2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI) | 978-1-6654-6603-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/IRI54793.2022.00027

Multilayer Meta-Learning Approach to Forecasting Air Pollutants

Samuel A. Ajila, *Senior Member IEEE*
Department of Sytems and Computer Engineering
Carleton University, Ottawa, Ontario, Canada
ajila@sce.carleton.ca

***Abstract*— Air pollution forecasting is an important factor in the analysis of air quality and it can be used to achieve an increase in the air quality index. The process is likely to contribute to environmental and human health improvement. In this research paper, we apply multilayer meta-learning forecasting using Beijing air quality datasets. The dataset consists among others six air pollutant attributes - PM2.5, PM10, SO_2, NO_2, CO and O_3 that are considered important factors in calculating the Air Quality Index. Major air pollutants, especially fine particles such as PM2.5 (particulate matter with diameter less than 2.5 µm), are generally associated with adverse health effects that includes cardiac and respiratory morbidity. The resulting performance accuracy (MAE and RMSE values) show a huge improvement over our previous results that uses a compositional learning model. Additionally, the results obtained are considerably better than using "single learners" – Random Forest and Random Committee.**

Keywords*— Meta-learning, Air Quality Index, Bagging, Boosting, Random Forest, Random Committee

I. Introduction

Air pollution is a serious issue around the world because of the threat it poses to human health and the destruction of the environment. Air pollution is particularly noticeable in some parts of the world. For instance, in 2016, data collection shows that only 84 of the 338 prefecture-level or high-density cities in China attained the national standard for air quality [1]. Air pollution is a serious environmental issue because of the pollutants such as fine particles with diameter less than 2.5 µm (PM2.5) that are associated with adverse health effects which include respiratory and cardiac problems. Independent research in environmental impacts shows that air pollution caused more than a million deaths per year [2]. In addition, high level of gas compounds such as NO_2, O_3, and SO_2 is very dangerous for the environment as well as human beings. If people with comorbidity are exposed to these toxic gases can lead to very serious health problems or even death.

This paper focuses on the design and use of a novel Multilayer meta-learning Approach (MMA) for forecasting air pollution. The goal is to compare the performance accuracy of MMA with previous work which was based on Compositional Learning method [4] in terms of MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) values. The main proposition in this paper is that *using the Beijing Air-Quality Dataset [1], [2], [4], can we improve the performance accuracy of the prediction and forecast of the future trend of air pollutants using a Multilayer Meta-learning Approach (MMA)*? The long-term objective in this paper is to be able to forecast the different components of the air pollutants at once and for a foreseeable future. The results of the forecasts can then be used to mitigate air pollution and thereby improve the environment as well as human health.

The remainder of the paper is organized as follows. Section II presents the background, previous work, and methodology. Section III discusses the dataset analysis and pre-processing while section IV presents the Multilayer Meta-learning Approach. Section V discusses the experiments and results. Conclusions are given in section VI.

II. Background and Privious Work

A. Linear and Additive Regressions

Linear regression (LR) models the relationship between dependent variable and independent variables. [6], [7]. A simple LR is a situation when there is only one independent variable and multiple LR if more than one independent variables. Multivariable LR is where multiple correlated dependent variables are predicted compared to a single variable. *LR* can be used easily for numeric attributes. It can also be used for any classification as long as regression is performed for each class by setting to "1" for those instances that belong to the class and "0" for those that do not. The problem with LR is that the membership values produced are not proper probabilities, because they can fall outside of the range of 0 and 1. However, a related technique called *Logistic Regression* does not have these problems because it builds a linear model based on the transformed target variable instead of approximating the values 0 and 1 directly [6], [7], [8]. In *Logistic Regression*, an attempt is made to produce accurate probability estimates that lead to accurate classifications. *Additive method* (AM) can be considered a form of boosting. AM is simply a method of generating predictions by adding up contributions obtained from other learning models. It forms an ensemble of base models that optimizes predictive performance based on a specified criterion.

