Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2019) 27: 1461 – 1476

doi:10.3906/elk-1803-25

A hybrid model for the prediction of aluminum foil output thickness in cold rolling process

Ali ÖZTÜRK[1,*], **Rifat ŞEHERLİ**[2]

[1]Department of Computer Engineering, Faculty of Engineering, Karatay University, Konya, Turkey
[2]Panda Aluminium Co., Kahramankazan, Ankara, Turkey

Received: 05.03.2018 • **Accepted/Published Online:** 28.01.2019 • **Final Version:** 22.03.2019

Abstract: This study proposes a hybrid model composed of multiple prediction algorithms and an autoregressive moving average (ARMA) module for the thickness prediction. In order to attain higher accuracy, the prediction algorithms were globally combined by simple voting to reduce the effect of the inductive bias imposed by each algorithm on the dataset. The global multiexpert combination (GMEC) system included the multilayer perceptron neural network (MLPNN), radial basis function network (RBFN), multiple linear regression (MLR), and support vector machines (SVM) algorithms. The proposed hybrid model extends the GMEC system by integrating an ARMA module for the output. On the test dataset, the mean absolute error (MEA) and root mean squared error (RMSE) were better for the hybrid model than the GMEC system. The GMEC system had approximately twice better performance than the MLPNN, which was the best among the learners. The performance was significantly improved via the hybrid model in terms of correlation coefficient (R). The results suggested that the proposed hybrid model can be used for more accurate and precise prediction of aluminum foil output thickness.

Key words: Prediction, global expert combination, autoregressive moving average, aluminum foil

1. Introduction

Aluminum foils are widely used in aeronautics and astronautics, transportation, packaging and printing, industrial robots, energy engineering, ocean exploitation, etc. [1] due to their water-, air-, and light-proof features. The thinner the foil, the more expensive it is, and it can be used in more specific application areas. Furthermore, it is an obligation by standards to use aluminum foil of 6.35 micron thickness in cigarette packages and liquid product packages. The quality of the whole process, beginning from foundry to the rolling, is very important to achieve the desired thickness of the aluminum foil [2] and improving the precision of products is always a hot research topic in the field. In the literature, data-driven modeling and control algorithms were proposed or investigated to control and predict the thickness of the metal strip in the rolling process. For example, a multiple support vector machine (MSVM) modeling approach was proposed by Liang et al. [3] to build a strip thickness model in a hot rolling automatic gauge system. Hydraulic automatic gauge control (HAGC) system identification and control were achieved by Mei et al. [4] based on radial basis function and backpropagation neural networks in an automatic position control system. Marcellos et al. [5] proposed a method to obtain the torque and rolling mechanical power estimates in real time without utilization of lamination process models. The power estimates were obtained from stator electrical variables, which were readily available in AC drives.

*Correspondence: ali.ozturk@karatay.edu.tr

The determination of input and output strip thickness from torque and rolling mechanical power estimates was investigated. A method based on the least squares (LS) method was introduced by Zheng and Qu [6] to analyze and calculate the characteristic parameters of periodic export thickness error of the rolling mill, and a new solution was proposed to control the rolling mill and to realize the compensation of the export thickness error. Another study [7] used MLPNN for the prediction of aluminum foil output thickness using the armature and field currents of the cold rolling machine. An extreme learning machine (ELM) and clustering forecast method were applied by Wang et al. [8] for the rolling strip thickness prediction. An application of SVM regression for HAGC for steel strip thickness prediction was proposed by Wei et al. [9]. They modeled the rolling gap with SVM regression for prediction. Varol et al. [10] applied the MLPNN for the thickness prediction of coatings fabricated by mechanical milling. A two layer feedforward network with sigmoid activation function was used. The short-term prediction of aluminum strip thickness with SVM was investigated by Öztürk and Şeherli [11] with various kernel functions. In a recent study by Öztürk and Şeherli [12], the nonlinearity analysis of thickness time series was made and many steps ahead short-term prediction via global regressor combination was achieved with significant accuracy.

The precision of the thickness prediction plays an important role in aluminum foil production. Therefore, any improvement in the accuracy of prediction is crucial. The method proposed in this study uses the armature and field current parameters of the cold rolling machine motors for instantaneous and precise prediction of the aluminum foil output thickness. Various machine learning algorithms were compared in terms of thickness prediction accuracy. After that, they were globally combined in order to improve the results by selecting the output of the learner that had the minimum absolute error for the current input. An ARMA module was used in the output of the GMEC system to obtain a hybrid model in order to further improve the accuracy. The experimental results indicated that the hybrid model had significantly higher prediction performance, especially for the nonstationary and irregular parts of the data.

2. The cold rolling process

The cold rolling machine used in this study is equipped with several instruments, actuators, and control devices, as well as hydraulics and oiling systems. The cold rolling machine is formed by various rolls, which are driven by direct current (DC) motors [13]. In Figure 1, the schematic diagram of the cold rolling machine is given. The six main motor types in the machine are rolling mill motor, front coiling motor, back coiling motor, first front uncoiling motor, first back uncoiling motor, and second uncoiling motor. These motors drive the tension mills and the rolling mills, which are shown in Figure 1.

The rotation of the DC motors is controlled by armature and field currents. These currents are supplied by the analog-to-digital converter (ADC) drivers, which are controlled by a programmable logic controller (PLC). As the rotation speed of the motors is under control, the stable tension of the foil under the rollers is protected [14].

The voltage equation for the armature circuit in a DC motor is

$$V = E + I_A R_A, \tag{1}$$

where I_A is armature current, R_A is armature resistance, and E is (in unit of voltage)

$$E = K \varnothing \Omega, \tag{2}$$

where K is a constant, ø is field flux (proportional to field current), and Ω is the speed of the motor (radian/s). The generated torque (T) value is (in Nm)

$$T = K\phi I_A. \tag{3}$$

The angular speed of the motor is handled in units of radian/s as in the following, by combining Eq. (1) and Eq. (2):

$$\Omega = \frac{V - I_A R_A}{K\phi}. \tag{4}$$

According to Eq. (4), the speed of the motor changes directly proportional to armature current and inversely proportional to field current.

