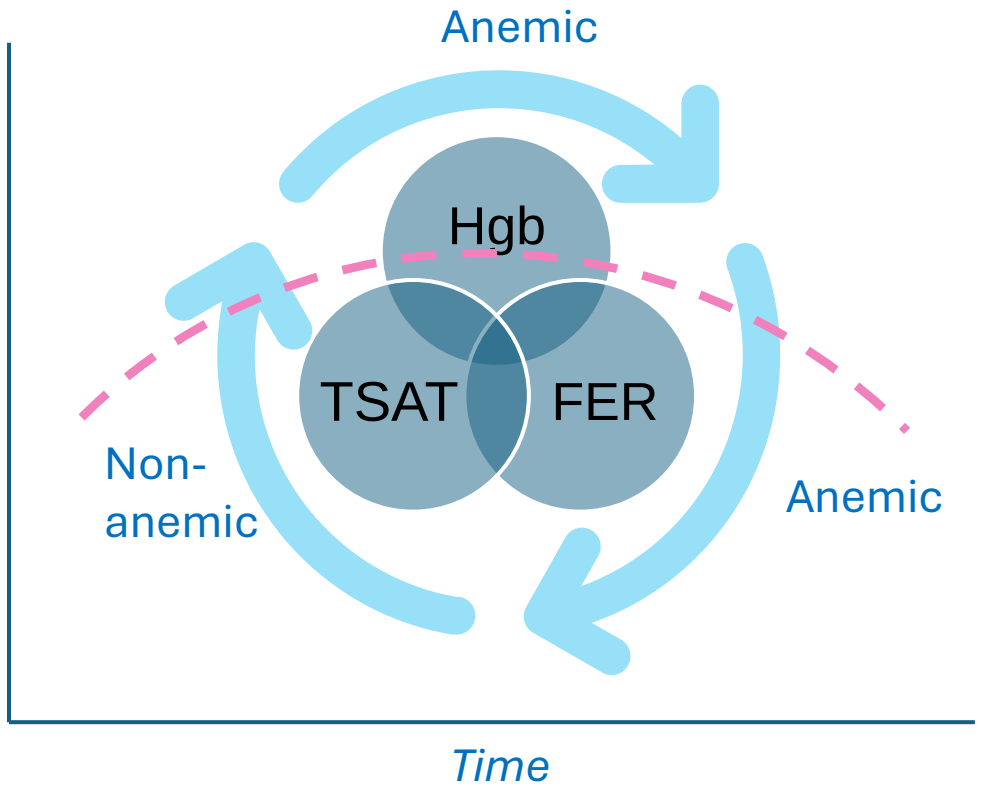
Early Identification of Patients at Risk for Iron Deficiency Anemia Using Deep Learning Techniques

Estefanie Garduno Rapp, MD,MSHI

Introduction: What is IDA and why is it important to diagnose it early?

