

Supplementary Material

0.1 G-Bert exclusion

A model we also considered was G-BERT[2], which, to our knowledge, was the first BERT model applied to electronic health records. G-BERT augments the BERT encoder with a Graph Neural Network that is trained to encode the hierarchical structure of medical codes. However, G-BERT operates at the visit level rather than the patient level, making it structurally incompatible with our multi-visit framework. Furthermore, its GNN component introduces an additional architectural variable that would confound a fair comparison of sequence encoding strategies, which is the central goal of this study. For these reasons, we exclude G-BERT from our comparison.

0.2 Ki-Thrust Feature Importance

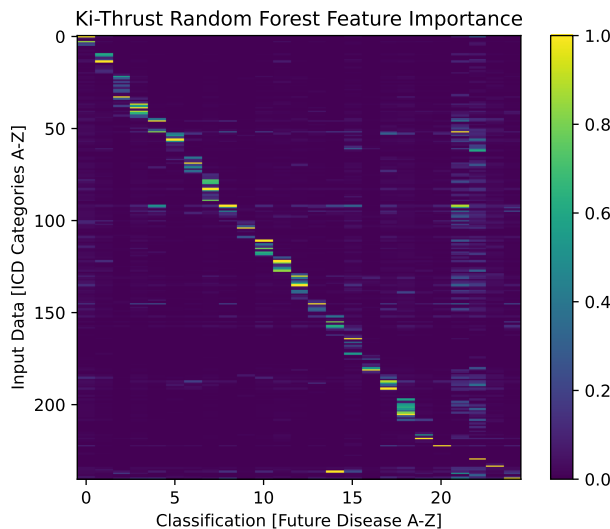


Figure 1: **categorical encoder feature importance on KI-THRUST**. The y-axis shows the 241 input ICD category features, and the x-axis shows the future disease classification targets.

0.3 Implementation Details

For the architecture of ICD2BERT, we initialized from the same architecture and the pretraining weights of the original BERT paper[1], specifically the `bert_base_uncased` model. The hidden size is set to 240 dimensions with all other parameters left at their default values ($L = 12, H = 240, A = 12$). The original vocabulary is replaced with ICD codes. While the token embedding layer is randomly initialized to accommodate the new vocabulary, the pretrained transformer weights provide a strong initialization for the self-attention layers, reducing the amount of pretraining data required. For Med-BERT and BEHRT, we used ICD2BERT as the base model, adding on the sequence, age, and visit encodings from their respective papers. For pretraining these models, we use masked language modeling, masking out 15% of the tokens. We used the AdamW optimizer with a starting learning rate of $5e - 5$ and linear decay. We used the Huggingface implementation of BERT, specifically from their transformers package version 5.4.0. For each model’s pretraining, we used a single Tesla V100 GPU with 32 GB of memory on the MIMIC-IV dataset and an Nvidia RTX A6000 GPU with 48GB of memory on the KI-THRUST dataset. The models were trained for

100 epochs on MIMIC-IV or for 16 epochs on KI-THRUST, roughly 6 million samples. The batch size was set to the highest multiple of 8 that still fit within the GPU’s memory, resulting in a batch size of 48. For the implementation of Pat2Vec, we used the gensim package version 4.3.3, and for their optimization framework, we used optuna version 3.6.1. We used the same hyperparameter ranges as the original paper and ran the optimization on a single 96-core Intel Xeon CPU for 1 day. The epoch count was also a hyperparameter, ranging from 5 to 40 on MIMIC-IV and from 1 to 6 on KI-THRUST. For CLMBR, we used the pretrained clmbr-t-base implementation from HuggingFace with femr version 0.2.3.

For the Logistic Regression and Random Forest classifiers, we use the scikit-learn implementation with version 1.6.0. We used the following hyperparameters for the grid searches on all datasets:

- Logistic Regression Classifier
 - C: [0.01, 0.1, 1, 10, 100]
 - penalty: [l1, l2]
 - solver: **saga**
 - class_weight: **balanced**
 - max_iter: 200
- Random Forest Classifier
 - n_estimators: [10, 25, 50, 100, 150]
 - max_depth: [2, 5, 10, None]
 - class_weight: **balanced**

ablation	future_disease	mortality	rehospitalization
full	0.78 (0.00)	0.79 (0.00)	0.68 (0.00)
icd 10	0.42 (0.01)	0.61 (0.00)	0.53 (0.00)
categories	0.44 (0.01)	0.61 (0.00)	0.53 (0.00)
icd diagnosis	0.78 (0.00)	0.80 (0.00)	0.67 (0.00)
icd 10 diagnosis	0.43 (0.01)	0.62 (0.00)	0.53 (0.00)
diagnosis length 2	0.78 (0.00)	0.80 (0.00)	0.67 (0.00)
diagnosis length 3	0.78 (0.00)	0.80 (0.00)	0.67 (0.00)

