Multimodal Knowledge Graph for Link Prediction and Node Classification in ADNI and PPMI Datasets

NTU-CE7455: Project Final Report

Ong Zhi Lin, Charlene
School of Computer Science and Engineering
Nanyang Technological University
{ongz0070}@ntu.edu.sg

Abstract

Real world data often comes in the form of multiple modalities, such as images, text and numerical data. These data are often relational, such as a patient medical records. However, there has not been much research in modelling multimodal data, especially in the medical imaging domain. In this work, we attempt to address this through the construction of our own knowledge graph consisting of medical images, medical notes and patient metadata, which is then encoded with a relational graph convolutional network for the tasks of link prediction and node classification. We demonstrated our approach on the Alzheimer's Disease Neuroimaging Initiative Dataset and the Parkinson's Progression Markers Initiative Dataset and obtained a test accuracy of 0.637 for node classification and a test mean reciprocal rank of 0.101 for link prediction on a small dataset of 509 patients.

1 Key Information

- Mentor: Chan Yi Hao
- Publication plan: I plan to submit an improved version of this work to AAA1 2022 on September 9, 2021.

2 Introduction

In the recent years, there has been an advancement of efforts to develop approaches to model relations among different entities, which are often best represented using a graph structure, as well as an explosion of knowledge bases such as YAGO [1] and FB15k[2]. However, there remains a sparsity of approaches which are developed to model different data types, such as numerical, text and images. In addition, many of the approaches that have been developed are applied to large standard datasets such as YAGO, and there has been a dearth of efforts that attempt to address, or even build up, knowledge graphs in other domains. In particular, in the medical domain, knowledge graphs have mainly been developed on electronic health records. [1], but few have been applied to the medical imaging domain.

In this work, we attempt to utilise the recent advances in knowledge graphs embeddings and graph convolutional networks on a multimodal dataset from the Alzheimer's Disease Neuroimaging Initiative [2] and the Parkinson's Progression Markers Initiative to identify possible missing entities in the knowledge graph, i.e., the task of link prediction, and to predict the possibility of a patient possibly having Alzhiemer's or Parkinson's disease, i.e. the task of entity classification. We also demostrated the potential applicability of knowledge graph approaches to model multimodal data, and even on small datasets.

[1] https://yago-knowledge.org/
[2] https://deepai.org/dataset/fb15k-237

3 Related Work

There has been a growing interest in multimodal and multi-relational approaches on knowledge bases over the years. In statistical relational learning (SRL), knowledge bases can often be thought as to consists of a multitude of triples, which are made up of a subject, predicate and object, which is denoted by $\langle s, p, o \rangle$ [3]. An example of a triple could be {*John*, *has_age*, *10*}, where both *John* and *10* are entities while the relation is *has_age*. Categories can also be assigned to the entities, for example, *John* is a *child*. In multimodal entities, *John* could not only have an age, but also a photo of himself, which would be in the form of an image. There hence exists multiple relations for John, such as *has_age* and *has_photo*.

There have been two main approaches to model multimodal knowledge graph. The first approach primarily involves the use of graph convolutional networks by utilising the adjacency matrix of the graph. The use of a graph convolutional network to model relational data is first proposed by Schlichtkrull et al[1], which encodes the entities in the relational graph for both entity classification and link prediction. This is then applied by Wilcke et al [3] for multiple modalities, where a multimodal node embedding $\mathbf{H}^0$ is fed together with the adjacency matrix $\mathbf{A}^r$ into the relational graph convolutional network.

In the second approach, shallow and linear models [4] are usually used to learn the embeddings of all entities and relations within a knowledge graph. [5] They often work by defining a scoring function on triples that outputs a score in $[0, 1]$ or $[-1, 1]$, which scores the relation between the subject and the object. [6, 7] Some commonly used knowledge graph embedding models include TransE [8], DistMult[9] and ComplEx[10]. Peseshkpour et al[11] adopted this approach by introducing the use of additional neural encoders to embed multimodal data and then utilising ConvE or DisMult to produce a score to indicate the probability of triple being correct.

