

# Stepping towards Medical Language Understanding

Aman Sinha

30th November 2023

- 3rd year PhD Student @ Université de Lorraine(UL), Nancy, France

- 3rd year PhD Student @ Université de Lorraine(UL), Nancy, France
  - Supervisors : Marianne Clausel, Mathieu Constant, and, Xavier Coubez

- 3rd year PhD Student @ Université de Lorraine(UL), Nancy, France
  - Supervisors : Marianne Clausel, Mathieu Constant, and, Xavier Coubez
  - Affiliated to IECL@UL (Mathematics lab), ATILF@UL (NLP lab) and ICANS Strasbourg (Cancer Research Institute)

- 3rd year PhD Student @ Université de Lorraine(UL), Nancy, France
  - Supervisors : Marianne Clausel, Mathieu Constant, and, Xavier Coubez
  - Affiliated to IECL@UL (Mathematics lab), ATILF@UL (NLP lab) and ICANS Strasbourg (Cancer Research Institute)
- *Multi-source deep learning models for medical domain.* More broadly, I am studying the topic of medical language understanding.

- 3rd year PhD Student @ Université de Lorraine(UL), Nancy, France
  - Supervisors : Marianne Clausel, Mathieu Constant, and, Xavier Coubez
  - Affiliated to IECL@UL (Mathematics lab), ATILF@UL (NLP lab) and ICANS Strasbourg (Cancer Research Institute)
- *Multi-source deep learning models for medical domain.* More broadly, I am studying the topic of medical language understanding.
- **Research Interest**  
Medical language understanding, Knowledge Graphs for enhancing NLP models, and Graph representation learning.

## Language Understanding

# Understanding Language ?



Can I have some ?



**Pragmatics:**  
What does it do?

**Semantics:**  
What does it mean?

**Syntax:**  
What is grammatical?

Natural language utterance



# Expectations versus Reality

## HAL



**David Bowman:** Open the pod bay doors, HAL.

**HAL:** I'm sorry, Dave, I'm afraid I can't do that.

**David:** What are you talking about, HAL?

**HAL:** I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.

## Siri (2011)



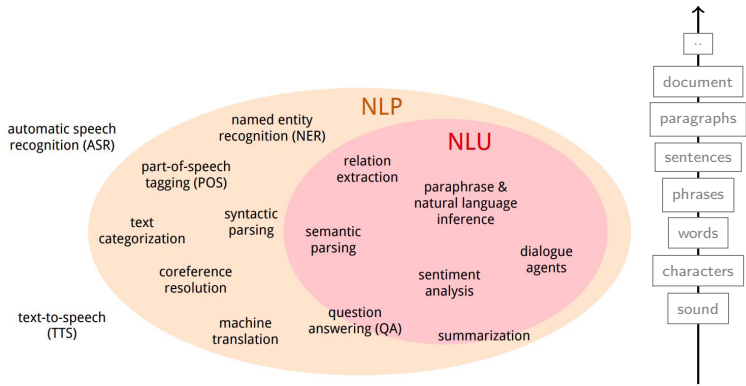
**Colbert:** ... I don't want to search for any write the show!

**Siri:** Searching the Web for "search for ar to write the shuffle."

**Colbert:** ... For the love of God, the came me something?

**Siri:** What kind of place are you looking fi stores or churches?

# Breaking down Language Understanding



## Medical Language Understanding

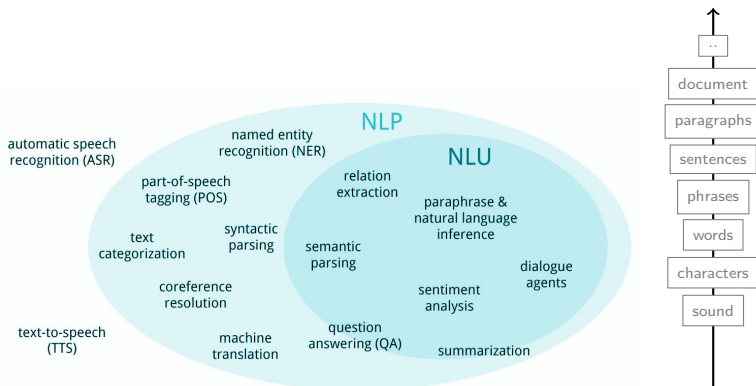
# Medical language is a *sublanguage*

- A *sublanguage* is a technical language that is used by the various actors in the technical field to pass specific messages. A technical language presents **some characteristics that differentiate it from the general language**. It is **easier to build linguistic tools for sublanguages** than for general language. [Spyns, 1996]

