iteratively updated. Here, α denotes the learning rate, controlling the extent to which new information overrides old, while γ represents the discount factor for future rewards, influencing the agent's long-term planning horizon. In summary, this reinforcement learning approach empowers the agent to autonomously acquire optimal decision-making abilities, dynamically adjusting to the complexity of the data. It is important to emphasize that the models presented have been rigorously trained directly on raw time series data, avoiding the need for cumbersome pre-processing steps. This strategic choice was driven by the desire to preserve the simplicity and universality of these models, while avoiding potential pitfalls associated with the introduction of additional parameters. In essence, this article not only highlights the enormous potential of using deep learning for ECG classification, but also underscores the need for continued research and innovation in the field. By exploring new frontiers and pushing the boundaries of existing methods, it paves the way for a future where the interpretation of ECG data is not only more accurate, but also more accessible to clinicians worldwide.

2.4 Hyperparameter Tuning

Hyperparameter	**Tested Ranges**	**Selection Criterion**	**Chosen Result**
Number of convolutional layers	2-4	Impact on the model's ability to extract relevant features	3
Number of filters per layer	16, 32, 64, 128	Impact on model complexity and its ability to capture patterns	32
Kernel size	3, 5, 7	Influence on the size of the convolution windows and pattern capturing	3
Pooling (pool size)	2, 3, 4	Impact on dimensionality reduction and preservation of important features	2
Number of neurons in dense layers	64, 128, 256	Influence on the model's ability to perform final classification	128
Dropout rate	0.2, 0.5, 0.7	Impact on model regularization and prevention of overfitting	0.5
Optimizer	Adam, RMSprop, SGD	Fast convergence and good performance on the test set	Adam

Loss function	sparse categorical crossentropy, categorical crossentropy	Adaptation to multi-class classification task and performance on the test set	sparse categorical crossentropy
Number of epochs	5-20	Performance optimization without overfitting	10
Batch size	16-64	Good performance with satisfactory training efficiency	32

Table 1: Hyperparameter Tuning for CNN

2.5 Results Analysis and Discussions

The PTB-XL dataset represents a diverse collection of electrocardiograms (ECGs) accompanied by various labels and metadata, providing a rich substrate for in-depth analysis. In this section, we present a detailed investigation of classification experiments performed on the PTB-XL dataset, with the aim of providing insights into the performance and effectiveness of different deep learning models. The evaluation metrics used cover a range of performance indicators, including accuracy, precision, recall and F1 score. These metrics serve as important benchmarks for assessing the effectiveness of classification models in accurately categorizing ECG data. Table 1 summarizes the performance results of different deep learning models used in the classification experiments on the PTB-XL dataset. The CNN, DNN, U-Net and LSTM models show commendable accuracy levels, indicating their ability to handle classification tasks. In particular, the RL with LSTM model emerges as the frontrunner with an outstanding accuracy of 98%, indicating its robustness in ECG classification.

A comprehensive examination of model performance reveals some interesting insights. While A comprehensive examination of model performance, validated by a cardiologist, reveals some interesting insights. While the CNN, DNN, U-Net and LSTM models deliver commendable results, the RL with LSTM model emerges as the standout performer, underlining its superiority in handling complex time series data such as ECGs. This nuanced analysis highlights the strengths and weaknesses of each model.

Hyperparameter	Tested Ranges	Selection Criterion	Chosen Result
Number of Dense layers	1-3	Impact on the model's ability to learn hierarchical representations	3
Number of neurons per layer	64, 128, 256	Influence on the model's ability to learn complex patterns	128
Dropout rate	0.1, 0.2, 0.3	Impact on model regularization and prevention of overfitting	0.2
Optimizer	Adam, RMSprop, SGD	Fast convergence and good performance on the test set	Adam
Loss function	categorical crossentropy, binary crossentropy	Adaptation to multi-class classification task and performance on the test set	categorical crossentropy
Number of epochs	5-20	Performance optimization without overfitting	10
Batch size	16-64	Good performance with satisfactory training efficiency	32

Table 2: Hyperparameter Tuning for DNN

In addition to traditional deep learning architectures, alternative classifiers such as RCNN and TCNN were investigated. However, these models failed to outperform the baseline deep learning models, highlighting the unique advantages of architectures designed specifically for sequential data analysis.