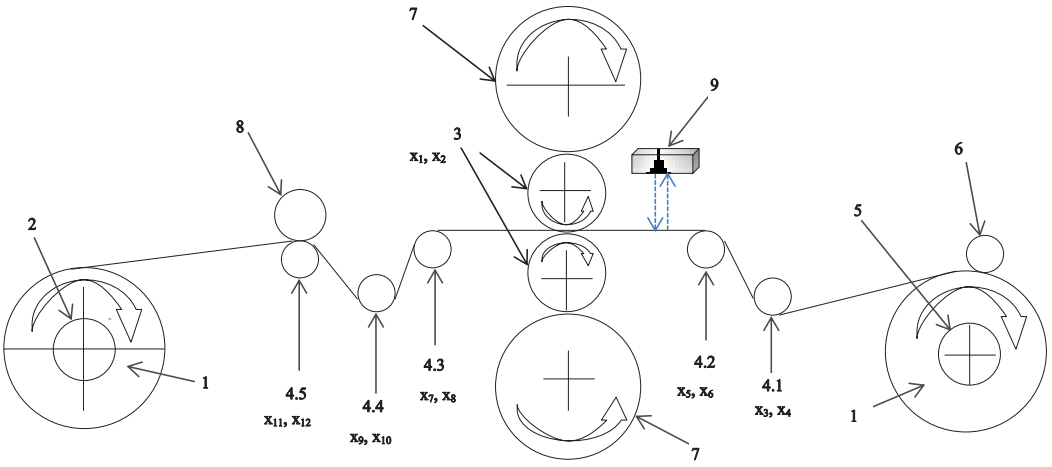

Figure 1. The schematic appearance of the cold rolling machine. 1: Aluminum folio roll, 2: uncoiling mill, 3: rolling mills, 4: tension mills (from 4.1 to 4.5), 5: coiling mill, 6: pressure mill, 7: supporting mills, 8: disk shears, 9: X-ray thickness measurement device.

During the rolling process, in order to keep the rolling speed constant, the angular rotation speed of the uncoiling motors should decrease. Likewise, the angular rotation speed of the coiling motors should increase. Otherwise, the tension will rupture the foil. The regulation of the tension is handled using the changes of the coiling and uncoiling roll diameters. The principle is to keep the coiling motor power P_{motor} constant during the rolling process according to the formula $P_{motor} = M \times W$. Here, M is the static moment of the roller, defined as $M = T \times (D/2)$, where D is the roll diameter and T is the tension. The angular speed of the roller is $W = (2 \times \pi \times n)/60$. Here, n is the rotations per minute (rpm) of the roller, defined as $n = (60 \times V_s)/(\pi \times D)$, where V_s is the speed of the roller. After simplification, W becomes $W = (2 \times V_s)/D$ and so P_{motor} becomes $P_{motor} = T \times V_s$, which means the electrical power on the coiling motor is the multiplication of tension by the speed of the roller.

The PLC pane used in this study included 1 Siemens 6ES7 405-10A01-0AA0 PLC, 1 Siemens 6ES7 952 1KK00-0AA0 flash memory, 1 Siemens 6ES7 414-2XK05-0AB0 CPU, and 1 Siemens 6GK7 443-1EX20-0XE0 Ethernet card. The communication with Siemens drivers was provided by the PROFIBUS protocol. ADC input/output (I/O) cards were used on the Siemens ET 200S I/O module for analog-to-digital I/O operations. A Siemens SIMOREG 6RA7075-6DV62-0 device was used as ADC on all driver panes.

3. Data collection

The armature and field current values for all motors of the rolling machine and the output thickness of the aluminum foil were continuously recorded on a PC. The output thickness of the aluminum foil was measured via X-ray device while the foil was passing through the rolling mill. The measurement range, measurement interval, and measurement accuracy were 0.002–0.2 mm, 100 mm, and 99.95% for the X-ray device, respectively. The flow rate of the foil was 550 meters per minute during the rolling process of the 6.35 micron aluminum foil. The surface roughness, i.e. profile roughness, of the rollers was 10–12 R_a, where R_a is a quantitative calculation of the relative roughness of the surface. Since there are 6 tension mills in the cold rolling machine, there are 12 input features in the dataset. The variables $x_{i=1..12}$ in Figure 1 correspond to the armature and field current values that drive the tension mills. These values constitute the input features and the thickness value is the output feature. In Table 1, the correspondence between the input variables $x_{i=1..12}$ and the armature and field current parameters of the DC motors is given.

Table 1. The correspondence between input variables and currents of the DC motors.

Variable	Physical parameter
x_1	Armature current for rolling mills motor
x_2	Field current for rolling mills motor
x_3	Armature current for front coiling motor
x_4	Field current for front coiling motor
x_5	Armature current for back coiling motor
x_6	Field current for back coiling motor
x_7	Armature current for first front uncoiling motor
x_8	Field current for first front uncoiling motor
x_9	Armature current for first back uncoiling motor
x_{10}	Field current for first back uncoiling motor
x_{11}	Armature current for second uncoiling motor
x_{12}	Field current for second uncoiling motor

The data were continuously recorded every second on the PC during production if it was not set otherwise. The data used for this study include 7250 instances that correspond to approximately 121 min of recording. The portions where rupture occurred were excluded from the recorded thickness data beforehand.

4. Prediction algorithms

The output thickness prediction performances of various algorithms were compared against real output thickness data. The algorithms used in this study were MLPNN, MLR, SVM, and RBFN.

MLPNN with backpropagation is a feedforward network where modification on the weights is based on the difference between the computed and actual values of the output nodes. The main idea is to minimize the mean squared error between the actual and computed output values in an iterative manner [15]. The stopping criterion is the reduction of the total network error to a predefined level. The error is computed as $E = \frac{1}{2}\sum_j (x_o - x_t)^2$, where x_o is the output value and x_t is the actual (target) value. The output value for neuron x_j is calculated as $X_j = \theta\left(\sum_i^n x_i w_{ij}\right)$, where θ is the transfer function and w_{ij} is the weight between the input neurons x_i and the output neuron x_j. θ is the sigmoid function and defined as $\theta(x) = 1/(1+e^{-x})$.