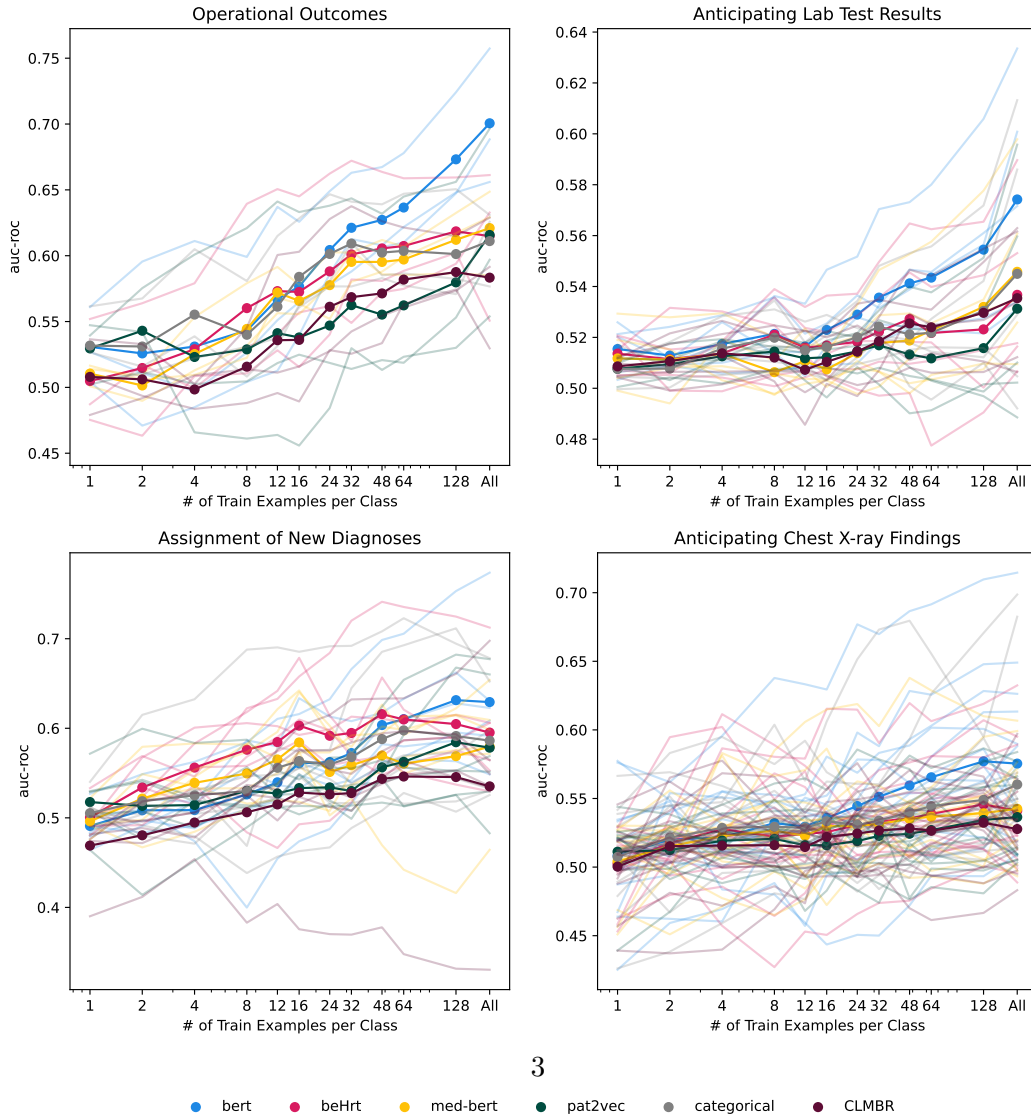
Table 1: **AUC-ROC scores for ablation study on MIMIC classification tasks.** Mean AUC-ROC scores with standard deviations (in parentheses). All encoders are 240-dimensional BERT models pretrained on the MIMIC-IV training set. Ablation refers to the type of ablation applied to the training dataset. Full refers to the model trained on the un-ablated dataset.

ablation	future_disease	mortality	rehospitalization
icd 10	0.66 (0.00)	0.76 (0.00)	0.64 (0.00)
full mimic	0.67 (0.00)	0.78 (0.00)	0.64 (0.00)
icd diagnosis	0.67 (0.00)	0.76 (0.01)	0.64 (0.00)
full ki-thrust	0.70 (0.00)	0.82 (0.00)	0.66 (0.00)
icd 10 diagnosis	0.66 (0.00)	0.78 (0.00)	0.64 (0.00)
diagnosis length 3	0.56 (0.00)	0.58 (0.00)	0.50 (0.00)
diagnosis length 2	0.51 (0.00)	0.54 (0.00)	0.58 (0.00)

Table 2: **AUC-ROC scores for ablation study on KI-THRUST classification tasks.** Mean AUC-ROC scores with standard deviations (in parentheses). All encoders are 240-dimensional BERT models pretrained on the MIMIC-IV training set. Ablation refers to the type of ablation applied to the training dataset. Full refers to the model trained on the un-ablated dataset.

Logistic Regression - AUC-ROC

□



training dataset	model type	dimensionality	future_disease	mortality	rehospitalization
MIMIC-IV	BEHRT	240	0.78 (0.01)	0.76 (0.04)	0.67 (0.00)
Ki-Thrust	BERT	240	0.73 (0.01)	0.63 (0.08)	0.58 (0.00)
MIMIC-IV	BERT	240	0.82 (0.00)	0.82 (0.00)	0.67 (0.00)
Ki-Thrust	BERT	768	0.71 (0.01)	0.63 (0.06)	0.56 (0.00)
MIMIC-IV	BERT	768	0.59 (0.01)	0.66 (0.03)	0.52 (0.00)
CLMBR	CLMBR	768	0.65 (0.00)	0.61 (0.03)	0.57 (0.00)
None	Categorical	240	0.76 (0.00)	0.60 (0.00)	0.56 (0.00)
None	Categorical	640	0.80 (0.00)	0.78 (0.00)	0.65 (0.00)
MIMIC-IV	Med-BERT	240	0.78 (0.00)	0.76 (0.01)	0.66 (0.00)
Pat2Vec	Pat2Vec	10	0.50 (0.00)	0.51 (0.01)	0.50 (0.00)
Ki-Thrust	Pat2Vec	240	0.74 (0.00)	0.66 (0.00)	0.55 (0.00)
MIMIC-IV	Pat2Vec	240	0.66 (0.00)	0.71 (0.00)	0.59 (0.00)

Table 3: MIMIC results, AUC-ROC, Logistic Regression.

