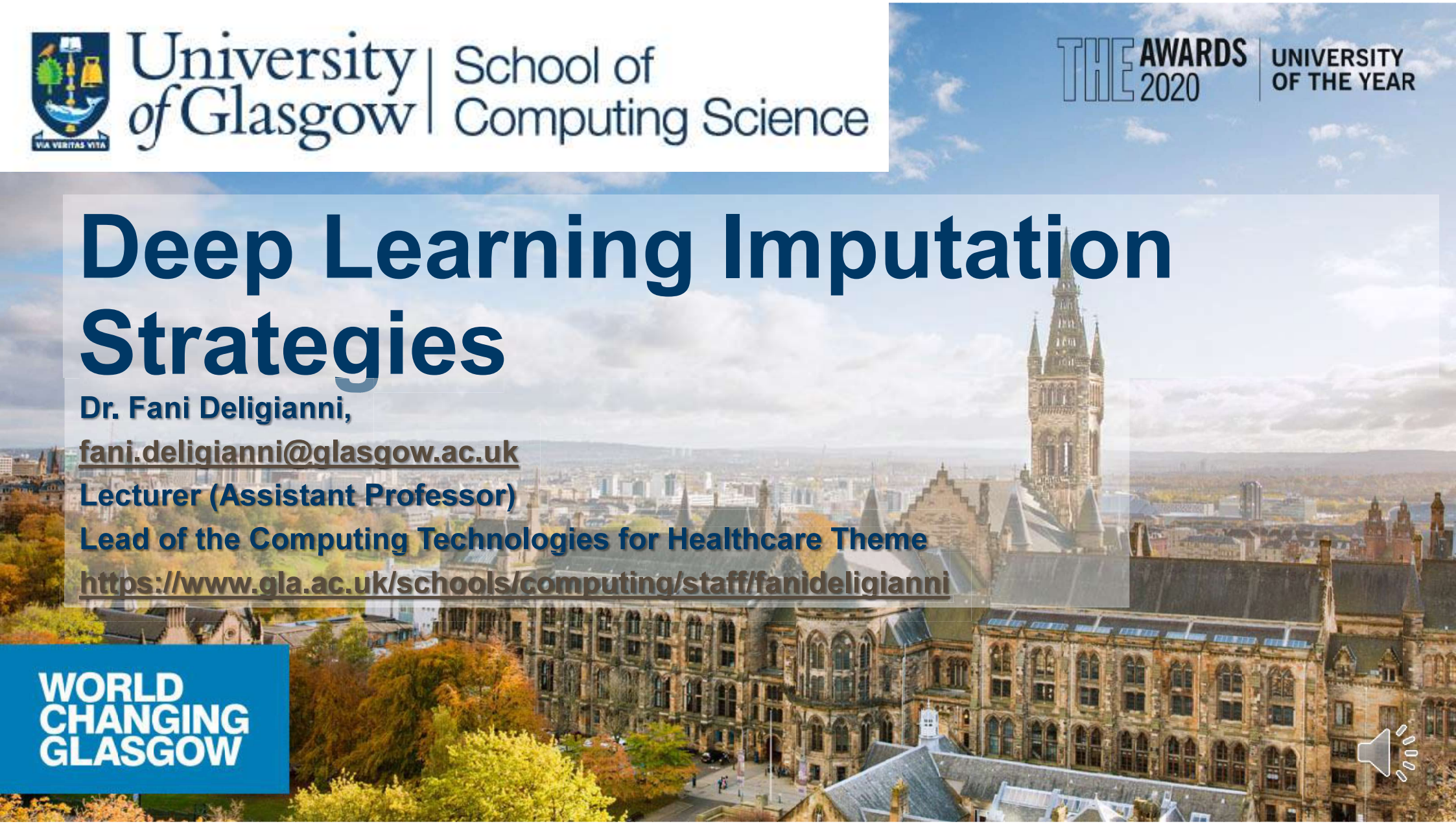
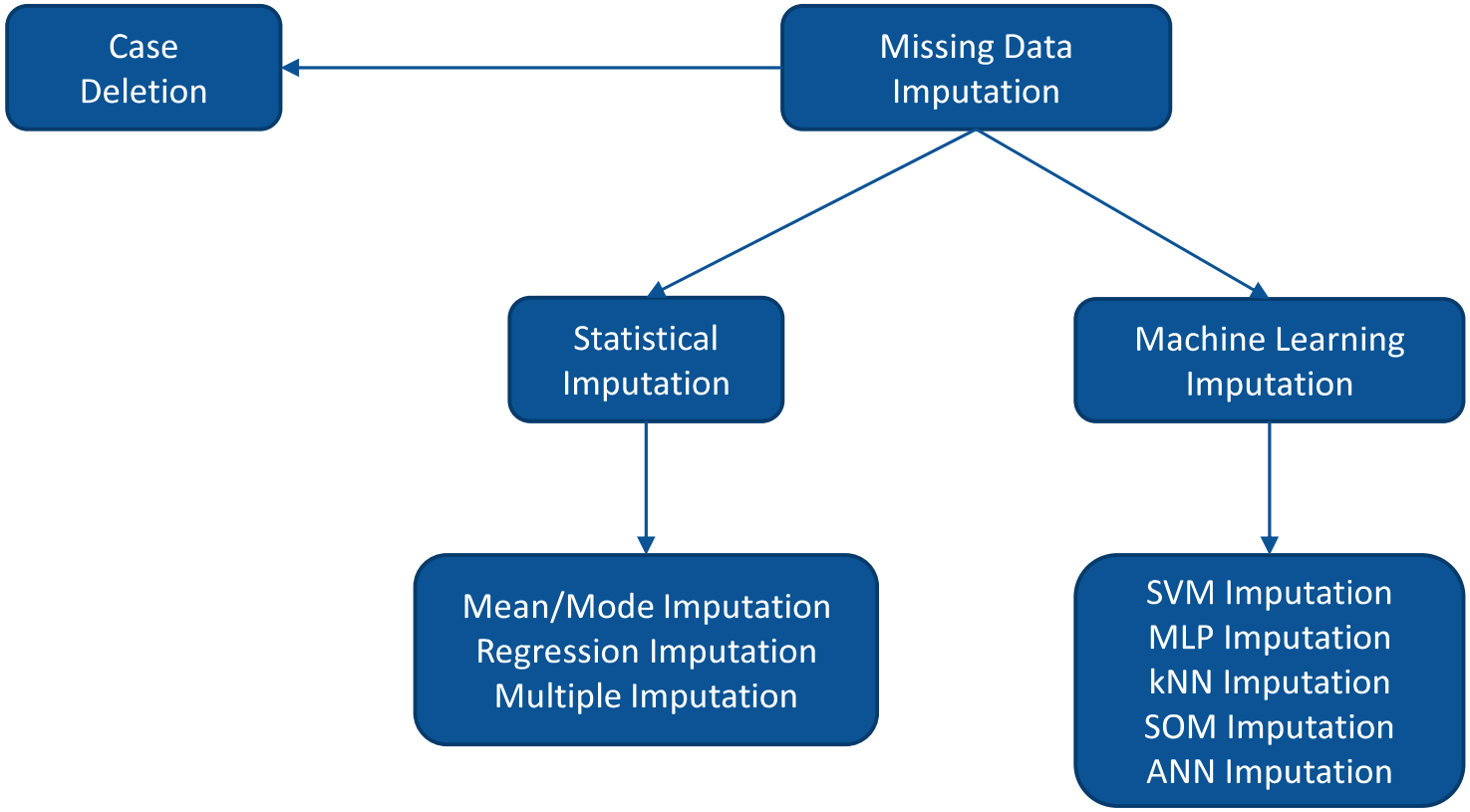


