Prediction of mechanical properties of wood fiber insulation boards as a function
of machine and process parameters by random forest

M. Schubert^a, M., Luković^a and H. Christen^b

^a*Empa, Swiss Federal Laboratories for Materials Science and Technology, Laboratory for
Cellulose & Wood Materials, Group WoodTec, Überlandstr. 129, 8600 Dübendorf, Switzerland*

^b*Pavatex SA, Knonauerstrasse 51, 6330 Cham, Switzerland*

Correspondence to: Dr. Mark Schubert

Phone: +41 58 765 76 24

E-mail: mark.schubert@empa.ch

This document is the accepted manuscript version of the following article:
Schubert, M., Luković, M., & Christen, H. (2020). Prediction of mechanical properties of wood
fiber insulation boards as a function of machine and process parameters by random forest. *Wood
Science and Technology*, 54(3), 703-713. <https://doi.org/10.1007/s00226-020-01184-3>

Abstract (Limited to 250 words)

In this case study, machine and process variables were extracted from the process control system (Prod-IQ) and combined with tested mechanical properties of wood fiber insulation boards according to product type and time of manufacture. The boards were taken from the production line (dry process) and the internal bond strength (σ_{mt}) and the compressive strength at 10% deformation (σ_{10}) were determined according to the European Standard EN 826 and 1607. The complete data set was preprocessed and split into training and test sets using k-fold cross-validation. The performance of the random forest algorithm (RF) was evaluated with the correlation coefficient (R), the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute percentage error (MAPE) and compared with artificial neural networks (ANN), and support vector machines (SVM). Forward feature selection was used to reduce input dimensionality and improve the generalizability of the algorithms. All machine learning algorithms predicted the mechanical properties with high accuracy, but the RF algorithm revealed the best generalization performance (σ_{mt} : $R=0.960$, $R^2=0.916$, RMSE=4.05, MAPE=12.11; σ_{10} : $R=0.981$, $R^2=0.963$, RMSE=17.19, MAPE=5.64). Our work demonstrates that machine learning can be applied to predict relevant properties of wood fiber boards for an improved quality control in real time.

Keywords: Artificial neural network, machine learning, mechanical properties, quality control, random forest, support vector machine, wood fiber boards

1. Introduction

The engineered wood industry is undergoing change in the form of corporate disposals and consolidations, rising costs of raw materials and energy and strong international competition (André et al. 2008). Maintaining the competitiveness of companies by reducing costs and optimizing product quality is therefore crucial for this industry. An accurate prediction of mechanical properties in real time could result in cost savings by reducing waste, customer complaints, downtime and wood consumption.

Thanks to the programmable logic controllers and the distributed supervisory control and data-acquisition systems present in modern process control systems, the manufacturing industry is experiencing an unprecedented increase in available data and the wood-based panel industry is no exception in this sense. Currently, one of the most pressing challenges is to gain a comprehensive understanding of the complexity of the process variables and to determine their influence on the quality characteristics of the end product. In this regard, the availability of quality-relevant data offers the potential to sustainably improve process and product quality (Kusiak 2017).

Today's data availability and the increase in computational power make machine learning techniques valid candidates to overcome such major challenges related to complex manufacturing systems (Escobar et al. 2018, Wuest et al. 2016). These data-driven approaches are able to find highly complex and non-linear patterns in data of different types and sources and automatically create a model that can be applied for prediction, detection, classification, regression, or forecasting (Wuest et al. 2016).

The use of machine learning could be particularly advantageous for the wood-based panel industry. In order to meet the quality criteria required for a specific wood-based product at optimum throughput, the process settings must be constantly adjusted because of the variability of wood and

specific requirements of the end product. In addition, at the time of production, the quality of the wood fiber boards is unknown, thus the quality control normally takes place after the production process as an event using destructive tests in the laboratory. The time span between testing the boards may be as long as several hours, during which a significant amount of production occurs. During this time, boards of low or even insufficient quality may be produced. These are serious problems for the engineered wood industry, where maintaining the competitiveness of companies depends heavily on reducing costs and optimizing product quality. The introduction of machine learning into the manufacturing process of wood fiber boards would provide the opportunity to predict in real-time the performance characteristics of the final product, to carry out production according to certain specifications and to reduce part of the raw material consumption and/or increase overall throughput (Escobar et al. 2018, Kwak et al. 2012, Lu 1990).