Furthermore, the comparative analysis of deep learning models on the PTBXL dataset confirms the effectiveness of RL with LSTM in ECG classification tasks, as validated by a cardiologist. This finding suggests its potential for real-world applications.

In the future, further exploration of ensemble techniques, hyperparameter tuning and alternative architectures promises to unlock even greater performance gains and advance the state of the art in ECG analysis and diagnosis.

3. Conclusion

Electrocardiography (ECG) plays a crucial role in medical diagnosis, both in hospitals and in doctors' offices. Automated ECG interpretation using algorithms has significant potential in various medical fields. However, despite progress, the development of these technologies is currently hampered by the lack of well-defined reference datasets and evaluation procedures. This study explores a variety of classification models on the PTB-XL dataset to provide a better model for classification and a foundation for future research in this area.

Hyperparameter	Tested Ranges	Selection Criterion	Chosen Result
Number of Convolu- tional layers	2-4	Impact on the model's ability to extract relevant features	3
Number of filters per layer	32, 64, 128	Influence on the model's complexity and its ability to capture patterns	32
Kernel size	2, 3, 4	Influence on the size of the convolution window and pattern capturing	3
Pooling size	2, 3, 4	Impact on dimensionality reduction and preservation of important features	2
Number of neurons in dense layers	64, 128, 256	Influence on the model's ability to perform final classification	64
Dropout rate	0.1, 0.2, 0.3	Impact on model regularization and prevention of overfitting	0.2
Optimizer	Adam, RMSprop, SGD	Fast convergence and good performance on the test set	Adam
Loss function	sparse categorical crossentropy, categorical crossentropy	Adaptation to multi-class classification task and performance on the test set	sparse categorical crossentropy
Number of epochs	5-20	Performance optimization without overfitting	20
Batch size	16-64	Good performance with satisfactory training efficiency	32

Table 3: Hyperparameter Tuning for U-Net

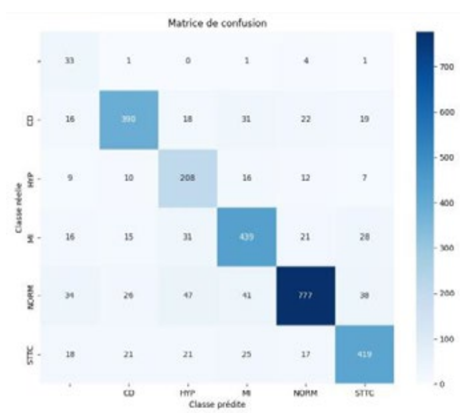

Figure 2: Confusion Matrix for CNN

The first results highlight the effectiveness of temporal classification algorithms. In particular, we observed that the RL with LSTM model stands out for its remarkable performance, but the paper shows that other models such as CNN, DNN, LSTM, and U-Net also show promising competitiveness.

Hyperparameter	Tested Ranges	Selection Criterion	Chosen Result
Units (LSTM Units)	32, 64, 128	Best compromise between precision, recall, and F1-score on validation set	64
Dropout	0.1, 0.2, 0.3	Best generalization and performance on test set	0.2
Optimizer	Adam, RMSprop, SGD	Fast convergence and good performance on test set	Adam
Loss Function	categorical crossentropy, binary crossentropy	Adaptation to multiclass classification task and performance on test set	categorical crossentropy
Number of Epochs	5-20	Optimization of performance without overfitting	10
Batch Size	16-64	Good performance with satisfactory training efficiency	32

