ISSN 2603-4697 (Online)

Complex Control Systems

Volume 4, Issue 1, 2022

Bulgarian Academy of Sciences

Institute of Robotics

COMPLEX CONTROL SYSTEMS

Volume 4/ Issue 1/ May 2022

Institute of Robotics – Bulgarian Academy of Sciences

Sofia

COMPLEX CONTROL SYSTEMS

Volume 4/ Issue 1/ April 2022

Institute of Robotics – Bulgarian Academy of Sciences

ISSN: 2603-4697 (Online)

Volume 4, Issue 1, 2022

Table of Contents

Preface

to the Proceedings of the International Conference Statistics and Machine Learning in Electronics

Dear colleagues,

The ***International Conference Statistics and Machine Learning in Electronics*** is organized according to the scientific project “Exploration and application of statistics and machine learning in electronics”, supported by Bulgarian National Science Fund under contract number КП-06-Н42/1. It was conducted in hybrid format on 12-13 May 2022 as the host organization was Technical University of Sofia, Sofia, Bulgaria.

The aim of the forum was to gather together students and experienced researchers in order to report and discuss contemporary solutions, best practices, and advancements regarding the application of machine learning techniques in the area of Electronics. It also stimulates international cooperation in order to be extended for obtaining new knowledge, for exchanging research ideas and promoting scientific results.

The participated authors were from University of Mostar, Bosnia and Herzegovina, University of Split, Croatia, University of Nis, Republic of Serbia, University of the Basque Country, Spain, TECNALIA, Spain, University of Applied Sciences Offenburg, Germany, Changshu Institute of Technology, China, Bulgarian Academy of Sciences and Technical University of Sofia, Bulgaria as the presented topics were from applications of statistics and machine learning in inertial-sensor based hand gesture recognition, simultaneous localization and mapping system, rehabilitation system for cerebral palsy through modeling a system that predicts value of one pore pressure cell, investigation the dimension reducing and dynamic pruning to improving the production process of electronic components, enhancing the vendor’s quality management assurance in automotive electronic products and organizing the supply chain management of electronic industry.

Our great keynote speakers present the current research, important issues and future directions for development. The talk of **prof. Petia Georgieva** from University of Aveiro, Portugal was about machine learning in industry - from model-driven to data-driven approach. **Prof. Stefan Hensel**, University of Applied Sciences in Offenburg, Germany discusses application of artificial intelligence in machine vision and autonomous systems applications.

The publication of the accepted papers in an open access journal ***Complex Control Systems*** of Bulgarian Academy of Sciences gives a nice opportunity for spreading the project mission, goals and results worldwide and we believe that they will reach all colleagues interested in this new and progressive topic.

We can say that the conference was fruitful, connecting theory and practices and allowing collaboration among experienced and inexperienced scientists.

ICSMLE 2022 Organizing Committee

Conference Chairs

Malinka Ivanova, Technical University of Sofia

Marin Marinov, Technical University of Sofia

Stefan Hensel, University of Applied Sciences Offenburg, Germany

Local Organization Committee

Valentin Videkov, Technical University of Sofia

Valentin Tsenev, Technical University of Sofia

Borislav Ganev, Technical University of Sofia

Zravka Tchobanova , Technical University of Sofia

Petya Petkova, Technical University of Sofia

Programme Committee

Desislava Ivanova, Technical University of Sofia

Anna Rozeva, Technical University of Sofia

Igor Vrublevsky, Belarusian State University of Informatics and Radioelectronics, Belarus

Petia Georgieva, University of Aveiro, Portugal

Hiroaki Wagatsuma, Kyushu Institute of Technology, Japan

Lyuba Alboul, Sheffield Hallam University, UK

Veselka Boeva, Blekinge Institute of Technology, Sweden

Galina Bogdanova, Bulgarian Academy of Sciences, Bulgaria

Maya Dimitrova, Bulgarian Academy of Sciences, Bulgaria

Marco Temperini, Sapienza University of Rome, Italy

Miona Stošović, University of Niš, Serbia

Svetozar Ilchev, Bulgarian Academy of Sciences, Bulgaria

Stoyan Stoyanov, University of Greenwich, UK

Zhifeng Wang, Florida State University, USA

CONTENT

Research paper ISSN 2603-4697 (Online) Complex Control Systems Vol. 4, No. 1, 2022, 1-6

Off-line Inertial-Sensor Based Hand Gesture Recognition and Evaluation

Mirela Kundid Vasić(1), Tamara Grujić(2), Ivo Stančić(2), Josip Music(2), Mirjana Bonković(2),

(1)Faculty of Mechanical Engineering, Computing and Electrical Engineering, University of Mostar, Mostar, B&H
(e-mail: mirela.kundid.vasic@fsre.sum.ba)
(2)Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture - FESB, Department of Electronics and Computing, University of Split, Split, Croatia
(e-mail: tamara.grujic@fesb.hr, ivo.stancic@fesb.hr, josip.music@fesb.hr, mirjana.bonkovic@fesb.hr)

***Abstract*— Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures with computer programs and many different algorithms. Any bodily motion or state that most commonly originates from the face and/or hand can be interpreted as a human gesture. Most of the research today focuses on emotion detection and recognition of hand gestures using cameras and computer vision algorithms. Gesture recognition can be seen as the way computers begin to understand human body language. There are many different areas this topic of computer science can be applied to; the main field is human-computer interaction interfaces (HCI). Today the main interaction tools between computers and humans are still keyboard and mouse. Gesture recognition can be used as a tool for communication with the machine and interact without any mechanical device such as keyboard or mouse. In this paper, we present the results of a comparison of five different machine learning classifiers in the task of human hand gestures recognition. Gestures were recorded by using inertial sensors, gyroscopes, and accelerometers placed at the wrist and index finger. One thousand and eight hundred (1800) hand gestures were recorded and labelled. Six important features were defined, for the identification of nine different hand gestures, using five different machine learning classifiers: Logistic Regression, Random Forests, Support Vector Machine (SVM) with linear kernel, Naïve Bayes classifier, and Stochastic Gradient Descent.**

***Keywords*—**hand gestures, inertial sensors, machine learning algorithms, off-line classification, evaluation.

I. Introduction

Nowdays, in a world where computers become more and more pervasive in culture, the need for efficient human-computer interaction is increasing at a rapid pace. The most commonly used way of human-computer interaction (HCI) is a graphical user interface (GUI) which requires the use of additional devices, e.g. mouse, keyboard, etc. Researchers in academia and industry are increasingly looking for ways to make human-computer interaction easier, safer, and more efficient. Consequently, new interaction styles have been explored. One of them is hand gesture-based interaction, which allows users a more natural way to communicate, without any extra devices, which is much simpler and intuitive than using graphical interfaces or text input. For example, such type of interaction is used in applications like controlling smart interactive television [1] or enabling a hand as a 3D mouse.

Recognized gestures also can be used for controlling a robot [2] or conveying meaningful information. It can be very useful in particular situations, e.g. with robots designed to assist disabled people and helping them with personal and professional tasks on a daily basis, or for real-time mobile robot control [2].

The key problem in hand gesture-based interaction is how to make gestures easily understandable and accurately interpretable to computers. Thus, hand gesture recognition is an area of active research in the field of computer vision and machine learning. Accordingly, different approaches have been considered and all of them can be mainly divided into two groups: vision-based [3], [4] and inertial sensors-based methods, which are reliable, accurate, and robust, for work in real-time [2], [5]. These methods involve the use of inertial sensors placed on the user's body.

In gesture recognition, two main popular approaches are available: body-worn systems that track user gesture and motion using body worn sensors attached to the human body, or external systems based on external sensors. Sensors like Microsoft Kinect allow human identification and tracking in 3D space using RGB-D sensors which combine RGB colour information with per-pixel depth information, to get information about objects in 3D space. Many researchers have reported successful usage of hand gesture recognition tasks for control of a variety of robots such as mobile robots [2], [6], or humanoid robots [7]. Most of the research is based on inertial sensors that use machine learning methods for gesture identification. The gesture identification task consists of several steps, including the recording of hand gestures and their classification, which is called hand gesture modelling. The process of hand gesture modelling is the main topic of this research and implies the selection of the right identification units of the movement also called features.

II. Features Definition and Selection

Feature selection is the most important task in classification, and the main goal is to determine the right set of identification units for the particular classification task, or in this case, movement, to most accurately determine the difference between movements in the particular set. The movements that we are trying to model are shown and described in detail in [2]. Nine gestures are simple to perform since they consist of few elementary hand motions and are devised in a manner that it is possible to extract at least one significant unique feature from each of the nine described gestures.

The quality of the features is the most important factor in the classification task. The hand gesture modelling includes the identification process of the features that are most discriminative using the subset of initially proposed features.Features used in this approach rely on data obtained from a combination of gyroscopes and accelerometers signals. The inertial sensors (accelerometers and gyroscopes) are connected to the wrist and index finger to obtain the empirical data needed to recognize the motion. Wearable sensors should be small and lightweight, in order to be fastened to the human body without compromising the user's comfort and allowing her/him to perform the movement under unrestrained conditions as much as possible.

A pattern recognition machine does not perform classification tasks working directly on the raw sensor data. Usually, before the classification, data representation is built in terms of feature variables [8].

The features used in this research are listed in Table I, [2]. The suggested features are extracted from gyroscopes and accelerometers and are hand-labelled for certain hand motions/gestures. The first feature is gesture duration, which in some cases can't be a discriminating feature because it could be the same duration for different gestures. By further analysis, additional features are obtained, like the second feature that contains local extremes of gyroscope differential data (number of extremes), while the gyroscope axis ratio is the third feature implemented. The fourth and fifth features are derived from accelerometers data and that is accelerometer axis ratio, which represents absolute acceleration, and movement energy. The last feature is the magnitude of the first significant extreme in the gyroscope data.

TABLE I
FEATURES USED FOR OFF-LINE EVALUATION WITH SELECTED CLASSIFIERS [2]

Feature name	Feature description	Sensors used *
Gesture duration	Gesture duration in ms	G1, G2
Number of extremes	Number of extremes from differential gyroscope data (DGD)	G1, G2
Gyroscope axis ratio	Mean ratio of axis of DGD, detects direction of motion	G1, G2
Accelerometer ratio	Mean ratio A1 axis, detects hand orientation	A1
Movement energy	Integrates absolute A1 and A2 magnitude over the duration of the whole gesture	A1, A2
First rotation direction (flexion or extension)	Magnitude of the first large DGD peak, detects hand rotation direction	G1, G2

*As in Fig. 2: G1 is the wrist gyroscope, G2 is the index finger gyroscope, A1 is the wrist accelerometer, A2 is the index finger accelerometer

Experimental setup and generation of the model for all nine gestures included twenty participants, each of them had ten measurements of all nine performed gestures, to generate the reference point for every gesture. Therefore, the proposed model included 1800 labelled hand gestures to form the baseline for the gestures included in process of classification. The hand gestures labelled data is used for the model generation using five different classifiers that will be explained in Section III.

III. Classification Task

The classifiers used in this research are the Naïve Bayes classifier, Support Vector Machine (SVM) with linear kernel, Logistic Regression, Random Forests, and Stochastic Gradient Descent. The classification task for each of five different classifiers is the same and consists of two phases – the training phase and the classification phase. Please note that the off-line classification approach is used, and not a real-time one. The data was collected during the feature extraction phase and then used to determine the best classifier based on the scoring classifier phase. The scoring phase first divided the dataset into training and testing sets in a 50:50 percent ratio. Then it trained the selected classifier and tested the classifier on the test set. For each of the classifiers the precision, recall, and F-Score have been calculated, to show which classifier will achieve the best score. Classification of data collected from the hand gesture collection phase has been done using SkLearn library with Python programming language.

A. *Naïve Bayes Classifier*

The Naive Bayes (NB) [9] classifier method consists of two phases: the training phase and the classification phase. The testing has been done on the test set of data from a total of nine hand gestures by the Naive Bayes classifier. Naive Bayes classifier gives a statistical dimension to the made conclusions. Membership to each cluster (class) is determined by the distribution of probabilities. Therefore, optimal classification can be determined by taking into consideration the distribution of probabilities to which each vector belongs (aligning each feature in each group). Data is presented as an n-dimensional vector; the classifier's task is to predict a group of testing data based on equation (1).

$$\underset{V_j \in V}{argmax}\, p(v_j | a_1, a_2, a_3, \dots, a_n) \qquad (1)$$

If Bayesian theorem is applied to Equation (1), Equation (2) is obtained.

$$V_{NB} = \underset{V_j \in V}{argmax}\, p(v_j) \prod p(a_i | v_j) \qquad (2)$$

Set V_{NB} denotes the classified instance and its probability of belonging into a certain class, in this case, one of nine hand gestures.

B. *Support Vector Machine*

Support Vector Machine (SVM) [10] represents a set of related supervised learning methods (supervised learning) that are used for classification and regression. The SVM is a binary classifier i.e. probabilistically classifies into two categories. The SVM classification is based on a division of all points in space into two categories according to the margin between support vectors, and the algorithm searches for the largest gap between the two categories.

This procedure is referred to as the so-called linear classification, however, to perform the classification of several categories, we perform a kernel trick that implicitly maps the inputs into the multidimensional space. The trick avoids explicit mapping, which is necessary to get a linear learning algorithm to be trained with a nonlinear function. Classifier training creates boundary decisions separate margin that must be maximum, and the input data set, we can get a linear distribution of the two classes. Computing the SVM classifier amounts to minimizing an expression of the form denoted in equation (3).

$$\left[\frac{1}{n}\sum_{i=1}^{n} \max\left(0, 1 - y_i(w * x_i + b)\right)\right] + \Lambda \left|\left|w\right|\right|^2 \qquad (3)$$

The y_i are labels for each data sample x_i, w is normal vector that separates the data into two planes and b is the margin between hyperplane and classified data. The parameter Λ denotes the trade-off between increasing the margin size and ensuring that x_i is inside the right plane.

C. *Logistic Regression*

Logistic regression (LR) [11] is a regression model where the dependent variable is categorical. The binary logistic model is used to estimate the probability of binary response, based on one or more predictor variables (features). LR measures the relationship between the categorical dependent variable and one or more independent variables using a logistic function. Conditional distribution can be Bernoulli or Gaussian because the outcome of the event can be binary (the dependent variable can have only two values). Mathematically, LR is the task of estimating log odds of a certain event, and it estimates multiple linear regression functions as denoted in equation (4).

$$logreg(p) = log\left(\frac{p(y=1)}{1-(p=1)}\right) = \beta_0 + \beta_1 * x_{i1} + \beta_2 * x_{i2} + \dots + \beta_p * x_{ip} \quad (4)$$

D. Random Forests Algorithm

Random forest classifier (RFC) [12] algorithm is a notion of the general technique of random decision forests that is a learning method for classification, regression, and other tasks. Random forest is a collection of decision trees whose results are aggregated into one final result. The algorithm operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Given a training set $X = [x_1, x_2, \dots, x_n]$ with corresponding labels $Y = [y_1, y_2, \dots, y_n]$, the algorithm selects a random sample with replacement of the training set and fits trees to these samples and does that repeatedly B times. After training, unseen samples x' can be classified by averaging predictions on individual trees (regression), as shown in equation (5), or by taking the majority vote in decision trees.

$$f' = \frac{1}{b}\sum_{b=1}^{B} f'_b(x') \quad (5)$$

E. Stochastic Gradient Descent

Stochastic gradient descent (SGD) [13] is a gradient descent optimization method for minimizing an objective function that is written as a sum of differentiable functions. In both gradient descent and stochastic gradient descent, a set of parameters is updated, in an iterative manner, to minimize an error function. SGD is one of the fastest training algorithms. SGD is popular within training wide range of models in machine learning and is a de-facto standard for training artificial neural networks. This problem is considered as the problem of minimizing an objective function as denoted in equation (6). The parameter w is to be estimated, whilst Q_i is associated with observations in the dataset.

$$Q(w) = \sum_{b=1}^{B} Q_i(w) \quad (6)$$

IV. Classification Results, Comparison, and Discussion

The comparison of the different classifiers has been done using the confusion matrices for each of the defined classifiers. The main concepts of confusion matrices are false positive observations (hereinafter denoted as FP), false negatives (FN), true positives (TP), and true negatives (TN). The evaluation measures for the scoring of the particular classifier were Precision, Recall, and F1 – score.

The results of the classification are generated by Python programming language using SkLearn library to calculate the classes based on generated models. Five different classifiers were compared, based on precision, recall, and F1 score, as shown in Table II.

TABLE II
PRECISION, RECALL AND F1 SCORE MEASURES FOR CLASSIFIERS COMPARISON

	Precision	**Recall**	**F1-score**
Random forest classifier (RFC)	0.974	0.973	0.973
Logistic regression (LR)	0.973	0.972	0.972
Linear SVM	0.956	0.955	0.955
Naïve Bayes (Gaussian)	0.936	0.934	0.934
Stochastic gradient descent (SGD)	0.288	0.412	0.290

When the problem of classification involves the search for the positive class samples and they are very rare compared to the negative classes, the precision and recall approach is used. This method for evaluation of classifiers is more useful in "needle-in-haystack" type problems where the positive class is more "interesting" then the negative class. When it is needed to emphasize negative class, the Receiver Operating Characteristic (ROC) plot is used. ROC curves represent a graphical plot of True Positive Rate (TPR) as the function of False Positive Rate (FPR). ROC curves for all five classifiers are shown in Fig. 1 - Fig. 5. Figures show ROC curves for all nine classes (nine hand gestures).

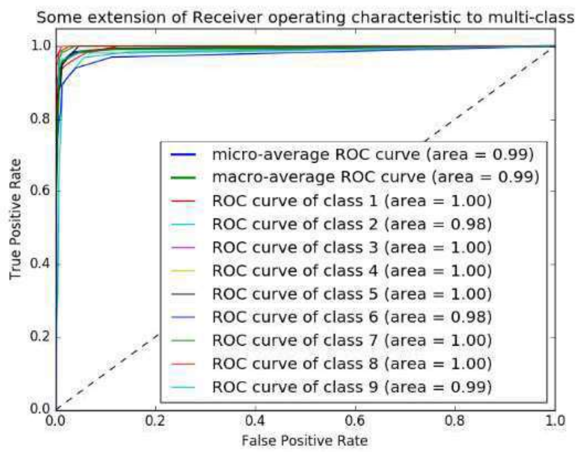


Fig. 1 ROC plot for Random forests classifier (micro averaged 99%)

The third classifier is SVM, whose advantages are the high accuracy, nice theoretical guarantees regarding overfitting, and with an appropriate kernel, they can work well even if data is not linearly separable in the base feature space. SVM is especially popular in text classification problems where very high-dimensional spaces are the norm (sparse data). In this case, where the data is clearly linearly separable, SVM with the right parameters can achieve the best results, but the main disadvantage is the speed due to which it is impossible to use this classifier in real-time scenarios.

One of the simplest classifiers is the Naïve Bayes algorithm, which is also, as Logistic regression, probabilistically oriented. Naïve Bayes classifier will converge quicker than discriminative models like Logistic regression, so less training data is needed. However, this algorithm is less accurate, as shown in PR values and ROC plots.

The last classifier that is tested in this comparison, Stochastic Gradient Descent (SGD), is the worst based on used performance parameters. SGD is one of the on-line algorithms that uses a combination of linear functions to solve the minimization problem. Despite the obvious lack of precision, recall, and F1-score in this off-line example, the main advantage of SGD is the light computational cost that is suitable for on-line approaches.