B. Support Vector Machine

Support Vector Machine (SVM) is used for many machine learning tasks such as pattern recognition, object classification, and regression analysis in the case of the time series prediction. Support Vector Regression (SVR) is the methodology by which a function is estimated by using the observed data. In this paper the SVR and the SVM terms are used interchangeably. SVM

978-1-6654-6603-5/22/$31.00 ©2022 IEEE
DOI 10.1109/IRI54793.2022.00027

uses the equations $f(x) = (w \cdot x) + b$ and $f(x) = (w \cdot \varepsilon(x)) + b$ to define the prediction functions for the linear and the non-linear regression models, respectively. w is a set of weights, b is a threshold, and ε is a kernel function. If the time-series is not linear, the regression model maps the time-series x to a higher dimension feature space by using kernel function $\varepsilon(x)$. Then the prediction model performs the linear regression in the higher dimensional feature space. The goal of the SVM training is to find the optimal weights w and the optimal threshold b.

C. Random Forest and Random Committee

Random Forest (RF) algorithm is a supervised machine learning algorithm, that uses a group of decision trees for prediction. A decision tree [7], [8], [9] is a map of the possible outcomes of a series of related choices. However, decision trees can become excessively complex for large number of features. Thus, Random Forest mitigates the disadvantages of decision tree. The Random Committee (RC) algorithm uses the decision tree for prediction similar to Random Forest [8], [9], [10]. However, Random committee constructs a number of base classifiers using unique random number seed values and the final classification result is provided by computing the average of the predictions generated by the individual base classifiers [11], [12].

D. Performance Measure Metrics

In this paper, we have used two common performance measures – mean absolute error and root mean square error to evaluate the accuracy of the different algorithms.

Given n set of predictions $y_1, \ldots, y_n$, made by a model M, we can define the following [4], [6]:

Mean Absolute Error $(MAE) = \frac{\sum_{j=1}^{n} |y_j - M(d_j)|}{n}$

Root Mean Square Error $(RMSE) = \sqrt{\frac{\sum_{j=1}^{n} (y_j - M(d_j))^2}{n}}$

Where $M(d_1), \ldots, M(d_n)$ is a set of n predictions for a set of test instances $d_1, \ldots, d_n$.

The lower the value of MAE the better the model as the error produced by the model will be low. Unlike MAE, RMSE accounts for both directions (positive or negative) and it provides the estimate of how large the errors are being dispersed. Since the square value is taken between the errors, RMSE tends to be higher than MAE when the sample size and the error increases.

E. Air Quality Index

Air Quality Index (AQI) [13], [14] is a metric used to communicate how polluted the current air is or forecast to become in the nearest future. In this paper we use the Beijing Air Quality dataset and AQI value is calculated per 24-hour period and as such P2.5, P10, SO_2, NO_2, and CO concentrations are measured as average per 24-hour period. However, O_3 concentration is measured as the maximum 24-hour moving average. AQI value calculated can be classified into different pollution levels [14], [15]. Pollution levels of concerns and its descriptions are given in Table I.

TABLE I. AQI POLLUTION LEVELS [4]

Levels of Concern	Values of AQI	Description
Good	0 to 50	Air quality is satisfactory, and air pollution poses little or no risk.
Moderate	51 to 100	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is less likely to be affected.
Unhealthy	151 to 200	Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy	201 to 300	Health alert: The risk of health effects is increased for everyone.
Hazardous	301 and higher	Health warning of emergency conditions: everyone is more likely to be affected.

F. Previous Works using the dataset

Beijing Air Quality Dataset has been used in many researches among which are [1, 2, 4, 15, 16, 17]. Our aim in this section is to discuss the commonality between these research works (except [4]). The commonality is that the previous research works focused exclusively on one air particle concentration – in particular PM2.5 particle. Other particle concentrations (SO_2, NO_2, CO, O_3, and PM10) captured by the dataset are sometimes not analysed or simply ignored. This is understandable because – firstly PM2.5 is a fine grain particle and the notion is that it carries majority of the pollutants but there is no proof of this fact. One can even argue that the gaseous particles (e.g. NO2 and SO2) are deadlier! Secondly, it is relatively easy to model a single attribute in timeseries forecasting over time. The accuracy will be better and the tendency is that the different machine learning algorithms can be trained effectively based on single attribute. However, it is difficult to deal with models for forecasting multi-attributes. Our approach starting with our previous work on the dataset [4] is to look at the entire dataset features and in particular, all the six pollutants (SO_2, NO_2, CO, O_3, and P2.5, PM10). To the best of our knowledge, our research work (previous work [4] and this paper) will be the first time that a forecasting approach will address all the six pollutants in the dataset at once using novel methods. Additionally, we started by doing a thorough analysis and pre-processing of the dataset.

