DOI 10.36074/logos-14.02.2025.044

USAGE OF AI FOR BIOLOGICAL AGEING ESTIMATION WITH TRANSFER LEARNING AND SOCIO-ECONOMIC STATUS

Volkov Oleksandr Volodymyrovych[1], Poliahushko Liubov Hryhorivna[2]

1. Postgraduate student of the Department of Digital Technologies in Energy
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», UKRAINE
ORCID ID: 0009-0003-6834-8118

2. Ph. D. Eng., Docent,
Associate Professor of the Department of Digital Technologies in Energy
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», UKRAINE
ORCID ID: 0000-0003-3287-8523

Modern ageing studies are becoming more and more actual and important as people in developed countries tend to live longer. Longer lifespan results in health issues could be avoided if studied and addressed beforehand. Biological age (BA) can be used to give an estimation of health status at the moment [3] and create ageing trajectory from health perspective to address possible health issues in the future. BA estimation is usually based on biomarkers, genetics, epigenetics, etc. [5], but adding socio-economic status (SES) of a person can produce more precise results taking into account lifestyle factors that influence health [6].

But all that data is quite hard to collect, especially in countries with sparse health records of people, e.g. Ukraine. Many studies suffer from insufficient data, either incomplete or small datasets that lead to obstacles creating stable predictive models.

Transfer learning could be used to create trained models on bigger and complete datasets to transfer obtained knowledge onto lesser and incomplete datasets [4, 7]. This is especially useful for studies of BA of particular regional groups with specific SES. Training on bigger, well-structured datasets produces well-trained models that with some attunement could be used on smaller datasets. Combination of trained model and specific data for smaller region could produce valuable longevity prediction data if applied correctly, taking into account that data in different regions could vary a lot and dataset itself could be structured differently [2, 8].

The domain adaptation task is very important task in the transfer learning. Data across different domains – populations vary, has different structure and could drastically differ (e.g. in a region close to a sea body, seafood consumption expected to be much higher and in multimodal model based on other dataset, it could lead to unstable results). Regularization techniques and adaptive learning can help minimizing influence of difference between datasets from different regions and contribute to the stability of trained model.

The chosen model consists of two main data sources:

- medical data – epigenetics markers, genetics, collected biological samples (blood samples, ECG, retina photos, etc.), biomarkers;
- socio-economic data – information about household income, education level, medical service accessibility, resources accessibility, etc.

The goal is to create a model that predict BA, ageing trajectory, possible health risks on basis of both biological and social factors to create more precise predictions and overcome limitations of each individual set of data (e.g. medical data being not able to track overall life stress factors and SES being less describing health status overall). In this particular case, the process of modeling BA would consist of following steps:

Firstly, base model is created on bigger dataset (probably on a bigger population of more studied society). A deep neural network is studied on such dataset with medical and social data to produce main connections between different biomarkers, SES factors and BA [1].

Secondly, after obtaining basic patterns in BA estimations, the next step is made – creation of separate model that is based on studied previously, but is adapted to lesser or incomplete datasets that focused on specific, smaller social groups. For that process, methods of domain adaptation and regularization methods are used, that allows model to minimize differences influence on the BA estimation [8].

Finally, integration of biomedical and socio-economic characteristics of specific region. Special module should study model on regional data and adapt it to chosen region, taking into account interactions between SES and biomarkers to create more accurate BA predictions, also allowing creation of models for regional analysis.

Data for initial model training can be taken from open data banks. For medical data, open medical data banks could be used, e.g. The National Health and Nutrition Examination Survey (NHANES) that collects data from population of the United States. Such data banks can contain data about food habits as well. On the one hand, socio-economic data is harder to collect, because mostly it is person specific data that can be obtained only trough survey, but on the other hand, for general assessment of SES in region, general survey of people is much easier to conduct. Typical data source for social information could be World Bank Data and others.

Presented AI model usage concept with transfer learning for integration of medical data and socio-economic data for prediction of BA should be used to overcome limited data accessibility. This approach has potential to create accurate data models for regions with limited datasets on basis of data from regions with bigger and complete datasets.

This approach allows researches to conduct analysis with obtained models, reusing knowledge from other regions, on data from other more limited data from another region. Transfer learning raises accuracy and speed of research, even allows the possibility of conducting a research.

Another valuable factor is integration of multimodal data. BA is a complex estimation, that generally evaluates health of a person, so, many factors outside of biological samples could provide insight on ageing trajectory. Transfer learning allows researches to evaluate BA more accurately with regional data that is available only in specific region.

Improving transfer learning contains many potential steps, but obvious ones are following: datasets expansion, improvement of domain adaptation process, integration of other environmental factors, e.g. ecological for improvement of model precision.

This whole process of transfer is really important in regions with limited datasets like Ukraine. Obviously, regional datasets should be improved and that will only bring benefit to transfer learning algorithm, but in the meantime, it is valuable tool for analysis in the conditions we have.

REFERENCES:

[1] Bengio, Y., Courville, A., & Goodfellow, I. (2016). *Deep learning*. MIT Press.
[2] Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P.-M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., . . . Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of the Royal Society Interface*, *15*(141), 20170387. https://doi.org/10.1098/rsif.2017.0387
[3] Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115. https://doi.org/10.1186/gb-2013-14-10-r115
[4] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539
[5] López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*, *153*(6), 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039
[6] Marmot, M. (2005). Social determinants of health inequalities. *The Lancet*, *365*(9464), 1099–1104. https://doi.org/10.1016/s0140-6736(05)71146-6
[7] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/tkde.2009.191
[8] Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1). https://doi.org/10.1186/s40537-016-0043-6