Print ISSN: 0976-2876 Online ISSN: 2250-0138

Available online at: http://www.ijsr.in

INDIAN JOURNAL OF SCIENTIFIC RESEARCH

DOI:10.32606/IJSR.V12.I2.00015

Received: 13-08-2021 Accepted: 07-01-2022 Publication: 31-01-2022

Indian J.Sci.Res. 12 (2): 91-97, 2022 Original Research Article

ANALYSING MULTIMODEL ENSEMBLE COMBINATION AND MODEL ARCHITECTURES

VISHAL JHA[a1] AND UNNATI SADH[b]

[ab]Department of Computer Science, SRM IST, Modinagar, Uttar Pradesh, India

ABSTRACT

In this paper, we will be combining multiple different and similar model architectures namely CNN, ANN and DNN - for a comparative analysis over three different datasets while focusing on relative comparison more and individual model accuracy and generalization less.The initial parameters induced from initial testing and training will be carried forward to nexttesting on similar datasets to test for the ability of model to provide effective results on similarand different datasets with similar model architecture parameter constraints.

KEYWORDS: Multi-model Ensemble, Convolution Neural Network, Recurrent Neural Networks, Deep Neural Networks, MNIST, CIFAR-10, Malaria

Neural networks are data-driven, self-adaptive nonlinear methods that do not require specific assumptions about the underlying model. Instead of fitting the data with a prespecified model form, neural networks let the data itself serve as direct evidence to support the model's estimation of the underlying generation process. Ensemble learning is a general meta-approach to machine learning that seeks better predictive performance by combining the predictions from multiple models. The ensemble methods are more likely to make stable predictions and less likely to make catastrophic predictions than any single network used in isolation.In this paper, a combination of 3 types of models (Convolution neural network, Recurrent neural network and Deep neural network) were used with variation in their architecture and combinationin super ensemble to analyse on what can be the ideal combination, and can we carry forward thesame constraints of parameters on a different similar and different dataset to achieve a similar result. The objective will be to use each model's unique features to capture different patterns in the data.The method of prediction and selection will be similar to the method proposed by Zhang1 after analysis of his drawbacks with Keep-the-best method and seeking multiple alternative solutions and combining them in a systematic manner to make predictions so as to improve the generalization and prediction ability. The paper focuses on classification dataset as the main objective is to conduct a comparative analysison different model performances. And sticking to a single type of major problem set will help in quality comparative analysis.

MATERIALS AND METHODS

Datasets

In the presented paper, the basic structure of model is visualized and developed using the MNIST hand written digit dataset. The dataset has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image.The above dataset was selected due to easy pre-processing and lower memory utilization capabilities achieved with its use. Since the aim of the paper is to get a comparative analysis between different structures and combinations of neural networks in a super ensemble, using a simple and readily available dataset was analytically a better choice for initial developments and testing. It saved us a lotof time and were able to extract a lot of data through testing.In the next phase of testing on model with developed constraints identified from MNIST datasettesting, CIFAR-10 dataset was used which a multiclass labelled dataset. The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images. The classes are completely mutually exclusive. The above dataset was selected due to its similarity in class size with the previous dataset butdifference in colour set which moved the data from a black-and-white set to a coloured dataset, allowingthe model to be tested on a slightly more complicated dataset. The constraints for number of layersand nodes were carried from previous testing for comparative analysis of model performance on different datasets with similar

[1]Corresponding author

structure and combination. In further testing phase, a more complicated dataset of diagnosis of malaria from segmented cellsfrom a blood smear slide images were used from National Institute of Health, Maryland. The datasetcontains a total of 27,558 cell images with equal instances of parasitized and uninfected cells.The above dataset was selected to get a more real-world performance comparison for model andanalyse its performance on binary classification dataset.

Calculation of Error for Super Model Ensemble Combinations

The mean error of the numerical model is calculated using the root mean square error metric, measuring how far are the regression residuals from the data points.

$$\mathrm{RMSE}_{fo} = [\sum_{i=1}^{N} (z_{f_i} - z_{o_i})^2 / N]^{1/2}$$

The squared difference indicates the tendencies of the model to under estimate or overestimate the values of the given data points. RMSE in each combination was calculated for each individual model trained and evaluated on a different subset of sample or full sample. The with method of simple averaging the RMSE for the whole combination was calculated and logged. Error metric was crucial in analysing on how different combinations with different model structures we're performing on the given data sample. Additional loss metric was calculated for individual models on validation sets indicating the difference between predicted and actual values. It contributed in analysing the different possible model structures and selection process.