# Medical language is a *sublanguage*

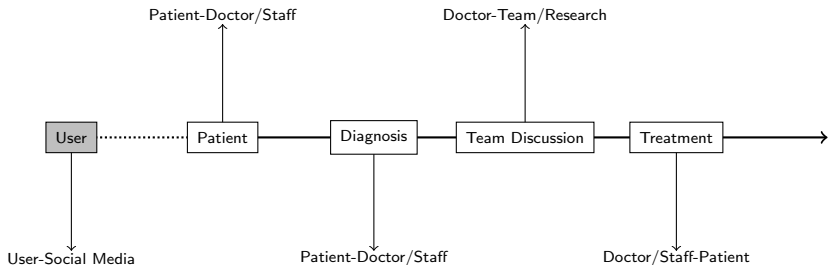
- A *sublanguage* is a technical language that is used by the various actors in the technical field to pass specific messages. A technical language presents **some characteristics that differentiate it from the general language**. It is **easier to build linguistic tools for sublanguages** than for general language. [Spyns, 1996]
- Much of the available clinical data are in narrative form as a result of transcription of dictations, direct entry by providers, or use of speech recognition applications. This free-text form is convenient to express concepts and events, but is **difficult** for searching, summarization, decision-support, or statistical analysis. [Meystre et al., 2008]

# Breaking down Medical Language Understanding



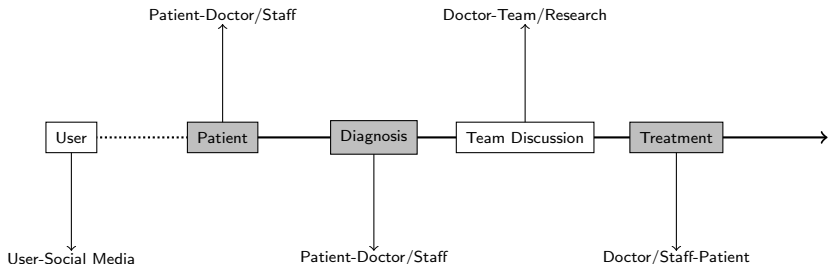
**Challenges** : variations, ambiguity, uncertainty, polysemy, vagueness, world knowledge (and many more)

# Different data sources



- **Social Media**[SOC]: Twitter, Reddit, FB, review-forums, etc.

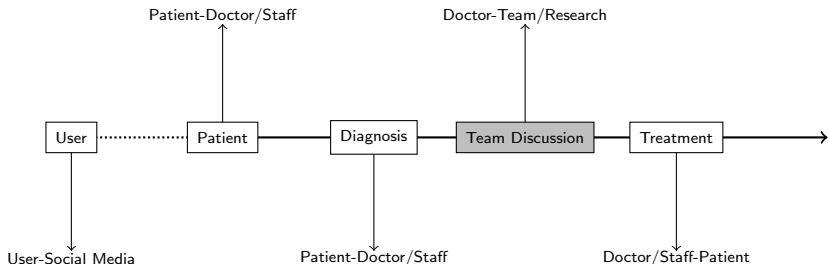
# Different data sources



- **Social Media**[SOC]: Twitter, Reddit, FB, review-forums, etc.
- **Clinical data**[CLIN]: Clinical notes, clinical diagnosis data, etc.

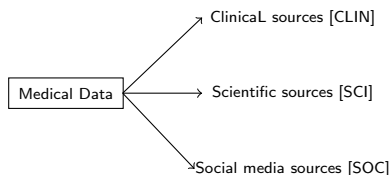


# Different data sources



- **Social Media**[SOC]: Twitter, Reddit, FB, review-forums, etc.
- **Clinical data**[CLIN]: Clinical notes, clinical diagnosis data, etc.
- **Scientific data**[SCI]: Research articles, citation graphs, etc.

# General Formalization



## Overall data view

For any source  $D \in \{SOC, CLIN, SCI\}$ , we study a collection of data points  $\{(x_i, y_i) : i=1 \dots N\}$  where  $x_i \in X$  which denotes the numerical representation of any raw data (for eg. tweet, clinical note, article),  $y_i \in Y$  which correspond to any NLP/NLU task (for eg. disease mention detection, mortality prediction, clustering) based labels, and  $N$  refers to number of data points.

# Temporal characteristics of medical data

## Overall data view

For any source  $D \in \{\text{SOC, CLIN, SCI}\}$ , we study a collection of data points  $\{(\mathbf{x}_i, y_i) : i=1 \dots N\}$  where  $\mathbf{x}_i \in X$  which denotes the numerical representation of any raw data (for eg. tweet, clinical note, article),  $y_i \in Y$  which correspond to any NLP/NLU task (for eg. disease mention detection, mortality prediction, clustering) based labels, and  $N$  refers to number of data points.

Note :

- $\mathbf{x}_i$  can also have temporal aspect i.e.  $\mathbf{x}_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,T_i}\}$ , where  $T_i$  is the number of timestamps associated with  $i$ th data point.