G. Compositional Leaning Method for air pollution forecasting

This paper is sequel to our previous paper [4]. This section summarizes our previous work on the dataset so that the current results can be compared with the previous ones. The previous work [4] applied a modified compositional learning model with disentanglement using optimized hyper parameters to forecast air pollution. The results show a marked improvement in the range of 3.34% to 78% in terms of MAE and RMSE compared to two “single learner” (Random Forest RF and Random

Committee RC) models. The summary of the results is given in Table II.

TABLE II. COMPOSITIONAL LEARNING MODEL [4]

Features	MAE			RMSE		
	RF	RC	Compositional Learning	RF	RC	Compositional Learning (CL)
PM2.5	55.405	58.651	39.2378	74.539	78.134	53.9483
PM10	58.028	58.667	46.2151	79.531	82.240	60.6940
SO_2	11.043	11.081	08.8364	14.124	14.468	13.6525
NO_2	30.135	27.489	16.2288	34.306	33.276	21.3743
CO	1240.897	1147.630	252.4395	1428.121	1403.337	334.1110
O_3	72.740	66.895	29.4006	95.493	88.635	39.3750

H. Methodology

Our methodology is to first analyse the dataset for missing data using MICE (Multivariate Imputation by Chained Equations) imputation method [4]. Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations [19]. Secondly, the dataset is further analysed to understand the data distribution using probability distribution and scatter plot matrix. Thirdly, we develop a novel Multilayer Meta learning Approach (MMA), train three meta learners, test, and use the trained meta learners for forecasting. Both MAE and RMSE are used as performance metrics.

III. DATASET PREPROCESSING

A. Beijing Air Quality Dataset

Beijing Air Quality Dataset [13], [14], [15] is divided into four regions Aotizhongxin, Changping, Dingling and Dongsi. The time period is from March 1st, 2013 to February 28th, 2017. The attributes of the dataset are given in Table III.

TABLE III. DATASET CHARACTERISTICS [4]

Attributes	Description
No	Row Number
year	Year of the data
month	Month of the data
day	Day of the data
hour	Hour of the data
PM2.5	PM2.5 Concentration (ug/m^3)
PM10	PM10 Concentration (ug/m^3)
SO_2	SO_2 Concentration (ug/m^3)
NO_2	NO_2 Concentration (ug/m^3)
CO	CO Concentration (ug/m^3)
O_3	O_3 Concentration (ug/m^3)
TEMP	Temperature (degree Celsius)
PRES	Pressure (Pa)
DEWP	Dew point Temperature (degree Celsius)
RAIN	Precipitation (mm)
wd	Wind direction
WSPM	Wind speed (m/s)
station	Name of the air-quality monitoring site

B. Data Exploration and Understanding

In order to explore the dataset and its quality, histograms of the continuous features are generated (see Fig. 1), and the mean and standard deviation of each feature are used to get a sense of the central tendency and variation within the dataset. The histograms of the continuous features give an easy way to understand how for each feature the values are distributed across range. These shapes (i.e. Fig. 1) relate very well with the standard probability distributions. It shows four unimodal distributions (PM10, NO_2, CO and O_3), and two exponential-like distributions (PM2.5 and SO_2). Unimodal distribution is a type of normal distribution with single peak around the central tendency. A normal distribution is preferred because many of the machine learning techniques will work very well with such dataset.

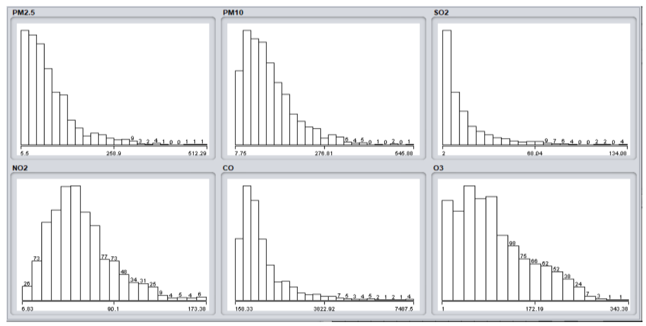
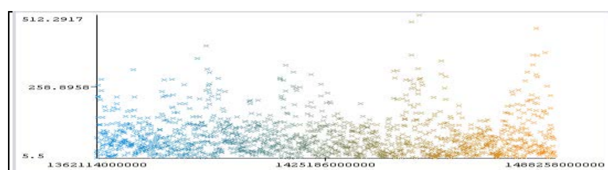
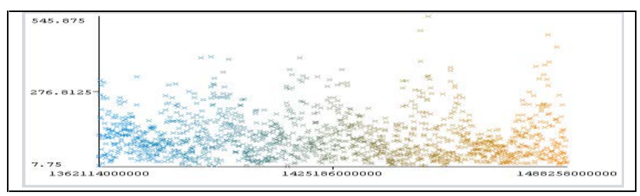


