In [20]:
```
import pandas as pd
import numpy as np

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer
from sklearn.linear_model import LogisticRegression, LogisticRegressionCV
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_sco
from sklearn.utils import resample, shuffle

from xgboost import XGBClassifier
import matplotlib.pyplot as plt
%config InlineBackend.figure_format='svg'
plt.rcParams.update({
    'text.usetex':False
})
import seaborn as sns

import warnings
warnings.filterwarnings("ignore", category=UserWarning)
warnings.filterwarnings("ignore", category=FutureWarning)
warnings.filterwarnings("ignore", category=RuntimeWarning)
```

Loading the dataset, pre-processing, and analysing the data

In [21]:
```
cohort_data = pd.read_csv('cohort_data.csv')
cohort_data
```

Out[21]:

	subject_id	hadm_id	icustay_id	gender	age	icu_los_hours	min_wbc	max_wbc
0	28162	194362	285686	F	18	49.0	NaN	NaN
1	22190	166880	290052	F	18	117.0	0.0	0.0
2	7717	159770	260370	F	18	34.0	9.0	9.0
3	69145	128969	213687	F	19	82.0	NaN	NaN
4	68035	145990	288524	F	19	22.0	NaN	NaN
...	...	...	...	...	...	...	...	...
29278	26644	170124	296631	M	89	0.0	0.0	0.0
29279	21296	172695	266878	M	89	18.0	NaN	NaN
29280	27515	123997	232528	M	89	93.0	NaN	NaN
29281	23531	193022	234645	M	89	234.0	NaN	NaN
29282	23248	170106	268726	M	89	134.0	NaN	NaN

29283 rows × 92 columns

In [22]:
```
# Data available in most of the patients:
df = cohort_data.dropna(subset=['icu_los_hours', 'min_hemoglobin', 'min_hematocr

# Drop sparse columns (missing values for many ICU entries)
df = df.dropna(axis=1, how='any')
df
```

Out[22]:

	subject_id	hadm_id	icustay_id	gender	age	icu_los_hours	min_hemoglobin	n
1	22190	166880	290052	F	18	117.0	9.4	
2	7717	159770	260370	F	18	34.0	11.6	
3	69145	128969	213687	F	19	82.0	10.8	
5	42842	162017	267868	F	20	403.0	4.6	
6	88518	158955	224060	F	20	15.0	12.1	
...	...	...	...	...	...	...	...	
29268	26701	189710	242094	M	77	310.0	11.9	
29275	19609	155814	274163	M	84	27.0	11.5	
29277	28848	137939	219453	M	88	277.0	12.1	
29280	27515	123997	232528	M	89	93.0	12.9	
29281	23531	193022	234645	M	89	234.0	11.7	