The backpropagation algorithm uses gradient descent to move the weight vectors in order to decrease error E through the error surface to the global minima. The weights are updated according to the following formula:

$$\Delta W_{ji}(n+1) = \eta \delta_{pj} O_{pi} + \alpha \Delta W_{ji}(n), \tag{5}$$

where η is the learning rate, α is the momentum coefficient, and δ_{pj} is the error value for the neuron on the Lth layer.

δ_{pj} is calculated differently for output layer neurons and hidden layer neurons. For the output layer neurons $\delta_{pj} = \left(O_{t_{pj}} - O_{pj}\right) O_{pj}\left(1 - O_{pj}\right)$ and for the hidden layer neurons $\delta_{pj} = O_{pj}\left(1 - O_{pj}\right) \sum_k \delta_{pk} w_{kj}$.

The MLPNN with one hidden layer is capable of approximating uniformly any continuous multivariate function to any desired degree of accuracy. This implies that any failure of a function mapping by a multilayer network must arise from inadequate choice of parameters or an insufficient number of hidden nodes [16].

Unlike a simple linear regression model where a single response measurement Y is related to a single predictor X for each observation, multiple linear regression (MLR) is a statistical model that determines a mathematical relationship between several independent variables and a dependent variable by fitting a linear mathematical equation to the observed data [17] as in Eq. (6):

$$E(Y|X) = \alpha + \beta_1 x_1 + \ldots + \beta_n x_n, \tag{6}$$

where α is called the intercept and β_{j} are coefficients.

MLR is less complicated than nonlinear prediction methods, demands less computational power, and requires fewer parameters prior to the application. In this study, the M5 algorithm was used as the attribute selection method in MLR implementation. This algorithm builds trees whose leaves are associated to multivariate linear models and the nodes of the tree are chosen over the attribute that maximizes the expected error reduction as a function of the standard deviation of output parameter. M5 removes the attribute with minimum standardized coefficient at each step until no improvement is observed in Akaike information criterion (AIC) error estimation [18].

The support vector machine (SVM) is based on statistical learning theory [19] and the basic idea is to map the input dataset X into a high dimensional feature space F via a nonlinear mapping function by constructing an optimal hyperplane in this new space [20]. When the SVM is used for regression, a hyperplane is constructed that lies close to as many points as possible. In the general form, the estimation function is $f(x) = (w \times \varnothing(x)) + b$, where w and b are the coefficients to be estimated from data, and $\varnothing(x)$ is the nonlinear function in feature space.

The risk function to be minimized is

$$R(w, \varrho^*) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N} (\varrho_i + \varrho_i^*), \tag{7}$$

and $d_i - w\varnothing(x_i) - b_i \leq \varepsilon + \varrho_i$, $(w\varnothing(x)) + b - d_i \leq \varepsilon + \varrho_i^*$ where $\varrho_i, \varrho_i^* > 0$.

The support vector regression estimation function is

$$f(x) = \sum_{i=1}^{NSV} (\alpha_i - \alpha_i^*) K(X, X_i) + b, \tag{8}$$

where α_i and α_i^* are Lagrange coefficients that force the predictions to move towards the target value d and NSV is the number of support vectors. The kernel function $K(X_i, X_j) = \varnothing(X_i)\varnothing(X_j)$ is any function that satisfies Mercer's theorem. In this study, the Pearson VII function-based kernel (PUK) proposed by Üstün et al. [21] was used because it was found more successful than the other two kernel functions (polynomial kernel $[1 + (XX_i)]^p$ and RBF kernel $\left(1/2\sigma^2\|X - X_i\|^2\right)$, respectively). The C, ε , and kernel parameters p and σ^2 are specified by the user. The complexity parameter C is used for controlling the smoothness of the approximating function and ε determines the margin in which the round-off error is tolerated [22].

RBF networks are feedforward neural networks that have a single hidden layer of nonlinear units whose activation functions are Gaussian or some other basis kernel function. They are trained with a supervised training algorithm, but much faster than backpropagation networks. Because of the radial basis function hidden units, they are less susceptible to nonstationary inputs [23]. The center of the basis function for a node i at the hidden layer is a vector c_i whose size is as the input vector u and normally there is a different center for each hidden unit. First, the radial distance d_i between the input vector u and the center of the basis function c_i is computed using the Euclidean distance for each hidden unit i as $d_i = \|u - c_i\|$. The output h_i of each hidden unit i is then computed by applying the basis function G to this distance as $h_i = G(d_i, \sigma_i)$. Then the output for the jth unit in the output layer is computed as

$$o_j = f_j(u) = w_{0j} + \sum_{i=1}^{L} w_{ij} h_i, \ j = 1, 2, \ldots, M. \tag{9}$$

The training of the RBF network is formulated as a nonlinear unconstrained optimization problem. Given input output training patterns (u^k, y^k), k = 1,2,... ,K, $w_{\mathrm{i,j}}$, and c_{i} = 1,2,... ,L, j = 1,2,..., M, the aim is to minimize $J(w, c) = \sum_{k=1}^{K} \|y^k - f\left(u^k\right)\|^2$.

5. The hybrid model

According to the no-free-lunch theorem, no single algorithm in a domain always induces the most accurate learner [15]. The inductive bias of each prediction algorithm leads to error if its assumptions do not hold for the data. Furthermore, each algorithm normally converges to a different solution and can fail under different circumstances. By combining multiple prediction algorithms suitably, accuracy can be improved.

The multiexpert combination method used in this study assumes that the prediction algorithms work in parallel. Given an input, each algorithm generates an output, and a linear combination of these outputs is taken by simple voting. The main idea is the fusion of the prediction algorithms in order to attain a higher accuracy level. The accuracy level will certainly be better than the most accurate prediction algorithm, because the method chooses the best prediction with minimum error for each instance in the test dataset found by each algorithm individually. In Figure 2, the global multiexpert combination (GMEC) model with simple voting is shown. For each input $\mathrm{x_{i,j=1..12}}$, the global function $f(\cdot)$ linearly combines the outputs of the prediction algorithms by selecting the one with minimum absolute error (min AE) and generates the output $\mathrm{y_i}$. Let $MLPNN_{AE} = |MLPNN_{x_{t+d\tau}} - x_{t+d\tau}|$, $SVM_{AE} = |SVM_{x_{t+d\tau}} - x_{t+d\tau}|$, $RBF_{AE} = |RBF_{x_{t+d\tau}} - x_{t+d\tau}|$, and $MLR_{AE} = |MLR_{x_{t+d\tau}} - x_{t+d\tau}|$, and then the $f(\cdot)$ selects the output of the regressor with minimum AE as $y_{t+d\tau}$.