Table 4: Hyperparameter Tuning for LSTM

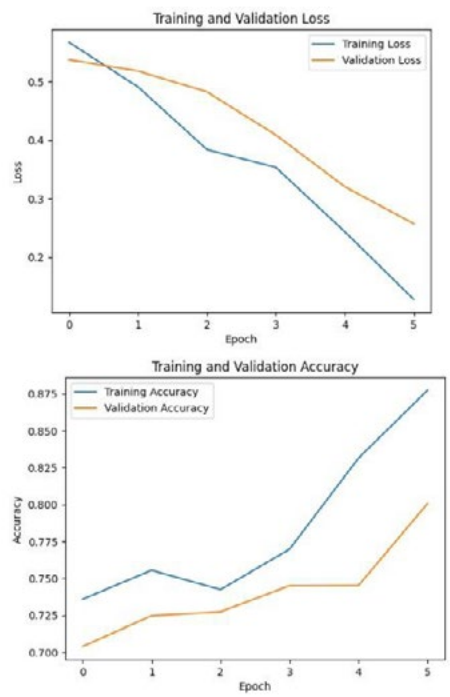

Figure 3: Training and Validation Loss/ Accuracy for CNN

These results suggest that algorithms have the potential to revolutionize ECG interpretation by providing faster and more accurate solutions.

Hyperparameter	Tested Ranges	Selection Criterion	Chosen Result
Number of LSTM neu-rons per layer	64, 128, 256	Influence on the model's ability to capture sequential patterns	128
Number of training episodes	5-20	Impact on agent convergence and model learning	10
Activation function of the last layer	Sigmoid, Softmax	Adaptation to classification task and model performance	Sigmoid
Loss function	categorical crossentropy, binary	Adaptation to multiclass classification task crossentropy	binary crossentropy
Optimizer	Adam, RMSprop, SGD	Fast convergence and good performance on the test set	Adam
Batch size	32, 64, 128	Impact on training efficiency and learning stability	32

Reward	0, 1, 2	Impact on reinforcement learning and agent's final performance	1

Table 5: Hyperparameter Tuning for RL-LSTM

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
CNN	0.80	0.79	0.80	0.79
DNN	0.75	0.57	0.75	0.65
U-Net	0.77	0.73	0.58	0.61
LSTM	0.75	0.56	0.75	0.66
RL-LSTM	0.98	0.67	0.75	0.65

Table 6: Obtained Results

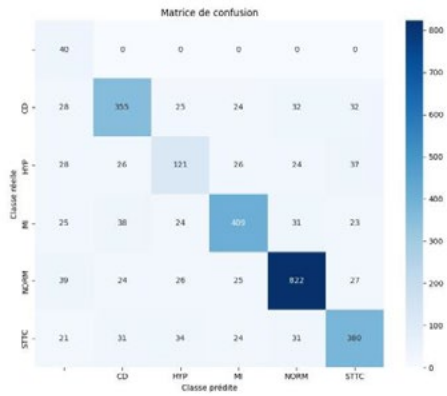

Figure 4: Confusion Matrix for DNN

This advancement is particularly crucial for the medical field, especially in the diagnosis of cardiovascular pathologies. The study underscores the importance of such work in aiding decision-making processes in healthcare.

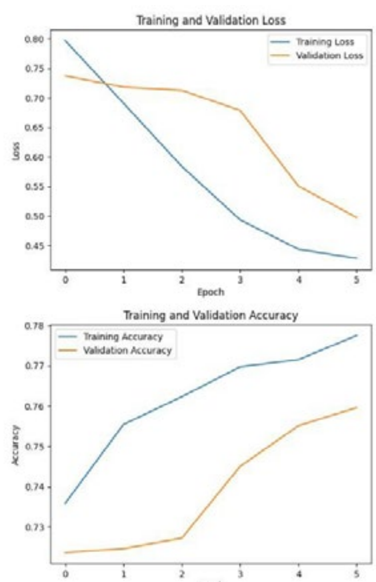

Figure 5: Training and Validation Loss/ Accuracy for DNN

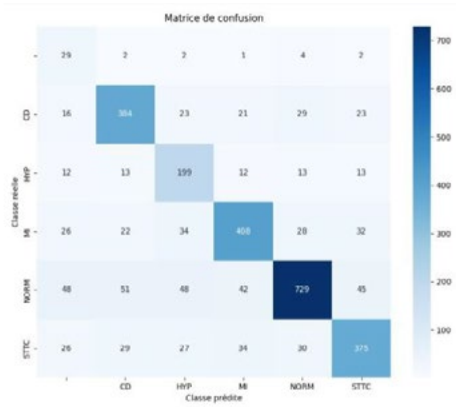

Figure 6: Confusion Matrix for U-net

However, further efforts are needed to consolidate these advances, including the development of larger reference datasets and the establishment of rigorous evaluation standards. In addition, this study provides insights into the capabilities of automated ECG interpretation algorithms, highlighting both the progress made and the challenges ahead for wider adoption of these technologies in medical practice.

