Importance-driven bottleneck model: a multi-stage deep learning approach for analyzing swelling-shrinkage behavior of wood species and their

Citation for the original published paper (version of record):

Movahedi-Rad, A., Grabner, M., Subhani, M. et al (2026). Importance-driven bottleneck model: a multi-stage deep learning approach for analyzing
swelling-shrinkage behavior of wood species and their mechanical properties. Journal of Wood Science, 72(1). http://dx.doi.org/10.1186/s10086-026-02266-9

Movahedi-Rad *et al. Journal of Wood Science* (2026) 72:21
https://doi.org/10.1186/s10086-026-02266-9

Journal of Wood Science

ORIGINAL ARTICLE

Open Access

Importance-driven bottleneck model: a multi-stage deep learning approach for analyzing swelling–shrinkage behavior of wood species and their mechanical properties

A. Vahid Movahedi-Rad[1*], Michael Grabner[2], Mahbube Subhani[3], Seyed Mohsen Moosavi-Dezfooli[4], Ingo Burgert[1,5] and Mark Schubert[5*]

Abstract

Experimental data were collected from defect-free specimens of 57 wood species in Central Europe, encompassing their structural, physical, chemical characteristics, and mechanical properties. The dataset highlights the complex relationships between physical properties (shrinkage–swelling behavior in three orthotropic directions), chemical composition (holocellulose, lignin, extractive content), structural characteristics (density, fiber length), and mechanical properties (compression, tensile and bending strength, hardness, and impact resistance), emphasizing the need for efficient and interpretable methods to identify their correlations. To address this issue, we developed an importance-driven bottleneck model, by revisiting the concept bottleneck model, which consists of a system of three connected neural networks; auxiliary, input, and output, that collaboratively work to investigate the relationship between the physical, structural, chemical, and mechanical features. The input and output networks are linked by a concept c bottleneck layer, with its features defined by analyzing wood's structural characteristics and composition via feature importance analysis conducted in the auxiliary network. Density and fiber length were identified as the most important concept c features, yielding testing R^2 scores above 0.70 for predicting mechanical properties. Predictive accuracy increased to over 0.85 when a third component, either holocellulose or lignin content, was included in the bottleneck. By adding all five chemical and structural features used in concept c layer, the prediction accuracy was slightly increased. While the importance-driven bottleneck model with three or more concept c features was narrowly outperformed by an end-to-end network using shrinkage–swelling as direct inputs, it offered a superior balance between performance and interpretability. This mathematical strategy, forcing information through a concept bottleneck, creates an interpretable physics-inspired AI framework.

Keywords Wood, Physics-inspired AI framework, Shrinkage–swelling behavior, Mechanical properties

*Correspondence:
A. Vahid Movahedi-Rad
avahidmovahedirad@ethz.ch
Mark Schubert
mark.schubert@empa.ch
Full list of author information is available at the end of the article

Introduction

Wood is a popular choice as a structural and construction material due to its desirable properties, such as high strength-to-weight ratio and stiffness [1, 2]. These desirable properties make it widely used in applications ranging from buildings to structural sandwich panels and scaffolds in composite materials [3–5]. Analyzing wood properties is essential for designing and developing engineered materials due to the significant variations between species, and even individual specimens from the same tree [1, 6–8].

Wood is primarily comprised of three macromolecules, namely, cellulose, hemicellulose, and lignin [9]. These chemical components, along with structural features such as tissue structure, fiber length, and density influence wood's physical and mechanical properties [7, 8]. Wood as a hygroscopic material is capable of adsorbing and releasing moisture from the cell wall, which directly influences its dimensions and significantly alters its mechanical properties [9–12]. As a result, the relationship between the moisture content and mechanical properties of wood has been a topic of extensive research [10, 13, 14].

The shrinkage and swelling behavior of wood, which is influenced by factors, such as wood species, microstructure, chemical composition, and anatomical direction, also influences its mechanical properties [8, 11, 13]. Despite numerous studies linking these factors, the inherent variability, non-homogeneity, and anisotropy of wood continue to present challenges in fully unraveling the complex inter-relations between chemical composition, structural characteristics, and mechanical performance [10].

Machine learning approaches offer opportunities to deal with such multi-aspect issues as they can handle large and complex datasets due to their ability to learn and improve from experience without being explicitly programmed [15–18]. In these methods, the relationships between wood properties and parameterized input data are explored via dynamic optimization of the model parameters, by learning from the parameterized input data. Researchers utilized different learning methods to predict the mechanical properties of wood and wood-based composites [19–22]. Among different machine learning methodologies, neural networks are widely used due to their ability to learn and adapt from vast amounts of data and to handle the nonlinear relationships within data. In addition, neural networks can perform feature selection implicitly during the learning process, identifying the most informative features for a given task while disregarding irrelevant or redundant ones [23].

High accuracy in predictions and good performance of neural networks are often associated with a lack of human interpretability, referred to as black-box models, since it can be challenging to understand how the network arrives at its predictions, making it difficult for researchers and engineers to interpret and retrace the results [24]. In numerous studies, interpretability was addressed using techniques, such as feature importance analysis [25, 26]. Nevertheless, the reliability of these analyses in neural networks is not always given and depends on various factors, including the dataset, the model architecture, the model performance, and the method used for feature importance estimation [27]. To address these issues, new approaches were suggested, aiming at improving the interpretability of neural networks [28, 29]. One of these is the emerging paradigm of theory-guided data science (TGDS). In this framework, the neural network structure is informed by a physical understanding of the phenomena under investigation [28]. Early efforts toward interpretable learning leveraged attribute- and semantic-based representations, enabling models to reason over human-understandable properties rather than raw features [30, 31]. Building on these ideas, Koh et al. [29] introduced the concept bottleneck models. In this method, the model first predicts the concepts selected by utilizing domain knowledge or interpretable parameters, using an input neural network, and afterward uses only those predicted concepts to make a final prediction using the output neural network. Concept bottleneck models were employed in computer vision-related studies [29]. Nevertheless, such an enhancement in interpretability is often accompanied by a reduction in accuracy. This results in a trade-off between these two aspects [24, 29].

The aim of this study is to conduct an exploratory investigation by leveraging neural networks to uncover correlations between the chemical composition, structural characteristics, shrinkage–swelling behaviors in the three orthogonal directions (longitudinal, radial, and tangential) and mechanical properties of the investigated wood species, using the defined dataset. Another objective of this work is to address the lack of interpretability of the typical neural network models, commonly used in existing research. This is done by modifying the concept bottleneck model [29] by determining the concept c features, referring to a specific set of interpretable features, through an auxiliary neural network. This mathematical strategy establishes a physics-inspired AI framework forcing information through a concept bottleneck, with a focus on balancing interpretability and accuracy. The results obtained were analyzed and compared with those from a conventional end-to-end neural network model.

Computational procedure

Artificial neural network

Artificial Neural Networks (ANNs) comprise interconnected structure, containing numerous simple processing units known as neurons, which are organized into layers in the network [32, 33]. Neural networks are computing models for information processing and are particularly useful for identifying the fundamental relationships between a set of variables or patterns in the data. The multi-layer perceptron (MLP) neural network, trained with an end-to-end manner, is considered a common choice among ANNs, typically structured with an input layer, an output layer, and one or more hidden layers, enabling the network to comprehend the connections between input and output variables [32, 33]. The effectiveness of neural networks in training high precision predictors has been repeatedly demonstrated across diverse domains [34–38]. In spite of the advantages, it is accompanied by the user's lack of insight into the underlying processes behind the specific outputs generated by the end-to-end neural networks [39, 40]. This is because the valuable information from the problem decomposition is often ignored during the end-to-end training, creating a challenge of interpretability for users [29, 40].

Importance-driven bottleneck model

To address the interpretability issue associated with neural network models, concept bottleneck models were introduced [29]. These models involve revisiting the simple idea of first predicting an intermediate set of human-specified concepts c, and then using it to predict the target y [29–31]. Such interventions cause richer human–model interaction and turn the models more interpretable in terms of high-level concept, nevertheless, they lead to a perceived trade-off between accuracy and interpretability [29]. The decision regarding the selection of concepts c is considered as a hyperparameter, having a significant effect on both the performance and interpretability of the model [41]. Supposing that the concepts c is chosen effectively, it will generate predictions that correspond closely to the true concepts. This indicates that the selected concept is defined as critical factors, which accurately capture the essence of a bottleneck situation. Conversely, if concepts c are not well-selected, meaningful results from interventions cannot be anticipated [29].

In this study, the concept bottleneck model is adapted to investigate the inter-relation between mechanical properties of wood species and their shrinkage–swelling behavior. The features representing the concepts c are selected through a feature importance analysis conducted on an auxiliary neural network, where the significance of each feature is ranked. The features of the auxiliary neural network represent potential key parameters that can influence the prediction of the final targets. The importance-driven bottleneck model aims to improve the interpretability of neural networks, alignment with the domain of knowledge, and aiming to reduce the manual tuning of the concept bottleneck model.

Experimental data

A list of the 57 wood species existing in Central Europe, summarized in Fig. 1, is accompanied by a comprehensive dataset containing extractive content, holocellulose content, lignin content, fiber length, density, and swelling behavior in the longitudinal, radial, and tangential directions. In total, this dataset comprises 2487 data points for the aforementioned properties, along with 2205 measurements of shrinkage behavior in the longitudinal, radial, and tangential directions. In addition, mechanical properties of these species were determined, leading to a total of 2487, 663, 2487, 587, and 1580 measurements for compression and tensile strength in the longitudinal direction, bending strength and impact resistance in the tangential direction as well as hardness in radial direction. All measurements were experimentally conducted at *University of Natural Resources and Life Sciences Vienna* and the detailed descriptions of the experimental procedure are provided in Appendix, Sect. *Experimental Procedure*. A statistical summary of the dataset is presented in Table 1.

Model applications

Data organization

The properties of each wood species' data were organized separately in ascending order in which the shrinkage values presented with negative signs and swelling values signified by positive signs. Since the dataset from the mechanical experiments had uneven amounts of data, in each wood species, the data were distributed evenly across all properties to match the size of the property dataset with the highest number of data points; however, this process led to some missing values in the dataset. The missing values for specific properties of individual wood species were imputed by utilizing linear interpolation, data resampling, only if the number of these non-missing values exceeded six data points. This approach was chosen to ensure that the nature of the data remained unchanged. For properties with non-missing values were fewer than six available data points, the entire property dataset was masked and excluded from both the training and testing datasets. The masking technique was done by identifying non-missing values, calculating the error only for those non-missing values in the loss function during the training process. The threshold of six observations was selected to ensure a minimal level of statistical representativeness at the species–property level. On

Table 1 Statistical summary of dataset

Property	Mean	Standard deviation	First quartile	Median	Third quartile
Extractive content (%)	7.96	3.23	5.90	6.80	9.87
Holocellulose content (%)	39.88	3.63	37.58	38.87	42.32
Lignin content (%)	18.43	3.68	15.89	17.70	19.00
Fiber length (μm)	1210.13	686.69	871.07	1071.20	1283.91
Density (g/cm^3)	0.70	0.12	0.60	0.70	0.79
Radial swelling (%)	6.08	2.04	4.63	5.87	7.35
Longitudinal swelling (%)	0.56	0.30	0.34	0.52	0.74
Tangential swelling (%)	11.85	3.20	9.39	11.85	14.07
Radial shrinkage (%)	−6.05	1.97	−7.30	−5.88	−4.63
Longitudinal shrinkage (%)	−0.59	0.34	−0.75	−0.53	−0.35
Tangential shrinkage (%)	−10.87	2.70	−12.76	−10.84	−8.91
Compression strength (MPa)	52.84	9.88	45.45	51.92	60.53
Tensile strength (MPa)	144.86	49.67	107.74	146.35	178.86
Bending strength (MPa)	112.55	26.05	95.60	112.70	129.72
Hardness (N/mm^2)	31.45	9.69	24.37	31.55	38.31
Impact resistance (kJ/m^2)	60.07	33.37	36.04	53.96	76.08

the other hand, a higher threshold would increase the number of missing values (NaNs), specifically for Tensile Strength (see Fig. SI). To mitigate the impact of masking properties, a weighting factor was introduced into the loss function to improve the robustness of model training. To clarify the effect of species-level resampling on the underlying data distribution, p values comparing the original and resampled datasets were calculated. As shown in Table SI, all p values exceed 0.05, indicating that the distribution functions as a result of the resampling process have not changed significantly.

