Chapter 10: Structured + Unstructured Data

Seminar - Multimodal Deep Learning
Master - Seminar (Summer Semester 2022)

Speaker: Rickmer Schulte

Supervisor: Daniel Schalk

21.07.2022

Outline

- Taxonomy: Structured vs. Unstructured Data
- Fusion Strategies
- Multimodal DL in Survival + Economics
    - Proposed Methods
    - Applications
    - Critical Assessment
- Conclusion and Outlook

Structured vs. Unstructured Data

Structured Data

vs.

Unstructured Data

Images Text Video Audio

Structured vs. Unstructured Data

Structured Data

vs.

Unstructured Data

Images Text Video Audio

- Differ regarding dimensionality and interpretability
- Blurred lines: Few selected biomarkers/genes vs. multiple thousand biomarkers/genes
- Structured data traditionally studied throughout all field of science vs.
  Unstructured main studied data source in DL
- <u>Idea:</u> Fuse different data modalities and methods from DL and classical Statistics

Fusion Strategies

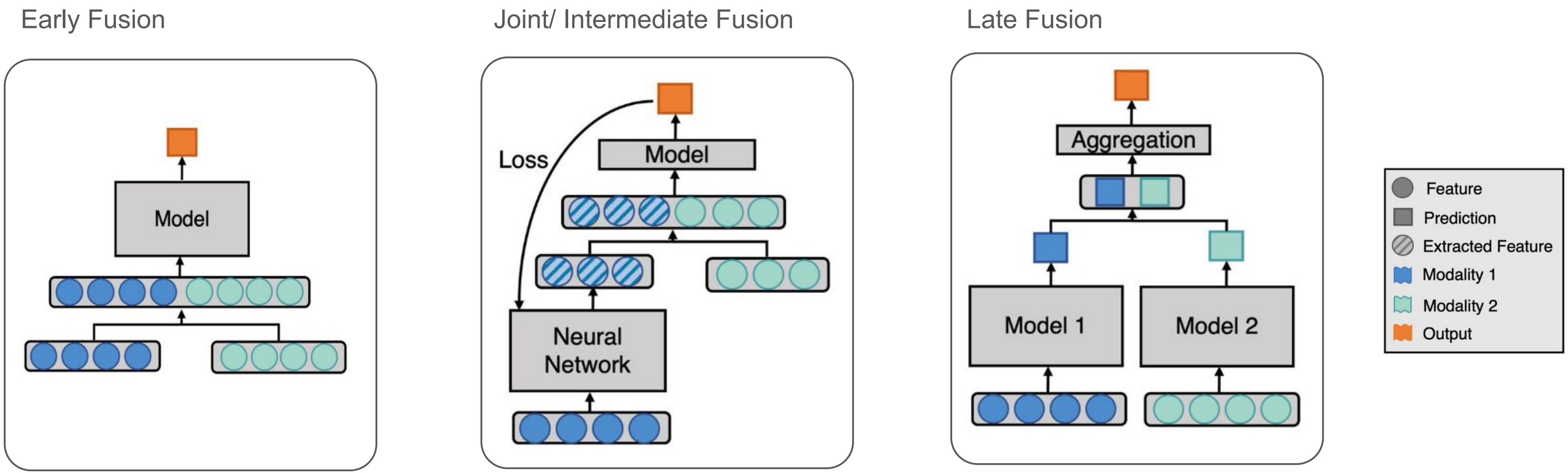
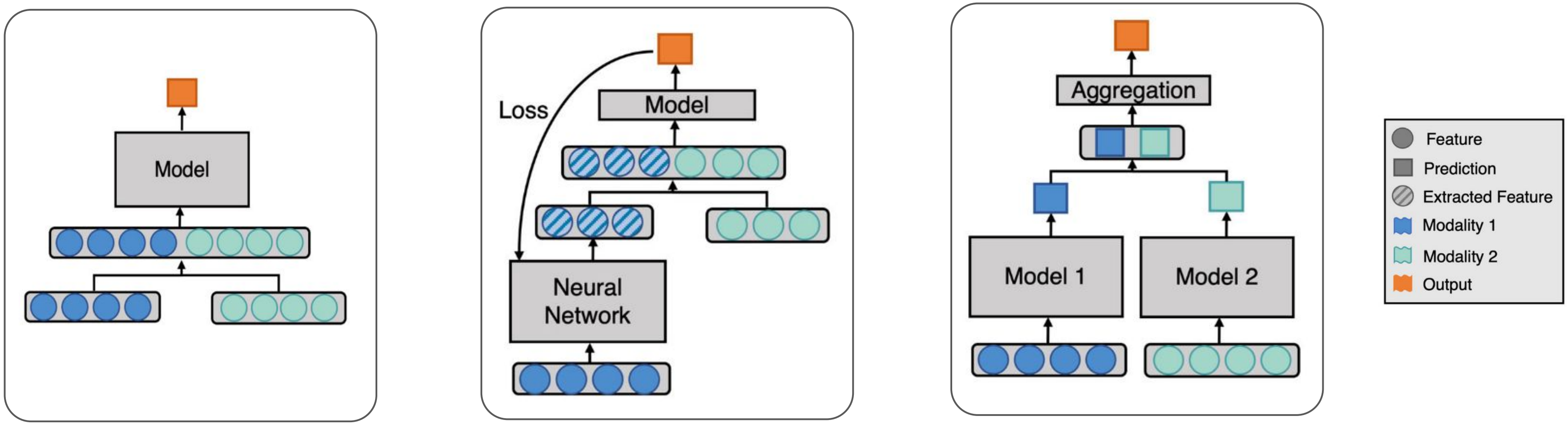


Fig.1: Fusion Strategies Graphics adopted from (Huang et al., 2020)

Fusion Strategies

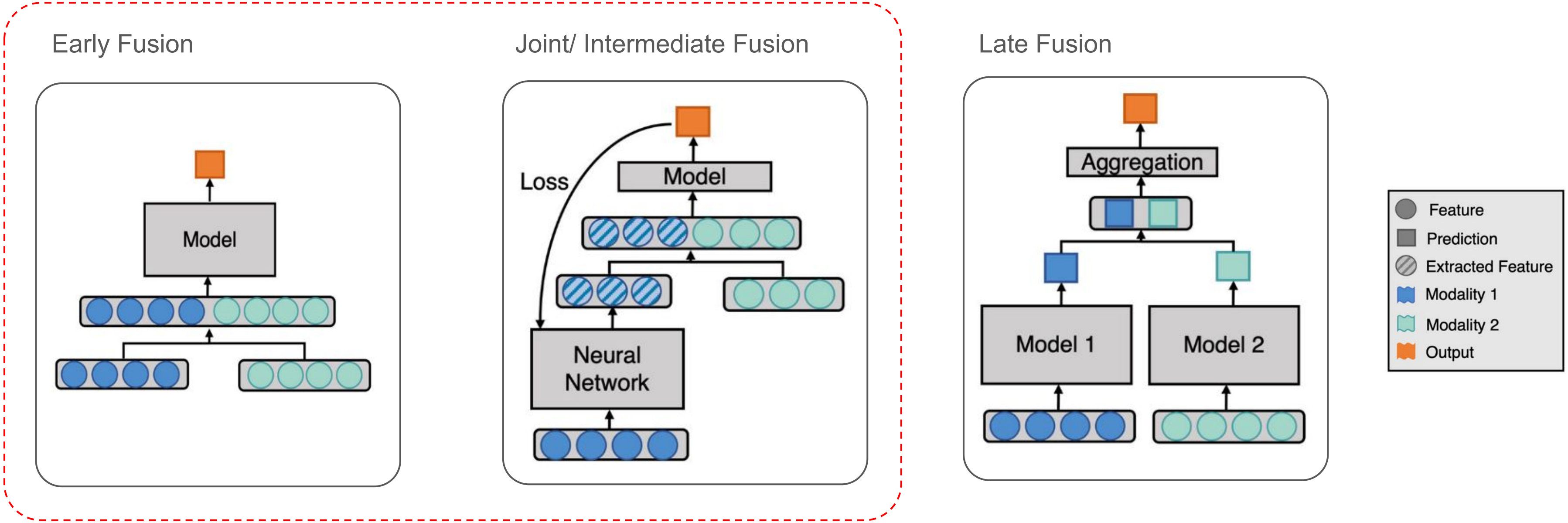
Early Fusion | Joint/ Intermediate Fusion | Late Fusion