training dataset	model type	dimensionality	future_disease	mortality	rehospitalization
Ki-Thrust	BERT	240	0.74 (0.00)	0.78 (0.03)	0.60 (0.02)
MIMIC-IV	BERT	240	0.69 (0.00)	0.77 (0.02)	0.61 (0.02)
Ki-Thrust	BERT	768	0.71 (0.00)	0.76 (0.03)	0.62 (0.02)
MIMIC-IV	BERT	768	0.69 (0.00)	0.77 (0.02)	0.61 (0.02)
None	Categorical	240	0.78 (0.00)	0.78 (0.02)	0.68 (0.00)
Pat2Vec	Pat2Vec	10	0.50 (0.00)	0.50 (0.01)	0.50 (0.00)
Ki-Thrust	Pat2Vec	240	0.50 (0.00)	0.49 (0.00)	0.50 (0.00)
MIMIC-IV	Pat2Vec	240	0.67 (0.00)	0.78 (0.00)	0.64 (0.00)

Table 4: KI-Thrust results, AUC-ROC, Logistic Regression.

ablation	future_disease	mortality	rehospitalization
full	0.82 (0.00)	0.82 (0.00)	0.67 (0.00)
icd 10	0.77 (0.00)	0.71 (0.00)	0.59 (0.00)
categories	0.75 (0.01)	0.66 (0.01)	0.58 (0.00)
icd diagnosis	0.82 (0.00)	0.81 (0.00)	0.67 (0.00)
icd 10 diagnosis	0.75 (0.01)	0.67 (0.02)	0.59 (0.00)
diagnosis length 2	0.80 (0.00)	0.74 (0.03)	0.68 (0.00)
diagnosis length 3	0.80 (0.00)	0.75 (0.03)	0.68 (0.00)

Table 5: MIMIC ablation, AUC-ROC, Logistic Regression.

ablation	future_disease	mortality	rehospitalization
icd 10	0.71 (0.00)	0.82 (0.01)	0.66 (0.00)
full mimic	0.69 (0.00)	0.77 (0.02)	0.61 (0.02)
icd diagnosis	0.70 (0.00)	0.83 (0.00)	0.66 (0.00)
full ki-thrust	0.74 (0.00)	0.78 (0.03)	0.60 (0.02)
icd 10 diagnosis	0.68 (0.00)	0.74 (0.03)	0.60 (0.01)
diagnosis length 3	0.61 (0.01)	0.71 (0.10)	0.60 (0.03)
diagnosis length 2	0.52 (0.01)	0.52 (0.04)	0.57 (0.02)

Table 6: KI-Thrust ablation, AUC-ROC, Logistic Regression.

Logistic Regression - F1

□

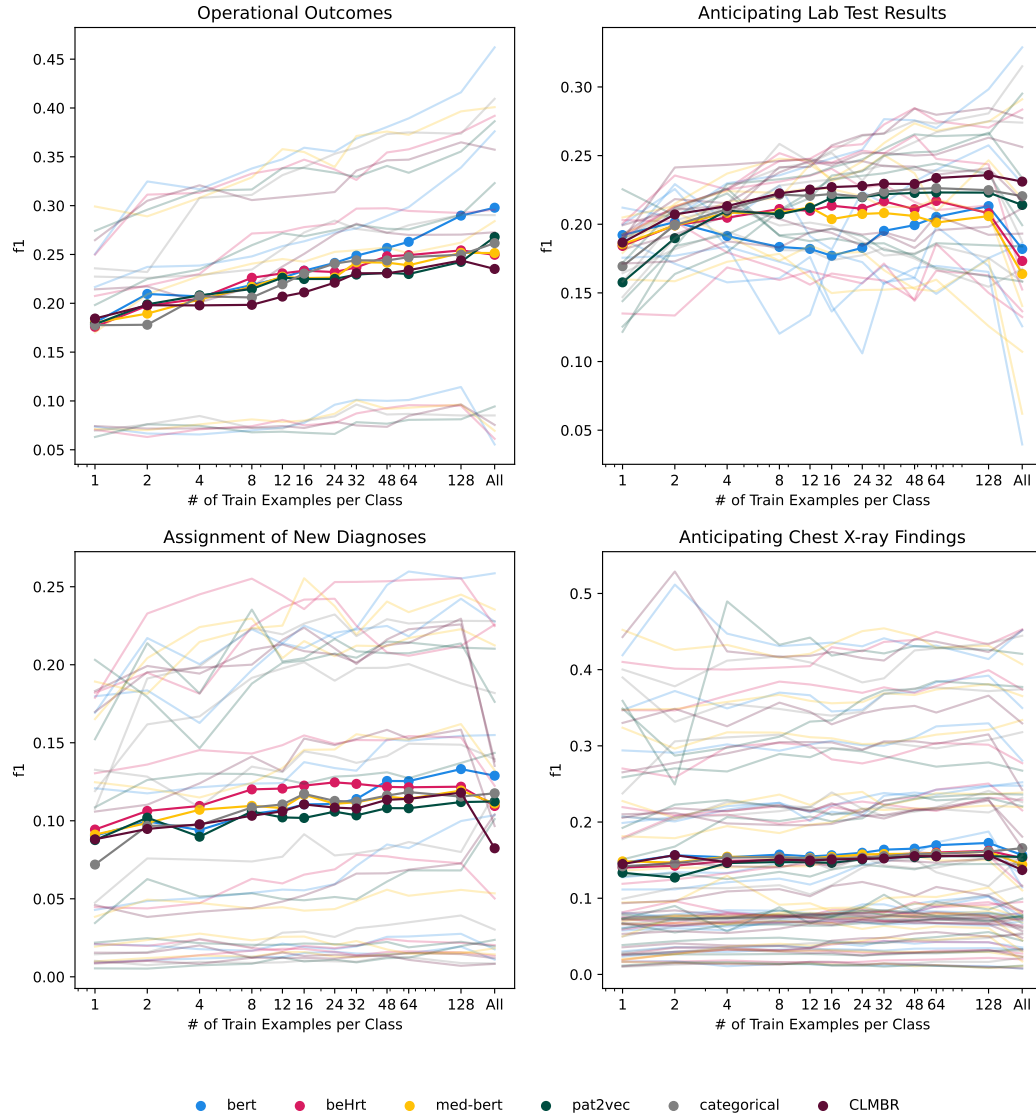


Figure 3: EHRSHOT plots, F1, Logistic Regression.