Importance-driven bottleneck model

The flow chart of the suggested concept bottleneck model is shown in Fig. 1, comprising three parts, attributing to the auxiliary, input and output networks. First, the selected features of the auxiliary neural network; extractive, holocellulose lignin content, density, and fiber length, as well as the targets; shrinkage–swelling in three orthogonal directions, were standardized, according to Eq. 1, and imported to the network. The most important features of the trained neural network were then determined by performing feature importance analysis methods. The feature importance analysis provides insight into which aspects of wood composition or microstructure, employed as features, have the most significant impact on its predicted targets. Four known feature importance analysis methods; gradient-based, weight-based, SHAP (Shapley Additive Explanations), and permutation importance, which have been commonly used by other researchers [42, 43], were employed in this work to provide a more comprehensive understanding of how different features contribute to the model's predictions. The selected features were considered as targets in the second neural network, named the input neural network, in which the standardized shrinkage–swelling behavior along with the wood species identifier were considered as the features. Since species identity is a strictly categorical variable, one-hot encoding was employed to identify the wood species. After the training of the input neural network, the prediction results were standardized and imported as features to the third or output neural network, where the mechanical properties were the targets. The mechanical properties were resampled to ensure an equal number of measurements across different species. The suggested architecture of the importance-driven bottleneck model, comprising of three neural networks is illustrated in Fig. 2.

The data implemented in all three neural networks was divided into two sets at the wood species level. For each wood species, 70% of the samples were randomly assigned to the training set and the remaining 30% to the test set, such that samples from the same species appear in both sets. This strategy was chosen, because the objective of this study is to predict mechanical properties based on material features within known species, rather than to evaluate generalization to unseen species. The auxiliary network was structured with five hidden layers comprising 1024, 512, 256, 128, and 64 neurons, while the input and output neural networks were configured with five hidden layers containing 512, 256, 128, 64 and 32 neurons. To ensure reliable

Wood number	Wood species	Wood number	Wood species
1	*Abies alba Mill.*	30	*Picea abies (L.) Karst.*
2	*Aesculus hippocastanum L.*	31	*Pinus cembra L.*
3	*Ailanthus altissima (Mill.) Swingle*	32	*Pinus sylvestris L.*
4	*Alnus glutinosa (L.) Gaertn.*	33	*Platanus L.*
5	*Amelanchier ovalis M.*	34	*Populus L.*
6	*Berberis vulgaris L.*	35	*Prunus americana Marsh.*
7	*Betula L.*	36	*Prunus avium L.*
8	*Buxus sempervirens L.*	37	*Prunus domestica L.*
9	*Carpinus betulus L.*	38	*Prunus mahaleb L.*
10	*Castanea sativa Mill.*	39	*Prunus padus L.*
11	*Colutea arborescens L.*	40	*Prunus spinosa L.*
12	*Cornus mas L.*	41	*Pseudotsuga menziesii (Mirb.) Franco*
13	*Cornus sanguinea L.*	42	*Pyrus L.*
14	*Corylus avellana L.*	43	*Quercus L.*
15	*Crataegus (L.) Tourn.*	44	*Rhamnus cathartica L.*
16	*Euonymus europaeus L.*	45	*Robinia pseudoacacia L.*
17	*Fagus sylvatica L.*	46	*Rosa canina L.*
18	*Frangula alnus Mill.*	47	*Sambucus nigra L.*
19	*Fraxinus excelsior L.*	48	*Sorbus aria (L.) Crantz*
20	*Hippophae rhamnoides L.*	49	*Sorbus aucuparia L.*
21	*Ilex aquifolium L.*	50	*Sorbus domestica L.*
22	*Juglans regia L.*	51	*Sorbus torminalis (L.) Crantz*
23	*Juniperus communis L.*	52	*Staphylea pinnata L.*
24	*Laburnum anagyroides Medik.*	53	*Syringa vulgaris L.*
25	*Larix decidua Mill.*	54	*Taxus baccata L.*
26	*Ligustrum vulgare L.*	55	*Tilia L.*
27	*Lonicera L.*	56	*Viburnum lantana L.*
28	*Malus Mill.*	57	*Viburnum opulus L.*
29	*Morus L.*	* Soft wood and hard wood	

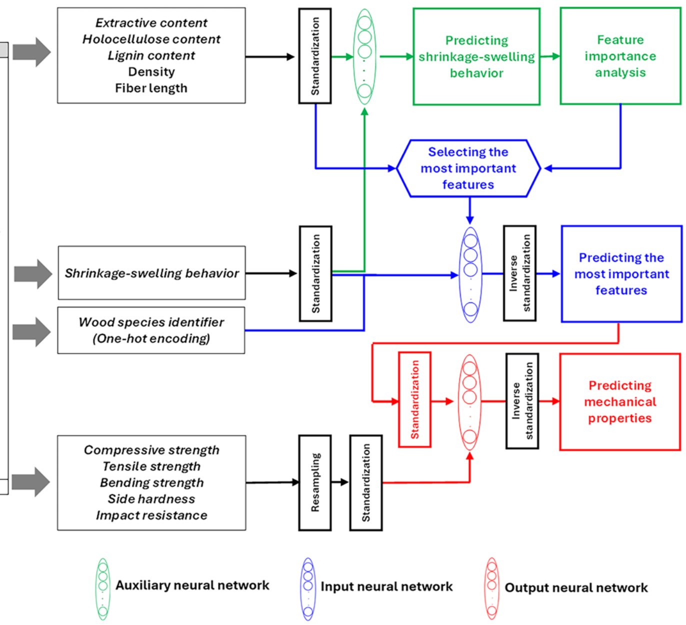

Fig. 1 Flow chart for importance-driven bottleneck model, used to predict mechanical properties of wood species. Integrated table on the left: Latin names of softwood species are marked in red and hardwood species in blue

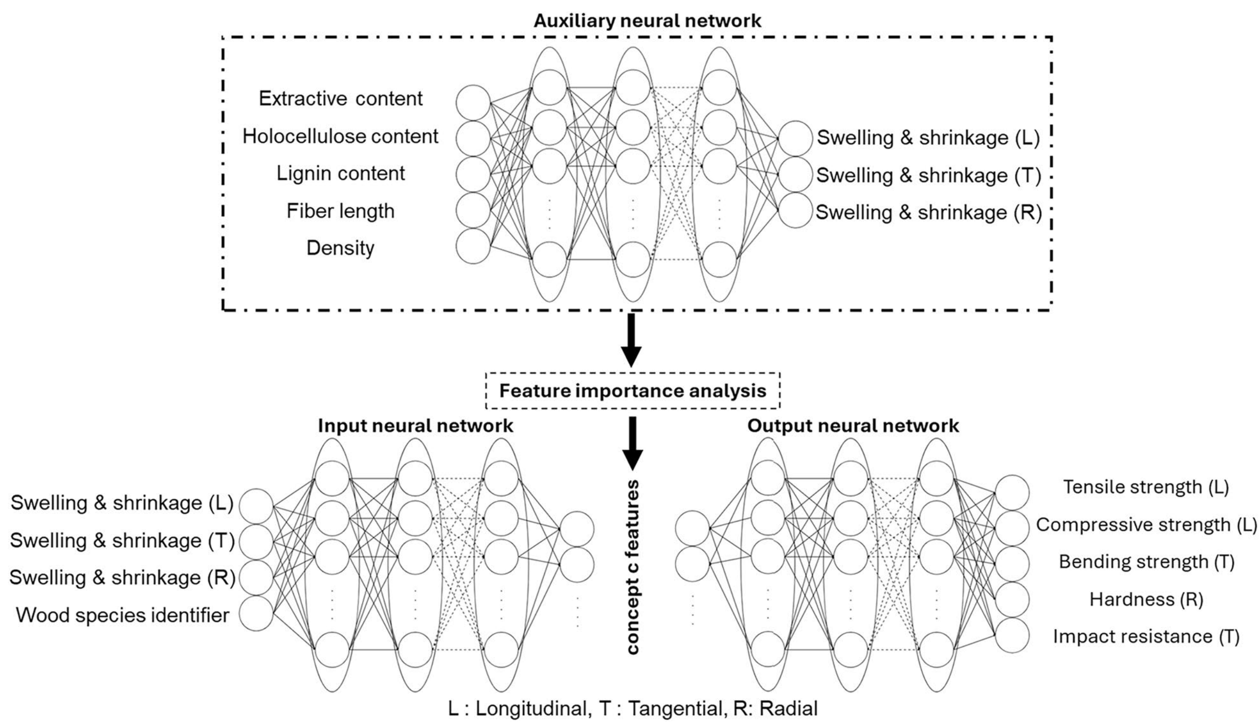

Fig. 2 Suggested architecture used in an importance-driven bottleneck model to predict mechanical properties of wood species

performance assessment, all three networks underwent training using a tenfold cross-validation approach, dividing the dataset into subsets for training and validation. The Rectified Linear Unit (ReLU) activation function was employed uniformly across networks to introduce non-linearity and enable feature learning.

The batch sizes were chosen for each network, with values of 128, 64, and 16 for the auxiliary, input, and output networks, respectively, to balance generalization, computational efficiency and convergence speed. In addition, a light regularization was applied to both input and output neural networks to prevent overfitting. In all neural networks, a learning rate of 0.001 was programmed. Optimization was facilitated by the Root Mean Square Propagation (RMSProp) optimizer across all networks, with the Mean Squared Error (MSE) serving as the loss function to quantify the disparities between predicted and actual values. RMSProp adapts the learning rate for each parameter based on the magnitudes of recent gradients. This adaptivity can be beneficial, because it allows the learning rate to vary across parameters, enabling faster convergence and more stable training. Furthermore, early stopping mechanisms with different epochs of patience were implemented to prevent overfitting and optimize training in the three mentioned neural networks. The implemented hyperparameters are summarized in Table S2.

The performance evaluation of all three neural networks was done by comparing each predicted property with the experimental data in both the training and testing datasets, which involved determining the accuracy of the predictions and quantifying the amount of error, see Sect. *prediction accuracy assessment*. Furthermore, the experimental and predicted distribution functions were compared.

End-to-end model

In contrast to the proposed the importance-driven bottleneck model, an end-to-end neural network model directly analyzes the inter-relation between mechanical properties of wood species and their shrinkage–swelling behavior, as shown in Fig. 3. The architecture of the end-to-end neural network was the same as the one used in the output neural network in the importance-driven bottleneck model (see Table S2).

Results and discussion

Importance-driven bottleneck model

Auxiliary neural network

The variation of the loss and accuracy curves for both the training and validation sets is shown in Fig. S2. The training process was characterized by a rapid initial decrease in loss for both sets, followed by smooth convergence. While a narrow gap between the training and validation loss curves was maintained, the model's complexity appeared well-matched to the data patterns. In addition, it was observed that the validation accuracy almost overlapped the training accuracy, indicating that the model generalized well without overfitting. A comparison between the predicted results and the experiments for the training and testing datasets is shown in Fig. 4a, b, and c for longitudinal, tangential, and radial directions, respectively. High training R^2 scores of 0.992, 0.999, and 0.999 related to three anatomical directions also indicate that the suggested neural network architecture has been successful in learning the relationship between the different input features (extractive, holocellulose and lignin content, fiber length, and density) and three output targets (shrinkage and swelling in three orthogonal directions). In the testing dataset, R^2 scores of 0.960, 0.993, and 0.989 were obtained in predicting the shrinkage–swelling in the longitudinal, tangential, and radial directions, respectively. Such high prediction accuracies indicate that the neural network could accurately predict the swelling–shrinkage behavior of the wood species, closely matching the experimental values by effectively generalizing from the training data to the testing data. To evaluate the role of the activation function, another network with Linear activation functions was programmed, and shown in Fig. S3. While comparing Fig. S3 against Fig. 4a–c, it is observed that the programmed network was not able to detect the pattern between the features and swelling–shrinkage behavior, turning the optimization process ineffective due to the complexity of the task.

