EXPLAINABLE AUTOMATED CODING OF CLINICAL NOTES

Seminar Advanced Machine Learning in Big Data Analytics

Presentation by Lisa Heihoff

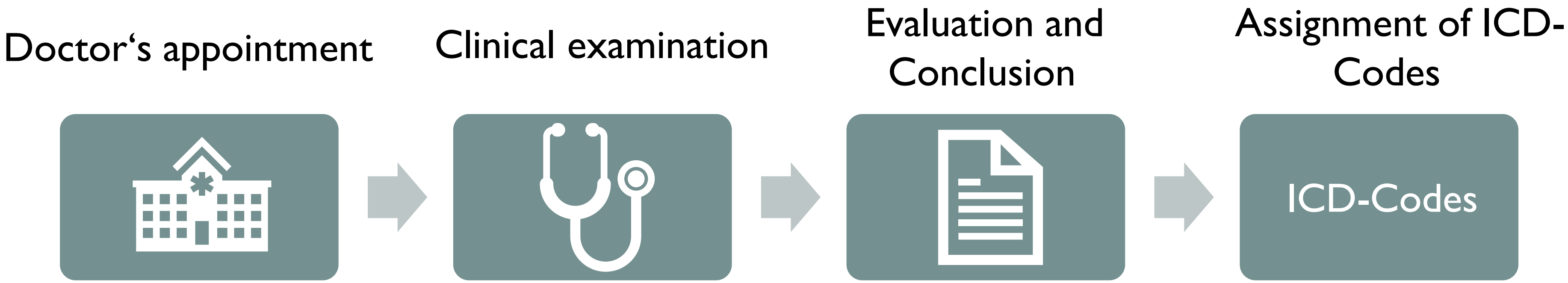
CONTENT

- Background Information
- Recent approaches and Problems
- HLAN
- Experiments and Results
- Further research and Conclusion

11-ICD-CODES

- International Classification of Disease
- Used in 117 countries
- Common language for reporting and monitoring disease
- Classification, standardizing and documentation of diagnosis and symptoms
- Usage in research, statistical analysis and billing health insurance

FROM DOCTORS APPOINTMENT TO ICD-CODE



Doctor‘s appointment

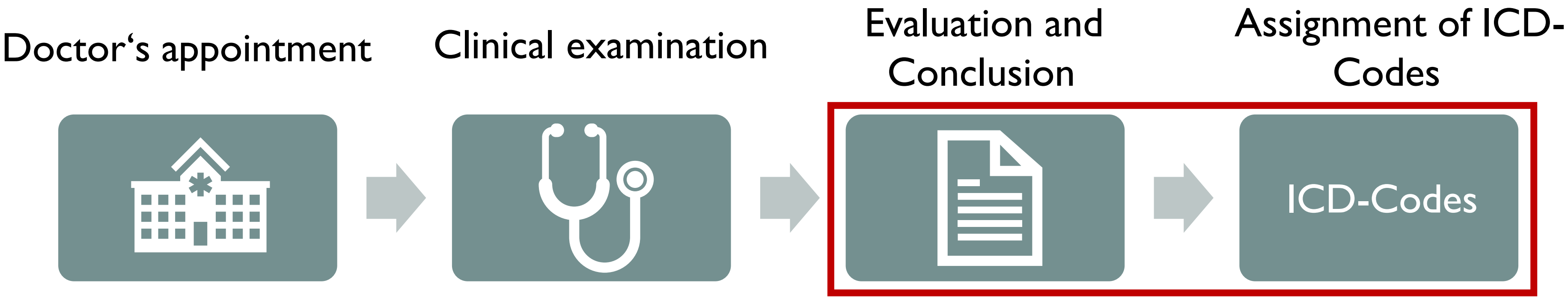
Clinical examination

Evaluation and Conclusion

Assignment of ICD-Codes

ICD-Codes

FROM DOCTORS APPOINTMENT TO ICD-CODE



Doctor‘s appointment

Clinical examination

Evaluation and Conclusion

Assignment of ICD-Codes

ICD-Codes

MANUAL CODING

- Doctor of medical personal writes codes for reports manually
- Advantage: Professionals understand complex cases and dependencies
- Problems
  - Relies heavenly on coder‘s knowledge and attention to detail
  - Different interpretation -> inconsistent
  - Needs much human effort -> not efficient
  - Issues in accuracy