14899 rows × 59 columns

Summary of the filtered data

In [23]:
```
print(df.info())        # Data types and non-null counts
print(df.describe())    # Summary stats for numeric columns
print(df.describe(include='object'))
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 14899 entries, 1 to 29281
Data columns (total 59 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   subject_id           14899 non-null  int64  
 1   hadm_id              14899 non-null  int64  
 2   icustay_id           14899 non-null  int64  
 3   gender               14899 non-null  object 
 4   age                  14899 non-null  int64  
 5   icu_los_hours        14899 non-null  float64
 6   min_hemoglobin       14899 non-null  float64
 7   max_hemoglobin       14899 non-null  float64
 8   mean_hemoglobin      14899 non-null  float64
 9   last_hemoglobin      14899 non-null  float64
 10  min_hematocrit       14899 non-null  float64
 11  max_hematocrit       14899 non-null  float64
 12  mean_hematocrit      14899 non-null  float64
 13  last_hematocrit      14899 non-null  float64
 14  min_platelet count   14899 non-null  float64
 15  max_platelet count   14899 non-null  float64
 16  mean_platelet count  14899 non-null  float64
 17  last_platelet count  14899 non-null  float64
 18  min_sodium           14899 non-null  float64
 19  max_sodium           14899 non-null  float64
 20  mean_sodium          14899 non-null  float64
 21  last_sodium          14899 non-null  float64
 22  min_potassium        14899 non-null  float64
 23  max_potassium        14899 non-null  float64
 24  mean_potassium       14899 non-null  float64
 25  last_potassium       14899 non-null  float64
 26  min_chloride         14899 non-null  float64
 27  max_chloride         14899 non-null  float64
 28  mean_chloride        14899 non-null  float64
 29  last_chloride        14899 non-null  float64
 30  min_bicarbonate      14899 non-null  float64
 31  max_bicarbonate      14899 non-null  float64
 32  mean_bicarbonate     14899 non-null  float64
 33  last_bicarbonate     14899 non-null  float64
 34  min_glucose          14899 non-null  float64
 35  max_glucose          14899 non-null  float64
 36  mean_glucose         14899 non-null  float64
 37  last_glucose         14899 non-null  float64
 38  min_creatinine       14899 non-null  float64
 39  max_creatinine       14899 non-null  float64
 40  mean_creatinine      14899 non-null  float64
 41  last_creatinine      14899 non-null  float64
 42  min_urea nitrogen    14899 non-null  float64
 43  max_urea nitrogen    14899 non-null  float64
 44  mean_urea nitrogen   14899 non-null  float64
 45  last_urea nitrogen   14899 non-null  float64
 46  min_ph               14899 non-null  float64
 47  max_ph               14899 non-null  float64
 48  mean_ph              14899 non-null  float64
 49  last_ph              14899 non-null  float64
 50  min_po2              14899 non-null  float64
 51  max_po2              14899 non-null  float64
 52  mean_po2             14899 non-null  float64
 53  last_po2             14899 non-null  float64
 54  min_pco2             14899 non-null  float64
```

```
 55  max_pco2            14899 non-null  float64
 56  mean_pco2           14899 non-null  float64
 57  last_pco2           14899 non-null  float64
 58  target              14899 non-null  int64
dtypes: float64(53), int64(5), object(1)
memory usage: 6.8+ MB
None
         subject_id        hadm_id     icustay_id           age  \
count  14899.000000   14899.000000   14899.000000  14899.000000
mean   37907.054702  150196.574669  250188.808913     63.125780
std    29274.311419   28883.645108   28830.430269     16.955309
min        3.000000  100009.000000  200003.000000     18.000000
25%    13752.000000  124954.000000  225343.500000     53.000000
50%    27555.000000  150509.000000  250142.000000     65.000000
75%    62414.000000  175412.500000  275221.000000     77.000000
max    99995.000000  199998.000000  299998.000000     89.000000

       icu_los_hours  min_hemoglobin  max_hemoglobin  mean_hemoglobin  \
count   14899.000000    14899.000000    14899.000000     14899.000000
mean      118.738103       10.043016       11.460588        10.720599
std       156.972040        2.009306        1.900922         1.734816
min         0.000000        0.000000        5.700000         5.100000
25%        36.000000        8.600000       10.200000         9.500000
50%        66.000000        9.900000       11.300000        10.500000
75%       127.000000       11.350000       12.500000        11.800000
max      3694.000000       19.600000      108.000000        44.400000

       last_hemoglobin  min_hematocrit  ...       last_ph       min_po2  \
count     14899.000000    14899.000000  ...  14899.000000  14899.000000
mean         10.787435       29.869971  ...      7.292823    104.430700
std           1.665622        5.540298  ...      0.456273     59.268979
min           2.200000        2.500000  ...      5.000000      0.000000
25%           9.600000       26.000000  ...      7.350000     69.000000
50%          10.600000       29.400000  ...      7.390000     92.000000
75%          11.800000       33.400000  ...      7.430000    127.000000
max          19.600000       61.800000  ...      9.000000    617.000000

            max_po2      mean_po2      last_po2      min_pco2      max_pco2  \
count  14899.000000  14899.000000  14899.000000  14899.000000  14899.000000
mean     242.690382    161.593390    123.401705     36.446406     46.049534
std      133.273513     74.002005     61.693503      7.705479     10.510201
min       14.000000     14.000000     14.000000      0.000000     10.000000
25%      128.000000    105.000000     85.000000     32.000000     40.000000
50%      214.000000    151.000000    110.000000     36.000000     45.000000
75%      354.000000    207.136364    149.000000     40.000000     51.000000
max      775.000000    617.000000    617.000000    108.000000    154.000000

          mean_pco2     last_pco2        target
count  14899.000000  14899.000000  14899.000000
mean      40.891639     40.327740      0.067656
std        7.600651      7.820584      0.251162
min       10.000000      8.000000      0.000000
25%       36.666667     36.000000      0.000000
50%       40.333333     40.000000      0.000000
75%       44.000000     44.000000      0.000000
max      108.000000    126.000000      1.000000

[8 rows x 58 columns]
       gender
count   14899
```

```
unique      2
top         M
freq     8948
```

In [24]:
```
df = df.drop('subject_id', axis=1)
df = df.drop('hadm_id', axis=1)
df = df.drop('icustay_id', axis=1)
df = df.drop('gender', axis=1)
df
```

Out[24]:

	age	icu_los_hours	min_hemoglobin	max_hemoglobin	mean_hemoglobin	last_l
1	18	117.0	9.4	11.9	10.266667	
2	18	34.0	11.6	12.4	12.000000	
3	19	82.0	10.8	10.9	10.850000	
5	20	403.0	4.6	10.8	8.522222	
6	20	15.0	12.1	12.1	12.100000	
...	...	...	...	...	...	
29268	77	310.0	11.9	12.1	12.000000	
29275	84	27.0	11.5	11.5	11.500000	
29277	88	277.0	12.1	12.1	12.100000	
29280	89	93.0	12.9	12.9	12.900000	
29281	89	234.0	11.7	12.6	12.150000	

14899 rows × 55 columns

In [25]:
```
dataset = np.array(df)
y = np.array(df['target'])
X = np.array(df.drop('target', axis=1))
print(np.shape(dataset))
print(np.shape(X))
print(np.shape(y))

# Random shuffle and split 70-30 into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_
print(f'% Readmissions in Train: {np.mean(y_train) * 100}')
print(f'% Readmissions in Test: {np.mean(y_test) * 100}')
```

```
(14899, 55)
(14899, 54)
(14899,)
% Readmissions in Train: 6.66410969412216
% Readmissions in Test: 7.0022371364653235
% Readmissions in Train: 6.66410969412216
% Readmissions in Test: 7.0022371364653235
```