The distribution functions of the predicted and experimental values from the testing dataset were compared and illustrated in Fig. 4d, e, and f, for longitudinal, tangential, and radial directions, respectively. Visually, it can be seen that the experimental distribution function was characterized by a bimodal form, attributed to the shrinkage and swelling behavior, where shrinkage is plotted in the negative axis and swelling is illustrated in the positive axis. It is observed that the predicted distribution function almost overlapped the experimental one in both shrinkage and swelling regions in the three directions. Due to the high performance of the auxiliary neural network in predicting shrinkage–swelling behavior across a wide variety of wood species, the use of the trained model in feature importance analysis can be considered to be reliable in determining the contribution level of each feature in training the model. Figure 5 shows the results of the feature importance analysis, highlighting that the most important features vary depending on the method used. Such observations have also been reported by Saarela and Jauhiainen [44], as each of these methods relies on a specific methodology to identify the most important features of a trained neural network. Therefore, combining several feature importance analysis techniques could yield more reliable and trustworthy results. For the gradient- and weight-based approach, the highest importance is given to fiber length followed by density. The SHAP method identified density as the

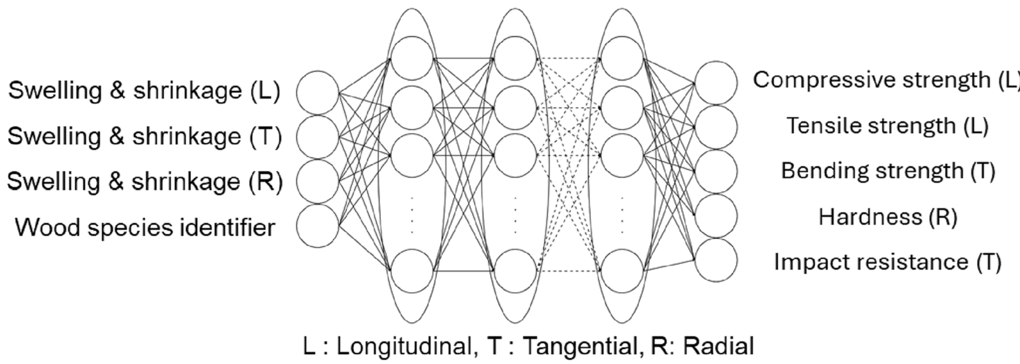

Fig. 3 Suggested architecture used in an end-to-end model to predict mechanical properties of wood species

most important feature, followed by fiber length; in contrast, the permutation approach also ranked density first but identified lignin content as the second most significant variable. Nonetheless, the permutation approach assigned nearly equal importance to lignin content, holocellulose content, and fiber length. Accordingly, considering all methods, it was revealed that density and fiber length were the most important features determining the swelling–shrinkage behavior of the wood species.

Input neural network

The swelling–shrinkage behavior in three directions together with the wood species identifier were used as features in the input neural network to unravel the relationship between these features and chemical and structural characteristics. The training and validation curves for both loss and accuracy sets are shown in Fig. S4. The model exhibited high efficiency, characterized by a nearly instantaneous transition from initial learning to peak performance. This suggests that the network effectively captured the defining features of the dataset with high confidence and no evidence of overfitting. Figure 6a, b demonstrates the predicted density and fiber length results with respect to their counterpart experiments. It can be observed that R^2 scores for density and fiber length predictions on the training dataset are 0.997 and 0.999, respectively, indicating that the input neural network model has been properly trained. Furthermore,

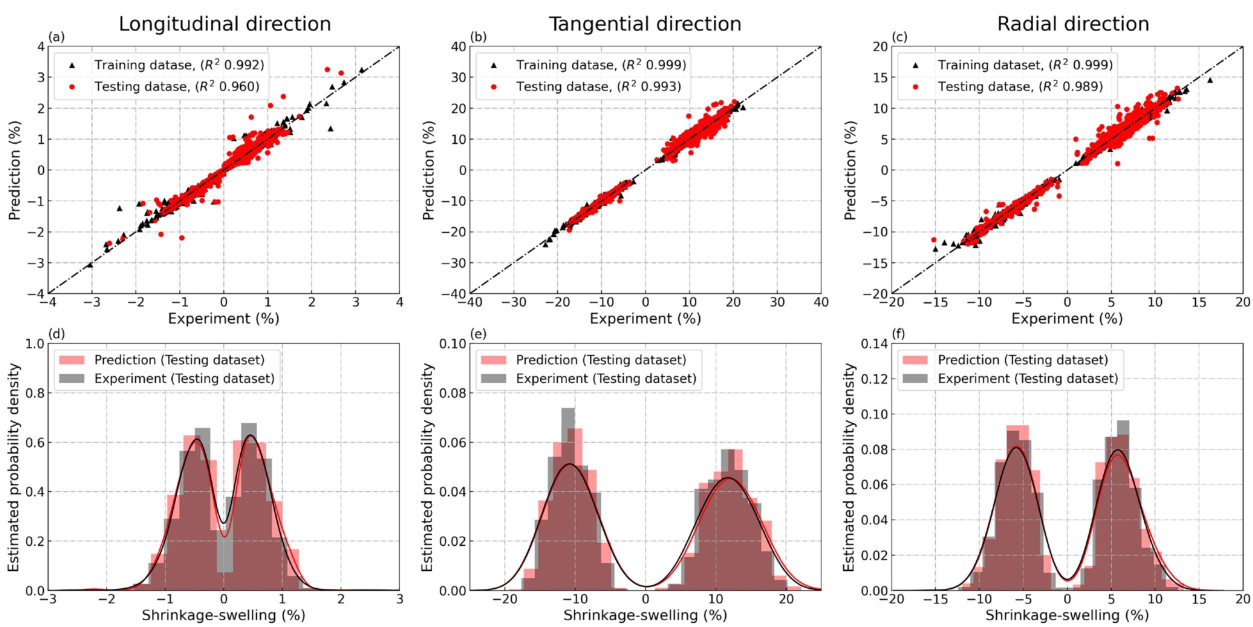

Fig. 4 Correlation between predicted and experimental swelling–shrinkage behavior in **a** longitudinal, **b** tangential, and **c** radial directions for both training and testing dataset together with distribution function and histograms of predicted and experimental in testing dataset in **d** longitudinal, **e** tangential, and **f** radial directions

high R^2 scores of 0.972 and 0.996 in predicting the density and fiber length in the testing dataset proved that the trained model was performing well without any overfitting. The R^2 scores for chemical features; holocellulose, lignin, and extractive content, were determined and are summarized in Table S3 (see six-observation threshold), demonstrating high predictive accuracy for both the training and testing datasets. Figure 6c, d compares the distribution function of the predicted density and fiber length in the testing dataset. Accordingly, a very high agreement between the predicted and experimental distribution function was observed.

Output neural network

The predicted chemical compositions and structural features were used individually as single concept features to predict mechanical properties. Tables S4 and 2 show the variation in prediction accuracy when restricted to a single feature for training and testing datasets, respectively. The results demonstrate that the model was unable to establish a functional correlation using only one feature, as evidenced by the low R^2 scores on the training dataset. Consequently, the model failed to generalize, leading to poor predictive performance on the testing dataset. This indicates that the interaction between shrinkage–swelling behavior and the mechanical properties of wood cannot be captured by an isolated variable.

As a next step, the predicted density and fiber length were considered as two parameters in concepts c were used as features of the output neural network. Training dynamics of the output neural network, described by the evolution of the loss function and training accuracy using density and fiber length as concept c features, are shown in Fig. 7a, b. It is observed that while the model continues to learn from the training dataset, a generalization gap persists due to the significantly lower and highly volatile performance on the validation data. Thus, it can be concluded that although the network was able to learn the patterns between the density and fiber length data with the mechanical properties, nonetheless, it could not well generalize its pattern. Figure 8 shows the correlation between predicted and experimental values in both the training and testing datasets for different mechanical properties. It is seen that the R^2 scores in the training dataset are in the range of 0.85 to 0.89, indicating that the neural network model has effectively learned the existing patterns and relationships present between features and targets in the training dataset. Regarding the testing dataset, the highest prediction accuracy was observed for hardness, where the model achieved an R^2 score of 0.840. Predictive accuracies for compressive strength and bending strength yielded R^2 scores of 0.750 and 0.753, respectively. Regarding impact resistance, a score of 0.764 was obtained, illustrating the model's ability to predict this property with reasonable precision. The lowest accuracy belonged to the prediction of the tensile strength, with an R^2 score of 0.700. The lower accuracy in predicting tensile strength can be attributed to the greater variability in tensile strength, compared to other properties, results in a wider distribution, making it more challenging for the ANN to generalize effectively.

To determine the effect of the observation threshold (see Sect. *Data organization*) on model performance, a sensitivity analysis was conducted, the results of which are shown in Table S3. No significant changes were observed in the predicted mechanical properties when using alternative thresholds of 4 and 8.

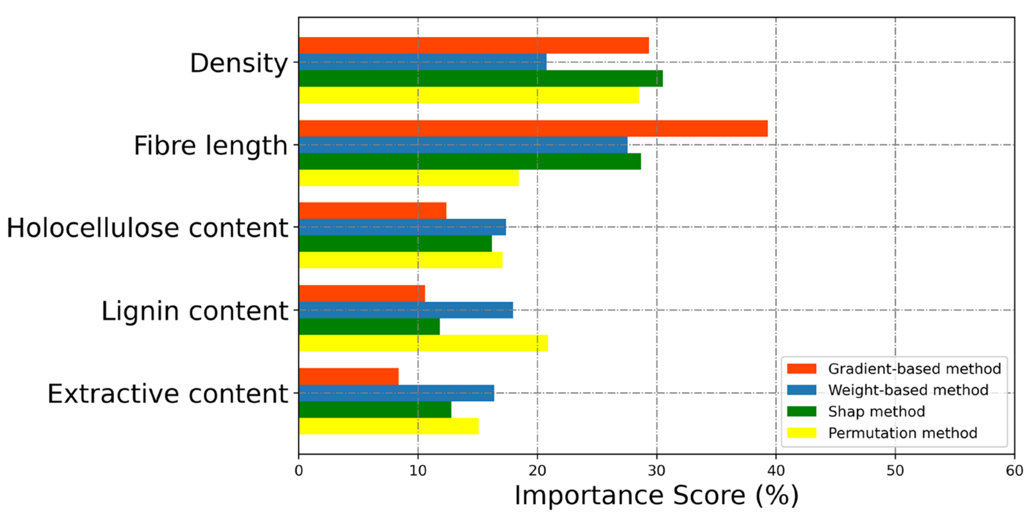

Fig. 5 Feature importance analysis on auxiliary neural network using different methods

