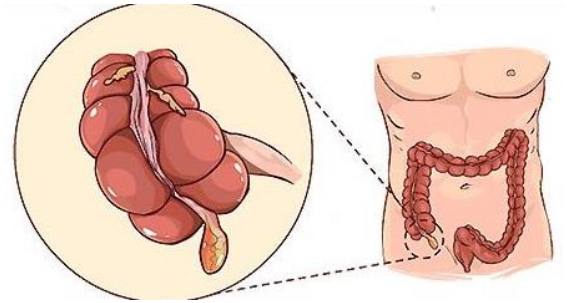
Multimodal Learning for Appendicitis Diagnosis

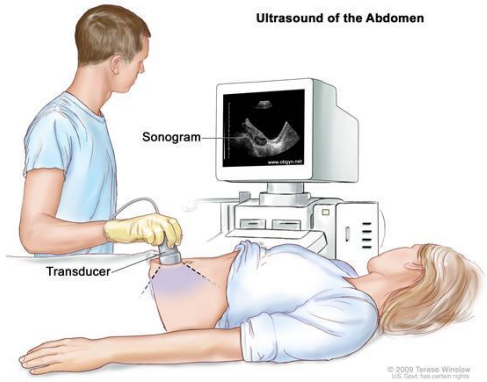
Saroj Baral

April 2, 2025

Introduction



Borrowed from indushealthplus (a)



Borrowed from nibib (b)

Borrowed from mayoclinic(c)

- **Appendicitis**: Inflammation of Appendix
- Generally, occurs in young individuals (ages 10–30)

Ultrasound

- Less expensive
- Uses Sound Wave
- No Radiation Exposure

Computed Tomography

- More expensive than Ultrasound
- Uses X rays to make detail images
- Radiation Exposure

a) https://www.indushealthplus.com/front/media/article_img/appendicitis-causes-symptoms-prevention.jpg
b) https://www.nibib.nih.gov/sites/default/files/inline-images/Ultrasound_Terese%20Winslow.jpg
c) https://www.mayoclinic.org/tests-procedures/ct-scan/about/pac-20393675

Overall Goal and Challenges

Ultrasound Images

Tabular Data

- Demographic (Age, Sex)
- Physical (BMI, Weight)
- Lab finding (WBC, RBC counts, etc.)