4 Approach

We propose the use of a multimodal relational graph convolutional network for the task of node classification and link prediction. Our methodology consist of several steps, from preparation of triples and their respective embeddings to training a relational graph convolutional network. For the task of link prediction, we utilise a DisMult decoder to produce a score regarding the validity of the triple.

4.1 Preparation of Triples

The dataset must first be represented in the form of triples formulated as $\langle s, p, o \rangle$. A record of nodes and their data type, as well as a record on the relations present in the model, are constructed. The data type refers to a label which denote the category of the node, whether it is categorical, numerical, text and images. The dataset consisting of entities which are patients, are categorised according to diseased or non-diseased. Pruning of the dataset is also performed where the number of triples within 2-hop neighbourhood around the subject is kept, to save memory. The entities of the dataset is grouped by data type for the creation of embeddings.

4.2 Creation of Embeddings

Numerical and Categorical Data For categorical data such as age, we embed them with an embedding layer. A simple feedforward layer is used to embed numerical data.

Image Data As the images used in our methodology are three dimensional, a pretrained 3D SqueezeNet is used as the embedding for the images [3]. The pretrained SqueezeNet is pretrained on the Kinetics Dataset. [12] Principal Component Analysis (PCA) is then performed on the feature representations of the images to reduce the embeddings to a lower dimension.

Text Data In the metadata for the patients in the Parkinson's Progression Markers Initiative Dataset, there are medical notes regarding the current medical conditions of the patient. To embed

[3]The pretrained model is downloaded from https://github.com/okankop/Efficient-3DCNNs

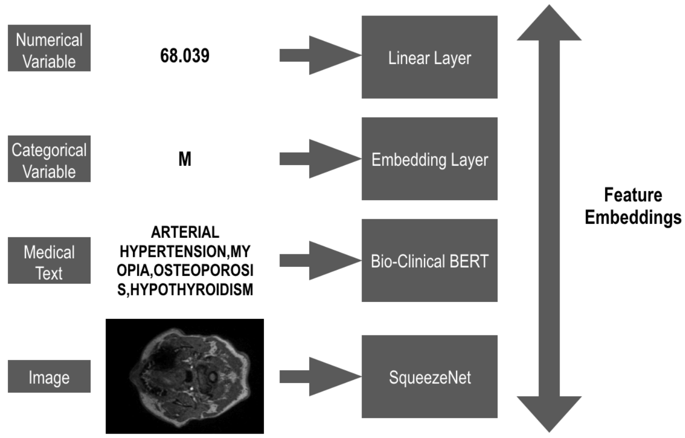

Figure 1: Embeddings of Multimodal Data

these text, the Bio_ClinicalBERT tokenizer and model is used. [13] The Bio_ClinicalBERT is pre-trained on all the notes from MIMIC III database, which is a database consisting of electronic health records from ICU patients from Boston, MA. [14] Similar to image data, PCA is also performed on the text embeddings to reduce a lower dimension.

4.3 Preparation of Adjacency Matrix

We consider the possibility of inverse relations and self loops and enriched the relations with them. A sparse adjacency matrix is then constructed based on the subject, predicate and object entities in the triples, which is then passed into the graph together with the initialised embeddings on the multimodal data. This is done by performing a matrix multiplication of the adjacency matrix and the feature embeddings, which will then be passed into a graph convolutional layer.

4.4 Encoding with Relational Graph Convolutional Network

Our methodology is a relational graph convolutional network which consists of two graph convolutional layers. The network is adapted from Schlichtkrull et al[1], where the forward pass for the entity through a single graph convolutional layer can be expressed with the following equation.

$$h_i^{(l+1)} = \sigma(\sum_{r \in R} \sum_{j \in N_i^r} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)}) \tag{1}$$

where $h_i(l+1)$ is the next hidden representation of the graph convolutional network. In this forward pass, the network is summing over all types of relations R and all the neighbor nodes N_i^r of node i with the relation R. The presence of $W_0^{(l)} h_i^{(l)}$ accounts for self loops. A matrix multiplication is then performed between the embedding of the first graph convolutional layer and the adjacency matrix, before being passed into the next layer. Finally for the second layer, which is the last layer, a softmax activation is used. The model is trained by minimising the cross entropy loss on the labeled entities in the training dataset, with the predicted probabilities from the network. A penalty in the form of L2 regularisation is introduced during model training.

