Random Multimodel Deep Learning for Classification

Project Report

Submitted By:

1. **Mayank Gupta(M20MA005)**
2. **Sanyam Jain(P20QC001)**
3. **Swatik Sahoo(M20CS063)**

Contents:

Introduction:

In the present era, the advancement of complex datasets for machine learning projects has made researchers think of robust techniques for accurate data classification. Various deep learning approaches have produced good results in text and image classification. However, finding a suitable structure for a combination of text and image data has been a challenge for researchers. The ensemble technique presented in this paper has solved the problem of finding the suitable structure and simultaneously improving accuracy and robustness.

RMDL consists of a combination of DNN(deep neural network), CNN(Convolutional neural network), and LSTM-GRU. It combines their results using the high voting methodology to provide a suitable architecture for the given combination of the dataset.

Architecture and Working:

As already mentioned it is an ensemble of DNN, CNN, LSTM-GRU, so it behaves like an ensemble model. The architecture is shown in the below figure. It takes a combination of text, image data, and this input is fed to all three networks simultaneously. The output of each individual model is represented by Y. Here ‘N’ RMDL models are created and each model consists of N/3 DNN, LSTM-GRU, CNN. The final output of the ‘N’ RMDL models is calculated using the majority of the votes.

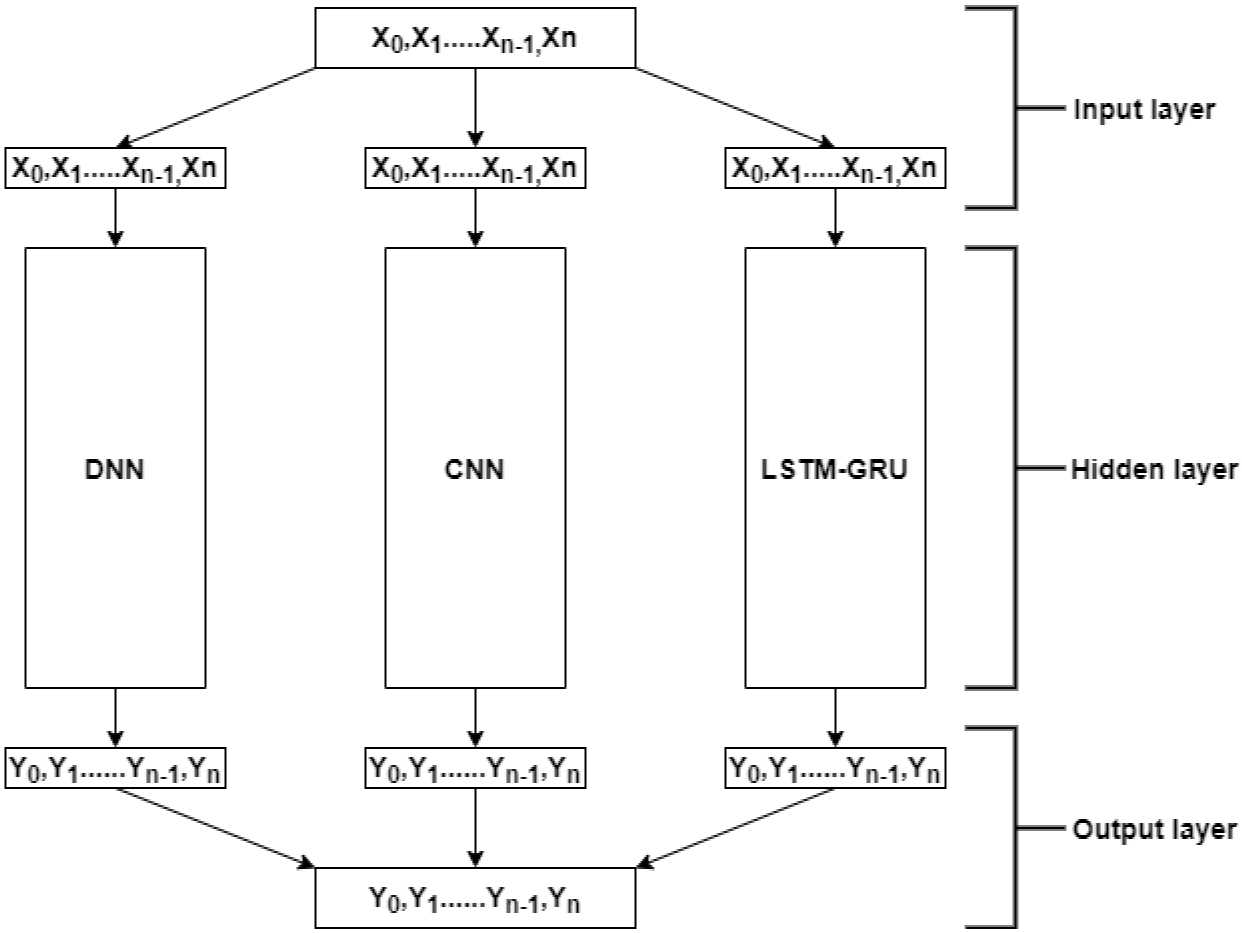

Fig 1: Architecture of RMDL

Fig 2 represents the combination of ‘N’ RMDL models and final output prediction using the high voting methodology. High voting means among all the RMDL models, the max value will be termed as the final prediction. Before finding the argmax, softmax is also applied. The optimization is being done using SGD, Adam, RMSprop, NAdam, and Adagrad. These optimizers are selected randomly by rand function and the model selects the best optimizer for the given input(dropout concept). Authors have used multiple optimizers as in multimodal there are chances that only one optimizer might not perform well. The RMDL model has the capability to use any kind of optimizer.