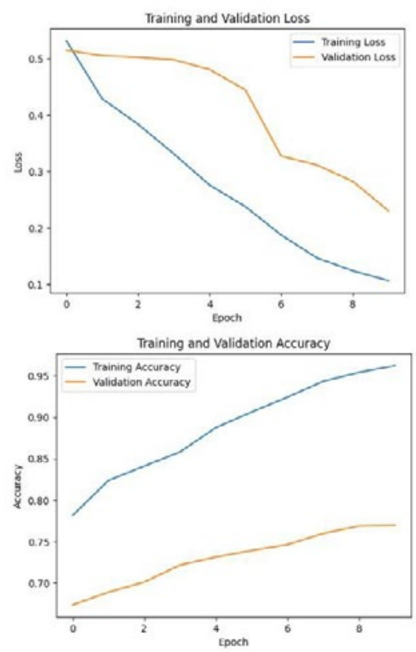

Figure 7: Training and Validation Loss/ Accuracy for U-net

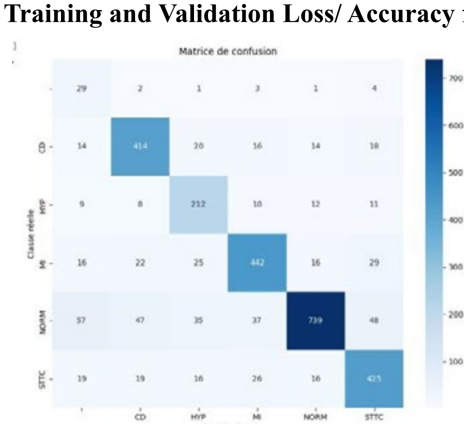

Figure 8: Confusion Matrix for LSTM

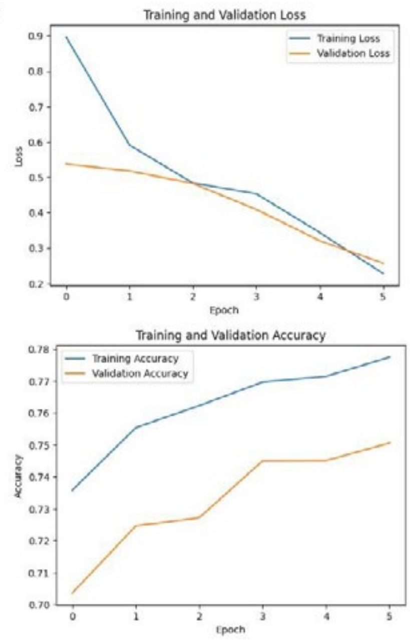

Figure 9: Training and Validation Loss/ Accuracy for LSTM

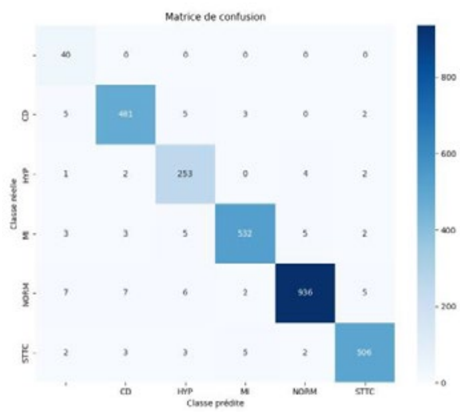

Figure 10: Confusion Matrix for RL-LSTM

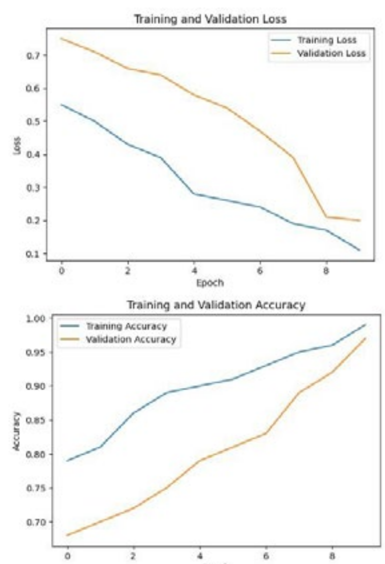

Figure 11: Training and Validation Loss/ Accuracy for RL-LSTM

References

1. Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., & Friedman, P. A. (2019). Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nature medicine, 25*(1), 70-74.
2. Attia, Z. I., Friedman, P. A., Noseworthy, P. A., Lopez-Jimenez, F., Ladewig, D. J., Satam, G., & Kapa, S. (2019). Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circulation: Arrhythmia and Electrophysiology, 12*(9), e007284.
3. Attia, Z. I., DeSimone, C. V., Dillon, J. J., Sapir, Y., Somers, V. K., Dugan, J. L., & Friedman, P. A. (2016). Novel bloodless potassium determination using a signal-processed single-lead ECG. *Journal of the American heart Association, 5*(1), e002746.
4. Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., & Friedman, P. A. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet, 394*(10201), 861-867.
5. CDC. (2019). National ambulatory medical care survey: 2016 national summary tables. Technical report. Centers for Disease Control and Prevention.
6. Dagenais, G. R., Leong, D. P., Rangarajan, S., Lanas, F., Lopez-Jaramillo, P., Gupta, R., & Yusuf, S. (2020). Variations in common diseases, hospital admissions, and deaths in middle-aged adults in 21 countries from five continents (PURE): a prospective cohort study. *The Lancet, 395*(10226), 785-794.
7. Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer methods and programs in biomedicine, 161*, 1-13.
8. Fent, G., Gosai, J., & Purva, M. (2015). Teaching the interpretation of electrocardiograms: which method is best? *Journal of electrocardiology, 48*(2), 190-193.
9. Hannun, A. Y., et al. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med, 25*(1), 65-69.
