Deep learning framework for predicting intelligence with multimodal neuroimage as networks

Sangyoon Bae
Interdisciplinary Program of Artificial Intelligence
16-M519, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
stellasybae@snu.ac.kr

Florian Brandl
Computer Science and engineering
301-316 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
flo_bra@snu.ac.kr

Gun Ahn
Interdisciplinary Program in Bioengineering
1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
geonahn@snu.ac.kr

Thomas Lykke Rasmussen
Department of Computer Science and engineering
301-316 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
thomaslykke@snu.ac.kr

Abstract

This study examines the effect of combining different modalities of brain imaging on predicting intelligence. More specifically, we took sMRI, DTI, and fMRI scans, converted them into graphs, and used them as the input for two different models. Both were then evaluated against a baseline which used TabNet and achieved an $RMSE = 0.921$*. The first proposed model used Graphormer, an extension of the classic Transformer architecture that is able to specifically encode graph structural information. Due to the complexity of this model and scarcity of the data, overfitting was experienced independent of hyperparameter configuration. In a second approach the graphs were encoded via Graph2Vec into vector embeddings, that were consequently fed into a Multilayer Perceptron. This model managed to outperform the baseline by achieving an* $RMSE = 0.708$ *when using only sMRI data. However, multimodality and transfer learning still remain further topic of research. The code is made publicly available through our github.*

1. Introduction

Intelligence is the ability of an organism to observe the outside world, think abstractly, argue, plan, learn, and solve problems. Intelligence explains much of human behavior and cognition. Therefore, understanding the development of intelligence and its neurological mechanism in children and adolescents is a very important research topic in developmental neuroscience. In addition, low intelligence in childhood and adolescence is associated with mental disorders such as schizophrenia and major depressive disorders in adulthood [14], and recent clinical neuroscience studies of brain cognitive development will provide important clues to psychiatric research.

Then what characteristics of the brain can explain intelligence? We chose an approach of viewing the brain as a network. This is because intelligence can be explained by network characteristics of the brain, such as connection strength and connection efficiency. [9] [15] [10] [27] In fact, the brain is in the form of a "small-world & scale-free network" in which numerous neurons form a local modular structure but also connect to distant neurons. [6]

Multimodality refers to various types of information that describe the same object, and in this study, it refers to different types of brain imaging data obtained from the human brain. Brain images obtained from humans include three major modalities: structural MRI (hereinafter referred to as sMRI), diffusion MRI (hereinafter referred to as dMRI), and functional MRI (hereinafter referred to as fMRI). sMRI represents the volume and distribution of gray matter, white matter, and cerebrospinal fluid (CSF), dMRI represents the distribution of white fiber traces, and fMRI can indirectly measure neural activity through changes in blood flow. Fusing multimodal neuroimages is in the spotlight for many neuroscientists because other brain images can provide additional information that one brain image does not have. [24]

On the other hand, we are going to add the reason for the attempt to combine multimodal neuroimages beyond the complementarity of information that has been dealt with a lot before. Structural data, sMRI, and dMRI, show a strong correlation because they measure the biological structure of the brain in different ways. Functional data, fMRI, is the data measured above the biological structure of the brain. To explain in more detail, from a perspective of statistical physics, functions arise spontaneously from dynamic changes in structure, and expanding this notion to neuroscience, cognitive function emerges from the dynamic of extended physical and subcortical networks [20].

In fact, structural and functional networks share hubs, and experiments predicting functional connectivity with structural connectivity showed high levels of consistency. [6] That is, brain image data of different modalities are closely related to each other, and this relationship between data must be considered when building deep learning models using multimodal neural image data. However, in current studies using multimodal neuroimage data, data of different modalities tend to be treated as data of the same modality, such as using model stacking algorithms that assume that the data input to the weak learner is the same. [21]

There are two main ways to fuse data of different modalities. The first is joint presentation, which is a method of creating a fancy presentation by combining several modalities. In a joint presentation, data is combined in the input stage, or features are extracted by applying a model suitable for each format to single model data, and the features are converted into the input in other models. The second is coordinated presentations, which create presentations in multiple modalities, and then create interactions between them. In contrast to the framework of multimodal deep learning, which combines different data obtained from the same target, there are also attempts to combine information on different scales obtained from the same target. HiDENN [22] complementarily combines two different scales of information - experimental data and physical data through transfer learning.