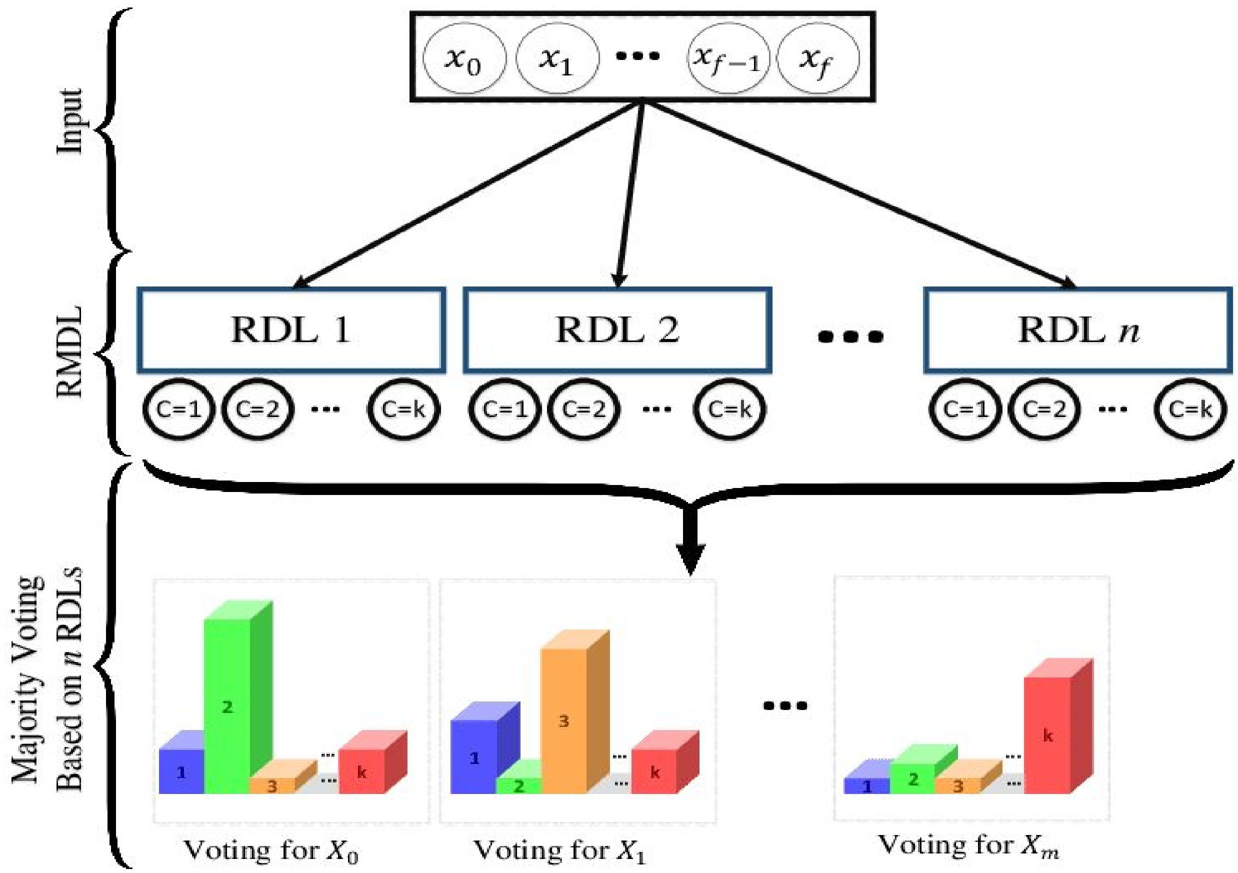

Fig 2: RMDL high voting methodology

Source: https://www.groundai.com/project/rmdl-random-multimodel-deep-learning-for-classification/1

Importance of DNN: As authors are trying to solve the problem of analyzing complex datasets and make a robust model that performs better on all kinds of datasets (which they envision), So authors have come up using multi-class DNNs where each learning model is chosen selectively. RMDL searches across randomly generated hyperparameters for the no. of hidden layers and nodes in layers.

Role of LSTM-GRU in image classification: The idea to develop a single model that can take image data as well as text data so that the ensemble model itself processes all the information itself and gives 3 results. The better results in the case of text will be RNN and for image data, it will be CNN. So with this intuition, and collecting the results from each of the sub-model, we can perform a poll and get the results. The paper mentions this as ”High Voting Methodology”.

Novelty:

This type of methodology brings the novelty under the course that it has an ensemble of multiple SOTA models which are then evaluated with voting criteria. The RMDL comprises RDL (Random Deep Learning) which is a depth of the participating model. A dataset is evaluated on 3 different models and chosen with the highest confidence. This reduces the future working for a particular dataset. When the analysis is performed on an unknown dataset, then this kind of approach works well where the dataset is a black box.

Results and analysis:

- **Datasets used in the paper:** IMDB and MNIST. Authors have run the model for 500 epochs with 3,9,15 combinations of each DNN, LSTM-GRU, and CNN but it will take many days to run for this number of epochs. To produce the results we have used 10 epochs for MNIST and 5 for the IMDB dataset with a combination of two from each of the architectures.

