Heart Failure Prediction

Deep Learning Methods with RNN-GRU, RNN-LSTM, and Transformers

Yue (Yelena) Yu, MSCS San Jose
Principal Investigator: Samir Farooq, Arcadia.io
Professor: Lee, Jeongkyu, Northeastern University

Northeastern University
Khoury College of Computer Sciences

BACKGROUND

Heart failure (HF) is a major public health concern with high mortality and healthcare costs, emphasizing the importance of accurate early prediction. This research explores the application of deep learning models, specifically **Recurrent Neural Networks with Long Short-Term Memory (RNN-LSTM)**, **Gated Recurrent Units (RNN-GRU)**, and **transformer** architectures to predict heart failure risk using patient time-series data. The goal is to reduce morbidity and improve patient outcomes.

Existing Work

Recurrent Neural Networks (RNNs), including GRU and LSTM, have demonstrated effectiveness in capturing sequential dependencies in clinical data, while transformer models have shown promise for their ability to handle long-term dependencies.

For example, Rahman et al. (2024) achieved high accuracy (96.51%) using self-attention-based transformer models for heart disease prediction, demonstrating their capability to identify critical features and offer interpretability. Antikainen et al. (2023) demonstrated the superiority of XLNet over BERT for six-month cardiovascular mortality prediction using heterogeneous EHRs, capturing complex interdependencies in multimodal data. Similarly, Choi et al. (2017) highlighted the efficacy of GRU-based RNNs in early heart failure detection, achieving an AUC of 0.883 over an 18-month observation window.

Novelty

Our study utilizes the largest and most diverse real-world patient dataset to date, with over 16,000 records from Arcadia.io with diverse demographic and vital sign data. Our data orchestration pipeline ensures seamless integration of sequential and non-sequential data, providing a comprehensive input structure for advanced deep learning models.

DATA PREPARATION

- Vital signs were processed into a time-series matrix with 12 intervals at 3-month duration over 3 years. Demographic information was handled separately as non-sequential data.
- The study aimed to predict heart failure probabilities for the next year within timeframes of 0–1, 0–3, 0–6, and 0–12 months, aligning with healthcare industry standards.

Data Source: sponsored by Arcadia.io

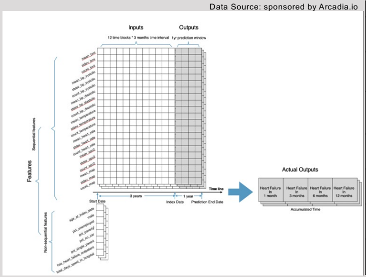

Figure 1: Input-Output Data Representation

Sequential Features		Non-sequential Features
BMI	Mean	Age
	Std Dev	Gender
	Count	pct unemployed
bp systolic	Mean	pct no car
	Std Dev	pct single parents
	Count	pct poverty
bp diastolic	Mean	has heart failure before
	Std Dev	days spent in hospital
	Count	
temperature	Mean	
	Std Dev	
	Count	
heart rate	Mean	
	Std Dev	
	Count	
blood oxygen	Mean	
	Std Dev	
	Count	
Mean arterial pressure	Mean	
	Std Dev	
	Count	

Table 1: Sequential and Non-Sequential Features

APPROACHES

- **Model 1:** Recurrent Neural Networks **with Long Short-Term Memory** (RNN-LSTM)
- **Model 2:** Recurrent Neural Networks with **Gated Recurrent Units** (RNN-GRU)
- **Model 3:** Transformer – Multi-head Attention Mechanism (**Decoder-Only Architecture**)
- **Model Performance Evaluation:** Area Under Curve (AUC)
- **Loss Function:** Binary Cross Entropy (BCE)
- **Hyper-Parameter Tuning:** Optuna
- **Optimize:** Adam

