Can you increase the predictive performance of models using binning methods or temporal weights?

**Background:** Electronic Health Record (EHR) data is considered both to have a lot of potential for clinical prediction modelling, yet complex and challenging to model. These complexities have led most researchers to bin covariates across a whole observation period or parts of an observation period, ignoring temporal information present in the data.

**Temporality of EHR data**
- **Sparsity**
  A patient goes to a clinician for a certain problem which is related to specific covariates, so not all covariates are recorded at every visit.

- **Irregular Intervals**
  Patients do not attend clinical visits regularly, meaning data is not sampled at regular intervals. But, increased visits may hint towards a patient’s health worsening.

- **High-dimensionality**
  In each visit different combinations of features are measured. These features are measured at different levels of granularity, and multiple features can represent the same clinical concept.

**Patient Level Prediction (hospital readmission)**
- **Index date (t=0) (hospital admission)**
  The point at which one predicts if an outcome will occur.

- **Observation Period**
  (one year prior admission)
  Where predictors are observed, albeit asynchronously and inconsistently.

- **Time-At-Risk**
  (one month post admission)
  Post the index date, in which there is or is not a predicted outcome.

**Methods**
- **One Window**
  The observation period acts as one window.

- **Overlapping Window**
  One or more windows that are anchored at the index date.

- **Distinct Window**
  T distinct windows over the observation period.

**Binning Strategies**
- **One Window**
  The observation period acts as one window.

**Weighting Strategies**
- **Knowledge Based**
  Non-chronic related covariates are discounted faster than chronic related covariates.

- **Probabilistic Weights**
  A probability density function is fitted to each covariate, this gives the probability of a covariate occurring. These probabilities are used as weights.

- **Learned Weights**
  Coefficients estimated through a regression are normalised and combined with reciprocal temporal discounting.