- Latent feature representation: Hand-crafted vs. Separately learned vs. End-to-end learning
- Model Block: DNN vs. Interpretable statistical model (Linear, Logistic, GAM etc.)

Fig.1: Fusion Strategies Graphics adopted from (Huang et al., 2020)

Fusion Strategies



- Latent feature representation: Hand-crafted vs. Separately learned vs. End-to-end learning
- Model Block: DNN vs. Interpretable statistical model (Linear, Logistic, GAM etc.)

Fig.1: Fusion Strategies Graphics adopted from (Huang et al., 2020)

How were papers selected?

- Introduced new methodological approaches
- Showed promising applications in new fields
- Used DL on unstructured data in fields where structured data is predominately used

Multimodal DL in Survival

Traditional Survival Analysis (Cox Proportional Hazard Model)

- Hazard function: $$h(t|x) = h_0(t)exp(\beta^T x)$$
- Loss function: $$l(\beta) = -\sum_{i=1}^{n} \delta_i \left(\beta^T x_i - log \sum_{j \in R(t_i)} exp(\beta^T x_j) \right)$$

Fig.2: Equations adopted from (Zhu, Yao and Huang, 2016)

Multimodal in Survival

Traditional Survival Analysis (Cox Proportional Hazard Model)

- Hazard function: $$h(t|x) = h_0(t)exp(\beta^T x)$$
- Loss function: $$l(\beta) = -\sum_{i=1}^{n} \delta_i \left(\boxed{\beta^T x_i} - log \sum_{j \in R(t_i)} exp(\boxed{\beta^T x_j}) \right)$$

Multimodal Survival Analysis

- Loss function: $$l(w,b) = -\sum_{i:R_i=1} \boxed{h_{w,b}^{last}(x_i)} + log \sum_{j:t_j>=t_i} exp(\boxed{h_{w,b}^{last}(x_j)}).$$

Fig.2: Equations adopted from (Zhu, Yao and Huang, 2016)

DeepConvSurv + DeepCorrSurv

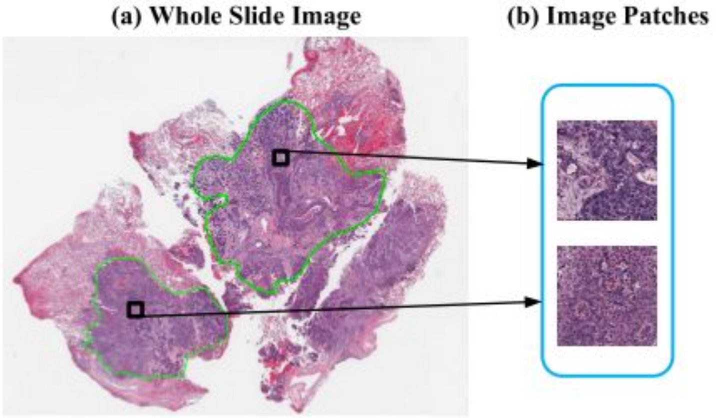


Fig.3: Sampling of Patches (Yao et al., 2016)

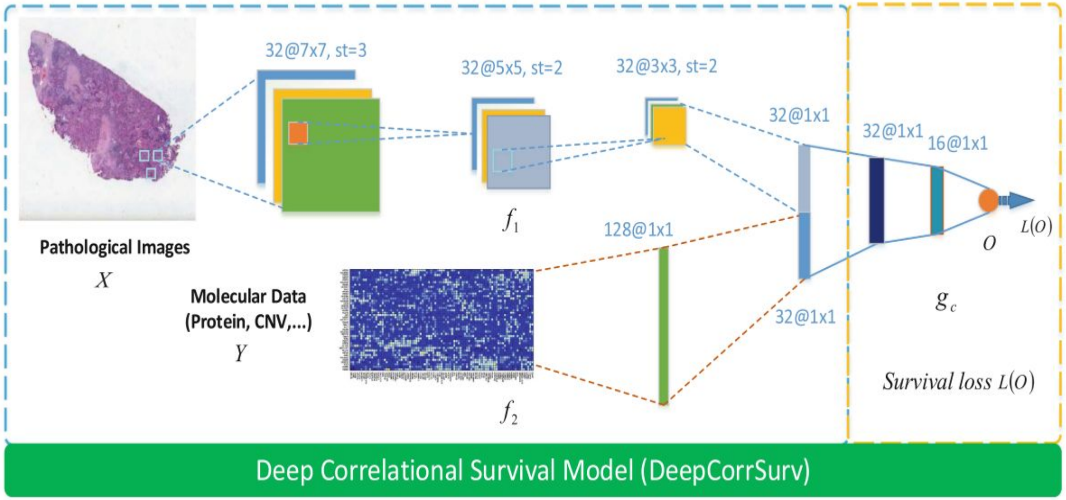


Fig.4: Architecture of the DeepCorrSurv (Yao et al., 2017)