Fig. 1 Probability distributions of the Pollutants

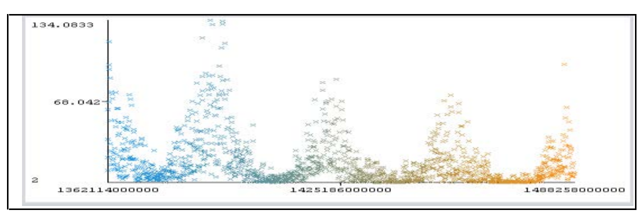
However, unimodal distributions (PM10 and CO) skewed to the right have tendencies toward very high values. Since this stage of data exploration is for information gathering and understanding the distribution of the dataset, it is clear from Fig. 1 that the dataset feature distributions are mostly "normal" unimodal skewed to the right. This means that over time, you have peak and low points in the dataset features. In order to confirm this assertion, we carried out further analysis using scatter plot analysis (see Fig. 2 a, b, c, d, e, and f).



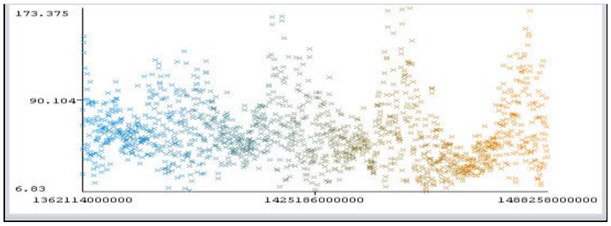
(a) – PM2.5: Peaks & Lows vs Time →



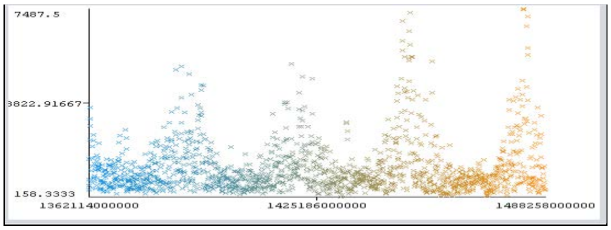
(b) – PM10: Peaks & Lows vs Time →



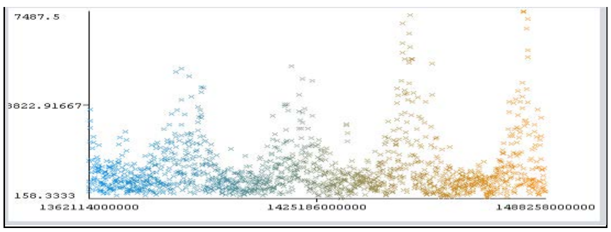
(c) – SO2: Peaks & Lows vs Time →



(d) NO2: Peaks & Lows vs Time →



(e) – CO: Peaks & Lows vs Time →



(f) – O3: Peaks & Lows vs Time →

Fig. 2 (a, b, c, d, e, and f) - Air Pollutants Scatter Plot Matrix

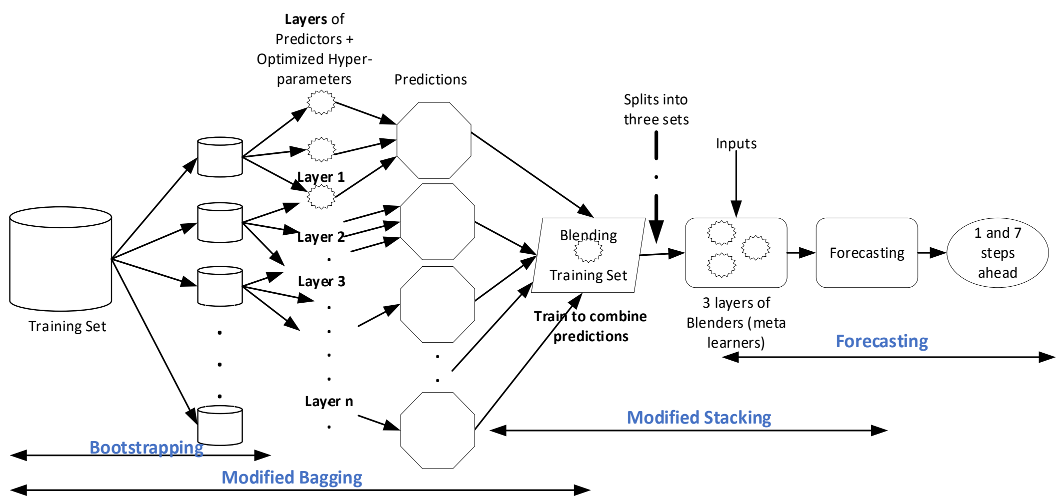


Fig. 3 – Multilayer meta-learning approach to Air pollution forecasting

Fig. 2 confirms our assertion that there are peaks and lows in the distributions of the dataset features over time. A further analysis (manually) of the dataset shows that high pollution concentration values are mostly obtained in the Winter months of Beijing in China which is December to February and minimum values in the Summer months from June to August. For example, the minimum and maximum PM2.5 concentration values are approximately around 10 ug/m^3 and 500 ug/m^3 in the 5 years period. This means that there are more pollutants in Winter compared to Summer season. These information (i.e. the scatter plots) will be used when interpreting the prediction and forecasting results in section V.

IV. Multilayer Meta-Learning Approach (MMA)

Our goal in this paper is to design a novel Multilayer Meta-learning Approach (MMA) for forecasting air pollution using all the pollutant features in the dataset. The proposition here is that the performance accuracy of this approach will be better compared to our former [4] work using a modified compositional model. Fig. 3 gives a graphical view of our approach. The steps are as follows.

1. Partition the dataset (already imputed using MICE) into training and testing sets based on 60%:40% splits.
2. Bootstrap the training set – this is a random sampling with replacement. This way, 10 random samples are created (i.e. n = 10, see Fig. 3).