The objective of the present work was (1) to develop a random forest algorithm (RF) to predict with high accuracy mechanical wood fiber board properties in the form of internal bond strength (σ_{10}) and compression strength at 10% deformation (σ_{mt}); (2) to compare the generalization performance of the RF algorithm with artificial neural networks (ANNs) and support vector machines (SVMs); and (3) to apply feature selection techniques to improve training efficiency, model interpretability and generalizability by reducing the input dimensionality.

2. Materials and methods

2.1 Quality control of the mechanical properties of wood fiber insulation boards

Softwood containing approximately 70 % w/w *Picea abies* and 30 % w/w *Abies alba* was used in a dry process to produce wood fiber insulation boards. The board samples were taken from the production line and were conditioned and equilibrated at 65 % RH and 22 C° prior to testing. The

specifications of the produced wood fiber insulation boards were in accordance with the European Standard EN 13171 (2012). In addition to testing the compressive strength of the boards at 10 % deformation σ_{10} (kPa) according to the European Standards EN 826 (1996), their internal bond strength (IB) σ_{mt} (kPa) was also tested according to EN 1607 (1996). This is because the IB is one of the most important board parameters in the industrial production of fiber boards (Dunky and Niemz 2002). Overall, 960 wood fiber insulation boards were tested for internal bond strength and 1210 for compressive strength.

2.2 Data preprocessing

The interdependent machine and process variables ($n=186$) of the refiner, process sections digester, and the gluing, drying, forming and pressing units, were extracted from the process control system (Prod-IQ) and combined with tested mechanical properties of wood fiber insulation boards according to product type and time of manufacture. These process variables comprised important parameters, described in detail by several authors, such as board density, moisture content, mat-forming variables, gluing, line speed, several pressing characteristics, drying temperature etc. (Lobenhoffer, 1982; Lobenhoffer 1986; Lobenhoffer 1987; Niemz and Schweitzer 1990; Dunky and Niemz 2002, Riegler et al. 2013).

The data were normalized between 0 and 1 in order to fit the requirements of the learning algorithms and to avoid the problem of prevailing high values, which would have a greater influence on the loss function because of their scale and not for being more important than the other values.

2.3 Random forest

A random forest (RF) is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees trained with the bagging (bootstrapping aggregation) method. Generally, a random forest consists of a collection of tree-structured classifiers $\{h(x, \Theta_k), k=1, \dots\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x (Breiman 2001).

The generalization performance of the random forest algorithm was compared with common machine learning algorithms such as artificial neural networks and support vector machines. A detailed mathematical description of the algorithms used can be found in the corresponding literature (Bishop 2006, Breiman 2001, Burden et al. 2009, Vapnik 1995, Wu et al. 2017).

The hyper-parameters of the random forest algorithm and the support vector machine were optimized using Bayesian optimization, in which a learning algorithm's generalization performance is modeled as a sample from a Gaussian process (Snoek et al. 2012). The optimal configuration of the artificial neural network, i.e., the choice regarding the type of training algorithm, activation function (e.g., radial basis, tanh, ReLU) and number of neurons per layer used, was obtained by trial and error until no further improvement could be achieved. All machine learning algorithms used were programmed in MATLAB R2018b using the statistics and machine learning toolbox.

2.4 Feature selection and performance assessment

The sequential feature selection was performed to exclude input parameters that might have contained information irrelevant to the case under study (Cai et al. 2018, Kira et al. 1992). In order to assess the influence of the hyper-parameters and the performance of the algorithms a k-fold cross-validation was employed (Refaeilzadeh et al. 2009). The evaluation of the algorithms and

the selection of the best model were based on performance metrics used during the training and testing phases, metrics such as the root-mean-square error (RMSE), the correlation coefficient R , the coefficient of determination R^2 and the mean absolute percentage error (MAPE), a statistical measure of predictive accuracy (goodness-of-fit).

3. Results and Discussion

The experimental methodology of cross-validation is a suitable and widely used approach to determine the generalization performance of machine learning algorithms with respect to the ‘unseen’ data and to adjust their hyperparameters effectively (Kohavi 1995, Mansfield Shawn et al. 2007). Therefore, to assess model performance, a stratified 10-fold cross-validation approach was used and the accuracy measures R , R^2 , RMSE, MAPE were obtained based on the average results of the training and test data sets (10 folds).

