Fusilli: A Python package for multimodal data fusion

Florence J Townend [1¶], **James Chapman** [1], **and James H Cole** [1,2]

1 Centre for Medical Image Computing, University College London, UK **2** Dementia Research Centre, Institute of Neurology, University College London, UK ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Ana Trisovic

Reviewers:

- @aaronhan223
- @felixkrones
- @cairola

Submitted: 12 January 2024
Published: unpublished

Summary

Multimodal data fusion is the integration of data from diverse sources, such as MRI scans, genetics, and clinical measures, to enable predictive analysis that leverages relevant information from all available data modalities. The terminology used to describe this approach varies widely; multimodal data fusion is also referred to as multi-view, cross-heterogeneous, and multi-source, among others. This nominative inconsistency makes it difficult for people to navigate current research and locate specific data fusion models. Moreover, many data-fusion models are underpinned by vastly different architectures, such as graph neural networks, autoencoders, and attention mechanisms. It remains unclear how to determine the most effective fusion model for a given analysis. Although previous research may indicate the superiority of one over another, comparisons are often made under different conditions. Crucially, the level of model complexity needed to optimise the information combination between modalities is unknown. It would be valuable to know the trade-off between model complexity and performance.

To address these issues, `fusilli` allows users to "fuse easily". It simplifies the comparison of various multimodal data fusion models in predictive tasks. Offering a collection of models designed for tabular-tabular and tabular-image fusion, `fusilli` operates as a comprehensive pipeline for training and assessing models across binary or multi-class classification, or regression tasks. Its user-friendly interface allows users to modify model structures to suit their specific requirements, empowering them to conduct direct comparisons within their unique settings.

Statement of need

Multimodal data fusion is applicable to any domain where multiple data modalities are collected. Its usage in healthcare and medical domains has increased notably since 2018 (Kline et al., 2022), owing to the multifactorial nature of medical conditions and the diverse means of assessing the human body. These medical domains include but are not limited to oncology (Lipkova et al., 2022), dermatology (Luo et al., 2023), and neurodegenerative disorders (Huang et al., 2023).

Data fusion has also been used in an agricultural context to predict crop yield (S. S. Gopi & Karthikeyan, 2023) or detect diseases (Patil & Kumar, 2022), and in robotics to interpret data from multiple sensors (Duan et al., 2022). Furthermore, data fusion can be used in analysing disaster response scenarios by integrating information from various sources, including social media posts, images, and audio (Algiriyage et al., 2021).

Due to the vast array of applications and the relative disconnect between them, there are many distinct machine learning architectures for multimodal data fusion. Deep learning models in particular are well-suited to multimodal data fusion, as they can learn complex non-linear relationships between modalities. It is, however, still not clear for researchers to know which models are best for their setting.

To address this, there have been several systematic reviews on the topic of multimodal data fusion (Cui et al., 2022; J. Gao et al., 2020; Stahlschmidt et al., 2022; X. Yan et al., 2021). However, these reviews are qualitative, and there is a lack of quantitative benchmarking of models due to non-standardised model implementations.

One solution to this lack of comparability is to create an application-agnostic resource for researchers to be able to easily compare different models in their setting.

Some multimodal data fusion architectures are publicly available (e.g. on GitHub). This is useful for researchers who want to use a specific model, but it would be cumbersome for a researcher to exhaustively find and implement all available models for comparison. Examples of some of these publicly available individual models include `image_tabular` (Tian, 2020), `MCVAE` (Antelmi et al., 2019), and `MADDi` (Golovanevsky et al., 2022).

Curated collections offer researchers diverse options for comparison without the need for extensive model sourcing and implementation. Some collections of multimodal data fusion models focus on non-deep learning models. For instance, `mvlearn` (Perry et al., 2021) is limited to tabular-tabular fusion and focuses on clustering and decomposition rather than deep learning approaches, and `scikit-fusion` (Zitnik, 2015) (no longer maintained) focuses on latent factor and matrix factorisation models.

As far as we are aware, there are three Python packages with collections of deep learning based multimodal data fusion models: `Multi-view-AE` (Aguila et al., 2023), `CCA-Zoo` (Chapman & Wang, 2021), and `pytorch-widedeep` (Zaurin & Mulinka, 2023). `Multi-view-AE` is a collection of autoencoder-based models and `CCA-Zoo` is a collection of fusion models based on canonical correlation analysis (CCA). `pytorch-widedeep` is a collection of models based on Google's Wide and Deep algorithm to combine tabular data with either text or images.