    1. **IMDB**: The model was trained and evaluated on the IMDB dataset. Five epochs run for a single combination. For two combinations ten epochs and for three architecture total of

thirty epochs will be there. The RMDL accuracy achieved for the ensemble model was around 87.26%.
Analysis: In the IMDB dataset, the task is to classify the sentiment as good or bad. So LSTM-GRU performs best as it combines both long term and short term memory. RMDL model also selects the best optimizer for LSTM-GRU architecture.

2. **MNIST**: In this dataset, digit classification has to be performed. The RMDL model gives a result on each DNN, LSTM-GRU, and CNN among all three best performances from the CNN model. The other two did not perform well as in image classification, the importance of long term or short term memory is less and moreover, RMDL selects the optimizer as per the input. So the best optimizer with the CNN model performs best.

Table 1: Showing results on both the dataset

	DNN	LSTM-GRU	CNN	Epochs	RMDL(2 combinations of each)	Dataset
Accuracy (%)*100	0.5685, 0.5902	0.5016, 0.8726	0.7986, 0.8296	5	0.8726	**IMDB**
Accuracy (%)*100	0.9041, 0.9174	0.989, 0.1135	0.9943, 0.9735	10	0.9943	**MNIST**

References:

- https://arxiv.org/pdf/1805.01890v2.pdf
- Kamran Kowsari, Mojtaba Heidarysafa, Donald E. Brown, Kiana Jafari Meimandi,Laura E. Barnes 2018. RMDL: Random Multimodel Deep Learning for Classification.
- Yoshua Bengio, Patrice Simard, and Paolo Frasconi. 1994. Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks 5, 2 (1994), 157–166.
- Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, and Yi Ma. 2015. PCANet: A simple deep learning baseline for image classification? IEEE Transactions on Image Processing 24, 12 (2015), 5017–5032.
- Ian J Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. 2013. Maxout networks. arXiv preprint arXiv:1302.4389 (2013).