Deep Learning Imputation Strategies

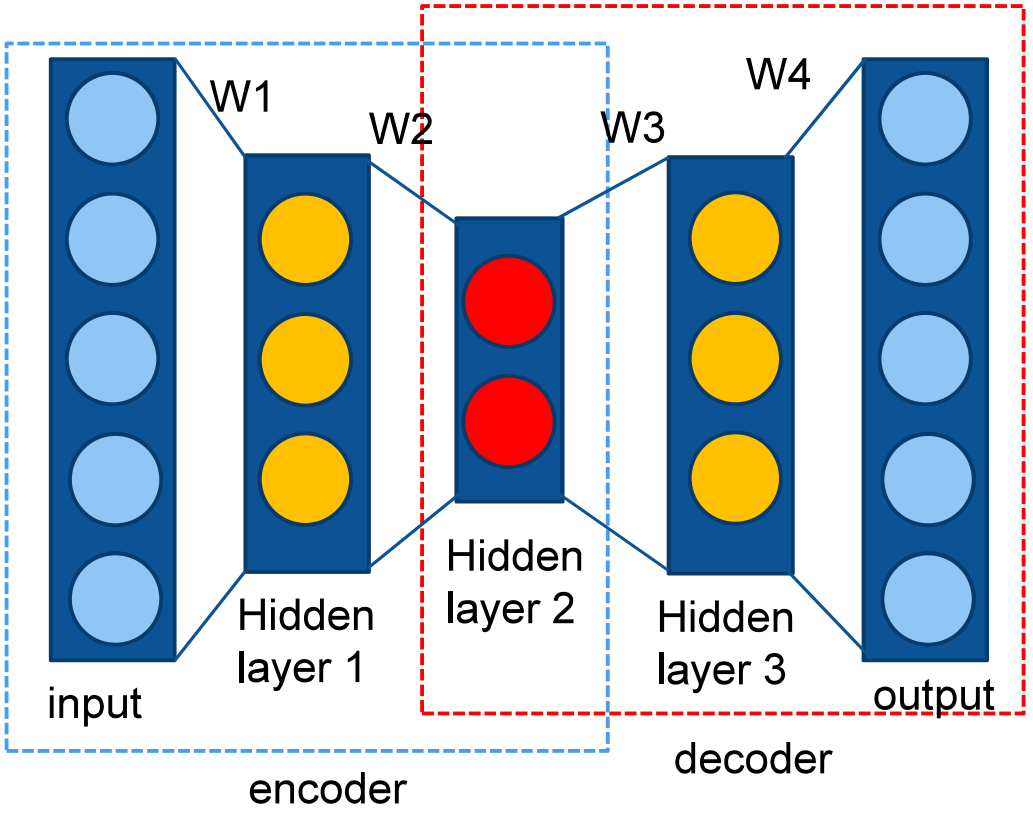
Dr. Fani Deligianni,
fani.deligianni@glasgow.ac.uk
Lecturer (Assistant Professor)
Lead of the Computing Technologies for Healthcare Theme
https://www.gla.ac.uk/schools/computing/staff/fanideligianni