For all three of these packages, the user is required to write their own script for training and evaluation, increasing the time, effort, and expertise needed to run experiments. However, `fusilli`'s pipeline is readily employable. Users can complete training and evaluation with just three function calls, while still having the option to extensively customise their experiment.

None of the current packages include models based on graph neural networks or attention mechanisms. `fusilli` has multiple variations of both of these models and more, covering a wide range of architectures and fusion types.

Additionally, unlike the other data fusion libraries, `fusilli` simplifies model comparison through built-in visualisation methods. It takes only one line of code to generate a clear figure showing model performances ranked based on the user's chosen performance metric, calculated from either validation or external test data.

Overall, `fusilli` differs from the existing fusion toolkits by providing a comprehensive and flexible pipeline for training, evaluating, and comparing state-of-the-art multimodal data fusion models.

Implementation

There are four main steps in the `fusilli` pipeline: experiment setup, data preparation, model training, and evaluation and comparison.

1. Experiment setup

- Choose the prediction task (binary, multi-class, or regression).
- Import the models to be trained.
- Choose whether to do train/test splitting or k-fold cross-validation.
- Define any model structure modifications.
- Specify experimental parameters, such as early stopping, batch sizes, how to log training, and input data file paths.

2. **Preparation of data**
    - Call `prepare_fusion_data` to obtain a PyTorch data module tailored to the model's format.
3. **Model training**
    - Call `train_and_save_models` to train a fusion model based on the experimental setup and prepared PyTorch data module.
4. **Evaluation and comparison**
    - Call `RealsVsPreds` or `ConfusionMatrix` to generate evaluation figures for a single model, either from validation data or external test data.
    - If multiple models have been trained, call `ModelComparison` to generate validation metrics for each fusion model and a figure comparing the models' performance.

Fusion models in `fusilli`

The table below shows the current list of models in `fusilli`. `fusilli` categorises models based on the type of fusion, following the taxonomy developed in (Cui et al., 2022). The models are also categorised by the modalities they fuse: tabular-tabular or tabular-image. Some tabular-tabular models have tabular-image counterparts, where the structure of the model lends itself to both types of fusion.

Most of the models in `fusilli` are inspired by methods found in the literature, and references are provided where this is the case. These models have been modified to suit the needs and format of `fusilli`, such as simplifying the model, rewriting in PyTorch, or adjusting the architecture to work with tabular-tabular and tabular-image data. Additionally, some of the models without references may have been used in literature, but they were not inspired by any specific paper because of their relatively ubiquitous implementation.

Importantly, `fusilli` also includes benchmark unimodal models to help users assess whether multimodal data fusion is beneficial for their task.

Model name (and reference where applicable)	Fusion Category	Modalities Fused
Tabular1 uni-modal	Unimodal	Tabular Only
Tabular2 uni-modal	Unimodal	Tabular Only
Image unimodal	Unimodal	Image Only
Activation function map fusion (Chen et al., 2023)	Operation	Tabular-tabular
Activation function and tabular self-attention (Chen et al., 2023)	Operation	Tabular-tabular
Concatenating tabular data	Operation	Tabular-tabular
Concatenating tabular feature maps (R. Gao et al., 2022)	Operation	Tabular-tabular
Tabular decision	Operation	Tabular-tabular
Channel-wise multiplication net (tabular) (Duanmu et al., 2020)	Attention	Tabular-tabular
Tabular Crossmodal multi-head attention (Golovanevsky et al., 2022)	Attention	Tabular-tabular

Model name (and reference where applicable)	Fusion Category	Modalities Fused
Attention-weighted GNN (Bintsi et al., 2023)	Graph	Tabular-tabular
Edge Correlation GNN	Graph	Tabular-tabular
MCVAE Tabular (Antelmi et al., 2019)	Subspace	Tabular-tabular
Concatenating tabular data with image feature maps (Li et al., 2020)	Operation	Tabular-image
Concatenating tabular and image feature maps (R. Gao et al., 2022)	Operation	Tabular-image
Image decision fusion	Operation	Tabular-image
Channel-wise Image attention (Duanmu et al., 2020)	Attention	Tabular-image
Crossmodal multi-head attention (Golovanevsky et al., 2022)	Attention	Tabular-image
Trained Together Latent Image + Tabular Data (Zhao et al., 2022)	Subspace	Tabular-image
Pretrained Latent Image + Tabular Data (Zhao et al., 2022)	Subspace	Tabular-image
Denoising tabular autoencoder with image maps (R. Yan et al., 2021)	Subspace	Tabular-image

Documentation

The fusilli documentation is hosted on Read the Docs (https://fusilli.readthedocs.io) and includes a guide to all the fusion models, installation instructions, tutorials on running experiments and modifying models, and guidance on contributing models to fusilli.