In order to address the issue of overfitting, basis decomposition is also introduced. Each of the weights of the RGCN is then modelled as follows.

$$W_r^{(l)} = \sum_{b=1}^{B} a_{rb}^{(l)} V_b^{(l)} \tag{2}$$

The weights are expressed as a linear combination of basis transformations $V_b^{(l)}$ with coefficients $a_{rb}^{(l)}$. B represents the total number of bases.

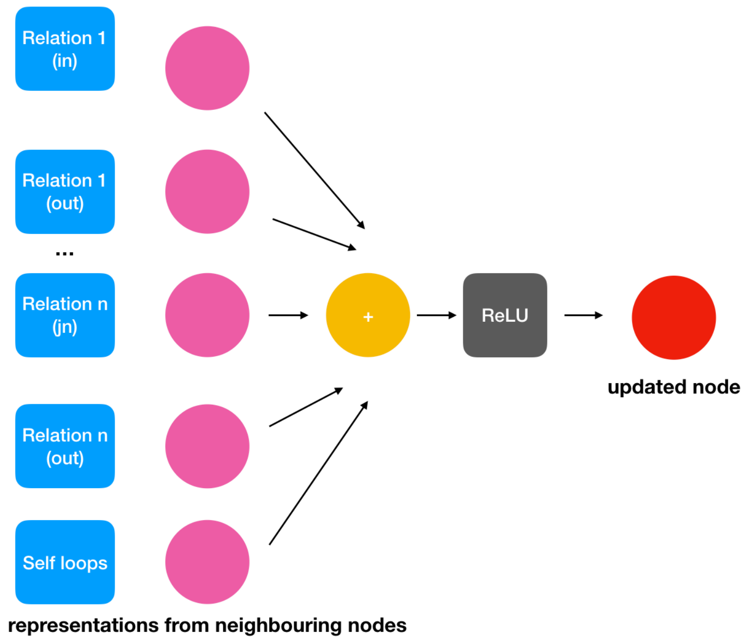

Figure 2: Forward Update of RGCN

4.5 Decoding with Knowledge Graph Embedding Models

For the task of link prediction, the encoded feature representation is then decoded with a knowledge graph embedding model, DisMult. [9] DisMult uses a basic bilinear scoring function where the relation is modelled with a diagonal matrix.

$$g_r^b(\mathbf{y}_{e1}, \mathbf{y}_{e2}) = \mathbf{y}_{e1}^T \mathbf{M}_r \mathbf{y}_{e2} \tag{3}$$

Where $g_r^b(\mathbf{y}_{e1}, \mathbf{y}_{e2})$ is the score, $\mathbf{y}_{e1}^T$ is the tensor of the first entity, $\mathbf{M}_r$ is a relational diagonal matrix, and $\mathbf{y}_{e2}$ is the tensor of the second entity.

Using negative sampling, the training triples are corrupted with a ratio of 5 to 1, i.e. for every 5 true triples, there is 1 corrupted triple. The triple is corrupted by randomly replacing either the subject or the object entity with any entity in the dataset. The ground truth labels are defined such that true triples are assigned a label of 1 and corrupted triples are assigned a label of 0. The model is then trained by minimising the binary cross entropy loss between the ground truth labels and the scores which are computed by DisMult for both the true and corrupted triples, such that we maximise the score given to the true triples and minimise the score that is given to the corrupted triples.

