2018 IEEE Conference on Decision and Control (CDC)
Miami Beach, FL, USA, Dec. 17-19, 2018

Multivariate Deep Causal Network for Time series Forecasting in Interdependent Networks

Lalitha Madhavi K.S[1], Mostafa Gilanifar[1] , Yuxun Zhou[2], Eren E. Ozguven[1], and Reza Arghandeh[3]

Abstract— **A novel multivariate deep causal network model (MDCN) is proposed in this paper, which combines the theory of conditional variance and deep neural networks to identify the cause-effect relationship between different interdependent time-series. The MCDN validation is conducted by a double step approach. The self validation is performed by information theory-based metrics, and the cross validation is achieved by a foresting application that combines the actual interdependent electricity, transportation, and weather datasets in the City of Tallahassee, Florida, USA.**

I. INTRODUCTION

Discovering causal models from observational and intervention data is an important first step preceding counter factual reasoning; however, in such a causal network, the extent to which the direction of edges can be identified from purely observational data is still limited in the literature. Techniques based on conditional independence such as Peter Clarke's tests can only discover edge directions within the limits of Markov equivalence [1]. Several causality methods have also been extensively studied in the past including Granger causality test [2], Pearl's causality model [3], and Structural Equation Modeling methods [4]. In order to determine the direction of a pairwise casual relationship, statistical techniques can, in certain conditions, augment the causal discovery process. Furthermore, the causal modeling methods have recently started utilizing emerging techniques such as neural networks-based ones[5]. In order to fill this gap, in this paper, we propose a novel approach, namely the Multivariate Deep Causal Networks (MDCN), to detect the causal relationships between multivariate series of datasets through the use of a deep neural network structure.

Deep Neural Networks (DNN) are multilayer neural networks that consist of more than one layer of hidden units as inputs and its corresponding outputs. One of the fundamental part of a DNN is the Activation Functions, which are the deciding factors for the activation and functioning of a node. Whether the received information by a neuron is relevant for the given information or should it be ignored is also decided by the activation functions. The most popular activation functions (learning units) are namely Logistic Sigmoid (logsig), Hyperbolic Tangent, and Exponential Linear Unit (ELU). This paper utilizes a *first order derivative of the sigmoid hyperbolic activation function*, which can successfully determine both linear and non-linear relationships. To validate the proposed MDCN, various information theory-based distance metrics are utilized in this paper. The predicted outcome is validated by measuring the Kullback-Leibler Divergence (KLD) and the Maximum Mean Discrepancy (MMD) indices. KLD is a statistical distance matrix, which is utilized to measure the difference between different distribution functions considered over the same variable.

Furthermore, a cross validation is performed in order to evaluate the observed causal relationship. This involves performing a load forecasting study combining the direct causal variables as predictors in forecasting the electricity consumption. The established causality relationship thus helps in selecting the most informative and relevant variables in predicting the electricity consumption. The proposed causality method is further validated using synthetic datasets and compared with other well established causality methodologies.

Load forecasting is crucial for power system planning and operation. It helps in maintaining the balance between the electricity supply and demand. There have been many studies focusing on short term load forecasting combining various factors such as economic and weather parameters [6], [7]. In this paper, a real-world electricity load forecasting application is performed by combining electricity, transportation and weather data which is in continuation to our previous works [6],[8],[9]. The contributions of this paper can be listed as follows:

- This study develops a novel multi-variable causality analysis method using a Deep Neural Networks structure called MDCN;
- A novel activation function, the first order derivative of sigmoid hyperbolic, is proposed for Deep Neural Networks;
- MDCN is utilized for enhancing electricity load forecasting accuracy considering the interdependencies between electricity consumption, human mobility, and weather conditions.

In the next section, we focus on the state of the art causal models and present the proposed novel causality approach. Section III introduces the various validation techniques for the causal models. The results and conclusions are presented in the final section of the paper.