Multi-Model Ensemble Approach

The general method to create the multi-model ensemble used in this paper can be divided intomultiple steps,

1. Creating diverse individual models of different architectures and parameters.
2. Compare different individual models and combinations.
3. Training different selected models on boot strapped and non-boot strapped datasets.
4. Analysing the testing data and define a constraint/range for parameters for the multi-modelensemble.
5. Use the same ranged parameters to train similar models on different data for comparative analysis of model performance.

The above approach for defining a multi-model ensemble was inspired by the water resource search work by Chang Shu and Donald H. Burn.

RESULTS

In this section, the results obtained after performing experiments and evaluations are discussed.

Model Architecture and Super Ensemble Combinations

The basic model structures used in the paper are Convolution neural network, Recurrent neural network and Deep neural network. The mean square error (MSE) is used to gauge the performance ofthe network models. The models will not use perfectly optimized architectures and parameters as theuse case of the paper is to analyse relative performance of the compared ensemble approaches instead of their absolute performance.

CNN is a type of neural network model which allows us to extract higher representations for the image content. Unlike the classical image recognition where you define the image features yourself, CNN takes the image's raw pixel data, trains the model, then extracts the features automatically for better classification. Recurrent neural networks (RNN) are a class of neural networks that are helpful in modelling sequencedata. Derived from feedforward networks, RNNs exhibit similar behaviour to how human brains function. A neural network with some level of complexity, usually at least two layers, qualifies as a deep neuralnetwork (DNN), or deep net for short. Deep nets process data in complex ways by employing sophisticated math modelling. The super ensemble model was created by varying the number of nodes and number of layers for individual models and combining n number of distinct and different models. The combinations ranged from super ensemble of 3 models of 1 type each of CNN RNN and DNN to super ensemble 12 modelsof 4 models of each type.

Initially the models were built with the following constraints on their number of nodes and number of layers.

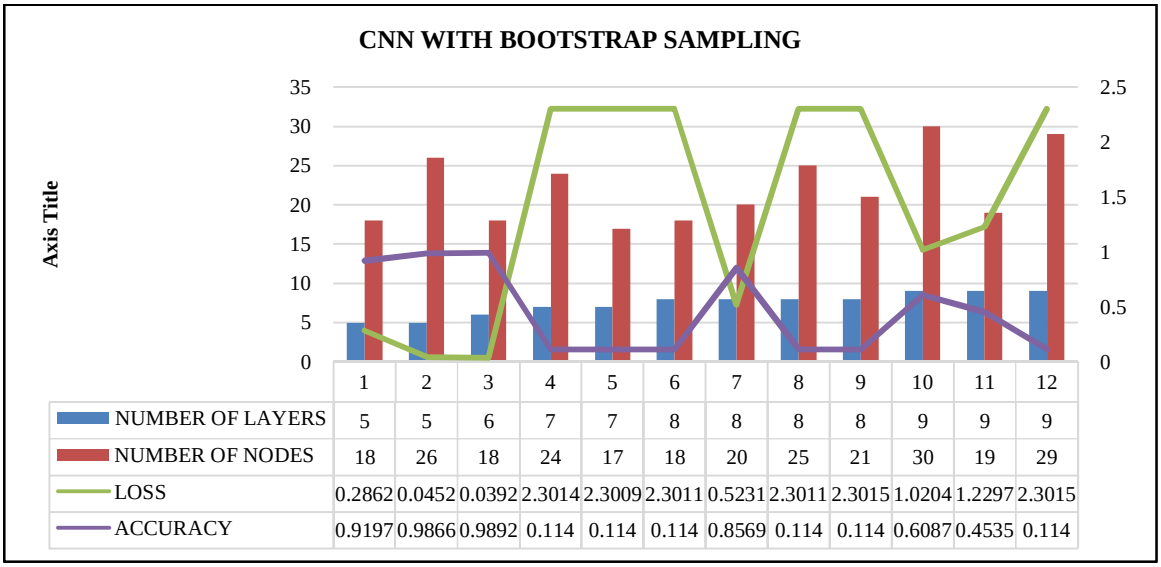
Table 1: Model configuration used during initial experimenting