Overview of imputation methods



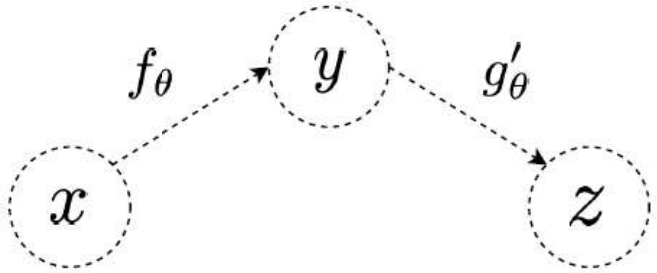
Deep Learning Approaches in Imputation

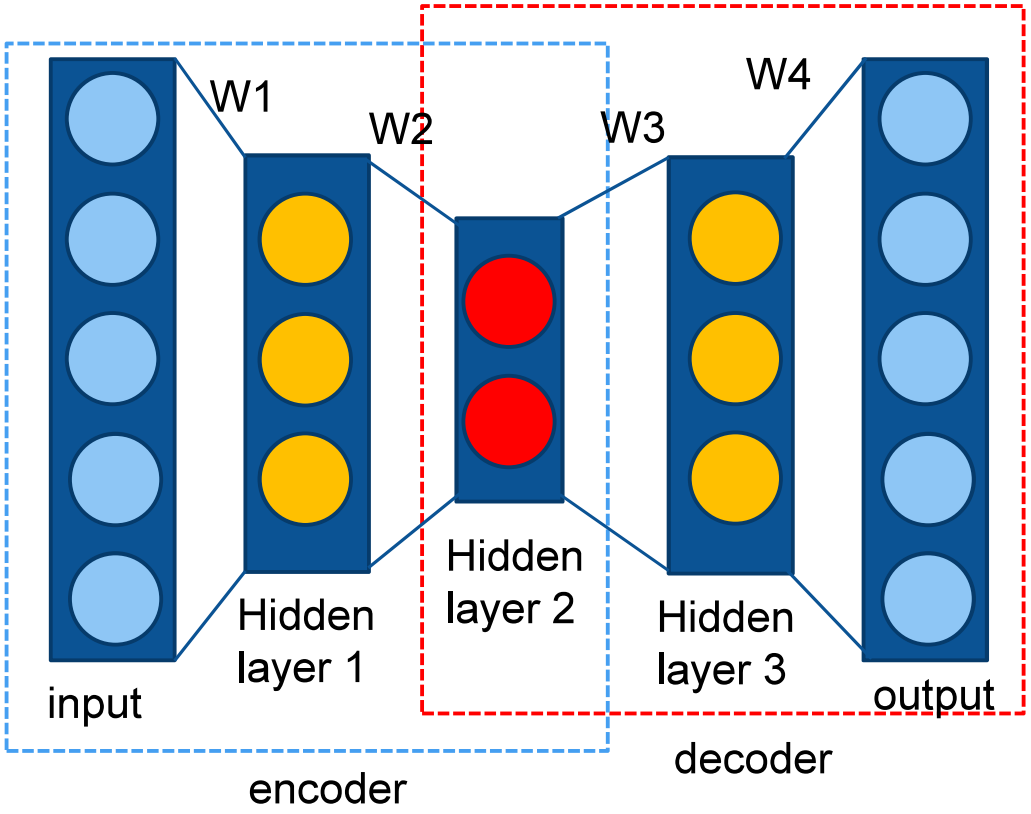
- **Autoencoder**
- Denoising autoencoders
- Variational autoencoders
- Generative adversarial networks