3. From step 2, each of the random sample is fed into a layer of three predictors – Linear Regression (LR), Support Vector Machine (SVM), M5 rule-based algorithm (M5P). The selection of these learning algorithms is based on the dataset analysis results and literature review. Note that each predictor will have a higher bias than if it were trained on the original training set. However, the aggregation will reduce both the bias and the variance. In this case, we have ten layers of "bootstrapping + aggregation" which we called "modified bagging." In general, bagging will allow training instances to be sampled many times for the same predictor. The most important thing at this point is that, the net results of these ten layers ensembles will have a lower variance than a single predictor on the original training set.
4. The predictions from step 3 are blended (i.e. combined) as the "training set" for the next stage. The best predictor out of the three algorithms from the step 3 is chosen for the blending (combining the predictions). This is done to improve the aggregation instead of using a trivial function for the blending. The blending stage signifies the end of the modified bagging and the beginning of the modified stacking process.
5. The result of the blending is splits into three sets based on the three best predictions and these predictors are used as base learners for the three layers of blenders. Note that the output of the previous stage is internal and the base learners

could be just one of the three algorithms or a combination of the three. In addition to the base learners, we choose three meta-learners: Additive Regression (AR), Random Forest (RF), and Random Committee (RC) and used 20% hold-out from the testing dataset to train the meta-learners. The remaining 20% is then used to test the forecaster using 1-step and 7-step ahead. The results of the experiment are discussed next.

V. EXPERIMENTS AND RESULTS

This section presents the effectiveness of Multilayer Meta-learner Approach (MMA) and compares the results against the results of modified compositional learning approach [4] (Table II). The experiments are setup using hardware environment consisting of *Intel Core i7-10700F processors, 16 CPUs, 64 cores, GeForce RTX 3060, and CentOS Linux 8 with 16G RAM, and 2 Terabytes storage*. Although we did not use Hadoop [22], the notion of MapReduce [22] was used to setup ten parallel “maps” for the ten layers of our modified bagging; three “reducers” for the three meta-learners; and one super node each for the combiner and forecasting. Each map, reducer, and super node is a CPU (4 cores). The GPU is used to coordinate the processing.