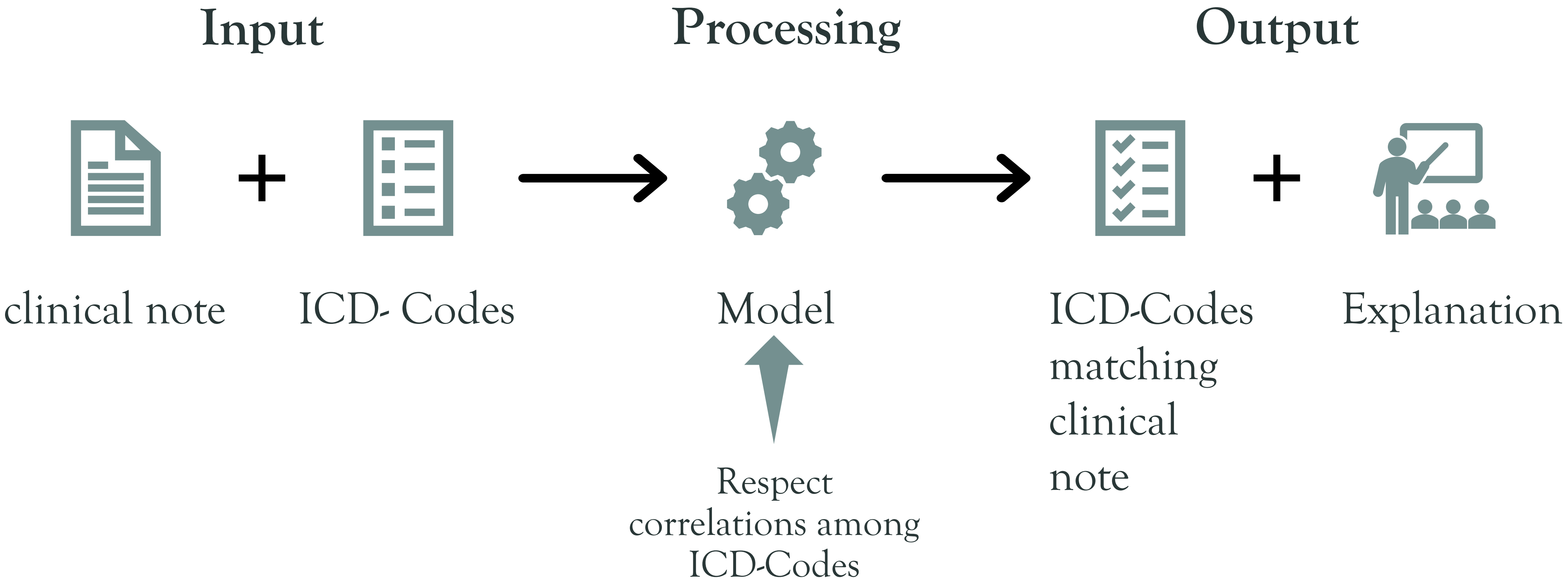
RECENT APPROACHES

- Early studies in 2010
  - Based on grammar, rules and string matching
- Based on deep learning models
  - Convolutional and Recurrent Neuronal Networks
  - Multi-label classification problem
  - Promising performance
  - Assume independence of ICD-Codes
  - Explainability is poor

PROBLEMS

- “Needle-in-a-haystack”-issue
  - Locate the key words and sentences relevant to each code
- Explainability
  - Important for clinical trust and ethical considerations (e.g., GDPR)
- Complex correlations of ICD-codes
  - Biological association among different diseases
  - Improve performance by representing correlations

PROBLEM



Input

clinical note + ICD- Codes

Processing

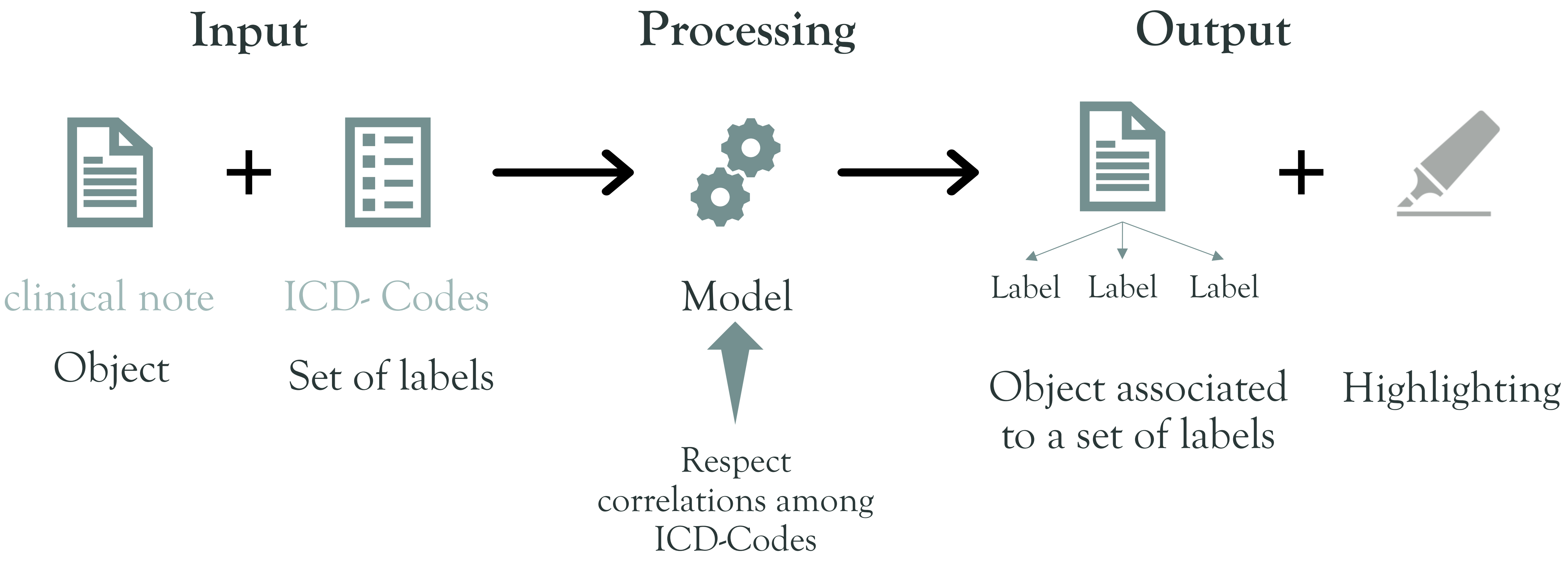
Model

Respect correlations among ICD-Codes

Output

ICD-Codes matching clinical note + Explanation

FORMALISED PROBLEM



Input

Processing

Output

clinical note

ICD- Codes

Object

Set of labels

Model

Respect correlations among ICD-Codes

Label

Label

Label

Object associated to a set of labels

Highlighting

HLAN

Hierarchical Label-wise Attention Network (HLAN) with label-wise word-level and sentence-level attention mechanisms