Concat + Cross Auto Encoders

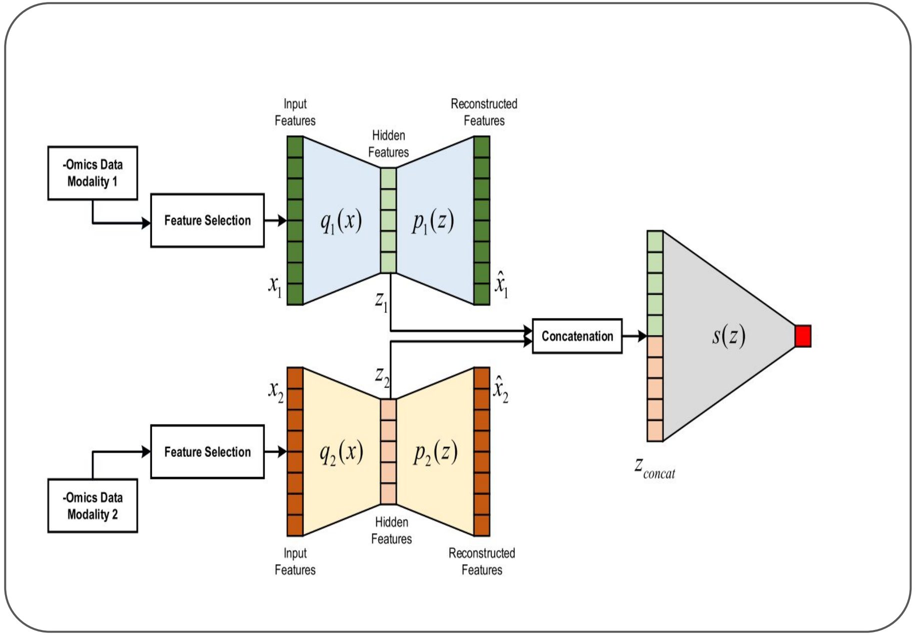


Fig.5: ConcatAE (Tong et al., 2020)

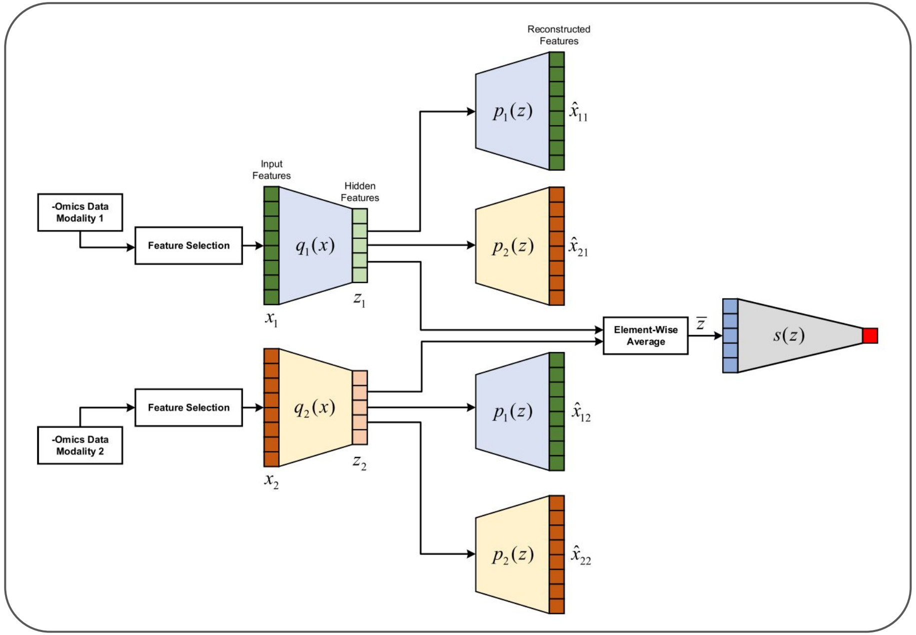


Fig.6: CrossAE (Tong et al., 2020)

Cheerla and Gevaert (2019)

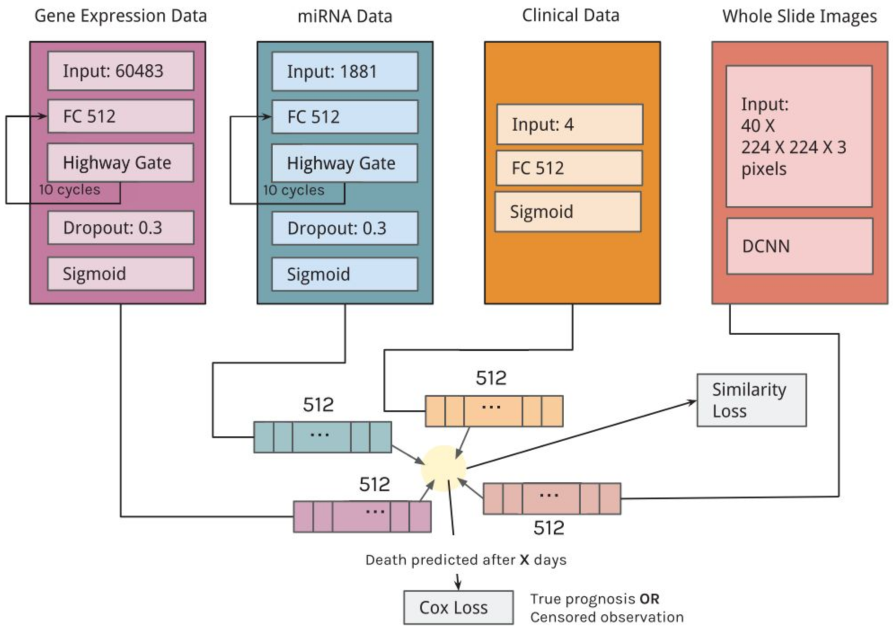


Fig.7: Architecture with Similarity Loss (Gevaert and Cheerla, 2019)

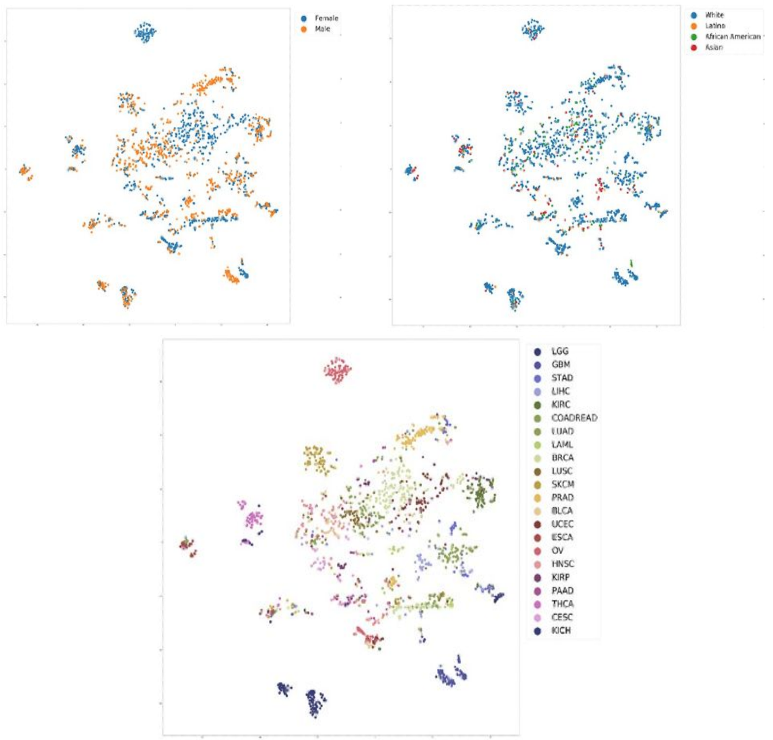


Fig.8: T-SNE-mapped representations of feature vectors
(Gevaert and Cheerla, 2019)