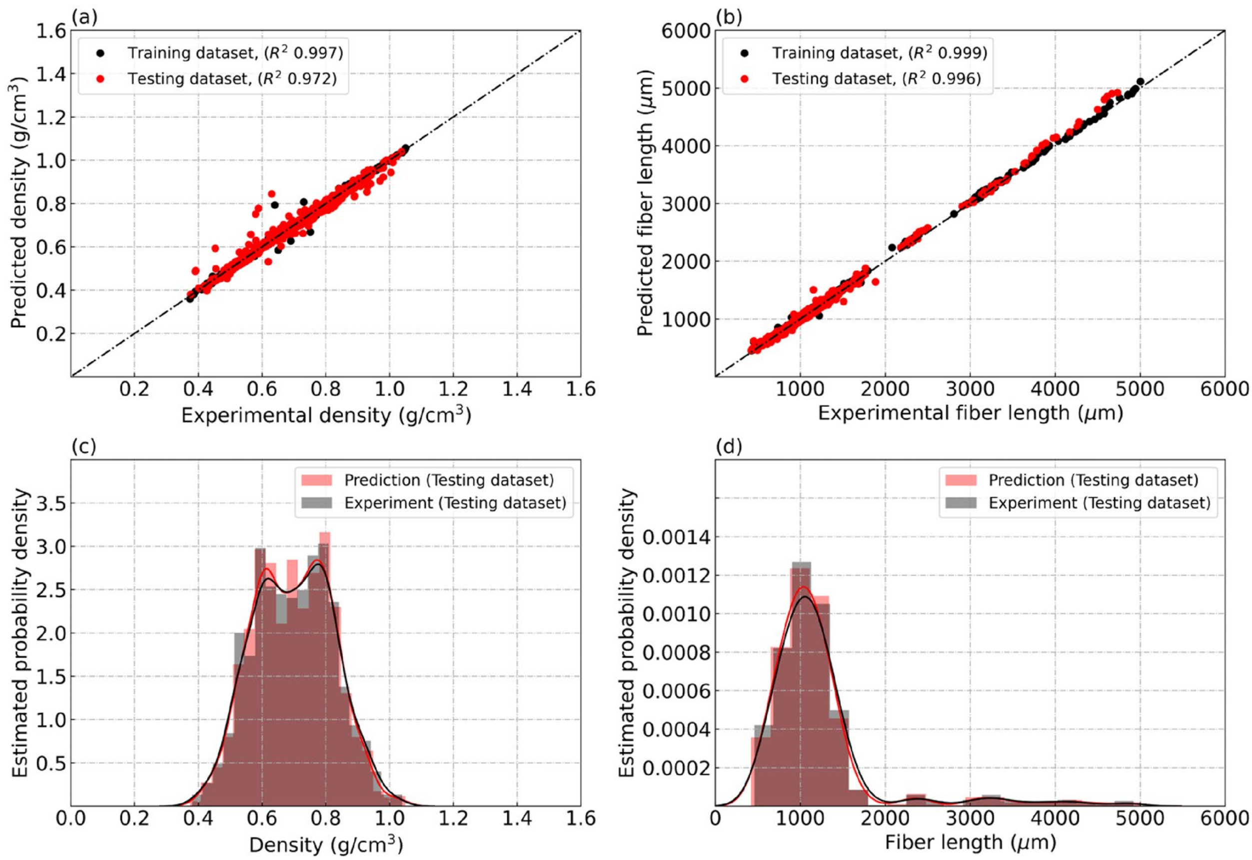

Fig. 6 Correlation between predicted and experimental values for **a** density and **b** fiber length for training and testing datasets, along with distribution functions and histograms of predicted and experimental values for **c** density and **d** fiber length in the testing dataset

To further investigate the effect of concept c features on final prediction accuracy, in addition to density and fiber length, holocellulose content, a third important parameter according to Fig. 5, was included in concept c features. The training dynamics of the output neural network are described by the evolution of the loss function and training accuracy, using density, fiber length, and holocellulose content as concept c features (Fig. 7c, d). Accordingly, it is observed that adding an additional concept c parameter led to a faster reduction in both the training and validation loss curves, narrowed the gap between them, and significantly enhanced the model's capability to learn and generalize underlying patterns. As a result of adding a new feature into concept c, a marked improvement in prediction accuracy of mechanical properties was also observed, as demonstrated in Fig. 9. The results showed a consistent improvement in the prediction accuracy across all mechanical properties for both the training and testing datasets, suggesting that implementing holocellulose content alongside density, fiber

Table 2 Summary of R^2 scores on the testing dataset using importance-driven bottleneck model, with individual concept c features

Properties	Concept c features				
	Extractive content	Lignin content	Holocellulose content	Density	Fiber length
Compression strength	0.29	0.46	0.52	0.36	0.18
Tensile strength	0.30	0.30	0.36	0.34	0.27
Bending strength	0.29	0.44	0.49	0.40	0.21
Hardness	0.48	0.68	0.75	0.65	0.10
Impact resistance	0.31	0.24	0.38	0.43	0.12

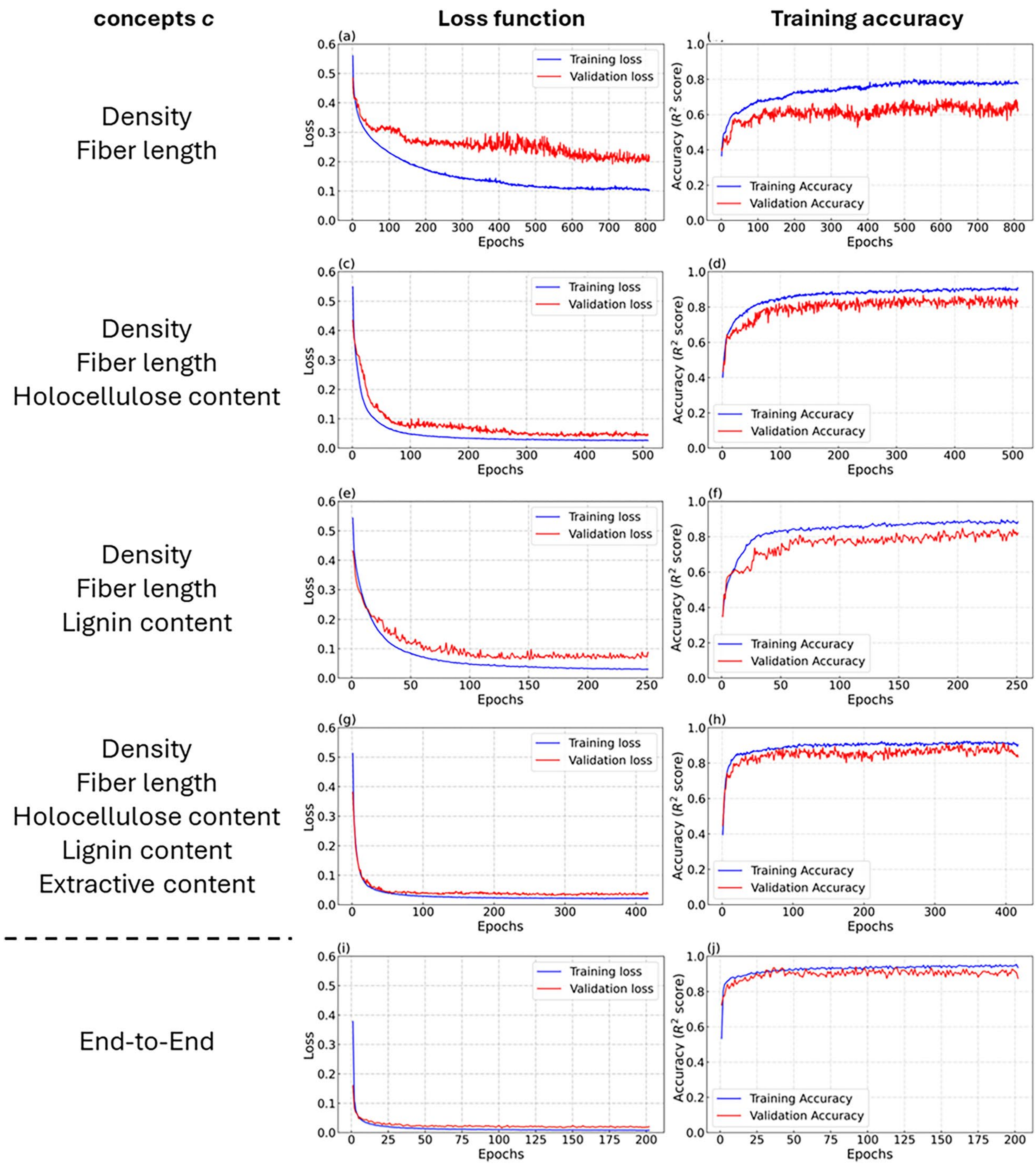

Fig. 7 Training dynamics of the output neural network of importance-driven bottleneck model, described by the evolution of the loss function and training accuracy, using **(a – h)** different concept c features, and **(i, j)** end-to-end model

length played an important role for capturing the complex behavior of the shrinkage–swelling behavior wood species and their mechanical properties.

Similar training dynamics were also obtained using density, fiber length, and lignin content as concept c features, as shown in Fig. 7e, f. This configuration resulted in high accuracy for the prediction of mechanical properties, as illustrated in Fig. S5. By incorporating all five chemical compositions and structural features, the gap between the training and validation curves narrowed further (see Fig. 7g, h), indicating that the model captured the underlying patterns more effectively and achieved more robust generalization. Ultimately, as shown in Fig. S6, this approach yielded the highest predictive accuracy for mechanical properties among all evaluated feature combinations.

To further extend the scope of the research, the analysis was continued by partitioning the data into two distinct datasets of softwood, hardwood, as classified in Fig. 1, and compared with all wood species combined, full dataset. This distinction was made to consider the potential effect of different fiber lengths (fibers, fiber–tracheids, and tracheids) of softwoods and hardwoods. To achieve this, the importance-driven bottleneck model utilized the same architecture introduced in Fig. 2; however, alternative hyperparameters were employed for hardwood and softwood due to the more limited data available, as summarized in Table S5.

The accuracy of the training dataset is shown in Fig. S7, indicating that, in general, as the number of features in concept c increased, the training process became more effective, leading to higher R^2 scores. This improvement was particularly evident in the full and softwood datasets, where a significant performance gain was observed when transitioning from two features to three or more across all mechanical properties, except for impact resistance (Fig. S7e), which showed performance increases across all three datasets. In addition, it was observed that for impact resistance, the training accuracy remained lower overall, particularly for the hardwood dataset, which exhibited the most limited predictive performance compared to other properties.

The R^2 scores of the testing dataset is shown in Fig. 10 for different mechanical properties under three classes of hardwood, softwood and full datasets. It is observed that in all three datasets, that incorporating a third concept c feature to the initial density and fiber length parameters enhanced the model's accuracy, especially pronounced in the softwood and full datasets. Beyond this point, predictive performance reached a stable plateau, suggesting a state of feature saturation. This indicates that the correlations between the shrinkage–swelling behavior of wood species and their mechanical properties are sufficiently captured by the first three concept c features. Among the evaluated metrics, the impact resistance of hardwood (Fig. S7e) consistently exhibited the lowest accuracy, likely due to a more complex or non-linear relationship that was less effectively captured during training.

Ablation analysis

To evaluate the relative importance of each chemical and physical parameter within the model, a feature ablation analysis was performed on both the training and testing datasets (Tables S6 and 3). This process involves systematically removing one concept c feature at a time and measuring the resulting drop in R^2 scores compared to the baseline model, which utilized all five chemical and structural features.