CNN:	Number of layers	5 - 24
	Number of nodes	Number of nodes
RNN:	Number of layers	1
	Number of nodes	16 – 256
DNN:	Number of layers	1 - 8
	Number of nodes	64 - 256

A constraint on number of models was kept due to lack of processing power and time constraints, but maximum possible combinations were tested with given environment. All combinations were first trained on MNIST dataset extensively. After using simple averaging tocombine the predictions as is it easy to suffer overfitting which was extensively described by Zhou inhis work, and then analysing, a range for all parameters were selected to improves focus point forfurther evaluations. Then new parameter ranges after MNIST data training were completed,

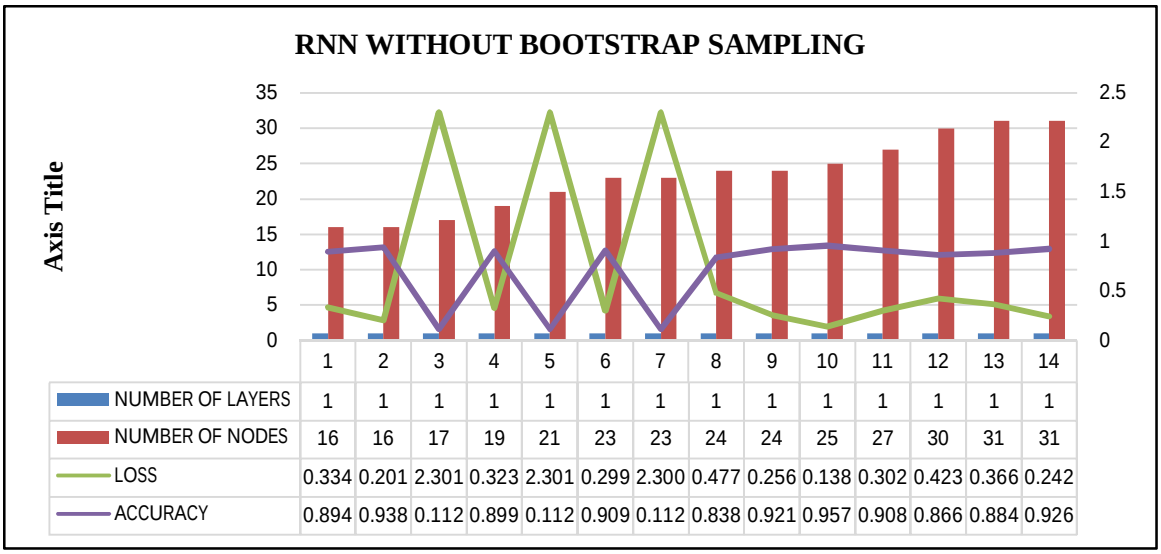
Table 2: Model configuration after initial experimenting

CNN:	Number of layers	8 - 12
	Number of nodes	15 - 26
RNN:	Number of layers	1
	Number of nodes	25 – 30
DNN:	Number of layers	9 – 11
	Number of nodes	72–125