The proposed hybrid model, which uses an ARMA model on the output of GMEC model, is shown in Figure 3. The ARMA models are an adaptation of discrete-time filtering methods developed by Wiener [24].

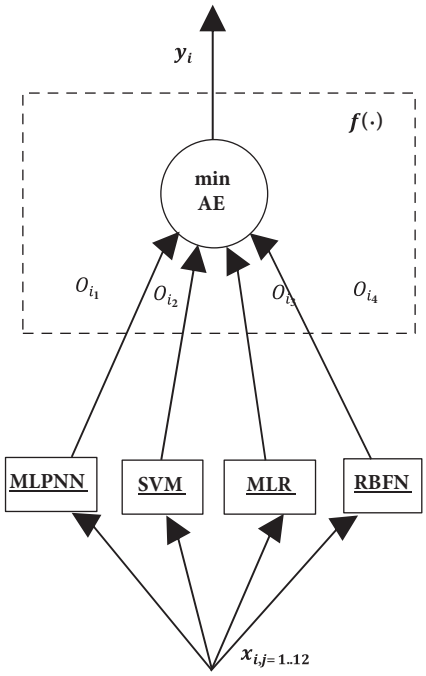

Figure 2. The GMEC model with simple voting.

They were applied to time series data obtained from various application areas by Box et al. [25]. In its compact form, the ARMA model is

$$A(q)\,y(t) = C(q)\,e(t), \tag{10}$$

where $y(t)$ is output at time t, $A(q)$ and $C(q)$ are polynomials given in Eq. (11) and Eq. (12), and $e(t)$ is the white-noise disturbance value:

$$A(q) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}, \tag{11}$$

$$C(q) = 1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}, \tag{12}$$

where n_a is the number of poles in the autoregressive part of the ARMA model and n_c is the number of C coefficients, which specifies the order of the moving-average part of the ARMA model. The output $y(t)$ depends on the previous outputs.

The flow chart of the hybrid model is given in Figure 4. After obtaining the GMEC model by globally combining the output of the learners, the orders of the ARMA model are found via partial autocorrelation function (PACF) and autocorrelation function (ACF) using error values of GMEC. The PACF and ACF are used to identify the order of the AR process and MA process, respectively. Then the hybrid model is used for the instantaneous prediction of the aluminum foil thickness using the new armature and field currents.

For nonstationary time series the terms of the ACF do not decay to zero, while they decay exponentially to zero for stationary time series. On the other hand, the partial autocorrelation of an AR(p) process is close to zero at lag p+1 and greater.

The previous inputs for the ARMA model are $y(t-1)\cdots y(t-p)$, which correspond to the difference values (error values) between actual and predicted thickness data of the global multiexpert system. Therefore,

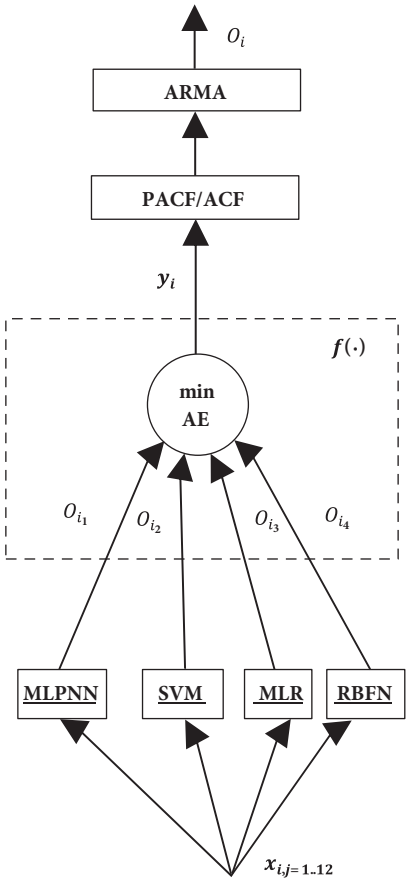

Figure 3. The proposed hybrid model for prediction.

the ARMA model is obtained by the error values produced by the global multiexpert system during each training session. The predicted error outputs of the ARMA model are generated by using the error values of the test data in each fold. These error values are subtracted from the estimated prediction values of the global expert system to generate the final thickness prediction of the hybrid system.

6. Experimental results and discussion

The Weka [26] package libraries, C# programming language, and MATLAB were used for the implementation of the hybrid model. The libraries of Weka were used for the application of the prediction algorithms and MATLAB was used for the implementation of ARMA.

For MLPNN implementation, several different alternative hidden layer neuron numbers between 5 and 30 were tried experimentally along with different learning rates (between 0.01 and 0.5) and momentum values (between 0.1 and 0.9) in order to find the most accurate thickness prediction. The MLPNN used in this study was a 3-layer feedforward backpropagation neural network with 8 hidden neurons. In the input layer, there were 12 input nodes, which correspond to twelve armature and field current values. The predicted thickness value was obtained in the single output node. The learning rate and momentum values that gave the best accuracy were found as 0.01 and 0.2, respectively. The network was trained with 1000 steps in each fold. More training steps did not improve the accuracy significantly.