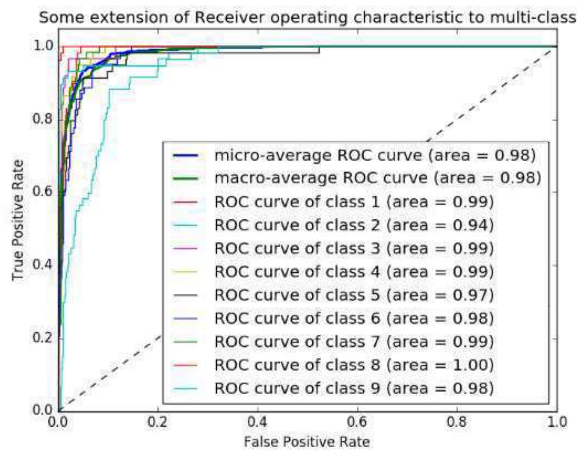


Fig. 2 ROC plot for Logistic regression classifier (micro averaged 98%)

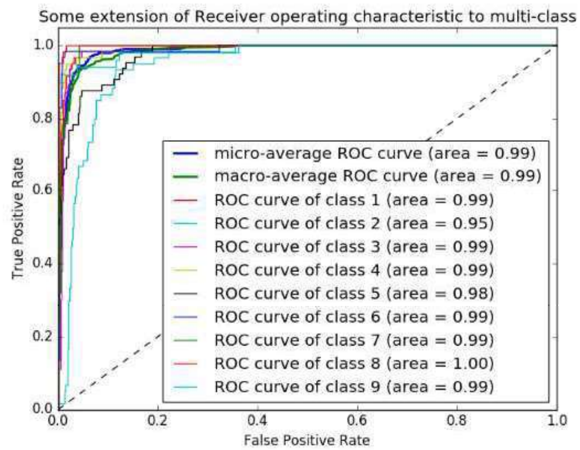


Fig. 3 ROC plot for Support Vector Machine classifier (micro averaged 99%)

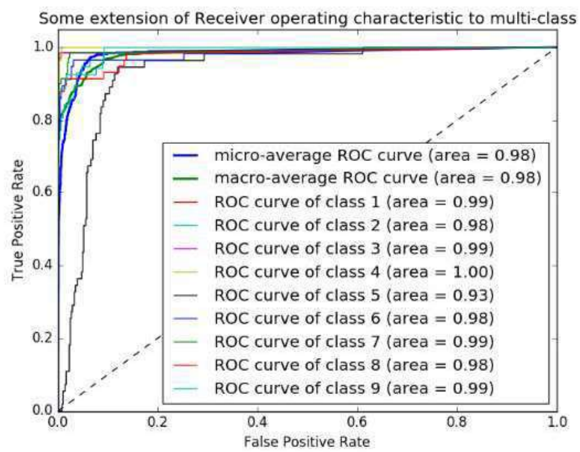


Fig. 4 ROC plot for Gaussian Naïve Bayes classifier (micro averaged 98%)

5

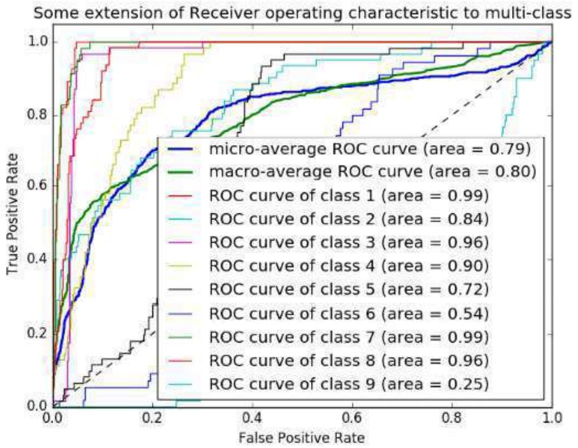


Fig. 5 ROC plot for Stochastic gradient descent classifier (micro averaged 79%)

V. CONCLUSION

The main goal of this research was to test the performance of five different machine learning algorithms to get a robust method for the classification of human hand gestures. All tested methods depend on the data that is presented as training data set, therefore the features that are used are a very important factor in a classification task. The data and features provided in [2] are highly discriminative and, as obtained results show, provide good classification results with tested machine learning (ML) algorithms.

All of the classifiers, except the SGD algorithm, showed very high F1-score and Area Under Curve (AUC) in ROC plots. The best classifier for the real-time classification scenario might be the Logistic Regression classifier, which showed almost identical results on testing data as Random forests classifiers. However, Random forests classifiers are not suitable for real-time classifications.

Better data often outperforms better algorithms, and thus the design of high-quality features is of great importance. With a sufficiently large dataset and high-quality features, different classification algorithms can perform well in terms of classification performances. Therefore, the choice of which algorithm to use should be based on performance speed or ease of use. In the case of Logistic Regression, different methods for model regularization are available, and the features correlation issue should not be a problem. Therefore, there are many arguments why and how this classifier can be successfully used in an on-line scenario of hand gesture recognition based on inertial sensors.

REFERENCES

[1] Vishwakarma D. K. and Kapoor R, "An Efficient Interpretation of Hand Gestures to Control Smart Interactive Television," *International Journal of Computational Vision and Robotics*, Vol. 7(4), pp. 454 – 471, 2017.

[2] Stančić I., Music J, and Grujić T., "Gesture Recognition System for Real-Time Mobile Robot Control Based on Inertial Sensors and Motion Strings," *Engineering Applications of Artificial Intelligence*, vol. 66, pp. 33-48, 2017.

[3] Kundid Vasić M., Galić I., and Vasić D., "Human Action Identification and Search in Video Files," in Proc. of 57th International Symposium on Electronics in Marine - ELMAR, IEEE, 2015.

[4] Choondal J. J. and Sharavanabhavan C., "Design and Implementation of a Natural User Interface Using Hand Gesture Recognition Method," *International Journal of Innovative Technology and Exploring Engineering,* vol. 10(3), pp. 249-254, 2013.

[5] Xu R., Zhou S., and Li W. J., "MEMS Accelerometer Based Nonspecific-User Hand Gesture Recognition," *IEEE Sensors Journal,* vol. 12(5), pp. 1166-1173, 2012.

[6] Filaretov V., Yukhimetsa D., and Mursalimov E., "The Universal Onboard Information-Control System for Mobile Robots," in Proc. of the *25th DAAAM International Symposium on Intelligent Manufacturing and Automation, DAAAM*, Vienna, 2014.

[7] Riek L, Rabinowitch T., Bremner P., Pipe A., and Fraser M., "Cooperative Gestures: Effective Signaling for Humanoid Robots," in *Proc. of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, 2010.

[8] Mannini A. and Sabatini A. M., "Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers," *Sensors,* vol. 10(2), pp. 1154-75, 2010.

[9] Pardos Z. and Heffernan N., "Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing," in *Proc. of the International Conference UMAP*, 2010.

[10] Cortes C. and Vapnik V., "Support - Vector Networks," *Machine Learning,* vol. 3, pp. 273–297, 1995.

[11] McCullagh P. and Nelder J., *Generalized Linear Models*, CRC press, 1989.

[12] Wai M. M. N., "Classification Based Automatic Information Extraction," in *ACELAE'11 Proc. of the 10th WSEAS International Conference on Communications, Electrical & Computer Engineering*, 2011.

[13] Chen J. and Melo G. D., "Semantic Information Extraction for Improved Word Embeddings," in *Proc. of NAACL-HLT*, Denver, Colorado, 2015.

Research paper ISSN 2603-4697 (Online) Complex Control Systems Vol. 4, No. 1, 2022, 7-10

Prediction of the Rockfill Dam Safety Using Long Short Term Memory

Miona Andrejević Stošović(1), Novak Radivojević(1), Jelena Marković Branković(2), Milica Marković(2), Srdjan Živković(2)

(1)University of Niš, Faculty of Electronic Engineering, Niš, Republic of Serbia
(2)University of Niš, Faculty of Civil Engineering and Architecture, Niš, Republic of Serbia

***Abstract*— Embankment dams are the oldest and the most common type of dam in use today. The rockfill embankment dam is defined as an embankment dam that relies on rockfill as major structural element. Mechanical stability of dam implies stability of slopes and foundation soil. In order to provide dam safety, pore pressures between the particles in the clay core and the phreatic line in the downstream support body should be reduced to a minimum. Measuring and analysing the values of the pore water pressure in the clay core are essential for the dam safety analysis. The idea in this paper was to predict the pore water pressure in the clay core based on the water level in the reservoir and the groundwater level. In this paper we will use a long short-term memory structure. We will try to model a system that predicts value of one pore pressure cell, and if we prove that this system has good characteristics, we can develop an overall system that has 11 pore pressures as its outputs.**

***Keywords*— long short-term memory, rockfill dam, structural safety**

I. Introduction

Dams are large and important objects not only because of their size, and the huge funds that need to be invested in their construction, but also in terms of participation in the economy of the region in which they are located [1].

According to the materials from which the dams were composed, they can be classified into two main categories - concrete and embankment dams.

Embankment dams are the oldest type of dam, at least 5000 years old, and further, the most common type of dam in use today. Mechanical stability of dam implies stability of slopes and stability of foundation soil.

In order to provide dam safety, pore pressures (PWP) between the particles in the impervious core and the phreatic line in the downstream support body should be reduced to a minimum [2], [3].

Measuring and analysing the values of the pore water pressure in the clay core are the most important for the dam safety analysis [4].

The idea in this paper was to predict the pore water pressure in the clay core based on the water level in the reservoir and the groundwater level.

We will use the long short-term memory structure of artificial neural networks and we will examine what is the optimum number of previous values of pore water pressures used as network inputs that gives the smallest prediction error, i.e. which is the number of previous days that has the most impact on network output.

II. Theoretical framework

In our previous work we tried to solve this problem in few different ways. First, a common multi-layer perceptron (MLP) was used, then a standard recurrent neural network (RNN), long short-term memory (LSTM), and finally gated recurrent unit (GRU) with a different number of inputs. Within this paper we will show our results regarding long short-term memory structure, shown in Fig. 1.

We chose LSTM structure [5] after considering disadvantages of standard RNNs. In fact, RNNs are very sensitive to input sequence length – the effects of inputs from distant moments in the past vanish as they propagate through the network. This problem with long sequences rises even more when the input sequence is too long. Then there exists a problem that a learning gradient either diverges or converges rapidly in parts of the network which are at a longer time distance from the output, so this causes that network stops learning. This makes RNNs successfully applicable only to sequences of short length.

So, in order to solve the presented problem, the LSTM incorporates a channel called "cell state" along the hidden state channel which is contained in a standard RNN.

The cell state connection spreads across consecutive LSTM cells and it is used to pass and update information that is important to many cells in the chain, no matter of their distance from the current time point.

The cell state is changed when applying activation functions and different matrix operations to its inputs, and also outputs from the previous time step and the previous cell state.

These functions are applied using three activation layers called gates: the forget gate, the input gate, and the output gate.

The function of forget gate is to direct the effect of the previous cell state on the next one - it decides if a piece of information that cell state contains need to be kept or removed.

The function of input gate is to decide if a new cell state should be affected by the input, i.e. if new information based on the input should be added to the cell state or the input should be ignored.

The output gate is of crucial importance since it generates the prediction by appointing the information that will be passed to the next long short-term memory cell depending on the new cell state and the network inputs [5].

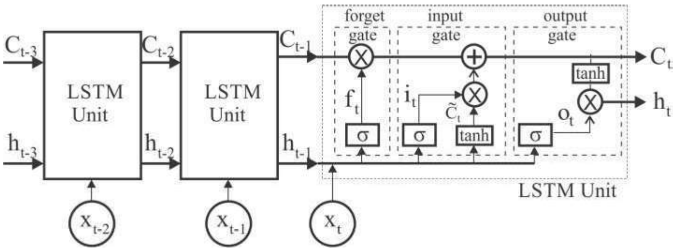


Figure 1 Long short-time memory structure

III. Development of ANN model

As a study embankment dam we used Bovan dam. It is a high rockfill dam with a clay core as impervious sealing element (Figure 2).

This dam is located at the south-east of Serbia, 50 kilometres north from Niš (Figure 3).

Figure 2 The Bovan dam

Figure 3 Location of the Bovan dam in Serbia

Based on available measured data for one year period (groundwater level, reservoir water level and air temperature), we tried to model a system that predicts value of one pore pressure cell. If we prove that this system has good characteristics, we can develop an overall system that has 11 pore pressures as its outputs.

By analysing measured data, we concluded that a system is also affected by values of pore pressures in previous time instants.

Based on these presumptions, we created an artificial neural network having all measured data and also previous values of pore pressures as inputs, and actual pore pressure as output.

In the Table 1 hydrometeorological data set for one week is given, in order to get insight into data representation. Measurement unit is meters above the sea level (MASL), for the reservoir water level (RWL), and water level in the piezometers in the dam crest (E1, E3, E4, E6, E7) and dam left and right banks (P2, P3 and P10) (Figure 4).

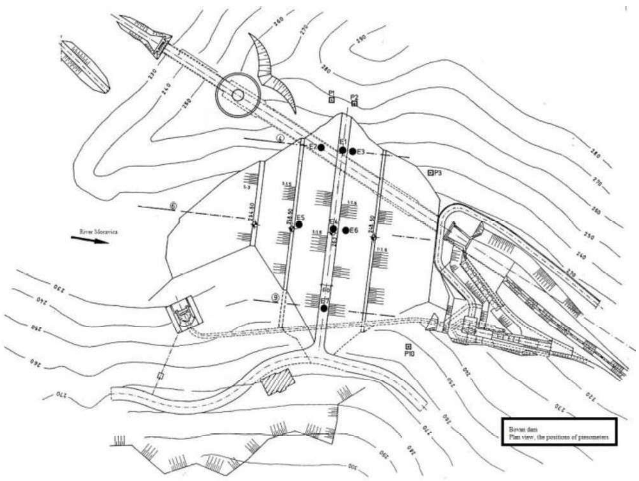


Figure 4 The positions of the piezometers

8

TABLE 1 WEEKLY HYDROMETEOROLOGICAL DATASET IN YEAR 2020

Day	RWL masl	P2 masl	P3 masl	P10 masl	E1 masl	E3 masl	E4 masl	E6 masl	E7 masl	T C°
	1	2	3	4	5	6	7	8	9	10
August 11th	253,22	245,11	221,17	218,83	261,31	244,59	235,06	219,50	249,06	26,00
August 12th	253,22	245,11	221,16	218,84	261,3	244,59	235,06	219,50	249,06	21,00
August 13th	253,28	245,12	221,16	218,84	261,3	244,59	235,06	219,49	249,06	26,00
August 14th	253,39	245,12	221,16	218,84	261,3	244,59	235,06	219,49	249,06	28,00
August 15th	253,32	245,12	221,16	218,84	261,29	244,59	235,06	219,49	249,06	28,00
August 16th	253,26	245,12	221,16	218,84	261,29	244,59	235,06	219,49	249,06	27,00
August 17th	253,17	245,11	221,15	218,84	261,29	244,59	235,06	219,49	249,05	23,00

As presented in the Table 1, hydrometeorological data are given per day, while pore water pressure data were available per hour, so we used average daily value for pore pressure.

We averaged pore water pressure as presented in the Table 2, but in neural network training process 24 hours available dataset is used, as well as average values for all other days.

Usually a test set takes 30% of overall dataset while the training set makes up 70% of overall dataset [6].

In order to estimate the values that influence the most on measured pore pressure, eight different network structures were analysed concerning numbers of previous days. We created structures with adopted data sets of 1, 3, 5, 10, 15, 20, 25 and 30 previous days.

TABLE 2 ADOPTION OF PWP VALUE FOR 24 HOURS

Date in 2020	Time (h)	PWP on PC 6-3 (kPa)
1	2	3
August 17Th	0.00	55,950
August 17Th	1.00	55,941
August 17Th	2.00	55,946
August 17Th	3.00	55,948
August 17Th	4.00	55,948
August 17Th	5.00	55,947
August 17Th	6.00	55,943
August 17Th	7.00	55,946
August 17Th	8.00	55,953
August 17Th	9.00	55,956
August 17Th	10.00	55,955
August 17Th	11.00	55,951
August 17Th	12.00	55,953
August 17Th	13.00	55,945
August 17Th	14.00	55,934
August 17Th	15.00	55,928
August 17Th	16.00	55,924
August 17Th	17.00	55,926
August 17Th	18.00	55,931
August 17Th	19.00	55,928
August 17Th	20.00	55,927
August 17Th	21.00	55,938
August 17Th	22.00	55,934
August 17Th	23.00	55,928
	Adopted PWP value	55,941

The purpose of adopting these various structures was to get results about the best structure, i.e. what is the optimal number of previous days that we use as network input, leading to achieving the best prediction. Results obtained after detailed analysis are given in the following text.

Each of these structures has 10 inputs as given in the Table 1 (RWL, P2, P3, P10, E1, E3, E4, E6, E7, T). Furthermore, each network has additional inputs, depending on its structure (data sets of 1, 3, 5, 10, 15, 20, 25 and 30 previous days).

For example, if we take into account 1 previous day, the network will have 1 additional input, determining pore pressure value for a previous day.

Similarly, if we want to include two previous days, the network will have 2 additional inputs, determining pore pressures for two previous days, and so on.

LSTM structures that we used have 2 hidden layers, where each of them has 128 hidden neurons.

This number of hidden neurons was selected after testing models with different numbers of neurons in the hidden layers. The training process consisted of iterative optimization of network parameters in order to obtain minimal root mean square error between the expected and predicted values. The network models are described in Python 3.7.9, using the Keras deep learning API from the TensorFlow platform for machine learning [7].

9

TABLE 3 RMSE FOR DIFFERENT NUMBER OF DAYS USED

Number of previous days used	RMSE for PC 6-3
1	0.058
3	0.064
5	**0.054**
10	0.073
15	0.055
20	0.055
25	0.092
30	0.122

RMSE values for all tested models are given in the Table 3. From this Table we can notice that minimum RMSE is achieved when we consider 5 previous days.

But, if that makes our network robust, we can use also networks with much simpler structure because it is obvious from the Table that error is not significantly different.

IV. CONCLUSION

The inflection of pore water pressure in the embankment dams is important for estimating dam's safety.

In this paper we considered dependency between pore pressure in the rockfill dam clay core, reservoir water level and groundwater data using LSTM structure. Based on this correlation, the ANN can predict pore water pressure or can assume what their values will be.

The obtained results show that a prediction can be made with negligible errors.

This method can be used for PWP data prediction in flood-rich period, i.e. in the spring when water level in the reservoir is on maximum. Flood periods occur mainly in early spring (March and April), the rainiest season in the Northern Hemisphere, as measured by the number of days with precipitation.

With this method, any anomaly in the PWP dataset can be detected, defining the "weak points" and possible cracks in the dam clay core.

V. ACKNOWLEDGMENTS

This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Markovic, M., Brankovic, J. M., Stosovic, M. A. et al. A New Method for Pore Pressure Prediction on Malfunctioning Cells Using Artificial Neural Networks. *Water Resour Manage* 35, 979–992 (2021). https://doi.org/10.1007/s11269-021-02763-0

[2] Ahmed, A., Elkatatny, S., Ali, A. et al. New Model for Pore Pressure Prediction While Drilling Using Artificial Neural Networks. *Arab J Sci Eng* 44, 6079–6088 (2019). https://doi.org/10.1007/s13369-018-3574-7

[3] Aniskin, N.A. et al. (2017). Filtration, pore pressure, and settling from consolidation of an ultra-high dam. *Power Technol Eng* 50, 600–605. https://doi.org/10.1007/s10749-017-0757-4

[4] Aniskin, N.A. et al. (2016). Seepage and pore pressure in the core of a earth-and-rockfill dam. Power Technol Eng, 50, 378–384. https://doi.org/10.1007/s10749-016-0717-4

[5] Hochreiter,S., Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* 9(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735.

[6] Jordan, J. Evaluating a machine learning model https://www.jeremyjordan.me/evaluating-a-machine-learning-model/ (accessed October 2021)

[7] TensorFlow: An end-to-end open source machine learning platform, https://www.tensorflow.org/ (accessed March 2021)..