MultiSurv

Architecture and latent feature representations

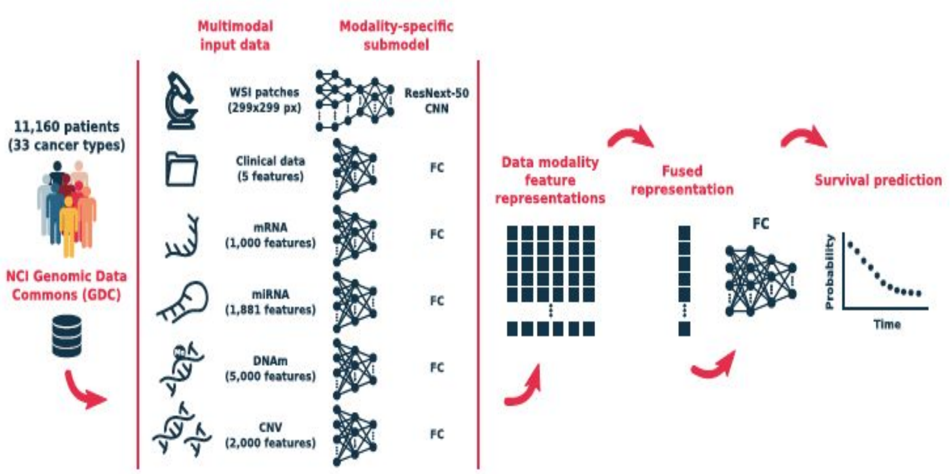


Fig.9: MultiSurv Model Architecture (Vale-Silva and Rohr, 2021)

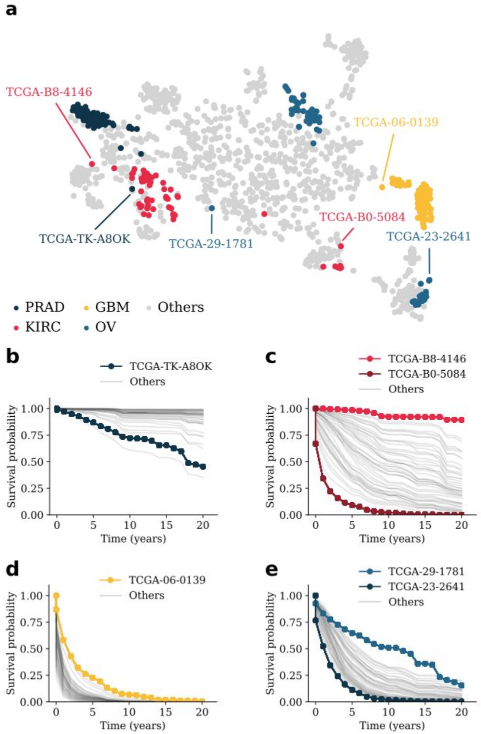


Fig.10: T-SNE-mapped representations of feature vectors (Vale-Silva and Rohr, 2021)

MultiSurv

Results

Metric	Data	Method CPH[6]
C^{td}	Clinical	*0.796* (0.779–0.813)
	mRNA	0.733 (0.712–0.755)
	DNAm	*0.739* (0.719–0.760)
	miRNA	0.676 (0.651–0.700)
	CNV	0.570 (0.543–0.599)
	WSI	–
IBS	Clinical	**0.143** (0.135-0.154)
	mRNA	**0.177** (0.165–0.190)
	DNAm	0.179 (0.165–0.192)
	miRNA	*0.186* (0.171–0.202)
	CNV	*0.214* (0.207–0.224)
	WSI	–

Fig.11: Results single modality models (Vale-Silva and Rohr, 2021)

Included data modalities							
Clinical	mRNA	DNAm	miRNA	CNV	WSI	C^{td} (95% CI)	IBS (95% CI)
•	•					**0.822** (0.805–0.837)	*0.138* (0.126–0.150)
•		•				0.808 (0.791–0.826)	**0.134** (0.125–0.148)
•			•			0.792 (0.775–0.810)	0.147 (0.136–0.161)
•				•		0.795 (0.778–0.812)	0.140 (0.131–0.152)
•					•	0.801 (0.783–0.817)	0.148 (0.140–0.158)
•	•	•				*0.810* (0.793–0.829)	0.146 (0.135–0.158)
•	•	•	•			0.798 (0.781–0.815)	0.153 (0.139–0.168)
•	•	•	•	•		0.802 (0.748–0.820)	0.149 (0.136–0.162)
•	•	•	•	•	•	0.787 (0.769–0.806)	0.152 (0.140–0.166)

Fig.12: Results multi-modality models (Vale-Silva and Rohr, 2021)

DeepPAMM

Deep Piecewise Exponential Additive Mixed Models

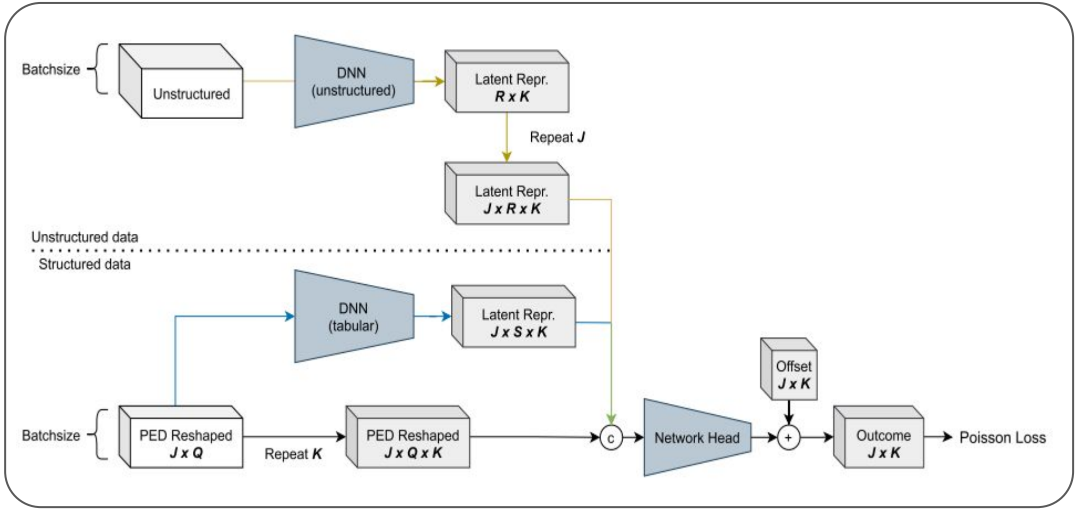


Fig.18: Architecture of DeepPAMM (Kopper et al., 2022)