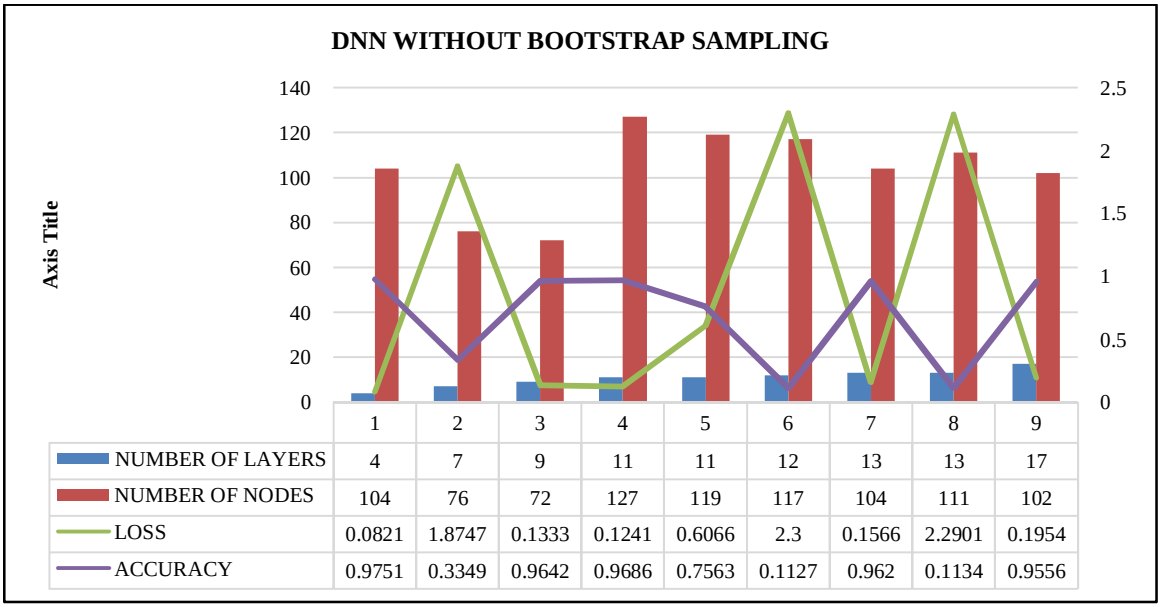
	1	2	3	4	5	6	7	8	9	10	11	12
NUMBER OF LAYERS	5	5	6	7	7	8	8	8	8	9	9	9
NUMBER OF NODES	18	26	18	24	17	18	20	25	21	30	19	29
LOSS	0.2862	0.0452	0.0392	2.3014	2.3009	2.3011	0.5231	2.3011	2.3015	1.0204	1.2297	2.3015
ACCURACY	0.9197	0.9866	0.9892	0.114	0.114	0.114	0.8569	0.114	0.114	0.6087	0.4535	0.114

Figure 1(a): Convolutional neural network models trained individually and as a part of ensemble on dataset without bootstrap sampling



	1	2	3	4	5	6	7	8	9	10	11	12	13	14
NUMBER OF LAYERS	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NUMBER OF NODES	16	16	17	19	21	23	23	24	24	25	27	30	31	31
LOSS	0.334	0.201	2.301	0.323	2.301	0.299	2.300	0.477	0.256	0.138	0.302	0.423	0.366	0.242
ACCURACY	0.894	0.938	0.112	0.899	0.112	0.909	0.112	0.838	0.921	0.957	0.908	0.866	0.884	0.926

Figure 1 (b): Recurrent neural network models trained individually and as a part of ensemble on dataset without bootstrap sampling



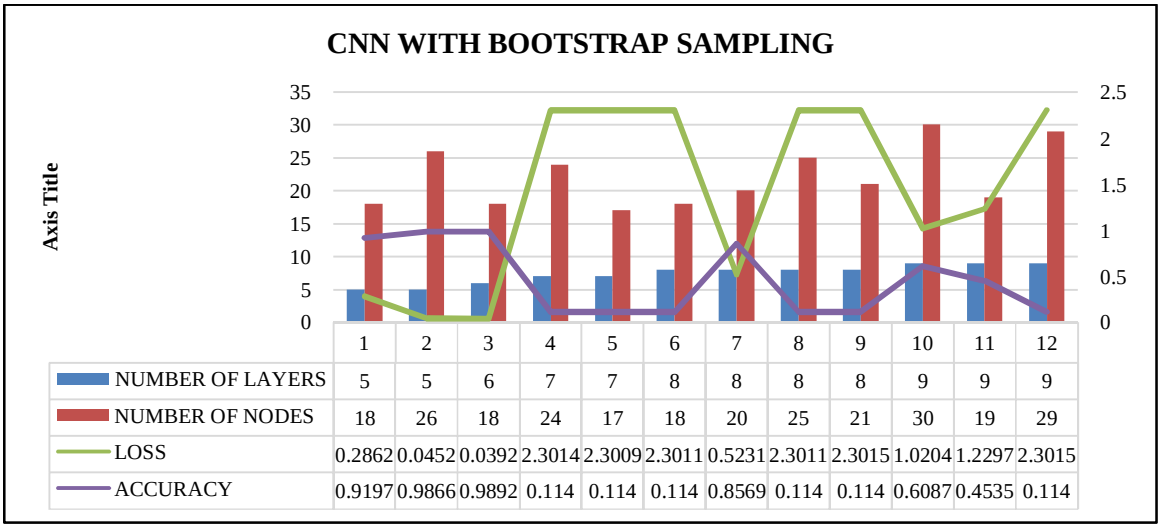
	1	2	3	4	5	6	7	8	9
NUMBER OF LAYERS	4	7	9	11	11	12	13	13	17
NUMBER OF NODES	104	76	72	127	119	117	104	111	102
LOSS	0.0821	1.8747	0.1333	0.1241	0.6066	2.3	0.1566	2.2901	0.1954
ACCURACY	0.9751	0.3349	0.9642	0.9686	0.7563	0.1127	0.962	0.1134	0.9556