Multimodal Neural Network

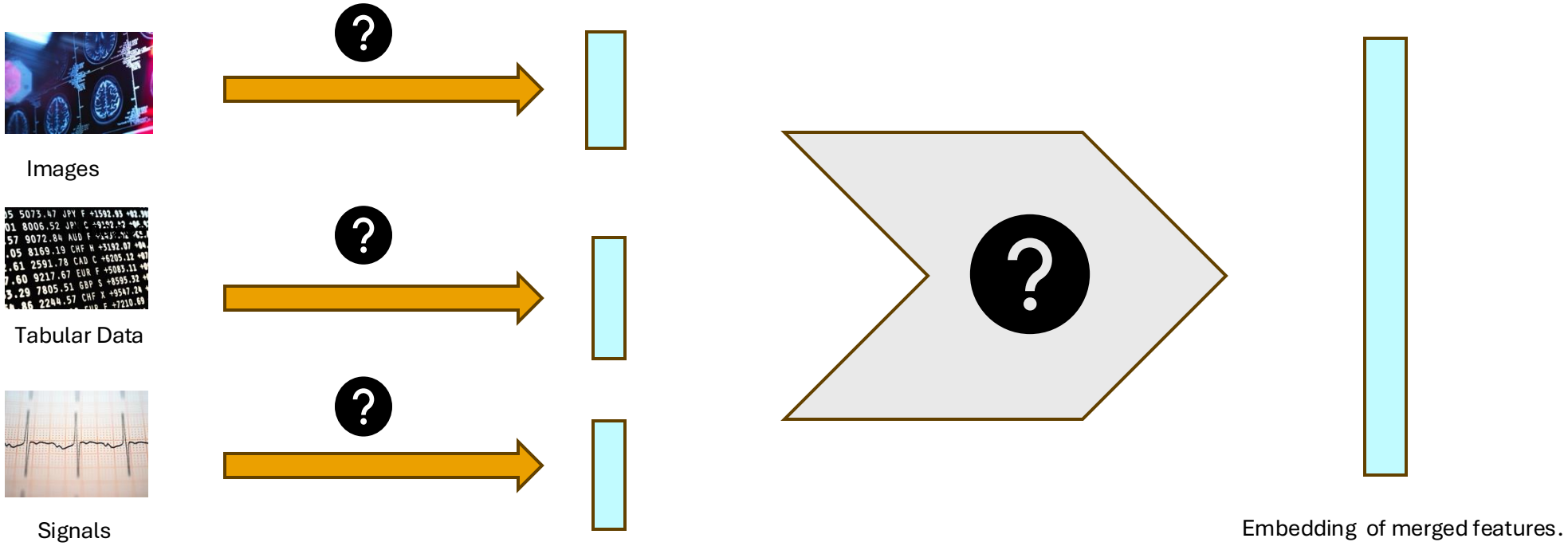
Diagnosis

Major Challenges

- Image preprocessing
- Small data size (579 records)
- Varying no of images (views) per record

Dataset Link: https://zenodo.org/records/7711412

Challenges in Multimodal Architecture



Representation of multiple modalities in vectors (embeddings)

Representation

How to encode data from different modality?

Fusion

How to combine information from different modality?

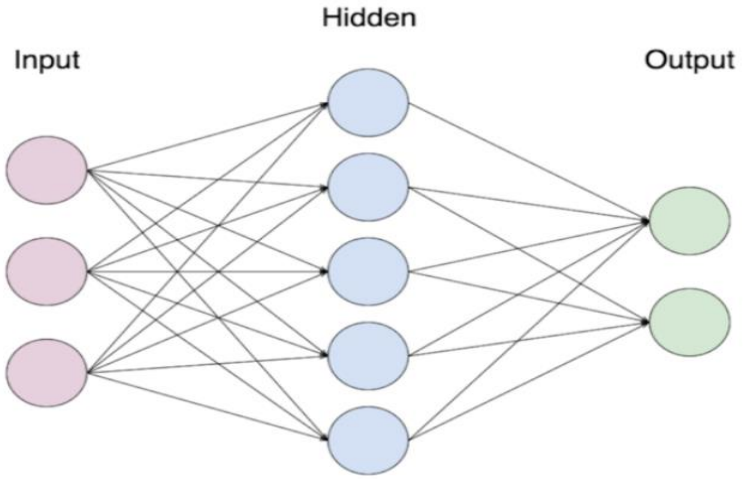
Proposed Model Architecture : Representing Tabular data

Input

- Lab reports (WBC, RBC count ,etc.)
- Physical (BMI, Weight)

Output

- Vector representing above feature (embedding)

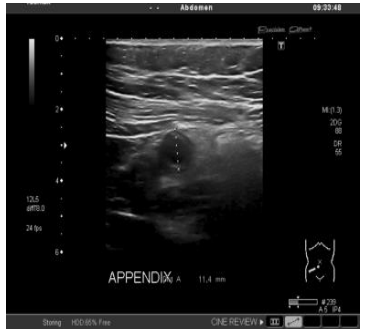


Borrowed from specbee (d)

Neural Network extracts useful information from data

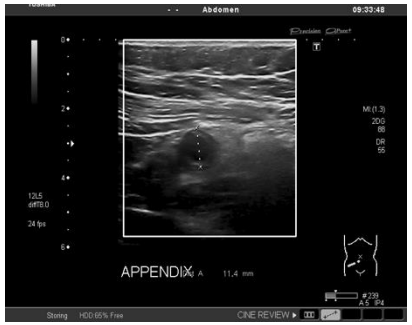
d) https://www.specbee.com/blogs/neural-networks-what-does-future-artificial-intelligence