*This work is partly supported by the US NSF award 1640587

[1] Lalitha Madhavi K.S, is with the Department of Electrical Eng, Mostafa Gilanifar is with Department of Industrial and Manufacturing Eng., and Eren E. Ozguven is with Department of Civil and Environmental Eng, Florida State University, Tallahassee, FL 32310, USA

[2] Yuxun Zhou is with the Department of Electrical Eng and Computer Science Dept., University of California at Berkeley, CA 94720, USA

[3] Reza Arghandeh is an Associate Professor in Department of Computing, Mathematics and Physics at the Western Norway University of Applied Sciences, Norway.

978-1-5386-1395-5/18/$31.00 ©2018 IEEE

II. METHODOLOGY

This section introduces the proposed methodology. In this paper, a causality analysis is conducted to relate electricity with transportation and weather parameters. The proposed causal model is also compared with other well established causal models. These causal models are then validated using Information theory based metrics. The causal relationship is utilized in load forecasting to ensure the use of most informative predictors.

A. Multivariate Deep Causal Network (MDCN)

A directed acyclic graph (DAG) is a graph with directed edges which are used in causality theory to illustrate direct causal relationships. A directed graph is represented as $D = (V, E, f)$ consisting of a set of nodes V, edges E and a mapping $f : E \rightarrow V * V$. The V consists of the various input data variables and the edges describe the relationship between the variables. A generic form of the proposed causal model is illustrated in Figure 1. Where, $X = X_1, X_2, ..., X_m$ are the inputs to the first batch of hidden layers, based on the low conditional variance between them. Similarly, $Y = X_{m+1}, X_{m+2}, ..., X_n$ are inputs to the second set of hidden layers. These inputs are sorted based on conditional variance test. The inputs to each hidden layer are categorized based on conditional variance test.

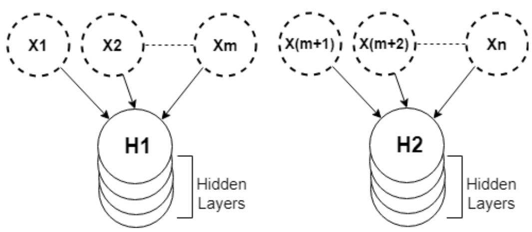

Fig. 1. Generic causal structures

The Conditional variance shows variance of a random variable given the values of one or more other variables. Consider two distinct variables X and Y. The conditional mean of Y given X = x is defined as

$$\mu_{Y|X} = E[Y|x] = \sum_y yh(y|x) \tag{1}$$

The conditional variance of Y given X = x is given as:

$$\sigma^2_{Y|X} = E[Y - \mu_{Y|x}]^2|x = \sum_y [Y - \mu_{Y|x}]^2 h(y|x) \tag{2}$$

After categorizing the inputs based on conditional variance as shown in Figure 2, activation functions are applied to initiate the functioning of the hidden layers of the MDCN. The two most common activation functions are the logistic sigmoid and hyperbolic ones [7] The logistic sigmoid activation function can be written as:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

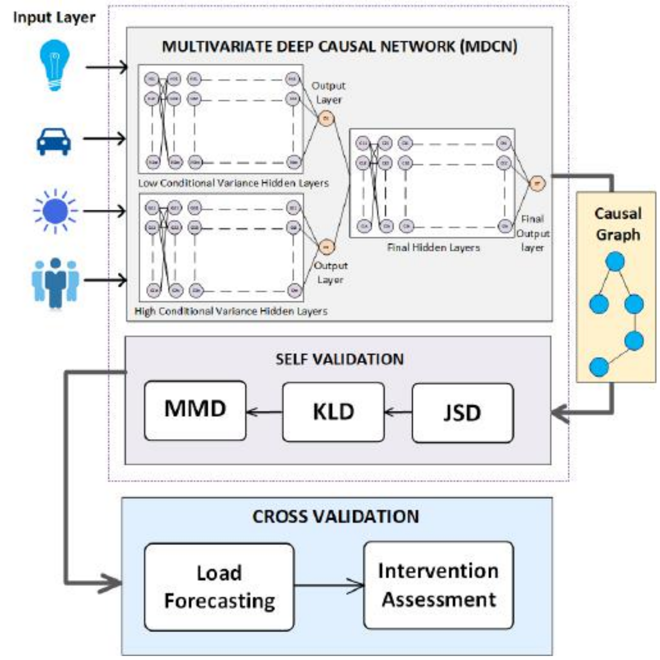

Fig. 2. The overall structure of proposed MDCN method

The hyperbolic activation function is given as:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{4}$$