Autoencoders

- **Autoencoder**
- Denoising autoencoders
- Variational autoencoders
- Generative adversarial networks

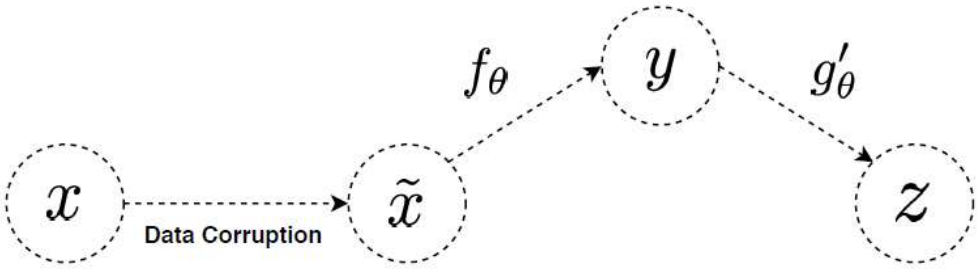


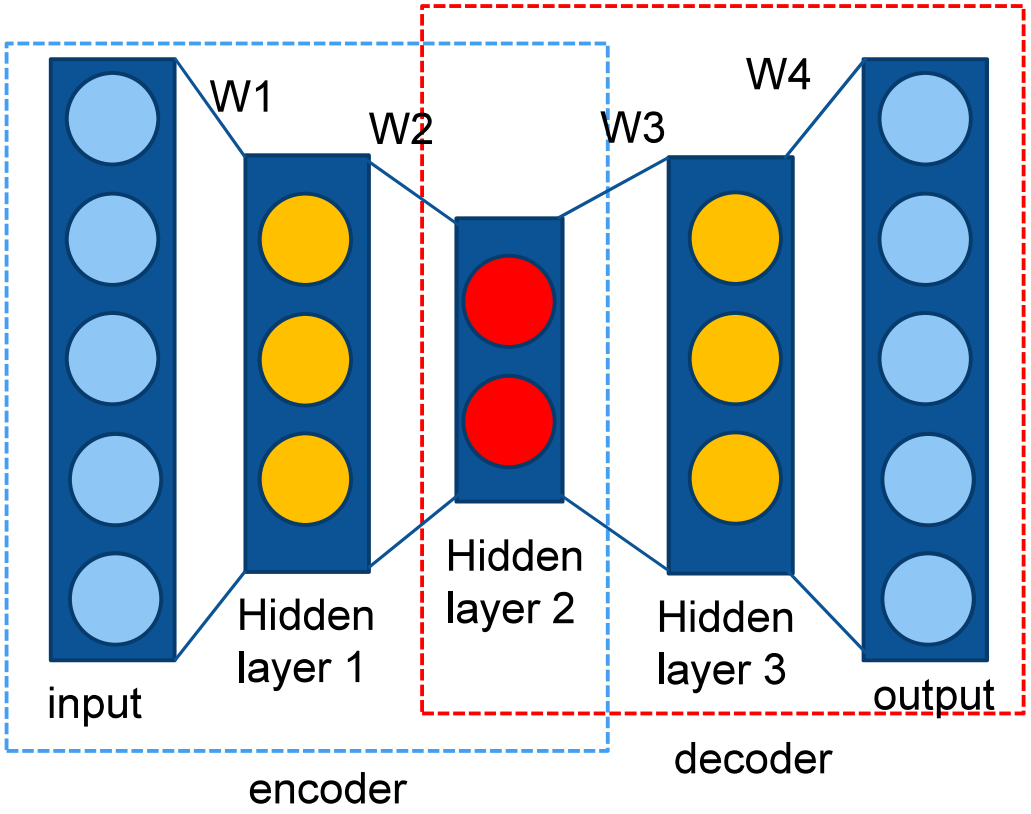


$$f_\theta(\mathbf{x}) = s(\mathbf{x}\mathbf{W}^T + \mathbf{b})$$

Denoising Autoencoders

- Autoencoder
- **Denoising autoencoders**
- Variational autoencoders
- Generative adversarial networks