10

Research paper ISSN 2603-4697 (Online) Complex Control Systems Vol. 4, No. 1, 2022, 11-15

Reviewing and discussing Graph reduction for prediction in the Edge Computing context

Asier Garmendia-Orbegozo (1), J. David Núñez-González (1), Miguel Angel Antón González (2)
(1) Department of Applied Mathematics, University of the Basque Country (UPV/EHU) Eibar, Spain
(2) TECNALIA, Basque Research and Technology Alliance (BRTA), San Sebastian, Spain

***Abstract*— Much effort has been devoted in transferring efficiently different Machine Learning (ML) algorithms and especially Deep Neural Networks (DNNs) to edge devices in order to fulfill real-time, storage and energy consumption issues, among others. Limited resources of edge devices and the necessity for energy saving to lengthen the durability of their batteries, has emerged an interesting trend on reducing neural networks and graphs while keeping almost untouched their predictability. In this work, latest works on this area are compared and analyzed in depth seeking to figure out the best techniques to reduce the dimension of those algorithms and maintain their ability at predicting. Most interesting ways of enhancing those skills are discussed, as well.**

***Keywords*— Artificial Intelligence, Edge Computing, Graph Reduction, pruning.**

I. Introduction

The use of DNNs in different scenery such as image classification, voice synthesis or object detection is undoubtedly one of the most effective ways to make predictions. The development of DNNs during the last years has evolved in such a way that nowadays neural network designs have billions of parameters with great capability of prediction, thus needing significant computation resources. Starting from huge amounts of data to be stored safely to powerful computation units, those could not be satisfied by current edge device by now. However, by reducing the size of these architectures in an efficient way it could be feasible their deployment in embedded systems.

Among others the most used and effective way to shrink these networks is the use of techniques such as pruning and quantization. The former one consists in removing parameters (neurons or weights) that have negligible contribution while maintaining the accuracy of the classifier. On the other hand, quantization involves replacing datatypes to reduced width datatypes, by transforming data to fit in new datatypes' shapes. By this way, reduced networks are able to compete with the original ones in terms of accuracy, even improving these in some cases in which overfitting issues were hindering their predictability. Moreover, by reducing the width of data edge devices could face the storage issue mentioned above and collect larger datasets in constrained memory sizes.

In this work different attempts to optimize these reduction techniques are described as well as possible future works that could be proposed to achieve even better results. The rest of this paper is organized as follows: section II introduces and analyzes the pruning process and most significant and attractive approaches made so far, section III does the same in case of quantization, and finally section IV concludes and outlines possible future research lines.

II. Pruning

Pruning consists in removing part of connections(weights) or neurons from the original network so as to reduce the dimension of the original structure by maintaining its ability to predict. The core of this technique resides on the redundancy that some elements add to the entire architecture. Memory size and bandwidth reduction are addressed with this technique. Redundancy is lowered and overfitting is faced in some scenarios. Different classifications of works based on this ability are made depending on:

- Element pruned.
- Structured / Unstructured
- Static / Dynamic

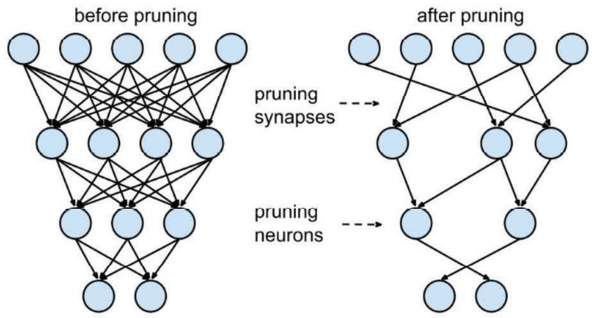


Figure 1: Different approaches for pruning. Source: https://towardsdatascience.com/ pruning-deep-neural-network-56cae1ec5505

The element pruned can be either a connection or a neuron in a pruning process. The difference between structured and unstructured pruning lies on whether the pruned network is symmetric or not. When we talk about static pruning we refer to the process in which all pruning steps are made before the inference time, while the dynamic pruning is performed during runtime.

A. Static Pruning

Static pruning is the process of removing elements of a network structure offline before training and inference processes. During these last processes no changes are made to the network previously modified. However, after removing different elements of the architecture it is interesting a fine-tuning or retraining of the pruned network. This is due to the changes that suffer the network by removing big part of its elements. Thus, some computational effort is needed in order to reach comparable accuracy to the original network.

The pruning has been carried out by following different criteria. Some works have based on the magnitude of the elements themselves. It is undoubtedly true that near-zero values of weights make far less contribution to the results than others that surpass certain threshold value. By this way, removing connections that may appear unneeded the original network is shrunk. It is an interesting approach to develop this process layer-by-layer to not affect brutely to the performance of the resulting network. Because by removing elements of the entire network some connections or neurons may take different role in the resulting network, thus being interesting some fine-tuning or retraining.

In [1, 2] they used the second derivative of the Hessian matrix to reduce the dimension of the original architecture. Optimal Brain Damage (OBD) and Optimal Brain Surgeon (OBS) respectively function under three assumptions. Quadratic: the cost function is near quadratic. Extremal: the pruning is done after the network converged. Diagonal: sums up the error of individual weights by pruning the result of the error caused by their co-consequence. Additionally, OBS avoids the diagonal assumption and improves neuron removal precision by up to 90% reduction in weights for XOR networks. Taylor expansions of first order where also considered to reduce the network dimension in [3, 4], as a criterion to approximate the change of loss in the objective function as an effect of pruning.

Some other works have followed the idea of removing elements based on different penalization terms. Penalty-based training aims to modify or add an error function to modify weights during training process using a penalty value. At the end, near-zero values are pruned from the original network. LASSO [5] was introduced as a penalty term. It shrinks the least absolute valued feature's corresponding weights increasing weight sparsity. This operation has been shown to offer a better performance than traditional procedures such as OLS by selecting the most significantly contributed variables instead of using all the variables, achieving approximately 60% more sparsity than OLS. The problem with LASSO is that is an element-wise pruning technique leading to unstructured network and sparse weight matrices. By performing this technique group-wise as it does Group LASSO [6] removing entire groups of neurons and maintaining the original network's structure. Groups are made based on geometry, computational complexity or group sparsity among others.

Other alternatives have been proposed to carry out static pruning. In [7] was proposed a novel criterion for Convolutional Neural Network (CNN) pruning called Layer-wise relevance propagation. It is measured the contribution of each unit to the relevance of the decision making. By this way, the units that are below a predefined threshold are removed from the graph and finally the relevance of each unit is recomputed. For this last step, the total relevance per layer is calculated so that to keep it untouched during iterations. Thus, each unit's relevance is recalculated to maintain this value.

In [8] a technique to prune redundant features along with their related feature maps according to their relative cosine distances in the feature space is proposed, thus leading to smaller networks with reduced post-training inference computational costs and competitive performance. Redundancy can be reduced while inference cost (FLOPS) is reduced by 40% for VGG-16, 28%/39% for ResNet-56/110 models trained on CIFAR-10, and 28% for ResNet-34 trained on ImageNet database with minor loss of accuracy. To recover the accuracy after pruning, models were finetuned for a few iterations without the need to modify hyper-parameters.

Figure 2: Architecture of the proposed approach Sparse Low Rank Decomposition in [9].

In [9] combining the ideas of sparsity and existence of unequal contributions of neurons towards achieving the target, sparse low rank (SLR) method is presented, which sparsifies Single Value Decomposition (SVD) matrices to achieve better compression rate by keeping lower rank for unimportant neurons. By this way, it is possible to save 3.6 ×storage space of SVD without much effect on the model performance. The structured sparsity achieved by the proposed approach has also the advantage of speedup in the computation.

Another interesting approach to be taken into consideration is pruning filter-by-filter. Filter-wise pruning [10] uses the l1-norm to remove filters that do not affect the accuracy of the classification. Pruning entire filters and their related feature maps resulted in a reduced inference cost of 34% for VGG-16 model and 38% for ResNet-110 model on the CIFAR-10 dataset with improved accuracy 0.75% and 0.02%, respectively. ThiNet [11] adopts statistics information from the next layer to determine the importance of each filter. It uses a greedy search to prune the channel that has the smallest reconstruction cost in the next layer. During each training pruning is carried out more lightly to allow for coefficient stability. The pruning ratio is a predefined hyper-parameter and the runtime complexity is directly related to the pruning ratio. ThiNet compressed ResNet-50 FLOPs to 44.17% with a top-1 accuracy reduction of 1.87%.

Other research has been carried out attending activations, that may also be indicators to prune corresponding weights. Average Percentage of Zeros (APoZ) [12] was introduced to judge if one output activation map is contributing to the result. Some activation functions, particularly rectification functions such as Rectified Linear Unit (ReLU), may result in a high percentage of zeros in activations, being interesting their pruning.

After applying different techniques to reduce the amount of non-relevant elements from the original structure, it is essential a fine-tuning or retraining phase. It is shown [8] that by training a pruned structure from scratch less accurate results are obtained compared to the retraining processes in which weights from the original network are maintained for the new training phase. That is why iteratively a retraining or fine-tuning step is

12

followed after each pruning step is carried out. This iterative process is repeated until a desired number of elements is achieved.

B. Dynamic Pruning

Pruning a DNN dynamically offers several benefits compared to the same process carried out offline before both training and inference processes. Identifying at runtime which elements of the original structure are relevant and which ones are not, offers the possibility of solving different issues related with static pruning by adapting the network with the changes of input data.

This process is far more complex than the static one, so that various decisions are needed to make before starting it. In some cases makes sense considering additional networks or connections to further assist pruning process. Information input could be either layer-by-layer feeding a window of data iteratively to the decision system or by one-shot feeding. As well as in the static pruning, a score system and a comparative system (automatic or manual) must be established. Similarly, a stopping criterion must be imposed, and finally, the additional components have to be trained at the same time the network has been trained.

A negative impact to the system computation requirements is also needed to be taken into account. Additional bandwidth, computation and power resources are necessary while computing at runtime which elements to be pruned. At the same time, convolution operations with large number of features consume huge part of the bandwidth. Thus, a trade-off between dynamic pruning overhead, reduced network computation, and accuracy loss, should be considered. Different approaches have been developed during recent years, and the most significant ones are described below.

In [13, 14] they focused on conditional computing by activating relevant parts of the original network. The non-activated elements act as pruned ones enlightening the original structure.

The main advantage that dynamic pruning offers is the capacity of adapting the pruned network at runtime. By obtaining intermediate trained models while carrying out the whole process is an interesting way of applying a trade-off between accuracy and computation cost. In [15,16, 17] different alternatives of cascade networks were proposed. A cascade network consists of a series of networks that each of them has its output layer, instead of offering an output per-layer. Its main advantage is that it could offer an early exit if desired accuracy is achieved. On the contrary, some hyper-parameters need to be tuned manually. Moreover, in [18] Blockdrop was introduced as an Reinforcement Learning method that with an input image was able to deduce which blocks should participate in the whole process. They were able to achieve an average speed-up of 20% on ResNet-101 for ILSVRC- 2012 without accuracy loss. On the other hand, Runtime Neural Pruning (RNP) was proposed [19] based on a feature selection problem as a Markov Decision Problem (MDP) finding computation efficiency. A Recursive Neural Network (RNN) based network was used to predict which feature maps were necessary. They found 2.3x to 5.9x reduction in execution time with top-5 accuracy loss from 2.32% to 4.89% for VGG-16.

In [20] a novel dynamic pruning technique based on pruning and splicing was presented. On the one hand, pruning operations can be performed whenever the existing connections seem to become unimportant. On the other hand, the mistakenly pruned connections shall be re-established if they once appear to be important (splicing). Experimental results show that their method compressed the number of parameters in LeNet-5 and AlexNet by a factor of 108x and 17.7x, respectively, with a better learning efficiency.

The negative point of RL techniques is their computation expense. Alternatively, differentiable approaches have been made to solve this issue. Using Dynamic Channel Pruning (DCP) in [21] they proposed a side network called Feature Boosting and Suppression (FBS) to decide which channel to skip. FBS achieved 5x acceleration on VGG-16 with 0.59% ILSVRC-2012 top-5 accuracy loss, and 2x acceleration on ResNet-18 with 2.54% top-1, 1.46% top-5 accuracy loss. Similarly, in [22] a channel-threshold weighting decision (T-Weighting) was used to prune dynamically channels. A T-sigmoid activation function, using as its entry a downsampling from a Fully Connected Layer (FCL), was used to calculate channels' score and decide which ones to prune.

Another interesting approach has been proposed in [23] to prune dynamically CNNs. They explore the manifold information in the sample space to discover the relationship between different instances from two perspectives, *i.e.*, complexity and similarity, and then the relationship is preserved in the corresponding sub-networks. An adaptive penalty weight for network sparsity is developed to align the instance complexity and network complexity, while the similarity relationship is preserved by matching the similarity matrices. Extensive experiments are conducted on several benchmarks to verify the effectiveness of this method. Compared with the state-of-the-art methods, the pruned networks obtained by this method can achieve better performance with less computational cost. For example, they can reduce 55.3% FLOPs of ResNet-34 with only 0.57% top-1 accuracy drop in ImageNet.

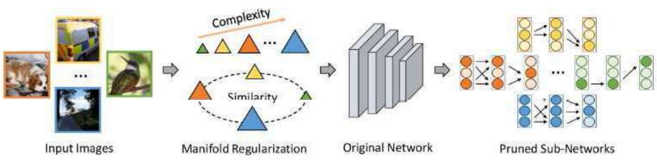


Figure 3: diagram of the architecture proposed in [23].

III. QUANTIZATION

Reducing the number of bytes used to represent data is crucial when transferring ML algorithms to resource constrained edge devices. When we mention quantization, we refer to the transformation of data from floating point representation to a determined value range. These former data could be either represented by predefined values or symbols. In the context of DNNs weights may be quantized by clustering processes, by lookup tables or using linear functions, all of them with the aim of reducing information width of data. Originally, most of the DNNs applied 32-bit floating point representation for weight parameters, but in most cases this falls excessive. In fact, 8-bit values can accelerate significantly inference process without observable loss in accuracy. Different alternatives have

13

been adopted in recent years, and some of the most interesting ones are described and analyzed below.

In [24] they conduct extensive experiments using incremental quantization on three applications: medical image segmentation, image classification, and automatic speech recognition. The main goal of incremental quantization is to convert 32-bit-floating-point weights (*W)* into low precision weights *W'* either power of 2 or zero with minimum accuracy loss. Each of *W'* is chosen from $Pl = \{\pm 2^{n1}, \dots, \pm 2^{n2}\}$, here $n1$ and $n2$ are two integer numbers determined by the max absolute value of *W* of each layer and expected quantized bit-width, and $n2 \leq n1$. Experimental results in FCN for biomedical image segmentation, CNN for image classification, and RNN for automatic speech recognition, show that incremental quantization can improve accuracy by 1%, 1.95%, and 4.23% on the three applications, respectively, with 3.5x to 6.4x memory reduction.

An adaptive quantization method was introduced [25] to enhance quantization results on MNIST, CIFAR-10 and SVHN datasets, finding a unique, optimal precision for each network parameter such that the increase in loss is minimized. Pruning unnecessary parameters or quantizing them they showed DNN model sizes can be reduced significantly without loss of accuracy. The resulting models were significantly smaller than state-of-the-art quantization technique.

Following the trend of mixing pruning and quantization techniques, in [26] they presented a training acceleration framework able to speed up training process while compressing DNN for mitigating transmission overhead. FL-PSQU is a Federated Learning mechanism that is divided in three steps. First, a one-shot pruning is done by the server to generate general models for all clients and after quantizing it, it is transferred to each client. Then, each client updates its model and depending on its update's significance it is transmitted to the server or not, avoiding unnecessary communications by this way.

Other simpler approaches have been made that include binary and ternary quantization. Thus, only two or three possible values were assigned to each element of the architecture, with a vast reduction in memory size and computation effort. However, accuracy loss is not negligible in these techniques due to the hard generalization of weights. Binary Connect [27] used stochastic plus-minus quantization by assigning +1 to positive valued weights and -1 to negative valued ones with hard sigmoid probability σ(x) and 1- σ(x), respectively. Similarly, in [28] they binarized all weights of different architectures and afterward multiply with a scaling factor for all the weights of a layer. In [29, 30] showed the power of ternary quantization by including additional value (-1, 0 and +1) compared to the binary case. Its implementation in hardware must be efficient due to the fact that 0 value does not actually participate in computations. Ternary Weight Networks (TWN) adopts the l2-distance to find the scale and formats the weights into -1, 0 and +1 with a threshold generated by an assumption that the weighs are uniformly distributed such as in [-a, a]. This resulted in up to 16x model compression with 3.6% ResNet-18 top-1 accuracy loss on ILSVRC-2012.

In some scenarios could be of special interest the conversion of floating-point multiplication to bit-shift operations, such as in case of FPGAs. Here, the constraint for weights for being power of two, leverages training and inference effort and time. This approach was proposed in [31] by quantizing the representation of weights layer-by-layer. Likewise, Incremental Network Quantization [32] replaces all weights with powers of 2 iteratively, preserving in each iteration some weights in full precision and retraining them. After multiple iterations majority of the weights are converted to power-of-two. The final structure has weights from 2 to 5 bits with values near zero set to zero. Results of group-wise iterative quantization show lower error rates than a random power-of-two strategy.

IV. CONCLUSION AND FUTURE WORK

Among different approaches that have been made so far in the literature to reduce the dimension of the original network size and optimize inference time, memory usage and computational cost, the ones mentioned above are the most interesting ones from our point of view. When transferring deep networks usually trained on cloud to edge devices, it is of great interest a reduction in network size to adapt these architectures to the constraints of such devices. Thus, pruning original network is undoubtedly an essential step to fit deep models in resource constrained devices. Dynamic pruning offers the possibility of removing unnecessary elements of the network at inference time, thus, offering way of achieving the most accurate reduction for the input data. In fact, these techniques are computationally more expensive than the static ones, and quite a bit complex. For some applications some approaches will suit better than others and it is up to the final user the election among them. However, achieving a computationally effortless way of pruning networks dynamically would be an interesting future research line.

On the other hand, quantization techniques would not be underestimated specially when transferring models to microcontrollers that lack of operating systems. Specially when those microcontrollers own FPGAs, a conversion from floating-point multiplications to bit-shift operation could alleviate a great computational cost and reduce runtime, as well. In the majority of scenarios reducing the number of representative bits to 8 is enough for avoiding a significant loss in terms of accuracy while reducing vastly memory usage. Nonetheless, some other approaches like binary quantization are too hard in many cases leading to a excessive accuracy drop.

REFERENCES

[1] Y. LeCun, J.S. Denker, S.A. Solla, Optimal Brain Damage, Advances in Neural Information Processing Systems (NIPS) (1990) *598–605*, https://doi.org/ 10.5555/109230.109298.

[2] Hassibi, B., et al. Optimal brain surgeon and general network pruning. 10.1109/icnn.1993.298572. 1993.

[3] P. Molchanov et al. Pruning convolutional neural networks for resource efficient transfer learning. Proceedings of International Conference on Learning Representations (ICLR) 2017.

[4] C. Yu et al. Transfer channel pruning for compressing Deep domain adaptation models. Int. J. Mach. Learn. Cybern, 10 (11) (2019) *3129-3144.*

[5] R. Muthukrishnan, R. Rohini. LASSO: A feature selection technique in predictive modeling for machine learning, in: 2016 IEEE International Conference on Advances in Computer Applications (ICACA), IEEE, 2016, pp. *18–20*, https://doi.org/10.1109/ICACA.2016.7887916.