Performance assessment and comparison of the machine learning algorithms to predict the internal bond strength (σ_{mt}) and compressive strength at 10% deformation (σ_{10}) was based on both, evaluation indices and graphical plots (Table 1, Figures 1&2). As shown in Table 1, the results of the training and test data sets are in a similar range and the deviation of the maximum and minimum values for each of the 10 folds was very limited (data not shown), indicating that all machine learning models showed no tendency towards overfitting. Furthermore, as judged by the relevant indices, the predictions of all machine learning algorithms correlated properly with the experimental observations of the mechanical wood fiber properties, but the random forest model revealed the best agreement for the training and test data sets (Table 1).

Figures 1 and 2 show the bias (actual vs. predicted internal bond strength and compressive strength), and the corresponding residuals plots for the machine learning algorithms RF, SVM and

ANN respectively. The bias from the RF algorithm was slightly closer to zero with fewer outliers for both mechanical properties and the spread of the residuals closer to zero compared to the SVM and ANN (Figure 1&2). Moreover, the residuals for all machine learning algorithms, with one exception, were distributed symmetrically around zero, indicating no tendency to either over- or underestimate the mechanical properties. Only the SVM showed a slight underestimation due to the asymmetry of the compressive strength residuals (Figure 2d).

High-dimensional data analysis is a challenge in the fields of machine learning and data mining. Feature selection provides an effective way to solve this problem by removing irrelevant and redundant input features, which can reduce computation time, improve learning accuracy, and facilitate the understanding of the learning model (Cai et al. 2018). In this study, the sequential forward feature selection method resulted in a significant reduction in the required training time and a slight improvement in generalization performance with the exception of ANN for compressive strength (Devaney et al. 1997). However, after dimensionality reduction, the random forest algorithm still showed the best performance compared to SVM and ANN (Table 2). In the case of RF, from 186 original input parameters, a total of 18 features were selected for internal bond strength and 14 for compressive strength, resulting in an 87.3% reduction in training time and slightly improved generalizability (σ_{int} : $R=0.960$, $R^2=0.916$, $\text{RMSE}=4.05$, $\text{MAPE}=12.11$; σ_{10} : $R=0.981$, $R^2=0.963$, $\text{RMSE}=17.19$, $\text{MAPE}=5.64$). Furthermore, the method of feature selection made it possible to identify important process variables that influence the mechanical properties of wood fiber insulation boards. Not surprisingly, the method revealed that, apart from the density, the selected process variables were also related to several other parameters, including those of the refiner (e.g. aux_ steam_energy, Ref1_cooker_pres_set), the fiber moisture content, the gluing process (e.g. M2_PC 1_glue_fact), the line speed and especially those of the press, such as

press_total_pressure, press_dist_kolben_2 and press_belt. These results are in good agreement with the observations of other authors who have investigated in detail the influence of process parameters on particleboard quality (Lobenhoffer, 1982, Lobenhoffer 1986, Lobenhoffer 1987, Niemz and Schweitzer 1990; Dunky and Niemz 2002, André et al. 2008, Riegler et al. 2013). Although the advantages of feature selection are apparent, one should also keep in mind that the reduction of dimensionality might also cause an inadvertent loss of information, especially if the parameters are dependent on each other.

Overall, all the machine learning algorithms used and the RF algorithm in particular, revealed very good prediction accuracies and abilities to generalize. This is remarkable since no data cleaning was carried out in the present study. Usually, data preprocessing involves the detection and deletion of outliers from the input parameter space to homogenize the data and to get better results. This is one advantage of machine learning algorithms compared to common statistical techniques such as multiple linear regression (MLR), which has been intensively investigated in the past for its suitability for the optimization of the particleboard production (Lobenhoffer, 1982, Lobenhoffer 1986, Kruse et al. 1997, Ritter et al. 1990). Compared to MLR techniques, machine learning algorithms are more robust against outlier and, more importantly, they can also deal with nonlinearities and are able to analyze simultaneously numerous highly multi-collinear, noisy predictor variables and also model several predictor variables simultaneously.