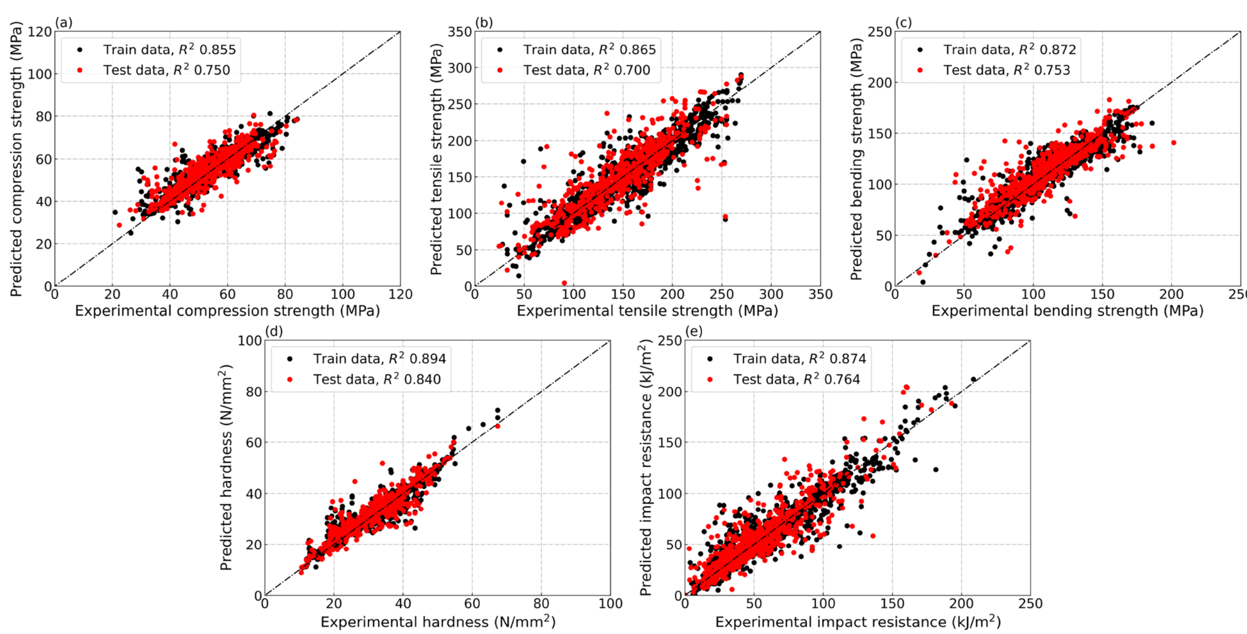

Fig. 8 Correlation between predicted and experimental values for **a** compression strength, **b** tensile strength, **c** bending strength, **d** hardness and **e** impact resistance, using density and fiber length as concept c features in importance-driven bottleneck model

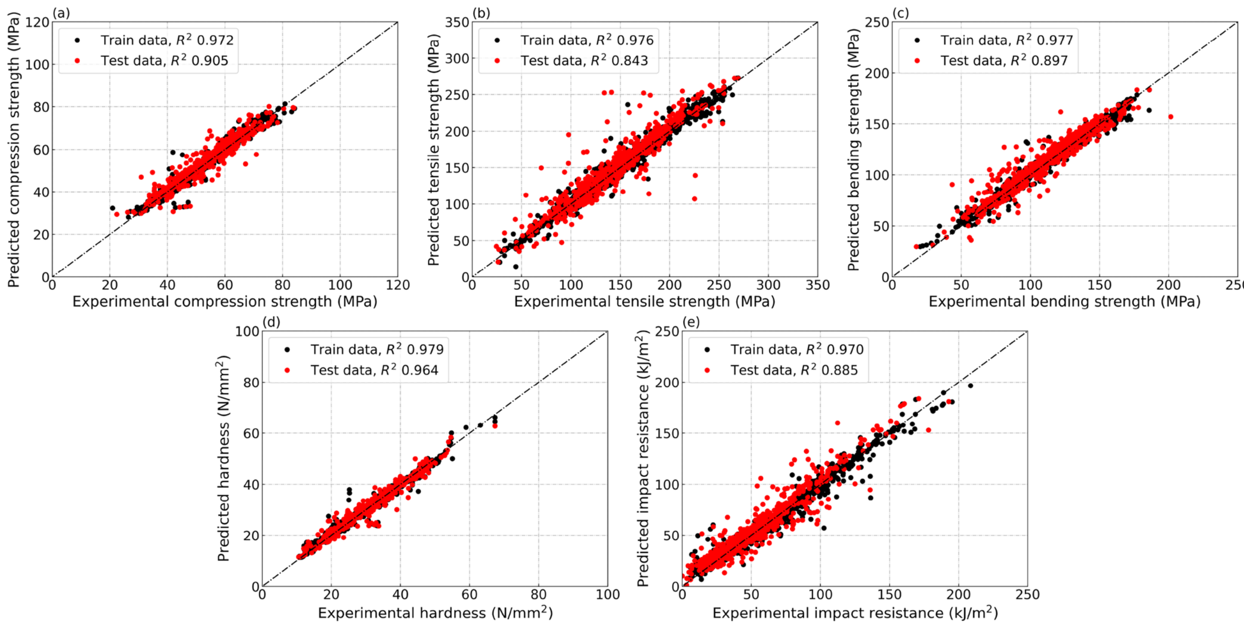

Fig. 9 Correlation between predicted and experimental values for **a** compression strength, **b** tensile strength, **c** bending strength, **d** hardness, and **e** impact resistance, using density, fiber length, and holocellulose content as concept c features in importance-driven bottleneck model

Within the training dataset, removing density resulted in the most significant and consistent drop in performance, identifying it as the most critical feature for predicting mechanical properties. Furthermore, the absence of fiber length in the concept c layer particularly reduced the accuracy of compression strength predictions. The results of testing dataset revealed that density is the most critical feature of predictive accuracy; its removal caused the most significant degradation across all mechanical properties, particularly for impact resistance, where the testing R^2 dropped by over 11.24%, and 6.98% drop in tensile strength. Fiber length also showed an influence specifically in compression strength, where its removal led to a 3.23% reduction in performance. Among chemical constituents, extractive content was found to be influential for impact resistance. It is noted that the higher baseline scores confirm that the synergy of all five chemical and structural features is necessary to achieve the model's peak predictive performance.

End-to-end neural network

In the end-to-end neural network, as described in Sect. *End-to-end model*, the shrinkage–swelling behavior of the wood species was directly employed to analyze its interrelation with the mechanical properties of the investigated wood species. The evolution of loss and accuracy during the training and validation phases for the end-to-end model is illustrated in Fig. 7i, j. As the number of epochs increased, the end-to-end model learned rapidly, reaching convergence significantly faster than the importance-driven bottleneck models. It achieved lower loss values and higher accuracy with a small gap between the training and validation curves, demonstrating that the model generalized effectively without signs of overfitting. Figure 11 illustrates the predicted mechanical properties with respect to their experimental counterparts in both training and testing datasets. It can be seen that the R^2 scores obtained from predicting the mechanical properties in the training dataset are very high, indicating that the model well-learned the patterns between the shrinkage–swelling behavior of wood species and their mechanical properties. Considering the testing dataset, it was observed that the R^2 scores were higher than those predicted by the importance-driven bottleneck models (Fig. 8, Fig. 9, Fig. S5); nonetheless, it became more comparable to the bottleneck configuration utilizing all five concept c features (see Fig. S6).

Tables S7 and 4 summarize the RMSE values of investigated mechanical properties using both the importance-driven bottleneck model with different concept c features and end-to-end neural network for training and testing datasets, respectively. RMSE values across both datasets decreased as the number of concept c features increased; notably, with five features, the RMSE was only marginally higher than that of the end-to-end neural network across various mechanical properties.

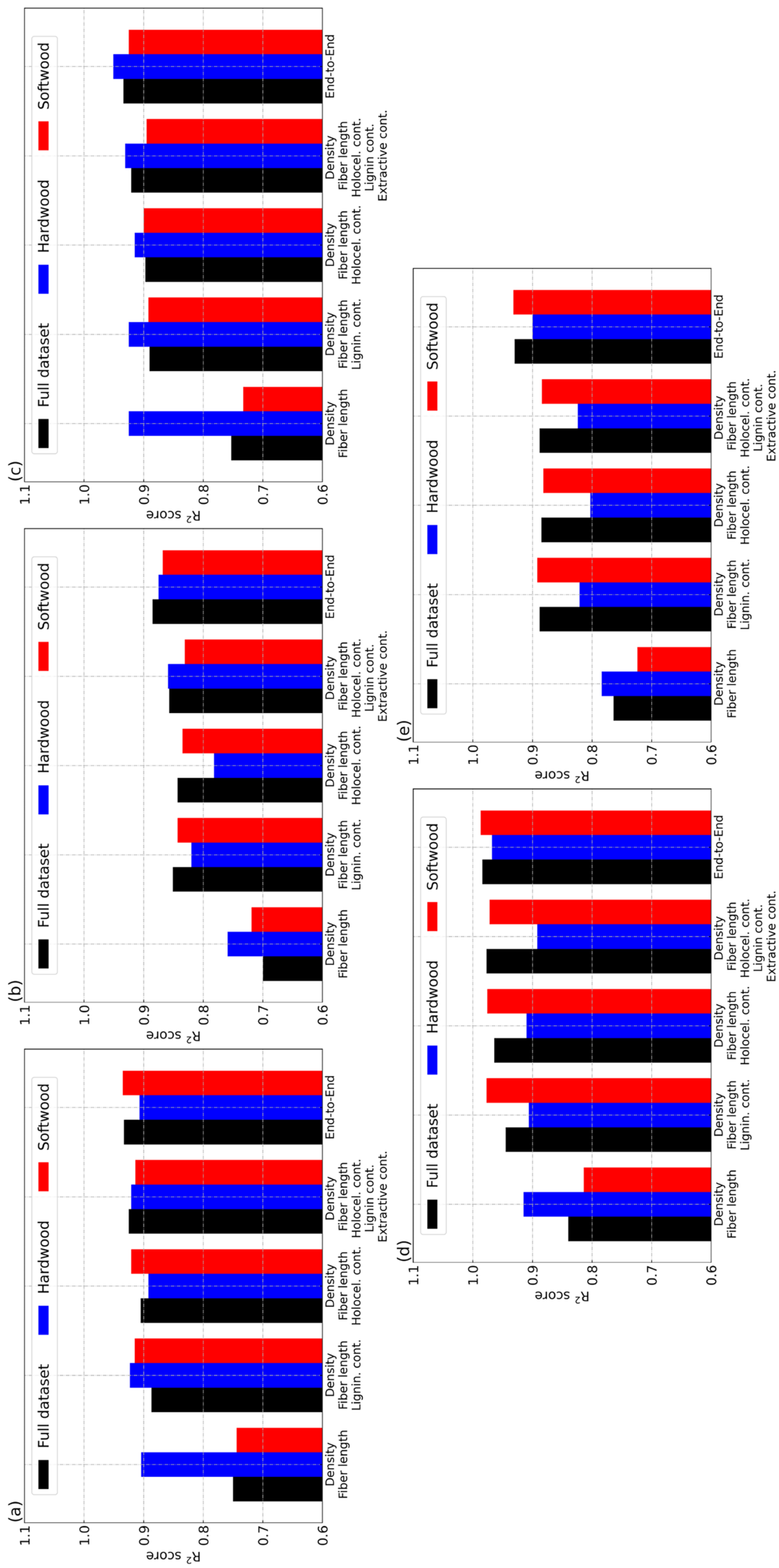

Fig. 10 Prediction accuracy, evaluated by the R^2 score in testing dataset using importance-driven bottleneck model, as a function of different concept c features and end-to-end model, considering full dataset, hardwood, and softwood for **a** compression strength, **b** tensile strength, **c** bending strength, **d** hardness, and **e** impact resistance

Table 3 Feature ablation analysis in testing dataset, using importance-driven bottleneck model

Properties	Baseline	No density		No fiber length		No holocellulose		No lignin		No extractive	
	R^2	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ
Compression strength	0.93	0.89	−4.30	0.90	−3.23	0.92	−1.08	0.92	−1.08	0.92	−1.08
Tensile strength	0.86	0.80	−6.98	0.87	1.16	0.85	−1.16	0.85	−1.16	0.85	−1.16
Bending strength	0.92	0.88	−4.35	0.91	−1.09	0.91	−1.09	0.92	0.00	0.91	−1.09
Hardness	0.98	0.95	−3.06	0.97	−1.02	0.98	0.00	0.97	−1.02	0.98	0.00
Impact resistance	0.89	0.79	−11.24	0.88	−1.12	0.88	−1.12	0.89	0.00	0.86	−3.37

The experimental distribution function of the investigated mechanical properties for testing dataset together with the predicted ones are shown in Fig. 12. By considering the form and modality of the distribution functions, it can be seen that the distribution function predicted by the end-to-end model and importance-driven bottleneck model with five concept c features more accurately follow the experimental distribution function in comparison with those predicted by importance-driven bottleneck model with density and fiber length as concept c features.