Figure 1(c): Deep neural network models trained individually and as a part of ensemble on dataset without bootstrap sampling

With the above parameters, the model performed very well and was successful in achieving a generalized weight set. The best results were obtained when 3 models each of CNN RNN and DNN were combined to make the super ensemble, and achieved and average accuracy of 79.852% with slightly higher loss in data of 59.92% than the other models. A stable score was observed when the number of models were of equal types where, CNN model structure layers and nodes greater than 8 and 20 respectively, RNN model structure had layers and nodes greater than 1(fixed) and 24 respectively and DNN had layers in range of 9 to 11 and nodes greater than 72 to 125. Also, to check how many numbers of models and of which type in what ratio, all high performing model combinations of multiple variations tabulated and a comparative analysis was done to check for an ideal combination. From the set of models, an ideal combination of 9 models of 3 models of each type was selected as ideal due to its consistent low loss metric and higher average accuracy on the dataset. Zhou *et al.* (2002) had effectively mentioned in his work that ensembling many of the available neural networks may be better than ensembling all of those neural networks so as to constitute an ensemble according to some evolved weights that could characterize the fitness of including the networks in the ensemble. The same approach is being utilized here by analysing the effective fitness of different model combinations and eliminating parts of model to achieve a better performing ensemble.

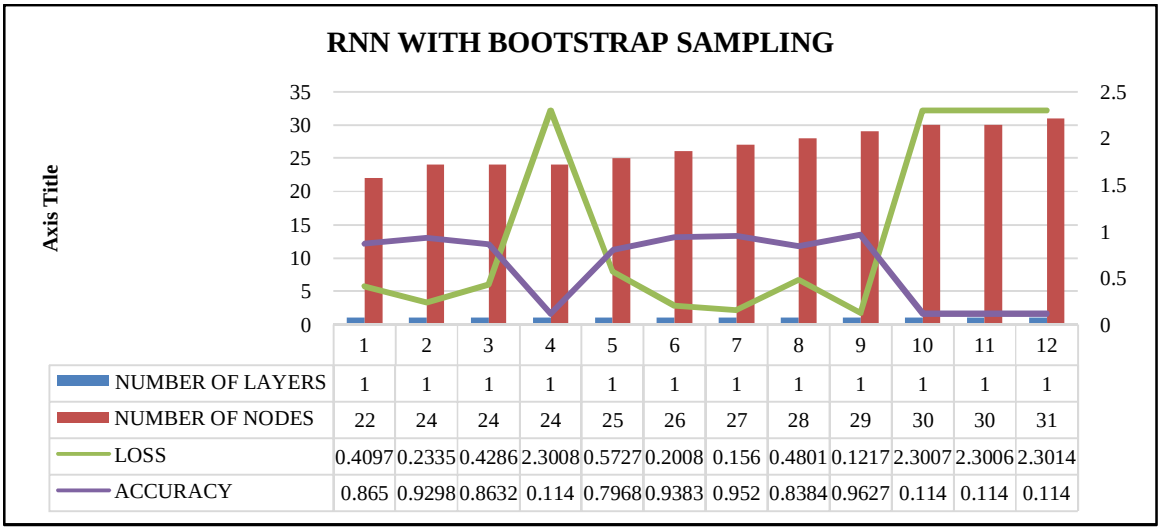
Also, recent researches in climate modelling methods used to combine forecasts suggests that combination schemes with unweighted means provide better results better than weighted combination methods in terms of model performance.

Same combinations were also trained on bootstrapped samples to check if better results could be achieved if the diversity of the models were increased by training them on different folds of data. But only marginal difference of not more than 3.36% improvement on an average was seen, which not a big improvement given that fairly simple structured data is used in all instances.