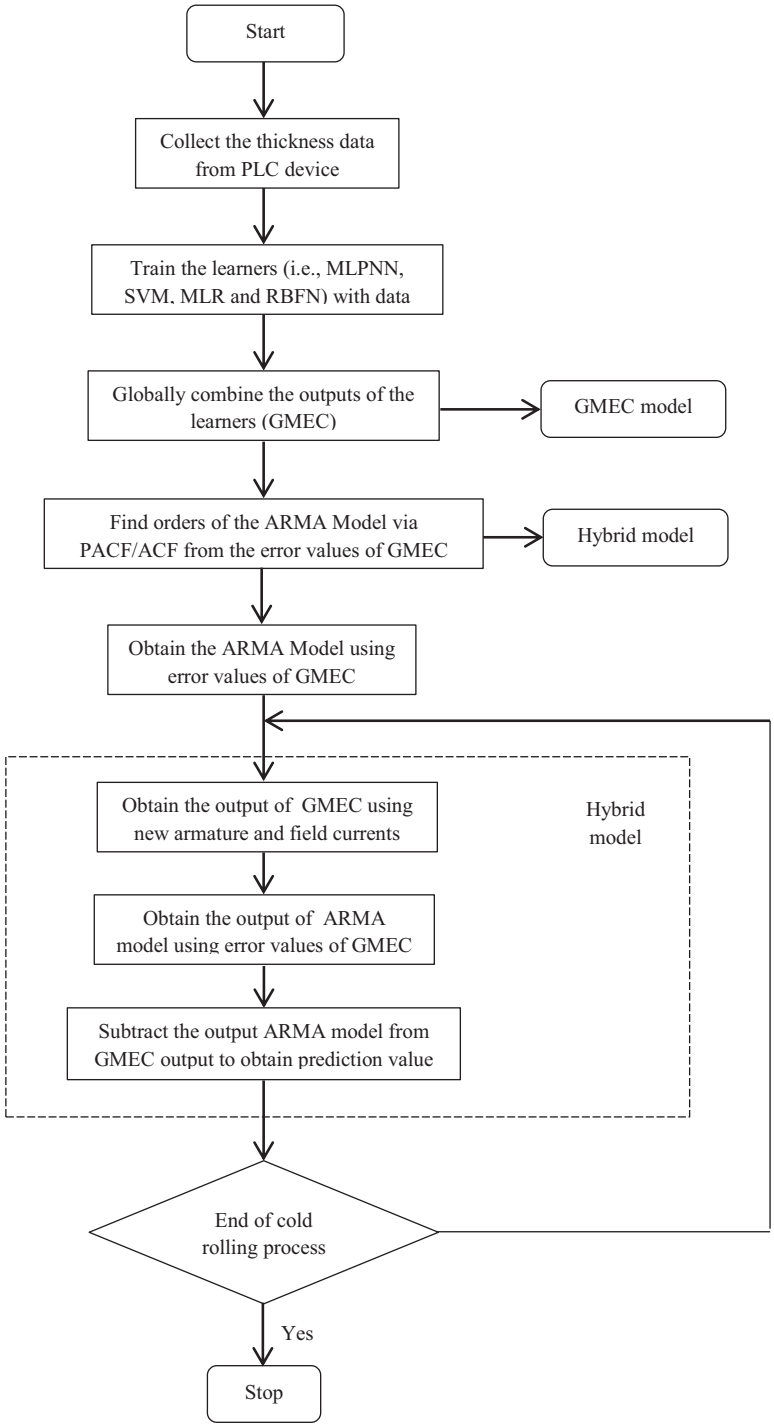

Figure 4. The flow chart of the hybrid model.

The activation function of RBF neurons in the implemented RBF network is $\varphi(x) = e^{-\beta\|x-\mu\|^2}$, where x is the input vector of armature and field current values, μ is the center of the Gaussian kernel function, β is the width of the bell curve, and $\|.\|$ is the Euclidean distance. The centers of the Gaussian kernel functions were found with k-means clustering and the weights of the output layer were obtained by linear regression.

Several different neuron numbers (between 2 and 100) and standard deviation values (between 0.01 and 0.5) of the Gaussian kernel function were tested to find the best accuracy for the RBF network. The architecture with 12 neurons with standard deviation of 0.1 gave the most accurate prediction values.

In the SVM implementation, the PUK function was found more successful than the other kernel functions (i.e. polynomial and RBF). The polynomial kernel function gave better accuracy than the RBF kernel function. The exponent in the polynomial kernel function was found to be p = 2, which gave the best accuracy. Increasing this exponent did not improve the accuracy. Fine-grain values (e.g., 0.7, 0.5, 0.3) and coarse-grain values (e.g., 5, 10) other than 1 were applied to complexity parameter C. However, the accuracy did not improve. Likewise, different values were given for ε , such as 0.01, 0.1, and 1, other than 0.001. Higher values of ε decreased the accuracy of the SVM model.

In the MLR implementation, the relationship between input vector X and output thickness value Y was defined as $E(Y|X) = \alpha + \beta_1 x_1 + \ldots + \beta_n x_n$, where α is called the intercept and β_j are coefficients. The greedy algorithm and M5 algorithm were compared for the attribute selection in the MLR model. The greedy algorithm selected fewer attributes; however, the accuracy was not affected significantly.

It takes 218.13 s, 113.45 s, and 52.35 s on average for building the SVM model with PUK, poly kernel, and RBF kernel, respectively. Testing times for these kernels are 0.42 s, 0.01 s, and 0.53 s, respectively. SVM with PUK takes the longest to be built; however, its accuracy is twice better than the other algorithms. It takes 15.2 s to build the MLPNN model and the testing time is very close to 0 on average, which means testing is very fast. The MLR model building time is 0.02 s and testing time is 0.01 s on average. For the RBF network, the model building time is 0.13 s and testing time is 0.01 s on average.

The k-fold validation method avoids overfitting and is a standard method used to find out-of-sample prediction error. The prediction algorithms were evaluated using leave-one-out 10-fold cross-validation. Therefore, the whole dataset of 7250 instances was divided into a training set of 6525 instances and testing set of 725 instances for each fold. In each fold, a sliding window of 6525 data point size is used to select the training set in the whole dataset in a circular manner. Thus, each data instance had the chance to be tested in a specific fold.

The steps for the evaluation of the hybrid model in aluminum foil thickness prediction are given below.

Step 1. Collect the thickness data from the PLC device.

Step 2. Split the data into 10 different training and validation subsets for 10-fold cross-validation.

Using the corresponding training subset in each fold, do steps 3–6.

Step 3. Train the learners (e.g., MLPNN, SVM, MLR, and RBFN).

