A two-stage feature selection method for neural network predictive models for AGV[1⋆]

Olena Pavliuk[1,2,*,†], Myroslav Mishchuk[2,†], Mykola Medykovskyy[2,†] and Rafal Cupek[1,†]

[1] *Silesian University of Technology, Gliwice 44-100, Poland*
[2] *Lviv Polytechnic National University, Lviv 79000, Ukraine*

Abstract

This study addresses the challenge of enhancing the performance of predictive artificial neural network models through effective feature selection. We introduce a novel two-stage feature reduction method based on the synergy of correlation analysis and the Random Forest (RF) algorithm. This method is based on the ability to identify correlational interdependencies between the characteristics of each observation, with further feature importance selection using the RF algorithm. The efficiency of the proposed approach was tested using a recurrent neural network to predict the battery charge level of an automated guided vehicle, Formica 1. The accuracy was compared using 6 different error metrics, training time of the predictive recurrent neural network model, and coefficient of determination to assess the adequacy. It was established that the proposed feature selection method increases the accuracy of the predictive model by 30%. In addition, it increases the neural network's learning speed by 5 times. However, the two-stage preprocessing method increased the data preprocessing time by 5 seconds.

Keywords

AGV; state of charge, battery; feature selection, ML; RNN; prediction

1. Introduction

The advancement of Industry 4.0 has created a demand for the extensive use of unmanned ground vehicles across various fields. The significant increase in computing power of modern equipment has enabled the usage of mobile platforms not only for transportation but also as a source of a large amount of information. Utilising this information, it is possible to conduct intelligent data analysis and optimise the process of assigned task execution prior to the automated guided vehicles (AGVs) [1, 2]. In general, intelligent data analysis involves three main stages: preliminary processing of collected data; selection and application of an optimal machine learning (ML) model for analysis and forecast; and evaluation of the performance of the model [3, 4, 5].

SMARTINDUSTRY-2024: International Conference on Smart Automation & Robotics for Future Industry, April 18 - 20, 2024, Lviv, Ukraine
* Corresponding author.
† These authors contributed equally.
olena.m.pavliuk@lpnu.ua (O. Pavliuk); myroslav.mishchuk.mknus.2023@lpnu.ua (M. Mishchuk); mykola.o.medykovskyi@lpnu.ua (M. Medykovskyy); rafal.cupek@polsl.pl (R. Cupek)
0000-0003-4561-3874 (O. Pavliuk); 0000-0001-8723-2514 (M. Mishchuk); 0000-0003-2492-8578 (M. Medykovskyy); 0000-0001-8479-5725 (R. Cupek)

The first stage, data pre-processing (which may include data consolidation, deduplication, interpolation, augmentation, detection and removal of anomalies and omissions, feature selection, and normalisation), is crucial in intellectual analysis, as the outcomes of all subsequent stages depend on it. A significant contribution of the data preprocessing stage is the improvement of the accuracy of classifiers and regressors built on the basis of the data [6, 7]. An important task at this stage is the selection of the minimum number of relevant features needed to construct predictive models with satisfactory accuracy. This publication is focused on an important step at the data processing stage—determining dimensionality or feature selection. A recurrent neural network (RNN) is used to predict time sequences using a dataset collected for the AGV Formica 1 at the AIUT enterprise in Gliwice, Poland [8].

The essence of most feature selection methods is to assess the accuracy of the forecast on different subsets of data [9, 10]. The most common methods are cross-validation, greedy feature selection, recursive feature elimination, random forest (RF), gradient boosting (GB), LASSO, ridge, etc., which use statistical approaches. They automatically consider the impact of each feature on the model's quality and efficiency. Consequently, the model uses only significant features while excluding less important ones. Most methods use an iterative approach where features are added, removed, or evaluated one by one or in groups, taking into account the properties of a particular model for data analysis. Specific criteria or metrics are used to assess the importance or quality of each feature. Additionally, most methods allow for the setting of a desired number of features to be included in the feature subset. Therefore, they can be quite effective for large datasets with a significant number of features. The methods with built-in feature importance are particularly effective.

The choice of the feature selection method must be carried out with consideration of the characteristics of the selected model since the features that are significant for one model may not be important for another. Nevertheless, each of these methods may result in the loss of important information, as features that have a significant impact on the model can be eliminated. Some methods are sensitive to the initial selection of features (e.g., greedy feature selection) since their results heavily rely on the features included in the initial step. On the other hand, methods that use the iterative approach (e.g., recursive feature elimination, sequential feature selection, forward selection) are computationally expensive, especially for large datasets. Feature selection can also make a predictive model less resilient to noise in the data. This may be due to the removal of some features that may be useful in low-noise environments. Some methods, such as regularisation methods, require predefined information about parameters or feature importance, which is sometimes very difficult to provide. Many methods have hyperparameters that must be configured. Their incorrect selection can lead to an underestimation or overestimation of the importance of features. Correlation dependencies reflect the importance of features for the model, as they only take into account linear dependencies. Therefore, the choice of a feature selection method depends very much on the specific task in a specific subject area, namely the properties of the data and the model.