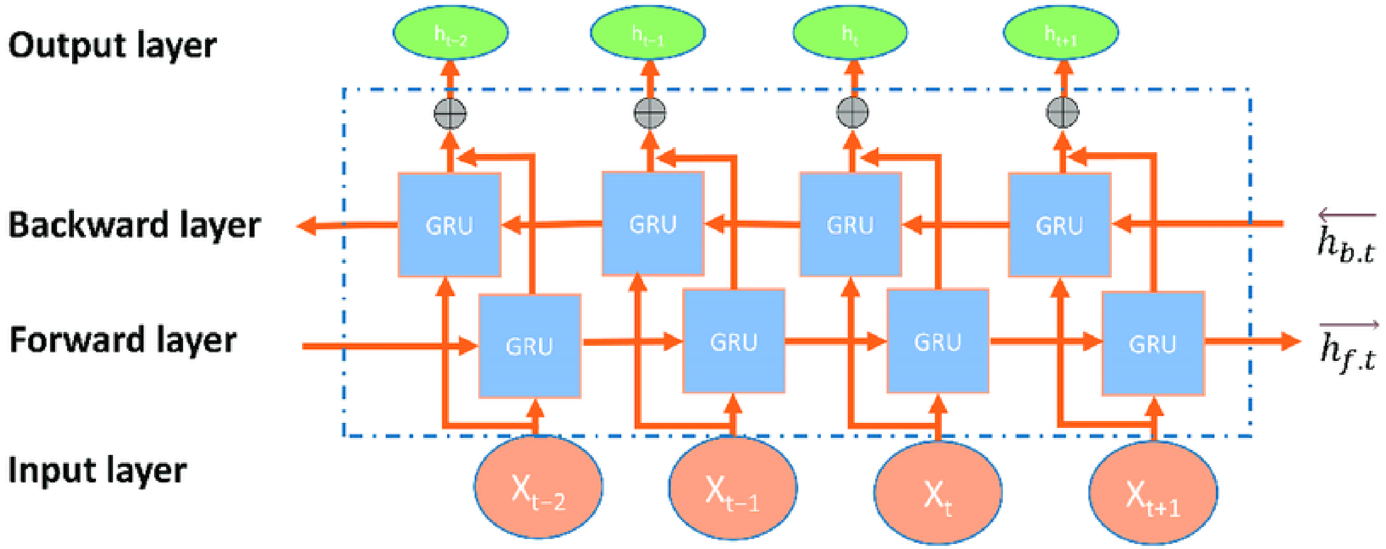
Based on

-Hierarchical Attention bi-directional Gated Recurrent Units (HA-GRU)

-Hierarchical Attention Networks (HAN)

BI-GRU

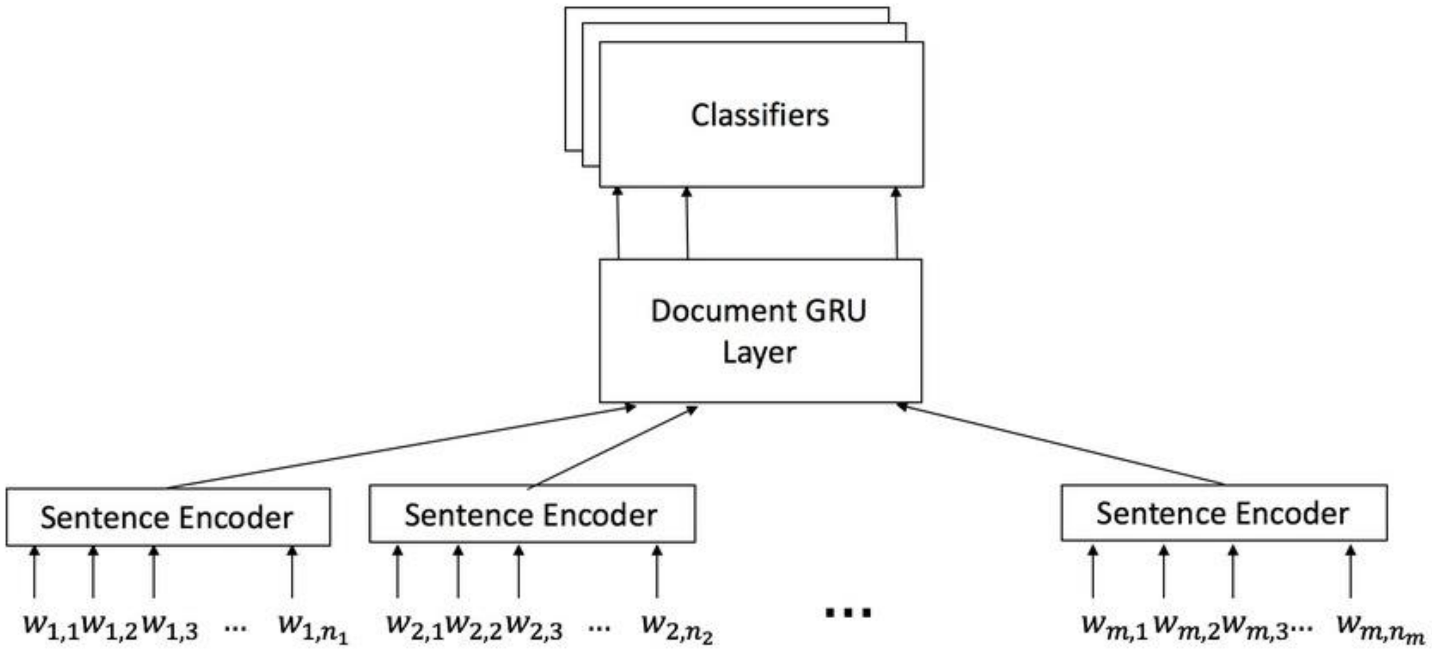
- used in HLAN, HA-GRU and HAN
- Capture long term dependencies
- Reads each token one by one and produces a hidden state h^t
- Reads sequence forwardly and backwardly



PICTURE'S SOURCE: HTTPS://WWW.RESEARCHGATE.NET/PUBLICATION/371937679/FIGURE/FIG2/AS:11431281170996173@1687971900816/BI-GRU-STRUCTURE-DIAGRAM.PNG

HA-GRU

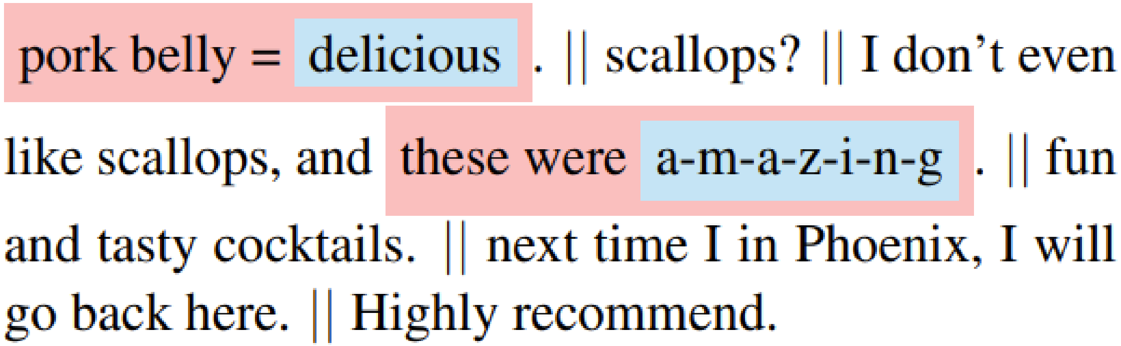
- Recent model
- sentence level explanation for each label
- No specific essential words leading to decision for each code
- For good explanation sentence and word level necessary