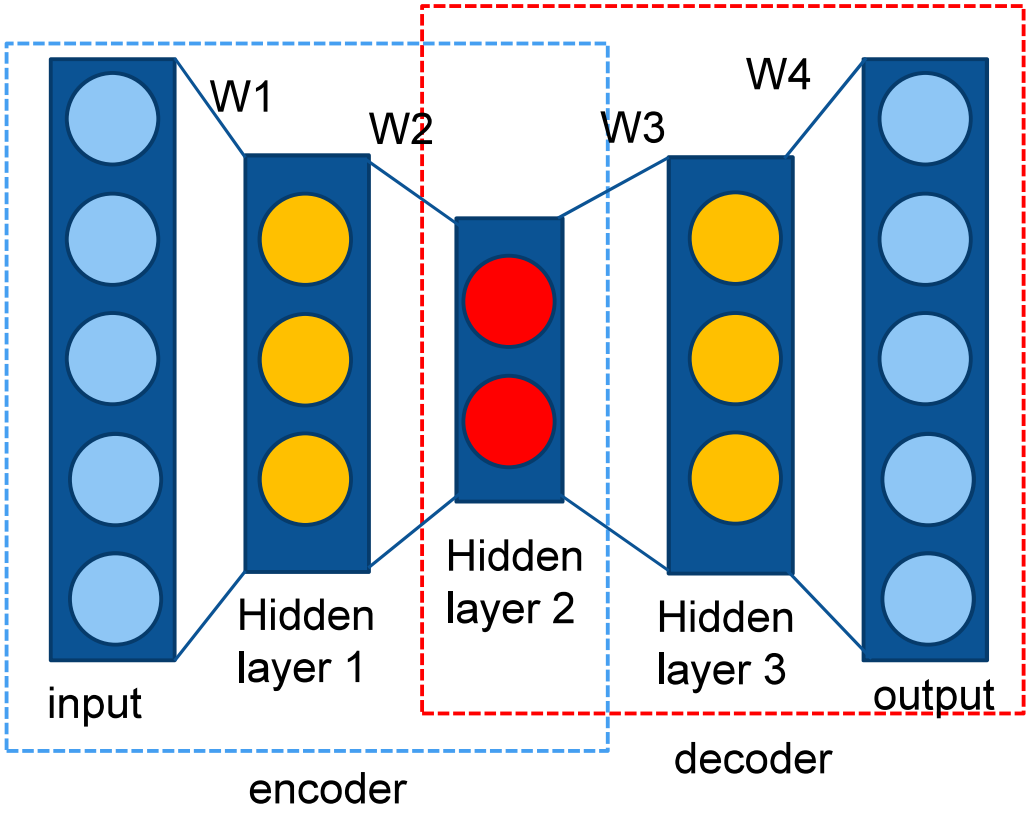




$$f_\theta(\mathbf{x}) = s(\mathbf{x}\mathbf{W}^T + \mathbf{b})$$

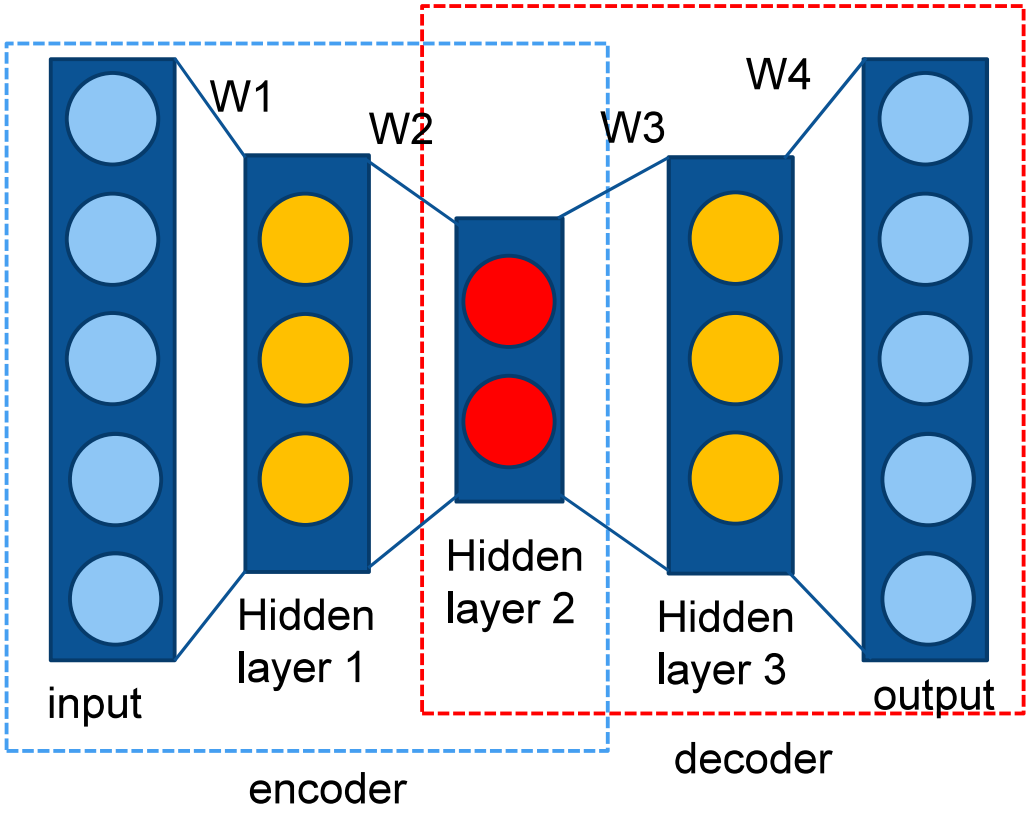
Variational Autoencoders

- Autoencoder
- Denoising autoencoders