Taking into account the advantages and disadvantages of each of the above methods, it is possible to achieve a positive result from their joint use. The purpose of this work is to develop a new method based on the synergistic effect of the joint use of known methods for feature selection. For example, in the RF method, there is a problem of feature importance distortion that arises due to the correlation between features. As a result, important features may be

discarded because the algorithm determines the importance of each feature independently of the others. Possible options for avoiding this problem are:

- preliminary use of feature selection methods to remove redundant or highly correlated features;
- reducing the number of features by analysing the correlation matrix between features, leaving only one that is highly correlated with others, or combining them into a new feature;
- evaluate the importance of features based on correlation when running the RF algorithm, for example, the parameter `max_features` of the scikit-learn library;
- use of dimensionality reduction methods, such as Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA).

For effective real-time situational management, it is necessary to forecast the parameters of AGVs. Hence, effective predictive models that work quickly and with high precision are required. For a high-precision model, it is necessary to select features that have the greatest impact on the result, and the number of features should be optimal. As the number of features increases, the computational and time costs also increase. If the features are not chosen optimally, then the accuracy of the forecast will be low and unsatisfactory for the production enterprise. Also, the accuracy of calculations depends on the quality and quantity of input data. Therefore, the task of reducing the number of features is very important for this domain.

This publication proposes the use of correlation analysis to establish correlations between pairs of parameters from an AGV dataset [11, 12]. A RF method will be used to establish the principal components to generate a predictive AGV battery discharge model. Based on the main components, excluding those that are highly correlated, we build a predictive model using RNN [13, 14, 15]. The results of the study were tested with data on the state of charge of the AGV Formica 1 battery.

2. The State-of-the-Art methods of feature selection for numerical datasets

The feature selection method proposed in this study was developed using data collected from the Formica 1 AGV manufactured by AIUT Gliwice, Poland. In publications [5, 11, 12], the authors forecast various AGV parameters but do not justify the choice of parameters for the predictive model. An analysis of methods that can be used to reduce the number of features, as well as examples of their use, is provided below.

2.1. Greedy Feature Selection

The greedy feature selection is a method used for feature selection in ML tasks [16, 17]. The algorithm starts working from an empty set (forward selection) or, conversely, from a full set of features (backward elimination). To improve model performance, features are gradually added or removed one by one based on given criteria (i.e., coefficients in linear models or feature importance in tree-based models), depending on the initial set of features. Criteria may include error metrics, regression coefficients, the importance of a feature for the model, or other performance metrics. After that, the model is evaluated each time with a new set of features.

This process is considered "greedy" because it makes local decisions at each step based on the current performance of the model. The process is iteratively repeated until certain terminating criteria are met, such as improved performance on the test data, reaching a given number of features, or other set constraints. This method is quite simple to use and implement and can mitigate overfitting the predictive model and increase its speed. The greedy feature selection method can be used for both a small and a large set of features. It can be particularly effective on small sets where each feature can have a significant effect on the result. Reducing the number of features can help reduce the risk of overfitting, especially if feature selection takes their interrelation into account. On large feature sets, greedy feature selection is computationally expensive, but it can be used to improve the computational efficiency of models by applying them to feature subsets.

Researchers use different combinations of features and different selection criteria for greedy feature selection to work effectively. Then, they choose the optimal set of features for a specific task. For example, the promising applications for the AGV Formica 1 are:

- the selection of the most important sensors or features that help the vehicle respond adequately to the surrounding road and obstacles;
- selection of the best visual surveillance features for tasks such as recognition and identification of road markers;
- selection of optimal sensors and features that best affect the quality of the monitoring and decision-making system;
- selection of the best features and parameters of these systems to ensure reliable communication and security;
- selection of the most important sensors that affect energy consumption during braking and acceleration;
- determination of which parameters or indicators affect the energy consumption of the vehicle (speed of movement, type of road, or weight of cargo transported).Use an explicit mark to indicate the affiliations.

2.2. Recursive Feature Elimination

Recursive Feature Elimination (RFE) [18, 19] is an iterative approach to determining the best subset of features that fits the predictive model. RFE is trained on the entire dataset with all features. The importance of each feature is estimated by using the importance coefficients for linear models or feature importance in tree-based models. Then, one or more of the least important features are removed from the dataset. The process is iteratively repeated until it reaches the desired number of features, which are considered the best for the predictive model.

An important advantage of the method is that it takes into account the importance of features in the selection process, so it is suitable for models with built-in importance of features. The disadvantage is the long computational time, which is especially weighty with a significant number of features. RFE takes feature importance into account during selection, but it is dependent on the chosen model and feature importance metric. With the help of this method, it is possible to avoid overfitting due to the exclusion of less important features. Still, due to high computing volumes, the algorithm takes significant computational time on large data sets.

In general, the algorithm is sensitive to hyperparameters, feature importance metrics, and feature removal orders. Also, when removing a feature, there is always a risk of losing important information. For data with a large number of correlated features or textual data, RFE may be less efficient.