[6] M. Yuan, Y. Lin. Model selection and estimation in regression with grouped variables, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68 (2006) *49–67*, https://doi.org/10.1111/j.1467-9868.2005.00532.x, http://doi.wiley.com/10.1111/j.1467-9868.2005.00532.x.
[7] S.K. Yeom et al. Pruning by explaining: A novel criterion for Deep neural network pruning. Pattern Recognition 115 (2021) 107899.
[8] B. O. Ayinde et al. Redundant feature pruning for acelerated inference in deep neural neutworks. Neural Networks 118 (2019) *148-158.*
[9] S. Swaminathan et al. Sparse low rank factorization for deep neural network compression. Neurocomputing 398 (2020) *185-196*.
[10] H. Li et al. Pruning Filters for Efficient ConvNets, in: International Conference on Learning Representations (ICLR), 2017, https://doi.org/10.1029/2009GL038531.
[11] J.H.H. Luo, J. Wu, W. Lin. ThiNet: A Filter Level Pruning Method for Deep Neural Network Compression, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2017-Octob, *5068–5076*, https://doi.org/10.1109/ICCV.2017.541.
[12] H. Hu et al. Network Trimming: A Data-Driven Neuron Pruning Approach towards Efficient Deep Architectures. 2016. ArXiv preprint http://arxiv.org/abs/1607.03250. .
[13] Y. Bengio. Estimating or Propagating Gradients Through Stochastic Neurons. 2013. ArXiv preprint http://arxiv.org/abs/1305.2982. .
[14] A. Davis, I. Arel. Low-Rank Approximations for Conditional Feedforward Computation in Deep Neural Networks, International Conference on Learning Representations.
[15] S. Leroux et al. The cascading neural network: building the Internet of Smart Things. Knowl. Inf. Syst. 52 (2017) *791–814,* https://doi.org/10.1007/s10115-017-1029-1, http://link.springer.com/10.1007/s10115-017-1029-1.
[16] T. Bolukbasi et al. Neural Networks for Efficient Inference, in: Thirty-fourth International Conference on Machine Learning, 2017.
[17] A. Odena et al. Changing Model Behavior at Test-Time Using Reinforcement Learning, in: International Conference on Learning Representations Workshops (ICLRW), International Conference on Learning Representations, ICLR. 2017. http://arxiv.org/abs/1702.07780. .
[18] Z. Wu et al. BlockDrop: Dynamic Inference Paths in Residual Networks, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2018, pp. *8817–8826*, https://doi.org/10.1109/CVPR.2018.00919.
[19] J. Lin et al. Runtime Neural Pruning. Advances in Neural Information Processing Systems (NIPS). 2017. *2178–2188*, https://papers. nips.cc/paper/6813-runtime-neural-pruning.pdf.
[20] Y. Guo et al. Dynamic Network Surgery for Efficient DNNs. 30th Conference on Neural Information Processing Systems (NIPS). 2016. Barcelona, Spain.
[21] X. Gao et al. Dynamic Channel Pruning: Feature Boosting and Suppression, in: International Conference on Learning Representations (ICLR), 2019, pp. *1–14*, http://arxiv.org/abs/1810.05331.
[22] Z. Chiliang et al. Accelerating Convolutional Neural Networks with Dynamic Channel Pruning, in: 2019 Data Compression Conference (DCC), IEEE, 2019, p. *563*, https://doi.org/10.1109/DCC.2019.00075.
[23] Y. Tang et al. Manifold Regularized Dynamic Network Pruning. Computer Vision Foundation. 2021.
[24] W. Chen et al. Quantization of Deep Neural Networks for Accurate Edge Computing. ACM Journal on Emerging Technologies in Computing Systems, Vol. 17, No. 4, Article 54. 2021. https://doi.org/10.1145/3451211
[25] S. Khoram and J. Li. Adaptive Quantization of Neural Networks, in: ICLR 2018.
[26] W. Xu et al. Accelerating Federated Learning for IoT in Big Data Analytics With Pruning, Quantization and Selective Updating. IEEE Access Vol. 9, 2021. *38457-38466.*
[27] M. Courbariaux, et al. Binarized Neural Networks: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1. 2016. ArXiv preprint https://github.com/MatthieuCourbariaux/ http://arxiv.org/abs/1602.02830.
[28] M. Rastegari et al. XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks, in: European Conference on Computer Vision, Springer. 2016. pp. *525–542*. http://arxiv.org/abs/1603.05279 http://link.springer.com/10.1007/978-3-319-46493-0_32, DOI: 10.1007/978-3-319-46493-0_32. .
[29] F. Li et al. Ternary Weight Networks, in: Advances in Neural Information Processing Systems (NIPS). 2016. http://arxiv.org/abs/1605.04711.
[30] C. Leng et al. Extremely Low Bit Neural Network: Squeeze the Last Bit Out with ADMM, in: The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), 2018.
[31] Z. Lin et al. Neural Networks with Few Multiplications, in: International Conference on Learning Representations (ICLR), 2016.
[32] A. Zhou et al. Incremental Network Quantization: Towards Lossless CNNs with Low-Precision Weights, in: International Conference on Learning Representations (ICLR), 2017.

15

Research paper ISSN 2603-4697 (Online) Complex Control Systems Vol. 4, No. 1, 2022, 16-21

Design and Implementation of a LIDAR Based Range Sensor System

Stefan Hensel(1), Marin B. Marinov(2), Markus Obert(1) and Dimitre Trendafilov(2)
(1)University of Applied Sciences Offenburg, Department for Electrical Engineering, Offenburg, Germany
(2)Technical University of Sofia, Department of Electronics, Sofia, Bulgaria

***Abstract*— Positioning mobile systems with high accuracy is a prerequisite for intelligent autonomous behavior, both in industrial environments and in field robotics. This paper describes the setup of a robotic platform and its use for the evaluation of simultaneous localization and mapping (SLAM) algorithms. A configuration using a mobile robot Husky A200, and a LiDAR (light detection and ranging) sensor was used to implement the set-up. For verification of the proposed setup, different scan matching methods for odometry determination in indoor and outdoor environments are tested.**

***Keywords*— LiDAR, LOAM (Lidar Odometry and Mapping), mapping, machine learning, mobile robotics, navigation, ROS, SLAM.**

I. INTRODUCTION

The range of areas in which autonomous mobile systems are used is growing continuously. The use of autonomous transport vehicles in industry or robots in private households has now become standard. The development of autonomous motor vehicles is also progressing steadily. Machine learning offers the field of robotics a set of tools for designing difficult and complex behaviors. On the other hand, the challenges of robotics-related problems also provide a positive impact on developments in robot learning.

In all use cases, the basic challenge is more or less the same. Autonomous mobile systems must simultaneously estimate their position in an unknown environment and simultaneously create a map of the environment. This challenge is also referred to as the simultaneous localization and mapping (SLAM) problem. While filter-based approaches were the most common solution for this problem before 2010, graph SLAM is now the most popular and efficient approach [1, 2]. In this approach, a robot's landmarks and poses are represented by a graph, which allows the SLAM problem to be solved via nonlinear optimization techniques.

The SLAM problem refers to the difficulty of locating and mapping a mobile robot in an unknown environment with its simultaneous positioning relative to this map [3]. When no other navigation capabilities, such as GNSS, are available, the SLAM problem becomes more important. While already solving the problem in simple applications, SLAM algorithms can be pushed to their limits by challenging dynamic robot motions or highly dynamic environments [4]. To obtain a map, sensors must be used to detect the structure of the environment. A variety of possible sensor types are available for this purpose.

Finally, by determining the position of the environmental features, it is possible to obtain a representation of the robot's environment and thus a map that can be used in various ways, such as localization. The basic problem within SLAM is to estimate the trajectory of the robot as well as the position of all environmental features without knowing the true position of the features or the robot itself [3, 5]. LiDAR sensors, in particular, play a key role in sensing the environment of mobile systems [6, 7].

II. RANGE SENSING BASICS

Range sensors are devices that capture the 3-D structure of surrounding objects from the sensor's perspective. They typically measure the distance to the nearest surfaces - that part of the scene is "visible" from the sensor. There is no full three-dimensional observation of all sides of the scene, and so in the field of sensory research is increasingly talking about two-and-a-half dimensional (2.5-D) data. (2.5-D).

The range data is a two-dimensional (2.5-D) or three-dimensional representation of the environment around the robot. The three-dimensional aspect arises because the coordinates (X, Y, Z) of one or more points in the scene are measured. But usually, only the fronts of objects are observed - that part of the scene that is visible from the robot. Typically, we do not have a full three-dimensional observation of all scene sides. Hence the term "2.5-D."

Two basic range measurement technologies exist *triangulation* and *time-of-flight* measurement, and there are multiple variations exist of each.

A. Triangulation

Triangulation sensors measure depth by determining the angle formed by rays from a world point to sensors located at a distance b. This so-called baseline of length b separates the sensors, assuming for simplicity that one of the rays forms a right angle with the baseline. Angle θ of the other sensor ray is then related to the depth Z perpendicular to the baseline through the relation:

$$\tan\theta = \frac{Z}{b}. \qquad (1)$$

An image sensor measures the angle θ by an offset in the image plane relative to the primary beam. This shift x is denoted as misalignment. If the image plane is assumed to be parallel to the baseline, then $\tan\theta = f/x$, and the basic equation of triangulation depth sensors is obtained

$$Z = \frac{fb}{x}. \quad (2)$$

B. Time of Flight

The method used in time-of-flight (TOF) sensors is like that of a radar: it measures the time it takes light to reach an object and return. Since light travels at about 0.3 m per nanosecond, a very accurate TOF measurement is needed.

As they measure time-of-flight, these sensors can theoretically have constant accuracy irrespective of the distance to the object - unlike triangulation sensors, whose accuracy decreases with the square of the distance to the object. However, TOF sensors cannot achieve the high accuracy of triangulation sensors for closely spaced objects, and so they are not typically used in close-range applications.

1) Direct Time of Flight

In direct TOF sensors, travel time is measured with a high-speed stopwatch. Direct TOF laser range sensors are also called LiDAR or LaDAR (laser radar sensors). The travel time multiplied by the speed of light (in a given medium - space, air, or water and corrected for the density and temperature of the medium) gives the distance

$$2d = ct, \quad (3)$$

where d is the distance to the object, c is the light speed, and t is the travel time. The error in the measurement of time t results in a proportional error in the distance. In practice, an attempt is made to measure the peak of the output pulse, which has a finite range; weak reflections from distant objects make this peak difficult to measure, and therefore the error tends to increase with the distance. Multiple readings averaging can reduce the random error in these measurements [8].

The simplest TOF sensors use only a single beam, so range measurements are obtained from only one point on the surface. Robotics applications typically need significantly more information. To obtain this information, the laser beam is moved across the stage. Typically, the beam is moved using a set of mirrors [8].

Typical ground based TOF sensors suitable for robotics applications have a range of $10 - 100$ m and an accuracy of $5 - 10$ mm. The volume of the scanned scene depends on the speed of the moving mirrors and the pulse rate, with typical values $1000 - 25000$ points per second.

The scanning multibeam LiDARs can increase the amount of information available. Companies such as Velodyne and Oster produce devices with 16, 32, 64, and 128 vertically aligned beams that acquire point data at up to 15 $scans/s$ (1.3 $MPixel/s$), with a full 360-inch horizontal scan and 27-inch vertical FOV from the laser array. The laser pulse length is 5 ns and the depth accuracy is approximately 2 cm. These devices are frequently used in environmental reconstruction, autonomous driving, and obstacle avoidance.

Flash LIDAR

Flash LIDARs have a two-dimensional detector array, unlike scanning devices. Instead of one or several laser beams, the light source pulse is shaped to cover a large area. All pixels start their timers when the pulse is initiated and measure the time it takes to receive the backscattered light. Typically, a few dozen samples are taken and averaged to reduce noise in the measurements - the amount of energy received is quite small since the laser is not focused on a beam. The pixels on the detector array are quite large because of the timing electronics; a typical ASC device has $128x128$ pixels and can capture data at up to 60 Hz. These devices are expensive and therefore not used in consumer applications [8, 9].

2) Indirect Time of Flight Sensors

In indirect TOF sensors, the distance is measured and transit times are determined from certain properties of the propagating beam. The two most important methods are based on modulation and phase differences and signal intensity [8].

III. Implementation

A. Hardware

1) LiDAR OS1

Fig. 1 shows the OS1 LiDAR from the company Ouster, which was used for this work. With the help of an interface box, the LiDAR sensor can be connected to a computer via a LAN cable. For the operation of the sensor, a 24 V power supply is also required.

The Ouster LiDAR is suitable for distances between 0.3 m to 100 m and has a vertical field of view of 45° (±22.5°). It has a vertical resolution of 128 lines and a configurable horizontal resolution of $512, 1024$, or 2048 lines. Depending on the resolution, the LiDAR sensor can scan its environment at 10 Hz or 20 Hz. Thus, at a resolution of 128x2048 and the frequency of 10 Hz, up to 2621440 points can be captured by the sensor within one second, corresponding to a data rate of up to 254 Mb/s [10].

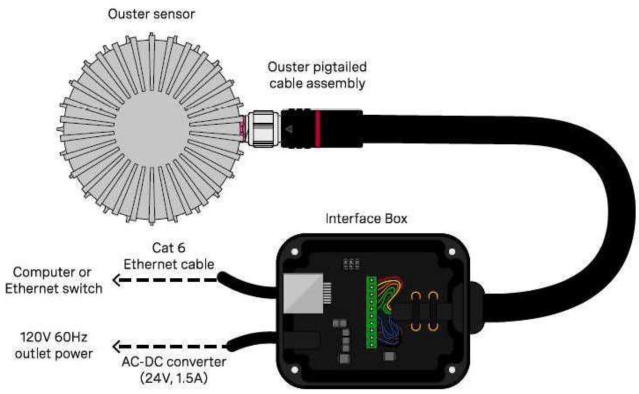


Fig. 1. Ouster OS1 with 128 lines [11].

For later outdoor applications, the raster size is also of interest. The determination of the raster size with which the environment is scanned, as a function of the distance, can be determined using trigonometric relationships. This relationship is shown in Fig. 2.

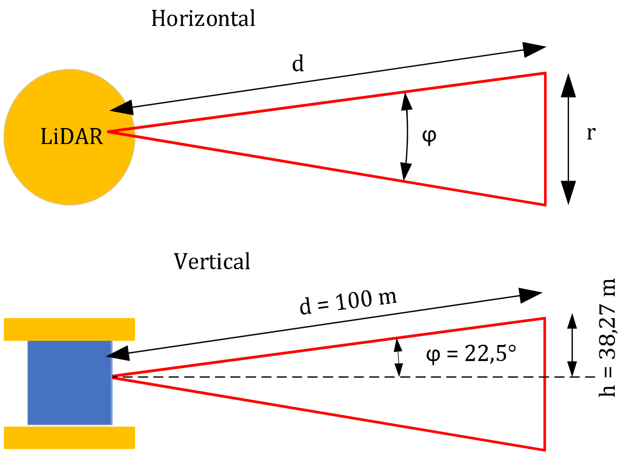


Fig. 2. Up: Horizontal distance between two laser beams as a function of distance; Down: Vertical field of view as a function of distance (100 m).

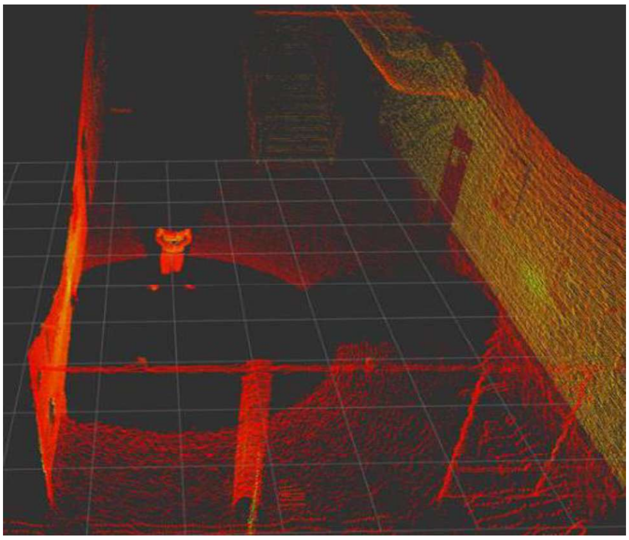
On the one hand, the maximum distance of 100 *m* is considered, which is particularly interesting for outdoor applications. On the other hand, a distance of 10 *m* is considered for the indoor area. The raster values, depending on the resolution and the distance, are listed in Table 1.

Table 1. Distance of the measuring points depends on the distance [12].

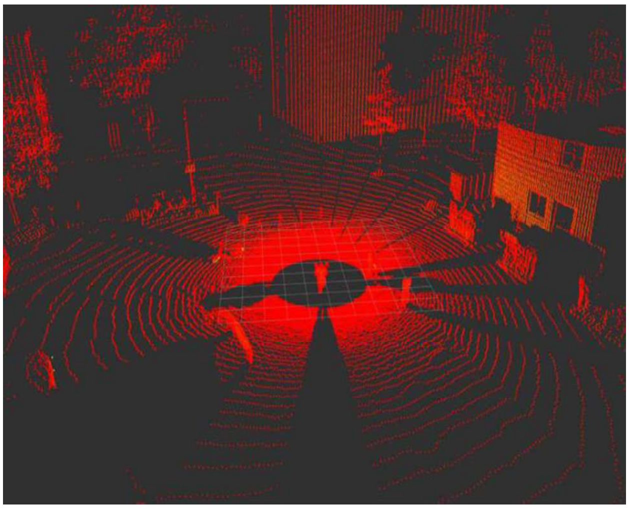
Alignment	Resolution, points	Field of view, °	Angle betw. 2 beams, °	Distance, m	Raster-size, cm
Vertical	128	45	0.35	100	61.36
				10	6.14
Horizontal	512	360	0.70	100	122.72
				10	12.27
	1028	360	0.35	100	61.12
				10	6.11
	2048	360	0.18	100	30.68
				10	3.07

For distances in the range of 10 *m*, a vertical grid of approx. 6 *cm* and a horizontal grid between 3 *cm* and 12 *cm* is obtained. In the outdoor area, a vertical grid of approx. 61 *cm* and a horizontal grid between 30 *cm* and 120 *cm* are obtained.

For indoor applications, where distances are smaller and there are many flat surfaces such as walls and floors, a horizontal resolution of 512 measurement points is sufficient. In the outdoor area, where there are considerably larger distances and increasingly complex structures such as trees, a horizontal resolution of 2048 measuring points is advantageous. This can be seen, for example, in the following images of the laser scanner in the indoor and outdoor areas (Fig. 3).



a)



b)

Fig. 3. Laser scanning in a) indoor and b) outdoor areas with the Ouster LiDAR OS1.

2) Husky A200

The UGV (Unmanned Ground Vehicle) "Husky A200" from Clearpath Robotics, which can be seen in Fig. 4, is used for this work.

The robot platform, which is equipped with an all-wheel-drive, can be remotely controlled with the aid of a controller. In this chapter, the structural design of the robot platform is described. First, the hardware structure is considered. This is followed by a brief general introduction to ROS (Robot Operating System) and the software structure.

18

Fig. 4. Laser scanning in indoor and outdoor areas with the Ouster LiDAR OS1 [12].

A structural diagram of the hardware setup is shown in Fig. 5. Inside the Husky is the main computer, which contains the basic system for the Husky. When the rover is switched on, the main computer and thus all necessary processes are started automatically. The MCU (Motor Control Unit) is also connected to the Husky computer via a serial RS232 interface, which controls the motors of the robot platform and receives the odometry data [13].