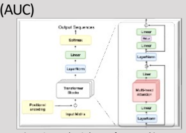

Figure 2: Decoder-Only Transformer Architecture

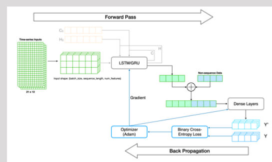

Figure 3: GRU/LST Model Architecture

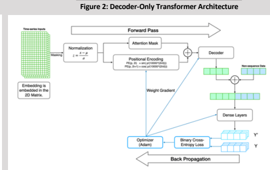

Figure 4: Transformer Model Architecture

Hyper-Parameter Tuning

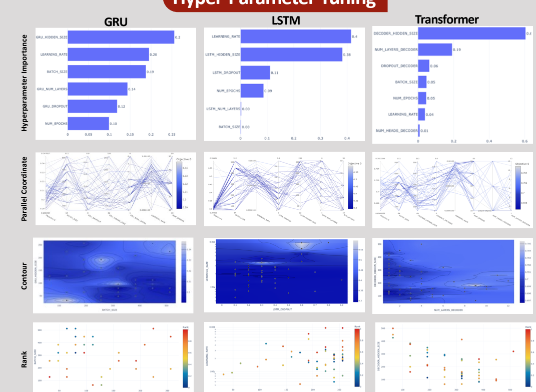

Figure 5: Hyper-Parameter Tuning Results

RESULTS

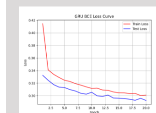

Figure 6a: GRU Binary Cross-Entropy (BCE) Loss Curve

Hyper-Parameters	GRU	LSTM	Transformer
Input Sequence Length	12	12	12
Number of Input Features *	21 + 8	21 + 8	21 + 8
(Decoder) Hidden Size	32	256	128
Number of Decoder Head (Transformer only)	/	/	3
Number of Decoder Layers (Transformer only)	/	/	6
Number of Layers	3	8	1
Dropout	0.3	0.3	0.3
Output Sequence Length	4	4	4
Learning Rate	0.0002	0.0002	0.0001
Batch Size	256	256	64
Number of Epochs	20	20	40
Optimizer	Adam	Adam	Adam
Loss Function	BCE**	BCE**	BCE**

* Input Features include 21 sequential and 8 non-sequential
** Binary Cross Entropy

Table 2: Optimal Hyper-Parameter Setting

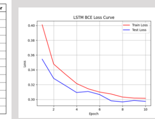

Figure 6b: LSTM Binary Cross-Entropy (BCE) Loss Curve

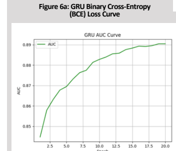

Figure 7a: GRU AUC Curve

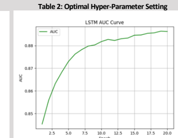

Figure 7b: LSTM AUC Curve

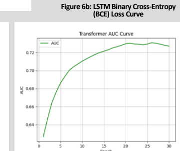

Figure 7c: transformer AUC Curve

GRU performs the best out of the three architectures

Result Examples

GRU:
Model Output: [1.3614e-05, 4.1598e-01, 9.9903e-01, 1.0000e+00]
Label: [0., 0., 1., 1.]

LSTM:
Model Output: [1.0392e-04, 1.7163e-01, 8.9943e-01, 9.9999e-01]
Label: [0., 0., 1., 1.]

Transformer:
Model Output: [1.1527e-05, 1.2703e-01, 9.9540e-01, 9.9998e-01]
Label: [0., 0., 1., 1.]

- GRU is better suited for short sequences and simpler data dependencies.
- Recent health conditions are more predictive of future outcomes, favoring RNN architectures.
- The simplicity and lower complexity of the dataset make RNNs more efficient and practical compared to transformers.

CONCLUSION & FUTURE EXPLORATION

- Transformers are **data greedy** and **require large, complex datasets** to perform optimally.
- Transformers generate numerous **sequence combinations**, which can overwhelm smaller datasets.
- Transformer architectures require **extensive design, validation, and hyper-parameter tuning**.

- Investigating transformer architecture to better align with our dataset and prediction goals.
- Introducing **embedding layers to enhance the representation** of input features.
- Implementing **L2 regularization** in hidden layers to prevent overfitting.
- Applying **batch normalization** to stabilize training dynamics and accelerate convergence.
- Conducting **extensive hyperparameter tuning** to unlock the full potential of transformer.
- **Expanding the feature set and increasing the sequence length** to allow the model to capture more complex temporal patterns and contextual dependencies.
- **Scaling up the dataset size** to better utilize the data-hungry nature of transformers.

References

[1] Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems.
[2] Rahman, A.U., Alsenani, Y., Zafar, A. et al. Enhancing heart disease prediction using a self-attention-based transformer model. Sci Rep 14, 514 (2024).
[3] Antikainen, E., Linnosmaa, J., Umer, A. et al. Transformers for cardiac patient mortality risk prediction from heterogeneous electronic health records. Sci Rep 13, 3517 (2023). https://doi.org/10.1038/s41598-023-30657-1
[4] Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc. 2017 Mar 1;24(2):361-370. doi: 10.1093/jamia/ocw112. PMID: 27521897; PMCID: PMC5391725.

CS 7980 Research Capstone