For AGV Formica 1, RFE can be used for the following:

- choosing the best set of sensors that are most suited for specific tasks, such as avoiding obstacles or recognising objects on the way;
- selection of the most important features for object recognition systems (people, other vehicles, etc.);
- determining the most important parameters of ML algorithms used to make decisions in the AGV system;
- selection of optimal features and parameters to reduce electricity consumption and improve autonomy;
- defining an initial set of features that can be potentially important for energy consumption analysis (vehicle parameters, delivery area, and various factors);
- assessment of the importance of each feature in predicting energy consumption.

2.3. Correlation approaches

Correlational approaches allow feature importance analysis using correlation [11, 19, 20]. To achieve this, they utilise a matrix that contains correlation coefficients between each pair of features and remove one or both features that are highly correlated with each other, thereby reducing multicollinearity.

This approach can be helpful for selecting a subset of features that contain meaningful information and help improve the accuracy of the predictive model. It also allows for the investigation of the correlation between pairs of features to identify relationships and dependencies between them. To determine the correlation with the target variable, it is necessary to establish how much each individual feature correlates with it. It is also necessary to analyse the correlation between the features themselves. To reduce multicollinearity, features that are strongly correlated with each other are excluded or combined into one feature. It is necessary to discard those features that may be highly correlated with noise rather than with real signals in the data, which may lead to overtraining of the model.

It is important to note that correlation does not always indicate causation, and a high correlation between traits does not necessarily mean that one trait causes the other. It indicates only the degree of linear relationship between them. Therefore, correlation analysis should be accompanied by further validation and study of relationships in the context of a specific ML task. A convenient method of visualising the correlation matrix is the heat map. With its help, it is easy to detect strong correlations between features and the target variable visually. The values of correlation coefficients can range from -1 to 1, with the highest directly proportional linear correlation indicated by the value 1 and inversely proportional -1. If the value of the correlation coefficient is equal to 0, then there is no linear relationship between the variables.

Pearson's correlation coefficient measures the linear correlation between each individual trait and the target variable. It does not take into account other types of relationships, such as non-

linear or monotonic relationships. Features with the highest value of the correlation coefficient are considered important.

Spearman's correlation coefficient measures the correlation between characteristics and the target variable and takes into account not only linear but also monotonic dependence. It determines how closely the ranks of the data in variable A are interconnected with the ranks of the data in variable B. It should be used for data that have a non-linear or monotonic dependence between variables, do not have a normal distribution, have a large number of outliers or anomalies in the data, and with variables measured on an ordinal or interval scale. It can help select features that interact well with the target variable, especially if linear relationships are implicit.

Kendall's tau rank correlation coefficient is a non-parametric indicator that measures the degree of monotonic dependence between two variables. It is not limited to linear relationships and does not require a normal distribution of data. When using the Kendall coefficient, the data are ranked by each variable. Correlation analysis using Kendall's correlation coefficient is useful for data that: have a monotonic dependence, but this dependence is not necessarily linear; do not have a normal distribution; have a large number of outliers or anomalies; and contain variables that are measured on an ordinal or interval scale.

Spearman and Kendall correlation coefficients help select traits that interact well with the target variable, especially in settings where linear relationships are difficult or impossible to detect. The correlation approach is not always effective, provided that the correlation does not reflect the importance of the features for the model. However, correlational approaches can be combined with other methods that do not take into account the correlation between features.

For AGV Formica 1, correlation analysis can be used to:

- estimate of the degree of dependence between energy consumption and each characteristic (which characteristics correlate with energy consumption, as well as the direction of this correlation).
- select features that have a high correlation with energy consumption.

2.4. Least Absolute Shrinkage and Selection Operator (LASSO)

Regularisation methods in linear models, which include LASSO (Least Absolute Shrinkage and Selection Operator) or Ridge, allow the automatic exclusion of insignificant features by setting the coefficients in front of them to zero [21]. LASSO is an automatic regularisation method used for linear models. By limiting the value of the coefficients in front of the features in the model, the possibility of overtraining is reduced, and the generalisation capabilities of the model are increased.

LASSO adds a term loss function ($\lambda * \Sigma|\beta i|$, where λ is the regularisation parameter and βi is the coefficient before the i-th feature), which limits the sum of the absolute values of the model coefficients. Most often, this function is the root mean square error of the regression. Due to the possibility of reducing some coefficients to zero, the algorithm has the property of automatic feature selection. It can also be used for compressed regression analysis. If the least significant features get a zero value, simpler models with fewer independent variables can be built. With the help of cross-validation in LASSO, you can find a balance between the accuracy of the model and its complexity. With a large value of λ, all coefficients βi become zero, and the

model will be simplified. If $\lambda = 0$, no regularisation is applied, and LASSO becomes a classical least squares method. By setting some coefficients to zero and selecting features, the method becomes more resistant to correlation between features. However, if the LASSO coefficients are not correctly set to zero, important information can be lost, and the predictive capabilities of the model can be degraded.