DeepPAMM & SSDDR

Deep Piecewise Exponential Additive Mixed Models

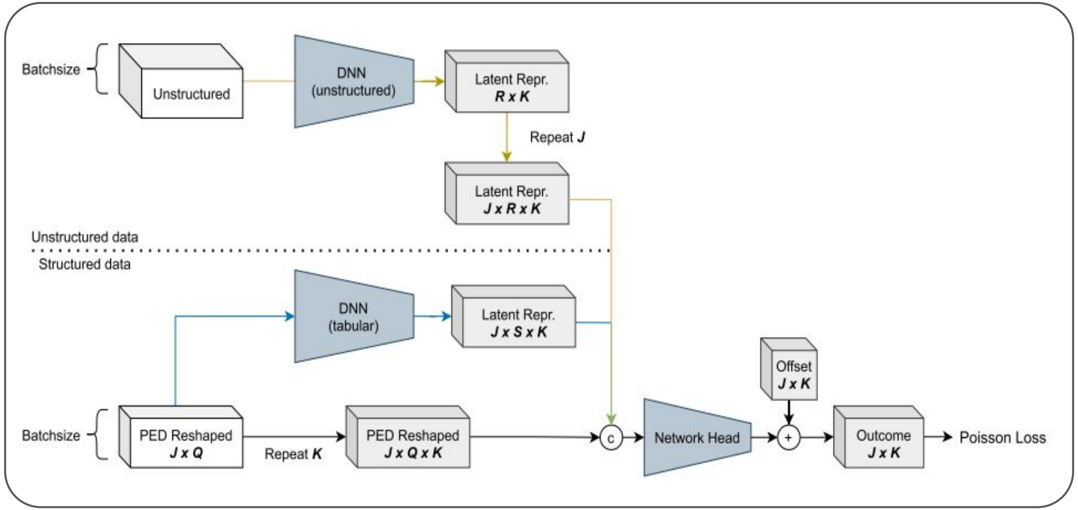


Fig.18: Architecture of DeepPAMM (Kopper et al., 2022)

Semi-Structured Distributional Regression

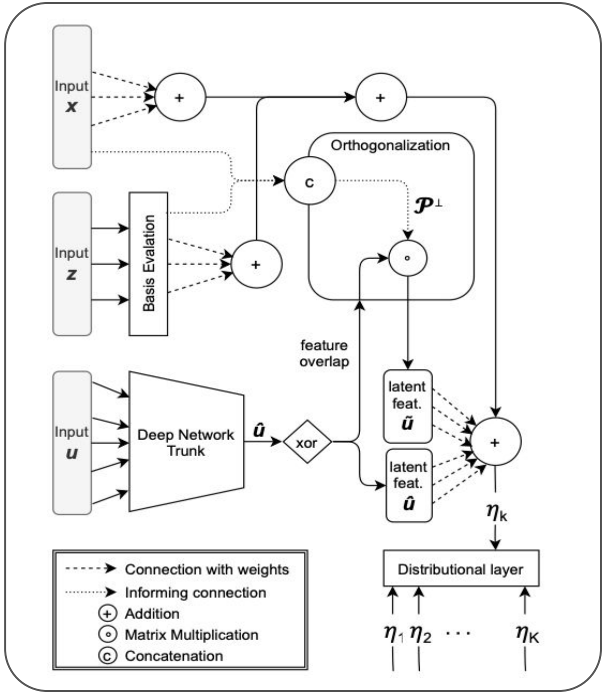


Fig.20: Exemplary architecture of SSDDR (Rügamer, Kolb and Klein, 2020)

Multimodal DL in Economics

Law, Paige and Russell (2019)

Using Street View and Satellite Images to Estimate House Prices

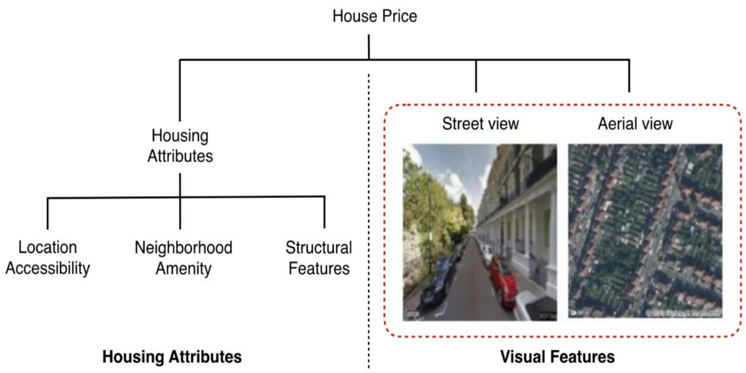


Fig.27: Concept of Model Architecture (Law, Paige and Russell, 2019)

1

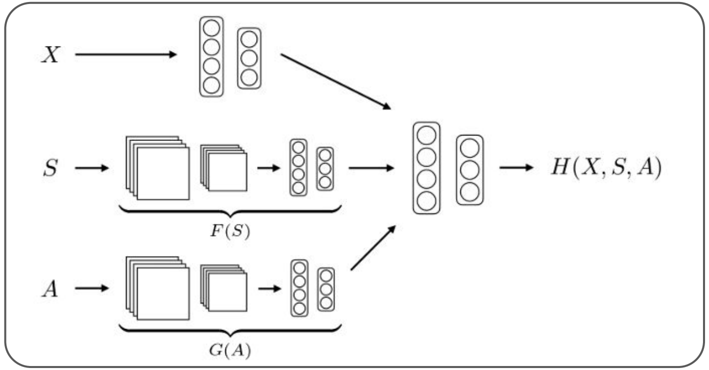


Fig.28: Fully nonlinear model network structure (Law, Paige and Russell, 2019)

2

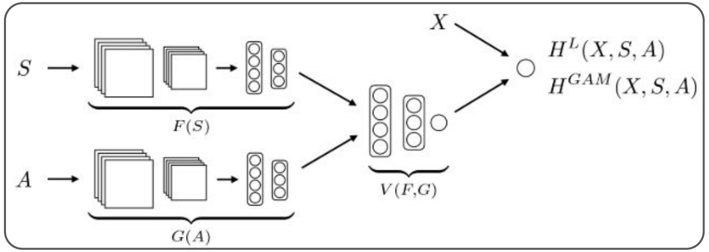


Fig.29: Semi-interpretable model network structure (Law, Paige and Russell, 2019)

Predicting Poverty using Satellite Imagery

What is the problem?

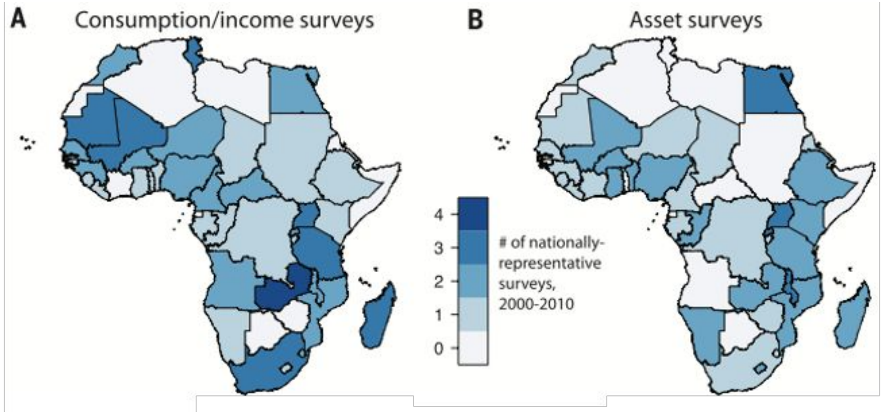


Fig.22: Number of nationally representative consumption and asset surveys occurring in each African country between 2000 and 2010 (Jean et al., 2016)