So, how can we model the way our brain's structural-functional network interacts? We can offer two possibilities here. First, structural connectivity and functional connectivity share intermediate information. If this hypothesis is correct, it can be implemented as a multi-channel multimodal model. Second, functional connectivity is accumulated on top of structural connectivity. If this hypothesis is correct, it can be implemented as transfer learning. Therefore, we would like to compare the performance of the four models - simply merged TabNet model (Figure 1A), single modality mode (Figure 1B,1C), multi-channel model (Figure 1D), and transfer-learned model (Figure 1E) to select one hypothesis.

In our study, we use a model called a Graphormer. First, in order to express the brain close to its natural state, a structural graph is using sMRI's volume and local thickness information as node, DTI's structural connectivity FA information as edge, and similarly, a functional graph is using sMRI as node and rsfMRI's functional connectivity information as edge. Since the network hub plays an important role as an intermediate feature in combining the structural network and the functional network, the network hub must be found. A Graphormer finds the centrality, which can explain the hub.

2. Method and Materials

2.1. ABCD

ABCD (Adolescent Brain Cognitive Development) study is a large, multi-site, longitudinal study that follows approximately 2000 9- and 10-year-old children through late adolescence to analyze factors that influence developmental trajectories from 21 research sites across the US. [13] ABCD study aims to examine how biological factors and the environmental factors interact and relate to developmental outcomes such as mental health and intelligence.

2.2. NIH Toolbox

The NIH toolbox is a set of simple, psychometrically sound measurement tools for evaluating the motor, emotion, sensation, and cognitive function of people aged 3 to 85. 104 verified measures and standard data are provided in English and Spanish. [11] We used the NIH Toolbox to represent an individual's intelligence. Some researchers found out that some networks in our brain were associated with fluid and crystallized intelligence. [19]

In the NIH toolbox, reading was chosen because it is a proxy for a wide range of cognitive, educational, and socioeconomic factors. The ability to pronounce low-frequency words in irregular spelling has been used as an estimate of overall intelligence. [7]. Vocabulary represents the vocabulary component of a language and is very related to the general measurement of "crystallized intelligence" [4]. Fluid intelligence evaluates the ability known to further reflect biological brain processes that change over life and are sensitive to potential acquired brain damage/disease. [3] Total Composite is a combination of the abbreviated crystallized and fluid scores. [8] Total Cognition Composite was used for our regression analysis because total intelligence is more suitable for common brain dynamics than fluid intelligence and crystallized intelligence which is focusing on specific tasks.