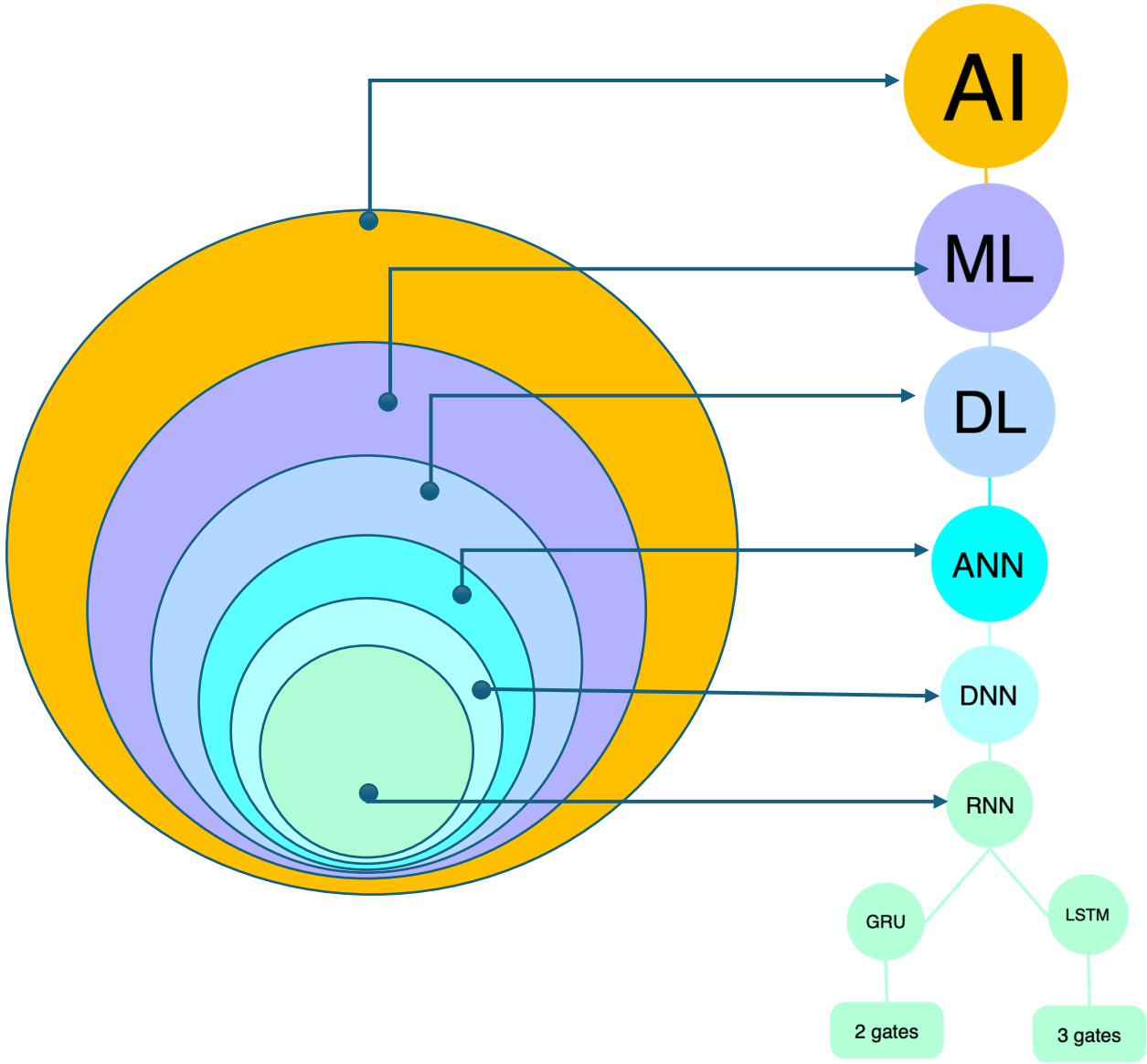
IDA is the most common extraintestinal symptom in patients with CRC

- Morbidity & Mortality
- Length of Hospital Stay
- Hospital Readmission (After Surgery)



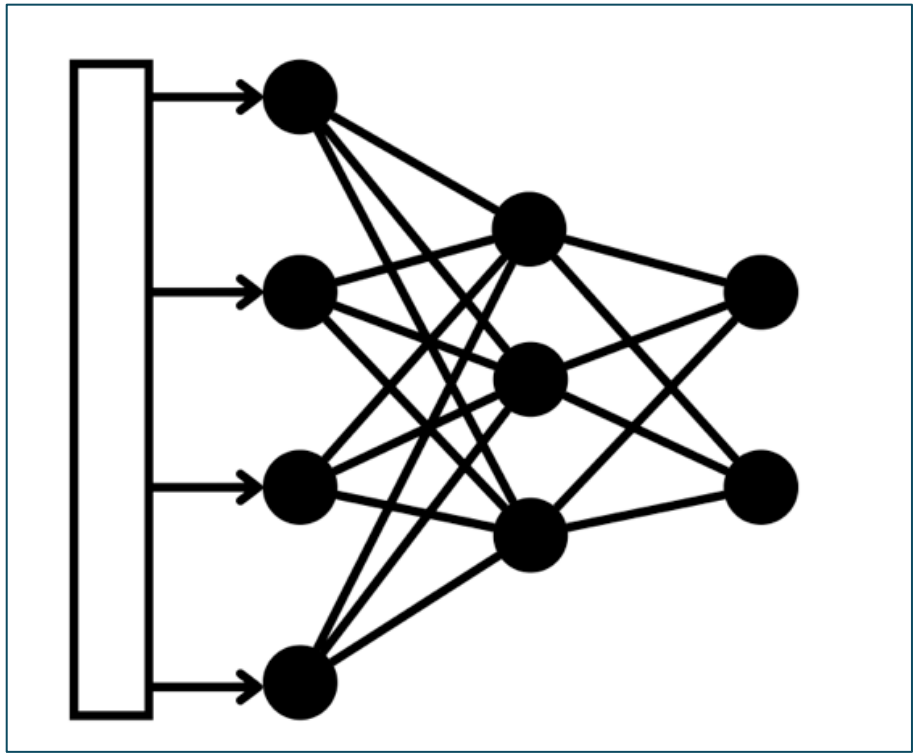
The insidious presentation of IDA can cause a significant treatment delay.