5 Experiments

5.1 Data

The datasets consist of the Alzheimer's Disease Neuroimaging Initiative Dataset (ADNI) [2] and the Parkinson's Progression Markers Initiative Dataset (PPMI) [15]. The datasets from Alzheimer's Disease Neuroimaging Disease Neuroimaging Initiative consist of preprocessed MRI 3D-scans in NifTI format, together with metadata from the images consisting of patient demographics, APOE allele score, cognitive function scores, depression scores and clinical scores. There are 282 patients in the dataset. In the PPMI dataset, there are similar preprocessed MRI 3D scans in NifTI format. These images are then matched to a number of demographic, olfactory, sleep, anxiety, depression and cognitive data for the same patient. Text records for the patient in the form of current medical conditions are also taken. There are 227 patients in this dataset. Both datasets are then combined, resulting in a total of 509 patients.

The patients are then assigned to either diseased or non-diseased. Diseased patients are patients which have Parkinson's Disease, patients which clinically appear to have Parkinson's but do not show a dopaminergic deficit (SWEDD) [16], patients with Alzheimer's Disease, patients with late mild cognitive impairment (LMCI) and patients with mild cognitive impairment (MCI). Non-diseased patients are patients which are cognitively normal (CN), have early mild cognitive impairment (EMCI) or with subjective memory concerns (SMC). Overall, the number of patients categorised as diseased is 255 while the number of patients categorised as diseased is 254. Feature engineering is performed on the raw data for Epworth Sleepiness Score (ESS), State and Anxiety Scores and REM Sleep Disorder Scores to obtain the total score to feed into the model development.

Table 1: Relations into the Model and Their Source

Predicate	Object	Source
hasImage	MRI scans	PPMI and ADNI
hasWeightKg	Weight	PPMI and ADNI
hasSubjectSex	Gender	PPMI and ADNI
hasSubjectAge	Age	PPMI and ADNI
hasAPOEA1	APOE A1	ADNI
hasAPOEA2	APOE A2	ADNI
hasMMSCORE	Mini-mental State Examination Score	ADNI
hasGDTOTAL	Geriatric Depression Score	ADNI
hasCDGTOTAL	Clinical Dementia Global Rating Score	ADNI
hasESS	Epworth Sleepiness Scale	PPMI
hasUPSIT	University of Pennsylvania Smell ID Test Score score	PPMI
hasMoCa	Montreal Cognitive Assessment Score	PPMI
hasSTAI	State and Anxiety Scores	PPMI
hasREM	REM Sleep Disorder Score	PPMI
hasCONDTERM	Text Embeddings	PPMI

5.2 Evaluation method

5.2.1 Node Classification

In the task of node classification, the patient is simply being classified as non-diseased or diseased. As the output of the network is the probabilities, the label with the maximum probability is taken to be the predicted label. The commonly used evaluation metric in node classification is accuracy, which is the fraction of correct predictions by the network.

5.2.2 Link Prediction

In the task of link prediction, the model's ability to predict a missing link is being evaluated. The evaluation is done by generating a set of corrupt triples by replacing either the head or tail entity with any entity present in the dataset. The DisMult score for both the corrupted and the non-corrupted triples is calculated. A ranking of the scores of the corrupted and non-corrupted triples is performed. In our approach, if the non-corrupted triple has the same score as the corrupted triple, we would consider the non-corrupted triple to have a better rank in half of the samples. The following metrics can then be computed based on the rank.

Mean Reciprocal Rank MRR computes the mean of the reciprocal ranks of the non-corrupted triples in the test dataset.

$$MRR = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{rank_i} \tag{4}$$

where n represents the number of non-corrupted test triples, and $rank_i$ represents the rank of the ith test triple.

Hits@k The metrics, Hits@1, Hits@3, Hits@10 and Hits@100 are computed. Hits@k is an evaluation metric that evaluates the percentage of non-corrupted test triples which have the first rank (i.e. most highly scored) in the event of Hits@1, have a rank within the top k possible triples for Hits@3, Hits@10 and Hits@100.

