ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

# Beyond SMOTE: Evaluating large language models and mixture of experts for prediction of surgical site infections

## Supplementary File

### S1. Feature list for all datasets (Table S1)

Table S1 shows a list of the electronic medical record (EMR) features. Emphasis was given to data elements that are readily available. Because many of the features are quantity or duration amounts, null handling consisted of imputing 0 values. Null or missing class values were imputed with the word "none."

### S2. Relevant International Classification of Diseases version 10 (ICD-10) diagnosis codes for surgical site infections (SSI)-positive encounters

Table S2 shows a list of relevant ICD-10 Dx codes. These core codes include the T81.4XXX series (*e.g.*, T81.41XA for superficial incisional SSI), K65.0 (generalized [acute] peritonitis), K65.1 (peritoneal abscess), K68.11 (postprocedural infection of other intra-abdominal organ or structure), and O86.03 (infection of obstetric surgical wound, organ, and space site). While Table S2 presents a broader list of ICD-10 diagnosis codes, it is important to note that the presence of general symptoms like R10.0 (acute unspecified abdominal pain) alone was not used to classify an encounter as SSI positive. These broader codes, such as R10.0, are included in Table S2 because they were found to be consistently present and frequently associated with the final coding of encounters that were definitively confirmed as SSI positive by the more specific diagnostic codes. This reflects the nuanced real-world clinical documentation where general symptoms often accompany or are documented alongside specific diagnoses of SSIs, particularly following procedures like colon surgery or hysterectomy.

### S3. Similarity threshold justification

A non-SSI case $e'_j$ is retained if $\text{Sim}(e_i, e'_j) \geq 0.70$ for at least one $e_i \in E_{SSI}$. We set the similarity threshold at 0.70 to retain non-SSI encounters, guided by prior applications in clinical similarity modeling where thresholds between 0.65 and 0.75 are considered effective for balancing specificity and coverage. Similarity metrics such as Jaccard and cosine distance have been widely used in patient similarity and clinical data matching tasks (*e.g.*, de-duplication of EMR notes using Jaccard at 0.7) and in case-based reasoning systems, achieving favorable sensitivity-accuracy trade-offs around 0.70. Thresholds much above 0.75 can exclude clinically analogous control exemplars, reducing cohort diversity, while lower thresholds (below 06.5) may inject heterogeneous majority-class contributors. Manual review of sample pairs across thresholds indicated that a 0.70 threshold preserved clinical plausibility and interpretability. The threshold represents a tunable compromise with transparent operationalization suitable for clinical deployments. With this threshold, the 151,733 encounters were reduced to 15,339 (5,282 SSI positive, ~34% and 10,057 clinically similar non-SSI).

### S4. Model imbalance configurations

We evaluated training sets across the following target-class proportions (positive: negative): 5:95, 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20, 90:10, 95:5, and the organic 34:66. For MoE, experts specialized on subsets (*e.g.*, 5–95, 50–50, and 95–5) and variants spanning 10–90 up through 90–10.

### S5. PRISMA-style flow diagram

A PRISMA-style flow diagram is adapted for this modeling study to ensure transparency in encounter selection and dataset construction.