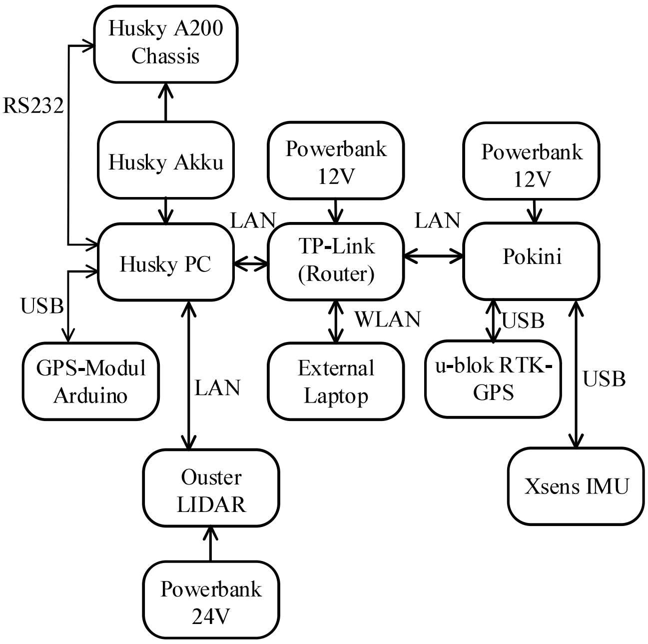


Fig. 5. The hardware structure of the robot platform (based on [13]).

In addition to the main computer, there is a second computer (Pokini) that is connected to the main computer via a router. The two computers are each connected to different sensors and have the necessary software packages accordingly. A GPS (Global Positioning System) module is connected to the Husky as well as the 128 lines LiDAR OS1 from Ouster. The Pokini is equipped with an IMU (Inertial Measurement Unit) from X-sense for inertial data acquisition and an RTK-GPS (Real-Time Kinematic-GPS) from u-blox. This will later be used as a reference to evaluate the accuracy of the trajectory estimation. The two computers can be accessed and communicated with via the WLAN router using an SSH connection.

B. *Software*

ROS is used as the framework for exchanging messages and controlling the Husky. In the following, the functionality of ROS is described first and then the structure of the ROS network of this robot platform is described.

1) *Robot Operating System (ROS)*

ROS (Robot Operating System) is an open-source platform for programming robot systems. It is a meta-operating system that provides various tools for simulation and visualization as well as libraries. The libraries mainly include hardware abstractions as well as device drivers for sensors and actuators [14].

ROS is a peer-to-peer network, which means that all participants have equal rights and that services can be offered and used. The individual participants can be distributed over several computers, whereby they are loosely coupled by the ROS communication infrastructure. To organize communication, there is exactly one master in each ROS network that manages communication and with which each node must be registered, as illustrated in Fig. 6.

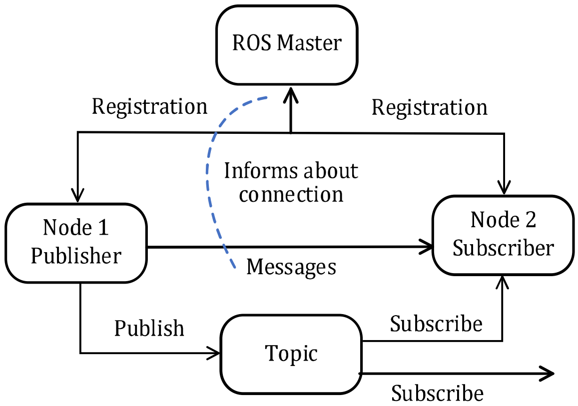


Fig. 6. Structural organization of ROS [14].

ROS nodes represent programs executable via a terminal, which can be compiled, executed, and managed individually. The organization of the Nodes takes place in packages, which contain the source code, Launch files, and configuration files [14, 15].

19

The nodes communicate with each other via so-called topics. Each node can publish different topics or subscribe to topics that are of interest to it. For example, sensor data can be shared via a topic and subscribed to by the nodes that require this sensor information.

2) ROS structure Husky

When the Husky is started, a ROS core, which represents the master, is automatically started on the Husky computer. All necessary ROS packages that are required for the operation of the Husky are executed. If the corresponding sensors are to be started or used, they must be started manually, for example via an SSH connection, on the respective computer on which the corresponding software packages are installed. For example, the node of the Ouster LiDAR sensor is started on the Husky computer. The exact description of which packages can be started on which computer and how is given in the GitLab repository. Finally, the hardware setup, as described in Fig. 5, allows for subscribing or publishing the desired topics from any computer, enabling easy cross-network data exchange.

IV. EXPERIMENTAL RESULTS WITH HDL_GRAPH_SLAM

A. *The parking lot of the University of Offenburg*

The first outdoor test was recorded in the parking lot of Offenburg University. On the day of the test, the parking lot was sparsely filled with cars, as can be seen in Fig. 7.

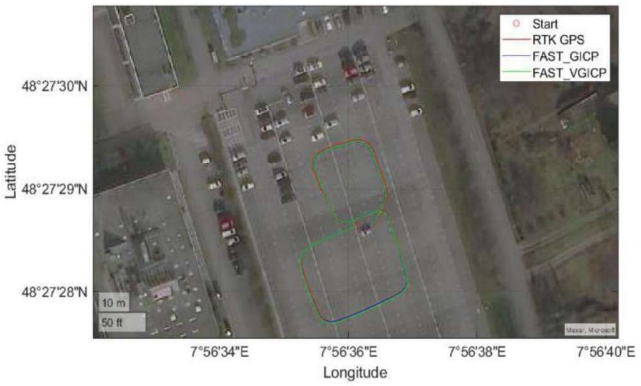


Fig. 7. Trajectories of the different scan matching methods (FAST_GICP, FAST_VGICP, and RTK GPS as ground truth) in the parking lot of the Offenburg University.

The parking lot is almost flat and very wide so that only ground surfaces are detected near the Husky robot, and trees and buildings can be detected by the LiDAR sensor from approx. 40 m. The LiDAR sensor is also able to detect trees and buildings. As a ground truth (reference), the trajectory was also determined using an RTK GPS. For this purpose, the base station was calibrated at the position of the car at the starting point. The RTK GPS can achieve the accuracy of 2 cm when the base station is calibrated. The total distance of the trajectory is 180 m with a duration of 223 s.

For verification, different scan matching methods for odometry determination are tested again. As in the indoor area, FAST_GICP and FAST_VGICP were shown to be the most suitable for this application. While the hdl_graph_slam with NDT_OMP deviates significantly from the ground truth up to the first loop closure, the deviation of the hdl_graph_slam with the other two algorithms is a maximum of 20 cm (Fig. 8).

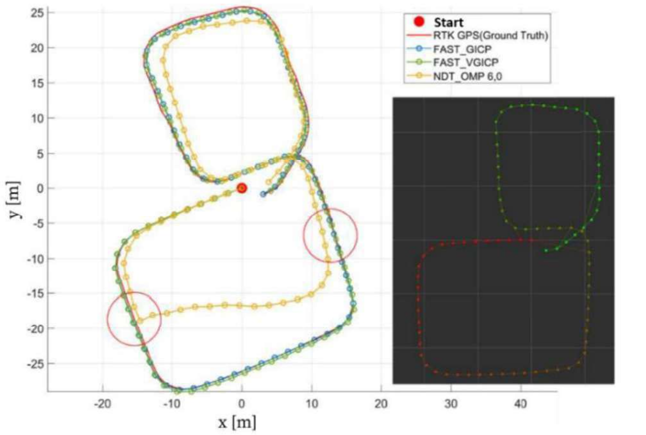


Fig. 8. Left: Trajectories of the different scan matching methods (FAST_GICP, FAST_VGICP, NDT_OMP, and RTK GPS as ground truth) in the parking lot of the University of Offenburg. Right: Trajectory with four loop closures found.

After the first Loop Closure, the deviation is reduced to 2 cm with all scan matching algorithms. In the second loop, the deviation with NDT_OMP is up to 3 m, but decreases again to a maximum of 1 m at the second loop closure. With FAST_GICP and FAST_VGICP, the deviation is 50 cm at the most and decreases again at the Loop Closure at the end to 20 cm at the most. The loop closures found by hdl_graph_slam can be seen on the right-hand side in Fig. 8.

B. *Check of the altitude course*

In addition to the trajectory, the altitude course is also checked. The RTK GPS cannot be used as a reference, since its elevation values vary by several meters (Fig. 9, b) and this is a flat parking lot. The values of the hdl_graph_slam, however, only vary by ±5 cm.

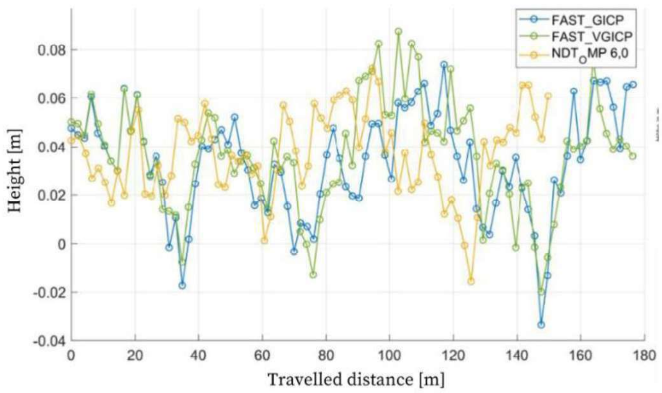


Fig. 9. Elevation course of the trajectory of the different scan matching methods (FAST_GICP, FAST_VGICP, and NDT_OMP) on the parking lot of the University of Offenburg.

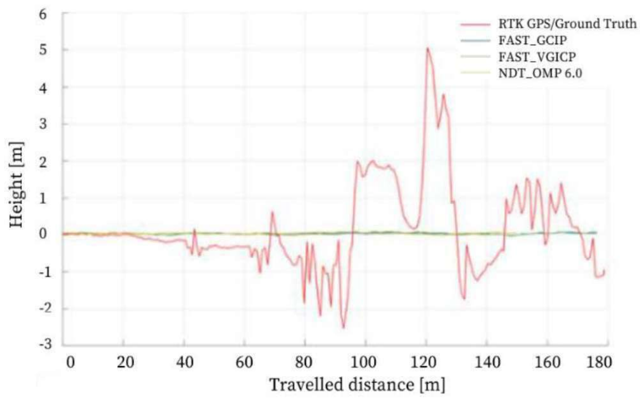


Fig. 10. Altitude course of the RTK GPS (red).

V. Conclusion

This paper presents a system for the implementation and evaluation of complex localization and mapping procedures. The considered methods are suitable for reliable localization in complex scenarios, such as logistics applications in Industry 4.0, exploration and sensing in field robotics, or applications in service robotics.

Currently, a major problem is the precise positioning of mobile systems, which is a prerequisite for any autonomous behavior in industrial environments and field robotics. The paper describes the setup of an experimental platform and its use for the evaluation of SLAM algorithms

With each scan of a LiDAR sensor, a point cloud with measurement points is obtained, as shown in the previous chapter. To be able to use these point clouds for mapping or localization in a map, the challenge is to "match" the current scan with an already known scan of the surrounding area. The process of matching two scans is also called registration. There are different scanmatching algorithms for registration.

Among the most common are the Iterative Normal Distribution Transform (NDT), Closest Point (ICP), and Voxelized Generalized Iterative Closest Point (VGICP) algorithms. In the next stages of the study, different algorithms will be investigated and compared in indoor and outdoor spaces.

Acknowledgments

This research is supported by the Bulgarian National Science Fund in the scope of the project “Exploration the application of statistics and machine learning in electronics” under contract number КП-06-Н42/1.

References

[1] M. Ivanova, P. Petkova and P. Petkov, "Machine Learning and Fuzzy Logic in Electronics: Applying Intelligence in Practice," *Electronics,* vol. 10, no. (22):2878, 2021.

[2] S. Hensel, M. B. Marinov, C. Kehret and M. Stefanova-Pavlova, "Experimental Set-up for Evaluation of Algorithms for Simultaneous Localization and Mapping. In: Yilmaz M., Niemann J., Clarke P., Messnarz R. (eds.) Systems, Software and Services Proces," *Systems, Software and Services Process Improvement. EuroSPI 2020. Communications in Computer and Information Science,* vol. 1251, pp. 433-444, 2020.

[3] H. Durrant-Whyte and T. Bailey, "Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms," *Robotics and Automation Magazine,* vol. 2, p. 1–9, 2006.

[4] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid and J. Leonard, "Past, Present, and Future of Simultaneous Localization and Mapping: Towards the Robust-Perception Age," *IEEE Transactions on Robotics,* vol. 32, no. 6, 2016.

[5] J. Zhang and S. Singh, "LOAM: Lidar Odometry and Mapping in Real-time," in *Robotics: Science and Systems Conference*, Berkeley, 2014.

[6] H. Weber, "Funktionsweise und Varianten von LiDAR-Sensoren," [Online]. Available: https://cdn.sick.com/media/docs/5/25/425/whitepaper_lidar_de_im00794 25.pdf . [Accessed Oct. 2021].

[7] I. Maksymova, C. Steger and N. Druml, "Review of LiDAR Sensor Data Acquisition and Compression for Automotive Applications," *Proceedings,* Vols. 2, 852, 2018.

[8] B. Siciliano and O. Khatib, Eds., Springer Handbook of Robotics, Springer, 2016.

[9] W. Hess, D. Kohler, H. Rapp and D. Andor, "Real-Time Loop Closure in 2D LIDAR SLAM," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[10] Ouster, "Ouster OS1: Mid-Range High-Resolution Imaging Lidar," [Online]. Available: https://ouster.com/products/os1-lidar-sensor/. [Accessed 20 Oct. 2021].

[11] Ouster, "OS1 Hardware User Manual," [Online]. Available: https://levelfivesupplies.com/wp-content/uploads/2019/03/OS-1-User-Guide-Hardware.pdf. [Accessed 20 Sept. 2021].

[12] M. Obert, "Inbetriebnahme und Evaluierung des hdl_graph_slam mit einem 128 Zeilen Ouster LiDAR -Sensor auf der Husky Roboterplattform von Clearpath," Hochschule Offenburg, Offenburg, 2021.

[13] C. Kupitz, "Inbetriebnahme und Verifizierung eines Kalman Filter zur Lagebestimmung des Clearpath Robotics Husky A200," 2021.

[14] E. Jelavic, "ETH Zürich: Programming for Robotics, Introduction to ROS," [Online]. Available: https://rsl.ethz.ch/education-students/lectures/ros.html. [Accessed Feb. 2022].

[15] Blasdel et al., "About ROS.Version: 2020," [Online]. Available: https://www.ros.org/about-ros/.

Review paper **ISSN 2603-4697 (Online) Complex Control Systems Vol. 4, No. 1, 2022, 22-25**

Construct Validity in Child-Robot Interaction Research

Maya Dimitrova(1), Snezhina Mihailova(2), Daniela Milanova(2) and Ivan Kolev(1)

(1)Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria
(1)Medical Center "Children with Developmental Problems", Sofia, Bulgaria

***Abstract*—The paper provides a review of existing approaches to understanding and defining validity in child-robot interaction research. Two main approaches are contrasted: construct validity in defining the phenomena, emerging in the process of interaction of the child with the robot, and system validity, defining the robot solely as a technological device. Examples are provided within a newly proposed approach, aiming at overcoming the existing theoretical debate in understanding construct validity in child-robot interaction research.**

***Keywords*—Child-robot interaction, construct, robot, validity.**

INTRODUCTION

The process of child–robot interaction is a focus of intensive research in the recent years. The motivation is twofold – on the one hand, the robotic technology is a radically new stimulus for the cognitive system of the child – it is attractive, capturing attention, novel, modifiable, "smart", "alien looking", etc. [1]. On the other hand, concerns are being raised stating that this novel technology may not be beneficiary to child's development [2]. Therefore, understanding the *construct* of the *communicative phenomenon,* underlying the interaction and defining the *validity* of child's responses to tasks performed with involvement of robots, is of paramount importance in research on pedagogical applications of robotic technology.

The paper describes our efforts to identify the pedagogically relevant *construct* of the *communicative phenomenon* underlying the child-robot interaction, which can be employed successfully in special education. Therefore, we are looking for relevant indicators of the validity of the proposed construct from the responses of the child in the investigated scenarios.

CONSTRUCT VS SYSTEM VALIDITY IN CHILD-ROBOT INTERACTION RESEARCH

A specific aim of the currently developed project CybSPEED (2017-2022) is providing a wider application of robotic technology than just improving the *problem solving* or *abstract thinking* abilities of children [3]. The research effort of the multidisciplinary team has been towards helping children better understand the complexity of the surrounding World, which is largely *socially-mediated.*

The complexity of the *human*-robot interaction case in general, and *child*-robot interaction in particular, falls out from the *dual nature* of the robot as technology. On the one hand, it is a complex technological device and is often tested within the *system validation* methodology [e.g. 4]. On the other hand, a large share of the attraction to the robots can be attributed to their *cognitive nature* [5]. Moreover, even if people focus on the *mechanical* nature of a humanoid robot, they acknowledge the robot's 'social presence' in the situation and expect manifestation of social behavior [6,7].

The *cognitive nature* of the robots – humanoid or abstract – provides the user/learner/child with a radically different and *novel* type of human-system interaction, transferring the theoretical approach to understanding it from *engineering,* in its essence, to *social-science based*, understanding of the observed interactive processes. The effect on the child's cognition cannot be described in *engineering* terms only, but have to be explained in *psychological* and *social science* terms. Therefore, a constellation of knowledge of probability theory, mathematical statistics and experimental psychology is necessary to be able to validate the main construct, defining the interaction of the child with the robot in pedagogical and special education scenarios.

A number of studies discuss the relations of different forms of validity tests in empirical and experimental research [8,9,10]. One particular study provides a classification, which is appropriate to illustrate some of our claims on *construct validity* testing in child-robot interaction. Figure 1 here presents an adaptation within the Creative Commons license of figure 2 from [11].

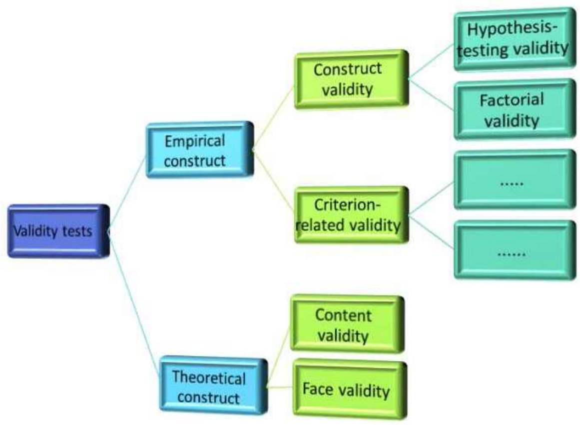


Fig. 1. Classification of validity tests in empirical research, adapted from [10] under the Creative Commons license.

DEFINITION OF CONSTRUCT

Construct is a concept, not easily formulated, especially when attempting to avoid circular definitions. For example, a

22

definition of the dictionary of the American Psychological Association (APA) is stating the following: *construct* is "an explanatory model based on empirically verifiable and measurable events or processes—an empirical construct—or on processes inferred from data of this kind but not themselves directly observable—a hypothetical construct. Many of the models used in psychology are hypothetical constructs [12]." In the context of figure 1 here these refer to the Empirical construct and the Theoretical construct, respectively. The *construct* under study is therefore a natural phenomenon that we aim to investigate/manipulate by applying a scientific approach – a learning process, a cognitive ability, an internal state or the like. The *natural* character of the internal state of the learner should not be ignored since it requires a specific type of manipulation, which is not straightforward (as any parent can confirm). This is the reason why the strictly engineering approach to testing validity of robotic technologies in education is insufficient.

What we validate in our child-robot interaction scenarios is always the underlying cognitive 'construct' and not some objectified entity like a 'social robot', for example. This is the main difference of the proposed here approach from the commonly implemented engineering approaches, labelling the types of robots with behavioral descriptions without rigorously testing the validity of these descriptions as 'constructs' in psychological terms.

It is not easy to define the construct when we try to formulate an experimental task for research. The CybSPEED main construct, defining the learning process, which we aim to modify in special education, is being formulated as: "ability for social learning … or ability for learning in social contexts… or socially-grounded learning ability (SGLA)."

