**Convolutional Neural Network and word2vec-based Model for gestational diabetes prediction**

**Group members:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Last Name** | **First Name** | **Email** | **ID** |
| Dou | Wanying | wanyingd@kean.edu | 1098621 |
| Chen | Xu | chexu@kean.edu | 1130033 |

Gestational diabetes is the type of diabetes diagnosed for the first-time during pregnancy. The incidence rate of Gestational diabetes reported worldwide is 1% to 14%. In the United States, about 3% to 8% of pregnant women are diagnosed with gestational diabetes. In China, the incidence rate is about 1% to 5%, and there has been a significant increase in recent years. Women with gestational diabetes have a higher risk of premature delivery and dystocia. Currently, the cause of gestational diabetes mellitus is unknown， so it's essential to uncover the phenotypes and the causes of gestational diabetes.

In our work, we plan to use the Word2Vec model, which is the NLP classic model developed by Google (Tomas Mikolov, 2013), to extract gestational diabetes-related information, including procedure items, medications, and patients, etc. from MIMIC-IV from PhysioNet. Critical dataset released by the MIT associated with over 40,000 patients). Then use the convolutional neural network CNN to accurately forecast the risk of gestational diabetes and find a new collection of phenotypes with similar ICU EHR patterns to aid in the care of Gestational diabetes patients.

References

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

Johnson, A., Pollard, T., & Mark, R. (2016). MIMIC-III Clinical Database (version 1.4). PhysioNet. <https://doi.org/10.13026/C2XW26>.

Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 26.