Classifiers
Document GRU Layer
Sentence Encoder
Sentence Encoder
Sentence Encoder
$w_{1,1} w_{1,2} w_{1,3} \cdots w_{1,n_1}$
$w_{2,1} w_{2,2} w_{2,3} \cdots w_{2,n_2}$
...
$w_{m,1} w_{m,2} w_{m,3} \cdots w_{m,n_m}$

PICTURE'S SOURCE: MULTI-LABEL CLASSIFICATION OF PATIENT NOTES A CASE STUDY ON ICD CODE ASSIGNMENT - SCIENTIFIC FIGURE ON RESEARCHGATE. AVAILABLE FROM: HTTPS://WWW.RESEARCHGATE.NET/FIGURE/HA-GRU-MODEL-ARCHITECTURE-OVERVIEW_FIG1_320075101 [ACCESSED 9 SEPT 2024]

HAN

- Word-level and sentence-level attention
- Used for document classification
- Highlights words and sentences leading to the classification
- Improves interpretability but lacks label-specific explanations
- Similar structure to HLAN



pork belly = delicious . || scallops? || I don’t even like scallops, and these were a-m-a-z-i-n-g . || fun and tasty cocktails. || next time I in Phoenix, I will go back here. || Highly recommend.

PICTURE’S SOURCE: HTTPS://WWW.CS.CMU.EDU/~./HOVY/PAPERS/16HLT-HIERARCHICAL-ATTENTION-NETWORKS.PDF

HAN STRUCTURE

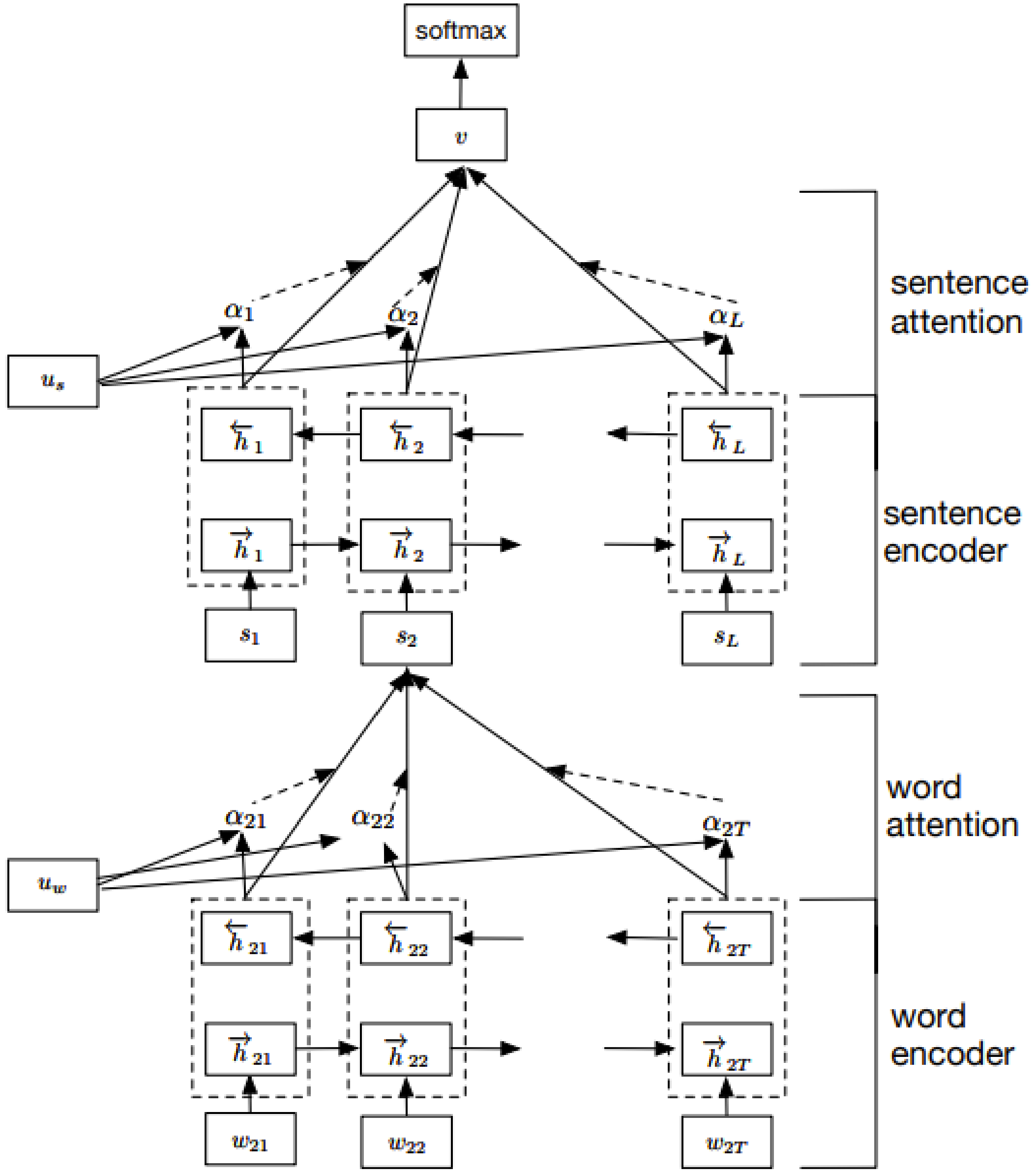
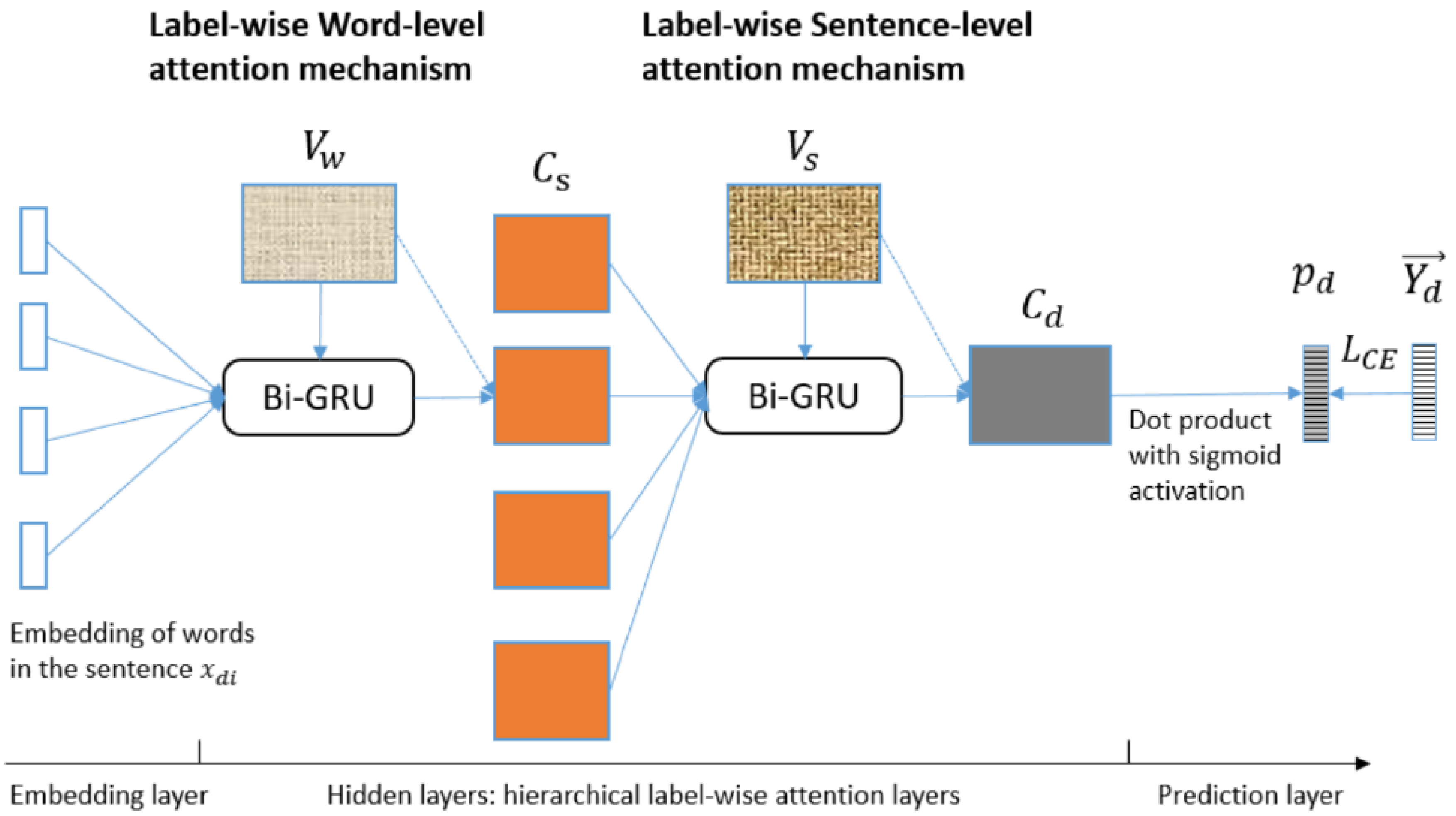