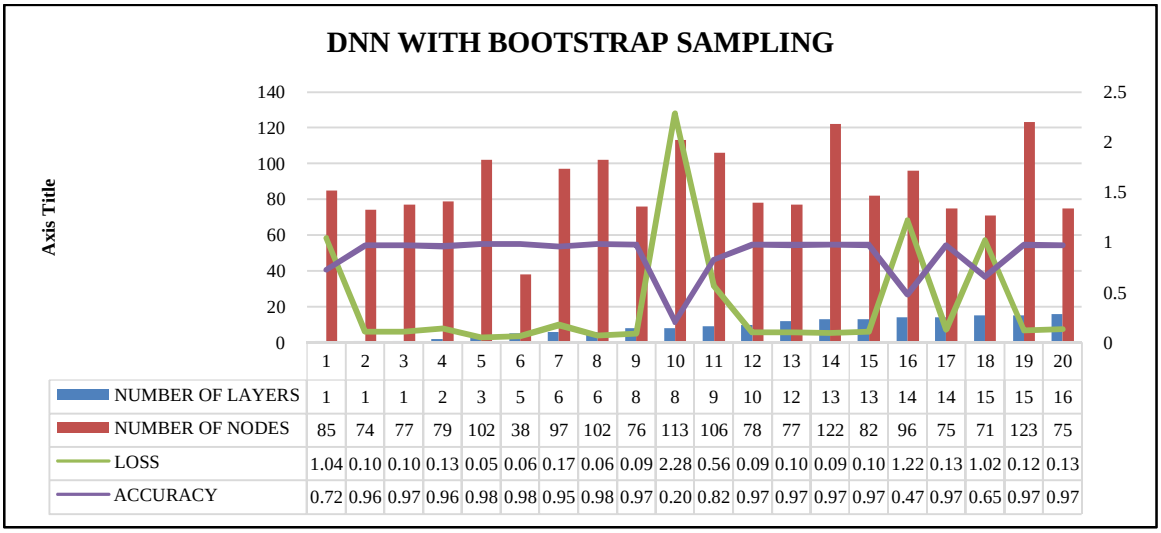
	1	2	3	4	5	6	7	8	9	10	11	12
NUMBER OF LAYERS	5	5	6	7	7	8	8	8	8	9	9	9
NUMBER OF NODES	18	26	18	24	17	18	20	25	21	30	19	29
LOSS	0.2862	0.0452	0.0392	2.3014	2.3009	2.3011	0.5231	2.3011	2.3015	1.0204	1.2297	2.3015
ACCURACY	0.9197	0.9866	0.9892	0.114	0.114	0.114	0.8569	0.114	0.114	0.6087	0.4535	0.114

Figure 2(a): Convolutional neural network models trained individually and as a part of ensemble on dataset with bootstrap sampling



	1	2	3	4	5	6	7	8	9	10	11	12
NUMBER OF LAYERS	1	1	1	1	1	1	1	1	1	1	1	1
NUMBER OF NODES	22	24	24	24	25	26	27	28	29	30	30	31
LOSS	0.4097	0.2335	0.4286	2.3008	0.5727	0.2008	0.156	0.4801	0.1217	2.3007	2.3006	2.3014
ACCURACY	0.865	0.9298	0.8632	0.114	0.7968	0.9383	0.952	0.8384	0.9627	0.114	0.114	0.114

Figure 2(b): Recurrent neural network models trained individually and as a part of ensemble on dataset with bootstrap sampling



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
NUMBER OF LAYERS	1	1	1	2	3	5	6	6	8	8	9	10	12	13	13	14	14	15	15	16
NUMBER OF NODES	85	74	77	79	102	38	97	102	76	113	106	78	77	122	82	96	75	71	123	75
LOSS	1.04	0.10	0.10	0.13	0.05	0.06	0.17	0.06	0.09	2.28	0.56	0.09	0.10	0.09	0.10	1.22	0.13	1.02	0.12	0.13
ACCURACY	0.72	0.96	0.97	0.96	0.98	0.98	0.95	0.98	0.97	0.20	0.82	0.97	0.97	0.97	0.97	0.47	0.97	0.65	0.97	0.97

Figure 2(c): Deep neural network models trained individually and as a part of ensemble on dataset with bootstrap sampling

Using Same Parameter Constraints on Dataset With Similar Features

Using the new ranged/constrained parameters, the super ensembles were evaluated on next to datasets of CIFAR-10 and Malaria dataset for comparative analysis. The new combinations and model structures are inherited from previous testing to check for compatibility of model and general ability to generalize.

