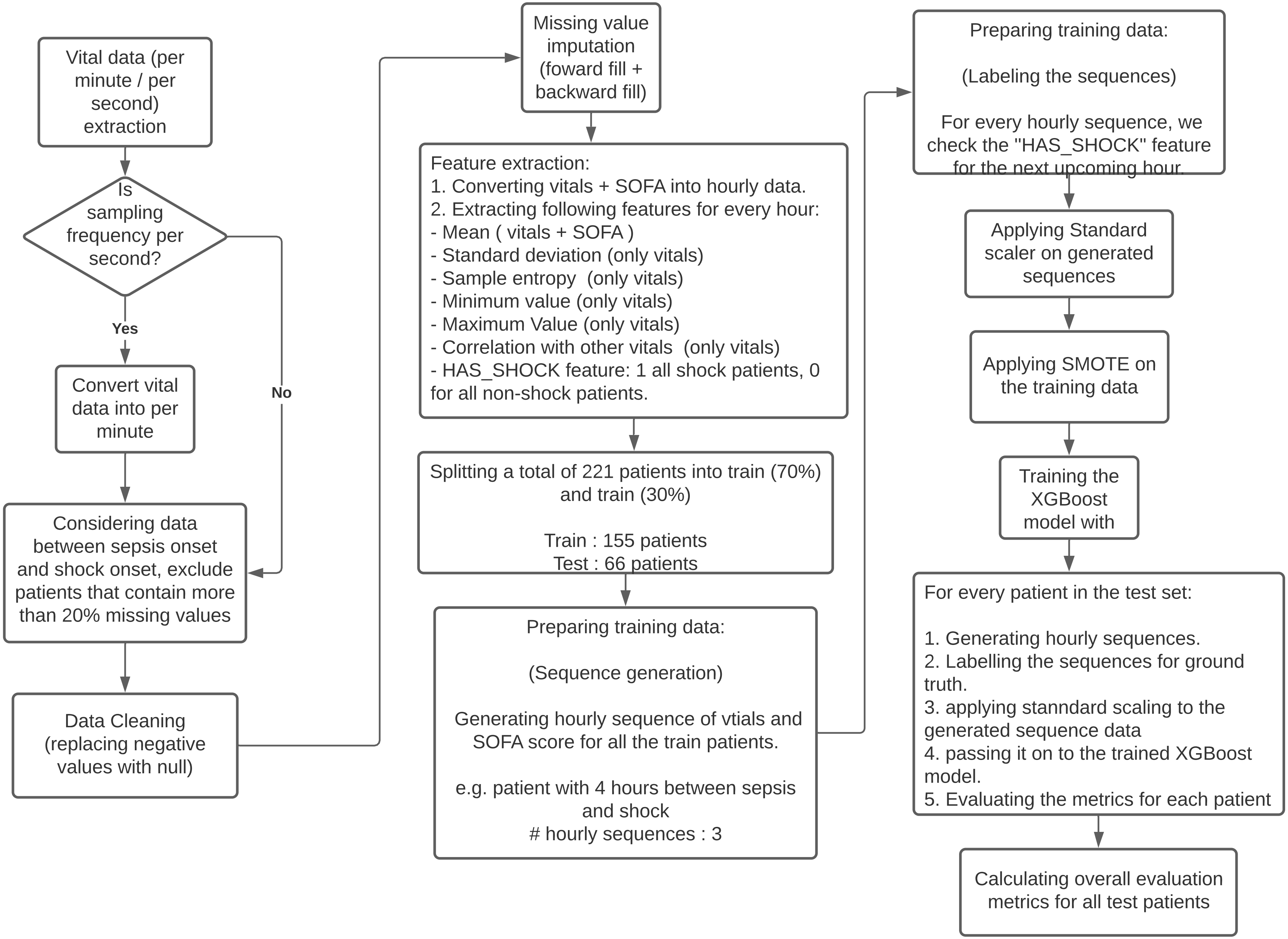

Vital data (per minute / per second) extraction

Is sampling frequency per second?

Convert vital data into per minute

Considering data between sepsis onset and shock onset, exclude patients that contain more than 20% missing values

Data Cleaning (replacing negative values with null)

Missing value imputation (foward fill + backward fill)

Feature extraction:
1. Converting vitals + SOFA into hourly data.
2. Extracting following features for every hour:
- Mean (vitals + SOFA)
- Standard deviation (only vitals)
- Sample entropy (only vitals)
- Minimum value (only vitals)
- Maximum Value (only vitals)
- Correlation with other vitals (only vitals)
- HAS_SHOCK feature: 1 all shock patients, 0 for all non-shock patients.

Splitting a total of 221 patients into train (70%) and train (30%)

Train : 155 patients
Test : 66 patients

Preparing training data:

(Sequence generation)

Generating hourly sequence of vtials and SOFA score for all the train patients.

e.g. patient with 4 hours between sepsis and shock
# hourly sequences : 3

Preparing training data:

(Labeling the sequences)

For every hourly sequence, we check the "HAS_SHOCK" feature for the next upcoming hour.

Applying Standard scaler on generated sequences

Applying SMOTE on the training data

Training the XGBoost model with

For every patient in the test set:

1. Generating hourly sequences.
2. Labelling the sequences for ground truth.
3. applying standdard scaling to the generated sequence data
4. passing it on to the trained XGBoost model.
5. Evaluating the metrics for each patient

Calculating overall evaluation metrics for all test patients