**Table S1. Feature descriptions for the prediction of SSIs**

| Feature name | Feature description |
|---|---|
| SSI | Target (0/1; 1=SSI positive) |
| LOS | Length of stay for current encounter |
| admissionType_key | Type of admission (elective, emergency, trauma, *etc.*) |
| DaysBtwPrevVisit | Days between last and current visit |
| QtyPriorIPvisits | Number of prior inpatient visits |
| VolProcs | Volume of procedures, current encounter |
| VolProcs_SSI | Volume of SSI-specific procedures, current encounter |
| VolProcs_Prev | Volume of procedures, previous encounter |
| VolProcs_SSI_Prev | Volume of SSI-specific procedures, previous encounter |
| proc1 | First SSI-specific PCS code, current encounter |
| proc2 | Second pSSI-specific PCS code, current encounter |
| proc1Prev | First SSI-specific PCS code, previous encounter |
| proc2Prev | Second pSSI-specific PCS code, previous encounter |
| Age | Age |
| Sex | Patient gender |
| Race | Patient race |
| drg | Surgical encounter (if available) Diagnosis-Related Grouping Code (DRG) |
| AcuityDRG | Surgical DRG acuity, *e.g.*, comorbid condition, major comorbid condition, or neither |
| maxAcuityDx | Maximum Dx code acuity, *e.g.*, comorbid condition, major comorbid condition, or neither |
| BMI | Body mass index (if available) |
| preOrHours | Hours between registration and 1st OR procedure |
| AnesMinsMaxSurgDur | Anesthesia duration (minutes) for the longest surgery |
| IntubateMins | Duration of intubation in OR |
| CardioPulClampMins | Duration of cardiopulmonary clamp (minutes) |
| dischargeDisposition | Discharge disposition |
| SurgHospSvc | Hospital service |
| SurgSpecialty | Surgical specialty |
| ClinHandOff | Clinical handoff in/after surgery (0/1; 1=True) |
| ANEPostOp | Anesthesia post op power plan (0/1; 1=True) |
| orQty | Quantity of OR records (surgeries) |
| orProcQty | Quantity of OR procedures |
| orProcCancelQty | Quantity of canceled OR procedures |
| maxOrProcMinutes | Longest OR procedure duration (minutes) |
| maxAnesDurationMins | Longest OR anesthesia duration (minutes) |

Abbreviations: BMI: Body mass index; SSI: Surgical site infections.

**Table S2. ICD-10 diagnosis codes frequently associated with colon or abdominal hysterectomy-related SSI encounters**

| Code | Description |
|---|---|
| T81.41XA | Infection following a procedure, superficial incisional surgical site, initial encounter |
| T81.43XA | Infection following a procedure, organ and space surgical site, initial encounter |
| T81.44XA | Sepsis following a procedure |
| T81.49XA | Infection following a procedure |
| T81.32XA | Disruption of internal operation (surgical) wound, not elsewhere classified, initial encounter |
| T81.9XXA | Unspecified complication of procedure, initial encounter |
| K65.0 | Generalized (acute) peritonitis |
| K65.1 | Peritoneal abscess |
| K65.9 | Peritonitis, unspecified |
| K68.11 | Postprocedural retroperitoneal abscess |
| K91.89 | Other postprocedural complications and disorders of digestive system |
| Z98.890 | Other specified postprocedural states |
| L02.211 | Cutaneous abscess of abdominal wall |
| R10.0 | Acute unspecified abdominal pain |
| O86.03 | Infection of obstetric surgical wound, organ, and space site |

Abbreviations: ICD-10: International Classification of Diseases version 10; SSI: Surgical site infections.

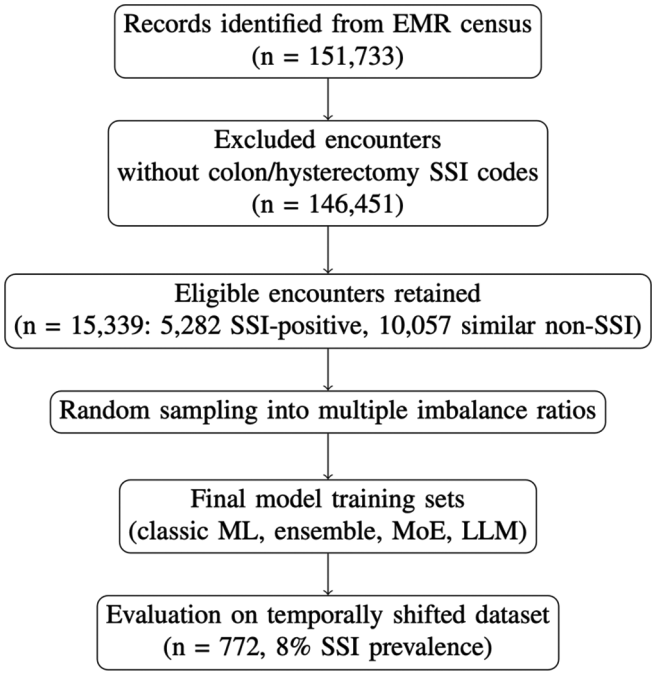


**Figure S1.** PRISMA-style flow diagram adapted for EMR cohort selection and modeling pipeline
Abbreviations: EMR: Electronic medical record; LLM: Large language model; ML: Machine learning; MoE: Mixture of Experts; SSI: Surgical site infection.