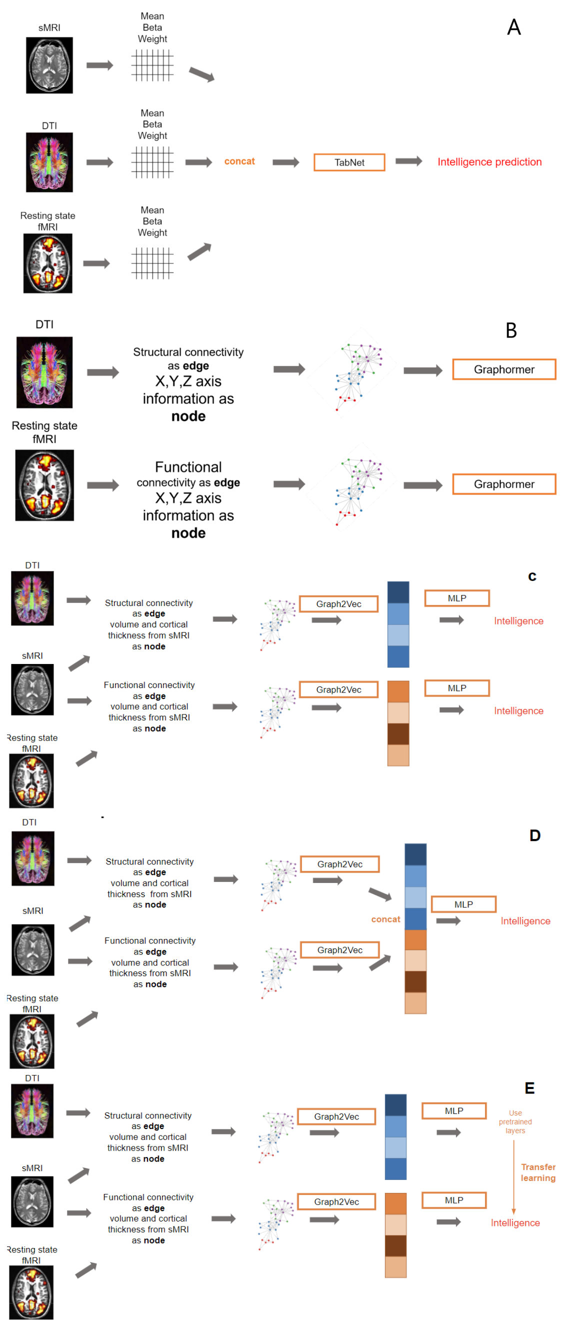

Figure 1. Proposed models: 1A: Merged TabNet model (baseline model), 1B: Single modality model, 1C: single modality model with Graph2Vec, 1D: concatenated modality model with Graph2Vec, 1E: Transfer-learned multimodal mode with Graph2Vec

2.3. TabNet (baseline)

Machine learning methods such as XGboost or LightGBM have been most used for tabular data input because of interpretability and less complexity. [12]. However, machine learning has a problem that it does not provide end to end learning. The recent development of TabNet has enabled the application of end-to-end deep learning with built-in interpretability. TabNet is based on the attention mechanism that softly selects features to reason from at each decision step and aggregates the processed information to make a final prediction. The model uses sequential attention to choose which features to deduce from at each decision step, enabling efficient learning. TabNet learns very efficiently in each decision-making step by explicitly selecting sparse features using sequential attention mechanisms to utilize relevant variables that yield high-performance model results for variables that are fully related to model capacity. The sparsity enables more interpretable decision-making through visualization of variable selection masks. [2]

2.4. Graphormer

Attention has become more and more the norm for various machine learning tasks. Nonetheless, it has not translated to the area of Graph Neural Networks. The recently proposed Graphormer [5] tries to bridge that gap by adding the structural information of the graph to the classical transformer [25] architecture. It does so by extending the original idea by adding three different encodings; The centrality encoding adds learnable centrality embeddings to the node features that are dependent on the number of edges attached to that specific node. By doing this, the Graphormer aims to give very connected nodes more importance. Secondly, spatial encoding adds a bias in the self-attention module that is determined by the shortest distance between two nodes. Two node-pairs that have the same shortest distance therefore get the same bias added, helping to encode spatial information. Lastly, the edge encoding adds a second term in the self-attention module, which lets the network adjust the weights for the k-nearest edges by aggregating them in the target node. These weights can be set to $-\infty$, the Graphormer therefore could only look at the 1-nearest edges and can therefore be understood as an extension of original Graph Neural Networks. [28]

2.5. Graph2Vec

Another model that was considered was a regular Multilayer Perceptron taking as input graph embeddings created by graph2vec. [18] In this paper, Narayanan et al. propose a embedding method very similar to that of word2vec [17] or doc2vec [16]. Compared to those approaches, graph2vec assumes that subgraphs compose a graph in a similar way than words compose a sentence or document. As such, the model takes all subgraphs as vocabulary and then trains a skipgram model that tries to maximize the probability that a certain subgraph is in the graph. As the vocabulary of all subgraphs can be huge, negative sampling is used: Only a few subgraphs are selected as negative samples and consequently only the embeddings of those are trained, not the embeddings of the whole vocabulary. To ultimately get the graph embedding, only the weights of the single hidden layers are kept, with which now the embedding can be computed.

One problem with using the graph2vec encoder is the fact that while it allows for node features, it does not take edge features into account. To account for this, we removed all edges that had less connectivity than a set hyperparameter(in our experiments 0.2) to make the otherwise nearly fully connected graphs more sparse.