The algorithm is only suitable for linear models but not for feature selection in non-linear models. Therefore, this method is inefficient for further use by neural networks because it does not take into account the interaction between features. At large values of λ, the algorithm becomes sensitive to noise in the data due to setting the coefficients in front of noise features to zero. In the case of collinearity, that is, the correlation between several features, the algorithm selects only one of them and sets the coefficients before the others to zero. Therefore, the algorithm has a high computational complexity.

LASSO can be used in AGV Formica 1 for:

- selection of the most important sensors and features that affect the safety and performance of the vehicle for the autopilot system;
- selection of the most informative data from sensors (lidars, radars, cameras, etc.) to improve navigation and vehicle control;
- selection of the most important parameters and features that affect the optimisation of the speed and power consumption of the vehicle;
- recognition of various objects on the way;
- traffic analysis and forecasting for AGV route optimisation;
- selection of the most important features that affect energy consumption.

2.5. Random Forest or Gradient Boosting embedded feature importance methods

Random Forest (RF) is a controlled algorithm with a marked target variable, and it is a popular tool in the field of ML [22, 23, 24]. It is well suited for tasks with a large number of features and complex relationships among them. It is based on the idea of reusing simple models, such as decision trees, to improve results. A peculiarity of RF is that it consists of multiple decision trees that are trained independently on different subsets of the training data. During the construction of each tree, a subset of data and features are randomly selected for training. That is, each tree sees only part of the total data. This makes RF resistant to overfitting and increases their generalisation abilities. For each node of the tree, a subset of features is randomly selected, so the model takes into account only certain features when making decisions about branches at each level of the tree. When decisions are made, the result is determined by a majority vote, which allows the model to avoid large impacts from individual trees, which may not correspond to the general trend. It is used in various subject areas for classification, regression, data analysis, forecasting, and anomaly detection tasks [22].

Due to the stochastic nature of the learning process, RF are generally less prone to overtraining, especially compared to single decision trees. RF can determine the importance of each feature for predictions. This makes the RF algorithm very efficient and powerful. It can improve the accuracy of the model and reduce the probability of overtraining. The main advantages of RF are that due to the use of a large number of decision trees, it gives a more optimal result for different subsets of data and features and has a lower tendency to overfit. It

also calculates the importance of each feature with high accuracy for a variety of tasks while showing robustness to noise and outliers in the data. The method is easy to implement and allows the processing of both numerical and categorical characteristics.

However, its important drawback is the need to select hyperparameters to achieve the best performance. The importance of features can be distorted if some features are correlated with each other. In tasks where the interpretation of the model is important, this method is less effective compared to others. For large data sets, it is time-consuming. It is also less efficient for tasks where some classes have a low number of instances (i.e., sparse data).

Using the RF method for Formica 1 ground AGV can be helpful for a variety of tasks. Due to the great flexibility and versatility of the RF algorithm, it can be used for various tasks in autonomous vehicles:

- classification of objects on the road (automatic recognition of vehicles and people; classification of markings and signals);
- forecasting traffic and traffic intensity (analysis and prediction of traffic flow for optimal management of traffic flow);
- determining a safe trajectory (choosing safe routes and managing traffic to avoid obstacles);
- detection of abnormal situations and accidents (response to danger, such as an emergency situation on the way or unexpected displacement of the vehicle from the correct route);
- stop-and-go control system (determining the moments of stop-and-go as well as optimising the speed of movement);
- motion trajectory prediction (analysis and prediction of motion trajectories of other objects on the way for efficient AGV route planning);
- optimisation of the braking and acceleration system (determining the optimal braking and acceleration points to ensure smooth and safe AGV movement);
- correction of sensor errors (correction of anomalies and errors in data coming from sensors to increase the reliability of the navigation system).

For the application of RF in AGV, large datasets collected from various sensors (lidars, radars, cameras) are usually used. Datasets can also contain geospatial data, data from other AGVs, etc. RF is often only part of a software complex that helps optimise the operation of the AGV control system. Other algorithms and technologies that implement ML methods can also be used for more complex data processing and decision-making tasks. The greatest effect can be obtained from the synergistic connection of using the RF with other methods.

Gradient Boosting (GB) method is based on the idea of creating a composition of decision trees in order to create a more powerful and accurate forecast model. This method combines several weak learners to create a robust model. Each weak node receives training to correct errors made by previous nodes. Gradient descent is used at each iteration to minimise errors. Due to this, GB usually achieves high prediction accuracy and avoids overfitting.

An important advantage of GB is the ability to determine the importance of each feature for prediction. It can work with both numerical and categorical features. It is often used to solve complex problems, including ranking and anomaly detection. However, like any iterative learning algorithm, GB takes a lot of computational time with large datasets and deep trees. The

result is highly dependent on the selection of hyperparameters (number of trees, tree depth, and learning rate). Additionally, if the depth of the trees is too large, there is a risk of overtraining the model. GB models are not always easy to interpret.

GB can be applied to unmanned ground vehicles for:

- analysis and forecasting of energy consumption for AGV Formica 1 cargo delivery;
- assessment of the importance of each feature in the GB model in order to find out which factors have the greatest impact on energy consumption;
- optimisation of delivery routes, taking into account the forecast of energy consumption and choosing the most efficient driving mode to reduce fuel consumption.