- **Variational autoencoders**
- Generative adversarial networks



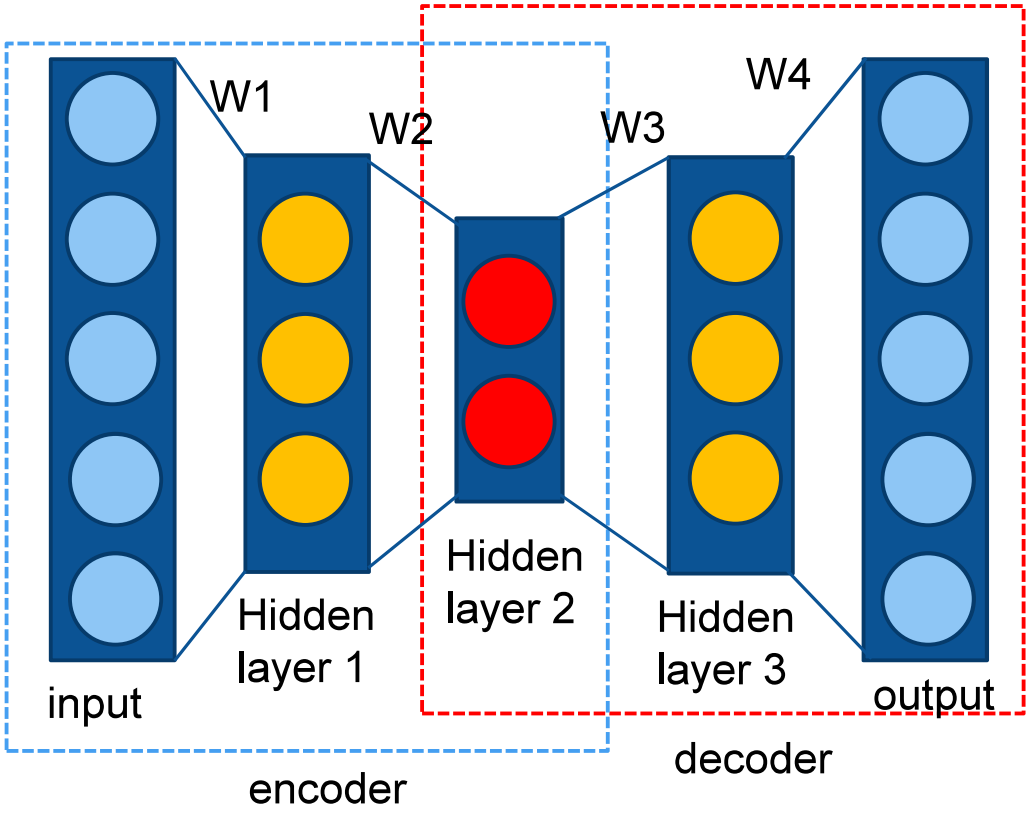
Network Structure and Parameters

- Number of layers
- Overcomplete representations
- Activation functions
- Loss functions