training dataset	model type	dimensionality	future_disease	mortality	rehospitalization
MIMIC-IV	BEHRT	240	0.38 (0.01)	0.18 (0.05)	0.51 (0.00)
Ki-Thrust	BERT	240	0.38 (0.01)	0.14 (0.04)	0.38 (0.00)
MIMIC-IV	BERT	240	0.39 (0.01)	0.16 (0.11)	0.51 (0.00)
Ki-Thrust	BERT	768	0.37 (0.00)	0.13 (0.02)	0.36 (0.00)
MIMIC-IV	BERT	768	0.21 (0.01)	0.01 (0.03)	0.39 (0.01)
CLMBR	CLMBR	768	0.28 (0.01)	0.16 (0.03)	0.48 (0.00)
None	Categorical	240	0.43 (0.00)	0.13 (0.00)	0.34 (0.00)
None	Categorical	640	0.42 (0.00)	0.25 (0.00)	0.50 (0.00)
MIMIC-IV	Med-BERT	240	0.37 (0.01)	0.20 (0.04)	0.51 (0.00)
Pat2Vec	Pat2Vec	10	0.19 (0.01)	0.07 (0.05)	0.39 (0.00)
Ki-Thrust	Pat2Vec	240	0.38 (0.00)	0.19 (0.00)	0.37 (0.00)
MIMIC-IV	Pat2Vec	240	0.28 (0.00)	0.21 (0.00)	0.45 (0.00)

Table 7: MIMIC results, F1, Logistic Regression.

training dataset	model type	dimensionality	future_disease	mortality	rehospitalization
Ki-Thrust	BERT	240	0.34 (0.01)	0.04 (0.01)	0.18 (0.01)
MIMIC-IV	BERT	240	0.29 (0.00)	0.05 (0.02)	0.16 (0.06)
Ki-Thrust	BERT	768	0.31 (0.00)	0.05 (0.01)	0.18 (0.01)
MIMIC-IV	BERT	768	0.29 (0.00)	0.05 (0.02)	0.16 (0.06)
None	Categorical	240	0.37 (0.00)	0.05 (0.01)	0.23 (0.00)
Pat2Vec	Pat2Vec	10	0.18 (0.00)	0.01 (0.01)	0.09 (0.07)
Ki-Thrust	Pat2Vec	240	0.18 (0.01)	0.01 (0.01)	0.13 (0.05)
MIMIC-IV	Pat2Vec	240	0.29 (0.00)	0.05 (0.00)	0.21 (0.00)

Table 8: KI-Thrust results, F1, Logistic Regression.

ablation	future_disease	mortality	rehospitalization
full	0.39 (0.01)	0.16 (0.11)	0.51 (0.00)
icd 10	0.41 (0.01)	0.17 (0.04)	0.39 (0.00)
categories	0.41 (0.01)	0.14 (0.02)	0.37 (0.00)
icd diagnosis	0.39 (0.01)	0.21 (0.06)	0.51 (0.00)
icd 10 diagnosis	0.40 (0.01)	0.16 (0.02)	0.38 (0.00)
diagnosis length 2	0.39 (0.01)	0.19 (0.02)	0.51 (0.00)
diagnosis length 3	0.39 (0.00)	0.20 (0.04)	0.51 (0.00)

Table 9: MIMIC ablation, F1, Logistic Regression.

ablation	future_disease	mortality	rehospitalization
icd 10	0.30 (0.00)	0.03 (0.02)	0.19 (0.02)
full mimic	0.29 (0.00)	0.05 (0.02)	0.16 (0.06)
icd diagnosis	0.30 (0.01)	0.03 (0.03)	0.18 (0.05)
full ki-thrust	0.34 (0.01)	0.04 (0.01)	0.18 (0.01)
icd 10 diagnosis	0.29 (0.01)	0.04 (0.02)	0.18 (0.01)
diagnosis length 3	0.25 (0.00)	0.04 (0.03)	0.17 (0.04)
diagnosis length 2	0.18 (0.01)	0.02 (0.01)	0.12 (0.06)

Table 10: KI-Thrust ablation, F1, Logistic Regression.