Step 4. Globally combine the output of the learners with simple voting depending on the MAE criterion.

Step 5. Apply ACF and PACF to the error values of GMEC to find the appropriate orders for the AR and MA processes of the ARMA model.

Step 6. Obtain the ARMA model using the error values of GMEC.

Using the corresponding validation subset in each fold, do steps 7–9.

Step 7. Obtain the output of GMEC.

Step 8. Obtain the output of the ARMA model using the error values of GMEC.

Step 9. Subtract the ARMA model output from GMEC predicted values to obtain the final prediction results.

In Figure 5, the thickness values of aluminum foil are shown for 7250 s of recording during the cold rolling process.

The PACF was close to zero at lag 4 for the test data, which is shown in Figure 6 as an example. Therefore, the order 4 was used for the AR process to produce the ARMA model for this portion of test data.

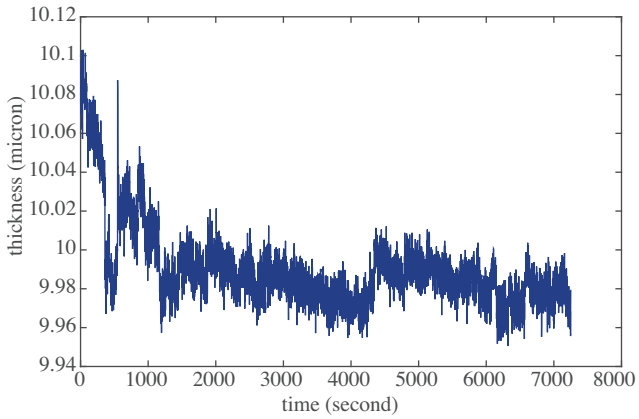

Figure 5. The thickness values of aluminum foil recorded during 7250 s of production.

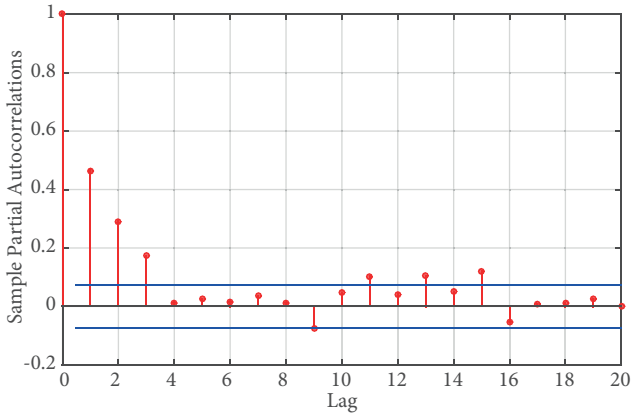

Figure 6. The PACF plot of thickness test data.

The ACF was close to zero at lag 9 for the test data, which is shown in Figure 7 as an example. Therefore, the order of 9 was used for the MA process to produce the ARMA model for this portion of test data.

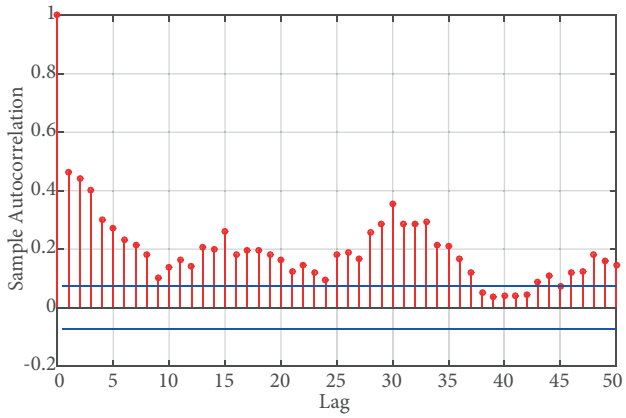

Figure 7. The ACF plot of thickness test data.

The mean absolute error (MAE), root mean squared error (RMSE), and correlation coefficient (R) equations used for the evaluation of the prediction algorithms are given as follows:

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|A_j - P_j|, \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(A_j - P_j)^2}, \tag{14}$$

$$R = \frac{\sum_{j=1}^{n}(A_j - A)(Pj - P)}{\sqrt{\sum_{j=1}^{n}(A_j - A)^2 \sum_{j=1}^{n}(P_j - P)^2}}, \tag{15}$$

where A_j and P_j are the actual and the predicted values, and A and P are the mean of actual and predicted values, respectively. The average MAE, RMSE, and R values obtained by using the prediction results of each individual fold are given in Table 2. The R values for GMEC and the hybrid model are significantly higher than those of the other prediction algorithms. The hybrid model has higher R values on average than GMEC, which means a better overall correlation. According to the results, MLPNN has the best accuracy among the other traditional prediction algorithms in terms of MAE and RMSE. R is also the highest for MLPNN with the value of 0.3181. However, the accuracy of the GMEC system is approximately twice better than MLPNN in terms of MAE and RMSE. The R value is significantly higher than MLPNN for the GMEC system. The hybrid model further improves the overall accuracy. These accuracy improvements are very important for the precision of the thickness prediction, especially when there is nonstationary behavior in the data.

Table 2. Average MAE, RMSE, and R values for the prediction algorithms, the GMEC system, and the proposed hybrid model for the test data.

Methods	MAE	RMSE	R
MLPNN	0.0116	0.0137	0.3181
SVM (PUK)	0.0118	0.0145	0.1804
RBF network	0.0154	0.0181	0.0972
MLR	0.0139	0.0170	0.0927
GMEC	0.0058	0.0080	0.8232
Hybrid model	0.0040	0.0052	0.8477

The overall accuracy was significantly improved by means of the hybrid model, especially for the portions of test data where the data exhibit nonstationary behavior. The comparison of MAE and RMSE values for the GMEC and hybrid model is given in Table 3 for all of the folds. For folds 1 and 10, the MAE values are above 0.01, and for fold 9 it is very close to 0.01. For folds 1, 9, and 10, the RMSE values of the GMEC model are above 0.01. The test data used in these 3 folds correspond to the nonstationary part of the thickness data. The error values above or close to 0.01 are not acceptable for 6.35 micron aluminum foil thickness prediction. The hybrid model results are obviously below 0.01 for all folds.

