

## ВИРІШЕННЯ ЗАДАЧІ ВІДНОВЛЕННЯ ВІДСУТНІХ ДАНИХ В МЕДИЦИНІ

Білобородова Т.О.

*Східноукраїнський національний університет ім. В. Даля,  
м. Сєвєродонецьк*

Відсутні дані є проблемою більшості медичних даних, що може привести до серйозних помилок при аналізі даних для прогнозування патологічних станів. Зокрема, вирішення цієї проблеми актуальне для систем акушерсько-гінекологічного профілю при прогнозуванні патологічних станів новонароджених на підставі даних перебігу вагітності. Ці дані представлені великою кількістю фізіологічних показників, що складаються з даних результатів періодичних оглядів та досліджень, наприклад, таких як артеріальний тиск, лабораторні аналізи, ультразвукові дослідження, тощо. Адекватність інтерпретації таких даних, за наявності в них відсутніх значень, залишається сумнівною.

Одним з можливих варіантів вирішення цієї проблеми є відновлення відсутніх значень за допомогою ідентифікації відносин між змінними, зокрема методу імпутації даних [1 – 4]. Визначення правильної стратегії щодо відсутніх даних при проведенні аналізу даних є основною метою дослідження, яке спрямоване на поліпшення прогнозу аналітики даних перебігу вагітності, що містять відсутні дані.

Основні задачі дослідження: оцінка ефективності, якості та чутливості методів відновлення відсутніх даних з використанням повного аналізу, множинної імпутації та при поєднанні імпутації та використання нейронних мереж в неконтрольованому зворотному поширенні. Дослідження проводиться на реальному наборі даних перебігу вагітності які містять відсутні значення.

У доповіді надано результати аналізу наявних даних з урахуванням відсутніх значень. Наведено результати перевірки відновлених даних наступними методами: аналіз повних спостережень, множинна імпутація з використанням бутстрапінга та EM-алгоритма, імпутації та використання нейронних мереж в неконтрольованому зворотному поширенні. Визначено якість досліджуваних методів за допомогою наступних показників: точність, специфічність, показник AUC. Визначено методологію та підхід до відновлення відсутніх значень в медичних даних.

В результаті роботи виділений метод по відновленню відсутніх значень, отримано набір даних з відновленими відсутніми значеннями для виконання подальшої задачі виділення релевантних факторів перебігу вагітності.

**Література:** 1. *Toutenburg H. Rubin. DB: Multiple imputation for nonresponse in surveys / H. Rubin Toutenburg. – 1990.* 2. *Batista G.E. An analysis of four missing data treatment methods for supervised learning / G.E. Batista, M.C. Monard //Applied artificial intelligence. – 2003. – Т. 17. – №. 5-6. – С. 519-533.* 3. *Razzaghi T. et al. Multilevel weighted support vector machine for classification on healthcare data with missing values / T. Razzaghi //PloS one. – 2016. – Т. 11. – №. 5. – С. e0155119.* 4. *Gashler M.S. et al. Missing value imputation with unsupervised backpropagation / M.S. Gashler et al. //Computational Intelligence. – 2014.*