2.6. Proposed Model

Our proposed model mainly leverages the Graphormer architecture, with additional layers to aggregate the features of the graphs from different modalities and predicting our single target, intelligence. Figure 1B illustrates the first approach, where we try to implement a single modality model that either takes the structural connectivity of DTI data or the functional connectivity of fMRI data to form a graph. These graphs have the connectivity value between nodes as the only edge feature, and the spatial coordinates X, Y, and Z as the three node features. The Graphormer then takes one of these graphs as input, attending to the most important features of the graph and ultimately regressing to the desired target. In the second step, this single modality model then gets compared to the performance of multi-channel models that incorporate all the data available. Figure 1C-E shows the approach with Graph2Vec.

Figure 1C shows usage of Graph2Vec that we made vectors that can be represented in embedding space. We made structural graph data by using structural connectivity (DTI) as edge volume and cortical thickness from sMRI as node. Also, We made functional graph data using functional connectivity (rsfMRI) as edge volume and cortical thickness from sMRI as node. After that, we put our embedding vector to MLP to predict individual's intelligence

Two multi-channel models are proposed: The first model is multi-channel, which can be seen in figure 1D and aggregates the embedding vector of both single modality and tries to regress the target. The second model uses transfer-learning, trying to use the weights of the first MLP with structural embedding vectors in shared layers to achieve a better performance on the training of the MLP with functional connectivity embedding vectors. This can be seen in figure 1E.

2.7. BrainNet

The brain networks were visualized with the BrainNet Viewer [26]. It needs node and edge information. BrainNet Viewer is a graph-theoretical network visualization toolbox which can help researchers to visualize structural and functional connectivity patterns. We used location of the DTI axis as node information, and correlation between each node as edge information. Each node has the same size, and depth of edges represent the correlation. The stronger relation it is, the thicker in the graph edge.

3. Results

Table 1. Total intelligence prediction with TabNet

Data	Test RMSE	# Features
Resting state fMRI	0.977	147
Diffusion MRI	0.928	1185
Structural MRI	0.921	1184
Whole Brain (Concatenated, RsfMRI+DTI+SMRI)	**0.910**	2516

We implemented baseline models, which simply concatenated mean beta weight of structural MRI, DTI, and resting state fMRI in tabular form. We used the root-mean-squared-error score as the metric of model performance. Individually, The RMSE in the test dataset of sMRI was 0.977, DTI was 0.928, and rsfMRI was 0.921. When we merged sMRI, DTI, rsfMRI dataset in the data input stage, the result of our simple multimodal model was 0.910.

We implemented the singular modal model for both the structural and functional data using Graphormer. Despite the fact that training loss decreased in all instances of our comprehensive and structured attempts to finetune the hyperparameters; dropout rate, attention dropout rate, input dropout rate, hidden dimension, and number of layers. The model did not gain any extra underlying information by being trained. More specifically figure 4 illustrates how the validation loss instantly increased when training began no matter what configuration of hyperparameters that were applied.

In our Graph2Vec+MLP experiment, The RMSE score in the test dataset of functional was 0.717, and struct was 0.708. When we merged the functional, struct dataset in the embedding stage, the result of our multimodal model was 0.866. In transfer learning that train in struct first and train again in functional later, the RMSE was 0.723. Although it was in embedding space, simply merging and transfer learning did not help to improve performance.