A. Results

The first experiment is the result (Fig. 4) of using the 20% test data of the meta learners. Fig. 4 depicts a distribution that confirms the lows during the Summer season (June to September) and the peak during Winter (December to February).

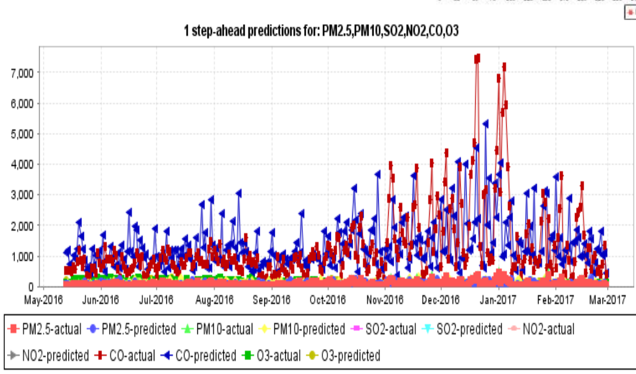


Fig. 4 – 1-step ahead test data predictions

The second set of experiments used the entire dataset to test the Multilayer Meta-learning Approach (MMA). Fig. 5 presents the result obtained and it depicts a distribution that closely resemble the distributions in Fig. 2 a, b, c, d, e, and f. Fig. 5 shows the same peaks and lows that characterise the dataset patterns for Winter when the pollution is very high and Summer when it is relatively low. The performance metrics will be discussed later in section V(B).

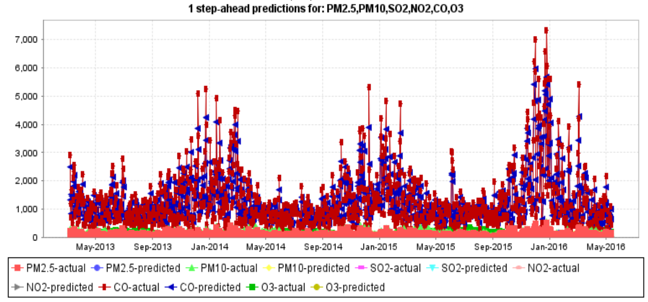


Fig. 5 – 1-step ahead predictions for the entire dataset

Next, we used the meta-learners to perform 1,186 1-step ahead forecasts (see Fig. 6). In addition, as a result of the dense distribution of Fig 6, the visualization for one pollutant CO is zoomed-out so that the distribution pattern can be clearly seen to be similar to both Figs. 2 and 4 which shows the peaks and lows of the forecast corresponding to relatively low pollution in Summer and high in Winter.