Accuracy–interpretability trade-off

The performance of the importance-driven bottleneck model and end-to-end neural network in predicting the different mechanical properties of the wood species, assessed by the R^2 scores as summarized in Table 2, Fig. 10, Table 4, and Fig. 12. The higher performance of the end-to-end neural network is attributed to the accuracy–interpretability trade-off with an inverse relationship between accuracy and interpretability [24, 29]. The end-to-end model outperforms importance-driven bottleneck model due to the direct optimization capacity and the absence of constraints for the flow of data during the propagation and backpropagation processes [32]. The performance of the importance-driven bottleneck model, however, was a function of both the number of features in concept c layer as well as the specific selection of those features. While configurations using only a single concept yielded insufficient accuracy (Table 2), two concept c features (density and fiber length) showed a significant improvement in prediction accuracies (Fig. 8), and it improved even more by expanding the concept c features (Fig. 10). This is because in the importance-driven bottleneck model, the flow of information is fully controlled by defining concept c features, guided by feature importance analysis as a supplementary tool. This mathematical strategy adds a new level of interpretability to machine learning models by forcing information to pass through a concept bottleneck, moving toward an interpretable and physics-inspired AI framework.

As a disadvantage, the importance-driven bottleneck model might lead to the loss of some relevant information for the final prediction [45]. In the case of defect-free wood specie, since their swelling–shrinkage behavior is closely related to their internal structure of the wood species, the predicted results by the end-to-end model were obtained by considering the interaction between the wood components beyond those selected for concept c features in the importance-driven bottleneck model, leading to higher performance of the end-to-end model.

The ablation analysis on concept c features (Table 3) indicated that density was the most critical concept c, and its absence led to the most significant decrease in prediction accuracy, which carries significant physical implications. It is well-documented that wood density is considered one of the most important characteristics affecting dimensional stability (linked to shrinkage and swelling properties) as well as mechanical properties of wood species [1, 6, 46]. The importance of fiber length for the mechanical properties of the wood species was the topic of some studies [47, 48] as well. Similar roles of fibers have been experimentally demonstrated also in other biological load-bearing materials, such as tissues and organs, where they significantly influence mechanical behavior [49]. Nevertheless, the effect of fiber length on dimensional stability of the wood species has not yet been fully investigated. By having all five structural and chemical features in concept c, the importance-driven bottleneck model reached a high performance nearly identical to the end-to-end neural network, indicating that once the bottleneck is informed by a sufficient range of informative features, the importance-driven bottleneck model becomes the superior choice, since it provides the predictive power of a 'black-box' system while maintaining the physics-inspired interpretability.

To examine the relationship between density and fiber length in terms of their combined effect on the mechanical properties, contour plots containing fiber length vs density vs mechanical properties are plotted in Fig. 13. As seen in the figure, no specific trend is observed between fiber length and density with

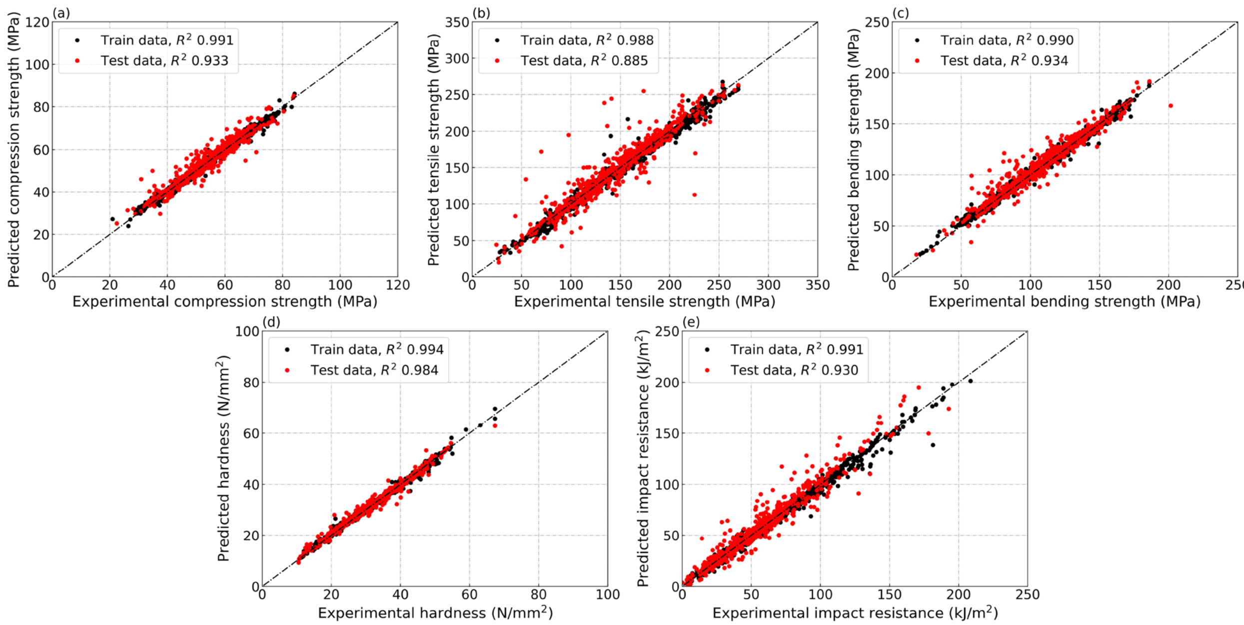

Fig. 11 Correlation between predicted and experimental values for **a** compression strength, **b** tensile strength, **c** bending strength, **d** hardness and **e** impact resistance, using end-to-end model

Table 4 Summary of the RMSE values, in testing dataset, using importance-driven bottleneck model and end-to-end model

Properties	Importance-Driven Bottleneck				End-to-End
	Den, Fib	Den, Fib, Hol	Den, Fib, Lig	Den, Fib, Hol, Lig, Ext	
Compression strength (MPa)	4.97	3.06	3.34	2.72	2.56
Tensile strength (MPa)	24.77	17.91	17.44	17.11	15.37
Bending strength (MPa)	13.22	8.54	8.80	7.46	6.82
Hardness (N/mm^2)	3.50	1.67	2.05	1.34	1.12
Impact resistance (kJ/m^2)	14.61	10.19	10.06	10.09	7.94

respect to mechanical properties, which possibly is due to their complex nonlinear relationship with the mechanical properties.

Conclusions

In this study, a dataset containing the physical, chemical, mechanical properties, and structural characteristics of defect-free specimens of 57 wood species, both softwoods and hardwoods, from Central Europe were utilized to comprehensively explore correlations between the mentioned properties. In addition, this work aimed to address the lack of interpretability in typical neural network models used in existing research. The following conclusions were drawn:

1. The importance-driven bottleneck model was proposed by modifying the concept bottleneck model, wherein the concepts c features were determined through the feature importance analysis of a trained auxiliary neural network. The selected properties for concepts c were assigned as targets for the input neural network, and their predicted values were used as features for the output neural network.
2. The shrinkage–swelling behavior of wood species was predicted with high accuracy using five structural and chemical characteristics: density, fiber length, lignin content, holocellulose content, and extractive content. Feature importance analysis was performed using gradient-based, weight-based,

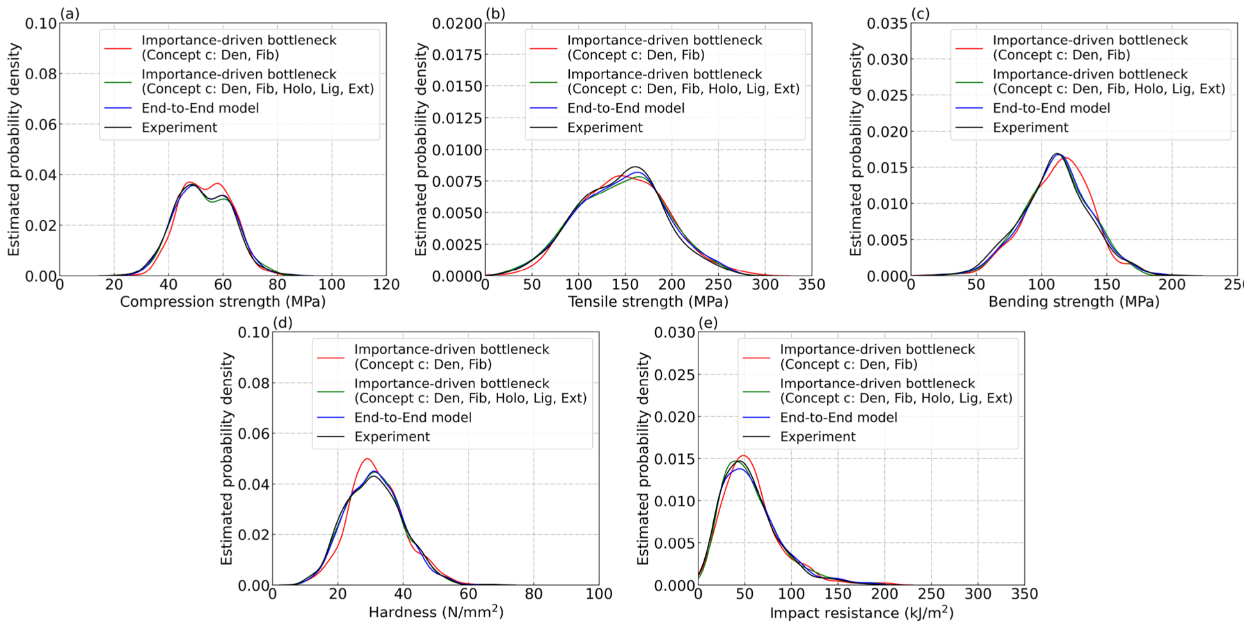

Fig. 12 Comparison of experimental with predicted distribution functions from the end-to-end model and importance-driven bottleneck model for **a** compression strength, **b** tensile strength, **c** bending strength, **d** hardness, and **e** impact resistance

SHAP, and permutation approaches. Across all evaluation metrics, density and fiber length were found to be the most important features, followed by holocellulose content, lignin content, and extractive content, which exhibited comparable levels of importance.

3. The input neural network of the importance-driven bottleneck model could accurately unravel the relationships between density, fiber length, lignin content, holocellulose content, and extractive content. based on shrinkage and swelling properties in all three directions.
4. For output neural networks, mechanical properties (bending, tensile, and compressive strength, hardness and impact resistance) of various wood species were predicted, by incorporating different chemical and structural characteristics into concepts c layer. The most significant performance gains occurred when transitioning from a single feature to the two features of density and fiber length, achieving testing R^2 scores above 0.70. The inclusion of a third chemical component further elevated these scores to over 0.85.
5. In comparison with the end-to-end neural network, the prediction accuracy of the importance-driven bottleneck model, using all five chemical and structural characteristics, was slightly lower; nevertheless, it offered higher interpretability.
6. Similar to full datasets, the softwood dataset also showed significant accuracy gains by adding a new feature into two concept c features of density and fiber length. In this case, the hardwood dataset for impact resistance exhibited the lowest performance due to more complex property relationships and data limitations.
7. Feature ablation analysis showed that density was the primary driver of model accuracy, as its removal caused the most performance degradation. Fiber length showed a pronounced influence on compression strength while extractive content on impact resistance.
8. The importance-driven bottleneck model establishes an interpretable, physics-inspired AI framework forcing information through a concept bottleneck. This approach allows the model to serve as a virtual laboratory to isolate and examine specific physical factors, a capability that standard regression and the end-to-end models cannot provide.
9. Finally, future work could investigate species-level generalization by extending the dataset to include a broader range of wood species.