3. Modelling and Results

Figure 1 shows AGV Formica 1, produced by AIUT, Gliwice, Poland. This is a mobile robotic platform weighing 600 kg, which can carry a total weight of 600 kg [11, 15]. The weight can be installed directly on the AGV or on trailers similar to those shown in Figure 1.

Figure 1: AGV Formica 1, AIUT, Gliwice, Poland.

The AGV Formica 1 route is based on a map from AIUT created by an operator. This map highlights all possible obstacles in the path of the AGV, such as structural elements of buildings, furniture, and workplaces. The route, traverse areas, and stop points of the AGV are also provided on the map.

All parameters of the TCP Frame Structure samples are shown in Figure 2.

Field	BEGINNING		ID		Status	TIMSTAMP														NUMBER				VER		LENGTH		DATA													ENDING		
Size (bytes)	3		2		2	12														4				2		2		various													3		
Byte number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32									40	41	42
Format	ASCII	ASCII	ASCII	INT		WORD		DTL												UINT				INT		INT		various													ASCII	ASCII	ASCII
Example	<	F	(																																					)	F	>

Figure 2: TCP Frame Structure.

Each type of signal can contain a large number of specific signals obtained from AGV sensors [25, 26]. Choosing the right parameters and their number is an important issue for building predictive models.

According to the proposed method, it is necessary to calculate the correlation between all parameters and to build a bar chart that reflects the importance of feature using the RF method.

If there is a high correlation dependence between certain parameters, and the results of the RF also confirm this, then these parameters are the minimum necessary for building a predictive model. If the number of parameters is excessive, it is necessary to discard the parameters that have a high correlation between them from the results of the RF method.

To build a prognostic model with the minimum number of features based on the synergy of the RF ensemble method and correlation analysis on the example of data for AGV Formica 1, it is necessary to first calculate the correlation matrix based on the correlation coefficients: Pearson, Spearman, and Kendall. Figure 3 shows the thermal matrices of these correlations.

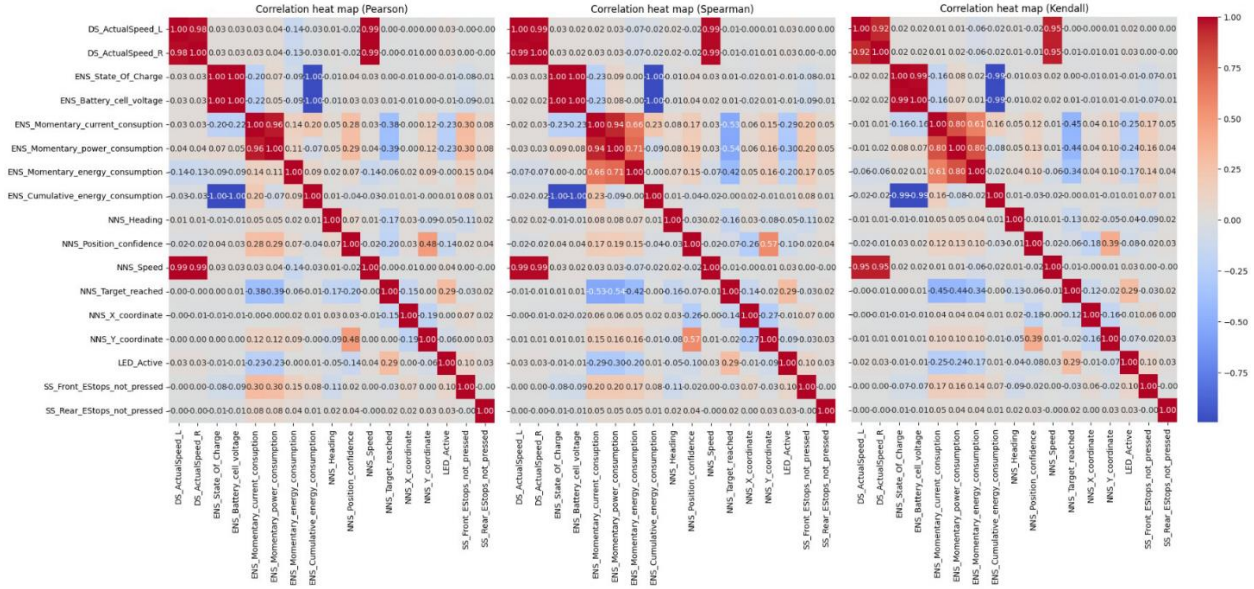

Figure 3: Correlation coefficients: Pearson, Spearman, Kendall.

As can be seen from Figure 3, according to the Pearson, Spearman, and Kendall correlation coefficients, parameters with a significant correlation dependence with ENS_State_Of_Charge are ENS_Battery_cell_voltage and ENS_Cumulative_energy_consumption with coefficients of 0.996984 and -0.999685, 0.999801 and -0.999841, 0.991193 and -0.991104, respectively.