Future Work

We would like to introduce more models to fusilli to broaden the available selection. Additionally, it would be a step forward to modify select models to be able to handle more than two modalities where feasible. Another objective is to enable users to input images in their original formats, such as JPGs or NIfTIs.

Conclusion

fusilli is a toolkit to compare diverse multimodal data fusion models for predictive tasks. It offers an array of models for tabular-tabular and tabular-image data fusion, operating as an efficient pipeline for training and evaluating models across binary, multi-class, and regression tasks. Users benefit from the ease of comparing various models within their settings and have the flexibility to adapt model structures to suit their specific requirements.

Acknowledgements

We would like to thank Sophie Martin and Ana Lawry Aguila for their advice and support in developing fusilli, and to Dr Paddy Roddy and Dr Philipp Göbl for their contributions during the 2023 Centre for Medical Image Computing Hackathon.

The Journal of Open Source Software

References

Aguila, A. L., Jayme, A., Montaña-Brown, N., Heuveline, V., & Altmann, A. (2023). Multi-view-AE: A Python package for multi-view autoencoder models. *Journal of Open Source Software*, *8*(85), 5093. https://doi.org/10.21105/joss.05093

Algiriyage, N., Prasanna, R., Stock, K., Doyle, E. E. H., & Johnston, D. (2021). Multi-source Multimodal Data and Deep Learning for Disaster Response: A Systematic Review. *SN Computer Science*, *3*(1), 92. https://doi.org/10.1007/s42979-021-00971-4

Antelmi, L., Ayache, N., Robert, P., & Lorenzi, M. (2019). Sparse Multi-Channel Variational Autoencoder for the Joint Analysis of Heterogeneous Data. *Proceedings of the 36th International Conference on Machine Learning*, 302–311. https://proceedings.mlr.press/v97/antelmi19a.html

Bintsi, K.-M., Baltatzis, V., Potamias, R. A., Hammers, A., & Rueckert, D. (2023). *Multimodal brain age estimation using interpretable adaptive population-graph learning*. arXiv. http://arxiv.org/abs/2307.04639

Chapman, J., & Wang, H.-T. (2021). CCA-Zoo: A collection of Regularized, Deep Learning based, Kernel, and Probabilistic CCA methods in a scikit-learn style framework. *Journal of Open Source Software*, *6*(68), 3823. https://doi.org/10.21105/joss.03823

Chen, Q., Li, M., Chen, C., Zhou, P., Lv, X., & Chen, C. (2023). MDFNet: Application of multimodal fusion method based on skin image and clinical data to skin cancer classification. *Journal of Cancer Research and Clinical Oncology*, *149*(7), 3287–3299. https://doi.org/10.1007/s00432-022-04180-1

Cui, C., Yang, H., Wang, Y., Zhao, S., Asad, Z., Coburn, L. A., Wilson, K. T., Landman, B. A., & Huo, Y. (2022). *Deep Multi-modal Fusion of Image and Non-image Data in Disease Diagnosis and Prognosis: A Review*. arXiv. https://doi.org/10.48550/arXiv.2203.15588

Duan, S., Shi, Q., & Wu, J. (2022). Multimodal Sensors and ML-Based Data Fusion for Advanced Robots. *Advanced Intelligent Systems*, *4*(12), 2200213. https://doi.org/10.1002/aisy.202200213

Duanmu, H., Huang, P. B., Brahmavar, S., Lin, S., Ren, T., Kong, J., Wang, F., & Duong, T. Q. (2020). Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Deep Learning with Integrative Imaging, Molecular and Demographic Data. In A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, & L. Joskowicz (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020* (pp. 242–252). Springer International Publishing. https://doi.org/10.1007/978-3-030-59713-9_24

Gao, J., Li, P., Chen, Z., & Zhang, J. (2020). A Survey on Deep Learning for Multimodal Data Fusion. *Neural Computation*, *32*(5), 829–864. https://doi.org/10.1162/neco_a_01273

Gao, R., Li, T., Tang, Y., Xu, K., Khan, M., Kammer, M., Antic, S. L., Deppen, S., Huo, Y., Lasko, T. A., Sandler, K. L., Maldonado, F., & Landman, B. A. (2022). Reducing uncertainty in cancer risk estimation for patients with indeterminate pulmonary nodules using an integrated deep learning model. *Computers in Biology and Medicine*, *150*, 106113. https://doi.org/10.1016/j.compbiomed.2022.106113