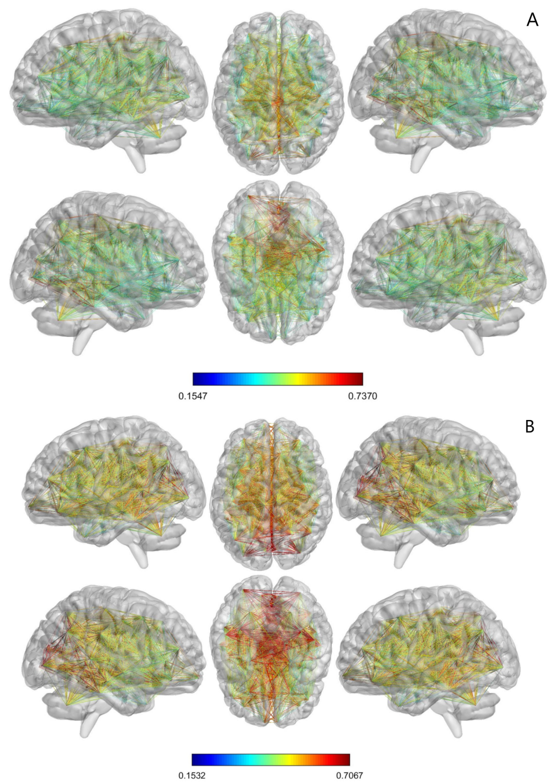

Figure 2. A The brain connectivity of the one who has the highest intelligence, Figure 2B:The brain connectivity of the one who has the lowest intelligence, Visualization of the brains showed that there are clear differences between individual connectivity. There are more connections through the brain in the lower intelligence brain than the one for the higher intelligence. For instance, the one who has the lowest intelligence has a total sum of connectivity of 1129, while the highest one has 939.

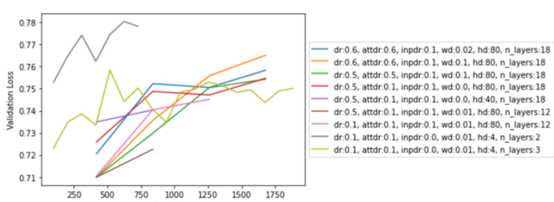

Figure 3. Validation loss change through the iterations various hyperparameters *dr: dropout rate, attdr: attention dropout rate ,inputdr: input dropout rate, hd: hidden dimensions , n_layer: number of layers

4. Discussion

In this paper, we tested deep neural networks trained on the largest youth brain multimodal MRI data to predict youth intelligence.

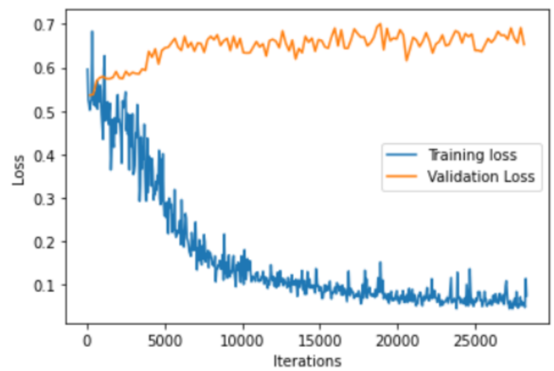

Figure 4. Training and validation loss through the iterations for structural fMRI for graphormer with dropout rate:0.1, attention dropout rate: 0.1, input dropout rate:0.1, hidden dimension:80 and number of layers:12.

Table 2. Total intelligence prediction with Graph2Vec

Data	RMSE	# Features
Functional(rsfMRI+SMRI)	0.717	128
Structural(DTI+SMRI)	**0.708**	128
Whole Brain (concatenated, Functional+Structural)	0.860	256
Whole Brain (Transferred, Structural to Functional)	0.723	128

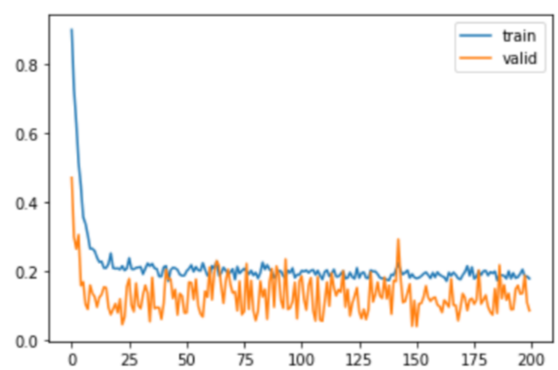

Figure 5. single modality model - structural (RMSE: 0.7078)

To begin with, in our baseline model, RMSE from the rsfMRI (0977) and DTI (0.928) was bigger than one from sMRI(0.921). When we simply merged the three datasets (0.910), RMSE got slightly smaller than one from the single modal model. In our result, compared to the added information, the decrease in RMSE was small. This is because the data we used is in tabular form, not reflecting the brain's natural state. Therefore, we needed to build a multimodal model which preserves the natural state of the brain like graph form. [6]

In visualization, we found that there are more connections through the brain in lower intelligence than in the brain with higher intelligence.[1] noted that the number of dendrites is important for efficient function to predict one's intelligence. Further, it is known that brain connectivity and IQ have negative correlation for boys. [23] Therefore, our findings through visualization are consistent with previous studies.