Introduction to Deep Learning



Hypothesis

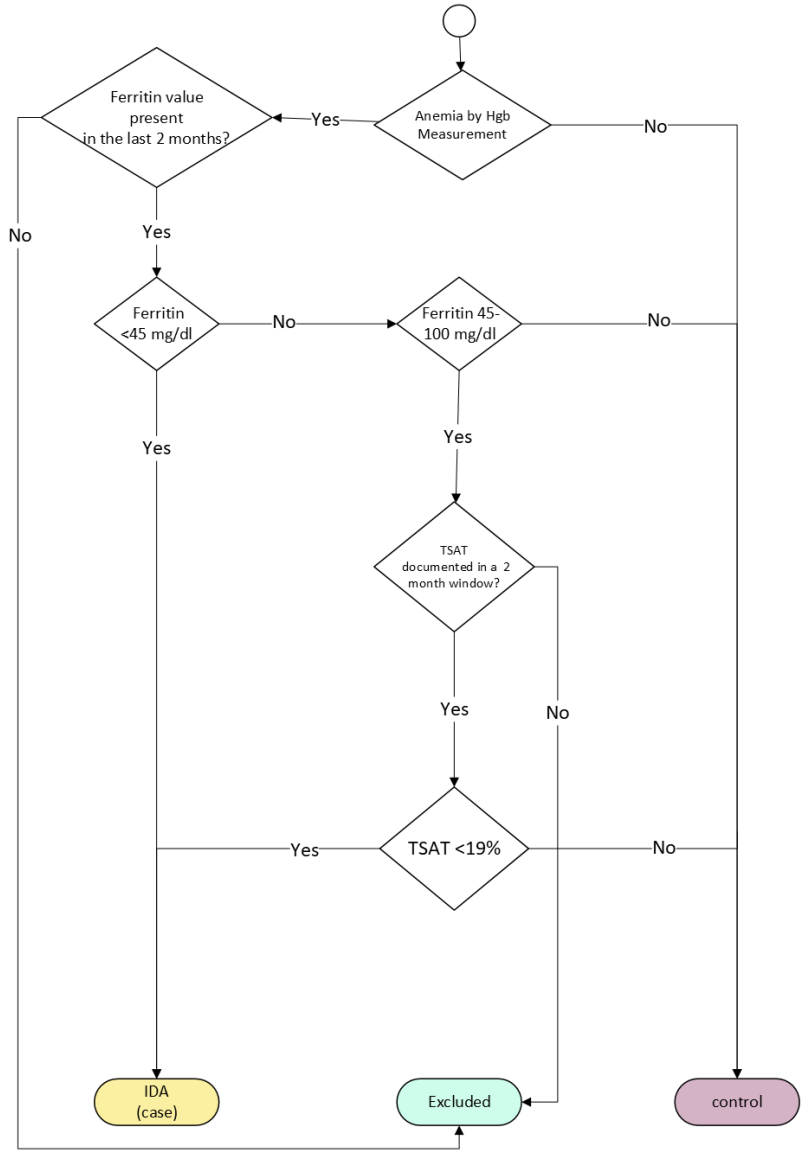
- The application of deep learning algorithms, particularly recurrent neural networks, will produce accurate predictions of the likelihood of patients developing IDA within a three to six-month timeframe, outperforming the diagnostic capabilities of traditional iron studies.