Considering the statistical significance of these parameters at a threshold value of 0.05, the correlation between ENS_Battery_cell_voltage and ENS_State_Of_Charge is statistically significant. The P-value is equal to 0.0 for all methods of establishing correlational dependencies. For Pearson, Spearman, and Kendall are: 0.9969836034863829, 0.9969836034863829, 0.9969836034863829, respectively. The correlation between ENS_Cumulative_energy_consumption and ENS_State_Of_Charge is also statistically significant and equals -0.9996846986127779 for all methods.

Figure 4 shows the resulting importance of features received using the RF method.

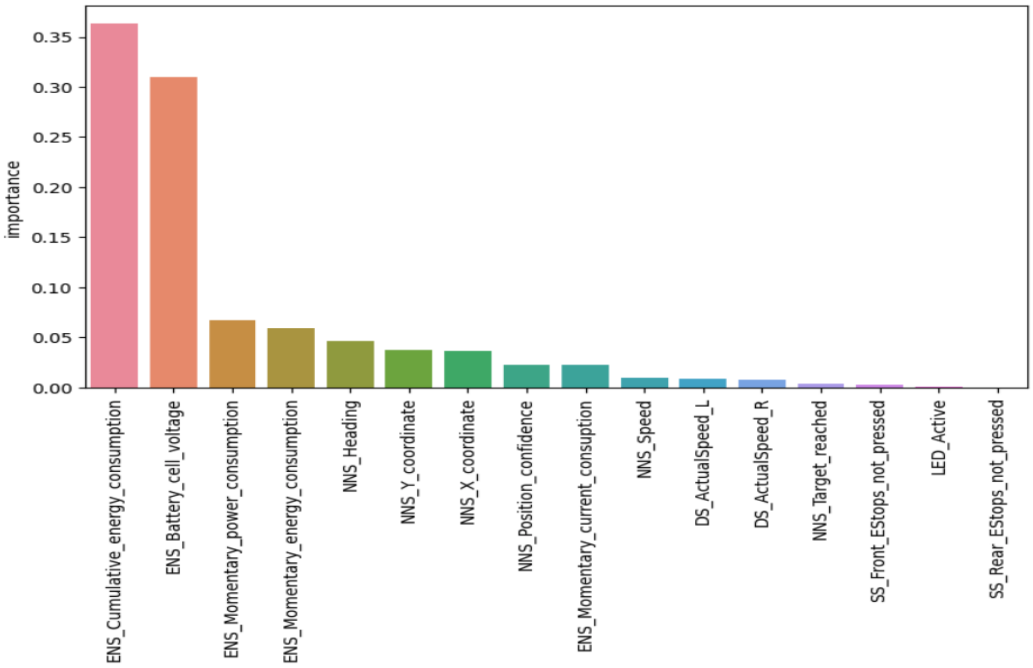

Figure 4: The importance of features received using the Random Forest method.

The results of the RF method and correlation analysis match and indicate that the least features needed to build a predictive model for ENS_State_Of_Charge are ENS_Battery_cell_voltage and ENS_Cumulative_energy_consumption. To confirm the forecast results, a recurrent neural network with the following parameters was used: the input layer is a SimpleRNN type with 12 neurons, which consist of 5-time steps and one feature. The number of training epochs is 100, and the batch size is 2. The model uses a sigmoid activation function in the last layer. Figure 5 presents the results of forecasting the ENS_State_Of_Charge parameter by the proposed method based on the synergy of the Importance of features of the RF method and limited by the Pearson correlation coefficient.

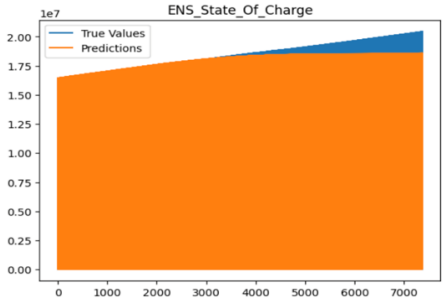

Figure 5: ENS_State_Of_Charge parameter prediction results for the proposed method.

In total, 36843 samples were used for data analysis. We used 80% for training (29474 samples) and the rest for verification and testing (7369 samples). The minimum required number of parameters gave a better result than all parameters for the predictive model using a recurrent neural network.

4. Comparison and Discussion

The effectiveness of the proposed feature reduction method was evaluated by comparing its accuracy with 8 parameters (6 error metrics, the training time of the predictive neural network model, and the determination coefficient to assess the adequacy). For repeatability of results, the model is tested on five different training samples, and the worst results are shown. All tests were performed on an Apple M2 Pro with 16 CPU cores without using a GPU. Table 2 shows the results of the prediction of the proposed method and the maximum number of features by RF methods.

Table 1
Results of the prediction of the proposed method and maximum number of features by Random Forest methods.