As we can see from the figure 3 and 4 validation loss was increasing throughout the iterations and the same phenomena occured in every trial with a differing combination of hyperparameters. [5] reported that Graphormer is more easily trapped in overfitting problems because of the large size of the model and the small size of the dataset. In addition, models with complex structures are known to tend to overfit training data to small datasets such as brain MRI. [1] To sum up, we concluded that Graphormer is too complex and heavy for our small MRI datasets.

In the Graph2Vec model, the RMSE got smaller ($0.70 \sim 0.86$) overall compared to the baseline model ($0.91 \sim 0.977$). This is because the model was trained while maintaining the properties of the networks as much as possible. [6] Contrary to our expectations, the method of simplifying the vector in the embedding space showed a decrease in performance. We concluded that It is hard to learn interaction between different modalities by simply merging. Brain data is especially known for big noise compared to the signal. [1] Therefore, if we can not get extra information about interactions in multimodalities, it might decrease performance because of noise effects. In a similar sense, transfer learning did not improve the performance as well.

5. Conclusion

To conclude, this paper was able to show that the performance of the simple TabNet baseline was able to be overcome by using Graph2Vec embeddings and leveraging the explicit structure of the graph. However, neither the approach of using multimodality nor transfer learning were able to outperform the single modality error of just sMRI and DTI as a graph data. Additionally, it could also be seen that some of the current state of the art models for graph classification and regression like Graphormer are very complex, requiring a lot of data to achieve good performance. With complex structures such as brain MRI being even more prone to overfitting, and the dataset only containing around 2000 graphs(when a lot of other common graph datasets have more than $40,000$ graphs), the size of the dataset seems to be one of the biggest limitations.

[1] (Vernon, P. A , 2013)

References

[1] Parnian Afshar, Konstantinos N. Plataniotis, and Arash Mohammadi. Capsule networks for brain tumor classification based on MRI images and course tumor boundaries. *CoRR*, abs/1811.00597, 2018. 6

[2] Sercan Ömer Arik and Tomas Pfister. Tabnet: Attentive interpretable tabular learning. *CoRR*, abs/1908.07442, 2019. 3

[3] S. Goodnight K. B. Casaletto A. Goldsmith R. K. Heaton A. W. K. WongC. M. Baum R. Gershon A. W. Heinemann Carlozzi, N. E. and D. S. Tulsky. Validation of the nih toolbox in individuals with neurologic disorders. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 32(5):555–73, 2017. 3

[4] Raymond Bernard Cattell. *Intelligence: Its structure, growth and action*. Elsevier, 1987. 2

[5] Shengjie Luo Shuxin Zhng Guolin Ke Di He Yanming Shen Chengxuan Ying, Tianle Cai and Tie-Yan Liu. Do transformers really perform bad for graph representation? *CoRR*, abs/2106.05234, 2021. 3, 6

[6] Bullmore E and Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Bullmore E, Sporns O.*, 10(3):186–198, 2009. 1, 2, 6

[7] E. Grober, M. Sliwinsk, and S. R. Korey. Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology*, 13(6):933–949, 1991. 2

[8] Tulsky D et al. Heaton RK, Akshoomoff N. Reliability and validity of composite scores from the nih toolbox cognition battery in adults. *Int Neuropsychol Soc.*, 20(6):588–598, 2014. 3

[9] Cornelis J. Stam René S. Kahn Heuvel, Martijn P. van den and Hilleke E. Hulshoff Pol. Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(23):76119–24, 2009. 1

[10] Matthias Ekman Christian J. Fiebach Hilger, Kirsten and Ulrike Basten. Efficient hubs in the intelligent brain: Nodal efficiency of hub regions in the salience network is associated with general intelligence. 60(1):10–25, 2017. 1