# Temporal characteristics of medical data

## Overall data view

For any source  $D \in \{\text{SOC, CLIN, SCI}\}$ , we study a collection of data points  $\{(\mathbf{x}_i, y_i) : i=1 \dots N\}$  where  $\mathbf{x}_i \in X$  which denotes the numerical representation of any raw data (for eg. tweet, clinical note, article),  $y_i \in Y$  which correspond to any NLP/NLU task (for eg. disease mention detection, mortality prediction, clustering) based labels, and  $N$  refers to number of data points.

Note :

- $\mathbf{x}_i$  can also have temporal aspect i.e.  $\mathbf{x}_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,T_i}\}$ , where  $T_i$  is the number of timestamps associated with  $i$ th data point.
- For eg. series of tweets by a user or series of clinical notes of a patient or series of scientific articles published about a topic.

# Temporal characteristics of medical data

## Overall data view

For any source  $D \in \{\text{SOC, CLIN, SCI}\}$ , we study a collection of data points  $\{(\mathbf{x}_i, y_i) : i=1 \dots N\}$  where  $\mathbf{x}_i \in X$  which denotes the numerical representation of any raw data (for eg. tweet, clinical note, article),  $y_i \in Y$  which correspond to any NLP/NLU task (for eg. disease mention detection, mortality prediction, clustering) based labels, and  $N$  refers to number of data points.

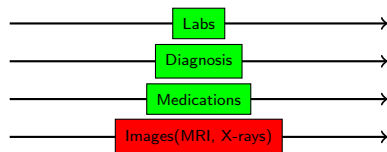
Note :

- $\mathbf{x}_i$  can also have temporal aspect i.e.  $\mathbf{x}_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,T_i}\}$ , where  $T_i$  is the number of timestamps associated with  $i$ th data point.
- For eg. series of tweets by a user or series of clinical notes of a patient or series of scientific articles published about a topic.
- These data are classified as temporal (or longitudinal) data.

## Longitudinal Medical Language Understanding

# Case Study : Clinical data setting

## Longitudinal Variables



## Static Variables

Demographics,  
Gender,  
Genes, socio-  
economics

In this study, we focus on clinical datasets: MIMIC-III [Johnson et al., 2016], Physionet 2012 [Goldberger et al., 2000], & Physionet 2019 [Reyna et al., 2019].

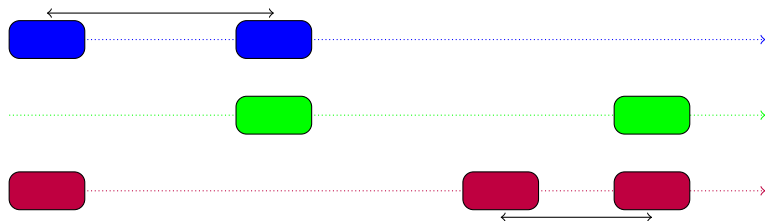
# Challenges in *longitudinal* Medical Language Understanding [Cascarano et al., 2023]

- **Dealing with missing values** : There are often missing measurements or dropouts in longitudinal data cohorts, while the time intervals between one measurement and another are not necessarily evenly distributed. These facts hamper an off-the-shelf application of time-series algorithms built on the assumptions of complete samples.
- *Dealing with patient trajectories* : Longitudinal data trajectories may be highly complex and non-linear (e.g., large variations between individuals)
- *Dealing with uncertainty* : The repeated measures can be subject to very different, and sometimes hard to estimate, uncertainties, which may also vary with time—from instrument inaccuracy to the specificity of the individual (e.g., different pain thresholds).



# Missing values in data

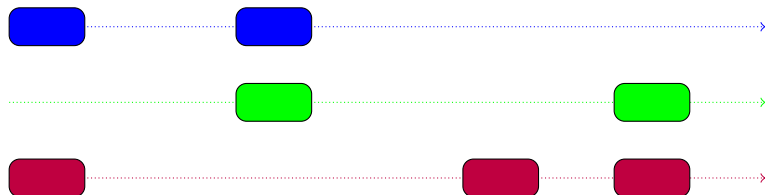
What and how?



- Such data is resulted due to complex generative processes in areas eg. user social-media activity, e-commerce transactions, industrial factories, clinical data, etc.
- These are multivariate time series recorded at inconsistent or non-uniform time intervals aka irregularly sampled time series or ISTS.

# Missing values in data

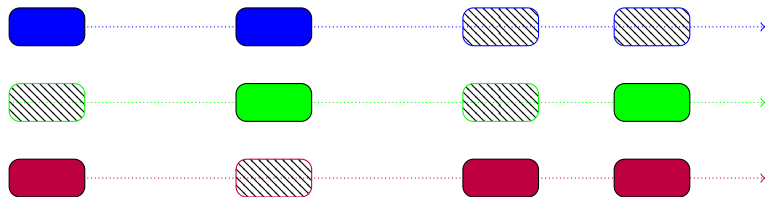
Missingness categorization [Rubin, 1976]



- Random missingness (missing at random [MAR] & missing completely at random [MCAR]) eg. fault in sensors.
- Non-random missingness (missing not at random [MNAR]) eg. subject to clinician.