Error	Pearson proposed method	Spearman proposed method	Kendall proposed method	Random Forest
MAPE	0.05	0.11	0.46	1.825e+20
RMSE	181049.1	347839.3	452789.9	3076850.18
MSPE	0.01	0.022	1.04	2.141e+30
RMSPE	6.65	14.76	101.98	1.46e+17
MBE	57404.7	93009.5	222501.6	822869.30
ME	820493.9	1215717.9	1281451.9	14596494.25
R2	0.99	0.99	0.99	0.66
Minutes	10.8	10.22	10.1	62.2

The proposed two-step method, based on the synergy of correlation analysis and RF, showed an accuracy increase of more than 30% compared to the most conventional RF method. Neural network training time decreased from 62.2 min to 10.8 min. Due to the reduction of features for the predictive model, the training time of the neural network has significantly decreased. The total time of data preprocessing, together with the neural network training time of the proposed method is about 11 minutes, and for the RF it is about 62 minutes. At the same time, the coefficient of determination showed that the model is adequate according to the proposed two-stage method, with R2 = 0.99. Considering the conventional RF method, the model is at the limit of adequacy with R2 = 0.66.

5. Conclusions

This study is focused on the problem of efficient data preprocessing to improve accuracy in predictive models for intelligent analysis. We developed a new two-stage feature reduction method based on the synergy of correlation analysis and Random Forest (RF). It is based on the possibility of considering the correlational interdependencies between the characteristics of each observation and the results of selecting the importance of features according to the RF method. The proposed approach was tested using a recurrent neural network to predict the state of charge of the AGV Formica 1 battery. For repeatability of the results, the model was tested on five different training samples. We compared the accuracy of the proposed method for selecting the minimum number of 6 errors and the coefficient of determination to assess the adequacy and

training time of the predictive neural network model. The adequacy of the model was also checked using the coefficient of determination.

It was established that the proposed method of feature selection increases the accuracy of the predictive model by 30%. In addition, it provided an increase in the learning speed of the neural network by more than 5 times. However, the data preprocessing time also increased by 5 seconds due to the two-stage method. Further research will be conducted to evaluate the accuracy of other types of artificial neural networks [27, 28, 29], particularly LSTMs and DNNs, based on the developed two-step feature reduction method for analysing large datasets.

Acknowledgements

The research leading to these results received funding from the Norway Grants 2014-2024, which the National Centre operates for Research and Development under the project "Automated Guided Vehicles integrated with Collaborative Robots for Smart Industry Perspective" (Project Contract no.: NOR/POLNOR/CoBotAGV/0027/2019 00). And also, the Polish-Ukrainian grant "Automated Guided Vehicles integrated with Collaborative Robots - energy consumption models for logistics tasks planning" (02/110/ZZB22/1022).

References

[1] Agarwal, D., & Bharti, P. S. (2022). A Case Study on AGV's Alternatives Selection Problem. International Journal of Information Technology, 14(2), 1011–1023. DOI: 10.1007/s41870-018-0223-z.

[2] Steclik, T., Cupek, R., & Drewniak, M. (2022). Automatic grouping of production data in Industry 4.0: The use case of internal logistics systems based on Automated Guided Vehicles. Journal of Computational Science, 62, 101693.

[3] Kumar, P., Kumar, Y., & Tawhid, M. A. (Eds.). (2021). Machine Learning, Big Data, and IoT for Medical Informatics; Intelligent Data Centric Systems. Academic Press.

[4] Ahsan, M. M., Mahmud, M. A. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. Technologies, 9(52).

[5] Cupek, R., Gólczyński, Ł., & Ziebinski, A. (2019). An OPC UA Machine Learning Server for Automated Guided Vehicle. Computational Collective Intelligence, pp. 218-228.

[6] Tlebaldinova, A., Denissova, N., Baklanova, O., Krak, I., & Györök, G. (2020). Normalization of Vehicle License Plate Images Based on Analyzing of Its Specific Features for Improving the Quality Recognition. Acta Polytech. Hung., 17, 193–206.

[7] Shakhovska, N., Yakovyna, V., & Kryvinska, N. (2020). An Improved Software Defect Prediction Algorithm Using Self-Organizing Maps Combined with Hierarchical Clustering and Data Preprocessing. In International Conference on Database and Expert Systems Applications (pp. 414–424). Springer International Publishing.

[8] Hu, Z., Ivashchenko, M., Lyushenko, L., & Klyushnyk, D. (2021). Artificial Neural Network Training Criterion Formulation Using Error Continuous Domain. IJMECS, 13, 13–22.

[9] Fang, L., Zhao, H., Wang, P., Yu, M., Yan, J., Cheng, W., & Chen, P. (2015). Feature Selection Method Based on Mutual Information and Class Separability for Dimension