[11] Landis SC; NIH Blueprint Hodes RJ, Insel TR. The nih toolbox: setting a standard for biomedical research. 80(11 Suppl 3):12–80, 2013. 2

[12] Jengei Hong. An application of xgboost, lightgbm, catboost algorithms on house price appraisal system. 2020. 3

[13] Chandni Sheth Perry F. Renshaw Deborah A. Yurgelun-Todd Huber, Rebekah S. and Erin C. McGlade. Suicide ideation and neurocognition among 9- and 10-year old children in the adolescent brain cognitive development (abcd) study. *Archives of Suicide Research: Official Journal of the International Academy for Suicide Research*, 9:1–15, 2020. 2

[14] Terrie E. Moffitt Andrea L. Roberts Laurie T. Martin Laura Kubzansky Honalee Harrington Richie Poulton Koenen, Karestan C. and Avshalom Caspi. Childhood iq and adult mental disorders: A test of the cognitive reserve hypothesis. *The American Journal of Psychiatry*, 166(1):50–57, 2009. 1

[15] Andreas Pedroni Lorena R. R. Gianotti Jürgen Hänggi Daria Knoch Langer, Nicolas and Lutz Jäncke. Functional brain network efficiency predicts intelligence. *Human Brain Mapping*, 33(6):1393–1406, 2012. 1

[16] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR, 2014. 4

[17] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013. 4

[18] Annamalai Narayanan, Mahinthan Chandramohan, Rajasekar Venkatesan, Lihui Chen, Yang Liu, and Shantanu Jaiswal. graph2vec: Learning distributed representations of graphs. *arXiv preprint arXiv:1707.05005*, 2017. 4

[19] Lindsay M. Squeglia Kara Bagot Joanna Jacobus Rayus Kuplicki Florence J. Breslin Jerzy Bodurka Paulus, Martin P. Screen media activity and brain structure in youth: Evidence for diverse structural correlation networks from the abcd study. *NeuroImage*, 185(1):140–53, 2019. 2

[20] Steven E. Petersen and Olaf Sporns. Brain networks and cognitive architectures. *Neuron*, 88(1):207–19, 2015. 2

[21] Narun Pornpattananangkul, Yue Wang, Richard Anney, Lucy Riglin, Anita Thapar, and Argyris Stringaris. Longitudinally stable, brain-based predictive models explain the relationships of childhood intelligence with socio-demographic, psychological and genetic factors. *bioRxiv*, 2021. 2

[22] Zhengtao Gan Lin Cheng Jiaying Gao Orion L. Kafka Xiaoyu Xie Hengyang Li Mahsa Tajdari H. Alicia Kim Saha, Sourav and Wing Kam Liu. Hierarchical deep learning neural network (hidenn): An artificial intelligence (ai) framework for computational science and engineering. *Computer Methods in Applied Mechanics and Engineering*, 373(1):113452, 2021. 2

[23] Vincent J. Schmithorst. Developmental sex differences in the relation of neuroanatomical connectivity to intelligence. *Intelligence*, 37(2):164–173, 2009. Intelligence and the Brain. 6

[24] Kâmil Uludağ and Alard Roebroeck. General overview on the merits of multimodal neuroimaging data fusion. *NeuroImage*, 102(1):3–10, 2014. 2

[25] Shazeer Noam Parmar Niki Uszkoreit Jakob Jones Llion Gomez Aidan N Kaiser Ł ukasz Vaswani, Ashish and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 3

[26] Jinhui Wang Xia, Mingrui and Yong He. Brainnet viewer: A network visualization tool for human brain connectomics. *PloS One*, 8(7), 2013. 4

[27] Won-Gyo Shin Tae Young Lee Ji-Won Hur Kang Ik K. Cho William Seunghyun Sohn Seung-Goo Kim Kwang-Hyuk Lee Yoon, Youngwoo Bryan and Jun Soo Kwon. Brain structural networks associated with intelligence and visuomotor ability. *Scientific Reports*, 7(1):2177, 2017. 1

[28] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81, 2020. 3