Logistic Regression - MCC

□

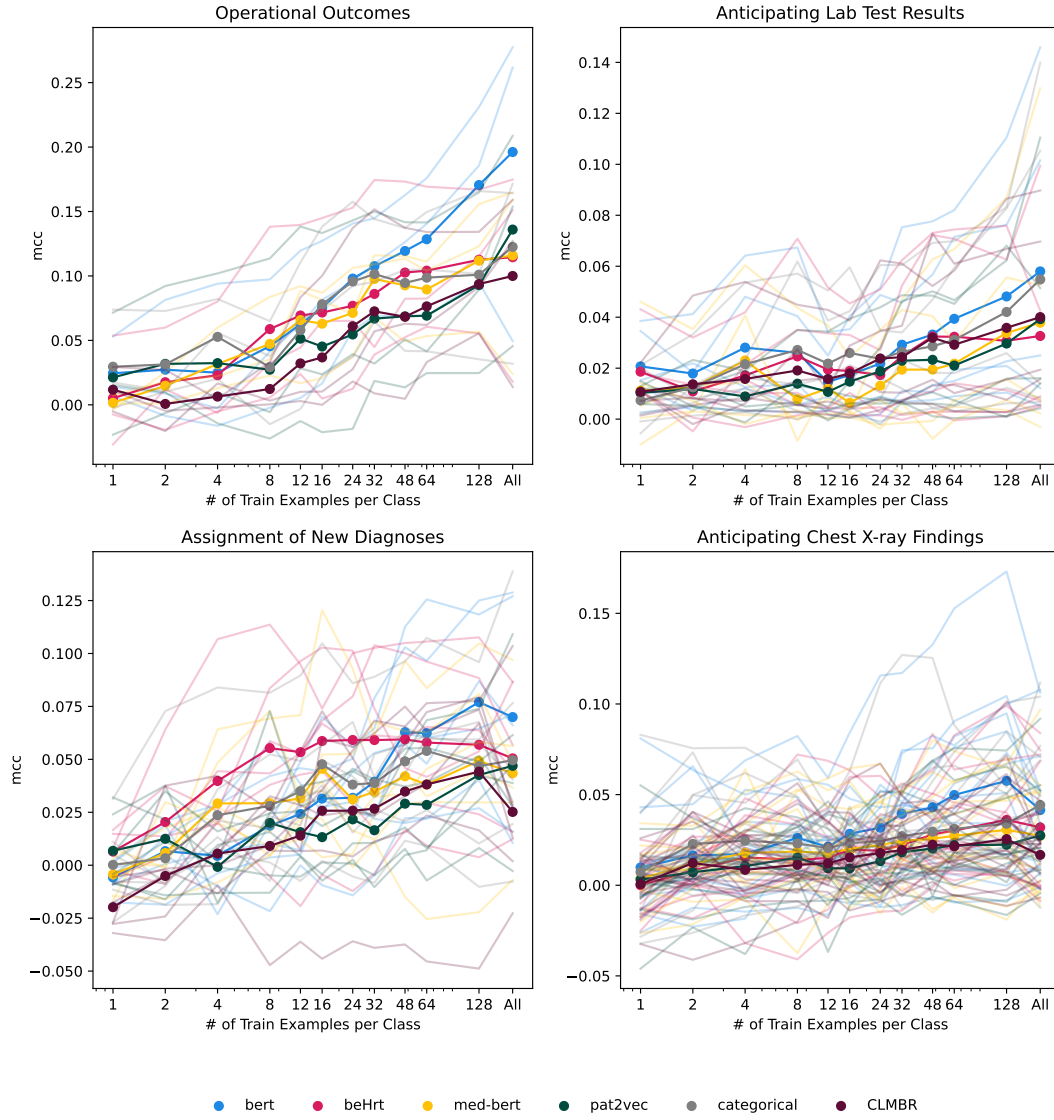


Figure 4: EHRSHOT plots, MCC, Logistic Regression.

training dataset	model type	dimensionality	future_disease	mortality	rehospitalization
MIMIC-IV	BEHRT	240	0.31 (0.01)	0.17 (0.04)	0.22 (0.00)
Ki-Thrust	BERT	240	0.31 (0.01)	0.09 (0.04)	0.13 (0.00)
MIMIC-IV	BERT	240	0.32 (0.01)	0.16 (0.12)	0.22 (0.00)
Ki-Thrust	BERT	768	0.29 (0.01)	0.08 (0.03)	0.11 (0.00)
MIMIC-IV	BERT	768	0.07 (0.00)	0.00 (0.00)	0.04 (0.01)
CLMBR	CLMBR	768	0.18 (0.01)	0.10 (0.05)	0.08 (0.01)
None	Categorical	240	0.37 (0.00)	0.07 (0.00)	0.13 (0.00)
None	Categorical	640	0.35 (0.00)	0.23 (0.01)	0.21 (0.00)
MIMIC-IV	Med-BERT	240	0.30 (0.01)	0.18 (0.03)	0.22 (0.00)
Pat2Vec	Pat2Vec	10	-0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)
Ki-Thrust	Pat2Vec	240	0.32 (0.00)	0.13 (0.00)	0.10 (0.00)
MIMIC-IV	Pat2Vec	240	0.16 (0.00)	0.17 (0.01)	0.14 (0.00)

Table 11: MIMIC results, MCC, Logistic Regression.

training dataset	model type	dimensionality	future_disease	mortality	rehospitalization
Ki-Thrust	BERT	240	0.25 (0.01)	0.09 (0.01)	0.08 (0.02)
MIMIC-IV	BERT	240	0.19 (0.00)	0.07 (0.02)	0.06 (0.04)
Ki-Thrust	BERT	768	0.21 (0.00)	0.07 (0.02)	0.09 (0.02)
MIMIC-IV	BERT	768	0.19 (0.00)	0.07 (0.02)	0.06 (0.04)
None	Categorical	240	0.29 (0.00)	0.09 (0.01)	0.15 (0.00)
Pat2Vec	Pat2Vec	10	0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
Ki-Thrust	Pat2Vec	240	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
MIMIC-IV	Pat2Vec	240	0.18 (0.00)	0.10 (0.00)	0.13 (0.00)

Table 12: KI-Thrust results, MCC, Logistic Regression.

ablation	future_disease	mortality	rehospitalization
full	0.32 (0.01)	0.16 (0.12)	0.22 (0.00)
icd 10	0.34 (0.01)	0.13 (0.04)	0.14 (0.00)
categories	0.34 (0.01)	0.09 (0.03)	0.14 (0.00)
icd diagnosis	0.32 (0.01)	0.20 (0.07)	0.24 (0.01)
icd 10 diagnosis	0.33 (0.01)	0.11 (0.02)	0.14 (0.01)
diagnosis length 2	0.32 (0.01)	0.17 (0.02)	0.24 (0.00)
diagnosis length 3	0.32 (0.00)	0.18 (0.04)	0.23 (0.00)