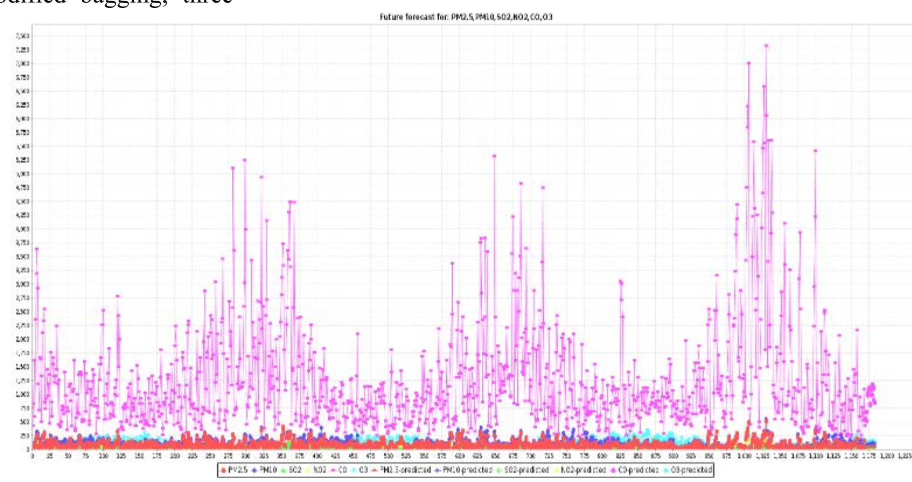


Fig. 6 - 1,186 1-step ahead future forecasts

B. Performance Metrics

MAE and RMSE are the performance metrics used in this paper. The metric values for the Multilayer Meta-learning Approach (MMA) is put side by side with that of the modified compositional learning model (CL), and single learners – Random Forest and Random Committee in Table IV. The values of MAE and RMSE for MMA are huge improvements compared to the values for CL. The metric values for Random Forest and Random Committee did not even come close to that of MMA. It is important to note the performance accuracy of three pollutants (SO_2, CO, and O_3) in terms of MAE (0.0124, 0.0002, and 0.2228) and RMSE (0.2104, 0.0003, and 0.4367) values. In order to make sure these values are not just chance, we examined how the meta learners handle the case of the attribute CO. In this case, MMA best meta learner for CO consists of a combination of *Meta Additive Regression* and *Random Forest* with *Greedy* search algorithm for attributes selection, *Dimensionality* reduction, and the estimated error rate is 2.6740. Note that in our MMA, we incorporated hyperparameters turning using AutoML based on Sequential Model-based Algorithm Configuration [23], [24], [25]. We then conducted ten simulations using the configuration produced by the MMA meta-learner for CO. The average estimated error came to 2.6862 which is sufficiently close enough to the estimated error rate (2.6740) obtained by MMA.

TABLE IV – COMPARING THE MULTILAYER META-LEARNING APPROACH MAE AND RMSE VALUES TO THAT OF COMPOSITIONAL LEARNING MODEL AND, RF & RC

Features	MAE				RMSE			
	RF	RC	Compositional Learning (CL)	**MMA**	RF	RC	CL	**MMA**
PM2.5	55.405	58.651	39.2378	**15.9901**	74.539	78.134	53.9483	**22.3228**
PM10	58.028	58.667	46.2151	**18.5086**	79.531	82.240	60.6940	**24.4219**
SO_2	11.043	11.081	08.8364	**0.0124**	14.124	14.468	13.6525	**0.2104**
NO_2	30.135	27.489	16.2288	**12.6741**	34.306	33.276	21.3743	**18.5930**
CO	1240.897	1147.630	252.4395	**0.0002**	1428.121	1403.337	334.1110	**0.0003**
O_3	72.740	66.895	29.4006	**0.2228**	95.493	88.635	39.3750	**0.4367**

VI. CONCLUSIONS

This paper presented a novel Multilayer Meta-learning Approach (MMA) for forecasting air pollutants in a multi-feature domain. It is based on modified bagging and stacking models. The MAE and RMSE values produced by MMA are compared with the values produced by our previous work [4] – modified Compositional Learning Model (CL) and with two single learners - Random Forest and Random Committee. The results of performance accuracy for MMA are superior to that of CL and the two single learners (Table IV).

Note that we also forecasted 7-step ahead and the results are essentially similar to that of 1-step ahead. We did not discuss it in this paper because of space.

In future work, the author will like to research the application of deep machine learning approach for data pre-processing, prediction, and forecasting. Furthermore, a different set of meta learning approaches will be considered.

REFERENCES

[1] Liang X, Zou T, Guo B, Li S, Zhang H, Zhang S, Huang H, Chen SX. 2015 Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating. Proc. R. Soc. A 471, 20150257. (doi:10.1098/rspa.2015.0257)

[2] Shuyi Zhang, Bin Guo, Anlan Dong, Jing He, Ziping Xu, Song Xi Chen, Cautionary tales on air-quality improvement in Beijing, Proc. R. Soc. A 473: 20170457, http://dx.doi.org/10.1098/rspa.2017.0457, April 2017

[3] M. Bentayeb, M. Simoni, D. Norback, S. Baldacci, S. Maio, G. Viegi, I. Annesi-Maesano, Indoor air pollution and respiratory health in the elderly, Journal of Environmental Science and Health, Part A 48 (14) (2013), 1783–1789.