When the same model parameters using the same constraints were used on CIFAR-10 dataset and malaria dataset, the model failed to converge on a higher accuracy combination within computational boundaries. On CIFAR-10 data, the model achieved an average accuracy of 32.613% with extremely high loss metrics of 1.8344. The results did not move towards better number after repeated trials of different model combinations and model architectures. This may be due CNN failing to provide better results in the super ensemble when compared to other two model types as it achieved the lowest average individual accuracy of 24.4%. On Malarial dataset, the model combinations achieved better average accuracy score of 60.4% on training data, but failed to generalize the model and also the loss metric was extremely high for DNN model at 5.240.

CONCLUSION

In this paper, a combination of 3 types of models (Convolution neural network, Recurrent neural network and Deep neural network) were used with variation in their architecture and combination in super ensemble to analyze on what can be the ideal combination, and can we carry forward the same constraints of parameters on a different similar and different dataset to achieve a similar result. All the predictions in all cases of different were combined via simple averaging method. After comparative analysis and extensive testing, we came up with certain model architecture parameter constraints for model building and an ideal combination for the 3 models to form a super ensemble. The super ensemble with above mentioned parameters and ideal combination gave good accuracy on the MNIST dataset, but transferring the same parameters to train a model on the CIFAR-10 dataset and Malaria dataset did not result in a good accuracy and loss metric. Hence, we cannot use same constraints for a super ensemble for a different dataset, even if the dataset present similar feature set.

REFERENCES

Canziani A., Paszke A. and Culurciello E., 2017. An Analysis of Deep Neural Network Models for Practical Applications.

Chai T. and Draxler R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? January 2014. Geo Scientific Model Development Discussions, **7**: 1525-34.

Krizhevsky A., 2009. Learning Multiple Layers of Features from Tiny Images.

Kukreja H., Bharath N., Siddesh C.S. and Kuldeep S., 2016. An introduction to artificial neural network. I.J.A.R.I.I.E., **1**(5): 27-30.

Krishnamurti T.N., Kishtawal C.M., Zhang Z., LaRow T., Bachiochi D., Williford E., Gadgil S. and Surendran S., 2000. Multimodel ensemble forecasts for weather and seasonal climate. Journal of Climate, **13**(23): 4196-4216.

LeCun Y., Bottou L., Bengio Y. And Haffner P., 1998. "Gradient-based learning applied to document recognition." Proceedings of the IEEE, **86**(11): 2278-2324.

Poon H., Christensen J., Domingos P., Etzioni O., Hoffmann R., Kiddon C., Lin T., Ling X., Mausam, Ritter A., Schoenmackers S., Soderland S., Weld D., Wu F. and Zhang C., 2010. Machine Reading at UW. NAACL HLT, pp. 87-95.

Rajaraman S., Antani S.K., Poostchi M., Silamut K., Hossain M.A., Maude R.J., Jaeger S. and Thoma G.R., 2018. Pre-trained convolutional neural networks as feature extractors toward improved Malaria parasite detection in thin blood smear images.

Rozante J.R and Moreira D.S, 2014. R.C.M Godoy, A.A Fernandes, Multimodel ensemble: technique and validation. Geoscientific model development 7/2333.

Sherstinsky A., 2020. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network Elsevier "Physica D: Nonlinear Phenomena" Journal, Volume **404**: Special Issue on Machine Learning and Dynamical Systems.

Shu C. and Burn D.H., 2004. Artificial neural network ensembles and their application in pooled flood frequency analysis. Water resources research, Vol **40**.

Yamashita R., Nishio M., Do R.K.G. and Togashi K., 2018. Convolutional neural networks: an overview and application in radiology. Insights Imaging, **9**: 611-629.

Zhang G.P. and Berandi V.L., 2001. Time series forecasting with neural network ensembles: an application for exchange rate prediction. Journal of the Operational Research Society, **52**: 652-664.

Zhou Z-H., Wu J. and Tang W., 2002. Ensembling neural networks: Many could be better than all. Artificial Intelligence, **137**: 239-263.