From the above, sigmoid hyperbolic function can be written as a product of Equation 3 and Equation 4:

$$f(x) = \frac{e^x(2 + e^x) - 1}{e^x + e^{-x} + e^{-2x} + 1} \tag{5}$$

The derivative of sigmoid hyperbolic function outperforms the logistic sigmoid activation function in terms of computation efficiency and thus is used as the activation function in this paper. The first order derivative of the sigmoid hyperbolic function can be written as shown below:

$$f'(x) = \frac{e^2x(e^{3x} + 3e^x + 2)}{(e^{2x} + 1)^2} \tag{6}$$

The number of hidden layers is an essential part in the overall functioning of the neural network. There are various approaches to choose the optimal number of hidden layers. This paper considers the heuristic forward approach where a small number of hidden neurons are selected. The numbers are gradually increased by evaluating the performance of the neural network. 6 hidden layers have been used in MDCN causal methodology. As illustrated in Figure 2 the outcomes of low conditional variance and high conditional variance DNN units are then fed into the final DNN unit, which results in the causal model (graph) that relates the input variables (in our use case electricity, traffic, and weather data). The resultant causal model is validated in two steps as discussed below.

B. State-of-the-art Causal Models

1) Peter Clark's causal model: : The Peter-Clarke's(PC) algorithm, begins with a graph G consisting of a group of vertices V that indicates the input data variables.A conditional independence (CI) test with a significance level $0 < \alpha < 1$ is conducted. Initially starting with an undirected graph, this algorithm shapes this graph by excluding edges with no CI relations. The structure if formed based on first order CI test [10].

2) Multivariate Granger Causality Test: A multivariate granger causality (MGC) approach can be developed by factorizing the spectral density matrix. The spectral density matrix denoted by $S(\omega)$ constitutes a major branch of the multivariate Granger causality, that provides reasoning using different tools for example auto-power, such as multiple and partial coherence etc [11].

3) Structural Equation Modeling: An approach for evaluating models containing linear relationships among multi variable systems. SEM models consist of manifest and latent variables, which can be independent (exogenous) or dependent (endogenous). The theory behind the SEM model is established by characterizing a pattern of linear relationships among a set of observable and exogenous variables. The disadvantages of using SEM includes the complexity of theory itself and the application, and varying robustness for different factor scores [12].

III. MDCN Causal Model Validation

The proposed MDCN model performance is first compared with the state of the art causality techniques using different information theory based metrics. An electricity load forecasting study is then performed as a cross validation to to show the impact of resultant causal outcomes.

A. MDCN Self Validation

For the purpose of self validation, the following distance metrics considered:

1) Kullback-Leibler Divergence: The Kullback-Leibler Divergence (KLD), denotes the difference between two probability distributions p and q. For example, KLD of q from p, written as *KLD(p, q)*, signifies the amount of information lost when q is used to approximate p. The Kullback-Leibler Divergence is computed as shown below:

$$KLD(p||q) = \sum_{x \epsilon X} p ln \frac{p}{q} \tag{7}$$

The KLD measures the expected number of extra bits required to code samples from p when performed based on the values of q. Basically, p represents the actual distribution of the data or a very close approximation of the distribution [13].

2) Jensen-Shannon Divergence: The Jensen–Shannon Divergence (JSD) is a technique used for measuring the correlation of different probability distributions. The JSD is defined as shown in equation below:

$$JSD[p,q] = \frac{1}{2} KLD[p||n] + \frac{1}{2} KLD[q||n] \tag{8}$$

In the above equation, $n = (p+q)/2$ is an equal blend of distribution functions p and q[14]. The optimization of JSD is performed by using the data density.