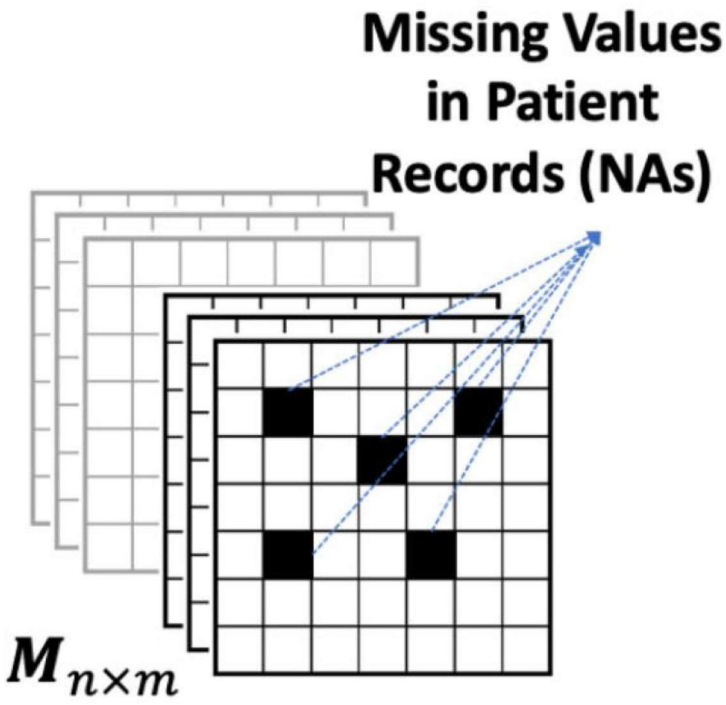
Network Structure and Parameters

- Regularisation
  - L2 regularization
  - Batch normalization
  - Dropout

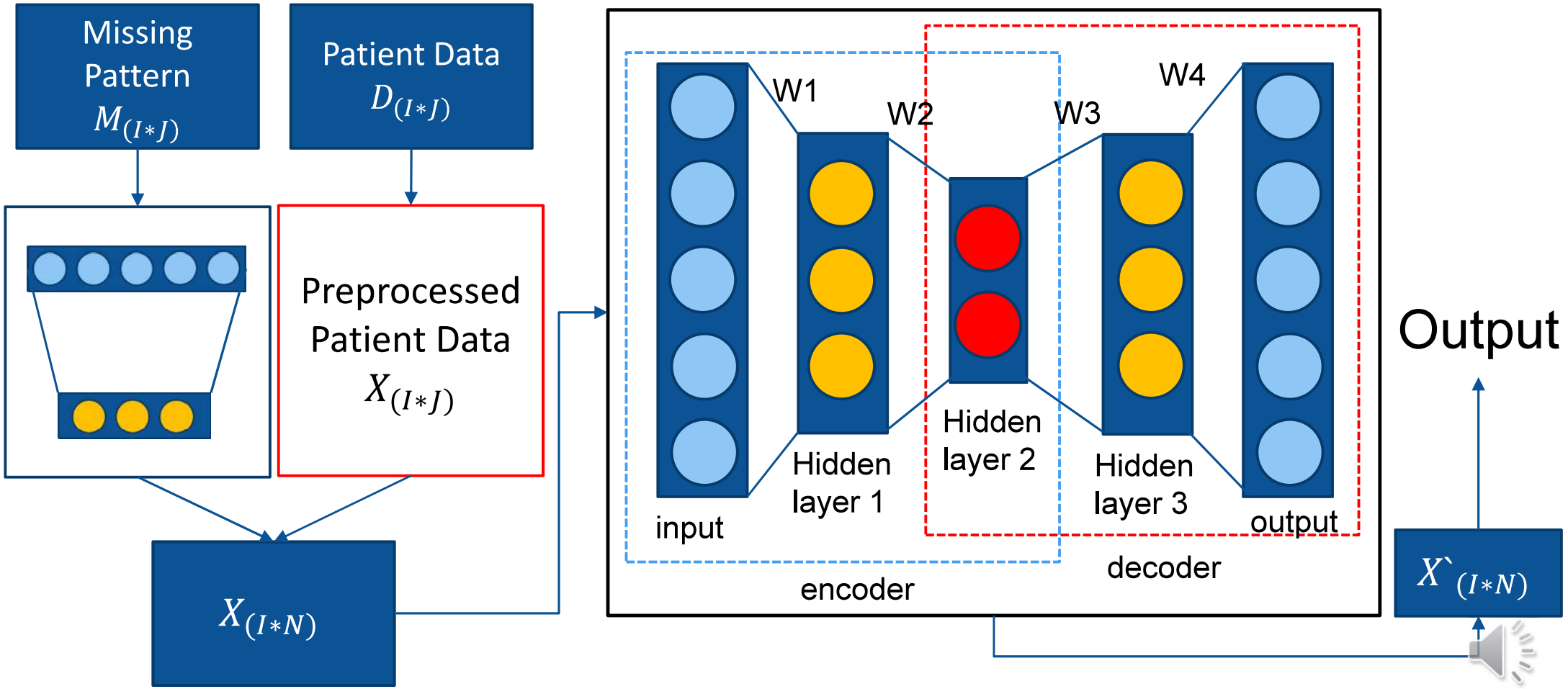


Overall Pipeline – Pre-imputation

- Aggregate patient data with respect to important clinical variables
- Z-score standardization
- One-hot encoding