XGBoost

In [28]:
```
import numpy as np
import optuna
```

```
import xgboost as xgb
from sklearn.metrics import roc_auc_score
from sklearn.model_selection import StratifiedKFold

RANDOM_STATE = 229
cv = StratifiedKFold(n_splits=10, shuffle=True, random_state=RANDOM_STATE)


def objective(trial):
    params = {
        "objective": "binary:logistic",
        "eval_metric": "auc",
        "use_label_encoder": False,
        "random_state": RANDOM_STATE,
        "n_jobs": -1,

        # Search around your known good values
        "learning_rate": trial.suggest_float("learning_rate", 0.01, 0.1, log=Tru
        "max_depth": trial.suggest_int("max_depth", 4, 12),
        "max_delta_step": trial.suggest_int("max_delta_step", 0, 5),
        "max_leaves": trial.suggest_int("max_leaves", 2, 10),
        "min_child_weight": trial.suggest_float("min_child_weight", 1, 8),
        "n_estimators": trial.suggest_int("n_estimators", 400, 1000),
        "alpha": trial.suggest_float("alpha", 0.1, 1.0),
        "lambda": trial.suggest_float("lambda", 0.8, 1.5),
        "scale_pos_weight": trial.suggest_float("scale_pos_weight", 0.5, 1.5),
        "subsample": trial.suggest_float("subsample", 0.7, 1.0),
    }

    aucs = []
    for train_idx, valid_idx in cv.split(X_train, y_train):
        X_tr, X_val = X_train[train_idx], X_train[valid_idx]
        y_tr, y_val = y_train[train_idx], y_train[valid_idx]

        model = xgb.XGBClassifier(**params)
        model.fit(X_tr, y_tr, eval_set=[(X_val, y_val)], verbose=False)
        y_pred = model.predict_proba(X_val)[:, 1]
        aucs.append(roc_auc_score(y_val, y_pred))

    return np.mean(aucs)

study = optuna.create_study(direction="maximize", sampler=optuna.samplers.TPESam
study.optimize(objective, n_trials=50, show_progress_bar=True)

print("\nBest Parameters Found:")
for k, v in study.best_params.items():
    print(f"{k}: {v}")

print("\nBest Cross-Validation AUC:", study.best_value)

best_params = study.best_params
best_params.update({
    "objective": "binary:logistic",
    "eval_metric": "auc",
    "use_label_encoder": False,
    "random_state": RANDOM_STATE,
    "n_jobs": -1,
})
```

```
final_model = xgb.XGBClassifier(**best_params)
final_model.fit(X_train, y_train)
```

```
[I 2025-11-11 18:24:27,143] A new study created in memory with name: no-name-d4e9
e242-aeb0-4e0a-a32b-75b0028f60e8
  0%|          | 0/50 [00:00<?, ?it/s]
```

```
[I 2025-11-11 18:24:35,429] Trial 0 finished with value: 0.6585032460759072 and p
arameters: {'learning_rate': 0.01250331001543945, 'max_depth': 10, 'max_delta_ste
p': 4, 'max_leaves': 4, 'min_child_weight': 6.456232181430629, 'n_estimators': 68
2, 'alpha': 0.4346791939075747, 'lambda': 1.2687132500980753, 'scale_pos_weight':
0.8541608132425591, 'subsample': 0.9897997300994384}. Best is trial 0 with value:
0.6585032460759072.
[I 2025-11-11 18:24:45,447] Trial 1 finished with value: 0.6447294852707957 and p
arameters: {'learning_rate': 0.039760682577662444, 'max_depth': 9, 'max_delta_ste
p': 5, 'max_leaves': 4, 'min_child_weight': 2.37463831194187, 'n_estimators': 81
7, 'alpha': 0.3181256942236034, 'lambda': 1.25254872668324, 'scale_pos_weight':
1.3110181408002357, 'subsample': 0.9008305830079517}. Best is trial 0 with value:
0.6585032460759072.
[I 2025-11-11 18:24:57,675] Trial 2 finished with value: 0.6130202946228439 and p
arameters: {'learning_rate': 0.09027408151278808, 'max_depth': 6, 'max_delta_ste
p': 0, 'max_leaves': 9, 'min_child_weight': 3.111073591668486, 'n_estimators': 60
5, 'alpha': 0.8717733066174138, 'lambda': 1.2877788303386297, 'scale_pos_weight':
1.349307324174672, 'subsample': 0.9794281769152999}. Best is trial 0 with value:
0.6585032460759072.
[I 2025-11-11 18:25:09,802] Trial 3 finished with value: 0.6578649012559893 and p
arameters: {'learning_rate': 0.012962741620244706, 'max_depth': 11, 'max_delta_st
ep': 2, 'max_leaves': 8, 'min_child_weight': 6.010620296479998, 'n_estimators': 7
55, 'alpha': 0.3273100266908442, 'lambda': 1.430609262293273, 'scale_pos_weight':
0.8875982893562331, 'subsample': 0.7154407987537261}. Best is trial 0 with value:
0.6585032460759072.