Figure 2: Hierarchical Attention Network.

PICTURE'S SOURCE: HTTPS://WWW.CS.CMU.EDU/~./HOVY/PAPERS/16HLT-HIERARCHICAL-ATTENTION-NETWORKS.PDF

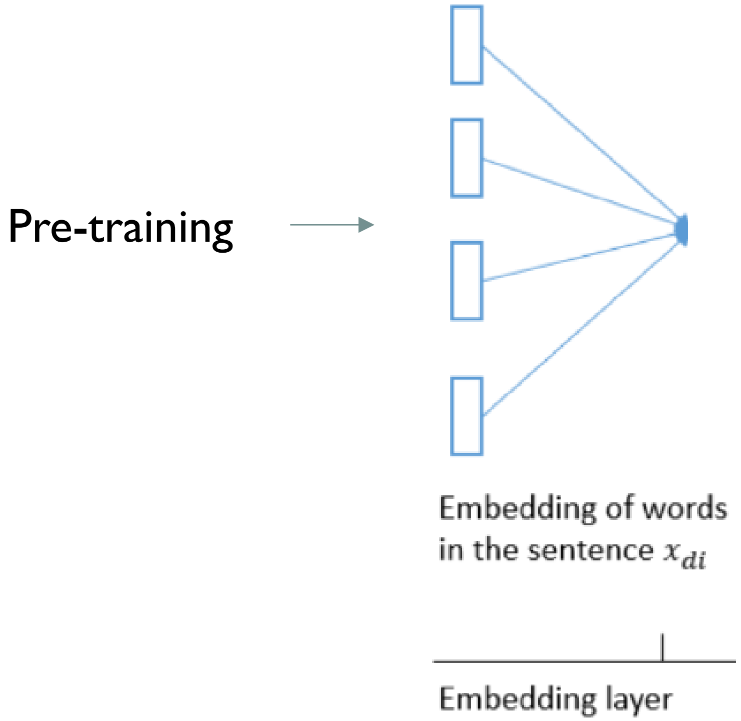
HLAN ARCHITECTURE


Label-wise Word-level attention mechanism
Label-wise Sentence-level attention mechanism
V_w
C_s
V_s
C_d
p_d
$\overrightarrow{Y_d}$
L_{CE}
Bi-GRU
Bi-GRU
Dot product with sigmoid activation
Embedding of words in the sentence x_{di}
Embedding layer
Hidden layers: hierarchical label-wise attention layers
Prediction layer