Image preprocessing



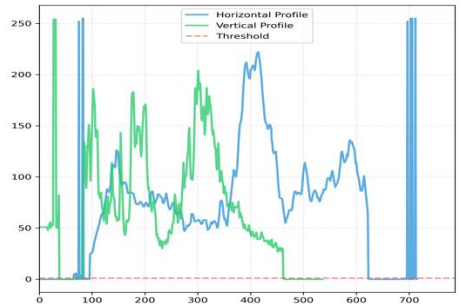
1. Original Image

- Annotations from ultrasound machine
- Markers by doctors
- Multiple in one file



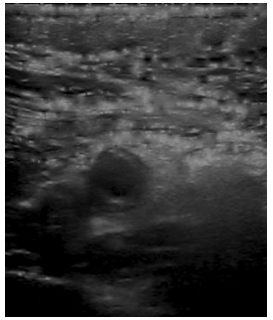
2. Original Image with boundaries

- Tried automated boundary detection
- Didn't work well for all
- Manually inspected and corrected



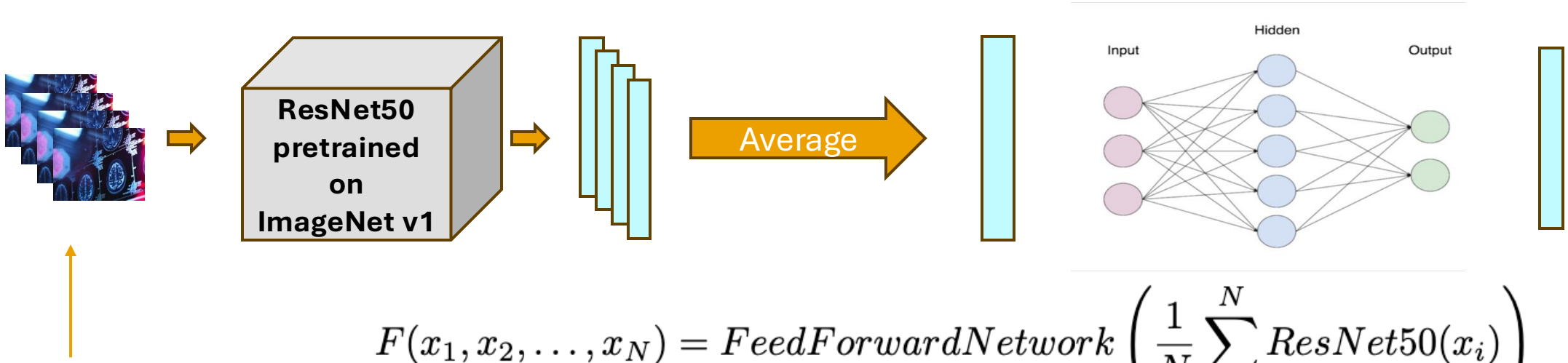
3. Intensity Profile

- Intensity profile is inspected
- Created mask for appropriate threshold



3. Final Image with annotations removed

Proposed Model Architecture : Representing Images



$$F(x_1, x_2, \ldots, x_N) = FeedForwardNetwork\left(\frac{1}{N}\sum_{i=1}^{N} ResNet50(x_i)\right)$$

Problem: Varying no. of pictures of a single person

From Deep Sets by Zaheer et.al [e]

Theorem 2 *A function $f(X)$ operating on a set X having elements from a countable universe, is a valid set function,* i.e., ***invariant*** *to the permutation of instances in X, iff it can be decomposed in the form $\rho\left(\sum_{x\in X}\phi(x)\right)$, for suitable transformations ϕ and ρ.*

e) Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. R., & Smola, A. J. (2017). Deep sets. *Advances in neural information processing systems*, *30*.

Model Architecture : Fusion Strategy

Fusion is way of combining the modality after representation.