Moreover, the performance of the RF algorithm is comparable or even better than models primarily based on ANNs used by other authors for predictive modeling of forest product characteristics (André et al. 2008, Barnes 2001, Cook et al. 1997, Cook et al. 2000, Esteban et al. 2009a, Esteban et al. 2009b, Esteban Luis et al. 2011, Gupta et al. 2007). In this context, it is also worth to mention that the authors of the mentioned references used considerably smaller data sets with only a few

hundred samples. For instance, Esteban Luis (2011) investigated the performance of an artificial neural network for the prediction of plywood bonding quality using 8 input variables and 139 observations (111 for training and 28 for testing).

In the present work, the full data set contained 186 variables (input parameter space) and 960 or 1210 observations, depending on the mechanical property investigated. Although the data set used can be considered as large and valid, we are aware of the fact that it is still not big data with more than hundreds of thousands or even millions of observations. In future studies, we would like to further improve quality control by analyzing big data with deep learning techniques based on deep artificial neural networks (Kusiak 2017, Wang et al. 2018). Nevertheless, it can be safely concluded, that our work using real plant process control data showed that the random forest algorithm can already be applied to predict with high accuracy relevant properties of wood fiber boards for improved quality control.

4. Conclusion

In the present paper, the challenges of the wood panel industry, such as increasing complexity, high data dimensionality and increased requirements for efficient product quality control were first pointed out. Subsequently, machine learning algorithms were examined for their suitability for these challenges. Interdependent machine and process variables were extracted from the process control system (Prod-IQ) and combined with tested mechanical properties of wood fiber insulation boards according to product type and time of manufacture. The performance of the random forest algorithm was evaluated and compared with a support vector machine and an artificial neural network using the k-fold cross-validation method. The random forest revealed the best generalization performance, enabling the prediction of the internal bond strength (σ_{mt}) and compressive strength at 10% deformation (σ_{10}) of wood fiber insulation boards. The high

219 predictive ability of the random forest algorithm allows real-time knowledge of the properties of
220 the final product and the ability to change process parameters online during the production process.
221 This will increase the efficiency of the production line and improve quality control by producing
222 according to selected specifications, reducing some of the raw material consumption and
223 increasing overall throughput.

224

225 **Acknowledgments**

226 The authors would like to thank Mr. G. Bernardy, Head of Department MES & Industrial IT,
227 Siempelkamp, Germany, for providing the process and machine data.

228

229 **References**

- 230 André, N., Cho, H.-W., Baek, S.H., Jeong, M.-K., Young, T.M. (2008) Prediction of internal bond
231 strength in a medium density fiberboard process using multivariate statistical methods and
232 variable selection. *Wood Sci Tec.* 42:521-534.
- 233 Barnes, D. (2001) A model of the effect of strand length and strand thickness on the strength
234 properties of oriented wood composites. *Forest Prod J.* 51:36-45.
- 235 Bishop, C.M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*.
236 Springer-Verlag, 2006.
- 237 Breiman, L. (2001) Random Forests. *Mach Learn.* 45:5-32.
- 238 Burden, F., Winkler, D. (2009) Bayesian Regularization of Neural Networks. In: *Artificial Neural*
239 *Networks: Methods and Applications*. Livingstone, D.J. Humana Press, Totowa, NJ. pp.
240 23-42.
- 241 Cai, J., Luo, J., Wang, S., Yang, S. (2018) Feature selection in machine learning: A new
242 perspective. *Neurocomputing.* 300:70-79.
- 243 Cook, D.F., Chin-Chou, C. (1997) Predicting the internal bond strength of particleboard, utilizing
244 a radial basis function neural network. *Eng Appl Artif Intell.* 10:171-177.
- 245 Cook, D.F., Ragsdale, C.T., Major, R.L. (2000) Combining a neural network with a genetic
246 algorithm for process parameter optimization. *Eng Appl Artif Intell.* 13:391-396.
- 247 Devaney, M., Ram, A., *Efficient Feature Selection in Conceptual Clustering*, in *Proceedings of*
248 *the Fourteenth International Conference on Machine Learning* 1997, Morgan Kaufmann
249 Publishers Inc. p. 92-97.
- 250 Dunky M, Niemz P (2002) *Holzwerkstoffe und Leime*. Springer, Berlin.
- 251 Escobar, C.A., Morales-Menendez, R. (2018) Machine learning techniques for quality control in
252 high conformance manufacturing environment. *Adv Mech Eng.* 10:1-16.
- 253 Esteban, L.G., Fernández, F.G., de Palacios, P. (2009a) MOE prediction in *Abies pinsapo* Boiss.
254 timber: Application of an artificial neural network using non-destructive testing. *Comput*
255 *Struct.* 87:1360-1365.
- 256 Esteban, L.G., García Fernández, F., Palacios, P.d., Conde, M. (2009b) Artificial neural networks
257 in variable process control: application in particleboard manufacture. *For Sys.* 18:92-100.
- 258 Esteban Luis, G., Fernández Francisco, G., de Palacios, P. (2011) Prediction of plywood bonding
259 quality using an artificial neural network. *Holzforschung.* 65:209-214.
- 260 European Standard EN 826 (1996). Thermal insulating products for building applications -
261 Determination of tensile strenght perpendicular to faces. European Commitee for
262 Standardization (CEN), Brussels, Belgium
- 263 European Standard EN 1607 (1996). Thermal insulation products for building applications -
264 Determination of compression behaviour. European Commitee for Standardization (CEN),
265 Brussels, Belgium
- 266 European Standard EN 13171 (2012). Thermal insulating products for buildings - Factory made
267 wood fibre (WF) products - Specification. European Committee for Standardization
268 (CEN), Brussels, Belgium
- 269 Gupta, A., Jordan, P., Pang, S., *Modelling of the Development of the Vertical Density Profile of*
270 *MDF during Hot Pressing*, in *Chemical Product and Process Modeling* 2007.
- 271 Kira, K., Rendell, L.A., *A practical approach to feature selection*, in *Proceedings of the ninth*
272 *international workshop on Machine learning* 1992, Morgan Kaufmann Publishers Inc.:
273 Aberdeen, Scotland, United Kingdom. p. 249-256.