PICTURE’S SOURCE:EXPLAINABLE AUTOMATED CODING OF CLINICAL NOTES USING HIERARCHICAL LABEL-WISE ATTENTION NETWORKS AND LABEL EMBEDDING INITIALISATION BY HANG DONGA,D, V´ICTOR SUAREZ-PANIAGUA ´ A,D, WILLIAM WHITELEYB,D, HONGHAN WUC,

PRE-TRAINING

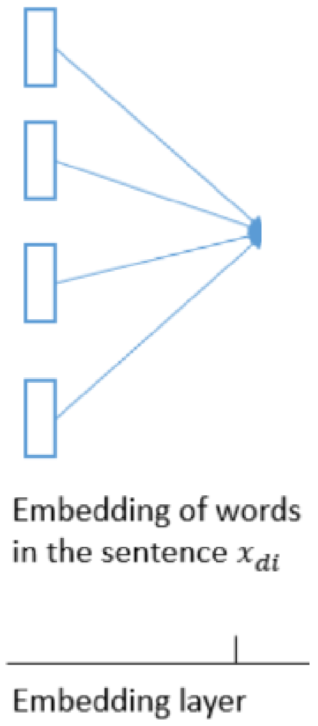
- Neural word embedding algorithm with label sets of training data as input
- Captures label-co-occurrences and correlations
- Resulting label embeddings used to initialise weights W_e for the neural network



PICTURE'S SOURCE:EXPLAINABLE AUTOMATED CODING OF CLINICAL NOTES USING HIERARCHICAL LABEL-WISE ATTENTION NETWORKS AND LABEL EMBEDDING INITIALISATION BY HANG DONGA,D, V´ICTOR SUAREZ-PANIAGUA ´ A,D, WILLIAM WHITELEYB,D, HONGHAN WUC

EMBEDDING LAYER

- Input: clinical note
- Each word x_{di} transformed into one-hot input representation u_{di}
- One-hot input representation u_{di} transformed into low-dimensional continuous vector $e_{di} = W_e u_{di}$



PICTURE'S SOURCE:EXPLAINABLE AUTOMATED CODING OF CLINICAL NOTES USING HIERARCHICAL LABEL-WISE ATTENTION NETWORKS AND LABEL EMBEDDING INITIALISATION BY HANG DONGA,D, V´ICTOR SUAREZ-PANIAGUA ´ A,D, WILLIAM WHITELEYB,D, HONGHAN WUC

HIDDEN LAYERS - WORD LEVEL

- One BI-GRU each sentence
- BI-GRU generates hidden states h_w for each word in both directions
- V_w := context matrix for word level attention mechanism
- Each row of V_w V_{wl} is context vector for label y_l
- Attention scores are generated for each word using V_{wl} and hidden state h_w

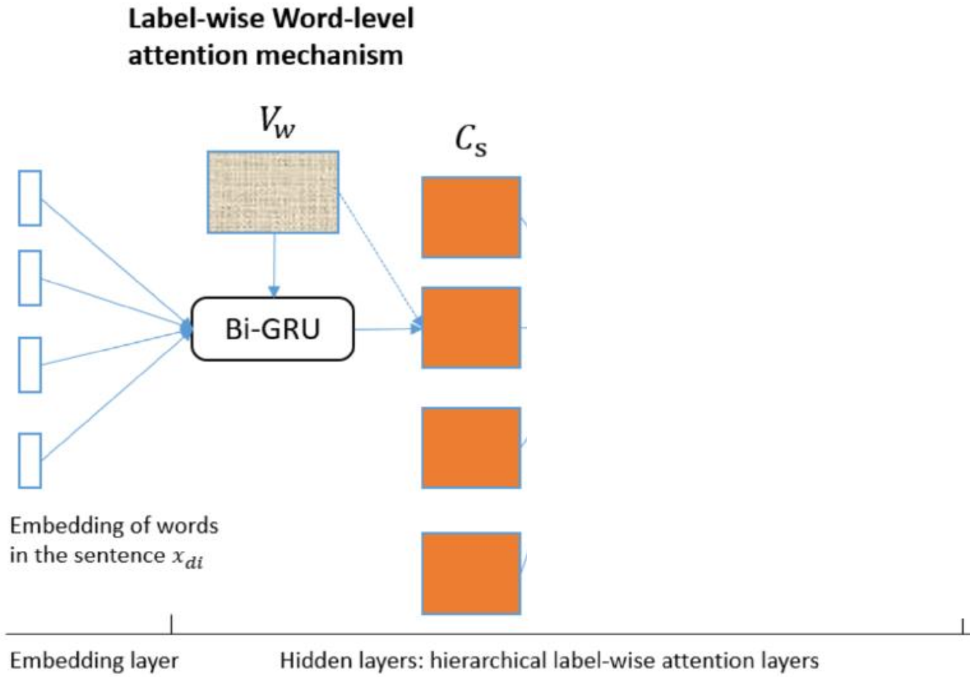
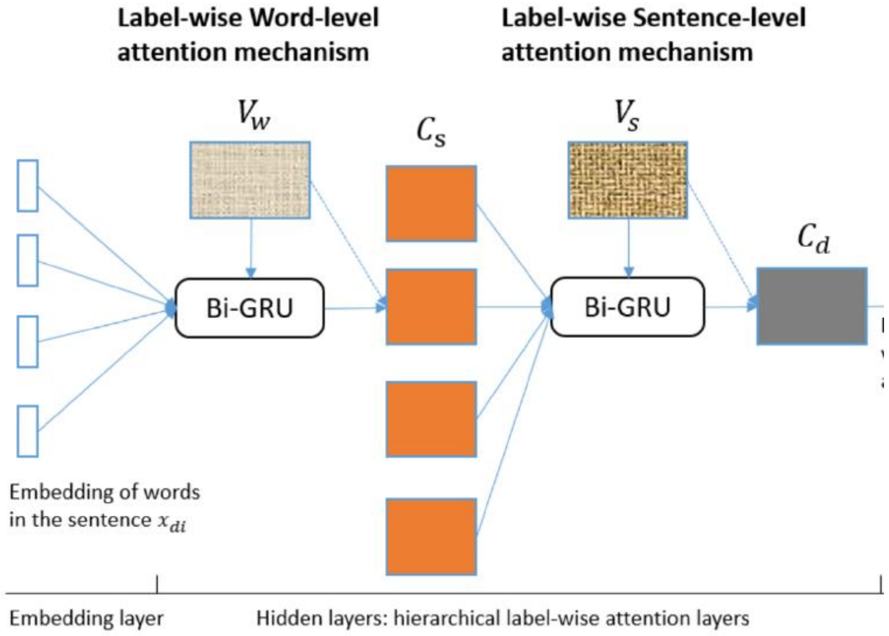