5.3 Experimental details

The numerical, categorical, image and text data are encoded with a 16-dimensional embedding vector for each entity. The RGCN network is initialised with 40 bases for basis decomposition. For the task of node classification, a penalty for L2 normalisation of the weights of the first graph convolutional layer is added to the training loss with a coefficient of 0.0005. The model is trained for 2000 epochs with early stopping on a NVIDIA Tesla P100 GPU. An adam optimizer is used with a learning rate of 0.001 and no weight decay. A patience of 20 epochs is introduced.

For the task of node classification, the dataset is split by patient into train and test with a ratio of 80% for train and 20% for test. For the train dataset, it is further split into training and validation dataset with a ratio of 80% to 20%. The model is trained on the training dataset and the best model in terms of the lowest validation loss is saved, and then tested on the test dataset.

For the task of link prediction, the dataset is split by triples into train and test with a ratio of 80% for train and 20% for test. Similar to the task of node classification, it is further split into a ratio of 80% train and 20% validation. The model is trained on the training dataset and the best model in terms of the highest Mean Reciprocal Rank for the validation dataset is saved. The best model is then tested on the test dataset.

5.4 Results

5.4.1 Node Classification

The size of the test dataset for node classification is 102 patients. The test accuracy for the test dataset is 0.637.

5.4.2 Link Prediction

There are 411 triples in the test dataset. The MRR for the test dataset is 0.101.

Table 2: Hits@k metrics

Hits@1	Hits@3	Hits@10	Hits@100
0.065	0.12	0.15	0.37

While the MRR and the Hits@k is not high, it must be noted that there is a large number of entities, 2446 in the dataset, with 4104 triples in total. Unfortunately, there are no baselines that we can currently compare our results with, as there have been no efforts so far on ADNI and PPMI datasets. However, if we compare to baselines performed on FB15k-237 datasets using RGCN and DisMult, the raw MRR reported by [17] is 0.156 and Hits@1, Hits@3 and Hits@10 is 0.153, 0.264 and 0.417 respectively. However, this is extremely dataset dependent, with MRR as high as 0.561 reported on the WN18 dataset.

6 Analysis

6.1 Modelling without Image Data

To test the effect of including MRI scans in our approach, we removed images and their relations from our dataset. The test accuracy for node classification is 0.588 while the test MRR for link prediction is 0.108.

It can be seen that while the test accuracy for node classification is slightly lower than the test accuracy with images, the model performs better in the task of link prediction for Hits@3, Hits@10 and Hits@100. It must be noted that the number of entities in our dataset after removing images

Table 3: Hits@k metrics when Images are Removed

Hits@1	Hits@3	Hits@10	Hits@100
0.040	0.13	0.23	0.52

is reduced to 1192, and hence it is expected that the model will perform better in link prediction as there are less entities.

6.2 Modelling without Text Data

To test the effect of including medical notes for PPMI patients in our approach, we removed medical notes and their relations in our dataset. The test accuracy for node classification is 0.627 while the test MRR for link prediction is 0.080.

Table 4: Hits@k metrics when Text are Removed

Hits@1	Hits@3	Hits@10	Hits@100
0.040	0.079	0.16	0.30

It can be seen that while the test accuracy for node classification is only marginally lower, there is a significant difference in the test MRR, Hits@1, Hits@3, and Hits@100 for the test dataset when we removed text. The number of entities is 2316 when we removed text, compared with 2446 when all the modalities are present.

6.3 Conclusion from Abalation Studies

Overall, from ablation studies when we removed text and images, it can be seen that incorporating multimodality data improved the model performance for link prediction and node classification.

6.4 Comparison with using SFCN Network for Image Embeddings

Currently, the pretrained convolutional neural network used for embedding the MRI scans is 3D SqueezeNet trained on Kinetics Dataset. We hypothesized that using the SFCN network for initialising image embeddings would result in a better performance as the SFCN is pretrained on brain MRI images from UK Biobank data, hence it is hypothesized that this will lead to better image embeddings. [18]

Table 5: Comparison of Node Classification Test Accuracies between SqueezeNet and SFCN

SFCN	SqueezeNet
0.656	0.637

From the comparison of metrics from link prediction and node classification in Table 5 and 6, it can be seen that using the SCFN network causes the model to perform marginally better in node classification but significantly worse in link prediction. The marginal improvement in node classification may just be due to the small test dataset size. Overall, it appears that using 3D SqueezeNet to SFCN provides a better image embedding for the model.