Possible solution

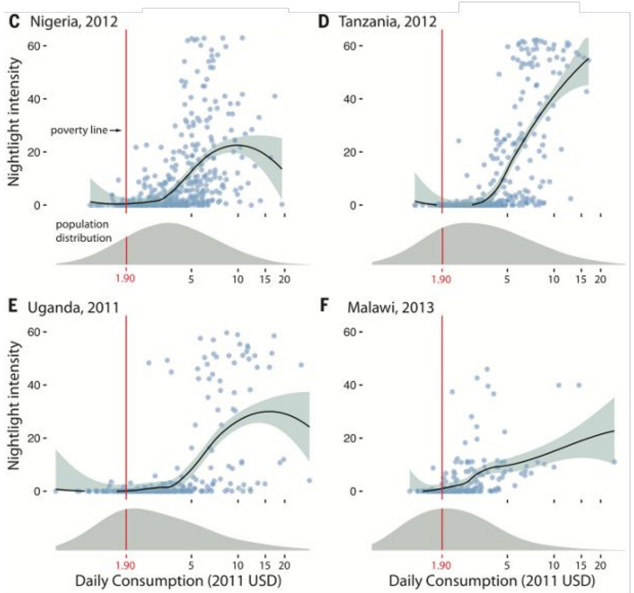


Fig.23: Relationship between per capita consumption expenditure and nightlight intensity at the cluster level (Jean et al., 2016)

Jean et al. (2016)

Methodology

1. Start with CNN that has been trained on ImageNet
2. Fine-tune CNN on new task to predict nighttime light intensities from daytime satellite imagery
3. Use ridge regression model to predict cluster-level consumption/ assets from image features extracted from daytime imagery by the CNN and mean cluster-level values from the survey data

- Nighttime light intensities noisy but globally consistent/ available proxy for economic activity

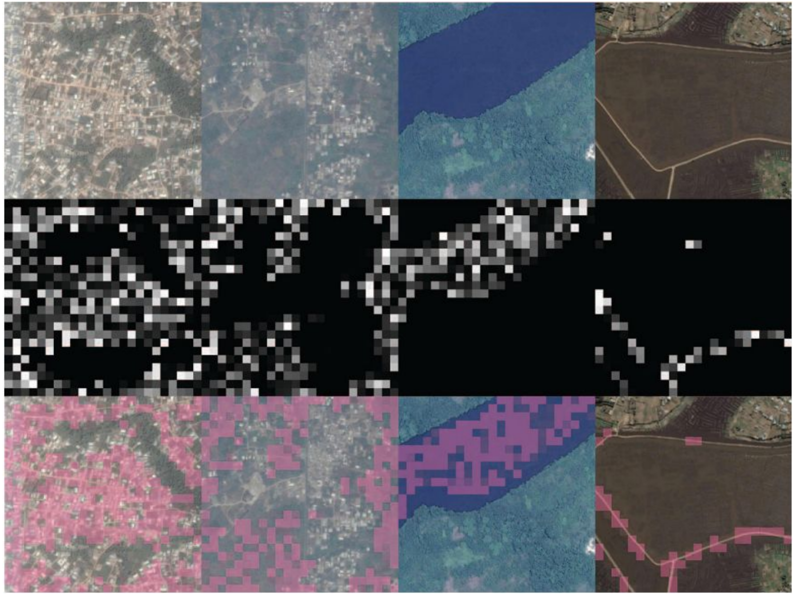


Fig.24: Visualization of activations of different convolutional filters in CNN (Jean et al., 2016)

Jean et al. (2016)

Results

- Transfer Learning improves performance
- Reasonable out-of-country predictions

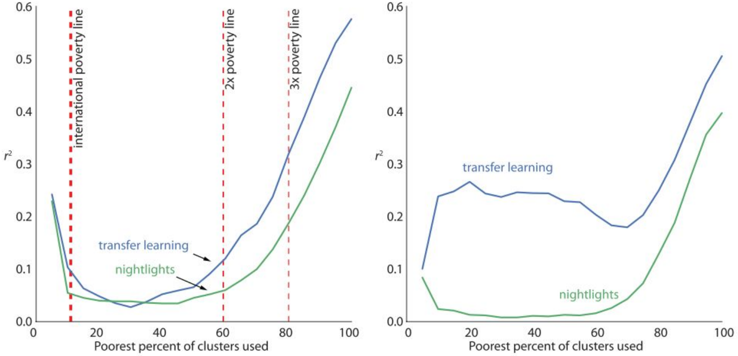


Fig.25: Performance of transfer learning model relative to nightlights for estimating consumption and assets (Jean et al., 2016)

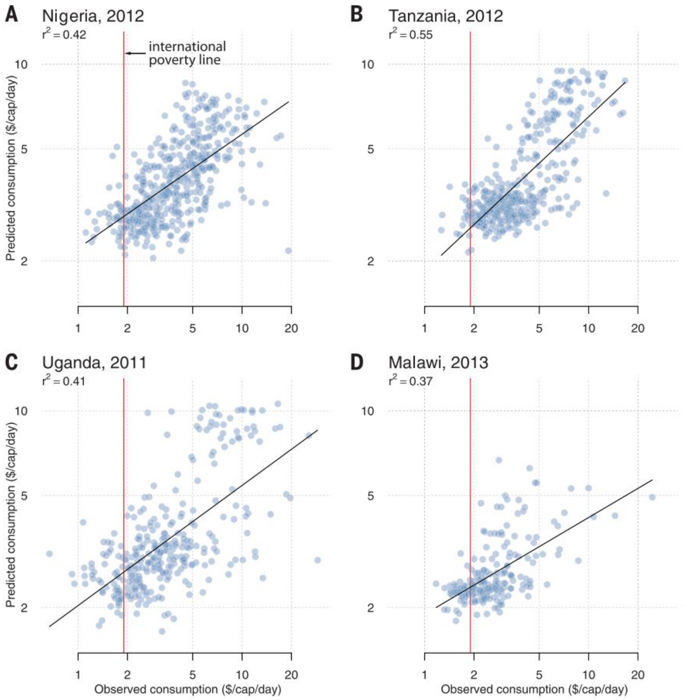


Fig.26: Predictions vs. survey-measured consumption at cluster level (Jean et al., 2016)

Conclusion & Outlook

Conclusion

Achievements

- Different ways to incorporate multimodal data using DL
- Performance increases on various prediction tasks
- Proof of concept: Using remote sensing data to replace costly collected tabular data

Conclusion

Achievements

- Testing many different ways to incorporate multi modal data using DL
- Performance increases on various prediction tasks
- Proof of concept: Using remote sensing data to replace costly collected tabular data

Major challenges

- Small sample size (often insufficient no. of image data)
- Unreliable Benchmarking (between proposed methods but also to single modality models)
- Large flexibility of tuning/ adjusting DL (architecture, fusion, hyperparameters)
- Marginal performance improvements over baselines
- Publication Bias

Outlook

Will multimodal DL replace classical models in regular scientific in context where good and interpretable frameworks for structured data are available?

Outlook

Will multimodal DL replace classical models in regular scientific in context where good and interpretable frameworks for structured data are available?