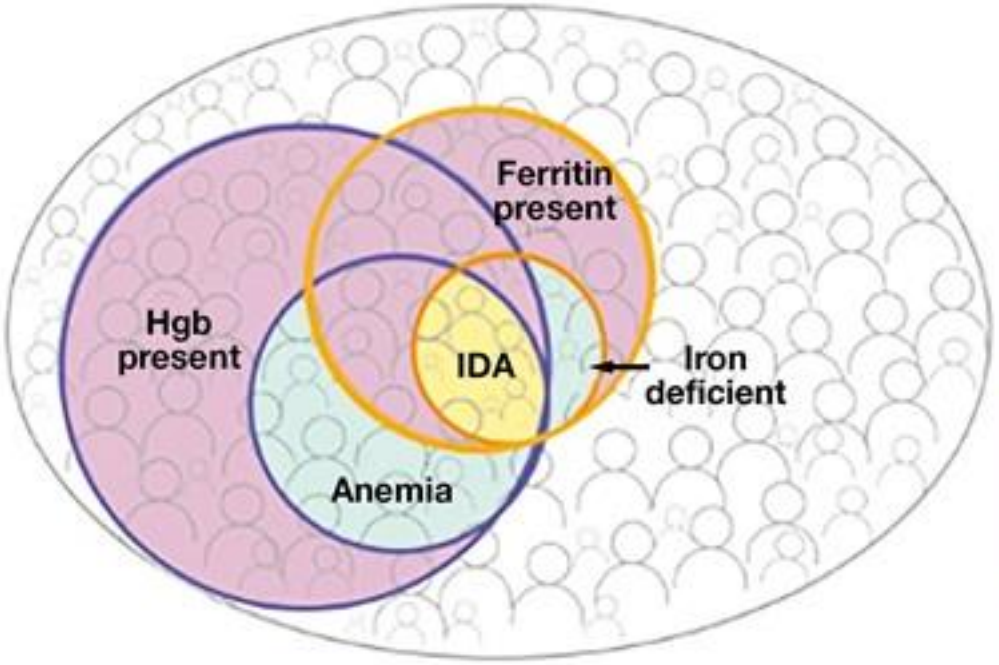


52 components

Iron saturation, percent, historical 1	1
Iron saturation, percent	1
% Saturation, total iron-binding capacity	1
Lymphocytes, percent, historical	2
Reactive lymphocytes, absolute count	3
Albumin	4
Alkaline phosphatase, historical	5
Alkaline phosphatase	5
Alkaline phosphatase S	5
Alanine aminotransferase	6
Anion gap	7
Aspartate aminotransferase	8
Bands, percent	9
Basophils, percent	10
Basophils, absolute count	11
Bilirubin, total	12
Blasts, percent, historical	13
Blasts, percent	13
Blood urea nitrogen	14
Calcium	15
Chloride	16
Carbon dioxide	17
Creatinine	18
Eosinophils, percent, historical	19
Eosinophils, percent	19
Eosinophils, absolute count	20
Red blood cells, absolute count	21
Ferritin	22
Ferritin, serum	22
Glucose	23
Granulocytes, absolute count	24
Hematocrit	25
Hemoglobin	26
Total iron-binding capacity, historical 1	27
Total iron-binding capacity, historical 2	27
Total iron-binding capacity, historical 3	27
Total iron-binding capacity	27
Unsaturated iron-binding capacity	28
Iron	29
Iron, serum	29
Lymphocytes, absolute count	30
Mean corpuscular hemoglobin	31
Mean corpuscular hemoglobin concentration	32
Mean corpuscular volume	33
Mean platelet volume	34
Metamyelocytes, absolute count	35
Metamyelocytes, percent	36
Monocytes, percent	37
Monocytes, absolute count	38
Myelocytes, absolute count	39
Myelocytes, percent	40
Neutrophils, percent	41
Neutrophils, absolute count	42
Nucleated red blood cells, percent	43
Nucleated red blood cells, absolute count	44
Platelets, absolute count	45
Potassium	46
Potassium, blood	46
Total protein	47
Red blood cell distribution width	48
Reticulocytes, absolute count	49
Reticulocytes, percent, historical	50
Reticulocytes, percent	50
Sodium	51
White blood cells, absolute count	52