Table 13: MIMIC ablation, MCC, Logistic Regression.

ablation	future_disease	mortality	rehospitalization
icd 10	0.20 (0.00)	0.07 (0.04)	0.11 (0.04)
full mimic	0.19 (0.00)	0.07 (0.02)	0.06 (0.04)
icd diagnosis	0.20 (0.01)	0.05 (0.05)	0.11 (0.03)
full ki-thrust	0.25 (0.01)	0.09 (0.01)	0.08 (0.02)
icd 10 diagnosis	0.18 (0.00)	0.06 (0.02)	0.07 (0.02)
diagnosis length 3	0.11 (0.01)	0.05 (0.04)	0.07 (0.04)
diagnosis length 2	0.01 (0.00)	0.00 (0.01)	0.03 (0.02)

Table 14: KI-Thrust ablation, MCC, Logistic Regression.

Random Forest - F1

□

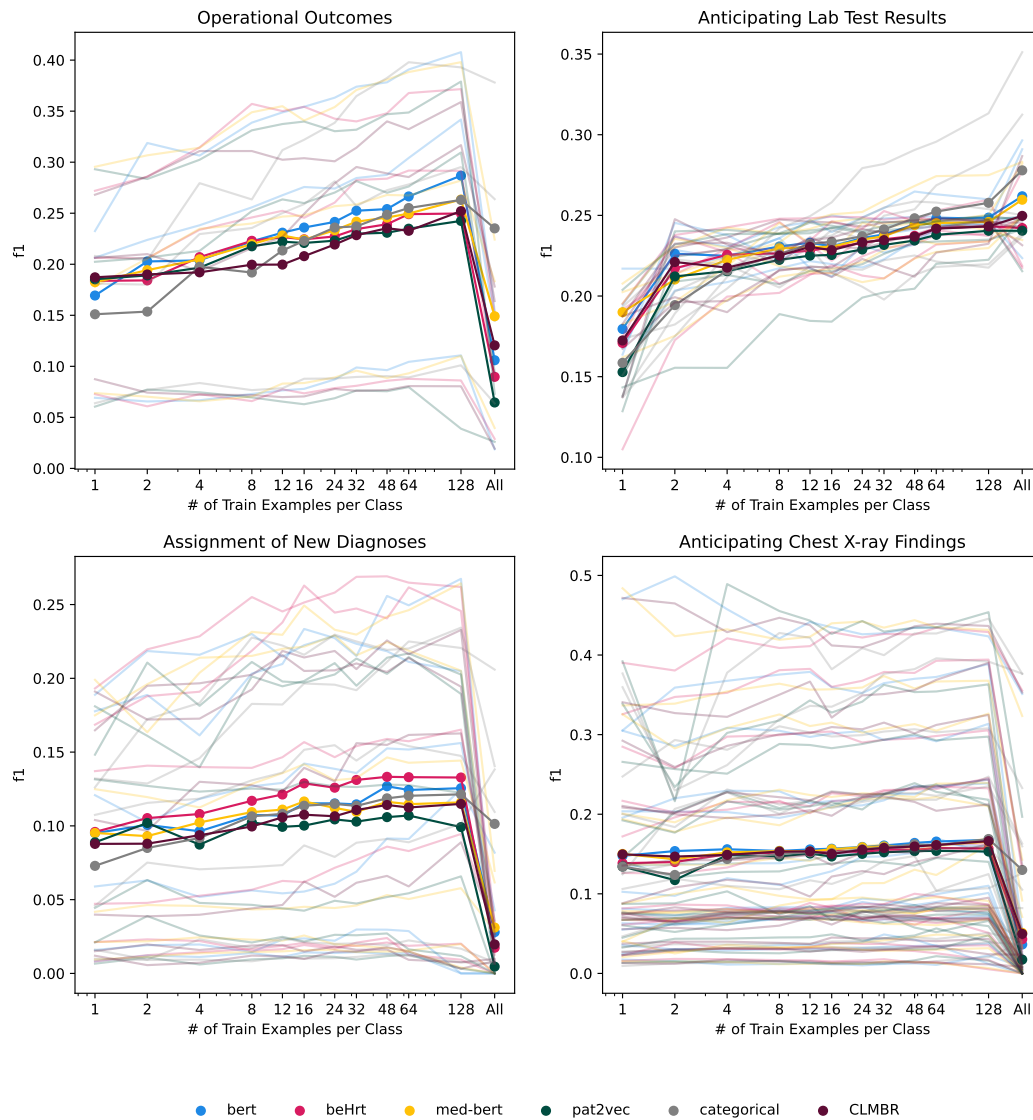


Figure 5: EHRSHOT plots, F1, Random Forest.

training dataset	model type	dimensionality	future_disease	mortality	rehospitalization
MIMIC-IV	BEHRT	240	0.21 (0.00)	0.00 (0.00)	0.23 (0.01)
Ki-Thrust	BERT	240	0.19 (0.00)	0.15 (0.01)	0.18 (0.01)
MIMIC-IV	BERT	240	0.18 (0.00)	0.01 (0.00)	0.21 (0.01)
Ki-Thrust	BERT	768	0.18 (0.00)	0.16 (0.00)	0.16 (0.01)
MIMIC-IV	BERT	768	0.06 (0.00)	0.00 (0.00)	0.03 (0.00)
CLMBR	CLMBR	768	0.21 (0.00)	0.00 (0.00)	0.41 (0.00)
None	Categorical	240	0.22 (0.01)	0.11 (0.00)	0.15 (0.01)
None	Categorical	640	0.26 (0.00)	0.00 (0.00)	0.28 (0.01)
MIMIC-IV	Med-BERT	240	0.23 (0.00)	0.00 (0.00)	0.25 (0.01)
Pat2Vec	Pat2Vec	10	0.02 (0.00)	0.00 (0.00)	0.04 (0.00)
Ki-Thrust	Pat2Vec	240	0.22 (0.00)	0.00 (0.00)	0.09 (0.00)
MIMIC-IV	Pat2Vec	240	0.07 (0.00)	0.00 (0.00)	0.05 (0.00)