[4] Samuel Ajila and Karthik Dilliraj, *Forecasting Air Pollution using a Modified Compositional Learning Approach*, 2021 IEEE International Conference on Big Data (IEEE BigData 2021), International Workshop on Big Data Analytics for Sustainability (BDA4S 2021), Dec. 15 – 18, 2021, pp. 3992 – 4001

[5] Beijing Multi-Site Air-Quality Data Set. (2019, Sep 20) Retrieved from https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data# , Accessed March 2022

[6] A. Das, S. A. Ajila, C.-H. Lung, A Comprehensive Analysis of Accuracies of Machine Learning Algorithms for Network Intrusion Detection, IFIP International Federation for Information Processing 2020, Published by Springer Nature Switzerland AG 2020, S. Boumerdassi et al. (Eds.): MLN 2019, LNCS 12081, pp. 40–57, 2020.

[7] J. D. Kelleher, B. Mac Namee, and A. D'Arcy, Fundamentals of Machine Learning for Predictive Data Analytics, The MIT Press, Cambridge, Massachusetts, USA, 2015, ISBN 978-0-262-02944-5

[8] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining – Practical Machine Learning Tools and Techniques, 4th ed., Morgan Kaufmann, Elsevier, New York, USA, 2017, ISBN 978-0-12-804291-5

[9] L. Breiman, Bagging predictors, Machine Learning 24 (2004) 123–140.

[10] L. Breiman, Stacked regressions, Machine Learning 24 (2004) 49–64.

[11] Paul Cohen, Empirical Methods for Artificial Intelligence, MIT Press, ISBN 0-262-03225-2, 1995.

[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, The MIT Press, Cambridge, Massachusetts, London, England, 2016, ISBN 9780262035613

[13] P. R. Bhardwaj, D. Pruthi, Evolutionary techniques for optimizing air quality model, Procedia Computer Science 167 (2020) 1872–1879.

[14] AQI Basics. Retrieved from https://www.airnow.gov/aqi/aqi-basics/, Accessed April, 2020

[15] Mingye Yang., A Machine Learning Approach to Evaluate Beijing Air Quality. (2018, June) Retrieved from https://www.math.ucdavis.edu/files/2015/2717/8083/Mingy_Yang_Spring_2018.pdf, Accessed March, 2020

[16] Shengdong Du, Tianrui Li, Yan Yang, and Shi-Jinn Horng, Deep Air Quality Forecasting Using Hybrid Deep Learning Framework, IEEE Transactions on Knowledge and Data Engineering, 2021, Vol. 33:6

[17] Hong Zheng, Haibin Li, Xingjian Lu, and Tong Ruan. A Multiple Kernel Learning Approach for Air Quality Prediction, Hindawi Advances in Meteorology, Volume 2018, Article ID 3506394

[18] James, G. (2015). An introduction to statistical learning: With applications in R.(pp. 303-335) New York: Springer.

[19] Cord, M. (2008)., Machine learning techniques for multimedia: Case studies on organization and retrieval ; 20 tables (pp. 21-49). Berlin: Springer.

[20] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[21] Lars Kotthoff, Chris Thornton, Holger H. Hoos, Frank Hutter, and Kevin Leyton-Brown, Auto-WEKA: Automatic Model Selection and Hyperparameter Optimization in WEKA, 2019 F. Hutter et al. (eds.), *Automated Machine Learning*, The Springer Series on Challenges in Machine Learning, https://doi.org/10.1007/978-3-030-05318-5_4, pp 81-95

[22] Silvano Gai, Building a Future-Proof Cloud Infrastructure – A Unified Architecture for Network, Security, and Storage Services, Addison-Wesley, ISBN: 978-0-13-662409-7, 2020

[23] F. Hutter, H. Hoos, and K. Leyton-Brown, Sequential Model-Based Optimization for General Algorithm Configuration, LION , volume 6683 of Lecture Notes in Computer Science, page 507-523. Springer, (2011)

[24] M. M. S. Lira, R. R. B. de Aquino, A. A. Ferreira, M. A. Carvalho, O. N. Neto and G. S. M. Santos, "Combining Multiple Artificial Neural Networks Using Random Committee to Decide upon Electrical Disturbance Classification," 2007 International Joint Conference on Neural Networks, Orlando, FL, 2007, pp. 2863-2868.

[25] J. Vanschoren, Meta-Learning, 2019 F. Hutter et al. (eds.), Automated Machine Learning, The Springer Series on Challenges in Machine Learning, https://doi.org/10.1007/978-3-030-05318-5_2 , pp 35-61