[I 2025-11-11 18:25:16,696] Trial 4 finished with value: 0.653485990452931 and pa
rameters: {'learning_rate': 0.020779468111306522, 'max_depth': 12, 'max_delta_ste
p': 1, 'max_leaves': 8, 'min_child_weight': 4.878211018624603, 'n_estimators': 43
5, 'alpha': 0.4874215174602171, 'lambda': 1.2243100189187586, 'scale_pos_weight':
1.4709699903899263, 'subsample': 0.7366733458174273}. Best is trial 0 with value:
0.6585032460759072.
[I 2025-11-11 18:25:25,377] Trial 5 finished with value: 0.6545128529343115 and p
arameters: {'learning_rate': 0.018212046620550208, 'max_depth': 10, 'max_delta_st
ep': 5, 'max_leaves': 9, 'min_child_weight': 1.46565483615348, 'n_estimators': 47
7, 'alpha': 0.5896753757268361, 'lambda': 0.9983208182892922, 'scale_pos_weight':
0.7419172784488272, 'subsample': 0.8811671691551128}. Best is trial 0 with value:
0.6585032460759072.
[I 2025-11-11 18:25:34,952] Trial 6 finished with value: 0.6551894937245891 and p
arameters: {'learning_rate': 0.015202923080984875, 'max_depth': 7, 'max_delta_ste
p': 3, 'max_leaves': 9, 'min_child_weight': 5.333461290299137, 'n_estimators': 51
7, 'alpha': 0.643818924966115, 'lambda': 0.9104178894465504, 'scale_pos_weight':
0.8550218769414866, 'subsample': 0.9738388228365443}. Best is trial 0 with value:
0.6585032460759072.
[I 2025-11-11 18:25:48,513] Trial 7 finished with value: 0.652339628141706 and pa
rameters: {'learning_rate': 0.01132351638392629, 'max_depth': 6, 'max_delta_ste
p': 2, 'max_leaves': 9, 'min_child_weight': 6.218433191284497, 'n_estimators': 72
7, 'alpha': 0.13174335652197058, 'lambda': 0.8201074718406658, 'scale_pos_weigh
t': 1.051957596534793, 'subsample': 0.9107315192248906}. Best is trial 0 with val
ue: 0.6585032460759072.
[I 2025-11-11 18:25:57,699] Trial 8 finished with value: 0.6436536057125487 and p
arameters: {'learning_rate': 0.03965537321649591, 'max_depth': 9, 'max_delta_ste
p': 5, 'max_leaves': 4, 'min_child_weight': 6.45462479366382, 'n_estimators': 74
8, 'alpha': 0.4230635229029006, 'lambda': 1.2982533902010795, 'scale_pos_weight':
1.0189934022674825, 'subsample': 0.9678517787597908}. Best is trial 0 with value:
0.6585032460759072.
[I 2025-11-11 18:26:04,779] Trial 9 finished with value: 0.6581363937235069 and p
arameters: {'learning_rate': 0.014856432422840802, 'max_depth': 6, 'max_delta_ste
p': 5, 'max_leaves': 4, 'min_child_weight': 1.6272796142214434, 'n_estimators': 5
84, 'alpha': 0.9616983599234089, 'lambda': 0.8522487752643214, 'scale_pos_weigh
t': 1.0151290219392148, 'subsample': 0.7984161834192552}. Best is trial 0 with va
lue: 0.6585032460759072.
```

```
[I 2025-11-11 18:26:14,123] Trial 10 finished with value: 0.6556491530719533 and
parameters: {'learning_rate': 0.028317472819862275, 'max_depth': 4, 'max_delta_st
ep': 3, 'max_leaves': 2, 'min_child_weight': 7.553373820280364, 'n_estimators': 9
99, 'alpha': 0.12689829217564597, 'lambda': 1.1013225004092586, 'scale_pos_weigh
t': 0.5789433796172965, 'subsample': 0.8128714735238762}. Best is trial 0 with va
lue: 0.6585032460759072.
[I 2025-11-11 18:26:23,133] Trial 11 finished with value: 0.6560873539525377 and
parameters: {'learning_rate': 0.022188527151386774, 'max_depth': 4, 'max_delta_st
ep': 4, 'max_leaves': 5, 'min_child_weight': 3.4588109219658767, 'n_estimators':
612, 'alpha': 0.8940747011863381, 'lambda': 1.4455470040826945, 'scale_pos_weigh
t': 1.119137114266418, 'subsample': 0.8030831210997967}. Best is trial 0 with val
ue: 0.6585032460759072.
[I 2025-11-11 18:26:28,954] Trial 12 finished with value: 0.6540999942530703 and
parameters: {'learning_rate': 0.010717065469400485, 'max_depth': 8, 'max_delta_st
ep': 4, 'max_leaves': 2, 'min_child_weight': 1.0273890260999945, 'n_estimators':
612, 'alpha': 0.7261684518983765, 'lambda': 1.1132035128244102, 'scale_pos_weigh
t': 0.7040637877455553, 'subsample': 0.7810079368916489}. Best is trial 0 with va
lue: 0.6585032460759072.
[I 2025-11-11 18:26:41,594] Trial 13 finished with value: 0.6105819202201975 and
parameters: {'learning_rate': 0.07772472678811199, 'max_depth': 6, 'max_delta_ste
p': 4, 'max_leaves': 6, 'min_child_weight': 7.635013640588602, 'n_estimators': 86
5, 'alpha': 0.9877400872142189, 'lambda': 1.0211035478495214, 'scale_pos_weight':