Two main strategies:

- Early Fusion (Feature level fusion)
  - Fuse at input to the model
- Late fusion
  - Fuse later in intermediate or end (ensemble)

Late fusion are slightly effective. [f]

Using Late fusion in the proposed model

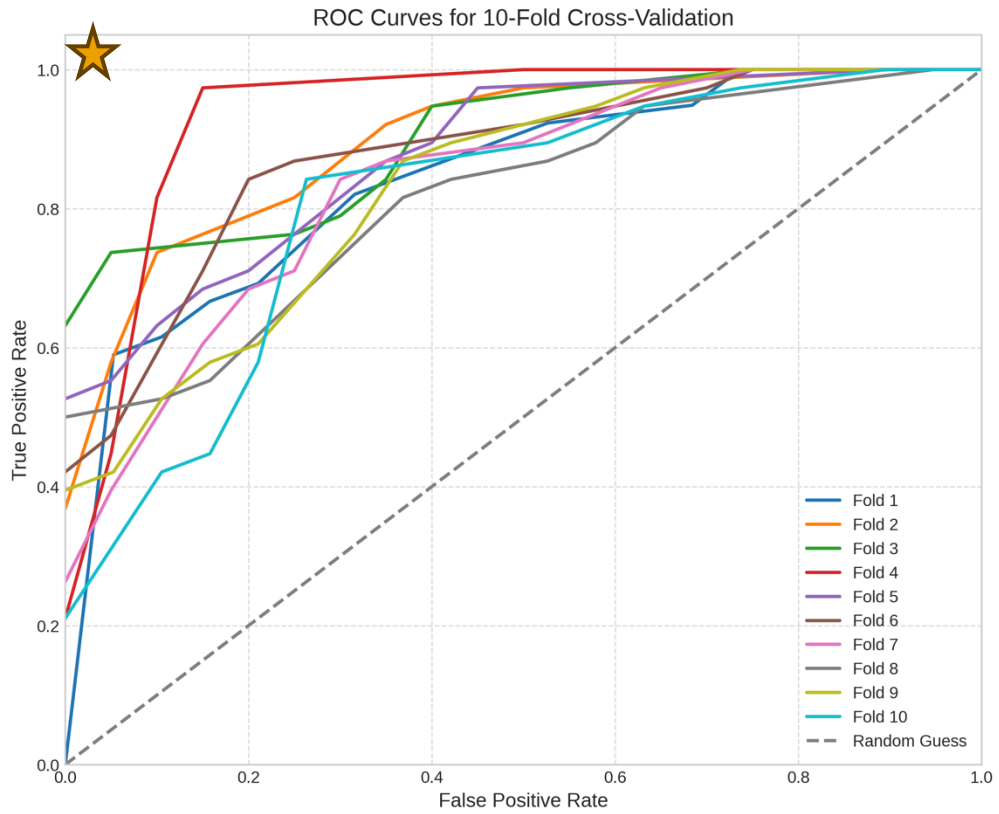
1. **Simple Concatation**
2. **Weighted Sum (For interpretability)**

Fused vector is then fed into another neural network for classification !!

f. Pawłowski M, Wróblewska A, Sysko-Romańczuk S. Effective Techniques for Multimodal Data Fusion: A Comparative Analysis. Sensors (Basel). 2023 Feb 21;23(5):2381. doi: 10.3390/s23052381. PMID: 36904585; PMCID: PMC10007548.

Results

Perfect Model



Comparing with related work (**AUROC**) on same dataset

Multimodal	**Images only (Marcinkevis et. al [g])**
0.84 ± 0.04	0.80 ± 0.06

Fusion Weights in 2nd fold

- W1 = 1.045 (tabular)
- W2 = 0.034 (images)

g. Marcinkevičs, R., Reis Wolfertstetter, P., Klimiene, U., Chin-Cheong, K., Paschke, A., Zerres, J., Denzinger, M., Niederberger, D., Wellmann, S., Ozkan, E., Knorr, C., & Vogt, J. E. (2024). Interpretable and intervenable ultrasonography-based machine learning models for pediatric appendicitis. In Medical Image Analysis (Vol. 91, p. 103042). Elsevier BV. https://doi.org/10.1016/j.media.2023.103042

Conclusions

This multimodal neural network shows promising potential for improving diagnostic accuracy over single-modal methods.

Doctors are better at distinguishing ultrasound images than neural networks.

The approach can be extended to other medical conditions that require combining imaging and tabular data.

Future work can explore alternative fusion techniques or models and larger datasets for further improvements

Thank you for listening