10. Hong, S., Zhou, Y., Shang, J., Xiao, C., & Sun, J. (2020). Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in biology and medicine, 122*, 103801.
11. ISO Central Secretary. (2009). Health informatics Standard communication protocol Part 91064: Computer-assisted electrocardiography. *International Organization for Standardization, Standard ISO 11073-91064:2009*.
12. Nagabushanam, P., Thomas George, S., & Radha, S. (2020). EEG signal classification using LSTM and improved neural network algorithms. *Soft Computing, 24*(13), 9981-10003.
13. What is reinforcement learning? Reinforcement learning explained - AWS. (n.d.). Amazon Web Services, Inc.
14. Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., & Ribeiro, A. L. P. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature communications, 11*(1), 1760.
15. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III* 18 (pp. 234-241). Springer International Publishing.
16. Salerno, S. M., Alguire, P. C., & Waxman, H. S. (2003). Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence. *Annals of Internal Medicine, 138*(9), 751-760.
17. Schläpfer, J., & Wellens, H. J. (2017). Computer-interpreted electrocardiograms: benefits and limitations. *Journal of the American College of Cardiology, 70*(9), 1183-1192.
18. Strodthoff, N., Wagner, P., Schaeffter, T., & Samek, W. (2020). Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *IEEE journal of biomedical and health informatics, 25*(5), 1519-1528.
19. Wagner, P., Strodthoff, N., Bousseljot, R. D., Kreiseler, D., Lunze, F. I., Samek, W., & Schaeffter, T. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Scientific data, 7*(1), 1-15.
20. Wagner, P., Strodthoff, N., Bousseljot, R. D., Kreiseler, D., Lunze, F. I., Samek, W., & Schaeffter, T. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Scientific data, 7*(1), 1-15.
21. Wagner, P., Strodthoff, N., Bousseljot, R. D., Kreiseler, D., Lunze, F. I., Samek, W., & Schaeffter, T. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Scientific data, 7*(1), 1-15.
22. Wilkins, E. (2017). European cardiovascular disease statistics. European Heart Network, Belgium.