DEFINITION OF CONSTRUCT VALIDITY

"Construct validity is the degree to which an instrument measures the trait or theoretical construct that it is intended to measure… It does not have a criterion for comparison rather it utilizes a hypothetical construct for comparison... It is the most valuable and most difficult measure of validity [11, p. 197]."

Testing the validity of the construct is essential in order to link meaningfully the hypothetical assumption about the mental processing of the child and some measurable indicator as a proof of the assumption. This indicator is called an *operational definition* of the measurable process. For example, an IQ test is considered an indicator of the 'intelligence' construct. The operational definition for intelligence, as measured by the IQ test, is 'success in solving a set of abstract tasks', which is measurable and formalisable, although it is a limited descriptor of the variety of human 'intelligence' aspects.

Defining an indicator for testing the validity of a construct is also a difficult task and this is where the creativity of the experimenter is often revealed. In our case this has to be linked to the number/frequency of the child's attempts to 'do something' that we want to observe, register and analyze (quantitatively/qualitatively). In the context of figure 1, we have attempted to link the theoretical and empirical constructs by testing the Face validity aspect of the game with a toy-like robot BigFoot, designed at IR-BAS [13] and the Hypothesis-testing aspect of the validity testing of the formulated construct "socially-grounded learning ability", manipulated in a special education learning context, as presented in the next section.

CONSTRUCT VALIDITY IN CHILD-ROBOT INTERACTION WITH THE WALKING ROBOT BIGFOOT

An approach to iteratively design educational scenarios in special education is proposed and tested in [14]. An interface is designed to help the child control a toy-like walking robot called BigFoot (figure 2). The child controls the robot with the keyboard arrows in a certain direction, designated by a coloured tile. The control could be performed with a joystick, too.

Fig. 2. An interface to control the movement of a walking robot towards a coloured goal.

Content validity of the SGLA construct. In the study of [14] a child with autism (ASC), who avoided any contact with other people or children, became interested in the BigFoot robot with the help of the psychologist. This was interpreted in support of the construct "socially-grounded learning ability" (SGLA), which had improved in 3 sessions. The internal mechanism was the following: the child established interest in the game through the trusted social mediator - the psychologist.

An important distinction can be made between the definitions of *validity* and *reliability* of the empirical tests of theoretical constructs [5]. Whereas a sample of data is needed for defining the *reliability* of a test, it is not the case with *validity*. In our case the construct proved *valid* for the particular child. This is why both measures are necessary in order to prove that the construct is valid in general, but not just in individual cases.

It was also shown in [14] that children enjoy collaborative play with humanoid robots such as NAO.

In [15] we extended the game towards collaborative play with the walking robot BigFoot (figure 3). Also, the interface was extended to be used with the gaze [16] (figure 4). The Ethics Committee for Scientific Research (ECSR) of IR-BAS gave permission to conduct the study with Protocol 4/10.02.2022.

The game involves 2 children – one is setting the goal and the other is controlling the robot, then they change places. Children are encouraged to help each other and teachers/parents are encouraged to help them with the game.

Face validity of the SGLA construct. Face validity is typically employed with novel experimental paradigms, such as ours. For this, experts are being interviewed or asked to fill in questionnaires in order to prove the correctness of implementing the experiment in relation to the investigated construct.

Fig. 3. Children playing in turn to control a toy-like robot.

Fig. 4. "Heat map" of focusing the gaze on the robot.

In our case we implemented a set of Likert scales. Teachers/parents were asked to assess the extent to which the game is interesting or motivating to the child, whether it helps the cognitive development, etc. [15].

The question of interest is what the main indicator for validation of the proposed construct SGLA can be. An indicator was implemented, which was first proposed in [17] and applied to a robotic scenario in [18]. It is called "self-initiated social contact" (SISC). This is the type of contact a child would direct to the other child if they ask about their intentions, feelings, etc., not just asking about daily routines or facts.

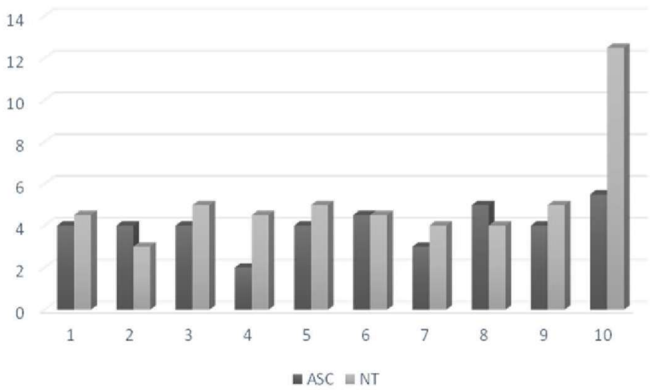


Fig. 5. Mean scores on different criteria in both groups of children. 10 – No of self-initiated social contacts; 5 – role in development of social abilities; 3 - role in development of cognitive abilities. Reproduced from [15] under the Creative Commons license.

The final question to the teachers/parents was about the number of SISC type of interactions, initiated by the individual child. The results of the comparison between 2 pairs of children – with and without symptoms of autism - are presented in [15]. Here figure 5 reproduces the chart of the results from [15] under the Creative Commons license. The clearly expressed difference in the number of SISC indicators between children of typical development (NT) and children with high-functioning autism (ASC) is evident from the figure – indicator 10 of the chart.

Another interesting observation was that children with ASC turned to the authority – parent or teacher - much more often than NT children, who preferred the company of peers. Correcting this tendency can be set as a further pedagogical goal in special education, supported by robotic devices.

CONCLUSION

The paper discusses the need to define the investigated *construct* when addressing the topic of child-robot interaction first of all, and implement a set of measures to validate it. This substantiates and confirms the plausibility of the currently designed novel educational scenarios with the walking robot BigFoot from a cyber-physical system perspective to pedagogical rehabilitation in special education.

24

Acknowledgement. We express our gratitude to assoc. prof. Ivan Chavdarov, Dr. Valentin Nikolov and assoc. prof. Aleksandar Krastev for their design of the robot and programming the interface. This research work has been partially supported by the National Science Fund of Bulgaria (scientific project "Digital Accessibility for People with Special Needs: Methodology, Conceptual Models and Innovative Ecosystems"), Grant Number KP-06-N42/4, 08.12.2020; EC for project CybSPEED, № 777720, H2020-MSCA-RISE-2017 and OP Science and Education for Smart Growth (2014-2020) for project Competence Center "Intelligent mechatronic, eco- and energy saving systems and technologies" № BG05M2OP001-1.002-0023.

REFERENCES

[1] Ramey, C.H. An inventory of reported characteristics for home computers, robots, and human beings: Applications for android science and the uncanny valley. Proc ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science, Vancouver, Canada, (2006) 21-25.

[2] Sharkey, N. The ethical frontiers of robotics, *Science*, 322, (2008) 1800-1801.

[3] https://cordis.europa.eu/project/id/777720

[4] Webster, M., Western, D., Araiza-Illan, D., et al. A corroborative approach to verification and validation of human–robot teams. The International Journal of Robotics Research. 2020, 39(1), 73-99. doi:10.1177/0278364919883338

[5] Reggia, J. A., Katz, G. E., & Davis, G. P. Humanoid cognitive robots that learn by imitating: implications for consciousness studies. Frontiers in Robotics and AI, 2018, 5, 1.

[6] Dimitrova, M. & Lekova, A. Security analysis and used acceptance of socially competent robotic systems. International Journal on Information Technologies and Security, 4(5), 2013, 37-46.

[7] Kaburlasos, V. G., & Vrochidou, E. Social robots for pedagogical rehabilitation: trends and novel modeling principles. In Cyber-physical systems for social applications (pp. 1-21). 2019, IGI global.

[8] Cronbach, L.J., & Meehl, P.E. Construct validity in psychological tests. Psychological Bulletin, 52(4), 1955, 281–302. https://doi.org/10.1037/h0040957

[9] Whitely, S. E. Construct validity: Construct representation versus nomothetic span. Psychological bulletin, 93(1), 1983, 179.

[10] Grimm, K.J., & Widaman, K.F. Construct validity. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics (pp. 621–642), 2012, American Psychological Association. https://doi.org/10.1037/13619-033

[11] Bolarinwa O.A. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. Niger Postgrad Med J, 2015, 22, 195-201. http://www.npmj.org/text.asp?2015/22/4/195/173959

[12] https://dictionary.apa.org/construct

[13] Chavdarov, I. Walking robot realized through 3d printing. Comptes rendus de l'Academie bulgare des Sciences, 69(8), 2016, 1069-1076.

[14] Dimitrova, M., Kostova, S., Lekova, A., Vrochidou, E., Chavdarov, I., Krastev, A., Botsova, R., Andreeva, A., Stancheva-Popkostadinova, V., Ozaeta, L.. Cyber-Physical Systems for Pedagogical Rehabilitation from an Inclusive Education Perspective. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 11, 2Sup1, LUMEN Publishing, 2020, 186-207, https://doi.org/10.18662/brain/11.2Sup1/104.

[15] Nikolov, V., Dimitrova, M., Chavdarov, I., Krastev, A., Wagatsuma, H.. Design of Educational Scenarios with BigFoot Walking Robot: A Cyber-physical System Perspective to Pedagogical Rehabilitation. 9th International Work-Conference on the Interplay between Natural and Artificial Computation, IWINAC, I, 13258, Springer-Verlag (LNCS), 2022, 1-11.

[16] Dimitrova, M., Sabev, N., Ozaeta, L., Nikolov, V., Krastev, A. Aspects of the Intrinsic Motivation as Accessibility Factors in the Inclusive "STEAM" Education, International Conference "Innovative STEM Education" (http://stemedu.eu/), 2022 (In Bulgarian) (in print).

[17] LeGoff, D.B. Use of LEGO© as a Therapeutic Medium for Improving Social Competence. J Autism Dev Disord 34, 2004, 557–571, https://doi.org/10.1007/s10803-004-2550-0

[18] Dimitrova, M., Vegt, N., Barakova, E. Designing a system of interactive robots for training collaborative skills to autistic children. IEEE 15th International Conference on Interactive Collaborative Learning (ICL), 2012, 1-8.

25

Combined Footprint – the Effect of Collecting of the Solder Material

Valentin Tsenev and Valentin Videkov
Technical University of Sofia

***Abstract*—In the production of electronic modules through soldering the use of PiP (Pin in paste) technology is increasingly required. With this technology bulk components are soldered using the advantages of SMT (Surface mount technology). The article examines a non-standard way to obtain successful soldering with PiP. The process of applying solder paste to obtain a combined footprint is considered and studied. When using this footprint research has been done and the results of its application have been reported. A major drawback of PiP for the use of all solder material has been solved by its optimal collection at the soldering point. The aim is to gather enough statistics to use the results of the study to create a model for machine learning.**

***Keywords*—Pin in paste, soldering, combined footprint, machine learning, statistics.**

I. Introduction

Unlike standard surface mounting there are cases where the solder paste is applied outside the contact area for soldering. A typical example is the printing technology using soldering in PiP through holes [1].

Here, an additional amount of paste is applied to provide the required amount of solder filling the hole [2]. The use of a thicker layer is not always appropriate due to the thickness limit of the metal mask. An option is to use double printing or forced application. Another possibility is to apply the paste on an area outside the contact area and then suck it into the hole [3, 4].

The standard topology of a footprint with an increased area can be with an evenly increased area relative to the contact pad - fig. 1.

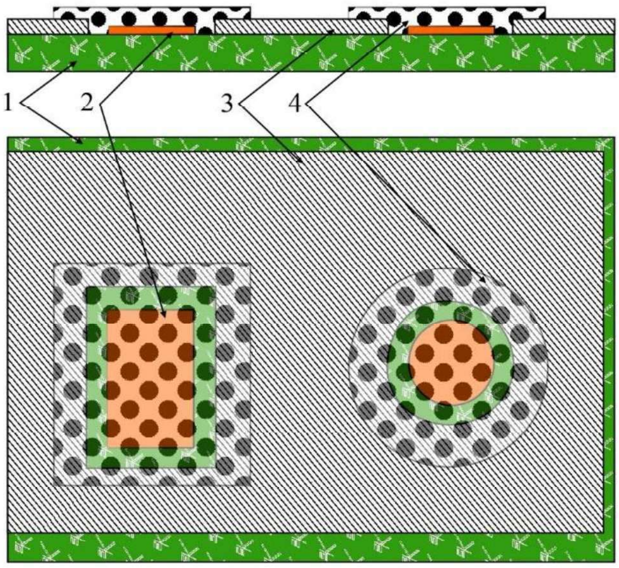


Figure 1. Overprinting footprint: 1- PCB, 2 - pad, 3 - solder mask, 4 – solder paste

In this way, the larger the area of the footprint, the greater the amount of solder material that is required for soldering the thru hole elements. This is possible because there is free space between the contact pads. Thus, electronic assembly of small and large elements can be performed. Small elements require imprints of small thickness, which is achieved with stencils of small height. Large elements are soldered with a large amount of solder material, which is possible thanks to the use of overprinting.

The second approach is asymmetric placement of the imprint in order to optimally use the free surface of the printed circuit board in accordance with the other contact pads - fig. 2.

26

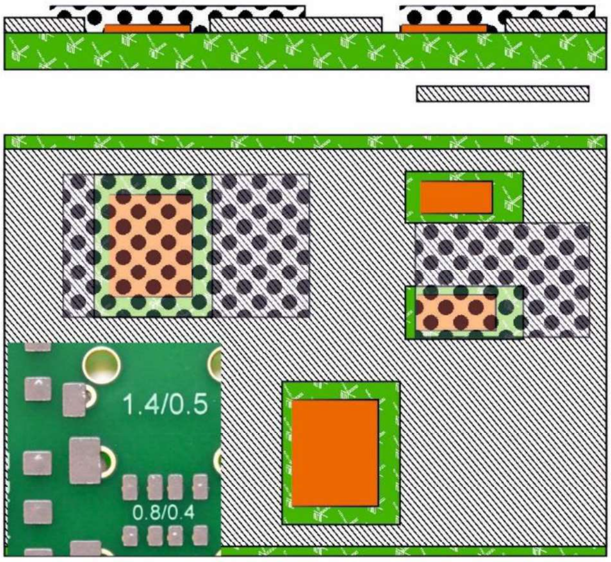


Figure 2. Overprinting footprint for PiP

II. Experiment

In the case of overprinting, a different shape of the footprint can be used depending on the topology - fig. 3.

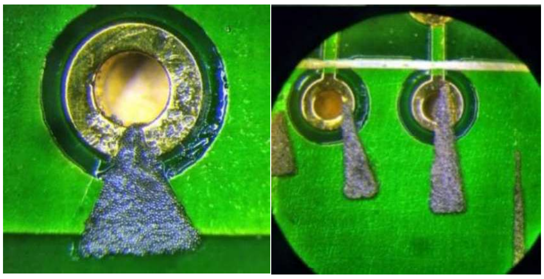


Figure 3. Non-standard overprinting

For the implementation of such footprints were designed, manufactured and used special stencils - fig. 4.

Figure 4. Stencils for nonstandard overprinting

Of particular interest is the process of melting and absorbing the paste. This process is timeless and differs from the melting process at a standard contact pad. Experiments have been conducted on the behavior of different footprint topologies in the fusion and drip mode.

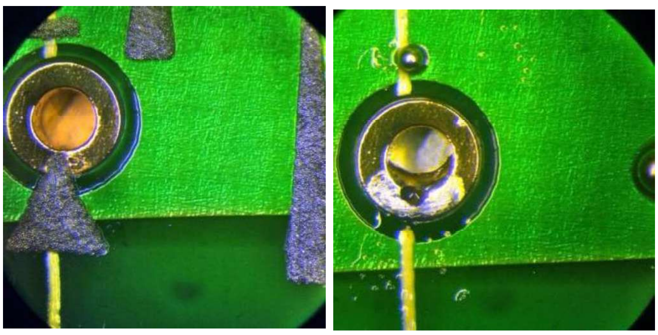
In the standard case we have melting and direct wetting by heat transfer through the contact area, the component outlet and heat transfer through the surface of the paste. In the cases under consideration, the process has two areas: a standard case area and a second area - heat supply through a printed circuit board and through the surface of the paste, which has a different heat transfer coefficient. The second difference is that the process of wetting consists in melting the paste - suction to the melt in the area of the solder - wetting. This leads to an increase in the time for complete wetting of the outlet and must be taken into account in the technological process.

Modeling the melting and suction process is possible but extremely complex. If you go down to the level of melting a single grain of solder, merging the individual grains, examining the shape of the melt and subsequent suction, it would be necessary to use a powerful calculation process, which is not justified.

A better option is to study the behavior of the process using basic forms of footprints and their subsequent application in machine learning.

The task can be divided into two:

1. Nature of melt formation in different imprint shapes as shown in Fig. 5.



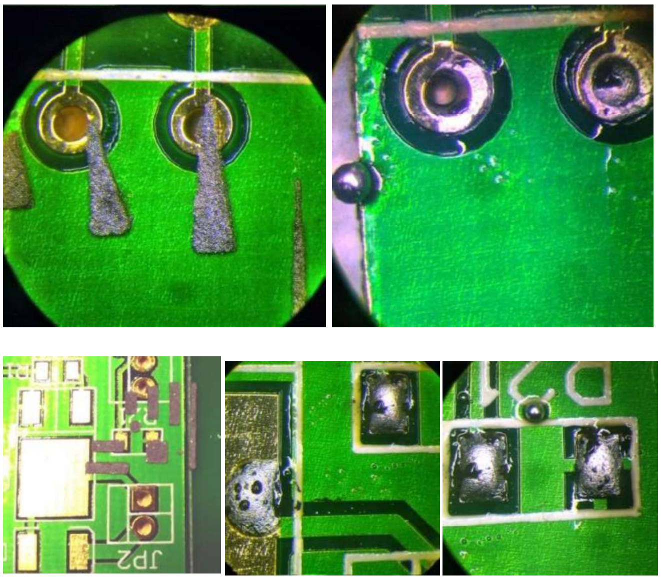
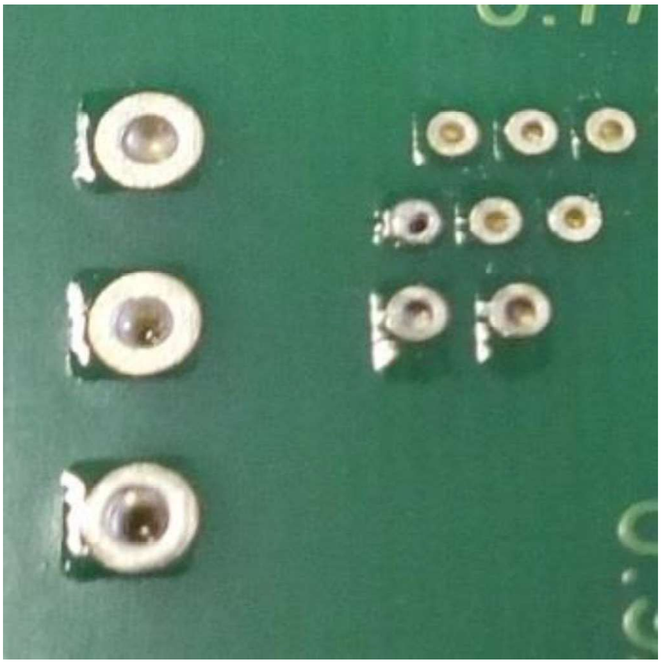


Figure 5. Solder with combined footprint and overprinting

2. Nature of suction at different contact of the imprint with the contact pad, as shown in fig. 6.



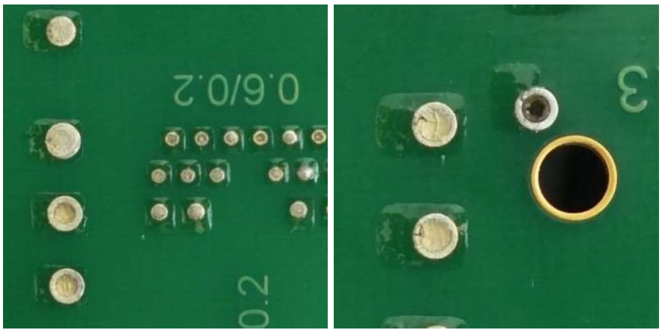


Figure 7. PCB – soldering with combined footprint and overprinting

For this purpose, conductive melting of various prints in shape and overlapping with a contact pad was performed. Conductive soldering was used to avoid the effect of fluid flow on droplet formation.