1.1744304196024518, 'subsample': 0.8370094264938139}. Best is trial 0 with value:
0.6585032460759072.
[I 2025-11-11 18:26:48,272] Trial 14 finished with value: 0.6588972336103084 and
parameters: {'learning_rate': 0.015925568167992215, 'max_depth': 8, 'max_delta_st
ep': 4, 'max_leaves': 4, 'min_child_weight': 4.110301367161153, 'n_estimators': 5
49, 'alpha': 0.7588376730444213, 'lambda': 0.8045584798417819, 'scale_pos_weigh
t': 0.8943596024204358, 'subsample': 0.7638249079955816}. Best is trial 14 with v
alue: 0.6588972336103084.
[I 2025-11-11 18:26:58,122] Trial 15 finished with value: 0.6459466663335519 and
parameters: {'learning_rate': 0.027008370724982196, 'max_depth': 10, 'max_delta_s
tep': 3, 'max_leaves': 6, 'min_child_weight': 3.8780704101531076, 'n_estimators':
674, 'alpha': 0.7836617927176942, 'lambda': 1.3693711241650293, 'scale_pos_weigh
t': 0.8690719991503694, 'subsample': 0.7500191569072924}. Best is trial 14 with v
alue: 0.6588972336103084.
[I 2025-11-11 18:27:04,681] Trial 16 finished with value: 0.6558073297433396 and
parameters: {'learning_rate': 0.010035515593729264, 'max_depth': 8, 'max_delta_st
ep': 4, 'max_leaves': 3, 'min_child_weight': 4.353462202374211, 'n_estimators': 5
31, 'alpha': 0.6488483787427823, 'lambda': 1.1923139080762348, 'scale_pos_weigh
t': 0.5789114980303893, 'subsample': 0.932956756315544}. Best is trial 14 with va
lue: 0.6588972336103084.
[I 2025-11-11 18:27:10,983] Trial 17 finished with value: 0.6582796620648017 and
parameters: {'learning_rate': 0.017680215766853278, 'max_depth': 12, 'max_delta_s
tep': 4, 'max_leaves': 5, 'min_child_weight': 5.455478046384526, 'n_estimators':
408, 'alpha': 0.48564581645281646, 'lambda': 0.9852390664544493, 'scale_pos_weigh
t': 0.7387926512214782, 'subsample': 0.8682599429079747}. Best is trial 14 with v
alue: 0.6588972336103084.
[I 2025-11-11 18:27:18,725] Trial 18 finished with value: 0.6464107178679874 and
parameters: {'learning_rate': 0.05420753113260418, 'max_depth': 10, 'max_delta_st
ep': 3, 'max_leaves': 3, 'min_child_weight': 6.896070383290729, 'n_estimators': 6
58, 'alpha': 0.30841462450624746, 'lambda': 1.3684095015769975, 'scale_pos_weigh
t': 0.9156844577497533, 'subsample': 0.8430085424995925}. Best is trial 14 with v
alue: 0.6588972336103084.
[I 2025-11-11 18:27:32,093] Trial 19 finished with value: 0.6466508534998987 and
parameters: {'learning_rate': 0.022800708880270342, 'max_depth': 9, 'max_delta_st
ep': 2, 'max_leaves': 6, 'min_child_weight': 4.440411002027704, 'n_estimators': 8
64, 'alpha': 0.751501156596042, 'lambda': 1.1556557326156178, 'scale_pos_weight':
1.1982176008204817, 'subsample': 0.7642401303755954}. Best is trial 14 with valu
e: 0.6588972336103084.
```

```
[I 2025-11-11 18:27:40,595] Trial 20 finished with value: 0.6584135251118698 and
parameters: {'learning_rate': 0.014021563695021489, 'max_depth': 11, 'max_delta_s
tep': 0, 'max_leaves': 7, 'min_child_weight': 2.677495018864424, 'n_estimators':
541, 'alpha': 0.39861533895978873, 'lambda': 0.9035593578214429, 'scale_pos_weigh
t': 0.6525596077572939, 'subsample': 0.7086727830546052}. Best is trial 14 with v
alue: 0.6588972336103084.
[I 2025-11-11 18:27:48,291] Trial 21 finished with value: 0.6590838474690132 and
parameters: {'learning_rate': 0.01403071068511437, 'max_depth': 11, 'max_delta_st
ep': 0, 'max_leaves': 7, 'min_child_weight': 2.6741884648252268, 'n_estimators':
508, 'alpha': 0.4022384686223349, 'lambda': 0.9323176655710569, 'scale_pos_weigh
t': 0.6583329361242028, 'subsample': 0.7021081885897614}. Best is trial 21 with v
alue: 0.6590838474690132.
[I 2025-11-11 18:27:55,269] Trial 22 finished with value: 0.6569426321531574 and
parameters: {'learning_rate': 0.01701138757063513, 'max_depth': 11, 'max_delta_st
ep': 1, 'max_leaves': 5, 'min_child_weight': 2.3144473766334297, 'n_estimators':
476, 'alpha': 0.5297459040390555, 'lambda': 0.8014198158059516, 'scale_pos_weigh
t': 0.7975310116394644, 'subsample': 0.7042905504447772}. Best is trial 21 with v
alue: 0.6590838474690132.
[I 2025-11-11 18:28:03,816] Trial 23 finished with value: 0.659383183730305 and p