Table 15: MIMIC results, F1, Random Forest.

training dataset	model type	dimensionality	future_disease	mortality	rehospitalization
Ki-Thrust	BERT	240	0.07 (0.00)	0.00 (0.00)	0.00 (0.00)
MIMIC-IV	BERT	240	0.08 (0.00)	0.00 (0.00)	0.00 (0.00)
Ki-Thrust	BERT	768	0.04 (0.00)	0.00 (0.00)	0.00 (0.00)
MIMIC-IV	BERT	768	0.08 (0.00)	0.00 (0.00)	0.00 (0.00)
None	Categorical	240	0.11 (0.00)	0.00 (0.00)	0.00 (0.00)
Pat2Vec	Pat2Vec	10	0.01 (0.00)	0.00 (0.00)	0.00 (0.00)
Ki-Thrust	Pat2Vec	240	0.01 (0.00)	0.00 (0.00)	0.00 (0.00)
MIMIC-IV	Pat2Vec	240	0.06 (0.00)	0.00 (0.00)	0.00 (0.00)

Table 16: KI-Thrust results, F1, Random Forest.

ablation	future_disease	mortality	rehospitalization
full	0.18 (0.00)	0.01 (0.00)	0.21 (0.01)
icd 10	0.19 (0.00)	0.16 (0.00)	0.22 (0.02)
categories	0.21 (0.00)	0.16 (0.01)	0.19 (0.01)
icd diagnosis	0.15 (0.00)	0.01 (0.00)	0.20 (0.01)
icd 10 diagnosis	0.20 (0.00)	0.16 (0.01)	0.21 (0.01)
diagnosis length 2	0.17 (0.00)	0.00 (0.00)	0.20 (0.00)
diagnosis length 3	0.15 (0.00)	0.00 (0.00)	0.19 (0.00)

Table 17: MIMIC ablation, F1, Random Forest.

ablation	future_disease	mortality	rehospitalization
icd 10	0.09 (0.00)	0.00 (0.00)	0.00 (0.00)
full mimic	0.08 (0.00)	0.00 (0.00)	0.00 (0.00)
icd diagnosis	0.09 (0.00)	0.00 (0.00)	0.00 (0.00)
full ki-thrust	0.07 (0.00)	0.00 (0.00)	0.00 (0.00)
icd 10 diagnosis	0.08 (0.00)	0.00 (0.00)	0.00 (0.00)
diagnosis length 3	0.17 (0.00)	0.00 (0.00)	0.03 (0.00)
diagnosis length 2	0.22 (0.00)	0.02 (0.00)	0.18 (0.00)

Table 18: KI-Thrust ablation, F1, Random Forest.