Reduction in Multidimensional Time Series for Clinical Data. Biomedical Signal Processing and Control, 21, 82–89.
[10] Han, M., & Liu, X. (2013). Feature Selection Techniques with Class Separability for Multivariate Time Series. Neurocomputing, 110, 29–34.
[11] Pavliuk, O., Steclik, T., & Biernacki, P. (2022). The forecast of the AGV battery discharging via the machine learning methods. 2022 IEEE International Conference on Big Data (Big Data), 6315–6324.
[12] Benecki, P., Kostrzewa, D., Grzesik, P., Shubyn, B., & Mrozek, D. (2022). Forecasting of Energy Consumption for Anomaly Detection in Automated Guided Vehicles: Models and Feature Selection. 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2073–2079.
[13] Teslyuk, V., Kazarian, A., Kryvinska, N., & Tsmots, I. (2021). Optimal Artificial Neural Network Type Selection Method for Usage in Smart House Systems. Sensors, 21(1), 47.
[14] Bykov, M. M., Kovtun, V. V., Smolarz, A., Junisbekov, M., Targeusizova, A., & Satymbekov, M. (2017). Research of Neural Network Classifier in Speaker Recognition Module for Automated System of Critical Use. In Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments (p. 1044521). International Society for Optics and Photonics.
[15] Pavliuk, O., Cupek, R., Steclik, T., Medykovskyy, M., & Drewniak, M. (2023). A Novel Methodology Based on a Deep Neural Network and Data Mining for Predicting the Segmental Voltage Drop in Automated Guided Vehicle Battery Cells. Electronics, 12(22), 4636.
[16] Alimardani, F., Boostani, R., Azadehdel, M., Ghanizadeh, A., & Rastegar, K. (2013). Presenting a New Search Strategy to Select Synchronization Values for Classifying Bipolar Mood Disorders from Schizophrenic Patients. Engineering Applications of Artificial Intelligence, 26(2), 913–923.
[17] Jabir, M., Afzal, M. K., Mudassar, R., Le, Y., & Yarong, C. (2024). Solving Line Balancing and AGV Scheduling Problems for Intelligent Decisions Using a Genetic-Artificial Bee Colony Algorithm. Computers & Industrial Engineering, 189, 109976.
[18] Dias, N. S., Jacinto, L. R., Mendes, P. M., & Correia, J. H. (2009). Feature down-selection in brain-computer interfaces dimensionality reduction and discrimination power. 2009 4th International IEEE/EMBS Conference on Neural Engineering (pp. 323–326). IEEE.
[19] Guo, J., Wang, Z., Jin, Y., Li, M., & Chen, Q. (2023). Predicting and Extracting Thermal Behavior Rules of Hydronic Thermal Barrier with Interpretable Ensemble Learning in the Heating Season. Energy and Buildings, 301, 113699.
[20] Yao, Y., Liu, Q., Fu, L., Li, X., Yu, Y., Gao, L., & Zhou, W. (2024). A Novel Mathematical Model for the Flexible Job-Shop Scheduling Problem With Limited Automated Guided Vehicles. IEEE Transactions on Automation Science and Engineering, 1–14.
[21] Kumar, M. P., Gao, Z.-J., & Chen, K.-C. (2023). Time Series-Based Sensor Selection and Lightweight Neural Architecture Search for RUL Estimation in Future Industry 4.0. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 13(2), 514–523.
[22] Trostianchyn, A., Duriagina, Z., Izonin, I., Tkachenko, R., Kulyk, V., & Lotoshynska, N. (2022). AN APPROACH TOWARD PREDICTION OF SM-CO ALLOY'S MAXIMUM

ENERGY PRODUCT USING FEATURE BAGGING TECHNIQUE. Acta Metallurgica Slovaca, 28(2), 91–96.

[23] Hanzel, K., Grzechca, D., Ziebinski, A., Chruszczyk, L., & Janus, A. (2023). Estimating the AGV Load and a Battery Lifetime Based on the Current Measurement and Random Forest Application. 2023 IEEE International Conference on Big Data (BigData), 5057–5063.

[24] Tkachenko R., Duriagina Z., Lemishka I., Izonin I., Trostianchyn A. (2018). Development of machine learning method of titanium alloy properties identification in additive technologies. Восточно-Европейский журнал передовых технологий. - 2018. - № 3(12). - С. 23-31.

[25] Shubyn, B., Kostrzewa, D., Grzesik, P., Benecki, P., Maksymyuk, T., Sunderam, V., Syu, J.-H., & Lin, J. C.-W. (2023). Federated Learning for Improved Prediction of Failures in Autonomous Guided Vehicles. Journal of Computational Science, 68, 101956.

[26] Szygula, J., Biernacki, P., Marek, D., Domanski, A., Sobczak, L., Flak, J., & Caban, D. (2022). Analysis of Web-Based Geo-Visualization Methods Applied for Automated Guided Vehicle Using Satellite Navigation Systems. 2022 IEEE International Conference on Big Data (Big Data), 6371–6377.

[27] Elsisi, M., & Tran, M.-Q. (2021). Development of an IoT Architecture Based on a Deep Neural Network against Cyber Attacks for Automated Guided Vehicles. Sensors, 21(24), 8467.

[28] Lai, J., Ren, Z., Wu, Z., Liu, Y., & Xie, S. (2020). Deep Neural Network-Based Real-Time Trajectory Planning for an Automatic Guided Vehicle with Obstacles. 2020 Chinese Automation Congress (CAC), 6311–6316.

[29] Vakaruk, S., Karamchandani, A., Sierra-García, J. E., Mozo, A., Gómez-Canaval, S., & Pastor, A. (2023). Transformers for Multi-Horizon Forecasting in an Industry 4.0 Use Case. Sensors, 23(7), 3516.