7 Conclusion

In conclusion, we have demostrated the use of a multimodal knowledge graph for link prediction and node classification on ADNI and PPMI datasets. The purpose of this work is a proof of concept to demostrate the feasibility of encoding patient images, medical notes and metadata and training a model on the constructed knowledge graph. Currently, the dataset used is small with many entities. In addition, the data is pre-encoded with pretrained embeddings for image and text, instead of being

Table 6: Comparison of Link Prediction Metrics between SqueezeNet and SFCN

Image Embeddings	MRR	Hits@1	Hits@3	Hits@10	Hits@100
SqueezeNet	0.101	0.65	0.12	0.15	0.37
SFCN	0.048	0.018	0.051	0.12	0.37

trained together with the RGCN. Nevertheless, we observed that the model is still able to learn important graph representations, achieving a MRR of 0.101 even with a large number of entities.

In the future, we would like to further extend the work by incorporating more relations and patients, as well as including additional modalities such as audio. It would also be interesting to explore improving the RGCN, such as by using Multi-head attention in the model.

References

[1] Maya Rotmensch, Yoni Halpern, Abdulhakim Tlimat, Steven Horng, and David Sontag. Learning a health knowledge graph from electronic medical records. *Scientific reports*, 7(1):1–11, 2017.

[2] the ADNI team. *ADNIMERGE: Alzheimer's Disease Neuroimaging Initiative*, 2021. R package version 0.0.1.

[3] WX Wilcke, Peter Bloem, Victor de Boer, RH van t Veer, and FAH van Harmelen. End-to-end entity classification on multimodal knowledge graphs. *arXiv preprint arXiv:2003.12383*, 2020.

[4] Aisha Mohamed, Shameem Parambath, Zoi Kaoudi, and Ashraf Aboulnaga. Popularity agnostic evaluation of knowledge graph embeddings. In *Conference on Uncertainty in Artificial Intelligence*, pages 1059–1068. PMLR, 2020.

[5] Matteo PALMONARI and Pasquale MINERVINI. Knowledge graph embeddings and explainable ai. *Knowledge Graphs for Explainable Artificial Intelligence: Foundations, Applications and Challenges*, 47:49, 2020.

[6] Eyvind Niklasson. Knowledge graph embeddings - what are they?, Mar 2019.

[7] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[8] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Neural Information Processing Systems (NIPS)*, pages 1–9, 2013.

[9] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*, 2014.

[10] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*, pages 2071–2080. PMLR, 2016.

[11] Pouya Pezeshkpour, Liyan Chen, and Sameer Singh. Embedding multimodal relational data for knowledge base completion. *arXiv preprint arXiv:1809.01341*, 2018.

[12] Joao Carreira, Eric Noland, Chloe Hillier, and Andrew Zisserman. A short note on the kinetics-700 human action dataset. *arXiv preprint arXiv:1907.06987*, 2019.

[13] Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. Publicly available clinical bert embeddings. *arXiv preprint arXiv:1904.03323*, 2019.

[14] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[15] Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, et al. The parkinson progression marker initiative (ppmi). *Progress in neurobiology*, 95(4):629–635, 2011.

[16] Fabienne S Sprenger, Klaus Seppi, Atbin Djamshidian, Eva Reiter, Michael Nocker, Katherina Mair, Georg Göbel, and Werner Poewe. Nonmotor s ymptoms in s ubjects w ithout e vidence of d opaminergic d eficits. *Movement Disorders*, 30(7):976–981, 2015.

[17] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer, 2018.

[18] Han Peng, Weikang Gong, Christian F Beckmann, Andrea Vedaldi, and Stephen M Smith. Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis*, 68:101871, 2021.