Defining Outcome Variable and Cohort of Interest

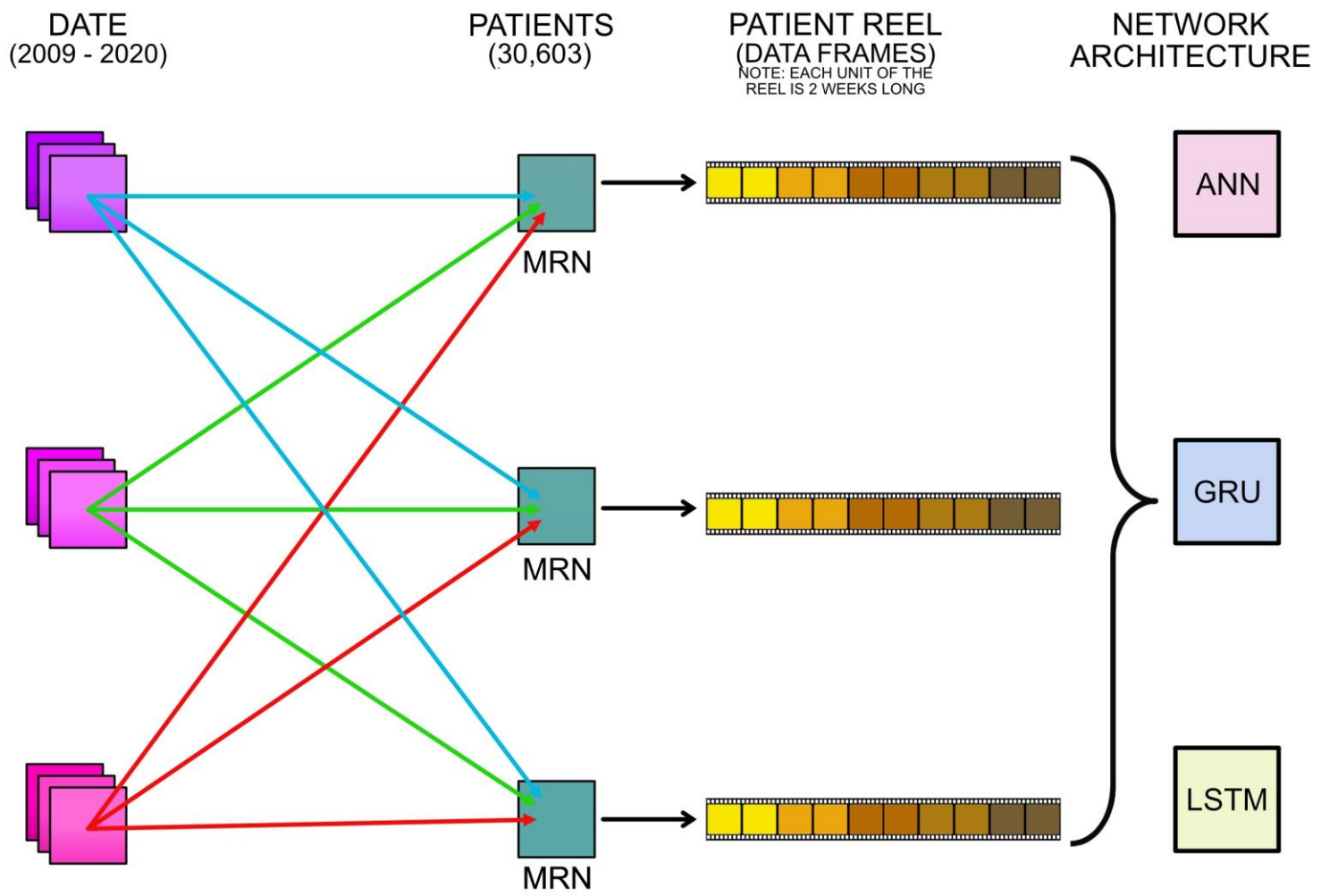




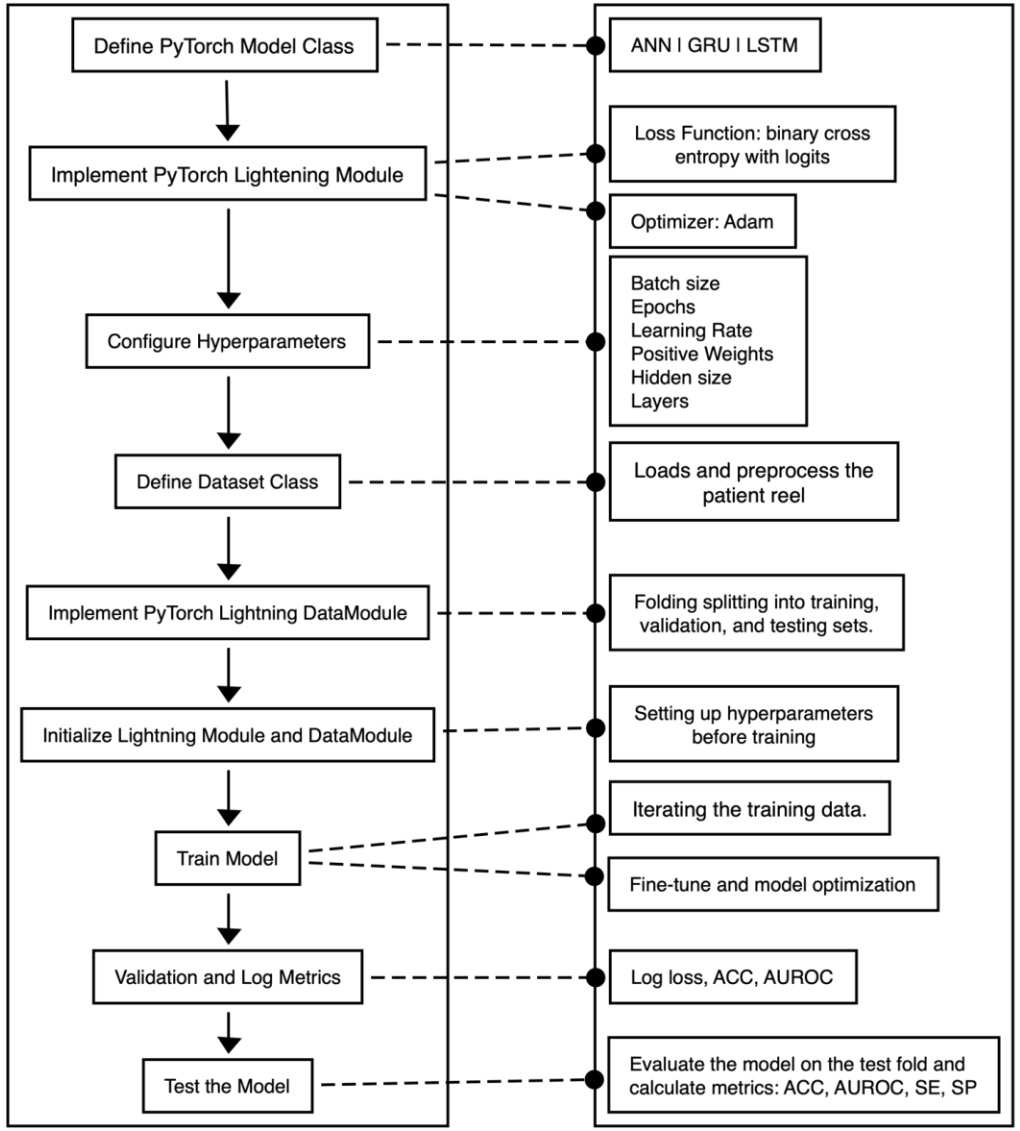
Hgb (Hemoglobin); TSAT (Transferrin Saturation); IDA (Iron Deficiency Anemia)