The temperature profile is controlled only until the moment of formation of the drop - wetting. Pastes with bismuth content are used for lower melting temperature.

The temperature profile used is shown in fig. 8.

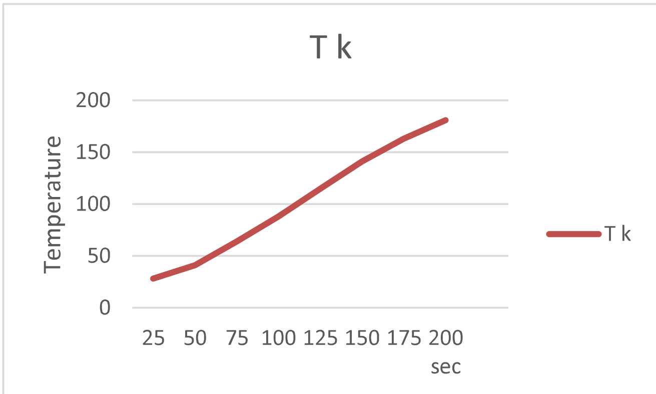


Figure 8. Temperature profile

III. Results

One of the obtained results is shown in fig. 9.

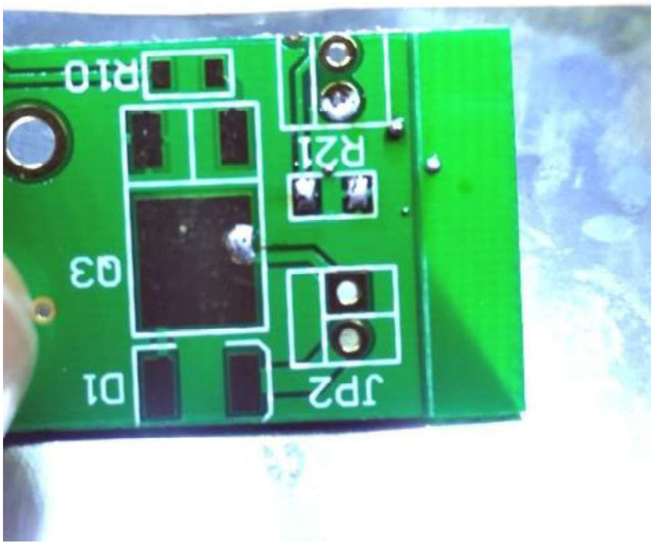


Figure 9. Collecting of soldering material

The obtained result shows excellent collection of the soldering material in different cases of combined footprint - only on solder mask, only on pad, on pad and solder mask.

The condition for good collection of the soldering material is that the combined footprint is well designed and dimensioned.

An even better result of collecting the solder material is obtained by making a combined footprint with two different pastes, as shown in fig. 10.

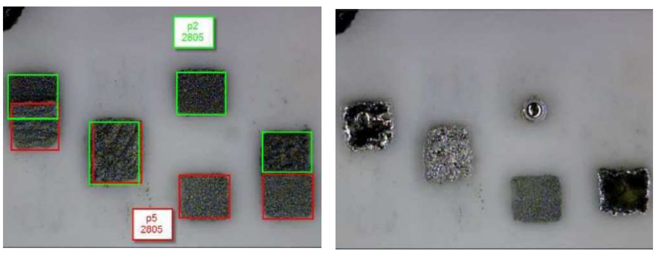


Figure 10. Combined footprint with two soldering pastes

Thanks to the ineffective soldering of the soldering materials (standard and low temperature) excellent collection of the melt in the place of the footprint is obtained.

IV. Conclusions

The obtained results are the basis for a more in-depth study of soldering using a combined footprint.

By collecting sufficient statistics on the results of soldering with different combined footprints, a machine learning model can be created. Thus, current standard printer machines can be easily reprogrammed to produce a combined footprint.

Soldering using a combined fingerprint will allow for:

- Stable and reusable PiP soldering
- No contamination with solder balls, which reduces the reliability of electronic products
- Avoid cleaning the electronic modules after soldering
- Creating strong solders
- Improved thermal conductivity of solders
- Low solder resistance
- Low cost due to the short production cycle

Acknowledgement

This research is supported by Bulgarian National Science Fund in the scope of the project "Exploration the application of statistics and machine learning in electronics" under contract number КП-06-Н42/1.

References

[1] Bob Willis, "PIHR technology", 2012.

[2] Oliver Krammer, New Method for Calculating the Necessary Amount of Solder Paste for Pin-in-paste Technology, 19th International Symposium for Design and Technology in Electronic Packaging (SIITME), 2013.

[3] Valentin Tsenev; Valentin Videkov; Aleksei Stratev; Mladen Mitov; Ana Bankova; Vasil Atanasov Influence of geometry on the applied solder paste in technology Pin in Paste(PiP) 2016 39th International Spring Seminar on Electronics Technology (ISSE) Year: 2016 Pages: 160 - 162, DOI: 10.1109/ISSE.2016.7563180.

[4] Valentin Tsenev ; Valentin Videkov, Experimental study of the melting of a solder paste, 2018 IEEE XXVII International Scientific Conference Electronics – ET, 13-15 Sept. 2018, DOI: 1109/ET.2018.8549606,

[5] R. B. Lima, V. F. Cardosoa and A. N. R. da Silva, Analysis of Pin in Paste Soldering on Printed Circuits Boards Assembly Using Lead Free Solder Paste, ECS Transactions, 39 (1) 417-422, 2011, DOI: 10.1149/1.3615221.

[6] Tony Lentz, Jasbir Bath, N. Pavithiran, How Does Printed Solder Paste Volume Affect Solder Joint Realiability, SMTAI, 2018.

Research paper ISSN 2603-4697 (Online) Complex Control Systems Vol. 4, No. 1, 2022, 30-35

Manufacturing Process Optimization Through Machine Learning and Analytical Prediction

Malinka Ivanova, Valentin Tsenev and Borislav Dimitrov
Technical University of Sofia

***Abstract*—A "smart" production is characterized with collection of a large amount of data and the application of machine and deep learning algorithms for the purposes of analytical prediction. The analysis supports the implementation of intelligent management and rapid response to changes in a manufacturing process. The paper proposes an approach for optimizing a robotic manufacturing line for electronic components through applying the failure mode and effect analysis and algorithm for deep learning. This approach is embedded in a software tool created through C#, Windows Forms technology and open source to assist identification of the potential risks by the responsible engineer.**

***Keywords*—big data, deep learning, FMEA, machine learning, optimization, statistics, "smart" manufacturing, robotic line.**

I. Introduction

Nowadays, the manufacturing process is considered in its dynamics, taking into account the influence of the factor time. With the time, various changes of some other factors can occur and this affect on the product quality.

It is known that the robotic manufacturing lines of electronic components collect a large amount of data, which makes it possible to monitor changes and take appropriate actions before failure.

The concept of "smart" manufacturing is discussed in [1] and it includes the use of new technologies, consideration of many parameters and a number of factors. An important part of a "smart" manufacturing is related to the techniques for processing and analysis of the collected data, including the application of statistical methods and algorithms for machine learning. Another paper points out the importance of large data sets for organizing "smart" manufacturing and using the advantages of predictive analysis for early detection of errors and shortcomings [2].

The most commonly used algorithms for machine learning in manufacturing processes and in the context of Industry 4.0 are discussed in [3]. Here they are summarized graphically and are presented through Figure 1. Also are shown the most commonly used algorithms for deep learning, which is summarized on Figure 2.

Deep learning is a part of machine learning and artificial intelligence, as the process of learning from data sets is based on artificial neural networks (ANNs) usage. The architecture of ANNs includes a number of layers depending on the specific task and on the requirement for finding a solution with high accuracy. Although artificial intelligence systems are currently part of "smart" manufacturing and certain techniques for data analysis are used, the researchers are still looking for solutions to improve and optimize manufacturing processes based on deep learning and analytical prediction.

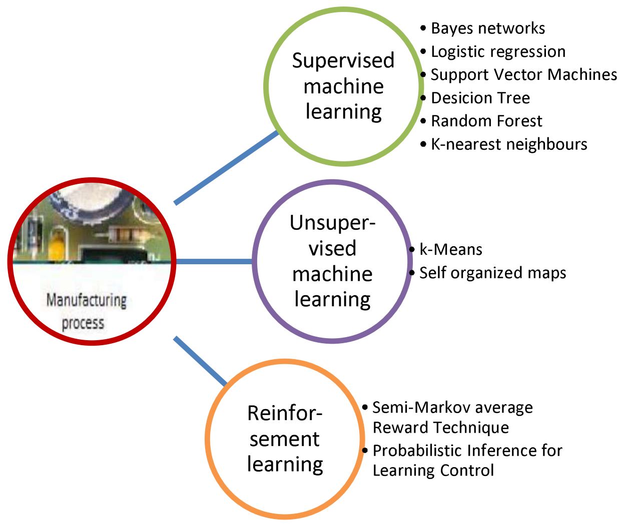


Figure 1. The most used algorithms for machine learning in a manufacturing process [3]

30

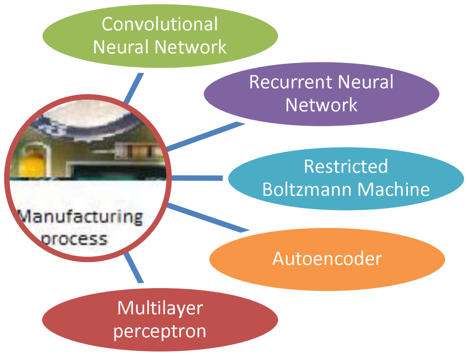


Figure 2. Deep Learning and its application in manufacturing process [3]

Statistical methods are also widely used for risk evaluation, control and optimization, as the most commonly used are: Gaussian process [4], [5], Failure Mode and Effect Analysis (FMEA) [6], [7], analysis of a tree with errors [8], repeatability and reproducibility [9].

The aim of the paper is to present an approach for optimizing a robotic manufacturing process for electronic components by comparing the FMEA and deep learning. The approach is implemented in a developed software tool to support timely informing the engineer in charge about the potential risks as well as to assist his decision-making process.

II. Research Methodology

The research methodology includes the following procedures: (1) A study of the FMEA and its applications for optimizing the manufacturing process of electronic components through the robot FANUC-M-10 is performed, which is documented through tables with information from experts. The activity *Component mounting* is discussed in more details as a demonstration example. Based on the collected information, data sets are formed, which are prepared for further processing by an algorithm for deep learning. Datasets are formatted according to the requirements of the .csv file. (2) A feed-forward artificial neural network with back-propagation is created, searching for the optimal neural network architecture according to the number of layers and neurons in each layer, as well as concerning the type of the activation function. Carrying out experiments in the environment of RapidMiner Studio on the data sets collected from the previous procedure as a deep learning is applied at the ratio of training and testing data: 70%/30%. Comparison between the FMEA results with deep learning is conducted. (3) Design and development of a software tool using C# programming language, Windows Forms technology and open source solutions for comparison of FMEA and analytical prediction is performed.

III. The Failure Mode and Effect Analysis

The FMEA is chosen as a statistical technique for performing risk evaluation in manufacturing using a FANUC-M-10 robot for placing and soldering electronic components on a printed circuit board. Failure regime means the situation in which a problem may occur. The problem is associated with errors, defects or outright failures. The analysis of the effects refers to the study of the consequences of these potential problems. Problems are prioritized according to how serious their consequences are, how often they occur, and how easily they can be identified. The purpose of the FMEA is some actions to be taken to eliminate or reduce failures, starting with those with the highest priority. FMEA is applied in order to prevent possible errors, defects and failures that may occur during the manufacturing. Thus, pre-identified problems at the earliest possible stage of the production process can save materials, resources and time. The advantages of FMEA are as follows: possibility for summarizing collective knowledge gained by experts; timely identification of risky manufacturing activities; reduction of the production cycle time; documenting possible risks of obtaining substandard or unusable products [10]. These benefits underlie the widespread use of FMEA to identify critical activities and priority risk.

This paper presents risk evaluation of manufacturing of electronic components, which includes activities for programming, starting, testing and stopping the robot, as well as the following main activities: component placing, soldering and external connections creation. All manufacturing activities are documented and evaluated according to three criteria: (1) *Severity* (S), indicating the influence of the failure effects on the subsequent manufacturing activities; (2) *Occurrence* (O), showing the probability of failure; (3) Detection (D), indicating how a particular control measure may contribute to the detection of a failure. Finally, the Risk Priority Number (RPN) is calculated as the product of above evaluated and described three criteria. RPN can take values from 0 to 100. If the value of the priority risk is higher, the greater is the probability of defect, error or failure occurrence. Then, activities are recommended, and after their implementation, the priority risk is recalculated. The resulting value of the risk is expected to be significantly lower, otherwise it is considered that the recommended activity has not led to reduction of the risk.

Tables document all activities related to this manufacturing process. Here is shown only one table (Table 1) with an evaluation of the manufacturing activity *Component mounting*. It consists of four potential failures: smaller diameter of the holes for the component mounting, larger diameter of the holes, no holes, no soldering square. Recommended activities to reduce the risk of potential problems are also shown. It can be seen that in this manufacturing activity the risk of failure is small and is minimized after the application of the recommended measures.

Data related to the rest of the main production activities (soldering and bonding) show that the priority risk may exceed 50 (from maximum 100 points). After the implementation of recommended activities, this risk can be significantly reduced.

Table 1. FMEA of manufacturing activity *Component mounting*

	Component mounting			
Activity	The component cannot be placed			
Potential failure	Smaller diameter of the holes	Greater diameter of the holes	Missing holes	Missing square for soldering
Potential effect	Another soldering cycle	Longer soldering time	Another soldering cycle	Another soldering cycle
S	2	4	3	5
Potential cause	Mistake of the programmer			
O	3	3	2	2
Current controls	Visual inspection			
D	2	2	2	3
RPN	12	24	12	30
Recommended activities	Preview of the printed circuit board			
	Proper selection of components before soldering			
	Corrections of the programming code			
Person in charge and date	Engineer production			
	Engineer production			
	IT expert			
Taken activities	Preview of the printed circuit board			
	Proper selection of the components			
	Programmer' training			
S	1	1	1	1
O	2	2	2	2
D	1	2	1	1
RPN	2	4	2	2

IV. DEEP LEARNING

Deep learning can be realized by applying a wide variety of algorithms and using ANNs with different architectures [11], [12].

In this work, a feed-forward artificial neural network with back-propagation is constructed, as shown in Figure 3. The ANN has 4 inputs: S, O, D and RPN and one output: the priority risk, which values can be: very low - at PRN = 1 ÷ 20, low - at PRN = 21 ÷ 40, medium - at PRN = 41 ÷ 60, high - at PRN = 61 ÷ 80) and very high - at PRN = 81 ÷ 100. The inputs are denoted with x_i and the output is y. Each neuron sums the input signals with a certain weight w_i and the result together with the deviation b is forward to the activating function AF, which can be linear or nonlinear [13]. Depending on this, linear or nonlinear regression or classification is performed. Our task here is a classification task to be solved. Then, the output y has the form: $y = AF\,(x_1w_1 + x_2w_2 + \ldots + x_nw_n + b)$.

The optimal architecture of an ANN is created after experimenting with the number of neurons and hidden layers with three different activating functions: tanh, rectifier and maxout (Table 2).

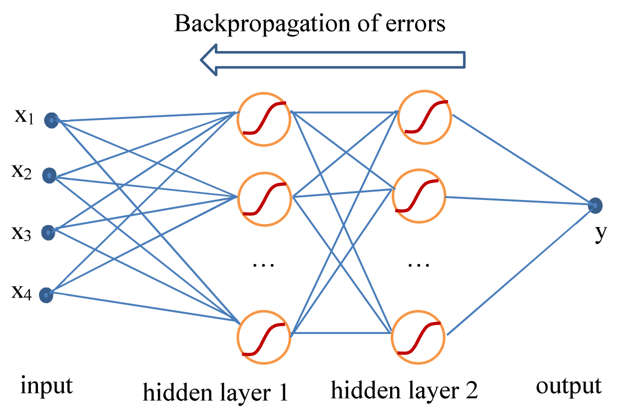


Figure 3. Feed-forward artificial neural network with back-propagation

Table 3 shows that with the highest accuracy of 96.81% is the ANN with 2 hidden layers, with 20 neurons in the first layer and 60 neurons in the second layer. The most suitable activating function for solving this classification problem is the hyperbolic tangent tanh.

Table 2. Activating functions

Hyperbolic tangent tanh	Rectified linear unit – ReLU	maxout
$\tanh(x) = (e^{x}-e^{-x})/(e^{x}+e^{-x})$	$ReLU = \begin{cases} 0, & x<=0 \\ x, & x>0 \end{cases}$	$maxout = \max(w_1^T x + b_1 + w_2^T x + b_2)$
Range in (-1,1)	Range in [0,∞)	Range in (-∞,+∞)

Table 3. ANN accuracy

	Accuracy		
Hidden layers/neurons	tanh	rectifier	maxout
2 layers /50/50	92.02%	85.11%	89.89%
2 layers /20/80	93.62%	91.49%	90.96%
2 layers /20/50	94.68%	83.51%	91.49%
2 layers /20/60	96.81%	82.45%	81.91%
2 layers /20/20	93.03%	87.77%	89.36%

V. Instrument for Comparison the Results from FMEA and Deep Learning

Comparison of FMEA results and deep learning is a possible approach for obtaining an objective evaluation and predictive analysis of critical activities in the manufacturing process of electronic components using the robot FANUC-M-10. This will support the responsible engineers to be timely informed about possible potential risks that can be avoided if appropriate measures are taken into account. Creating a software tool, implementing the proposed approach, could facilitate decision making and could prevent occurrence of critical problems.

The developed software tool incorporates the algorithm presented on Figure 4. Data from FMEA and deep learning can be entered manually via a form or via a pre-prepared .csv file. Numeric values for Severity - S, Occurrence - O, Detection - D, Risk Priority Number (RPNFMEA) are calculated after FMEA analysis and Risk Priority Number (RPNDL) is predicted by the neural network. Also, the evaluation is performed, as the risk is classified into five groups: very low risk, low risk, medium risk, high risk, very high risk.

The constructed software tool compares the results of FMEA and deep learning using two parameters RiskFMEA and RiskDL, and the comparison is presented in two ways: tabular and graphical. In the tabular and graphical representation, in addition to the parameters RiskFMEA and RiskDL, FMEAPoints and DLPoints are also shown, which provide statistical information on how many times the same risk priority number is obtained.