Label-wise Word-level attention mechanism
V_w
C_s
Bi-GRU
Embedding of words in the sentence x_{di}
Embedding layer
Hidden layers: hierarchical label-wise attention layers

Figure 1: Hierarchical Label-wise Attention Network (HLAN)

PICTURE'S SOURCE:EXPLAINABLE AUTOMATED CODING OF CLINICAL NOTES USING HIERARCHICAL LABEL-WISE ATTENTION NETWORKS AND LABEL EMBEDDING INITIALISATION BY HANG DONGA,D, V´ICTOR SUAREZ-PANIAGUA ´ A,D, WILLIAM WHITELEYB,D, HONGHAN WUC

HIDDEN LAYERS - SENTENCE LEVEL

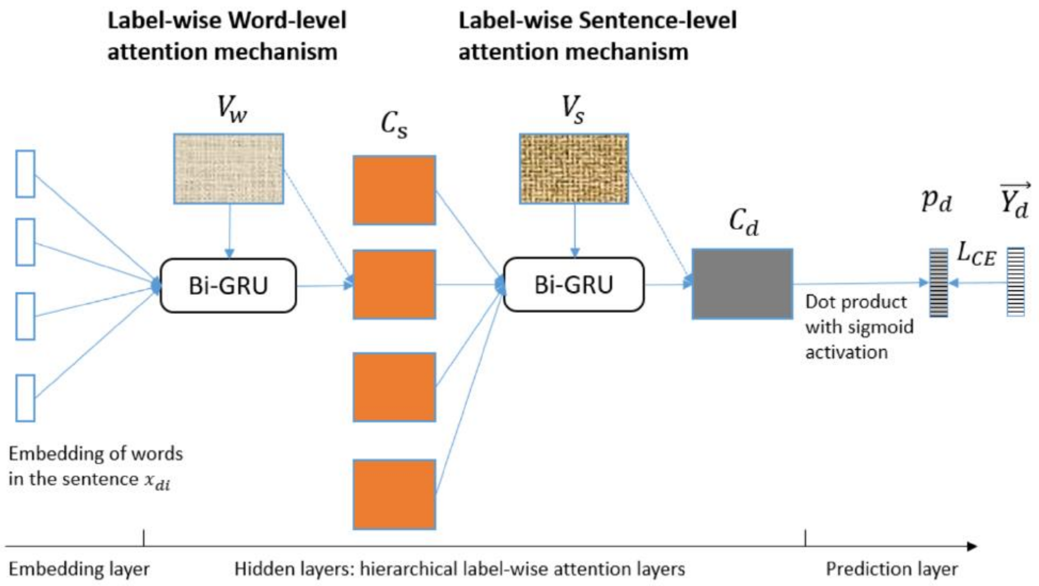
- BI-GRU generates hidden states h_s for each sentence in both directions
- V_s := context matrix for sentence level attention mechanism
- Each row of V_s V_{sl} is context vector for label y_l
- Attention scores generated for each sentence using V_{sl} and hidden state h_s of the sentence


Label-wise Word-level attention mechanism
Label-wise Sentence-level attention mechanism
V_w
C_s
V_s
C_d
Bi-GRU
Bi-GRU
Embedding of words in the sentence x_{di}
Embedding layer
Hidden layers: hierarchical label-wise attention layers

PICTURE'S SOURCE:EXPLAINABLE AUTOMATED CODING OF CLINICAL NOTES USING HIERARCHICAL LABEL-WISE ATTENTION NETWORKS AND LABEL EMBEDDING INITIALISATION BY HANG DONGA,D, V´ICTOR SUAREZ-PANIAGUA ´ A,D, WILLIAM WHITELEYB,D, HONGHAN WUC

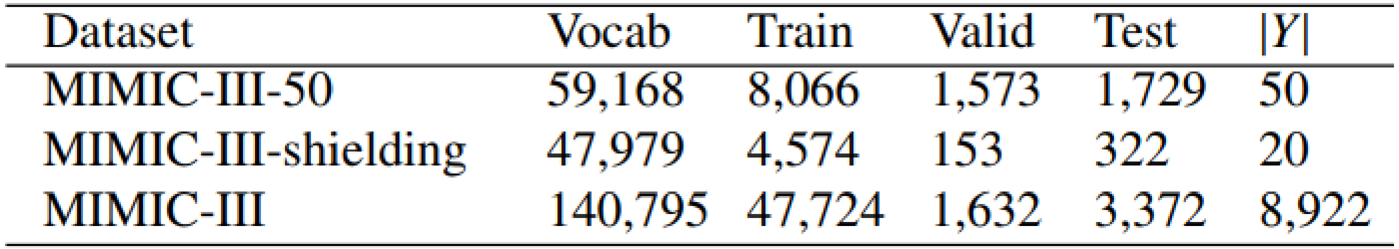
PREDICTION LAYER

- Attention-weighted representations combined to label-wise document representation
- Document representation used to predict the probability for each label
- Binary cross entropy loss function to optimise the predictions for multi-label classification


Label-wise Word-level attention mechanism
Label-wise Sentence-level attention mechanism
V_w
C_s
V_s
Bi-GRU
Bi-GRU
C_d
p_d
$\overrightarrow{Y_d}$
L_{CE}
Dot product with sigmoid activation
Embedding of words in the sentence x_{di}
Embedding layer
Hidden layers: hierarchical label-wise attention layers
Prediction layer

PICTURE'S SOURCE:EXPLAINABLE AUTOMATED CODING OF CLINICAL NOTES USING HIERARCHICAL LABEL-WISE ATTENTION NETWORKS AND LABEL EMBEDDING INITIALISATION BY HANG DONGA,D, V´ICTOR SUAREZ-PANIAGUA ´ A,D, WILLIAM WHITELEYB,D, HONGHAN WUC