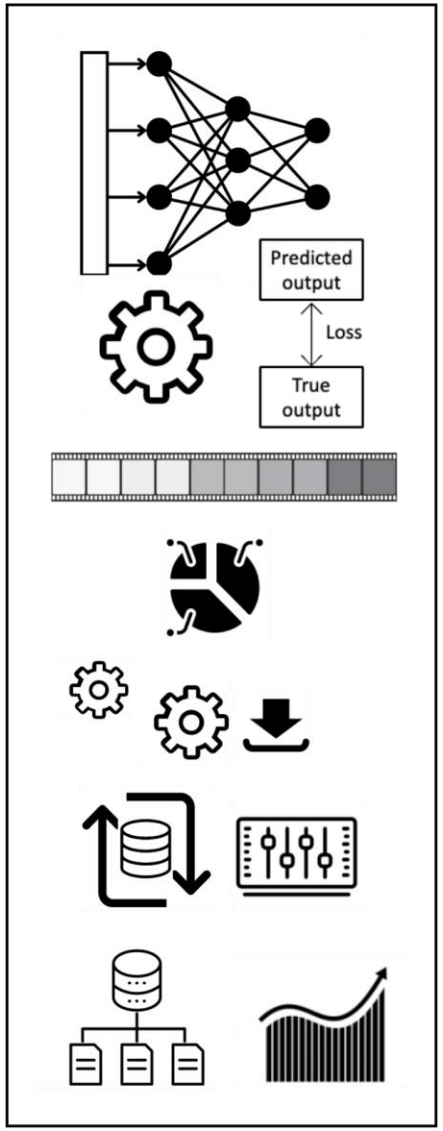
Objective and Methods

To develop and evaluate three DL models trained on retrospective longitudinal outpatient laboratory data to identify adult patients at risk for developing IDA three to six months prior to the establishment of the diagnosis by traditional iron studies.



Network Development Using the PyTorch Framework





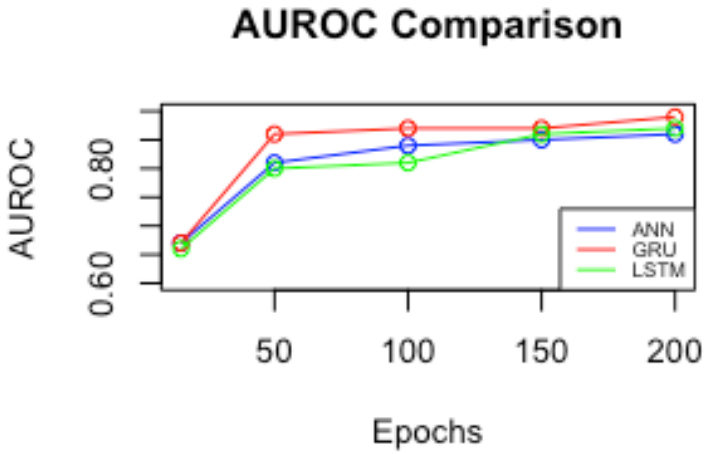
Best performing values per network			
Hyperparameter	**ANN**	**GRU**	**LSTM**
Batch size	800	800	800
Epochs	200	200	200
Learning Rate	0.001	0.001	0.0003
Positive Weights	2.0	2.0	2.0
Hidden size	16	64	64
Layers	3	2	2

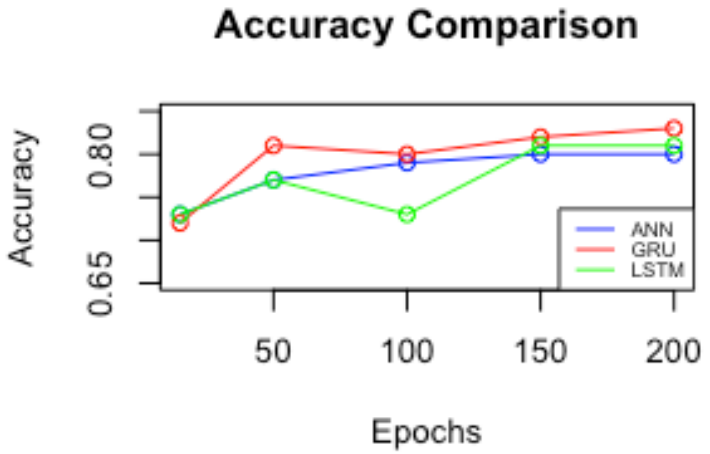
Brief Definitions
Batch Size: Number of examples in one training iteration
Epochs: Full pass through the entire training dataset.
Learning Rate: Determines step size for weight updates.
Positive Weights: Weights assigned to neuron connections.
Hidden Size: Neurons or units in a neural network's hidden layer.
Layers: The building blocks of a neural network.