Table 3. MAE and RMSE values for GMEC and hybrid model for the test data.

Methods	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10
GMEC MAE	**0.0126**	0.0064	0.0029	0.0023	0.0041	0.0025	0.0051	0.0036	**0.0094**	**0.0212**
Hybrid MAE	**0.0049**	0.0044	0.0028	0.0026	0.0039	0.0026	0.0042	0.0034	**0.0055**	**0.0063**
GMEC RMSE	**0.0123**	0.0087	0.0043	0.0032	0.0055	0.0037	0.0071	0.0053	**0.0124**	**0.0286**
Hybrid RMSE	**0.0052**	0.0056	0.0038	0.0033	0.0049	0.0036	0.0055	0.0046	**0.0072**	**0.0085**

The actual and predicted thickness values for one of the test sessions are given in Figure 8 as an example. In this figure, it can be observed how the prediction accuracy improves in the case of the hybrid model. The prediction performance of the hybrid model is very obvious, where the nonstationarity of the data is high. It can be observed from the figure how the predicted thickness values get closer to the actual values by means of the hybrid model. The smoothing effect of the ARMA module moves the prediction output of the global multiexpert system towards the mean.

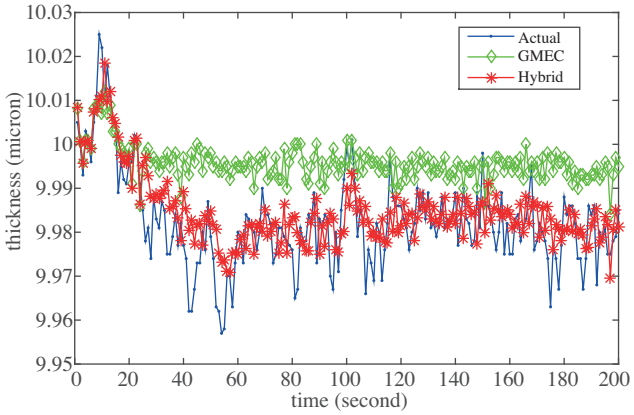

Figure 8. The actual and predicted thickness values for the test dataset.

The performances of the prediction algorithms as well as the ones proposed in this study are given in Figure 9 for an example portion of test data. As can be seen in Figure 9, the output of GMEC is mainly determined by the output of the SVM algorithm for this portion. The performance of RBF and MLPNN are nearly equally bad, while the MLR algorithm is the worst.

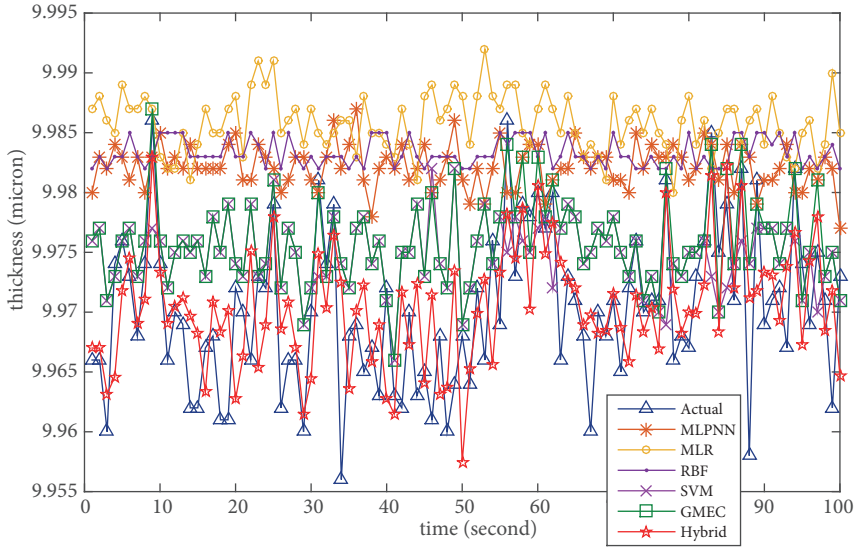

Figure 9. Comparison of the prediction performances of the algorithms for the test dataset.

The R value is the measure of how strongly the actual and and predicted thickness values are correlated. On the other hand, R^2 is a statistical measure of correlation determination of how close the actual data are to the regression model. R^2 is 0.7186 for the hybrid model, while it is 0.6777 for GMEC. A high value of R^2 alone does not necessarily indicate that the regression model has a good fit to the actual data. Furthermore, although the R^2 values of the hybrid model and GMEC are close, we need a formal hypothesis test to determine if they are statistically significant. In this study, we used the two-samples F-test to determine the statistical significance of the obtained results. The null hypothesis was that the actual and predicted thickness values come from normal distributions with the same variance. According to the test results, the P-value of the hybrid model was found as 0.0035. This means that if the alpha significance level is set small or equal to 0.0035 in

the F-test, the null hypothesis is not rejected. The SVM, RBF, and MLR algorithms had P-values of zero, which means that the null hypothesis is always rejected. The GMEC and MLPNN had P-values that were very close to zero, which means that even setting an alpha significance level was not meaningful. Therefore, we can conclude that only the hybrid model has a good fit to the actual thickness data.

Table 4. Mean and variance values of the actual thickness data and the output of the prediction algorithms.

Data	Mean	Variance
Actual thickness	9.9898	0.00045971
Hybrid model output	9.9900	0.00049231
GMEC output	9.9918	0.00069533
MLPNN output	9.9941	0.00087215
SVM output	9.9871	0.00017066
MLR output	9.9851	0.00013701
RBF output	9.9872	0.00007074

As can be seen from Table 4, the variances of actual thickness data and hybrid model output are very close. The variance of GMEC output is 50% over the actual thickness data variance.