EXPERIMENTS- DATASETS



Dataset	Vocab	Train	Valid	Test	$\|Y\|$
MIMIC-III-50	59,168	8,066	1,573	1,729	50
MIMIC-III-shielding	47,979	4,574	153	322	20
MIMIC-III	140,795	47,724	1,632	3,372	8,922

- MIMIC-III (Full codes, Top-50 codes, COVID-19 shielding codes)
- Full codes
  - Clinical data from adult patients between 2001 and 2012
  - ICD-9 codes annotated by professionals
- Top-50 codes by their frequencies
- For COVID-19 dataset ICD-9 codes which matched to patients with high risks during Covid-19
- Most label occurrences are from a few labels

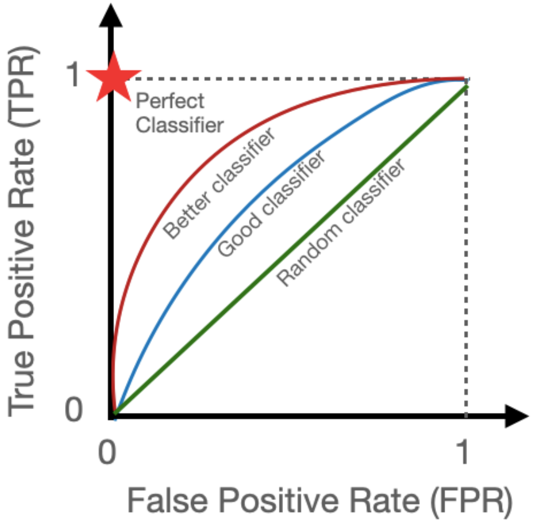
PICTURE'S SOURCE:EXPLAINABLE AUTOMATED CODING OF CLINICAL NOTES USING HIERARCHICAL LABEL-WISE ATTENTION NETWORKS AND LABEL EMBEDDING INITIALISATION BY HANG DONGA,D, V´ICTOR SUAREZ-PANIAGUA ´ A,D, WILLIAM WHITELEYB,D, HONGHAN WUC

EXPERIMENTS- SETTINGS

- CNN for text classification; CNN+att; Bi-GRU; HAN; HA-GRU
- All models with and without label embedding initialization (+LE)
- Evaluation Metrics:
    - Micro- and Macro- Averaging to AUC, Precision, Recall, F1-Score
    - Precision to the top 5 predicted labels

EXPERIMENTS- METRICS

- Recall: ratio of correctly predicted positive observations to all the actual positive observations
- Precision: ratio of correctly predicted positive observations to the total predicted positive observations
- F1 := harmonic mean of precision and recall
- Area under Curve (AUC): tradeoff between true positive rate and false positive rate



PICTURE'S SOURCE: HTTPS://MIRO.MEDIUM.COM/V2/RESIZE:FIT:494/1*EPMZI0GCGDLSTSJB6Q8E-W.PNG

RESULTS- PERFORMANCE

- HLAN performed best on MIMIC-III-50 dataset
    - Highest Micro-AUC (91.9%); Micro F1 (64.1%); Precision@5 (62.5%)
- MIMIC-III-shielding dataset: HAN and CNN performed best; HLAN achieved comparable results
- Full dataset: CNN+att with label embedding initialization best results

RESULTS- PERFORMANCE

- CNN and HAN better for smaller datasets
  - Fewer labels and documents favor simpler architectures
- HA-GRU did not perform better
  - Label wise word-level attention mechanism in HLAN improved performance
- Scalability of HLAN needs to be improved so it can process large label sizes
- Macro level metrics lower than Micro level metrics
  - String imbalance of labels in MIMIC-III

RESULTS- IMPACT OF LABEL EMBEDDING

- LE initialization improved performance for most models
- CNN+att model significant performance boost in MIMIC-III-shielding
- CNN, Bi-GRU and HA-GRU less affected
- No significant improvement for HLAN and HAN
  - Prior layers already learn label relations
- In general higher stability and low variance

RESULTS- EXPLAINABILITY

- HAN: same highlights in same document for different labels
- HA-GRU: same word-level but different sentence level highlights across labels
- HLAN most salient words and sentences
- Attention weights were unstable for CNN
  - ➢ Different results for different runs

➢ HLAN provided more meaningful interpretations by highlighting most salient words and sentences for each label

RESULTS EXPLAINABILITY

- Validation of results by experienced clinician
    - Potential reasons for false positives in different documents
- Potential reasons:
    - Missed coding of medical professionals
    - Past disease
    - Wrong correlations learned from the data
    - Subtle difference among Sub-type disease
        - Because of imbalance of vocabularies in training data

RESULTS- USAGE

- Explanation of HLAN cost further memory requirement and training time
- HLAN with fewer codes or specific tasks that require higher explainability
- HAN and HA-GRU for tasks with higher label size

FURTHER RESEARCH

- Enable application to large label size
- Incorporating external knowledge
  - To address wrong correlations
- Tests in real-world clinical setting
- Consult professionals to identify issues
- Improve efficiency and accuracy

CONCLUSION

- HLAN provides better or comparable results in comparison to state of the art models
- Label embedding initialisation boosts performance of deep learning models
- HLAN more suitable for medical coding because of model explainability
- Errors can be explained so that the model can be improved

SOURCES

- Details to HAN: https://www.cs.cmu.edu/˜./hovy/papers/16HLT-hierarchical-attention-networks.pdf [accessed 5th Sept 2024]
- Details to HA-GRU: Multi-Label Classification of Patient Notes a Case Study on ICD Code Assignment - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/HA-GRU-model-architecture-overview_fig1_320075101 [accessed 5th Sept 2024]
- BI-GRU: https://www.researchgate.net/publication/371937679/figure/fig2/AS:11431281170996173@1687971900816/Bi-GRU-structure-diagram.png [accessed 5th Sept 2024]
- Explainable Automated Coding of Clinical Notes using Hierarchical Label-wise Attention Networks and Label Embedding Initialisation Hang Donga,d, V´ıctor Suarez-Paniagua ´ a,d, William Whiteleyb,d, Honghan Wuc

THANK YOU FOR YOUR ATTENTION!

Questions?