3) Maximum Mean Discrepancy: The Maximum Mean Discrepancy (MMD) was first introduced as a distribution function matrix[15].

$$MMD[p,q] = (E_{p,q}[k(x,x') - 2k(x,y) + k(y,y')])^{1/2} \tag{9}$$

In the above equation, x and x' are the random variables with distribution p. y and y' are independent random variables with distribution q. Equation 9 represents the MMD of those distributions. Also, $E_p q$ denotes the expectations with respect to p and q respectively [16].

B. MDCN Cross Validation

The causal relationships thus achieved signifies information flow between directly causal variables. In order to quantify the flow of information between the causal variables, a predictive analysis has been performed. As a second step validation of the observed causal relationships, a short term electricity load forecasting is performed using the state-of-the-art forecasting methods. The resultant causal graphs denote the variables that have a direct cause on electricity consumption. Such variables are used as predictor variables to train the neural network. This paper studies different time series forecasting methodologies such as ARMA [17], Support Vector Machine (SVM) [18], Deep Neural Networks(DNN) [19] and Multi Linear Regression(MLR) [20].

IV. Data Description

The City of Tallahassee, Florida electricity consumption data which is used in this study has 30 min sampling rate for the entire year of 2015. Over 8 million data points were analyzed for the selected neighborhoods selected.

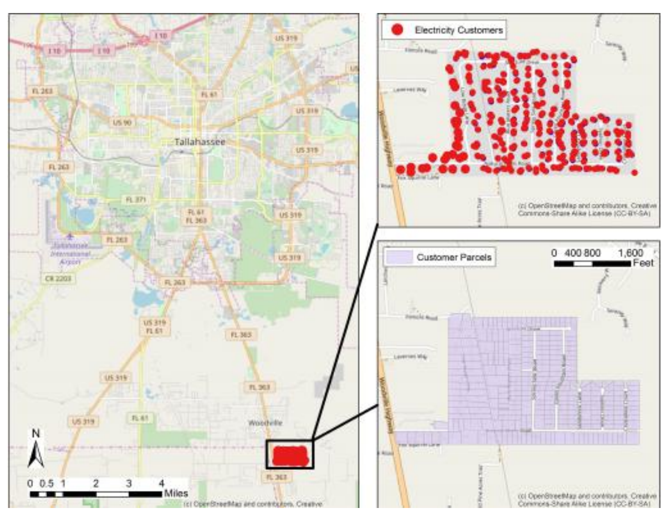

Fig. 3. Neighborhood considered for case study

For this paper, a neighborhood in the southeast part of the city is considered. The State of Florida consists of about 5000 Telemetered Traffic Monitoring Sites (TTMS) located around the state that streams and collects real time traffic flow information. Florida Department of Transportation (FDOT) receives high-volume and high-speed data collected for each and every lane with an interval of 1 hour. Figure 3 represents the neighborhood under consideration for the application of the MDCN causality and load forecasting.

V. Results and Discussions

Since electricity is highly impacted by weather changes, the causal models reveal a direct causal relationship between weather factors and electricity. Also, transportation which implies human mobility also influences electricity consumption. Therefore, north and south bound traffic also affects electricity consumption.

This section presents the result of the above mentioned methodology on actual electricity, transportation and weather data acquired from the city of Tallahassee.

