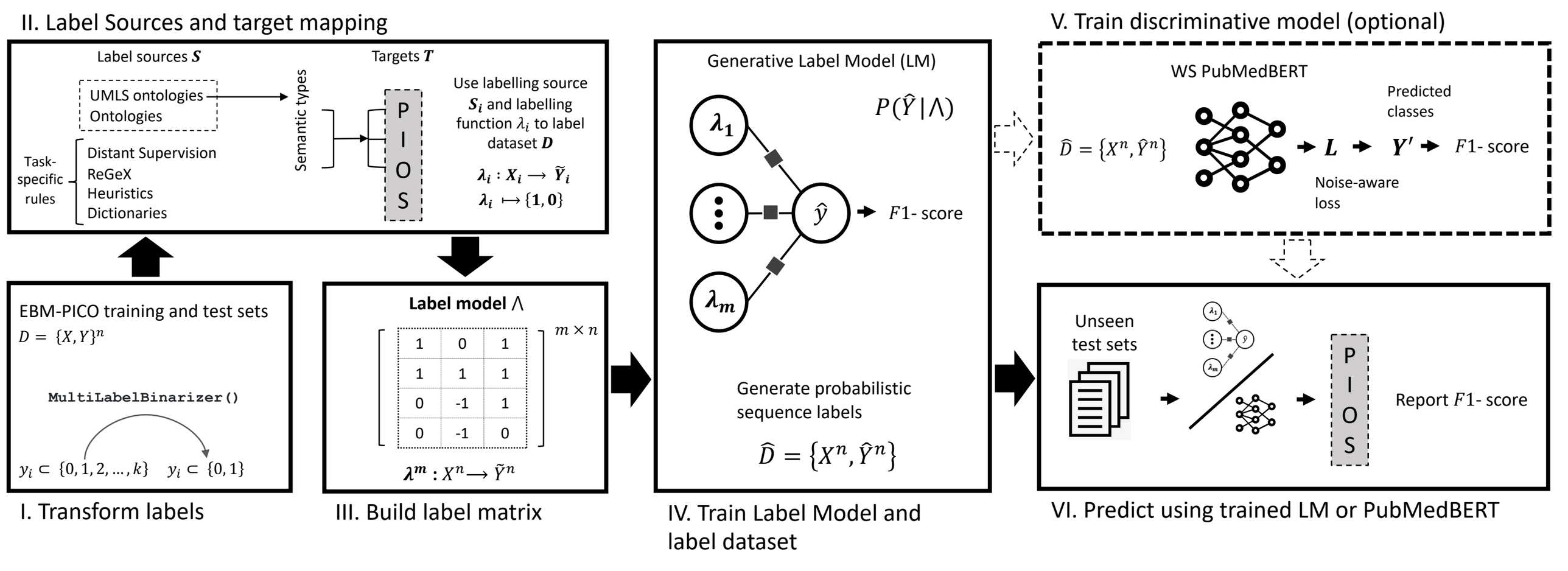
II. Label Sources and target mapping
Label sources S
Targets T
UMLS ontologies
Ontologies
Semantic types
Task-specific rules
Distant Supervision
ReGeX
Heuristics
Dictionaries
P
I
O
S
Use labelling source S_i and labelling function λ_i to label dataset D
$\lambda_i : X_i \longrightarrow \tilde{Y}_i$
$\lambda_i \longmapsto \{1, 0\}$
V. Train discriminative model (optional)
WS PubMedBERT
$\widehat{D} = \{X^n, \hat{Y}^n\}$
L
Y'
Predicted classes
$F1$- score
Noise-aware loss
EBM-PICO training and test sets
$D = \{X, Y\}^n$
MultiLabelBinarizer()
$y_i \subset \{0, 1, 2, \ldots, k\}$
$y_i \subset \{0, 1\}$
I. Transform labels
Label model Λ
$m \times n$
1 0 1
1 1 1
0 -1 1
0 -1 0
$\lambda^m : X^n \longrightarrow \tilde{Y}^n$
III. Build label matrix
Generative Label Model (LM)
λ_1
λ_m
$\hat{y}$
$P(\hat{Y} \mid \Lambda)$
$F1$- score
Generate probabilistic sequence labels
$\widehat{D} = \{X^n, \hat{Y}^n\}$
IV. Train Label Model and label dataset
Unseen test sets
P
I
O
S
Report $F1$- score
VI. Predict using trained LM or PubMedBERT