arameters: {'learning_rate': 0.011933657936689018, 'max_depth': 7, 'max_delta_ste
p': 1, 'max_leaves': 7, 'min_child_weight': 3.5523815019130254, 'n_estimators': 5
64, 'alpha': 0.23139266871319286, 'lambda': 0.8896247518672754, 'scale_pos_weigh
t': 0.5086661882302743, 'subsample': 0.7409378291980556}. Best is trial 23 with v
alue: 0.659383183730305.
[I 2025-11-11 18:28:10,920] Trial 24 finished with value: 0.6600137623958835 and
parameters: {'learning_rate': 0.012134116163552658, 'max_depth': 7, 'max_delta_st
ep': 1, 'max_leaves': 7, 'min_child_weight': 3.7673814119035045, 'n_estimators':
470, 'alpha': 0.20012105506607555, 'lambda': 0.9091879011562307, 'scale_pos_weigh
t': 0.5502549147709227, 'subsample': 0.7405828437185442}. Best is trial 24 with v
alue: 0.6600137623958835.
[I 2025-11-11 18:28:17,747] Trial 25 finished with value: 0.6612875706187362 and
parameters: {'learning_rate': 0.01005203554777857, 'max_depth': 7, 'max_delta_ste
p': 1, 'max_leaves': 7, 'min_child_weight': 3.392228725695112, 'n_estimators': 45
2, 'alpha': 0.215470249993757, 'lambda': 0.9357274317664734, 'scale_pos_weight':
0.5076342229293634, 'subsample': 0.731754782941457}. Best is trial 25 with value:
0.6612875706187362.
[I 2025-11-11 18:28:25,834] Trial 26 finished with value: 0.6583283948318389 and
parameters: {'learning_rate': 0.010155270901017886, 'max_depth': 7, 'max_delta_st
ep': 1, 'max_leaves': 10, 'min_child_weight': 3.620935967436947, 'n_estimators':
444, 'alpha': 0.2067525490468405, 'lambda': 1.0627655552166542, 'scale_pos_weigh
t': 0.5296443694851631, 'subsample': 0.740419199412363}. Best is trial 25 with va
lue: 0.6612875706187362.
[I 2025-11-11 18:28:31,930] Trial 27 finished with value: 0.6617636356962469 and
parameters: {'learning_rate': 0.012553497702095068, 'max_depth': 5, 'max_delta_st
ep': 1, 'max_leaves': 7, 'min_child_weight': 3.2131323422544567, 'n_estimators':
402, 'alpha': 0.19150418015918289, 'lambda': 0.9504312018278428, 'scale_pos_weigh
t': 0.5111128209832312, 'subsample': 0.7273431726126296}. Best is trial 27 with v
alue: 0.6617636356962469.
[I 2025-11-11 18:28:38,578] Trial 28 finished with value: 0.6545499960232889 and
parameters: {'learning_rate': 0.01990050392394721, 'max_depth': 5, 'max_delta_ste
p': 1, 'max_leaves': 8, 'min_child_weight': 3.0984632689139016, 'n_estimators': 4
12, 'alpha': 0.21761580013135917, 'lambda': 0.9606995613154572, 'scale_pos_weigh
t': 0.6038178492675177, 'subsample': 0.7815199869270455}. Best is trial 27 with v
alue: 0.6617636356962469.
[I 2025-11-11 18:28:45,803] Trial 29 finished with value: 0.6602513549250897 and
parameters: {'learning_rate': 0.012742410018921665, 'max_depth': 5, 'max_delta_st
ep': 2, 'max_leaves': 7, 'min_child_weight': 5.070507443208342, 'n_estimators': 4
71, 'alpha': 0.16568548747448592, 'lambda': 1.057327618121622, 'scale_pos_weigh
t': 0.5138270953998312, 'subsample': 0.7249492365827841}. Best is trial 27 with v
alue: 0.6617636356962469.
```

```
[I 2025-11-11 18:28:51,959] Trial 30 finished with value: 0.6581248455100279 and
parameters: {'learning_rate': 0.02571506652068143, 'max_depth': 5, 'max_delta_ste
p': 2, 'max_leaves': 6, 'min_child_weight': 4.935559069739544, 'n_estimators': 40
0, 'alpha': 0.1025641731251561, 'lambda': 1.0387496752196, 'scale_pos_weight': 0.
5019604893076678, 'subsample': 0.7230049804584942}. Best is trial 27 with value:
0.6617636356962469.
[I 2025-11-11 18:28:59,199] Trial 31 finished with value: 0.6615352385604802 and
parameters: {'learning_rate': 0.012368821969864758, 'max_depth': 5, 'max_delta_st
ep': 1, 'max_leaves': 7, 'min_child_weight': 3.1486743521423657, 'n_estimators':
483, 'alpha': 0.26419982133473463, 'lambda': 0.9517703255014193, 'scale_pos_weigh
t': 0.6212001230198433, 'subsample': 0.7280705850972236}. Best is trial 27 with v
alue: 0.6617636356962469.
[I 2025-11-11 18:29:06,352] Trial 32 finished with value: 0.6590972536802142 and
parameters: {'learning_rate': 0.012838740825566101, 'max_depth': 5, 'max_delta_st
ep': 2, 'max_leaves': 8, 'min_child_weight': 2.0104854962316603, 'n_estimators':
453, 'alpha': 0.2733391193955198, 'lambda': 1.0754981867288116, 'scale_pos_weigh
t': 0.6390462838681146, 'subsample': 0.7628828493920793}. Best is trial 27 with v