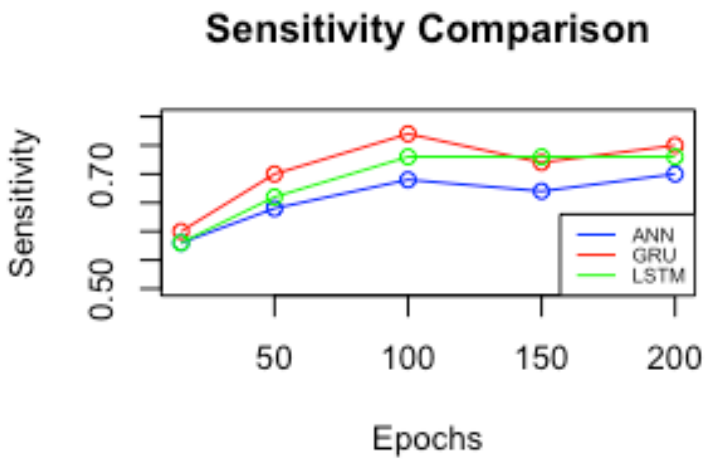
Summary of ANN Performance				
Epochs	AUROC	Accuracy	SE	SP
50	0.81	0.77	0.64	0.81
100	0.84	0.79	0.69	0.82
150	0.85	0.80	0.67	0.85
200	**0.86**	**0.80**	**0.70**	**0.84**
STDEV	0.02	0.01	0.03	0.02
Mean	0.84	0.79	0.68	0.83

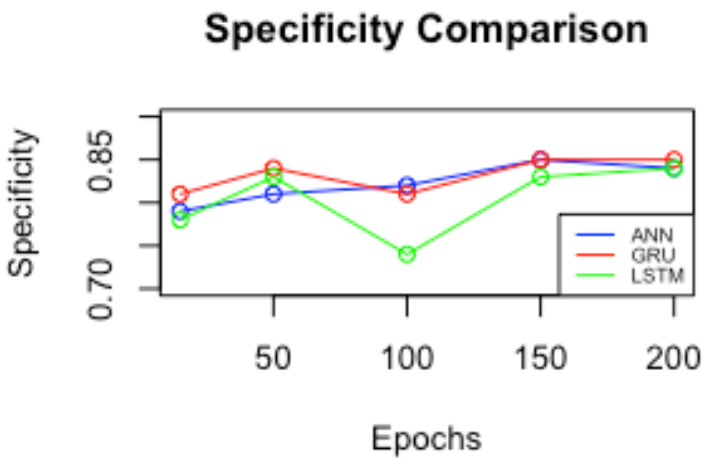
Summary of GRU Performance				
Epochs	AUROC	Accuracy	SE	SP
50	0.86	0.81	0.70	0.84
100	0.87	0.80	0.77	0.81
150	0.87	0.82	0.72	0.85
200	**0.89**	**0.83**	**0.75**	**0.85**
STDEV	0.01	0.01	0.03	0.02
Mean	0.87	0.81	0.74	0.84

Summary of LSTM Performance				
Epochs	AUROC	Accuracy	SE	SP
50	0.80	0.77	0.66	0.83
100	0.81	0.73	0.73	0.74
150	0.86	0.81	0.73	0.83
200	**0.87**	**0.81**	**0.73**	**0.84**
STDEV	0.04	0.04	0.04	0.05
Mean	0.84	0.78	0.71	0.81









Discussion

Deep learning can early identify patients at risk of iron deficiency anemia using outpatient laboratory data sequences.

The model uses features from standard outpatient lab data, demonstrating the feasibility of integrating a sequential model into clinical settings.

What makes this model unique?

Our model's ground truth effectively navigates the complexity of IDA while considering the non-linear trend of the disease in contrast to other models which use Hgb or Fe singularly to define IDA	Our model was able to predict IDA 3-6 months before conventional diagnostic methods. While other authors focus on predictions made from already anemic patients.	We utilized a robust dataset in conparison to other studies in which models were developed with extremely limited datasets.

Limitations

External Dataset

“Black box”

The Implementation of this model could

1. 
   Assist clinicians in identifying high-risk individuals, prompting timely interventions
2. Enable healthcare personnel to integrate predictive analytics into routine outpatient visits.
   