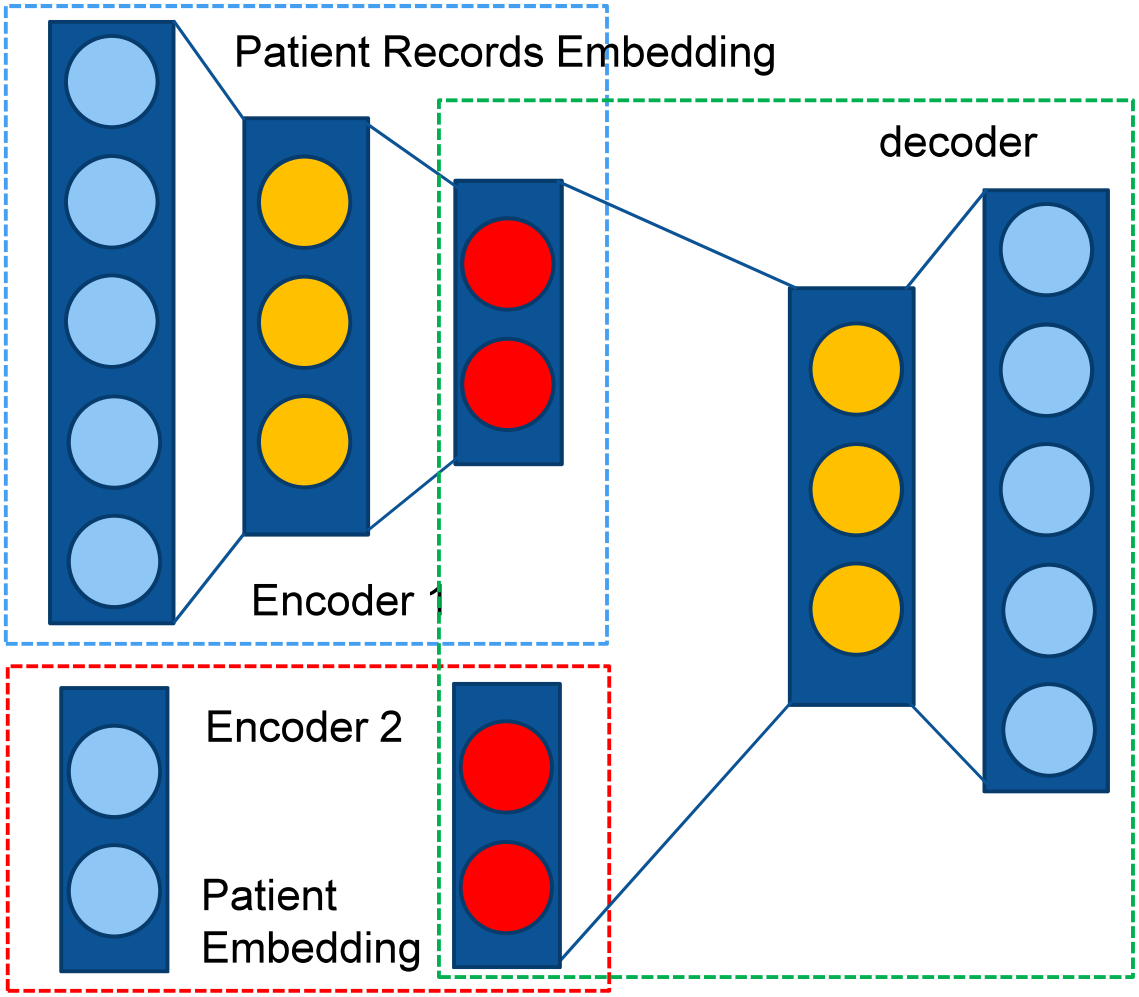


Examples of Autoencoder-based Imputation



Xu et al., Journal of Biomedical Informatics, 2020.

Imputation based on autoencoders



Xu et al., Journal of Biomedical Informatics, 2020.

Missing at Random

- Missing Completely At Random (MCAR)
- Missing At Random (MAR)
- Missing Not At Random (MNAR)

Summary

- Autoencoders and variants neural network structures outperform most statistical methods
- ***Studies do not address the missing data mechanisms***
- Most AE approaches require a pre-imputation step
- The performance of imputation in classification and regression tasks is understudied
- There is a lack of baseline comparisons

References

- Xu et al. A deep learning–based, unsupervised method to impute missing values in electronic health records for improved patient management, Journal of Biomedical Informatics, 2020.
- Pereira et al. Reviewing Autoencoders for Missing Data Imputation: Technical Trends, Applications and Outcomes, Journal of Artificial Intelligence Research, 2020.
- Carreras et al. ‘Missing not at random in end-of-life care studies: multiple imputation and sensitivity analysis on data from the ACTION study’, BMC Medical Research Methodology, 2021