- Kohavi, R., *A study of cross-validation and bootstrap for accuracy estimation and model selection*, in *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2* 1995, Morgan Kaufmann Publishers Inc.: Montreal, Quebec, Canada. p. 1137-1143.
- Kruse, K., Thömen, H., Maurer, H., Steffen, A., Leon-Mendez, R. (1997) Optimierung der Spanplattenproduktionsprozesses mit Hilfe der Prozessmodellierung. (Optimisation of particleboard production by means of process modelling). *Holz Roh- Werkst.* 55: 17-24.
- Kusiak, A. (2017) Smart manufacturing must embrace big data. *Nature.* 544:23-25.
- Kwak, D.-S., Kim, K.-J. (2012) A data mining approach considering missing values for the optimization of semiconductor-manufacturing processes. *Exp Sys Appl.* 39:2590-2596.
- Lobenhoffer, H. (1982) Auswertung von Prozessdaten mittels statistischer Methoden. (Evaluation of Process Data by Means of Statistical Methods). *Holz Roh- Werkst.* 40: 395-401.
- Lobenhoffer, H. (1986) Überprüfung eines Optimierungsverfahrens für einen Spanplattenformstrang durch Simulation mit Betriebs-Datensätzen. (Testing an optimization method for particleboard forming lines by simulation using real process data). *Holz Roh- Werkst.* 44: 55-56.
- Lobenhoffer, H. (1987) Verfahren zur optimalen Regelung eines Formstrangs zur Herstellung von Spanplatten. (Procedure for the optimal control of a mat forming line in particleboard production). *Holz Roh- Werkst.* 45: 429-337.
- Lu, S.C.Y. (1990) Machine learning approaches to knowledge synthesis and integration tasks for advanced engineering automation. *Comput Ind.* 15:105-120.
- Mansfield Shawn, D., Iliadis, L., Avramidis, S. (2007) Neural network prediction of bending strength and stiffness in western hemlock (*Tsuga heterophylla* Raf.). *Holzforschung.* 61:707-716.
- Niemz, P. Schweitzer, F. (1990) Einfluss ausgewählter Strukturparameter auf die Zug- und Druckfestigkeit von Spanplatten. (Influence of some selected structural parameters on tensile and compression strength of particle boards). *Holz Roh- Werkst.* 48: 361-364.
- Riegler, M., Spangl, B., Weigl, M., Wimmer, R., Müller, U. (2013) Simulation of a real-time process adaptation in the manufacture of high-density fibreboards using multivariate regression analysis and feedforward control. *Wood Sci Tec.* 47: 1243-1259.
- Ritter, C., Landmesser, W., Niemz, P., Hänsel, A. (1990) Vergleichende Untersuchungen zur Anwendbarkeit verschiedener Methoden für die Modellierung der Spanplattenfertigung. (Comparative studies on the applicability of different methods for modelling particleboard production). *Holztechnologie* 30: 23-25.
- Refaeilzadeh, P., Tang, L., Liu, H. (2009) Cross-Validation. In: *Encyclopedia of Database Systems*. Liu, L., Özsu, M.T. Springer US, Boston, MA. pp. 532-538.
- Snoek, J., Larochelle, H., Adams, R.P., *Practical Bayesian optimization of machine learning algorithms*, in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2* 2012, Curran Associates Inc.: Lake Tahoe, Nevada. p. 2951-2959.
- Vapnik, V.N. The nature of statistical learning theory. Springer-Verlag, 1995.
- Wang, J., Ma, Y., Zhang, L., Gao, R.X., Wu, D. (2018) Deep learning for smart manufacturing: Methods and applications. *J Manuf Syst.*
- Wu, D., Jennings, C., Terpenney, J., Gao, R.X., Kumara, S. (2017) A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. *J Manuf Sci Eng.* 139:1-9.