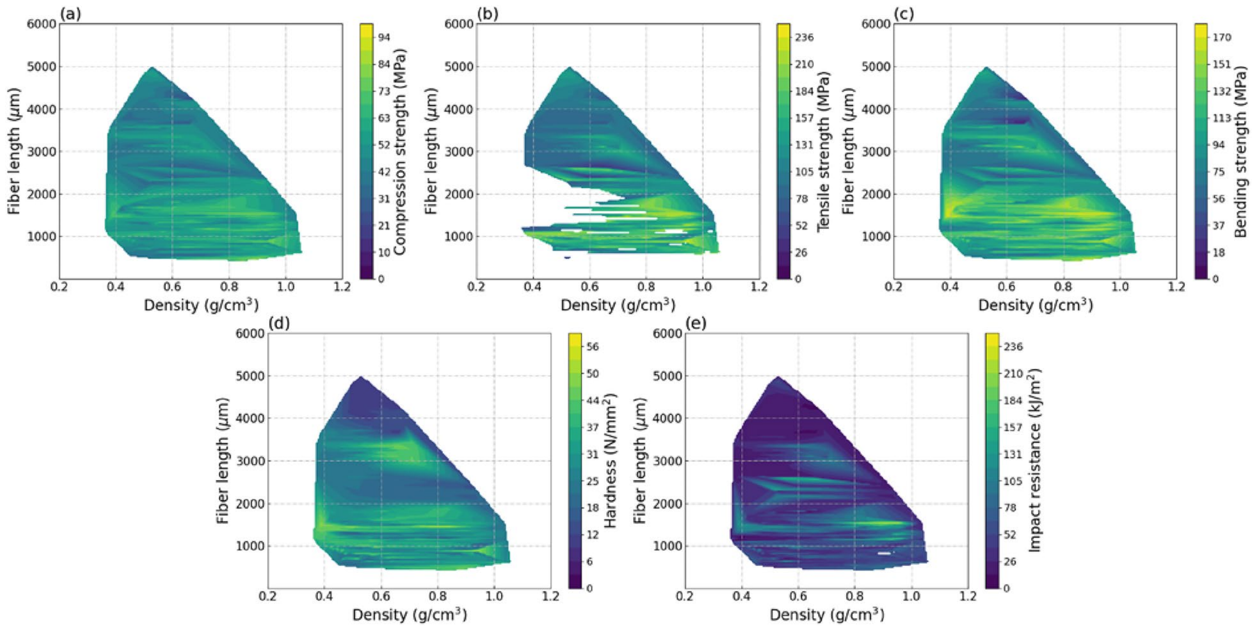

Fig. 13 Contour/response surface plot of **a** compression strength, **b** tensile strength, **c** bending strength, **d** hardness and **e** impact resistance vs fiber length and density

Appendix

Experimental procedure

Chemical composition assessment

Density measurements The determination of the wood density was carried out in accordance with DIN 52182 with reduced sample dimensions of $20 \times 10 \times 10\ mm^3$ (axial, radial, tangential) compared to those prescribed in the standard. The sample dimensions were measured with the help of a dial gauge with an accuracy of 1 μm, and the mass was determined to an accuracy of 1 μg using a laboratory balance. The density was calculated in standard climate (in accordance with DIN 50014 [50]; 20°C, 65% relative humidity; wood density) and in a kiln-dry state (dry wood density, see [51]) in accordance with the formula in the standard.

Extractive Material from several individuals per wood species was ground in two cycles to a particle size of 80 μm and in a speed extractor from the company "Büchi" with the solvents cyclohexane / ethanol (95%) (2:1), ethanol (95%) / water (1 :1) and water were extracted and, after subsequent drying of the extract substances, their mass is determined gravimetrically.

Lignin The lignin content was determined analogously to T222 and UM 250 (TAPPI) [52], after the extractive substances had been separated off (see method for determining the extractive content). For this purpose, after 2 h of hydrolysis with sulfuric acid (H_2SO_4; 72%) at 20°C and 4 h of hydrolysis with sulfuric acid (H_2SO_4; 3%) at 100°C, the acid-soluble lignin was measured photometrically at 205 nm and Klason lignin gravimetrically after previous drying at 105°C over 24 h.

Fiber length To measure the fiber or tracheid length, wood samples were macerated in a mixture of equal volumes of hydrogen peroxide (H_2O_2 30%), water and acetic acid (96%) for several hours (until the fiber composite disintegrated) [53]. Subsequently, the fiber lengths were measured under a transmitted light microscope using the image analysis software "ImageJ" [54].

Shrinkage and swelling measurements

The determination of the maximum swelling and shrinkage dimensions was based on DIN 52184 [55], the sample dimensions were reduced to $20 \times 10 \times 10\ mm^3$ (axial, radial, tangential) compared to the prescribed dimensions and were measured with the help of a dial gauge with an accuracy of 1 μm. Drying was carried out at 103°C until constant weight was reached, and the samples were then cooled in a desiccator using silica gel as a drying agent. Humidification to determine the maximum swelling dimension was also carried out in a desiccator and in analogy to the accelerated procedure described in the standard under vacuum.

Mechanical property measurements

Compression experiment The compression test parallel to the fiber direction was carried out under standard conditions (according to DIN 50014; 20°C, 65% relative humidity) in accordance with DIN 52185 [56]. The specimen dimensions were reduced to $20 \times 10 \times 10\,\mathrm{mm}^3$ (axial, radial, tangential) compared to those specified in the standard. The universal testing machine Z020 from Zwick Roell was used and the specimens were loaded within the required test time of 1.5±0.5 min (with test speed varying depending on the wood species) until the maximum force was reached.

Tensile experiment Reduced specimen dimensions compared to DIN 52188 [57] were used to measure the tensile strength. The specimens were waisted on all sides and had dimensions of 8x4 mm in the area of the smallest specimen cross section. The annual rings were aligned along the longer side of the specimen cross section to maximize the number of annual rings in the test specimen. The tensile strength (under standard climate conditions according to DIN 50014; 20°C, 65% relative humidity) was calculated from the maximum force and specimen cross section in accordance with the formula in the standard.

Bending experiment The bending test was carried out in accordance with DIN 52186 [58] as a three-point bending test on the "Z020" universal testing machine from Zwick Roell, the specimen deformation was determined two-dimensionally and contact-free with an accuracy of 0.3 µm using the "laserXtens" laser extensometer from Zwick Roell. Due to the reduced specimen dimensions of $180 \times 10 \times 10$ mm (axial, radial, tangential) compared to those specified in the above-mentioned standard, the support spacing was set at 150 mm in accordance with the required minimum length of 15 times the specimen height, and the radius of the steel support rollers was reduced to 7.5 mm. Depending on the wood species, the test speed was selected differently to achieve the maximum force within the required 1.5±0.5 min. The samples were loaded evenly and without impact in tangential direction (horizontal growth rings) under standard climatic conditions (in accordance with DIN 50014; 20°C, 65% relative humidity) until breakage and the bending strength was calculated in accordance with the calculation rules in the standard using the sample dimensions determined prior the test.

Impact experiment The impact bending test was carried out on the "HIT 50 P" testing machine from Zwick Roell in accordance with DIN 52189 [59] in a standard climate (in accordance with DIN 50014; 20°C, 65% relative humidity). Deviations from the standard were the smaller specimen size of $80 \times 10 \times 10\ \mathrm{mm}^3$ (axial, radial, tangential), the reduced distance between the abutments from 210 to 62 mm, and a smaller pendulum impact tester with a working capacity of 50 joules. Once the specimen dimensions have been determined, the specimen is placed to be struck with horizontal growth rings/tangential loading direction of the pendulum impact tester, and the impact energy is calculated in accordance with the formula of the standard.

Hardness experiment The Brinell hardness was measured analogue to ÖNORM EN 1534 under standard climate conditions (according to DIN 50014; 20°C, 65% relative humidity). The "Z020" universal testing machine from Zwick Roell was used with the hardness tester from the same company. The measurements of the diameters of the indentations remaining on the specimen after the test, which are necessary for the conventional calculation of the Brinell hardness, were carried out using the "Stemi 2000" stereomicroscope from "Zeiss" with a connected "Axiocam ERc 5" camera system from the same company and the "ImageJ" image analysis software [54] after prior calibration.

Data standardization

Data standardization, is one of the stages of data preparation used in machine learning methods, which involves transforming data into a consistent format, thereby guaranteeing uniformity across diverse datasets or variables. By performing the standardization, the values are centered around the mean with a unit standard deviation. Standardization can enhance the performance of machine learning models, mitigate the influence of outliers, and maintain data consistency across various scales:

$$X' = \frac{X - \mu}{\sigma} \tag{1}$$

where μ and σ are the mean and standard deviation of the values, respectively.

Cross-validation and hyperparameters

Upon searching for the best estimator, there is always a risk of overfitting during the learning step, which reduces the generality of the model and lowers prediction accuracy. To avoid overfitting, a cross-validation procedure was applied, in which the training set is split into k smaller sets, called k-fold CV. In this procedure, the model gets trained using k-1 folds of the training data and the obtained results are validated using the remaining fold, validation set. The final performance computed by the k-fold cross-validation is the average of all the values obtained of all the loops. The hyperparameters are determined according to the empirical observations or domain of knowledge. Selecting the best hyperparameters manually could take a considerable amount of time and resources. In this case, one of the solutions to find the optimal values of hyperparameters is via using hyperparameters tuning methods, e.g., grid search cross-validation method, in short GridSearchCV method, which automates the tuning process of hyperparameters. GridSearchCV method performs loops among a range of predefined hyperparameters and implements them into the training dataset. Afterward, the accuracy of the fitting is determined and the best parameters from the suggested hyperparameters are selected.

Prediction accuracy assessment

The assessment of the prediction accuracy was done in several approaches; determining the error of the predictions, calculating the prediction accuracy, and via the statistical analysis. The error in the predictions for all the investigated properties were determined based on the root mean squared error (RMSE), while the prediction accuracy was assessed using the coefficient of determination. RMSE is defined as the square root of all the squares of the distance divided by the total number of points, n, as shown in the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(y_{\exp} - y_{\mathrm{pre}})^2} \tag{2}$$

The coefficient of determination, R^2, is calculated according to Eq. 3, which shows in a range of 0.0, for the case that the model does not predict accurately, to 1.0, for a fully accurate prediction:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_{\exp} - y_{\mathrm{pre}})^2}{\sum_{i=1}^{N}(y_{\exp} - y_{\mathrm{mean}})^2} \tag{3}$$

where y_{mean} stands for the mean experimental value.

The probability distribution of a variable can be described using a distribution function represented by the estimated probability density, which determines the likelihood of a given observation falling within a certain range of values. In statistics, among the different techniques to estimate probability density, kernel density estimation is one of the common methods. Kernel density estimation method is a non-parametric method that estimates the probability density function of a random variable. In this work, the Gaussian kernel was used to plot the distribution function for the experiments and the predictions. To evaluate the distribution functions, they were visually inspected, and their mode values were determined. The mode is the most frequent value in a dataset and can be estimated from the peak of the probability density curve.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s10086-026-02266-9.

Supplementary Material 1.

Acknowledgements
Not applicable.

Funding
Open access funding provided by Swiss Federal Institute of Technology Zurich.

Data availability
Data will be made available on request.

Declarations

Competing interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author details
[1]Wood Materials Science, Institute for Building Materials, ETH Zurich 8093, Switzerland. [2]Wood Materials Science, Institute of Wood Technology and Renewable Materials, Department of Material Science and Process Engineering, University of Natural Resources and Life Sciences Vienna, 3430 Tulln a.d. Donau, Austria. [3]Division of Structural Engineering, Chalmers University of Technology, Gothenburg, Sweden. [4]Imperial College, London, United Kingdom. [5]Laboratory for Cellulose & Wood Materials, Group WoodTec, Empa, Swiss Federal Laboratories for Materials Science and Technology, Uberlandstr. 129, 8600 Dubendorf, Switzerland.