# Dealing with missing values in data

Is Imputation all we need?

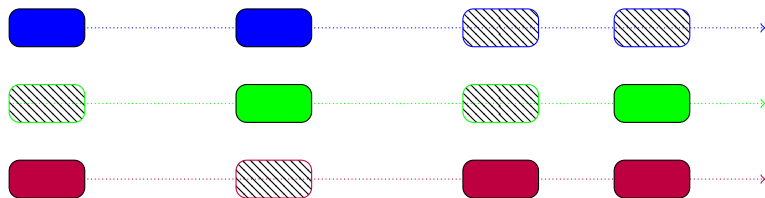


## Imputation based methods

- Many methods handle ISTS by filling missingness via imputation **converting ISTS to regularly sampled time series**, assuming an underlying missing mechanism .

# Dealing with missing values in data

Is Imputation all we need?

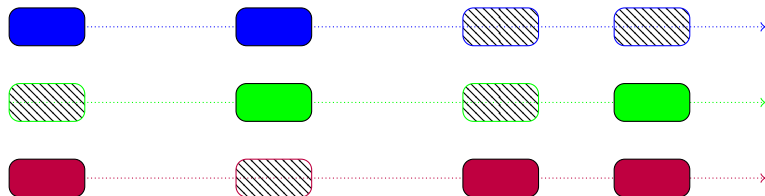


## Imputation based methods

- Many methods handle ISTS by filling missingness via imputation **converting ISTS to regularly sampled time series**, assuming an underlying missing mechanism .
- Methods such as using global mean value (GRU-mean), the last measured value (GRU-forward), interpolation (IP-Nets[Shukla and Marlin, 2019]), or learnable decay on the global mean (GRU-D[Che et al., 2016]).

# Dealing with missing values in data

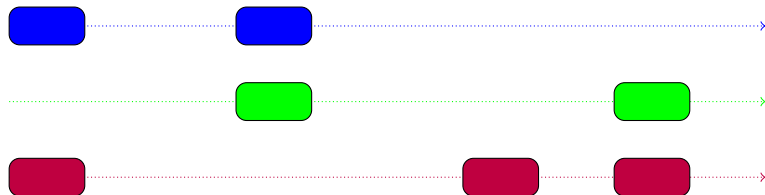
Is Imputation all we need?



## Imputation based methods

- Many methods handle ISTS by filling missingness via imputation **converting ISTS to regularly sampled time series**, assuming an underlying missing mechanism .
- Methods such as using global mean value (GRU-mean), the last measured value (GRU-forward), interpolation (IP-Nets[Shukla and Marlin, 2019]), or learnable decay on the global mean (GRU-D[Che et al., 2016]).
- MAR & MCAR : Works well ✓ MNAR : Caution !

# Dealing with missing values in data

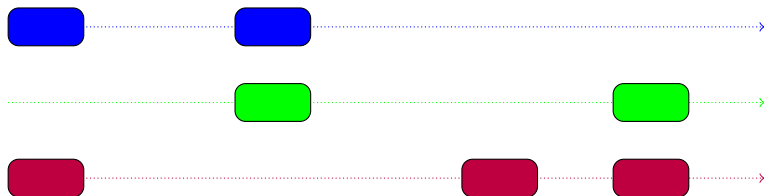


Why imputation in MNAR is **not** a good practise?

- Informative missingness [Rubin, 1976]
- Assumption about underlying process
- Exposure to biasness

# Dealing with missing values in data

What if we don't impute?

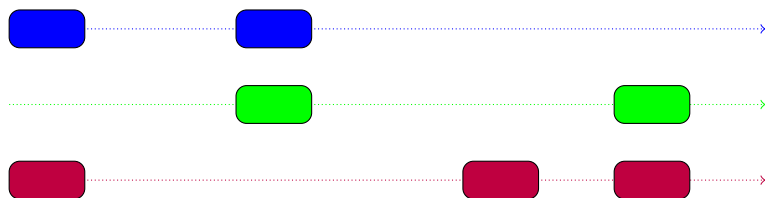


## Non-Imputation based methods

- Recent approaches have also explored **learning directly from the ISTS data** without any form of imputation.

# Dealing with missing values in data

What if we don't impute?



## Non-Imputation based methods

- Recent approaches have also explored **learning directly from the ISTS data** without any form of imputation.
- Use of time encoding embedding and self-attention (Transformers [Vaswani et al., 2017]), set-based data representations (SeFT [Horn et al., 2020]), sensor dependency graph via graph neural networks (RAINDROP [Zhang et al., 2021])



# Dealing with missing values in data

To impute or not is the question!