alue: 0.6617636356962469.
[I 2025-11-11 18:29:13,855] Trial 33 finished with value: 0.6607199498677925 and
parameters: {'learning_rate': 0.011866020956583765, 'max_depth': 4, 'max_delta_st
ep': 0, 'max_leaves': 7, 'min_child_weight': 3.154015313753165, 'n_estimators': 4
97, 'alpha': 0.16069143777205985, 'lambda': 0.9742471917018471, 'scale_pos_weigh
t': 0.609082066458168, 'subsample': 0.7283649503933967}. Best is trial 27 with va
lue: 0.6617636356962469.
[I 2025-11-11 18:29:22,134] Trial 34 finished with value: 0.6580222802694313 and
parameters: {'learning_rate': 0.0112506325867458, 'max_depth': 4, 'max_delta_ste
p': 0, 'max_leaves': 8, 'min_child_weight': 3.0481900372699746, 'n_estimators': 5
02, 'alpha': 0.27597739571956387, 'lambda': 0.9730317913076186, 'scale_pos_weigh
t': 0.6979412709860575, 'subsample': 0.820229537037639}. Best is trial 27 with va
lue: 0.6617636356962469.
[I 2025-11-11 18:29:29,224] Trial 35 finished with value: 0.6492709928083944 and
parameters: {'learning_rate': 0.03389207647256812, 'max_depth': 4, 'max_delta_ste
p': 0, 'max_leaves': 6, 'min_child_weight': 2.6802669561504766, 'n_estimators': 4
39, 'alpha': 0.3492697670010463, 'lambda': 0.871285957032021, 'scale_pos_weight':
0.7947686083452257, 'subsample': 0.7236129913068383}. Best is trial 27 with valu
e: 0.6617636356962469.
[I 2025-11-11 18:29:39,842] Trial 36 finished with value: 0.6580923264427545 and
parameters: {'learning_rate': 0.013862404375565665, 'max_depth': 5, 'max_delta_st
ep': 0, 'max_leaves': 7, 'min_child_weight': 2.0646673330888436, 'n_estimators':
498, 'alpha': 0.16877355107527528, 'lambda': 0.9469712546058118, 'scale_pos_weigh
t': 0.6000279946220233, 'subsample': 0.7768966498269575}. Best is trial 27 with v
alue: 0.6617636356962469.
[I 2025-11-11 18:29:51,755] Trial 37 finished with value: 0.6272639661392454 and
parameters: {'learning_rate': 0.050799639712969205, 'max_depth': 6, 'max_delta_st
ep': 1, 'max_leaves': 8, 'min_child_weight': 3.215056378020216, 'n_estimators': 6
46, 'alpha': 0.26485957842696534, 'lambda': 1.004916321129244, 'scale_pos_weigh
t': 0.7753257900024196, 'subsample': 0.7561821491527028}. Best is trial 27 with v
alue: 0.6617636356962469.
[I 2025-11-11 18:30:00,051] Trial 38 finished with value: 0.6516299673666758 and
parameters: {'learning_rate': 0.01969404884190564, 'max_depth': 4, 'max_delta_ste
p': 0, 'max_leaves': 10, 'min_child_weight': 4.079054098978261, 'n_estimators': 4
30, 'alpha': 0.344205245924615, 'lambda': 0.8649378039416089, 'scale_pos_weight':
1.48295248480213, 'subsample': 0.727807759443043}. Best is trial 27 with value:
0.6617636356962469.
[I 2025-11-11 18:30:10,010] Trial 39 finished with value: 0.6536308798452438 and
parameters: {'learning_rate': 0.016305824997727755, 'max_depth': 5, 'max_delta_st
ep': 1, 'max_leaves': 8, 'min_child_weight': 2.923917870193732, 'n_estimators': 5
71, 'alpha': 0.16017850074214796, 'lambda': 0.9419284348110127, 'scale_pos_weigh
t': 0.6258996060917354, 'subsample': 0.792609657585426}. Best is trial 27 with va
lue: 0.6617636356962469.
```

```
[I 2025-11-11 18:30:20,414] Trial 40 finished with value: 0.65694522235683 and pa
rameters: {'learning_rate': 0.011399362137907262, 'max_depth': 6, 'max_delta_ste
p': 0, 'max_leaves': 9, 'min_child_weight': 2.269264519369334, 'n_estimators': 48
7, 'alpha': 0.3725049274571888, 'lambda': 0.8334397542273106, 'scale_pos_weight':
0.6915107019575376, 'subsample': 0.9984793928630921}. Best is trial 27 with valu
e: 0.6617636356962469.
[I 2025-11-11 18:30:28,201] Trial 41 finished with value: 0.660223063890588 and p
arameters: {'learning_rate': 0.012873923806811027, 'max_depth': 5, 'max_delta_ste
p': 2, 'max_leaves': 7, 'min_child_weight': 4.923220430422032, 'n_estimators': 45
9, 'alpha': 0.16712425366386374, 'lambda': 1.0481408896110793, 'scale_pos_weigh
t': 0.5549620105751012, 'subsample': 0.7195114930778446}. Best is trial 27 with v
alue: 0.6617636356962469.
[I 2025-11-11 18:30:35,877] Trial 42 finished with value: 0.6619400698037214 and
parameters: {'learning_rate': 0.010188456768883 55, 'max_depth': 4, 'max_delta_ste