319 Wuest, T., Weimer, D., Irgens, C., Thoben, K.-D. (2016) Machine learning in manufacturing:
320 advantages, challenges, and applications. Prod Manuf Res. 4:23-45.
321

Figure 1: Comparison of predicted and observed internal bond strength (σ_{mt}) and the corresponding residuals according to the random forest (a, b), support vector machine (c, d) and artificial neural network (e, f) for the test data sets.

Figure 2: Comparison of predicted and observed compressive strength at 10% deformation (σ_{10}) and the corresponding residuals according to the random forest (a, b), support vector machine (c, d) and artificial neural network (e, f) for the test data sets.

330 **Table 1:** Descriptive statistics for the machine learning algorithms over the 10 folds

		^a Random forest		^b Support vector machine		^c Artificial neural network	
		^d σ_{mt}	^e σ_{10}	σ_{mt}	σ_{10}	σ_{mt}	σ_{10}
R	Training	0.993	0.996	0.949	0.977	0.959	0.983
	Testing	0.954	0.980	0.936	0.973	0.946	0.977
R²	Training	0.984	0.992	0.901	0.955	0.910	0.965
	Testing	0.910	0.960	0.865	0.947	0.894	0.954
RSME	Training	1.86	8.79	4.62	20.88	4.33	18.23
	Testing	4.27	18.02	5.08	21.34	4.59	19.79
MAPE	Training	5.29	2.28	14.20	8.27	13.89	6.10
	Testing	13.51	5.74	17.73	9.22	13.79	6.91

Hyperparameters:
^a = ‘Trees’: 3000; ‘mtry’: 62; ‘Sample size’: size(X_{train},1); ‘Minimal node size’: 5; ‘Splitting rule’: Randomly selected mtry; ‘MinLeafSize’: 1; ‘NumPredictorsToSample’: 30
^b = ‘KernelFunction’: linear; ‘Sigma’: 8.786e-04; ‘IterationLimit’: 1e6; ‘Solver’: SMO; ‘BoxConstraint’: 0.00132; ‘Epsilon’: 1.81
^c = ‘Architecture’: 186-[15-10-5]-1; ‘Training algorithm’: Bayesian regularization backpropagation; ‘Transfer function’: tansig (hidden layers), purelin (output layer); ‘Epochs’:10; ‘Learning rate’: 0.3
^d = Internal bond strength (kPa)
^e = Compressive strength at 10 % deformation (kPa)

331

332

333 **Table 2:** Descriptive statistics for the machine learning algorithms after feature selection

^a Random forest				^b Support vector machine		^c Artificial neural network	
		^d σ_{mt}	^e σ_{10}	σ_{mt}	σ_{10}	σ_{mt}	σ_{10}
R	Testing	0.960	0.981	0.942	0.975	0.951	0.976
R²	Testing	0.916	0.963	0.886	0.951	0.904	0.950
rsme	Testing	4.05	17.19	4.79	20.61	4.34	20.85
MAPE	Testing	12.12	5.64	16.10	8.75	12.19	7.54

^a18 features for internal bond strength and 14 for compressive strength

^b39 features for internal bond strength and 32 for compressive strength

^c12 features for internal bond strength and 6 for compressive strength

^d = Internal bond strength (kPa)

^e = Compressive strength at 10 % deformation (kPa)

334

335