## Limitations of Imputation based methods

- Assumption of missingness is at random or an underlying missing mechanism which leads to unwanted bias and potential distribution shift.
- We argue that learning **the imputation task is challenging because of the underlying missing mechanism** and is not required for the downstream task.

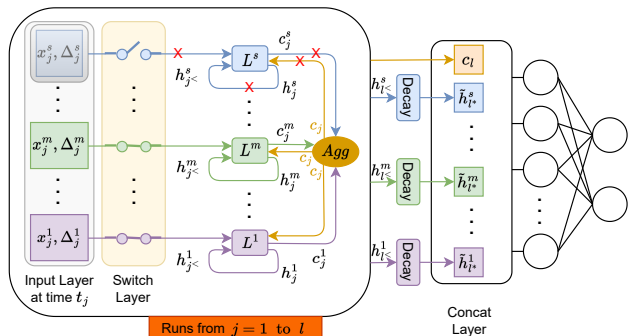
## Limitations of Non-Imputation based methods

- Permutation-invariant nature (Transformers); Order-invariant nature of set representation (SeFT) and, sensor dependency graph does not take into account the irregularity information (RAINDROP)
- MNAR datasets contain informative missingness [Rubin, 1976]. Therefore, specialized methods are required to handle such missingness **by exploiting the irregularity information.**

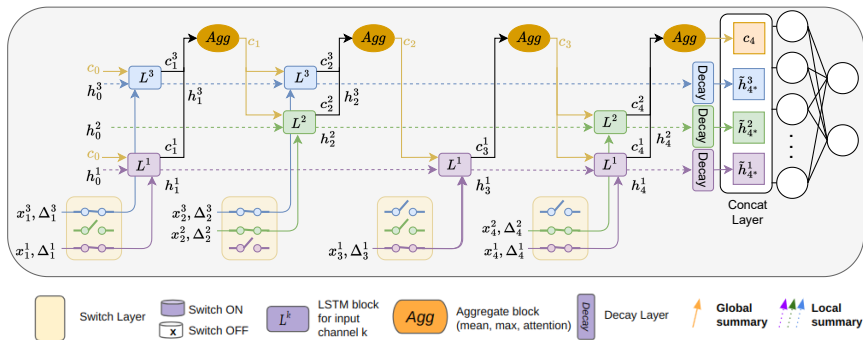
# SLAN (Switch LSTM Aggregate Network)

We propose a model that dynamically changes its architecture depending on the measured sensors at any time point.

- ★ SLAN contains a pack of LSTMs and a simple switch layer.
- ✓ eliminates the need for missing value imputation.
- ✓ No risk of bias or distribution shift.



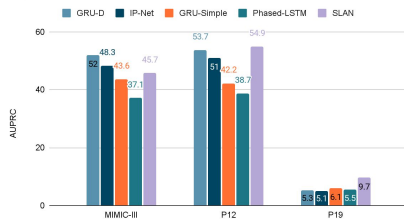
# Unrolled SLAN



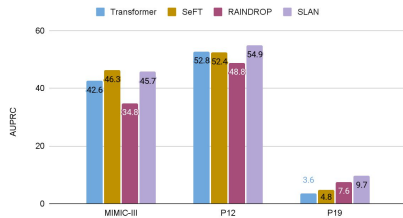
[Demo]

# Does SLAN work? (°\_°)

Imputation Based Methods



Non-Imputation based methods

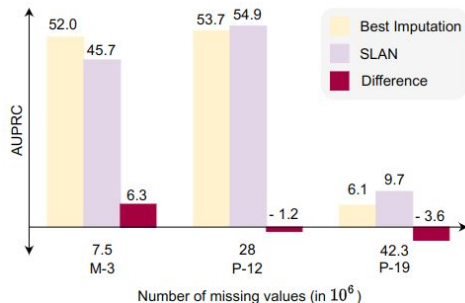


Downstream tasks: In-hospital mortality prediction (MIMIC, P12); Early Sepsis Prediction (P19)

For MIMIC, SLAN is overall 4th best.

For Physionet data (P12 and P19), SLAN outperforms all baselines.

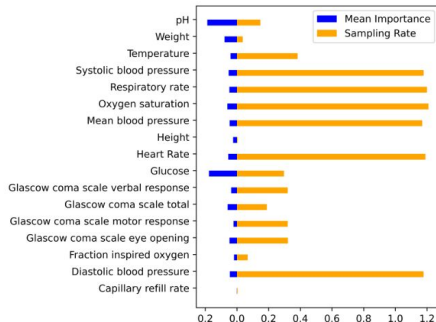
# An observation (\*\_\*)



As the number of missing values increases (from MIMIC-III to P-19), the performance of the best imputation model compared to SLAN **deteriorates**.