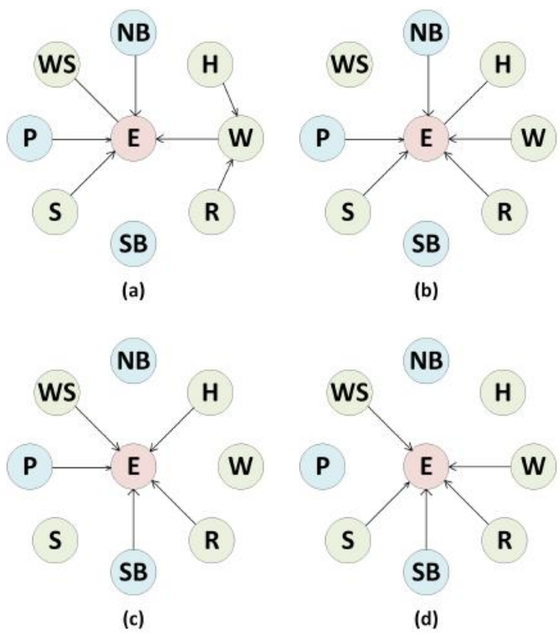

Fig. 4. Causal model outcomes from left: (a) MDCN, (b) PC, (c) SEM and (d) MG

To validate the performance of proposed MDCN causality model, we compare the resultant causal graph with three other causality methodologies including PC, SEM and MGC as illustrated in Figure 4. In this causal graph arrows denote direct causal relationship and line denotes dependency between the variables. In the causal graphs, E denotes electricity consumption, NB and SB represents north bound and south bound traffic flow, H and W represent the heat index and wind chill respectively. The rain rate, solar irradiation and wind speed are represented by R, S and WS respectively. For example, for the MDCN causal model as shown in Figure 4 (a), Electricity consumption, denotes a direct effect from North Bound traffic flow. Also weather parameters such as Wind Chill, Solar Irradiation and Wind Speed shows direct causal relationship with E. A study including the neighborhood population (P) as a variable in this analysis exhibited some causal relationship over a training dataset of 1 year. This can be explained since rate of change of population is significantly lower and hence population can be informative in a long term load forecasting.

For better comparing the various causal models, information theory based distance metrics such as MMD, KLD and JSD have been calculated as explained in Section II. The mean values of the distance metrics are given in the Table III:

TABLE I
Mean Distance Metrics Table for different causality approaches

Causal Approach	KLD	JSD	MMD
MDCN	0.21	0.18	0.23
SEM	0.62	0.47	0.72
PC	0.76	0.67	0.49
MGC	0.82	0.79	0.56

As observed in the table III, all the three distance metrics for the MDCN causal model outperforms considerably the other state-of-the-art causal models. For example, the KLD for MDCN is decreased by 66.13%, 72.37% and 74.39% when compared with SEM, PC and MGC causal models respectively. For further validation a load forecasting study was performed by combining the combination of inputs depicted to have a direct influence on electricity consumption.

To quantize the forecasting performance in this paper, we use two well known error indices. In order to compare each forecasting method, we used the Mean Absolute Percentage Error (MAPE) as defined by equation below:

$$MAPE = \frac{100}{m} \sum_{t=1}^{m} \left| \frac{x(t) - x'(t)}{x'(t)} \right| \tag{10}$$

Along with MAPE, Root Mean Square Error (RMSE) as defined by equation below, has also been considered as an additional index for measuring the error.

$$RMSE = \sqrt[2]{\frac{1}{m} \sum_{i=1}^{m} [x(t) - x'(t)]^2} \tag{11}$$

In equation (10) and (11), *x(t)* and *x'(t)* represent the original and predicted values respectively and *m* is the total number of observations. The resulting average MAPE and RMSE for load forecasting for a 1 week period using 3 months of training data are seen in Table II:

It is evident from table II that among the four selected load forecasting techniques, DNN outperforms the other methods. The table also shows the effect of combining all the variables as predictors without any causality techniques. It is observed that combining causality technique as the preprocessor for the load forecasting results in reduces the error in all cases. For example, the reduction in MAPE for DNN based load forecasting when combined with MDCN causality model

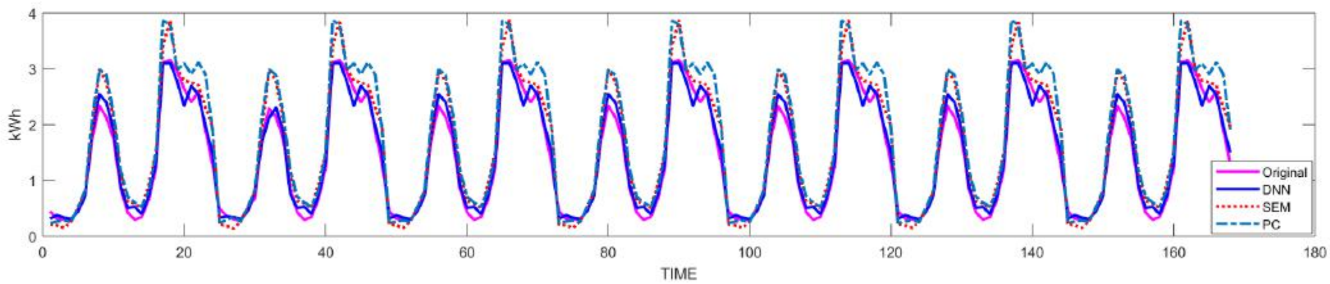

Fig. 5. 1 Week ahead forecasting for a week of January 2016

TABLE II
LOAD FORECASTING ERROR INDICES FOR 1 WEEK

Method	Error Index	No Causal	PC	SEM	MDCN
ARMA	MAPE	7.53	4.15	3.32	2.43
	RMSE	12.65	6.33	5.87	4.11
MLR	MAPE	6.32	4.42	5.62	4.21
	RMSE	11.41	6.27	7.61	5.93
SVM	MAPE	6.81	3.86	5.14	4.12
	RMSE	12.18	6.77	6.98	7.13
DNN	MAPE	5.61	2.96	3.22	2.11
	RMSE	11.12	4.27	3.61	2.23

results in error reduction in comparison with no causal, PC and SEM models are 62.38%, 26.72% and 34.47% relatively. Figure 6 illustrates the result of having different number of hidden layers. It is seen that, for MDCN, the number of hidden layers is inversely proportional to the electricity forecasting error. In this case study, 6 hidden layers are found to provide the most accurate result, therefore MDCN consists of 6 hidden layers containing 6 neurons each. It is also observed that on addition of more hidden layers, a saturation is achieved in the error percentages.

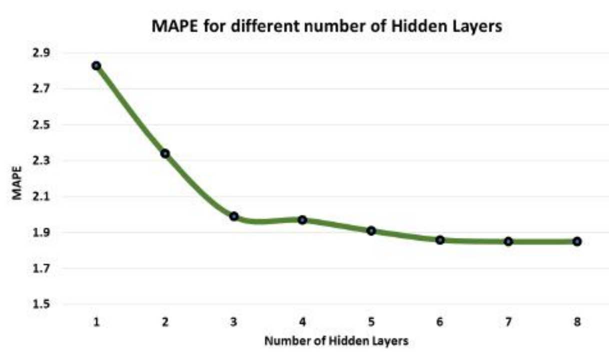

Fig. 6. Load Forecasting MAPE for various number of hidden layers in MDCN

The importance of analyzing causal relationships can be observed in Figure 7. It also depicts an Intervention Assessment based MAPE. The Intervention Assessment defines the process of combining all the variables as predictors irrespective of any causal significance. It is observed that, the MAPE in case of intervention assessment is higher, since the predictors use a combination of irrelevant variables as input predictors. The MAPE reduces when inputs combines past values of electricity consumption with weather parameters (E-W). Also combining electricity consumption with traffic flow information (E-T) makes a further reduction in error. However, the highest accuracy is achieved when historical values of electricity consumption is combined with weather parameters and traffic flow information (E-W-T). The relative decrease in MAPE when compared to E only, E-W, E-T is 84.14%, 71.71% and 58.58% respectively.

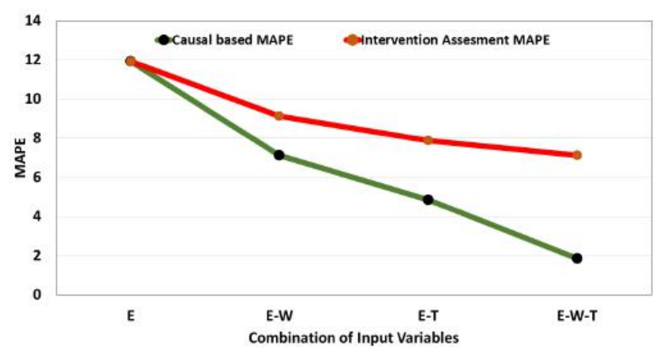

Fig. 7. MAPE for combination of different Inputs

VI. VALIDATION USING SYNTHETIC DATASET

The causal models are further validated using synthetic dataset. The datasets were generated following the methodology in [21]. The 5 series dataset consists of a ground truth causal relationship as shown in Figure 8.

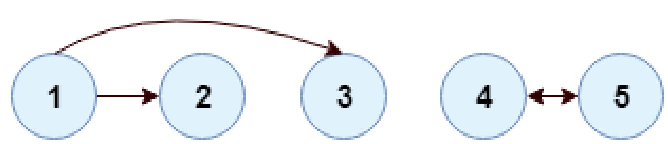

Fig. 8. Synthetic Data Ground Truth Causal Model

The performance of the causality tests were compared by using different forecasting methods: ARMA, SVM and MLR using the synthetic dataset. The Table below provides the comparison of the different causal models with the forecasting approaches. The causal graphs achieved on the synthetic dataset has also been illustrated in Figures 9-12.

TABLE III

FORECASTING MAPE FOR SYNTHETIC DATASET BASED COMPARISON OF VARIOUS CAUSAL MODELS

Causal Approach	ARMA	MLR	SVM
MDCN	0.16	0.13	0.11
SEM	0.33	0.37	0.41
PC	0.26	0.22	0.19
MGC	0.48	0.42	0.33

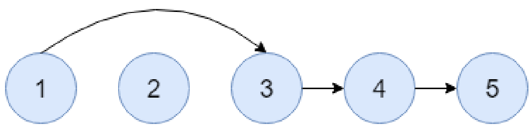

Fig. 9. Synthetic dataset on Multivariate Granger causality

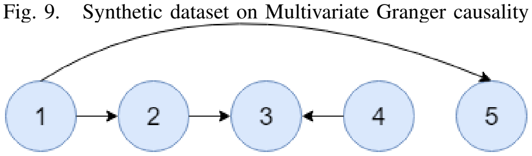

Fig. 10. Synthetic dataset on PC algorithm

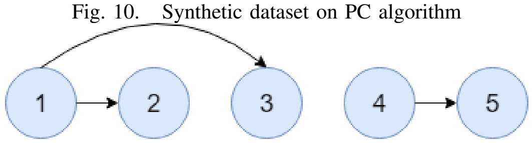

Fig. 11. Synthetic dataset on MDCN

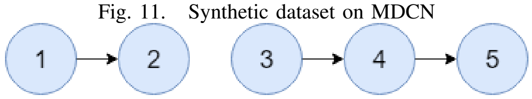

Fig. 12. Synthetic dataset on SEM

VII. CONCLUSIONS

We propose a novel approach for causality combining conditional variance and a deep neural network structure. The proposed MDCN causal approach outperformed the other state-of-the-art causality methods in the self validation as well as cross validation. Using the MDCN causal model as a preprocessor for the electricity load forecasting to combine transportation and weather along with electricity brings down the relative error percentage by 84.14% for the case study under consideration.

ACKNOWLEDGMENT

The authors would like to thank the City of Tallahassee for providing the data and expertise. Special thanks also goes to Prof. Mattia Zorzi and Prof. Alessandro Chiuso from the University of Padova, Italy for their advice and counsel about causality analysis.

REFERENCES

[1] J. Koster, '1. causation, prediction, and search. 2nd edn. peter spirtes, clark glymour and richard scheines, mit press, cambridge, ma, 2000. no. of pages: 543. isbn 0-262-19440-6', *Statistics in Medicine*, vol. 22, no. 13, pp. 2236–2237, 2003.

[2] C. W. J. Granger, 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969, ISSN: 00129682, 14680262.

[3] J. PEARL, 'Causal diagrams for empirical research', *Biometrika*, vol. 82, no. 4, pp. 669–688, 1995.

[4] D. Yang, H. Chen, Y. Song and Z. Gong, 'Granger causality for multivariate time series classification', in *2017 IEEE International Conference on Big Knowledge (ICBK)*, Aug. 2017, pp. 103–110.

[5] L. Q. Uddin, A. Clare Kelly, B. B. Biswal, F. Xavier Castellanos and M. P. Milham, 'Functional connectivity of default mode network components: Correlation, anticorrelation, and causality', *Human Brain Mapping*, vol. 30, no. 2, 2009.

[6] K. S. L. Madhavi, J. Cordova, M. B. Ulak, M. Ohlsen, E. E. Ozguven, R. Arghandeh and A. Kocatepe, 'Advanced electricity load forecasting combining electricity and transportation network', in *2017 North American Power Symposium (NAPS)*, Sep. 2017, pp. 1–6.

[7] S. R. Salkuti, 'Short-term electrical load forecasting using radial basis function neural networks considering weather factors', *Electrical Engineering*, Jan. 2018.

[8] L. M. K. Sriram, M. Gilanifar, Y. Zhou, E. E. Ozguven and R. Arghandeh, 'Causal markov elman network for load forecasting in multi network systems', *IEEE Transactions on Industrial Electronics*, pp. 1–1, 2018, ISSN: 0278-0046. DOI: 10.1109/TIE.2018.2851977.

[9] M. B. Ulak, A. Kocatepe, L. M. Konila Sriram, E. E. Ozguven and R. Arghandeh, 'Assessment of the hurricane-induced power outages from a demographic, socioeconomic, and transportation perspective', *Natural Hazards*, vol. 92, no. 3, pp. 1489–1508, Jul. 2018, ISSN: 1573-0840. DOI: 10.1007/s11069-018-3260-9. [Online]. Available: https://doi.org/10.1007/s11069-018-3260-9.

[10] P. Spirtes and K. Zhang, 'Causal discovery and inference: Concepts and recent methodological advances', *Applied Informatics*, vol. 3, no. 1, p. 3, Feb. 2016.

[11] R. R. Ahmed, J. Vveinhardt, D. Streimikiene and M. Fayyaz, 'Multivariate granger causality between macro variables and kse 100 index: Evidence from johansen cointegration and toda & yamamoto causality', *Economic Research-Ekonomska Istraživanja*, vol. 30, no. 1, pp. 1497–1521, 2017.

[12] C. Nachtigall, U. Kroehne, F. Funke and R. Steyer, '(why) should we use sem?—pros and cons of structural equation modelling', vol. 8, Jan. 2003.

[13] D. Polani, 'Kullback-leibler divergence', in *Encyclopedia of Systems Biology*, W. Dubitzky, O. Wolkenhauer, K.-H. Cho and H. Yokota, Eds. New York, NY: Springer New York, 2013, pp. 1087–1088.

[14] B. Fuglede and F. Topsoe, 'Jensen-shannon divergence and hilbert space embedding', in *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, Jun. 2004, pp. 31–.

[15] A. Smola, A. Gretton, L. Song and B. Schölkopf, 'A hilbert space embedding for distributions', in *Algorithmic Learning Theory*, M. Hutter, R. A. Servedio and E. Takimoto, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 13–31.

[16] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf and A. Smola, 'A kernel two-sample test', *J. Mach. Learn. Res.*, vol. 13, pp. 723–773, Mar. 2012, ISSN: 1532-4435.

[17] S.-J. Huang and K.-R. Shih, 'Short-term load forecasting via arma model identification including non-gaussian process considerations', *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 673–679, May 2003.

[18] J. Shi, W. J. Lee, Y. Liu, Y. Yang and P. Wang, 'Forecasting power output of photovoltaic systems based on weather classification and support vector machines', *IEEE Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064–1069, May 2012.

[19] D. L. Marino, K. Amarasinghe and M. Manic, 'Building energy load forecasting using deep neural networks', in *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, Oct. 2016, pp. 7046–7051.

[20] V. Bianco, O. Manca and S. Nardini, 'Electricity consumption forecasting in italy using linear regression models', *Energy*, vol. 34, no. 9, pp. 1413–1421, 2009.

[21] L. Luo, W. Liu, I. Koprinska and F. Chen, 'Discovering causal structures from time series data via enhanced granger causality', in *AI 2015: Advances in Artificial Intelligence*, B. Pfahringer and J. Renz, Eds., Cham: Springer International Publishing, 2015, pp. 365–378, ISBN: 978-3-319-26350-2.