Random Forest - MCC

□

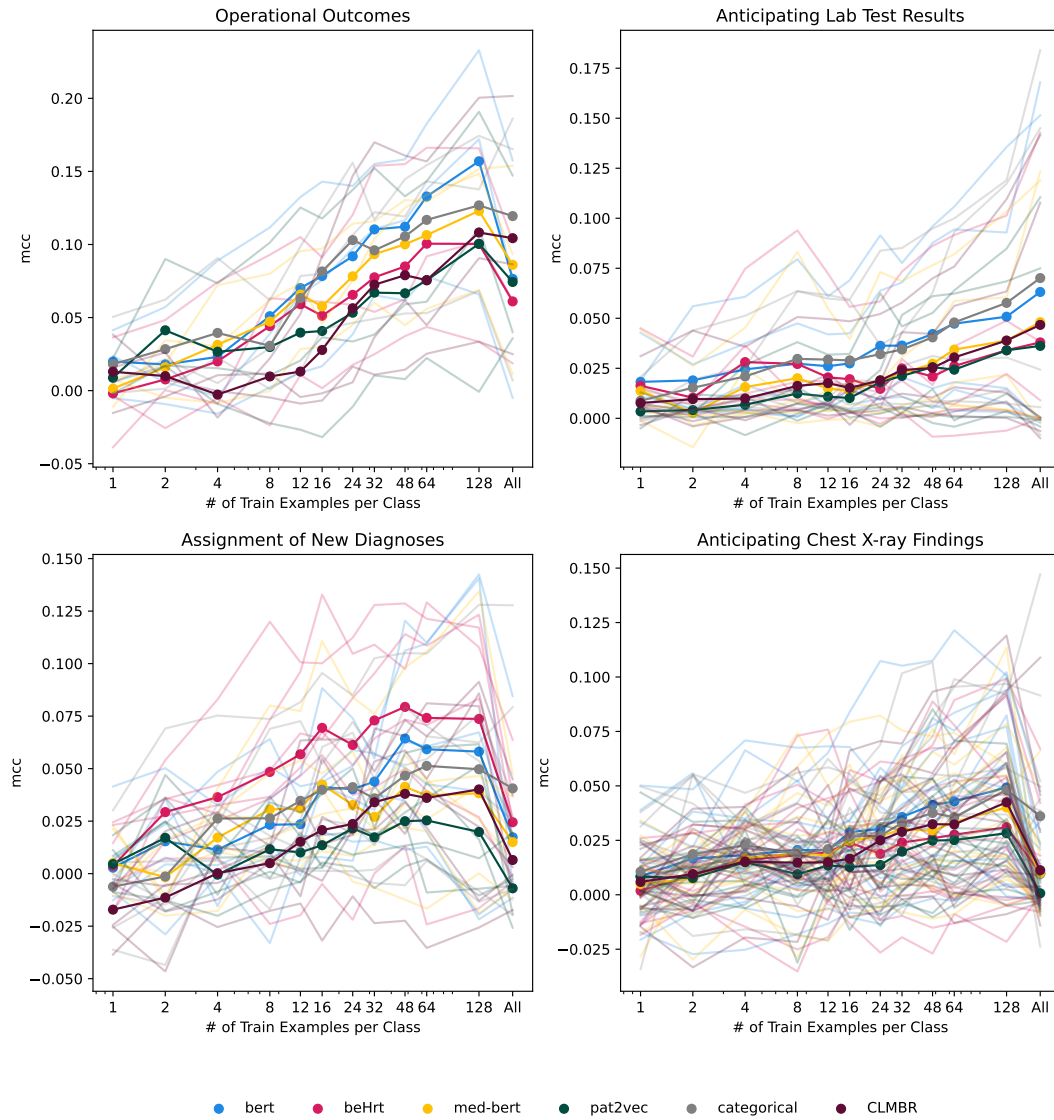


Figure 6: EHRSHOT plots, MCC, Random Forest.

training dataset	model type	dimensionality	future_disease	mortality	rehospitalization
MIMIC-IV	BEHRT	240	0.19 (0.00)	0.03 (0.02)	0.17 (0.00)
Ki-Thrust	BERT	240	-0.05 (0.00)	0.09 (0.01)	0.06 (0.01)
MIMIC-IV	BERT	240	0.16 (0.00)	0.03 (0.01)	0.19 (0.00)
Ki-Thrust	BERT	768	-0.07 (0.00)	0.10 (0.00)	0.06 (0.01)
MIMIC-IV	BERT	768	0.02 (0.00)	-0.00 (0.00)	0.07 (0.00)
CLMBR	CLMBR	768	0.06 (0.02)	0.01 (0.01)	0.02 (0.00)
None	Categorical	240	-0.02 (0.02)	0.02 (0.00)	0.12 (0.01)
None	Categorical	640	0.21 (0.00)	-0.00 (0.01)	0.18 (0.01)
MIMIC-IV	Med-BERT	240	0.20 (0.00)	0.02 (0.02)	0.18 (0.01)
Pat2Vec	Pat2Vec	10	-0.01 (0.00)	-0.00 (0.00)	0.08 (0.01)
Ki-Thrust	Pat2Vec	240	0.19 (0.00)	-0.00 (0.00)	0.12 (0.00)
MIMIC-IV	Pat2Vec	240	0.04 (0.00)	0.00 (0.00)	0.10 (0.00)

Table 19: MIMIC results, MCC, Random Forest.

training dataset	model type	dimensionality	future_disease	mortality	rehospitalization
Ki-Thrust	BERT	240	0.06 (0.00)	-0.00 (0.00)	0.00 (0.00)
MIMIC-IV	BERT	240	0.07 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Ki-Thrust	BERT	768	0.03 (0.00)	-0.00 (0.00)	0.00 (0.00)
MIMIC-IV	BERT	768	0.07 (0.00)	-0.00 (0.00)	-0.00 (0.00)
None	Categorical	240	0.09 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Pat2Vec	Pat2Vec	10	0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
Ki-Thrust	Pat2Vec	240	-0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
MIMIC-IV	Pat2Vec	240	0.04 (0.00)	-0.00 (0.00)	-0.00 (0.00)

Table 20: KI-Thrust results, MCC, Random Forest.

ablation	future_disease	mortality	rehospitalization
full	0.16 (0.00)	0.03 (0.01)	0.19 (0.00)
icd 10	-0.05 (0.00)	0.09 (0.00)	0.08 (0.01)
categories	-0.03 (0.00)	0.09 (0.01)	0.08 (0.01)
icd diagnosis	0.14 (0.00)	0.04 (0.01)	0.18 (0.01)
icd 10 diagnosis	-0.04 (0.00)	0.10 (0.01)	0.08 (0.01)
diagnosis length 2	0.14 (0.00)	-0.00 (0.00)	0.18 (0.00)
diagnosis length 3	0.13 (0.00)	-0.00 (0.00)	0.19 (0.00)

Table 21: MIMIC ablation, MCC, Random Forest.

ablation	future_disease	mortality	rehospitalization
icd 10	0.08 (0.00)	-0.00 (0.00)	-0.00 (0.00)
full mimic	0.07 (0.00)	-0.00 (0.00)	-0.00 (0.00)
icd diagnosis	0.07 (0.00)	-0.00 (0.00)	-0.00 (0.00)
full ki-thrust	0.06 (0.00)	-0.00 (0.00)	0.00 (0.00)
icd 10 diagnosis	0.07 (0.00)	-0.00 (0.00)	-0.00 (0.00)
diagnosis length 3	0.02 (0.00)	-0.00 (0.00)	-0.01 (0.00)
diagnosis length 2	0.01 (0.00)	0.01 (0.00)	0.07 (0.00)

Table 22: KI-Thrust ablation, MCC, Random Forest.

References

- [1] Jacob Devlin et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. NAACL-HLT 2019. Ed. by Jill Burstein, Christy Doran, and Thamar Solorio. Minneapolis, Minnesota: Association for Computational Linguistics, June 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.
- [2] Junyuan Shang et al. *Pre-training of Graph Augmented Transformers for Medication Recommendation*. Nov. 27, 2019. DOI: 10.48550/arXiv.1906.00346.