# Interpretation with SLAN (O\_O)

- Features with higher sampling rates are **oxygen saturation, respiratory rate, heart rate and temperature**. Whereas the most important features are **pH, glucose, weight, oxygen saturation and glasgow coma scale total**.
- Thus, **frequently measured sensors are not the most important feature**. Even though pH has the fifth-lowest sampling rate, it is the most important feature in providing inference.

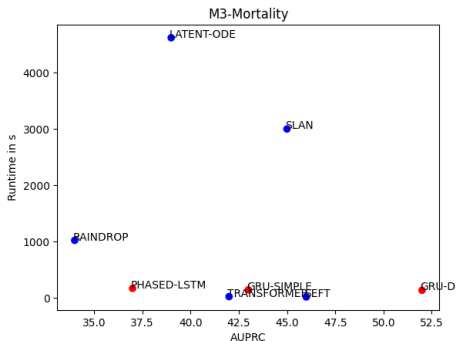


Note:

**Sampling rate:** the number of measurements per hour of a particular sensor.

**Feature Importance via Global summary:** Mean of the attention weights of each sensor across all the time steps.

# Limitations to SLAN (--)



- Algorithmic Complexity :  $O(npq)$  where  $n$  is #Instances,  $p$  is #Sensors and  $q$  is #Observations.
- Scalability to sensors : The time complexity is linearly dependent on the number of sensors. Therefore, SLAN may not be scalable to applications with many sensors.

- 1 SLAN is a simple switch-based adaptable architecture for MNAR (or any type of ISTS) data eliminating the need for missing value imputation.
- 2 There seems to be a tradeoff between MNAR data handling ( $\uparrow$ ) and model performance ( $\downarrow$ ).
- 3 To impute or not? still remains a question.  
[Ma and Zhang, 2021, Berrevoets et al., 2023]

Preprint : <https://arxiv.org/abs/2309.08698>

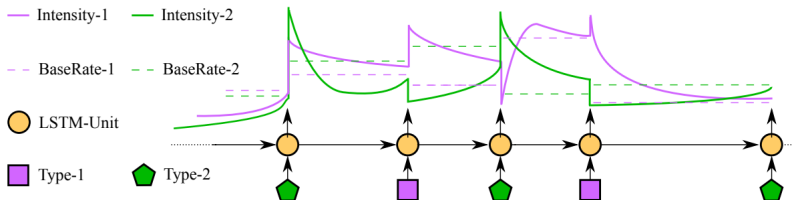
Github : <https://github.com/Rohit102497/SLAN>



# From SLAN to beyond..

Two line of thoughts:

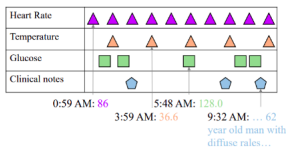
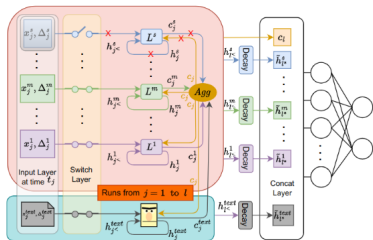
- 1 More sophisticated modeling technique that is suited for MNAR data, via neural point process modelling. [Mei and Eisner, 2017]



# From SLAN to beyond..

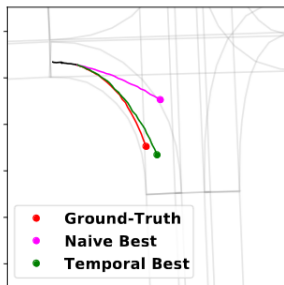
Two line of thoughts:

- 1 More sophisticated modeling technique that is suited for MNAR data, via neural point process modelling. [Mei and Eisner, 2017]
- 2 Including clinical notes along with physiological data for multimodal learning. [Zhang et al., 2022]

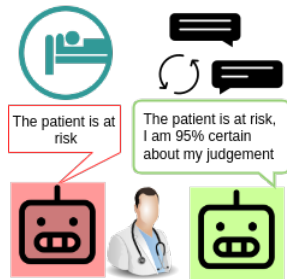


# More Medical Language Understanding

- Modeling complex patient trajectories [Miotto et al., 2016, Allam et al., 2021]





- Uncertainty in Medical domain [Peng et al., 2019, Ulmer et al., 2022]



# References I

-  Agarwal, R., Sinha, A., Prasad, D. K., Clausel, M., Horsch, A., Constant, M., and Coubez, X. (2023).  
Modelling irregularly sampled time series without imputation.  
*arXiv preprint arXiv:2309.08698*.
-  Allam, A., Feuerriegel, S., Rebhan, M., and Krauthammer, M. (2021).  
Analyzing patient trajectories with artificial intelligence.  
*Journal of medical internet research*, 23(12):e29812.
-  Berrevoets, J., Imrie, F., Kyono, T., Jordon, J., and van der Schaar, M. (2023).  
To impute or not to impute? missing data in treatment effect estimation.  
*In International Conference on Artificial Intelligence and Statistics*, pages 3568–3590. PMLR.