7. Conclusions and future work

The hybrid model proposed in this study can be used to predict the thickness of the aluminum foil in an acceptable range (with average MAE of 0.0040, average RMSE of 0.0052, and R of 0.8477) using the armature and field current values of the cold rolling machine motors. These prediction values can be used by the operators of the rolling machine in order to control and audit the production process. The prediction outputs can be used for diagnostic cross-checks to make sure that the PLC is providing the armature and field currents to the rolling machine motors correctly and on time, by comparing them with the real thickness measures of the X-ray device. Since one of the principal purposes of this study is to predict the thickness of 6.35 micron aluminum foil, the deviation in the prediction value should not exceed ± 0.0099. Therefore, any improvement in MAE, RMSE, or R values is important for the reliability of the thickness prediction. The convergence time of SVM is much longer than the other algorithms, so this affects the total training time of the hybrid model negatively. In order to reduce the training time, a faster algorithm than SVM may be considered in the model. However, the prediction times of all of the algorithms are under 1 s. Therefore, since the sampling period is 1 s in the cold rolling process, the real-time application of the hybrid model is not affected. As a future work, we will focus on the prediction of aluminum foil ruptures before they occur. The ruptures are caused by the unexpected tensions produced by the coiling and uncoiling motors.

Acknowledgment

This study was supported by Republic of Turkey Ministry of Science, Industry, and Technology under the SANTEZ project "Implementation and Know-how Transfer of the Production System of 6.35 micron Aluminum Foil" with project code 01078.STZ.2011-2.

References

[1] Sun M, Wang T, Zhang W. Research on cascade predictive control in hydraulic AGC of cold rolling mill. In: Proceedings of the 2nd IEEE Conference on Industrial Electronics and Applications; 23–25 May 2007; Harbin, China. New York, NY, USA: IEEE. pp. 2775-2780.

[2] Zarate LE. A predictive thickness control structure and decision about the better control parameter for the cold rolling process through sensitivity factors via neural networks. In: IEEE Mid-Summer Workshop on Soft Computing in Industrial Applications Conference; 28–30 June 2005; Espoo, Finland. New York, NY, USA: IEEE. pp. 17-23.

[3] Liang H, Tong C, Peng K. Data-driven modeling and online algorithm for hot rolling process. In: Proceedings of the 30th Chinese Control Conference; 22–24 July 2011; Yantai, China. New York, NY, USA: IEEE. pp. 1560-1564.

[4] Mei J, Wang H, Xiao J, Shan M. Study on sampling control in hydraulic AGC based on neural network. In: Proceedings of the 6th International Conference on Machine Learning and Cybernetics; 19–22 August 2007; Hong Kong, China. New York, NY, USA: IEEE. pp. 504-508.

[5] Marcellos NS, Denti JF, Saousa GC. Strip thickness estimation in rolling mills from electrical variables in AC drives. Latin American Applied Research 2009; 39: 353-359.

[6] Zheng G, Qu Q. Research on periodical fluctuations identification and compensation control method for export thickness in rolling mill. In: Proceedings of the 8th International Conference on Machine Learning and Cybernetics; 12–15 July 2009; Hebei, China. New York, NY, USA: IEEE. pp. 1972-1977.

[7] Öztürk A, Çınar K. Output thickness prediction of aluminium foil using artificial neural network in cold rolling process. In: ELECO; 27–29 November 2014; Bursa, Turkey. pp. 389-392.

[8] Wang L, Fan L, Lu N, Cui X, Xie Y. Rolling thickness prediction based on extreme learning and clustering. In: International Conference on Computer Science and Mechanical Automation; 23–25 October 2015; Hangzhou, China. New York, NY, USA: IEEE. pp. 30-35.

[9] Wei L, Xiaolan Y, Lei Y, Yue G. Application of SVM regression in HAGC system. In: 27th Chinese Control and Decision Conference; 23–25 May 2015; Qingdao, China. New York, NY, USA: IEEE. pp. 3490-3494.

[10] Varol T, Çanakçı A, Özşahin S, Erdemir F, Özkaya S. Artificial neural network-based prediction technique for coating thickness in Fe-Al coatings fabricated by mechanical milling. Particulate Science and Technology 2018; 36: 742-750.

[11] Öztürk A, Şeherli R. Short-term prediction of aluminium strip thickness via support vector machines. In: IEEE 23rd Signal Processing and Communications Applications Conference; 16–19 May 2015; Malatya, Turkey. New York, NY, USA: IEEE. pp. 208-211.

[12] Öztürk A, Şeherli R. Nonlinear short-term prediction of aluminum foil thickness via global regressor combination. Applied Artificial Intelligence 2017; 31: 568-592.

[13] Huang Q, Liang A. High-precision Rolling Technologies. Beijing, China: Metallurgy Industry Press, 2004.

[14] Beddoes J, Bibbly MJ. Principles of Metal Manufacturing Process. Burlington, MA, USA: Elsevier Butterworth-Heinemann, 2003.

[15] Alpaydın E. Introduction to Machine Learning. 2nd ed. Cambridge, MA, USA: MIT Press, 2010.

[16] Hassoun MH. Fundamentals of Artificial Neural Networks. Cambridge, MA, USA: MIT Press, 1995.

[17] Ruppert D, Wand M. Multivariate locally weighted least squares regression. Annual Statistics 1994; 22: 1346–1370.

[18] Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. San Diego, CA, USA: Academic Press, 2000.

[19] Mohandes MA, Halawani TO. Support vector machines for wind speed prediction. Renewable Energy 2004; 29: 939-947.

[20] Osowski S, Garanty K. Forecasting of daily meteorological pollution using wavelets and support vector machine. Engineering Applications of Artificial Intelligence 2007; 20: 745-755.

[21] Üstün B, Melssen WJ, Buydens LMC. Facilitating the application of support vector regression by using a universal Pearson VII function-based kernel. Chemometrics and Intelligent Laboratory Systems 2006; 81: 29-20.

[22] Radhika Y, Shashi M. Atmospheric temperature prediction using support vector machines. International Journal of Computer Theory and Engineering 2009; 1: 1793-8201.

[23] Halıcı U. Artificial Neural Networks – Lecture Notes. Ankara, Turkey: Middle East Technical University, 2004.

[24] Wiener N. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. New York, NY, USA: Wiley, 1949.

[25] Box GEP, Jenkins GM, Reinsel GC. Time Series Analysis: Forecasting and Control. 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1994.

[26] Frank E, Hall MA, Witten IH. The WEKA Workbench Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques. 4th ed. San Francisco, CA, USA: Morgan Kaufmann, 2016.