3. 
   Offer a novel and promising avenue for potential CRC screening during a clinic visit.
4. Identifies strategic information for population health initiatives
   

Conclusion

This framework extends beyond IDA, holding promise for predictive modeling in diseases requiring longitudinal blood tests such as diseases of the liver, thyroid, and kidney.

Acknowledgements

Read our Paper here!

Bibliography

1. Read AJ, Waljee AK, Chen CS, Holleman R, Kumbier KE, Saini SD. Prevalence of Appropriate Testing for Incident Anemia in the US Department of Veterans Affairs. *JAMA Netw Open*. Jan 4
2. Wang Y, Liu L, Wang C. Trends in using deep learning algorithms in biomedical prediction systems. *Frontiers in Neuroscience*. 2023;17doi:10.3389/fnins.2023.1256351
3. Schmidt RM. Recurrent Neural Networks (RNNs): A gentle Introduction and Overview. *ArXiv*. 2019;abs/1912.05911
4. Rinke ML, Singh H, Heo M, Adelman JS, O'Donnell HC, Choi SJ, Norton A, Stein REK, Brady TM, Lehmann CU,
5. Appiahene P, Asare JW, Donkoh ET, Dimauro G, Maglietta R. Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms. *BioData Mining*. 2023;16(1)doi:10.1186/s13040-023-00319-z
6. Asare JW, Appiahene P, Donkoh ET, Dimauro G. Iron deficiency anemia detection using machine learning models: A comparative study of fingernails, palm and conjunctiva of the eye images. *Engineering Reports*. 2023;5(11)doi:10.1002/eng2.12667
7. Azarkhish I, Raoufy MR, Gharibzadeh S. Artificial Intelligence Models for Predicting Iron Deficiency Anemia and Iron Serum Level Based on Accessible Laboratory Data. *Journal of Medical Systems*. 2012;36(3):2057-2061. doi:10.1007/s10916-011-9668-3
8. Andro M, Le Squere P, Estivin S, Gentric A. Anaemia and cognitive performances in the elderly: a systematic review. *Eur J Neurol*. Sep 2013;20(9):1234-40. doi:10.1111/ene.12175
9. Cook R, O'Dwyer N, Parker H, et al. Iron Deficiency Anemia, Not Iron Deficiency, Is Associated with Reduced Attention in Healthy Young Women. *Nutrients*. 2017;9(11):1216. doi:10.3390/nu9111216
10. Munoz M, Gomez-Ramirez S, Campos A, Ruiz J, Liumbruno GM. Pre-operative anaemia: prevalence, consequences and approaches to management. *Blood Transfus*. Jul 2015;13(3):370-9. doi:10.2450/2015.0014-15

- APPENDIX

Outcome variable definition

- The definition of our outcome variable (diagnosis of IDA) was informed by scientific literature and expert collaborators.

- IDA was defined as: 1) a hemoglobin below the lower reference limit AND 2a) a ferritin <45 µg/L OR 2b) a ferritin 45-100 µg/L AND a transferrin percent saturation ≤ 19%. All results had to be documented within a two-month window.

Normalization

We calculated how much the actual laboratory value differed from the lower reference limit. Then, we determined the total range within the reference values. We divided the first difference by the second to determine the relation of the laboratory value to the normal range. Finally, we subtracted 0.5 from the result to interpret and scale the values in the model. Negative values implied that the value was below the normal range, and positive values implied above, showing how much, it deviates from normal values. The calculation was computed like this:

$$Normalized\ Value = \frac{lab\ value - ref\ low}{ref\ high - ref\ low} - 0.5$$

Model development

The loss function that uses binary cross-entropy with logits is often referred to as "Binary Cross-Entropy Loss with Logits" or "Sigmoid Cross-Entropy Loss." It is commonly used in binary classification problems where the model outputs logits (log-odds or unnormalized probabilities) for two classes, and a sigmoid activation function is applied to convert these logits into probabilities.

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^{N} [y_i \cdot \log(\sigma(\hat{y}_i)) + (1 - y_i) \cdot \log(1 - \sigma(\hat{y}_i))]$$

Here:

- N is the number of samples in the batch.

- yi is the true label (either 0 or 1) for the i-th sample.

- ŷ is the logit output for the i-th sample.

- sigma is the sigmoid activation function.

- log is the natural logarithm.

This loss function penalizes the model based on the difference between its predicted probabilities (after applying the sigmoid activation) and the true binary labels. It's commonly used in logistic regression and other binary classification tasks.

Adam Optimizer

- Optimization algorithm commonly used in training artificial neural networks.
- Adjusts the learning rates for each parameter individually based on their historical gradients. This adaptability helps handle different scales and variations in the data.
- Adam uses a moving average of the gradients' past values to help accelerate convergence and overcome obstacles such as local minima.