Received: 14 November 2025 Accepted: 19 March 2026
Published online: 18 April 2026

References

1. Wangaard F (1950) The mechanical properties of wood. The mechanical properties of wood https://www.cabidigitallibrary.org/doi/full/10.5555/19500602031, Accessed 15 March 2026
2. De Borst K, Jenkel C, Montero C et al (2013) Mechanical characterization of wood: An integrative approach ranging from nanoscale to structure. Computers & Structures 127:53–67. https://doi.org/10.1016/j.compstruc.2012.11.019
3. Shams M, Yano H (2004) Compressive deformation of wood impregnated with low molecular weight phenol formaldehyde (PF) resin II: effects of processing parameters. J Wood Sci 50(4):343–350. https://doi.org/10.1007/s10086-003-0571-5
4. Wool R, Sun X (2005) Bio-based polymers and composites. Elsevier Amsterdam Netherlands. https://doi.org/10.1016/b978-0-12-763952-9.x5000-x
5. Movahedi-Rad AV, Ritter M, Kindler RO et al (2024) Reconstructing poplar wood into a high-performance fiber-reinforced biocomposite. ACS Sustain Chem Eng 12(40):14694–14703. https://doi.org/10.1021/acssuschemeng.4c04652
6. Rowell R, Banks W (1985) Water repellency and dimensional stability of wood. US Department of Agriculture, Forest Service, Forest Products

Laboratory, Madison, WI, Gen Tech Rep FPL 50:1–24. https://research.fs.usda.gov/treesearch/8912, Accessed 15 March 2026
7. Donaldson L (2008) Microfibril angle: measurement, variation and relationships–a review. Iawa Journal 29(4):345–386. https://brill.com/view/journals/iawa/29/4/article-p345_1.xml, Accessed 15 March 2026
8. Thibaut B, Gril J (2021) Tree growth forces and wood properties. Peer Community Journal 1:e46. https://doi.org/10.24072/pcjournal.48
9. Rowell R, Pettersen R, Han JS et al (2005) Cell wall chemistry. Handbook of wood chemistry and wood composites 2:33–72. https://doi.org/10.1201/9780203492437
10. Mvondo R, Meukam P, Jeong J et al (2017) Influence of water content on the mechanical and chemical properties of tropical wood species. Results Phys 7:2096–2103. https://doi.org/10.1016/j.rinp.2017.06.025
11. Ozyhar T, Hering S, Niemz P (2012) Moisture-dependent elastic and strength anisotropy of european beech wood in tension. J Mater Sci 47:6141–6150. https://doi.org/10.1007/s10853-012-6534-8
12. Gennari E, Picchio R, Monaco A (2021) Water absorption behaviour and dimensional stability of a thermally modified tropical hardwood (*Triplochiton scleroxylon* K. Schum). Environmental Sciences Proceedings 13(1):1. https://doi.org/10.3390/iecf2021-10822
13. Kretschmann D, Green D (1996) Modeling moisture content-mechanical property relationships for clear southern pine. Wood and fiber science 28(3):320–337. https://wfs.swst.org/index.php/wfs/article/view/216, Accessed 15 March 2026
14. Steele J, Ifju G, Johnson J (1976) Quantitative characterization of wood microstructure. J Microsc 107(3):297–311. https://doi.org/10.1111/j.1365-2818.1976.tb02450.x
15. Cook D, Chiu C (1997) Predicting the internal bond strength of particleboard, utilizing a radial basis function neural network. Eng Appl Artif Intell 10(2):171–177. https://doi.org/10.1016/s0952-1976(96)00068-1
16. Esteban L, Fernández F, Palacios P (2009) Moe prediction in abies pinsapo boiss. timber: Application of an artificial neural network using non-destructive testing. Computers & Structures 87(21-22):1360–1365. https://doi.org/10.1016/j.compstruc.2009.08.010
17. Esteban L, Fernández F, Palacios P (2011) Prediction of plywood bonding quality using an artificial neural network. Holzforschung 65:209–214. https://doi.org/10.1515/hf.2011.003
18. Movahedi Rad AV, Ritter M, Colmant A et al (2026) combined experimental and machine learning study on the interplay between delignification and mechanical properties for improved poplar wood reconstruction. ACS Applied Materials & Interfaces 18(14):5882–5895. https://doi.org/10.1021/acsami.5c20194
19. Goyal S (2013) Artificial neural networks in vegetables: a comprehensive review. Sci J Crop Sci 2(7):75–94. https://doi.org/10.5815/ijigsp.2014.05.07
20. Kuvendziev S, Lisichkov K, Zeković Z et al (2014) Artificial neural network modelling of supercritical fluid CO_2 extraction of polyunsaturated fatty acids from common carp (Cyprinus carpio L.) viscera. J Supercr Fluids 92:242–248. https://doi.org/10.1016/j.supflu.2014.06.007
21. Alves D, Tomaz R, Laurindo B et al (2017) Artificial neural network for prediction of the area under the disease progress curve of tomato late blight. Scientia Agricola 74:51–59. https://doi.org/10.1590/1678-992x-2015-0309
22. Miguel E, Melo R, Serenini Junior L et al (2018) Using artificial neural networks in estimating wood resistance. Maderas Ciencia y tecnología 20(4):531–543. https://doi.org/10.4067/s0718-221x2018005004101
23. Fathi H, Nasir V, Kazemirad S (2020) Prediction of the mechanical properties of wood using guided wave propagation and machine learning. Constr Build Mater 262:120848. https://doi.org/10.1016/j.conbuildmat.2020.120848
24. He C, Ma M, Wang P (2020) Extract interpretability-accuracy balanced rules from artificial neural networks: A review. Neurocomputing 387:346–358. https://doi.org/10.1016/j.neucom.2020.01.036
25. van Blokland J, Nasir V, Cool J et al (2021) Machine learning-based prediction of surface checks and bending properties in weathered thermally modified timber. Constr Build Mater 307:124996. https://doi.org/10.1016/j.conbuildmat.2021.124996
26. Nasir V, Fathi H, Fallah A et al (2021) Prediction of mechanical properties of artificially weathered wood by color change and machine learning. Materials 14(21):6314. https://doi.org/10.3390/ma14216314
27. Tian Y, Zalesky A (2021) Machine learning prediction of cognition from functional connectivity: Are feature weights reliable? Neuroimage 245:118648. https://doi.org/10.1101/2021.05.27.446059
28. Karpatne A, Atluri G, Faghmous JH et al (2017) theory-guided data science: A new paradigm for scientific discovery from data. IEEE Trans Knowl Data Eng 29(10):2318–2331. https://doi.org/10.1109/TKDE.2017.2720168
29. Koh P, Nguyen T, Tang Y et al (2020) Concept bottleneck models. International conference on machine learning 119:5338–5348
30. Kumar N, Berg A, Belhumeur P, et al (2009) Attribute and simile classifiers for face verification. 2009 IEEE 12th international conference on computer vision :365–372. https://doi.org/10.1109/iccv.2009.5459250
31. Lampert C, Nickisch H, Harmeling S (2009) Learning to detect unseen object classes by between-class attribute transfer. 2009 IEEE conference on computer vision and pattern recognition :951–958. https://doi.org/10.1109/cvprw.2009.5206594
32. Murphy K (2012) Machine learning: a probabilistic perspective. MIT press, Cambridge, MA
33. Awad M, Khanna R (2015) Efficient learning machines: theories, concepts, and applications for engineers and system designers. Springer nature Berkeley CA. https://doi.org/10.1007/978-1-4302-5990-9
34. Collobert R, Weston J, Bottou L et al (2011) Natural language processing (almost) from scratch. J Mach Learn Res 12:2493–2537
35. Mnih V, Kavukcuoglu K, Silver D et al (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533. https://doi.org/10.1038/nature14236
36. Wang H, Li B, Xuan F (2022) Fatigue-life prediction of additively manufactured metals by continuous damage mechanics (cdm)-informed machine learning with sensitive features. Int J Fatigue 164:107147. https://doi.org/10.1016/j.ijfatigue.2022.107147
37. Ehrhart T, Palma P, Schubert M, et al (2022) Predicting the strength of european beech (*Fagus sylvatica L.*) boards using image-based local fibre direction data. Wood Science and Technology 56(1):123–146. https://doi.org/10.1007/s00226-021-01347-w
38. Lukovic M, Ciernik L, Müller G et al (2024) Probing the complexity of wood with computer vision: from pixels to properties. J R Soc Interface 21(213):20230492. https://doi.org/10.1098/rsif.2023.0492
39. Dayhoff J, DeLeo J (2001) Artificial neural networks: opening the black box. Cancer: Interdisciplinary International Journal of the American Cancer Society 91(S8):1615–1635. https://doi.org/10.1002/1097-0142(20010415)91:8+<1615::aid-cncr1175>3.0.co;2-l
40. Glasmachers T (2017) Limits of end-to-end learning. Asian Conference on Machine Learning 77:17–32
41. Fong R, Vedaldi A (2017) Interpretable explanations of black boxes by meaningful perturbation. Proceedings of the IEEE international conference on computer vision :3429–3437. https://doi.org/10.1109/iccv.2017.371
42. da Costa N, de Lima M, Barbosa R (2021) Evaluation of feature selection methods based on artificial neural network weights. Expert Syst Appl 168:114312. https://doi.org/10.1016/j.eswa.2020.114312
43. Zhou W, Yan Z, Zhang L (2024) A comparative study of 11 non-linear regression models highlighting autoencoder, dbn, and svr, enhanced by shap importance analysis in soybean branching prediction. Sci Rep 14(1):5905. https://doi.org/10.1038/s41598-024-55243-x
44. Saarela M, Jauhiainen S (2021) Comparison of feature importance measures as explanations for classification models. SN Applied Sciences 3(2):272. https://doi.org/10.1007/s42452-021-04148-9
45. Zhai P, Zhang S (2022) Adversarial information bottleneck. IEEE Trans Neural Netw Learn Syst 35(1):221–230. https://doi.org/10.1109/TNNLS.2022.3172986
46. Christoforo A, Almeida T, Almeida D et al (2016) Shrinkage for some wood species estimated by density. Int J Mater Eng 6(2):23–27. https://doi.org/10.5923/j.ijme.20160602.01
47. Pittermann J, Sperry J, Wheeler J et al (2006) Mechanical reinforcement of tracheids compromises the hydraulic efficiency of conifer xylem. Plant, Cell Environ 29(8):1618–1628. https://doi.org/10.1111/j.1365-3040.2006.01539.x
48. Kiaei M, Sadegh A, Moya R (2013) Site variation of tracheid features and static bending properties in Pinus eldarica wood. Cellul Chem Technol 47(1):49–59
49. Elices M (2000) Structural biological materials: design and structure-property relationships. Elsevier, Oxford, UK

50. DIN 50014, (1985) Standard atmospheres for conditioning and/or testing - Specifications. Standard, Deutsches Institut für Normung, Berlin, Germany
51. Niemz P (1993) Physik des Holzes und der Holzwerkstoffe. DRW-Verlag Weinbrenner GmbH & Co, KG
52. Standard T (2002) Acid-insoluble lignin in wood and pulp. Technical association of the pulp and paper Industry (TAPPI), TAPPI Press, Atlanta, GA
53. Gärtner H, Schweingruber FH (2013) Microscopic Preparation Techniques for Plant Stem Analysis. Verlag Kessel, Remagen, Germany
54. Schneider CA, Rasband W, Eliceiri K (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9(7):671–675. https://doi.org/10.3410/f.717968566.793456800
55. DIN 52184, (1979) Testing of wood - determination of swelling and shrinkage. Standard, Deutsches Institut für Normung, Berlin, Germany
56. DIN 52185, (1976) Testing of wood - compression test parallel to grain. Standard, Deutsches Institut für Normung, Berlin, Germany
57. DIN 52188, (1979) Testing of wood - determination of impact bending strength. Standard, Deutsches Institut für Normung, Berlin, Germany
58. DIN 52186, (1978) Testing of wood - bending test. Standard, Deutsches Institut für Normung, Berlin, Germany
59. DIN 52189–1, (1981) Testing of wood - sampling and general requirements for physical and mechanical tests. Standard, Deutsches Institut für Normung, Berlin, Germany

Publisher's Note