Appendix I – Electronic Health Record Features Used for Classification

Variable name	Description	Data Type
Static Features		
first_admit_age	Age at hospital admission	Continuous
admission_type	Type of hospital admission	Categorical
admission_location	Location of hospital admission	Categorical
discharge_location	Discharge location	Categorical
Insurance	Type of insurance	Categorical
Religion	Religious affiliation, if any	Categorical
marital_status	Marital status in chart	Categorical
Ethnicity	Patient-reported ethnicity	Categorical
Gender	Patient gender	Categorical (Male/Female)
Antib	Text search of admission note identifies antibiotic use	Binary (1 - yes, 0 - no)
Kidney	Text search of admission note identifies comorbid kidney disease	Binary (1 - yes, 0 - no)
Cdiff	Text search of admission note identifies previous CDI	Binary (1 - yes, 0 - no)
Diab	Text search of admission note identifies comorbid diabetes	Binary (1 - yes, 0 - no)
Dynamic Feature Sets		
Day of hospital stay	Single feature	Integer
Laboratory items	HEMATOCRIT HEMOGLOBIN MCH MCHC MCV PLATELET COUNT	Continuous, normalized to 0-1 interval for algorithm

	RDW RED BLOOD CELLS WHITE BLOOD CELLS ANION GAP BICARBONATE CALCIUM, TOTAL CHLORIDE CREATININE GLUCOSE MAGNESIUM PHOSPHATE POTASSIUM SODIUM UREA NITROGEN	
Chart items	Arterial BP [Systolic] Arterial BP Mean" CVP Eye Opening Heart Rate Motor Response NBP [Systolic] NBP Mean Respiratory Rate SpO2 Temperature F Temperature F (calc) Verbal Response ABP Alarm [Low] HR Alarm [Low] NBP Alarm [Low] Resp Alarm [Low] SpO2 Alarm [Low] Arterial BP [Diastolic] NBP [Diastolic] ABP Alarm [High] HR Alarm [High] NBP Alarm [High] Resp Alarm [High] SpO2 Alarm [High] Heart Rate Heart rate Alarm - High Heart Rate Alarm - Low Arterial Blood Pressure systolic Arterial Blood Pressure	

	diastolic Arterial Blood Pressure mean Arterial Blood Pressure Alarm - Low Arterial Blood Pressure Alarm - High Central Venous Pressure Non Invasive Blood Pressure systolic Non Invasive Blood Pressure diastolic Non Invasive Blood Pressure mean Respiratory Rate O2 saturation pulse oximetry GCS - Eye Opening Non-Invasive Blood Pressure Alarm - High Non-Invasive Blood Pressure Alarm - Low Temperature Fahrenheit O2 Saturation Pulse oximetry Alarm - High O2 Saturation Pulse oximetry Alarm - Low GCS - Verbal Response GCS - Motor Response" Resp Alarm - High Resp Alarm – Low	
Services	*All service designations coded as binary variables (1 - patient assigned to service, 0 - all others) Each patient is assigned to one service.*	
	Cardiac Medical (CMED)	Cardiac-related, non-surgical
	Cardiac Surgery (CSURG)	Cardiac-related, surgical
	Dental (DENT)	Dental/jaw-related
	Ear, Nose, and Throat (ENT)	Conditions affecting the ear, nose, or throat
	Genitourinary (GU)	Urinary system/reproductive organs

	Gynecological (GYN)	Female reproductive systems and breasts
	Medical (MED)	General service for internal medicine
	Newborn (NB)	Infants born at the hospital
	Newborn baby (NBB)	Infants born at the hospital
	Neurological Medical (NMED)	Non-surgical, relating to the brain
	Neurologic Surgery (NSURG)	Surgical, relating to the brain
	Obstetrics (OBS)	Concerned with childbirth and the care of women giving birth
	Orthopaedic (ORTHO)	Surgical, relating to the musculoskeletal system
	Orthopaedic medicine (OMED)	non-surgical, relating to musculoskeletal system
	Plastic surgery (PSURG)	Surgical, restoration/reconstruction of the human body (including cosmetic or aesthetic)
	Psychiatric (PSYCH)	Mental disorders relating to mood, behaviour, cognition, or perceptions
	Surgical (SURG)	General surgical service not classified elsewhere
	Trauma (TRAUM)	injury or damage caused by physical harm from an external source

	Thoracic surgery (TSURG)	Surgery on the thorax, located between the neck and the abdomen
	Vascular surgery (VSURG)	Surgery related to the circulatory system
Medications Binary features recording whether a drug was prescribed that day.	high_antib: Antibiotics associated with greater CDI risk	Identified by searching drug name record for antibiotic names: o Cephalosporins o Fluoroquinolones o Macrolides o Penicillins o Sulfonamides o Tetracyclines
	antib: Any antibiotic use recorded in chart	Identified by searching drug names record for all antibiotics
	h2: Any H2 antagonist use recorded in chart	Identified by searching drug names record for all H2 antagonists.
	ppi: Any Proton Pump Inhibitor Use recorded in chart	Identified by searching drug names record for all proton pump inhibitors.
Microbiology items Binary features recording whether microbiology record includes a positive culture for this organism.	YEAST STAPHYLOCOCCUS COAGULASE NEGATIVE STAPH AUREUS COAG + GRAM NEGATIVE ROD(S) PSEUDOMONAS AERUGINOSA KLEBSIELLA PNEUMONIAE ENTEROCOCCUS SP. ESCHERICHIA COLI PROTEUS MIRABILIS ENTEROBACTER CLOACAE KLEBSIELLA OXYTOCA	Chart time records the record date.

	ACINETOBACTER BAUMANNII SERRATIA MARCESCENS ENTEROBACTER AEROGENES	
Temporal Features (constructed for each dynamic feature)		
	Feature Description	Formulation
Trend-Based x_i	Length of recordings	n
	Average of recordings	$\frac{1}{n}\sum_{i=1}^{n} x_i$
	Linear weighted average	$\frac{2}{n(n+1)}\sum_{i=1}^{n} i x_i$
	Quadratic weighted average	$\frac{6}{n(n+1)(2n+1)}\sum_{i=1}^{n} i^2 x_i$
	Standard deviation	σ
	Maximum recording	$\max_i x_i$
	Normalized maximum location	$\frac{1}{n} f(max_i\ x_i)$
	Minimum recording	$\min_i x_i$
	Normalized minimum location	$\frac{1}{n} f(min_i\ x_i)$
	Normalized first record location	$\frac{1}{n} f(x_1)$
	Normalized last record location	$\frac{1}{n} f(x_n)$

Fluctuation-Based	Mean absolute difference	$\frac{1}{n}\sum_{i=1}^{n-1} \mid x_i - x_{i+1} \mid$
	Number of increase patterns	$\frac{1}{n}\sum_{i=1}^{n-1} \mathbb{1}\left((x_i - x_{i+1}) > 0\right)$
	Number of decrease patterns	$\frac{1}{n}\sum_{i=1}^{n-1} \mathbb{1}\left((x_i - x_{i+1}) < 0\right)$
	Ratio of change in direction	$\frac{S(inc, dec) \cup S(inc, dec)}{n-1}$
Sparsity-based	Measurement frequency	$\frac{n}{los}$
	Proportion of missing values	$\frac{los - n}{los}$