-  Cascarano, A., Mur-Petit, J., Hernández-González, J., Camacho, M., de Toro Eadie, N., Gkontra, P., Chadeau-Hyam, M., Vitrià, J., and Lekadir, K. (2023).  
Machine and deep learning for longitudinal biomedical data: a review of methods and applications.  
*Artificial Intelligence Review*, pages 1–61.
-  Che, Z., Purushotham, S., Cho, K., Sontag, D. A., and Liu, Y. (2016).  
Recurrent neural networks for multivariate time series with missing values.  
*Scientific Reports*, 8.

-  Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000).




Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220.




-  Horn, M., Moor, M., Bock, C., Rieck, B., and Borgwardt, K. (2020).

Set functions for time series.




In *International Conference on Machine Learning*, pages 4353–4363. PMLR.




## References IV



-  Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L.-w. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., and Mark, R. G. (2016). Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9.
-  Ma, C. and Zhang, C. (2021). Identifiable generative models for missing not at random data imputation. *Advances in Neural Information Processing Systems*, 34:27645–27658.
-  Mei, H. and Eisner, J. M. (2017). The neural hawkes process: A neurally self-modulating multivariate point process. *Advances in neural information processing systems*, 30.

-  Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., and Hurdle, J. F. (2008).  
Extracting information from textual documents in the electronic health record: a review of recent research.  
*Yearbook of medical informatics*, 17(01):128–144.
-  Miotto, R., Li, L., Kidd, B. A., and Dudley, J. T. (2016).  
Deep patient: an unsupervised representation to predict the future of patients from the electronic health records.  
*Scientific reports*, 6(1):1–10.
-  Peng, Y., Yan, S., and Lu, Z. (2019).  
Transfer learning in biomedical natural language processing: An evaluation of bert and elmo on ten benchmarking datasets.  
*BioNLP 2019*, page 58.



-  Reyna, M. A., Josef, C., Seyedi, S., Jeter, R., Shashikumar, S. P., Westover, M. B., Sharma, A., Nemati, S., and Clifford, G. D. (2019).  
Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019.  
*In 2019 Computing in Cardiology (CinC)*, pages Page–1. IEEE.
-  Rubin, D. B. (1976).  
Inference and missing data.  
*Biometrika*, 63(3):581–592.
-  Shukla, S. N. and Marlin, B. M. (2019).  
Interpolation-prediction networks for irregularly sampled time series.  
*ArXiv*, abs/1909.07782.

-  Spyns, P. (1996).  
Natural language processing in medicine: an overview.  
*Methods of information in medicine*, 35(04/05):285–301.
-  Ulmer, D., Frelsen, J., and Hardmeier, C. (2022).  
Exploring predictive uncertainty and calibration in NLP: A  
study on the impact of method & data scarcity.  
In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 2707–2735, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
-  Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017).  
Attention is all you need.  
In *NIPS*.

-  Zhang, X., LI, S., Chen, Z., Yan, X., and Petzold, L. (2022). Improving medical predictions by irregular multimodal electronic health records modeling. *ArXiv*, [abs/2210.12156](https://arxiv.org/abs/2210.12156).
-  Zhang, X., Zeman, M., Tsiligkaridis, T., and Zitnik, M. (2021). Graph-guided network for irregularly sampled multivariate time series. *ArXiv*, [abs/2110.05357](https://arxiv.org/abs/2110.05357).

# Appendix-I

## ISTS baselines

- Via learnable decay on the global mean and the last measured value. (GRU-D)
- Temporal discretization and performing interpolation (IP-Nets)
- Using missing value indicator along with RNN-based filling (GRU-Simple, Phased-LSTM)
- replace the positional encoding with an encoding of time and model sequences using self-attention and concatenate it with the input representation. (Transformer)
- treating time series as an unordered set of measurements (SeFT)
- By GNN to learn a sensor dependency graph and leveraging inter-sensor dependency to train latent embeddings. (RAINDROP)

# Appendix-II

## Limitations of ISTS baselines

- assuming an underlying missing mechanism (GRU-D) or a predefined nonlinear form assuming that missingness is at random, thus inducing bias. (IP-Nets)
- lead to a potential distribution shift. (GRU-Simple, Phased-LSTM)
- the permutation-invariant nature of self-attention, which can be problematic in capturing dependencies within each time series (Transformer)
- the order-invariant nature of set representation fails to capture the irregularity information, which is order-variant and increases with time (SeFT)
- does not exploit the irregularity information of the sensors. (RAINDROP)

# Appendix-III

## Results - SLAN experiments

Table 2: Comparison of various methods on M-3, P-12 and P-19 datasets.  $\text{Imp}$  and  $\text{No-Imp}$  represent imputation models and non-imputation models, respectively. The **best** and 2<sup>nd</sup> best performance is represented by **bold** and UNDERLINE, respectively. The metric is reported as the mean  $\pm$  standard deviation of three runs with different seeds.