- Currently, at least questionable with the present setup (small sample sizes, marginal improvements) in many fields

However:

- More unstructured data will become available
  - Full potential (performance improvements) of DL in many fields yet to be discovered
  - Consolidation of Methods is in starting and will accelerate once proper benchmarking datasets are available
- Publicly available and up-to-date unstructured data (satellite imagery) might be an alternative to costly collected tabular data

Many thanks for listening!

Sources

- Cheerla A, Gevaert O, (2019). Deep learning with multimodal representation for pancancer prognosis prediction, *Bioinformatics*, Volume 35, Issue 14, Pages i446–i454, https://doi.org/10.1093/bioinformatics/btz342

- Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, and Hemal Shah. 2016. Wide & Deep Learning for Recommender Systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems (DLRS 2016). Association for Computing Machinery, New York, NY, USA, 7–10. https://doi.org/10.1145/2988450.2988454

- Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med. 2020 Oct 16;3:136. doi: 10.1038/s41746-020-00341-z. PMID: 33083571; PMCID: PMC7567861.

- Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S. Combining satellite imagery and machine learning to predict poverty. Science. 2016 Aug 19;353(6301):790-4. doi: 10.1126/science.aaf7894. PMID: 27540167.

- Kopper, Philipp & Wiegrebe, Simon & Bischl, Bernd & Bender, Andreas & Rügamer, David. (2022). DeepPAMM: Deep Piecewise Exponential Additive Mixed Models for Complex Hazard Structures in Survival Analysis.

- Law, S., Paige, B., and Chris Russell. 2019. Take a Look Around: Using Street View and Satellite Images to Estimate House Prices. ACM Trans. Intell. Syst. Technol. 10, 5, Article 54 (September 2019), 19 pages. https://doi.org/10.1145/3342240

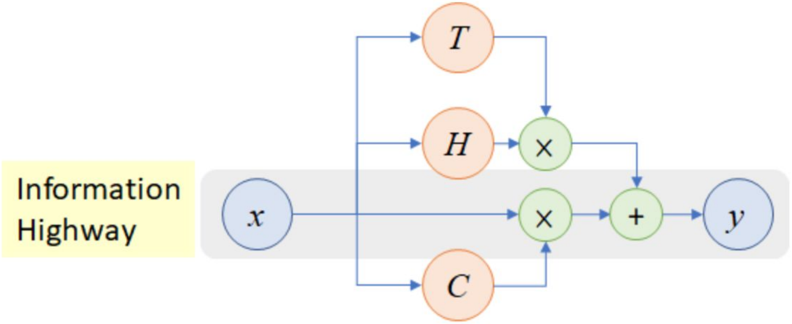
- Pölsterl, Sebastian & Sarasua, Ignacio & Gutierrez Becker, Benjamín & Wachinger, Christian. (2020). A Wide and Deep Neural Network for Survival Analysis from Anatomical Shape and Tabular Clinical Data. 10.1007/978-3-030-43823-4_37.

Sources

- Rügamer, David & Kolb, Chris & Klein, Nadja. (2020). A Unifying Network Architecture for Semi-Structured Deep Distributional Learning.
- Sardelich, Marcelo and Suresh Manandhar. “Multimodal deep learning for short-term stock volatility prediction.” *ArXiv* abs/1812.10479 (2018): n. Pag.
- Steele, Jessica & Sundsøy, Pål & Pezzulo, Carla & Alegana, Victor & Bird, Tomas & Blumenstock, Joshua & Bjelland, Johannes & Engø-Monsen, Kenth & Montjoye, Yves-Alexandre & Iqbal, Asif & Hadiuzzaman, Khandakar & Lu, Xin & Wetter, Erik & Tatem, Andrew & Bengtsson, Linus. (2017). Mapping poverty using mobile phone and satellite data. Journal of The Royal Society Interface. 14. 20160690. 10.1098/rsif.2016.0690.
- Tong L, Mitchel J, Chatlin K, Wang MD. Deep learning based feature-level integration of multi-omics data for breast cancer patients survival analysis. BMC Med Inform Decis Mak. 2020 Sep 15;20(1):225. doi: 10.1186/s12911-020-01225-8. PMID: 32933515; PMCID: PMC7493161.
- Vale-Silva, L.A., Rohr, K. Long-term cancer survival prediction using multimodal deep learning. *Sci Rep* 11, 13505 (2021). https://doi.org/10.1038/s41598-021-92799-4
- Yao, J., Zhu, X., Zhu, F., Huang, J. (2017). Deep Correlational Learning for Survival Prediction from Multi-modality Data. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D., Duchesne, S. (eds) Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017. MICCAI 2017. Lecture Notes in Computer Science(), vol 10434. Springer, Cham. https://doi.org/10.1007/978-3-319-66185-8_46
- You J, Li X, Low M, Lobell D, Ermon S (2017). Deep Gaussian process for crop yield prediction based on remote sensing data. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17). AAAI Press, 4559–4565.

Appendix

Highway Networks

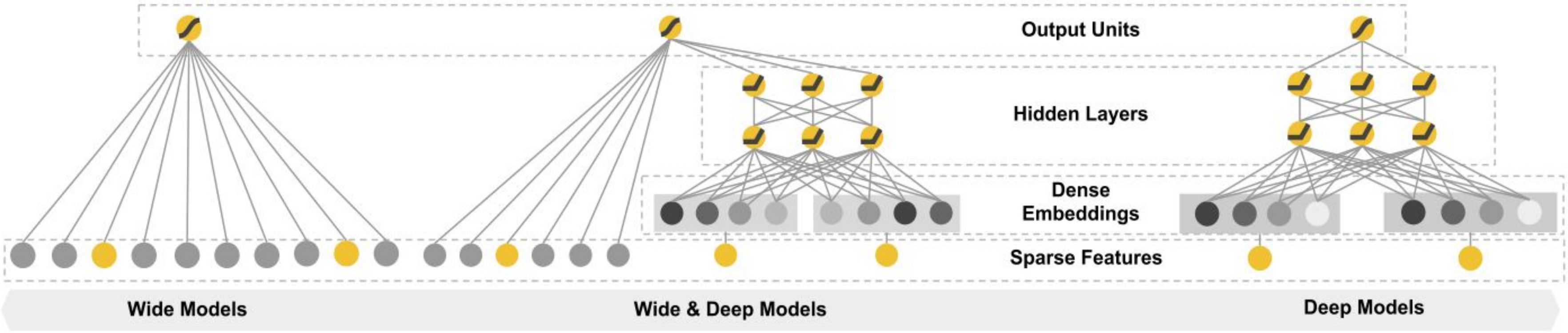


Highway Circuit

- $\mathbf{y} = H(\mathbf{x}, \mathbf{W_H}) \cdot T(\mathbf{x}, \mathbf{W_T}) + \mathbf{x} \cdot C(\mathbf{x}, \mathbf{W_C}).$
- $\mathbf{y} = H(\mathbf{x}, \mathbf{W_H}) \cdot T(\mathbf{x}, \mathbf{W_T}) + \mathbf{x} \cdot (1 - T(\mathbf{x}, \mathbf{W_T})).$
- $\mathbf{y} = \begin{cases} \mathbf{x}, & \text{if } T(\mathbf{x}, \mathbf{W_T}) = \mathbf{0}, \\ H(\mathbf{x}, \mathbf{W_H}), & \text{if } T(\mathbf{x}, \mathbf{W_T}) = \mathbf{1}. \end{cases}$