Golovanevsky, M., Eickhoff, C., & Singh, R. (2022). Multimodal attention-based deep learning for Alzheimer's disease diagnosis. *Journal of the American Medical Informatics Association*, *29*(12), 2014–2022. https://doi.org/10.1093/jamia/ocac168

Huang, G., Li, R., Bai, Q., & Alty, J. (2023). Multimodal learning of clinically accessible tests to aid diagnosis of neurodegenerative disorders: A scoping review. *Health Information Science and Systems*, *11*(1), 32. https://doi.org/10.1007/s13755-023-00231-0

Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F., & Luo, Y. (2022). *Multimodal Machine Learning in Precision Health*. arXiv. http://arxiv.org/abs/2204.04777

Li, W., Zhuang, J., Wang, R., Zhang, J., & Zheng, W.-S. (2020). Fusing Metadata and Dermoscopy Images for Skin Disease Diagnosis. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 1996–2000. https://doi.org/10.1109/ISBI45749.2020.9098645

Lipkova, J., Chen, R. J., Chen, B., Lu, M. Y., Barbieri, M., Shao, D., Vaidya, A. J., Chen, C., Zhuang, L., Williamson, D. F. K., Shaban, M., Chen, T. Y., & Mahmood, F. (2022). Artificial intelligence for multimodal data integration in oncology. *Cancer Cell*, *40*(10), 1095–1110. https://doi.org/10.1016/j.ccell.2022.09.012

Luo, N., Zhong, X., Su, L., Cheng, Z., Ma, W., & Hao, P. (2023). Artificial intelligence-assisted dermatology diagnosis: From unimodal to multimodal. *Computers in Biology and Medicine*, *165*, 107413. https://doi.org/10.1016/j.compbiomed.2023.107413

Patil, R. R., & Kumar, S. (2022). Rice-Fusion: A Multimodality Data Fusion Framework for Rice Disease Diagnosis. *IEEE Access*, *10*, 5207–5222. https://doi.org/10.1109/ACCESS.2022.3140815

Perry, R., Mischler, G., Guo, R., Lee, T., Chang, A., Koul, A., Franz, C., Richard, H., Carmichael, I., Ablin, P., Gramfort, A., & Vogelstein, J. T. (2021). *Mvlearn: Multiview Machine Learning in Python*.

S. S. Gopi, P., & Karthikeyan, M. (2023). Multimodal Machine Learning Based Crop Recommendation and Yield Prediction Model. *Intelligent Automation & Soft Computing*, *36*(1), 313–326. https://doi.org/10.32604/iasc.2023.029756

Stahlschmidt, S. R., Ulfenborg, B., & Synnergren, J. (2022). Multimodal deep learning for biomedical data fusion: A review. *Briefings in Bioinformatics*, *23*(2), bbab569. https://doi.org/10.1093/bib/bbab569

Tian, Y. (2020). *Image_tabular*. https://github.com/naity/image_tabular/

Yan, R., Zhang, F., Rao, X., Lv, Z., Li, J., Zhang, L., Liang, S., Li, Y., Ren, F., Zheng, C., & Liang, J. (2021). Richer fusion network for breast cancer classification based on multimodal data. *BMC Medical Informatics and Decision Making*, *21*(1), 134. https://doi.org/10.1186/s12911-020-01340-6

Yan, X., Hu, S., Mao, Y., Ye, Y., & Yu, H. (2021). Deep multi-view learning methods: A review. *Neurocomputing*, *448*, 106–129. https://doi.org/10.1016/j.neucom.2021.03.090

Zaurin, J. R., & Mulinka, P. (2023). Pytorch-widedeep: A flexible package for multimodal deep learning. *Journal of Open Source Software*, *8*(86), 5027. https://doi.org/10.21105/joss.05027

Zhao, D., Homayounfar, M., Zhen, Z., Wu, M.-Z., Yu, S. Y., Yiu, K.-H., Vardhanabhuti, V., Pelekos, G., Jin, L., & Koohi-Moghadam, M. (2022). A Multimodal Deep Learning Approach to Predicting Systemic Diseases from Oral Conditions. *Diagnostics*, *12*(12), 3192. https://doi.org/10.3390/diagnostics12123192

Zitnik, M. (2015). *Scikit-fusion*. https://github.com/mims-harvard/scikit-fusion/tree/master