p': 1, 'max_leaves': 7, 'min_child_weight': 4.621266425891584, 'n_estimators': 42
4, 'alpha': 0.24640722879569227, 'lambda': 1.0036241524325655, 'scale_pos_weigh
t': 0.548009153005773, 'subsample': 0.7285299637632721}. Best is trial 42 with va
lue: 0.6619400698037214.
[I 2025-11-11 18:30:42,653] Trial 43 finished with value: 0.6594522366761751 and
parameters: {'learning_rate': 0.01013699253447863, 'max_depth': 4, 'max_delta_ste
p': 1, 'max_leaves': 7, 'min_child_weight': 3.381018364064473, 'n_estimators': 42
5, 'alpha': 0.2507520138724106, 'lambda': 0.9967749475312624, 'scale_pos_weight':
0.5793596055087669, 'subsample': 0.747652950187229}. Best is trial 42 with value:
0.6619400698037214.
[I 2025-11-11 18:30:52,550] Trial 44 finished with value: 0.6584407420870166 and
parameters: {'learning_rate': 0.014721212597637152, 'max_depth': 4, 'max_delta_st
ep': 1, 'max_leaves': 8, 'min_child_weight': 4.648388757237187, 'n_estimators': 5
24, 'alpha': 0.30143438006639833, 'lambda': 0.9258297596719867, 'scale_pos_weigh
t': 0.5547244263089381, 'subsample': 0.7146347533606822}. Best is trial 42 with v
alue: 0.6619400698037214.
[I 2025-11-11 18:30:58,738] Trial 45 finished with value: 0.6583370110756265 and
parameters: {'learning_rate': 0.011172758771130787, 'max_depth': 4, 'max_delta_st
ep': 1, 'max_leaves': 6, 'min_child_weight': 5.792828739039822, 'n_estimators': 4
04, 'alpha': 0.11755469669885302, 'lambda': 0.9689358103876994, 'scale_pos_weigh
t': 1.3876496341984559, 'subsample': 0.701343731793264}. Best is trial 42 with va
lue: 0.6619400698037214.
[I 2025-11-11 18:31:10,618] Trial 46 finished with value: 0.6576363383929777 and
parameters: {'learning_rate': 0.012033355431642566, 'max_depth': 6, 'max_delta_st
ep': 0, 'max_leaves': 6, 'min_child_weight': 4.250328671482023, 'n_estimators': 7
99, 'alpha': 0.45722826844822706, 'lambda': 1.0973754626548384, 'scale_pos_weigh
t': 0.952258716860032, 'subsample': 0.7351376733463594}. Best is trial 42 with va
lue: 0.6619400698037214.
[I 2025-11-11 18:31:21,531] Trial 47 finished with value: 0.6049858063331515 and
parameters: {'learning_rate': 0.09572517551767505, 'max_depth': 5, 'max_delta_ste
p': 2, 'max_leaves': 9, 'min_child_weight': 3.815409717784502, 'n_estimators': 59
3, 'alpha': 0.1882657122397291, 'lambda': 1.0224222950690052, 'scale_pos_weight':
0.6810570260420835, 'subsample': 0.7717089811950363}. Best is trial 42 with valu
e: 0.6619400698037214.
[I 2025-11-11 18:31:35,502] Trial 48 finished with value: 0.6559220402611643 and
parameters: {'learning_rate': 0.015382285705321191, 'max_depth': 6, 'max_delta_st
ep': 1, 'max_leaves': 5, 'min_child_weight': 1.4311276828331685, 'n_estimators':
972, 'alpha': 0.2465375565596173, 'lambda': 1.1421555356397233, 'scale_pos_weigh
t': 0.6147569111580873, 'subsample': 0.7530046662481199}. Best is trial 42 with v
alue: 0.6619400698037214.
[I 2025-11-11 18:31:42,247] Trial 49 finished with value: 0.6616402603481077 and
parameters: {'learning_rate': 0.01088751706124153, 'max_depth': 7, 'max_delta_ste
p': 0, 'max_leaves': 7, 'min_child_weight': 2.7973069102257675, 'n_estimators': 4
50, 'alpha': 0.3153443954828679, 'lambda': 0.8852147402589927, 'scale_pos_weigh
t': 0.7436305177784995, 'subsample': 0.732976573938 4795}. Best is trial 42 with v
alue: 0.6619400698037214.
```

```
Best Parameters Found:
learning_rate: 0.01018845676888355
max_depth: 4
max_delta_step: 1
max_leaves: 7
min_child_weight: 4.621266425891584
n_estimators: 424
alpha: 0.24640722879569227
lambda: 1.0036241524325655
scale_pos_weight: 0.548009153005773
subsample: 0.7285299637632721

Best Cross-Validation AUC: 0.6619400698037214
```

Out[28]:

XGBClassifier

```
XGBClassifier(alpha=0.24640722879569227, base_score=None, booster=None,
              callbacks=None, colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric='auc', feature_types=None,
              feature_weights=None, gamma=None, grow_policy=None,
```

In [29]:

```
y_proba_test = final_model.predict_proba(X_test)[:, 1]
y_pred_test = (y_proba_test >= 0.5).astype(int)

test_auroc = roc_auc_score(y_test, y_proba_test)
print(f"\nFinal Test ROC AUC: {test_auroc:.4f}")
```

```
Final Test ROC AUC: 0.6390
```

In [ ]: