Examining nominal and ordinal classifiers for forecasting wind speed

George Georgoulas
Laboratory of Knowledge and Intelligent Computing
Department of Computer Engineering
Technological Educational Institute of Epirus
Arta, Greece
georgoul@kic.teiep.gr

Stavros Kolios
Laboratory of Knowledge and Intelligent Computing
Department of Computer Engineering
Arta, Greece
stavroko@gmail.com

Petros Karvelis
Cardiovascular Biology and Biomechanics Laboratory
Cardiovascular Division, University of Nebraska
Medical Center Omaha, USA
petros.karvelis@unmc.edu

Chrysostomos Stylios
Laboratory of Knowledge and Intelligent Computing
Department of Computer Engineering
Arta, Greece
stylios@teiep.gr

***Abstract*—Wind speed forecasting is an essential problem for many applications i.e. estimating the short-term energy production from wind farm operations. This work investigates and compares different approaches for the problem of wind forecasting in a simplified manner considering only the prediction of the range within which the mean wind value will fall in the next time step. The simplified forecasting approach which treats forecasting as a classification problem is suitable for this specific application. Moreover, there is a natural ordering of the predefined classes, thus in this work ordinal classification – also known as ordinal regression- approaches are tested and compared with conventional nominal classifiers. The preliminary results indicate that considering the natural ordering of the classes yields the best performance for the specific test site involved in this study.**

Keywords-Ordinal Classification/Regression; LS-SVMs; SVOREX; C4.5 Decision Trees; Logistic Regression Models; Weather Forecasting

I. Introduction

Meteorological parameters are essential for the study of a variety of phenomena taking place in the atmosphere and on the Earth surface. The climate change, the intensification of greenhouse effect, the local disturbances of climatic profiles etc., comprise only a small number of cases where monitoring and forecasting of the meteorological parameters is essential. Estimation of these parameters offers valuable information for sustainable environmental management and future planning, warning about extreme weather phenomena, estimation of short–term energy production for the case of wind farms, etc. Thus, analytical and continuous recording of meteorological parameters as well as their timely and accurate forecasting are of paramount importance and significance.

There is a large number of studies in the international literature that analyze time-series of various meteorological parameters from different points of view. In some studies, the goal is to classify the weather types and estimate potential long-term climate change [1]-[3] while in others the goal is the analysis of extreme values [4]-[7]. Furthermore, time-series of meteorological data are used either for short-term or long-term forecasting of the evolution of these parameters using and integrating different types of data and methods [8]-[13] with various degrees of success.

For forecasting applications, the local character and the complex nature of the meteorological parameters can affect the performance and the accuracy of statistical methods. For this reason, it is common to test various approaches regarding their efficiency in order to provide accurate weather forecasts from local up to regional scales.

Furthermore, the irregular variations of meteorological parameters, require novel, nonlinear and multi parametric statistical procedures to improve/complement conventional statistical methods. Artificial Neural Networks (ANN), fuzzy logic, genetic algorithms and other computational intelligence approaches are already applied in the field. Saima *et al.* presented a short review about the significant efficiency of hybrid models such as ARIMA models (Auto Regressive Integrated Moving Average) and neuro-fuzzy algorithms in forecasting meteorological parameters [14]. In that review [14], the wide use of different modern algorithms and the significant accuracy that can be achieved in weather forecasting is highlighted. Aggarwal and Kumar [15] examined the contribution of statistical based modelling approaches for weather prediction providing comparable accuracy with the physical based modelling, offering a less expensive and more easily to use solution for weather prediction. Bushara and Abraham [16] presented an overview regarding the efficiency of computational intelligent methods in weather forecasting, illustrating the good performance and reasonable prediction accuracy that can be achieved.

Almost all the aforementioned works, tackle the problem of forecasting meteorological parameters as a conventional time series prediction problem, with the meteorological parameters

978-1-5090-1353-1/16/$31.00 ©2016

treated as any real valued quantities. However, in some cases such a fine grain representation, is not necessary and coarser representations could comprise a useful transformation of the same problem into a simpler one. Such an approach pursued in [17] where wind prediction is cast as a classification problem where the inputs to the classifier are the measurements of the synoptic pressure. Moreover, in this work, both ordinal as well as nominal classification approaches are considered since a natural ordering exists among the different involved classes / wind levels.

In this study, different ordinal and nominal classification approaches for the short-range forecasting of wind speed (up to the next three hours), are applied. It is noteworthy that such approaches can provide accurate estimations [18]-[20] over short periods (up to the next several hours). For this study, Data records come from the meteorological station located near Agrinio city, providing archive datasets with contiguous three-hourly wind and relative humidity measurements.

The rest of the paper is structured as follows. Section II provides a short description of the ordinal regression/classification problem along with the performance measures that are usually involved to assess the effectiveness of such settings. Section III presents all six investigated classification schemes while section IV summarizes the available data for the study. Section V presents the classification results and section VI concludes the paper presenting the lessons learned as well as a roadmap for feature research.

II. ORDINAL REGRESSION PROBLEM DEFINITION

Regression is one of the most fundamental problems concerning machine learning. Many techniques such as Support Vector Machines (SVMs) [21] have been proposed in order to deal with this kind of problems. However, less attention has been paid to ordinal regression also called ordinal classification or ranking learning problem [22], where the labels of the target variable exhibit a natural ordering.

Ordinal problems are all around us. For example, student satisfaction surveys usually involve rating teachers based on an ordinal scale {poor, average, good, very good, excellent}. Hence, class labels contain order information, e.g. a sample vector associated with class label "average" has a higher rating (or better) than another from the "poor" class, but "good" class is better than both. On the medical field the condition of a patient could be considered as normal, suspicious or pathological [23], which indicates that there is an increase of the severity condition as it moves from the normal to the pathological class.

Formally speaking, the ordinal regression problem consists of predicting the label y of an input vector x, where $x \in X \subseteq \Re^K$ and a set of labels $y \in Y = \{C_1, C_2, ..., C_Q\}$. In other words, x lies in a K-dimensional input space and y is a label space of Q different labels/classes. The objective of ordinal regression/classification is to find a function $r : X \to Y$ able to predict the labels of new patterns, given a training set of N points, $D = \{(x_i, y_i),\ i = 1, ..., N\}$ and considering an ordering of the labels e.g. $C_1 \prec C_2 \prec ... \prec C_Q$ where $\prec$ defines the ordering relation between the labels. Moreover, in conventional classification paradigms, this ordering of the classes is not taken into consideration during the estimation of the error rate (which is equivalent to 1 – the classification accuracy) which is the most prevalent means for assessing the performance of a (conventional) classifier. Therefore, other measures could be considered [24], [25] when it comes to the assessment of different ordinal classifiers.

Thus given the predicted $\hat{y}_i$ and the actual labels y_i of a data set, the following measures are involved in this study:

Error Rate (*ER*) (1-Accuracy)

$$ER = \frac{1}{N}\sum_{i=1}^{N} I(\hat{y}_i \neq y_i) \tag{1}$$

with $I(\cdot)$ being the indicator function.

Mean Absolute Error (*MAE*)

MAE as its name implies, is the average deviation of the prediction from the true class

$$MAE = \frac{1}{N}\sum_{i=1}^{N} \left| s(\hat{y}_i) - s(y_i) \right| \tag{2}$$

where s is a score function which in its simplest form-also considered here- is given by $s(y_i) = i$, $1 \le i \le Q$

Average Mean Absolute Error (*AMAE*)

In the case of imbalanced classes (as in our case - see section IV) the aforementioned two measures may not be the most appropriate for evaluating the performance of a learning algorithm. Therefore, for that case, the Average MAE (AMAE) [26] is considered, which takes into account the differences in size of the involved classes:

$$AMAE = \frac{1}{Q}\sum_{j=1}^{Q} \frac{1}{N_j}\sum_{i=1}^{N_j} \left| s(\hat{y}_i) - s(y_i) \right| \tag{3}$$

where N_j is the number of cases belonging to class j.

III. CLASSIFICATION SCHEMES

In this work three nominal and three ordinal classification methods are tested. The set of nominal classifiers consists of the Multinomial Logistic Regression (MLR) [27], the C4.5 Decision Tree (DT) classifier [28], and the Least Squares SVM (LS-SVM) classifier [29]. The set of ordinal classifiers consists of the Proportional Odds Model (POM) [30], a simple ordinal scheme proposed in [31] involving the C4.5 as the base classifier, and the Support Vector Ordinal Regression with Explicit constraints (SVOREX) algorithm [32].

A. Nominal Classifiers

1) Multinomial Logistic Regression: Logistic regression is a popular method for binary classification, with MLR being its extension to multiclass problems. In MLR the posterior probabilities of Q classes are given using linear functions ofthe input vector x . Therefore, given an input vector x the model has the form:

$$\log\frac{\Pr\left(class{=}n\middle|X=x\right)}{\Pr\left(class{=}Q\middle|X=x\right)}=b_{n0}+b_n^T x\text{ , for }n=1,\ldots,Q-1 \qquad (4)$$

which can also be written as:

$$\Pr\left(class=n\middle|X=x\right)=\frac{\exp\left(b_{n0}+b_n^T x\right)}{1+\sum_{m=1}^{C-1}\exp\left(b_{m0}+b_m^T x\right)}, \qquad (5)$$
$$\text{for } n=1,\ldots,Q-1$$

$$\Pr\left(class=Q\middle|X=x\right)=\frac{1}{1+\sum_{m=1}^{k-1}\exp\left(b_{m0}+b_m^T x\right)}. \qquad (6)$$

The parameters of the model are usually estimated using the maximum likelihood technique [27].

2) C4.5 Classsifier: The C4.5 algorithm builds a DT based on a divide and conquer strategy. During the training phase each node of the tree is assigned a number of samples which are weighted to take into account unknown attribute values. Given that the set of samples associated with the node t is denoted as V , the weighted frequency $freq(i,V)$ of cases in V whose class is i , $i\in[1,..,Q]$ with Q being the number of classes, is computed. If V contains cases belonging to two or more classes then the information gain must be computed for each attribute

$$gain=\text{info}(V)-\sum_i\frac{|V_i|}{|V|}\times\text{info}(\text{V}_i) \qquad (7)$$

where $V_i, i=1,\ldots,s$ is the set of the splitting produced by the test on the selected attribute and s is the number of splitting of node t . Finally, we compute the entropy $\text{info}(V)$ of the set V :

$$\text{info}(V)=\sum_{j=1}^{Q}\frac{freq(j,\text{V})}{|V|}\times\log_2\left(\frac{f(j,V)}{|V|}\right) \qquad (8)$$

If V_i is not empty, the divide and conquer approach consists of recursively applying the same operations on the set consisting of V_i plus those cases in V with unknown value of the selected attribute. At this point, it is noted that apart from a binary decision, C4.5 can also provide the probability of an input vector belonging to a specific class, making it a valid candidate for its use as a base classifier for the case of the simple ordinal classification scheme proposed in [31].

3) LS-SVM: The LS-SVM classifier belongs to the family of SVMs but the optimization problem involved is linear instead of the usual quadratic one. To be more specific the optimization problem given a set of labeled instances $D=\{(x_i,y_i),\ i=1,\ldots,N\}$, with $x\in X\subseteq\Re^K$ and $y\in\{-1,1\}$, is the following [29]:

$$min_{w,b,e}\,F_2\left(w,e\right)=\frac{1}{2}w^T w+\frac{1}{2}\gamma\sum_{i=1}^{N}e_i^2 \qquad (9)$$

Subject to: $y_i\left[w^T\varphi\left(x_i\right)+b\right]=1-e_i,\ i=1,\ldots,N$

where e_i is an error variable and γ is the regularization parameter. As in the case of standard SVMs the above formulation leads to the construction of a decision function of the form:

$$y\left(x\right)=sign\left(\sum_{i=1}^{N}a_i y_i K\left(x_i,x\right)+b\right) \qquad (10)$$

where $K(\cdot,\cdot)$ is a kernel function that implicitly performs the mapping form the input to the high dimensional feature space and a_i are the Lagrange multipliers. In this work the following RBF kernel is used:

$$K\left(x_i,x_j\right)=exp\left(\frac{-\left\|x_i-x_j\right\|_2^2}{\sigma^2}\right) \qquad (11)$$

where σ is the spread parameter of the RBF kernel.

B. Ordinal Classifiers

In order to tackle the problem of ordinal regression a number of methods have been proposed. The most naïve idea is to transform the ordinal scales into real values, and then solve the problem as a standard regression problem using e.g. Support Vector Regression (SVR) [33]. Other methods simply ignore the ordering of the labels and consider the task as a standard nominal classification problem. A more advanced method considered in this group is the cost-sensitive learning [34]. These methods require the definition of different costs for different types of misclassification errors.

Another group of methods is based on decomposing the ordinal target variable into several binary ones which are then estimated by a single or multiple models. Different binary decompositions can be considered [24]: *OneVsAll*, *OneVsOne*, *OrderedPartitions*, *OneVsNext*, *OneVsFollowers*, *OneVsPrevious*.

Finally, the last set of algorithms assumes that an unobserved variable underlies the ordinal response. Such a variable is called a latent variable, and the methods that are based on this assumption are known as threshold models [24], [35].

1) Proportional Odds Model: POM was developed to deal with ordinal regression [30]. Unlike the standard MLR model where the probabilities of each class are modeled, in POMs the cumulative probabilities are estimated instead $\Pr\left(class\le n\middle|X=x\right)$ which can be subsequently transformed to standard class probabilities.

$$\Pr\left(class\le n\middle|X=x\right)=\sum_{j=1}^{n}\Pr\left(\text{class}{=}j\middle|X=x\right) \qquad (12)$$

$$\Pr\left(class = n \middle| X = x\right) = \Pr\left(class \le n \middle| X = x\right) - \Pr\left(class \le n\text{-}1 \middle| X = x\right) \tag{13}$$

with $n = 2,...,Q$ and

$\Pr\left(class = 1 \middle| X = x\right) = \Pr\left(class \le 1 \middle| X = x\right)$ and

$\Pr\left(class \le Q \middle| X = x\right) = 1$

Given an input vector x the POM has the form:

$$\ln\left(\frac{\Pr\left(class \le n \middle| X = x\right)}{1 - \Pr\left(class \le n \middle| X = x\right)}\right) = a_n + b^T x, \quad \text{for } n = 1,...,Q-1 \tag{14}$$

The parameters of the model can be estimated using the maximum likelihood technique.

2) Binary Decomposition using C4.5 as the base classifier

The method transforms the original Q class ordinal problem into $Q-1$ binary class problems and uses the probabilistic values of a classifier to predict the class value. Sequentially a model is built to predict what is the probability of a given instance to belong at any of the classes that are located higher than C_1, higher than C_2 etc. up to the probability of the instance belonging to the "highest" class C_Q. After that it is simple matter to find the class with the highest probability using the following set of equations:

$$\Pr\left(C_1 \mid x\right) = 1 - \Pr\left(class \succ C_1 \mid x\right) \tag{15}$$

$$\Pr\left(C_n \mid x\right) = \Pr\left(class \succ C_{n-1} \mid x\right) - \Pr\left(class \succ C_n \mid x\right), \quad \text{for } 1 < n < Q \tag{16}$$

$$\Pr\left(C_Q \mid x\right) = \Pr\left(class \succ C_{Q-1} \mid x\right) \tag{17}$$

In this work the C4.5 is involved as the base binary classifier within this simple ordinal classification scheme as in the original publication [31].

3) Support Vector Machines for Ordinal Regression

Because of their good generalization performance, SVM models are the most widely applied ones to ordinal regression, and their structure is easily adapted to that of latent variable models. Chu and Keerthi [32] adapted the well-known SVM formulation for the ordinal classification problem by defining a different threshold for each class. They did that by splitting different intervals of the projection real line $w^T x$ using a threshold vector b instead of deciding the class of the pattern by the sign of the projection $w^T x$. This results in defining parallel hyperplanes with the same w and different thresholds b_j. Two different implementations have been proposed [32]: SVOR with Explicit constraints (SVOREX) and SVOR with Implicit constraints (SVORIM). In our case, the SVOREX implementation is used, where explicit constraints are imposed on the optimization problem by considering adjacent labels for threshold determination. Let N_q be the number of patterns of class C_q and let x_i^q be the patterns that belong to class C_q then the SVOREX problem is defined as follows:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|_2^2 + C\sum_{q=1}^{Q-1}\left(\sum_{i=1}^{Nq}\xi_i^q + \sum_{i=1}^{Nq+1}\xi_i^{*(q+1)}\right) \tag{18}$$

s.t.

$$w \cdot x_i^j - b_q \le -1 + \xi_i^q,\ \xi_i^q \ge 0,\ i = 1,2,...,N_q \tag{19}$$

$$w \cdot x_i^j - b_q \ge +1 + \xi_i^{*(q+1)},\ \xi_i^{*(q+1)} \ge 0,\ i = 1,2,...,N_{q+1} \tag{20}$$

$$b_{q-1} \le b_q \tag{21}$$

where $q = 1,2,...,Q-1$ and $C > 0$.

IV. Data

The data used in this study comes from a meteorological station located closely to the center of Agrinio city (Western Greece) (lon.: 21°21'11", Lat.: 38°36'21", Height: 24 m). The greater area has a complex topography with four small lakes surrounding the city, located as a relatively small distance from the Ionian Sea (at about 25-30 kilometers) and with a relatively high mountainous ridge northeastern of the city (Fig. 1). Between the Ionian Sea and the city of Agrinio, there are extensive agricultural areas and with Acheloos River (one of the largest rivers in Greece) running through them. The topography affects significantly the surface thermal variation and consequently the lapse rate in the atmospheric boundary layer as well as the humidity in the lower parts of the troposphere. These factors can favor cloud formation, often of convective nature, while the thermal variations due to the different response of land cover types in the solar radiation can trigger local wind flows, which in combination with land-sea thermal distribution can affect significantly the intensity and the direction of wind fields in the lower parts of troposphere (inside the boundary layer) from local up to regional scales. This makes wind forecasting a very challenging task. In this region, wind parks are already in operation and a detailed analysis of the wind profile can have an added value. To be more specific, short-term wind forecasting and the corresponding energy production forecasting is very important for the stability of the grid of electric transmission system and in the near future for the electric energy market operations.

Figure 1. The greater area of Agrinio with the surrounding lakes, the Acheloos river and the Ionian Sea at the western part.

The data for this study comes from the Hellenic National Meteorological Service. Data from two consecutive years is used for training and tuning of the model parameters while the next year's data is used for the estimation of the performance of the different algorithms. The data set consists of 3-hourly records of mean wind (km/h) and mean relative humidity values (%). For the needs of the study, the original range is converted into three classes consistent with the initial three classes (labeled as "low", "moderate" and "high") of the Beaufort classification scale. Not all the available classes of this scale is used, because the measurements are mean three-hourly, thus very high values such as those corresponding to wind gust are highly unlikely to be recorded.

In all cases the input to the model consists of the previous three recorded values for the wind (W) and the relative humidity (RH) $[W(t), W(t-T), W(\mathrm{t}-2\mathrm{T}), RH(t), RH(t-T), RH(t-2T)]$ and the model is trained to predict the wind speed at $t+T$, where T=3h. Fig. 2 depicts a schematic of the proposed approach. The values for the time lags were empirically selected based on their correlation with the output variable evaluated using the training data.

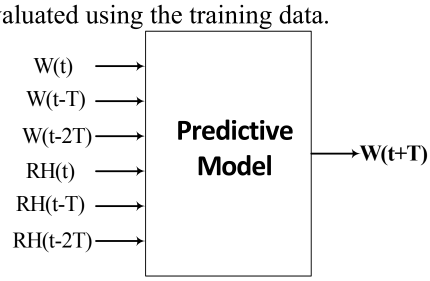

Figure 2. A schematic representation of the input/output relationship of the proposed wind forecasting approach.

Due to the heavily imbalance nature of the data set, as it is seen at histogram of Figure 3, the synthetic minority oversampling technique (SMOTE) is employed to tackle the highly skewed class distribution [37]. SMOTE is applied in such a way so as the three classes to have approximately the same number of cases in the training data set.

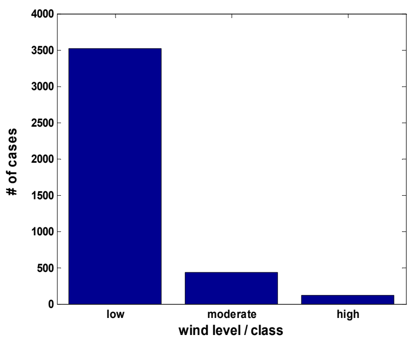

Figure 3. The distribution of cases among the three classes (wind levels) in the training set.

V. RESULTS

The performance of the different algorithms is summarized in Table I, where the best achieved values are in bold.

TABLE I. PERFORMANCE OF THE SIX CLASSIFICATION SCHEMES

Classifier	ER	MAE	AMAE
MLR	0.3158	0.3405	0.1135
C4.5	0.2471	0.2840	0.0947
LSSVM	0.2503	0.3062	0.1021
MOP	0.3158	0.3463	0.1154
ASA[1] C4.5	**0.2287**	**0.2510**	**0.0837**
SVOREX	0.2846	0.3151	0.1050

[1]ASA stands for "A Simple Approach" to ordinal classification adopted by [17].

Table I presents the performance results, where the best regarding the three measures is achieved using the simple approach to ordinal classification, the ASA C4.5, which proposed by Frank and Hall [31]. Then follows the C4.5 in its base classifier and then the two SVM approaches follow in performance with the nominal setting (LS-SVM) performing slightly better than the ordinal paradigm (SVOREX). The logistic models (MLP and MOP) had the worst performance among all six algorithms.

VI. CONCLUSIONS

In this work, six different classifications methods are tested for short-range wind forecasting. The wind values are aggregated to form three classes and so the forecasting problem is transformed into a classification one. To the best of our knowledge this is the first attempt that treats the wind forecasting problem in such a way using as inputs historical data on wind and relative humidity values.

Three of the involved classifiers take into consideration the ordinal nature of the three wind classes and three of them completely neglect that information. The results are a bit surprising since in a similar study [17] SVOREX consistently outperformed the simple approach to ordinal classification, which was proposed in [31] using C4.5 as its base classifier. In the present study however, the simple approach of ordinal classification achieved the best results compared to all other five algorithms. Moreover, the nominal classifiers achieve comparable, if not better classification performance compared to their ordinal counterparts, with the exception of C4.5 which behaves better under the ordinal scheme (in fact it outperforms all other algorithms). The lack of evidence that the ordinal approach offers an advantage for this specific setting is probably due to the small number of classes. As it was observed in [25] the ordinal classifiers have comparable performance to nominal ones for small number of classes. Further investigation is needed before safer conclusions can be drawn especially since the simplified approach, which achieved the best results, is usually less competitive than other more advanced methods [24].

In future work more data, spanning a longer time period, as well as data coming from different stations and including other measurements apart from wind and the relative humidity values will be tested, while a feature selection stage [38] will be included to improve efficiency. Furthermore, more data will be gathered to allow the experimentation with different class formations (using more than three classes).

Acknowledgments

This work was supported by the E.U. FP7–PEOPLE–IAPP–2009, Grant Agreement No. 251589, Acronym: SAIL under the national co-funded.

References

[1] A.R. Naik, and S.K. Pathan "Weather classification and forecasting using back propagation feed-forward neural network," Int. J. of Scientific and Research Publications, vol. 2, no. 12, pp. 2250-3153, 2012.

[2] F. Olaiya, and A.B. Adeyemo, "Application of data mining techniques in weather prediction and climate change studies," Int. J. of Information Engineering and Electronic Business, vol. 1, pp. 51-59, 2012.

[3] M. Saha, P. Mitra, and A. Chakraborty, "Fuzzy Clustering Based ensemble approach to predicting Indian Monsoon" Advances in Meteorology, 2015.

[4] K. Goubanova, and L. Li, "Extremes in temperature and precipitation around the Mediterranean basin in an ensemble of future climate scenario simulations," Global and Planetary Changes, vol. 57, pp. 27-42, 2007.

[5] A. Kalimeris, D. Founda, C. Giannakopoulos, and F. Pierros "Long term precipitation variability in the Ionian islands (Central Mediterranean): Climatic signal analysis and future projection," Theoretical and Applied Climatology, vol. 109, pp. 51–72, 2011.

[6] A.F. Karagiannidis, T. Karacostas, P. Maheras, and T. Makrogiannis, "Climatological aspects of extreme precipitation in Europe, related to mid-latitude cyclonic systems," Theoretical and Applied Climatology, vol. 107, pp. 165-174, 2012.

[7] M.S. Varfi, T.S. Karacostas, T.J. Makrogiannis, and A.A. Flocas, "Characteristics of the extreme warm and cold days over Greece," Advances in Geosciences, vol. 20, pp. 45-50, 2009.

[8] S.A.P. Kani, and M.M. Ardehali, "Very short-term wind speed prediction: A new artificial neural network-Markov chain model," Energy Conservation and Management, vol. 52, no. 1, pp. 738-745, 2011.

[9] Z. Guo, D. Chi, J. Wu, and W. Zhang, "A new wind speed forecasting strategy based on the chaotic time series modelling technique and the Apriori algorithm," Energy Conservation and Management, vol. 84, pp. 140-151, 2014

[10] A. Pierre, and V. Monbet, "Markov-switching autoregressive models for wind time series," Environmental Modelling and Software, vol. 30, pp. 92-101, 2012.

[11] S. Chattopadhyay, D. Jhajharia, and G. Chattopadhyay, "Univariate modelling of monthly maximum temperature time series over northeast India: neural network versus Yule-Walker equation based approach," Meteorological Applications, vol.18, pp. 70-82, 2011.

[12] F. Almonacid, P. Perez-Higueras, P. Rodrigo, and L. Hontoria, "Generation of ambient temperature hourly time series for some Spanish locations by artificial neural networks," Renewable Energy, vol. 51, pp. 285-291, 2013.

[13] K. Abhishek, M.P. Singh, S. Ghosh, and A. Abhishek, "Weather forecasting model using Artificial Neural Network" Procedia Technology, vol. 2, pp. 311-318, 2012.

[14] H. Saima, J. Jaafar, S. Belhaouari, T. Perak, and T.A. Jillani, "Intelligent methods for weather forecasting: A review," In Proceedings of the National Postgraduate Conference, (Tronoh Perak, Malaysia) Sept. 19-20, 2011.

[15] R. Aggarwal and R. Kumar, "A comprehensive review in weather prediction models," Int. J. of Computer Applications, vol. 74, no, 18, pp. 44-48, 2013.

[16] N.O. Bushara and A. Abraham, "Computational Intelligence in Weather Forecasting: A review," Journal of Network and Innovative Computing, vol. 1, pp. 320-331, 2013.

[17] P.A. Gutiérrez, S. Salcedo-Sanz, C. Hervás-Martínez, L. Carro-Calvo, J. Sánchez-Monedero, and L. Prieto, "Ordinal and nominal classification of wind speed from synoptic pressure patterns," Eng. Applications of Artificial Intelligence, vol. 26, no. 3, pp. 1008-1015, 2013.

[18] C. Wen-Yeau, "A literature review of wind forecastin methods," Journal of Power and Energy Engineering, vol. 2, pp. 161-168, 2014.

[19] C. Potter, M. Ringrose and M. Negnevinsky, "Short-term wind forecasting techniques for power generation," Australian Universities Power Engineering Conference (AUPEC), 26-29 September, 2004.

[20] A.M. Foley, P.G. Leahy, A. Marvuglia, and E.J. McKeogh, "Current Methods and Advances in Forecasting of Wind Power Generation" Renewable Energy, vol. 37, pp. 1-8, 2012.

[21] N.P. Vapnik, The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.

[22] W. Chu and S.S. Keerthi, "New approaches to support vector ordinal regression," in Proceedings of International Conference on Machine Learning (ICML), pp. 145-152, 2005.

[23] P. Karvelis, J. Spilka, G. Georgoulas, V. Chudáček, C.D. Stylios, and L. Lhotská, "Combining latent class analysis labeling with multiclass approach for fetal heart rate categorization," Physiological measurement, vol. 36, no. 5, pp. 1001-1024, 2015.

[24] A.P. Gutierrez, M. Perez-Ortiz, J. Sanchez-Monedero, F. Fernandez-Navarro, and C. Hervas-Martınez, "Ordinal regression methods: survey and experimental study," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 1, pp. 127-146, 2016.

[25] S. Janitza, G. Tutz, and A.L. Boulesteix, "Random forest for ordinal responses: Prediction and variable selection," Computational Statistics & Data Analysis, vol. 96, pp.57-73, 2016.

[26] J. Sánchez-Monedero, P. Gutiérrez, P. Tiňo, and C. Hervás-Martínez, "Exploitation of pairwise class distances for ordinal classification," Neural computation, vol. 25, no. 9, pp. 2450-2485, 2013.

[27] T. Hastie, R. Tibshirani, and J. Friedman, The elements of statistical learning, New York: Springer, 2009.

[28] J.R. Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.

[29] J.A.K. Suykens, J. Vandewalle, "Least Squares Support Vector Machine Classifiers," Neural Processing Letters, vol. 9, pp. 293-300, 1999.

[30] P. McCullagh, "Regression models for ordinal data," Journal of the royal statistical society," Series B (Methodological), pp. 109-142, 1980.

[31] E. Frank, and M. Hall, "A simple approach to ordinal classification", Lecture Notes in Computer Science, vol. 2167, pp. 145-156, 2001.

[32] W. Chu, and S.S. Keerthi, "Support vector ordinal regression," Neural Computation, vol. 19, no. 3, pp. 792-815, 2007.

[33] A. Smola, and B. Scholkopf, "A tutorial on support vector regression," Statistics and Computing, 14, 3, 199–222, 2004

[34] H.H. Tu, and H.T. Lin, "One-sided support vector regression for multiclass cost-sensitive classification," in Proceedings of the Twenty-Seventh International Conference on Machine learning (ICML2010), pp. 49–56, 2010.

[35] J. Verwaeren, W. Waegeman, and B. Baets, "Learning partial ordinal class memberships with kernel-based proportional odds models," Computational Statistics & Data Analysis, vol. 56, no. 4, pp. 928–942, 2012.

[36] J.E. Oliver, Encyclopedia of world climatology. Springer, 2005.

[37] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, pp. 321-357, 2002.

[38] Q. Hu, W. Pan, Y. Song, and D. Yu, D. 2012. "Large-margin feature selection for monotonic classification," Knowledge-Based Systems, vol. 31, pp. 8-18, 2012.