Model	MIMIC-III		Physionet 2012		Physionet 2019 (ESP)				
	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	B <sub>Acc</sub>	U <sub>norm</sub>	
$\text{Imp}$	GRU-D	<b>52.0<math>\pm</math>0.8</b>	<b>85.7<math>\pm</math>0.2</b>	<u>53.7<math>\pm</math>0.9</u>	<b>86.3<math>\pm</math>0.3</b>	5.3 $\pm$ 0.4	67.4 $\pm$ 1.2	57.4 $\pm$ 0.2	12.6 $\pm$ 1.1
	IP-Nets	48.3 $\pm$ 0.4	83.2 $\pm$ 0.5	51.0 $\pm$ 0.6	<u>86.0<math>\pm</math>0.2</u>	5.1 $\pm$ 0.8	74.2 $\pm$ 1.2	63.8 $\pm$ 0.9	-11.9 $\pm$ 4.0
	GRU-SIMPLE	43.6 $\pm$ 0.4	82.8 $\pm$ 0.0	42.2 $\pm$ 0.6	80.8 $\pm$ 1.1	6.1 $\pm$ 0.7	78.1 $\pm$ 1.5	<u>71.0<math>\pm</math>1.4</u>	26.9 $\pm$ 4.1
	Phased-LSTM	37.1 $\pm$ 0.5	80.3 $\pm$ 0.4	38.7 $\pm$ 1.5	79.0 $\pm$ 1.0	5.5 $\pm$ 0.9	75.4 $\pm$ 1.3	67.5 $\pm$ 1.7	20.2 $\pm$ 3.2
$\text{No-Imp}$	TRANSFORMER	42.6 $\pm$ 1.0	82.1 $\pm$ 0.3	52.8 $\pm$ 2.2	<b>86.3<math>\pm</math>0.8</b>	3.6 $\pm$ 0.9	65.8 $\pm$ 3.7	53.6 $\pm$ 1.7	-43.9 $\pm$ 10.0
	SeFT	46.3 $\pm$ 0.5	83.9 $\pm$ 0.4	52.4 $\pm$ 1.1	85.1 $\pm$ 0.4	4.8 $\pm$ 0.2	76.8 $\pm$ 0.9	70.9 $\pm$ 0.8	25.6 $\pm$ 1.9
	RAINDROP	34.8 $\pm$ 1.4	79.3 $\pm$ 0.9	48.8 $\pm$ 3.1	84.3 $\pm$ 1.1	<u>7.6<math>\pm</math>0.2</u>	<u>78.1<math>\pm</math>0.4</u>	69.3 $\pm$ 0.8	<b>48.4<math>\pm</math>0.1</b>
	<b>SLAN (Ours)</b>	45.7 $\pm$ 0.9	<u>84.9<math>\pm</math>0.2</u>	<b>54.9<math>\pm</math>0.4</b>	86.2 $\pm$ 0.2	<b>9.7<math>\pm</math>0.6</b>	<b>80.5<math>\pm</math>2.0</b>	<b>71.8<math>\pm</math>2.9</b>	<u>48.1<math>\pm</math>0.3</u>

# Appendix-IV

## Datasets - SLAN experiments

Table 1: Description of the MIMIC-III, Physionet 2012 and Physionet 2019 (early sepsis prediction) datasets. #Instances denotes the number of patient records in the datasets, #Sensors denotes the number of features/sensors in each instance, #Observations denotes the average number of observations recorded in each instance, i.e., the number of time steps, Measured ratio denotes the ratio of the total number of data measured by the total number of data measured in a fully available dataset, #Num-Imputation is the number of imputation or missing values and Imbalance denotes the percentage of instances with a minority class label.

Dataset	#Instances	#Sensors	#Observations(avg.)	Measured ratio (%)	#Num-Imputation	Imbalance (%)
MIMIC-III	21110	17	77.7	32.5	$7.5 \times 10^6$	13.22
Physionet 2012	11988	37	74.9	11.6	$28.0 \times 10^6$	14.24
Physionet 2019 (ESP)	40333	34	38.5	19.8	$42.3 \times 10^6$	1.79

Thank you for your attention !

Aman Sinha  
aman.sinha@univ-lorraine.fr

Website : amansinha09

Github : amansinha09

Twitter : amansinha09