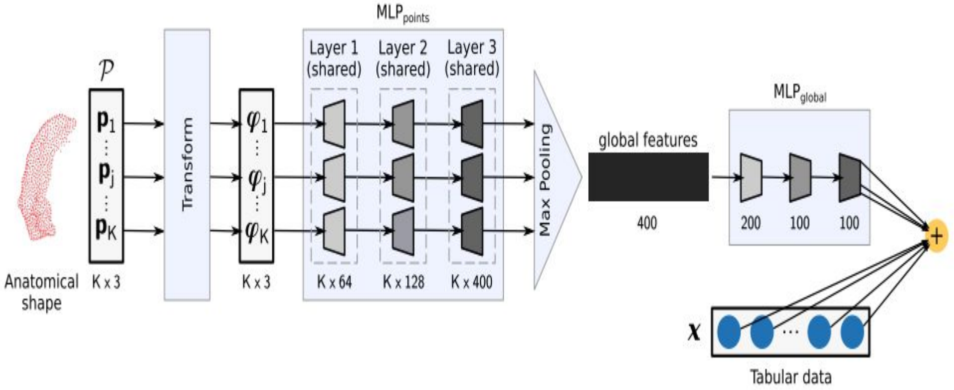
Source: Review Highway Networks (https://towardsdatascience.com/review-highway-networks-gating-function-to-highway-image-classification-5a33833797b5)

Wide & Deep Models

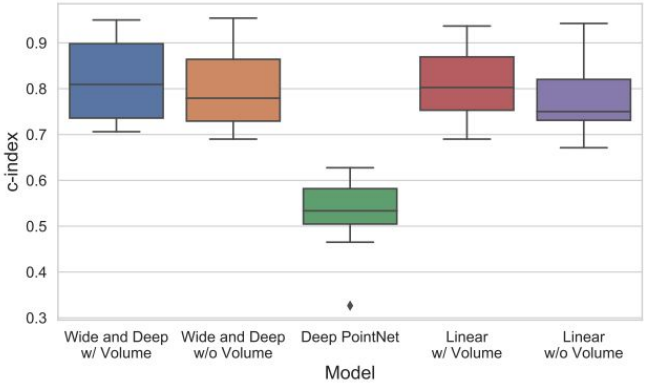


A-Fig.1: Spectrum of Wide & Deep models (Cheng et al. 2016)

Wide & Deep NN for Survival



A-Fig.2: Wide and Deep PointNet Architecture (Pölsterl et al., 2019)

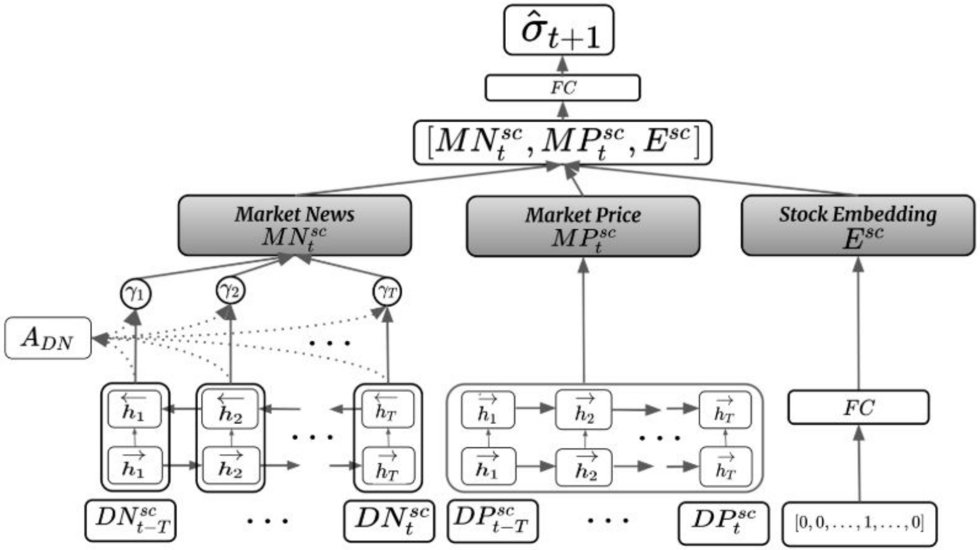


A-Fig.3: Performance Benchmarking (Pölsterl et al., 2019)

Multimodal DL for Stock Volatility Prediction

Hierarchical Network

- 1. Word Embedding
- 2. News Encoder (BiLSTM)
- 3. News Relevance Attention
- 4. News Temoral Context (Attention)
- 5. Price Encoder (LSTM)
- 6. Stock Embedding



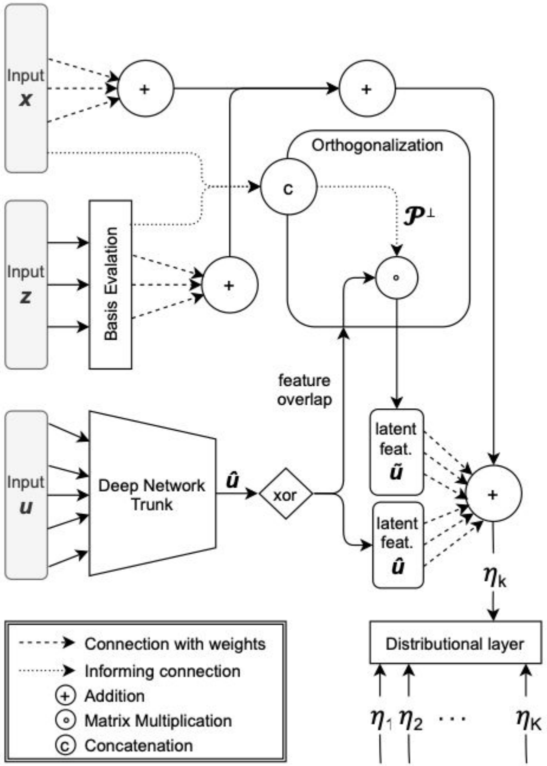
A-Fig.4: Hierarchical Neural Network Architecture (Sardelich and Manandhar, 2018)

SSDDR

Semi-Structured Distributional Regression

$$\eta_k = f_{k,0}(\boldsymbol{x}) + \sum_{j=1}^{r_k} f_{k,j}(z_j) + \sum_{j=1}^{g_k} d_{k,j}(\boldsymbol{u})$$

A-Fig.5: Additive predictor structure including struc. linear, struc. non-linear and unstruc. predictors (Rügamer, Kolb and Klein, 2020)



A-Fig.6: Exemplary architecture of SSDDR (Rügamer, Kolb and Klein, 2020)

Further Use of Remote Sensing

Poverty

- Mapping poverty using mobile phone and satellite data
  (Steele et al, 2017)

Corn Yield Prediction

- Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data
  (You et al, 2017)

Socioeconomic Survey Predictions

- Using Deep Learning and Google Street View to Estimate the Demographic Makeup of the US
  (Gebru et al, 2017)