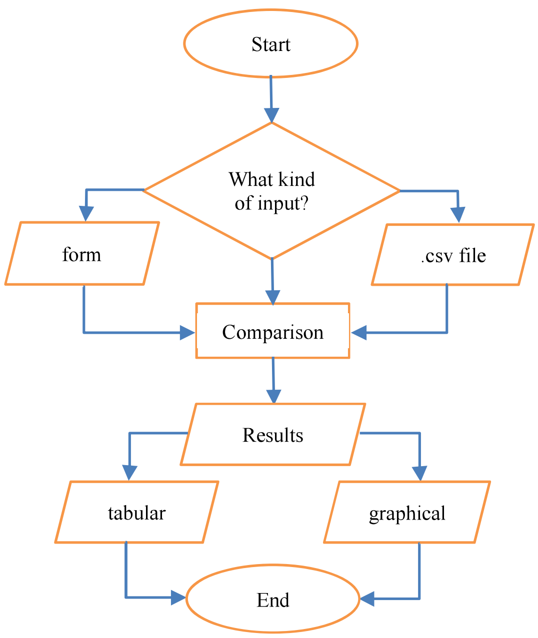


Figure 4. Algorithm behind the created software tool

Figures 5, 6 and 7 present screens from the implementation of the software tool with data taken from the demonstration example for the FMEA analysis concerning the manufacturing activity *Component mounting* and the four identified potential failures. From the obtained data, it can be seen that this activity does not carry a high potential risk of damage, defects or failures. However, there is a risk, which although minimal, must be taken into consideration. Using the *Add record* button, it is possible to enter data through the fields in the form, by the responsible engineer. Using the *Import .csv* button the engineer can read the data from a pre-prepared .csv file. The parameters Severity, Occurrence, Detection, RPNFMEA and RPNDL have a numerical expression, and the risk evaluation of FMEA RiskFMEA and the obtained prediction from deep learning RiskDL can be selected from five possible values. When the *Compare* button is clicked, a comparison is done between the results of the FMEA analysis and deep learning. The result of comparison is presented through tables and charts, showing the priority risk in the form of an evaluation in five groups: very low, low, medium, high and very high risk. According to the above mentioned demonstration related to the manufacturing activity *Component mounting*, the parameter RiskFMEA has values very low and low, repeated twice. The parameter RiskDL is characterized with very low values, repeated 3 times and low value predicted once.

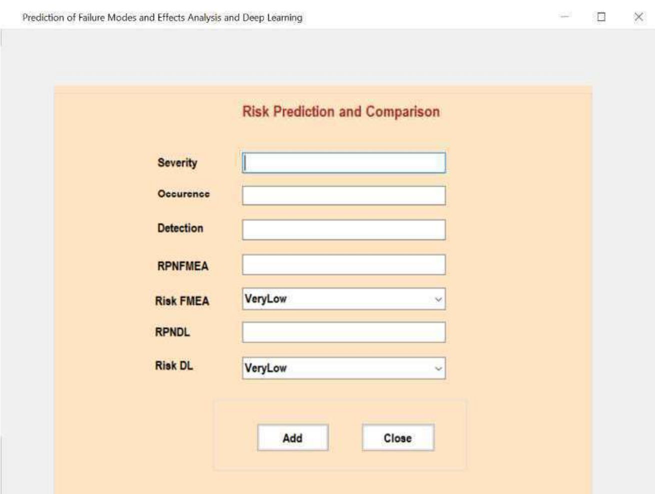


Figure 5. The form for data input

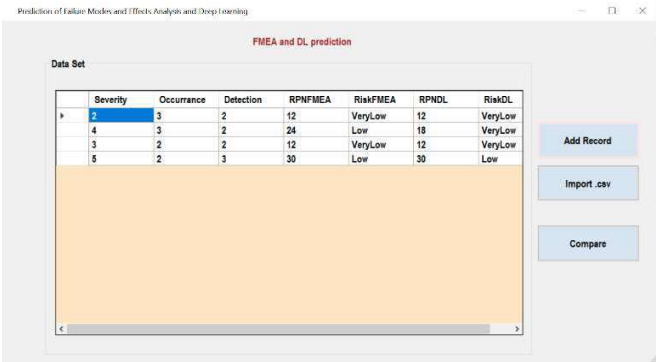


Figure 6. Entered data with possibility for comparison

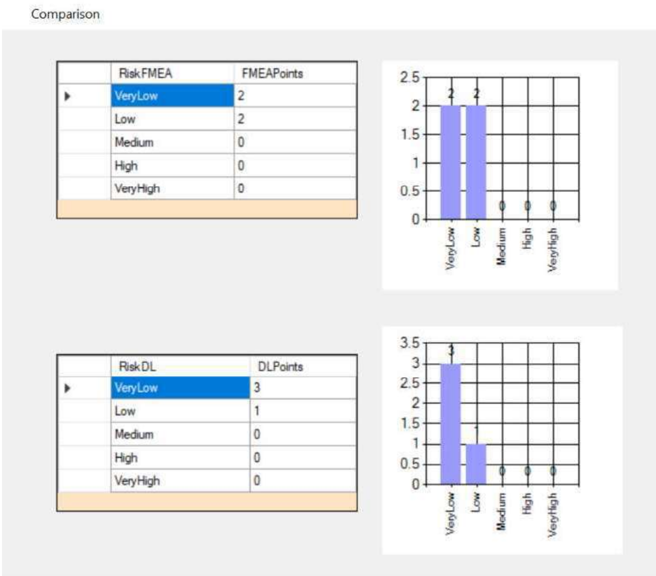


Figure 7. The results from comparison

VI. CONCLUSION

The FMEA is widely applied approach for evaluation of a manufacturing process and for identifyng potential failures. When the FMEA is combined with a deep learning algorithm it gives the opportunity for analytical prediction based on collected data during a dynamic manufacturing process. It is proved in this work that could be achieved very high accuracy of the predictive models and it depends on the constructed architecture of the ANN. Deep learning algorithm from supervised learning is tuned through usage of three different activation functions tanh, rectifier and maxout as the best accuracy 96.81% is obtained at ANN with 2 layers as the first one includes 20 neurons and the second one 50 neurons. The most suitable activation function for this classification task is tanh. Also, the developed software tool is presented, which compares the results from the FMEA and deep learning. There are huge possibilities for the FMEA to work in combination with deep learning algorithms in support of the responsible engineer, who can receive predicted analytics about potential critical manufacturing activities. This allows preventive actions to be taken and decisions to be made very quickly.

ACKNOWLEDGEMENT

This research is supported by Bulgarian National Science Fund in the scope of the project "Exploration the application of statistics and machine learning in electronics" under contract number КП-06-Н42/1.

REFERENCES

[1] Mittal, S., Khan, M. A., Romero, D., and Wuest, T., Smart manufacturing: Characteristics, technologies and enabling factors, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 233(5), 2019, 1342–1361, https://doi.org/10.1177/0954405417736547.

[2] Moyne, J., Iskandar, J., Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing, *Processes*, 5(3), 39, 2017, https://doi.org/10.3390/pr5030039.

[3] Kotsiopoulos, T., Sarigiannidis, P., Ioannidis, D., Tzovaras, D., Machine Learning and Deep Learning in smart manufacturing: The Smart Grid paradigm, *Computer Science Review*, vol, 40, May 2021, 100341, https://doi.org/10.1016/j.cosrev.2020.100341.

[4] Civera, M., Boscato, G., Fragonara, L. Z., Treed Gaussian Process for Manufacturing Imperfection Identification of Pultruded Gfrp Thin-Walled Profile, *Composite Structures*, vol. 254, 15 December 2020, https://doi.org/10.1016/j.compstruct.2020.112882.

[5] Hensel, S., Marinov, M. B., Schwilk, T., Nikolov, D., Application of Gaussian Process Estimation for Magnetic Field Mapping, In: Perakovic, D., Knapcikova, L. (Eds,), Future Access Enablers for Ubiquitous and Intelligent Infrastructures, FABULOUS 2021. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 382, 2021, Springer, Cham, https://doi.org/10.1007/978-3-030-78459-1_21.

[6] Baynal, K., Sarı, T. Akpınar, B, Risk management in automotive manufacturing process based on FMEA and grey relational analysis: A case study, *Advances in Production Engineering & Management*, 13(1), 2018, 69–80, https://doi.org/10.14743/apem2018.1.274.

[7] Wu, Z., Liu, W. and Nie, W., Literature review and prospect of the development and application of FMEA in manufacturing industry, *The International Journal of Advanced Manufacturing Technologies* 112, 2021, 1409–1436. https://doi.org/10.1007/s00170-020-06425-0.

[8] Yahmadi, R., Brik, K. and Ammar, F. ben, Causal tree analysis for quality control of the lead acid battery manufacturing process, *International Journal of Energy Research*, 42(4), 2018, 1738-1759, https://doi.org/10.1002/er.3987.

[9] Runje, B., Novak, A. H. and Razumić, A., Measurement system analysis in production process, *XVII International Scientific Conference on Industrial Systems (IS'17)*, Novi Sad, Serbia, 4–6 October 2017, 274-277, http://www.iim.ftn.uns.ac.rs/is17.

[10] Liew, C. F., Prakash, J., Kamaruddin, S. and Ong, K. S. Deployment of an FMEA-Integrated Framework to Improve Operational Performance in Semiconductor Manufacturing: A Case Study, IOP Conf. Series: *Materials Science and Engineering* 530(1), 2019, doi: 10.1088/1757-899X/530/1/012040.

[11] Smys, S., Chen, J. I. Z., Shakya, S., Survey on Neural Network Architectures with Deep Learning, *Journal of Soft Computing Paradigm* (JSCP), 2(3), 2020, 186-194, doi: https://doi.org/10.36548/jscp.2020.3.007.

[12] Singaravel, S., Suykens, J. and Geyer, P., Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction, *Advanced Engineering Informatics*, vol. 38, 2018, 81-90, https://doi.org/10.1016/j.aei.2018.06.004.

[13] Chigozie, N., Ijomah, W. L., Gachagan, A. and Marshall, S., Activation Functions: Comparison of trends in Practice and Research for Deep Learning, 2018, https://doi.org/10.48550/arXiv.1811.03378.

35

Research paper ISSN 2603-4697 (Online) Complex Control Systems Vol. 4, No. 1, 2022, 36-40

Interactive Interface Module for Cerebral Palsy Rehabilitation: Study on the Performance through Machine Learning

Zhenli Lu(1), Roumiana Ilieva(2), and Malinka Ivanova(3)

(1) Changshu Institute of Technology, Changshu, China
(2) Technical University of Sofia, Sofia, Bulgaria

***Abstract*—The paper presents the developed interactive interface module assisting the rehabilitation therapy of children with cerebral palsy. It is part of an intelligent training system and includes a brain-computer interface. The performance of the interactive module is evaluated through the methods typical for explainable artificial intelligence as in this work the Random Forest algorithm is applied. The findings show the better performance at smile recognition in comparison to recognition of the blink and muscle nose exercises.**

***Keywords*—brain-computer interface, cerebral palsy rehabilitation, interactive interface module, machine learning, intelligent training system**

I. Introduction

Nowadays, machine learning is used in a wide variety of areas in the case when a large amount of data is collected and there is a need for their exploration and better understanding. Such a way gives possibilities for extraction of patterns, tendencies or anomalies and very often this knowledge is used for making predictions and analysis [1].

Recently the topic about explainable artificial intelligence is discussed as a scientific field which looks for new methods with capability to interpret and explain different machine learning models. Linardatos et al. study these methods and propose a taxonomy which points out the purpose of interpretable methods and the direction for achieving this purpose [2]. It proves that machine learning is also applied for receiving good interpretability and explanation of created data models as a factor for evaluation their usefulness, sensitivity and scope. A very detailed investigation about the applications of explainable artificial intelligence is proposed in [3], where is outlined that 60% of the published papers are in the healthcare domain for mainly solving classification tasks (88.89%) in intelligent systems (50%) taking advantages of Artificial Neural Networks (81.25%), Support Vector Machines (12.5%), Bayesian networks (6.25%). This analysis reveals suitability of machine learning for explanation of a wide variety of healthcare problems, including for studying and evaluating the brain-computer interfaces [4], healthcare robots [5], neuromotor rehabilitation intelligent systems [6].

Cerebral palsy is related to disturbances which appear in early human biological growth or in the early child ages that could be in different forms and severity. The most discussed form concerns disorder in motor functions of children, reflecting on their movement and spatial orientation. Some research reports that cerebral palsy influences social behavior and communication, as well as cognition, sensation and emotions [7]. The children diagnosed with this disease do not possess correct spatial orientation, body parts perceiving, left and right directions determination.

The rehabilitation therapy must begin as early as possible and it could be very uncomfortable, fatiguing and even painful for children with cerebral palsy.

So, the most important arising question is how to motivate and stimulate children to go through such long, boring and difficult therapy. It is related not only to the children, but also to the therapists and parents. The practice shows that more and more researchers discover the benefits of robotic intelligent systems for establishment of social contact and information exchange through brain-computer interface with children. But it is not enough, there is a need for a human-computer interface for triggering the social interactions and for facilitating the connection between children and therapists. The evaluation of the performance of such interactive interface and workability of intelligent training system is a valuable task that can draw their usefulness, correctness and accuracy.

The aim of the paper is to present the designed interactive interface module for robotized training system, which is used in rehabilitation of children with cerebral palsy and to propose an approach for evaluation of its performance through explainable artificial intelligence methods and in particular through machine learning.

II. Machine Learning in Hardware Systems for Cerebral Palsy Rehabilitation

This section reviews and summarizes the application of machine learning techniques for explanation and interpretation of the performance of hardware based approaches for rehabilitation of children with cerebral palsy.

Ahmadi et al. use motion sensors to detect and assess the physical activities of the children with cerebral palsy as the machine learning algorithms SVM and Random Forest are utilized to classify the recognized activities in different types [8]. The created classification models for group, personalized-group and child-personalized activity are compared and evaluated and the results show that more accurate are models for identification of personalized-group and child-personalized activity.

In another work, Bedla et al. present a method for evaluation the Gross Motor Function Measure, taking into account the collected data from Zebris FMD-T device [9]. The authors conclude that the more complex exercises conducted by children lead to bigger estimation errors.

A methodology, which includes eye images and machine learning is proposed and applied by Illavarason et al. [10] for automatic diagnosis and further rehabilitation of children with mild motor difficulties. The models are characterized with high accuracy through usage of neural networks learner.

Martin et al. describe a solution for rehabilitation therapy by applying a social robotic system with possibility for children profiling and realization of rich human-computer interaction [11]. It includes machine learning for recognition of the user' pose in support of correct diagnose determination.

Usama et al. discuss the evaluation of the performance of brain-computer interface through applying neural networks for solving classification task [12]. The recognition rate of the patients' movement is high.

Faria et al. is talking about the assessment of brain-computer interface, which is used for facial emotion recognition and communication with an intelligent wheelchair [13]. Some algorithms of machine and deep learning are utilized for classification models creation: one for picking out the emotions on the face and another for thoughts determination. The achieved accuracy at emotions recognition is higher than the accuracy at thought identification.

The examined research shows the possibilities of machine learning to evaluate with high accuracy the hardware-based systems and their performance despite the hardware type (motion sensors, social robots, human-computer interface or brain-computer interface) used for rehabilitation of children with cerebral palsy. Such approach is utilized and applied in this work for studying and assessing the interactive interface module at robot-assisted therapy.

III. Interactive Interface Module

As before it has been discussed the rehabilitation therapy of children with cerebral palsy could be boring and hard for following. This is the reason for looking a motivational, interesting and interactive interface to drive the child to go through the whole rehabilitation process. Thus, the children in their early age could be attracted through interesting cartoon pictures that could improve their enthusiasm for training.

So, in order to reduce the workload of the rehabilitation therapist and for better guidance of children's training, an interactive interface is designed and presented here. The software is created in the Python environment, because Python provides several libraries for graphical interface development, including Tkinter, wxpthon and Jython. Tkinter does not need to be installed separately during use, and can run on Windows, UNIX and other operating systems.

37

The structure of the interface module is presented on Figure 1 and it consists of a main panel and some panels with cartoon pictures (smile, blink and muscle nose). The action switching button is set in the interface in help of the rehabilitation therapist, who has the possibility to adjust the training sequence of actions according to the actual situation during the training of the children.

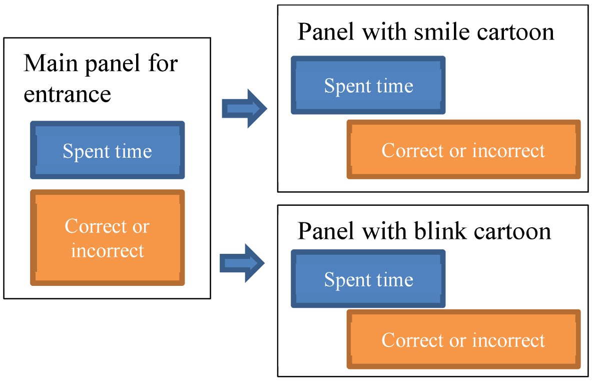


Figure 1. The structure of the software interface

Clicking on the switch button, the panel will switch to the one of the panels, where the action indication picture will change to smile or blink and muscle nose. In addition, the interface also contains the start and end buttons, and the time and action discrimination display function. Clicking the start button will trigger the start of the training and also the time of the training process is counting. The time is displayed in the time dialog box. Clicking on the end button will register the end of the training and the content of the dialog box will be cleared.

For experimentation and software system debugging a brain-computer interface is used in the form of 14 sensors. The workability of the sensors and the quality of the received signal is tested via the Emotiv software for 3D brain visualization. If the sensor has good contact, then the indicator of the light is green as it is shown on Figure 2.

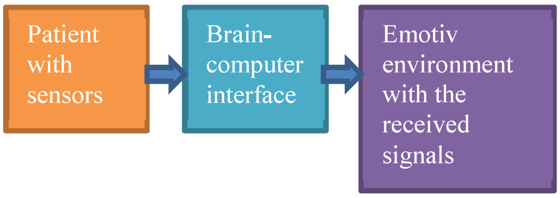


Figure 2. The signal quality test in the Emotiv environment

The interactive module is developed according to the architecture client/server. After the handshaking between server and client, the signal stream brain-computer is recorded in .txt files. Then, also the human-computer interface is tested for its normal workability. The overall test in the case of blink and muscle nose exercise is performed in laboratory settings.

The gathered data are in the form presented in Table 1. It includes the result of the emotion recognition on the child's face (smile or blink and muscle nose). Similar tables are created for all tested users.

Table 1. The result of facial emotion recognition through brain-computer interface for subject A

Time	Smile	Blink
1	Correct	Incorrect
2	Incorrect	Correct
3	Incorrect	Correct
4	Correct	Incorrect
5	Correct	Correct
6	Correct	Correct
7	Correct	Correct
8	Correct	Correct
9	Correct	Correct
10	Incorrect	Correct

IV. Results Evaluation

For evaluation the functionality and the performance of the developed interactive software module the gathered data tables are used. The method of explainable artificial intelligence is also applied. Two classification models are created: the first one evaluates the performance of the software interactive module for recognition of the smile emotion and the second is related to the blink and muscle nose

38

recognition. This is a two class classification task as each record is labeled to be from class correct (recognition) and incorrect (recognition). The utilized classifier is the machine learning algorithm Random Forest and it is chosen, because of its possibility to deal with different data sets, its flexibility and high accuracy.

The results show that the model for smile recognition is characterized with accuracy 83.33% and the model for blink and muscle nose recognition with accuracy of 66.67%. The smile on the children' face is better recognizable in comparison to the blink The results show that the model for smile recognition is characterized with accuracy 83.33% and the model for blink and muscle nose recognition with accuracy of 66.67%. The smile on the children' face is better recognizable in comparison to the blink and muscle nose exercise. It could be explained with the quality of the transmitted signal through the brain-computer interface as well as with the children' mistakes during the therapy.

It could be explained with the quality of the transmitted signal through the brain-computer interface as well as with the children' mistakes during the therapy.

V. Conclusion

In this paper, a brain-computer interface and developed interactive software module are presented as well as the test results, which are conducted in the environment Emotiv.

The rehabilitation training system completes the signal reception and processing in the Python programming environment, and realizes the communication.

Through the combination of imagination and movement, we can realize the training of attention and facial muscles in children with cerebral palsy.

Through interesting cartoon pictures and interactive interface, the therapy is assisted to improve the fun of the training process, which can provide new technical support for the rehabilitation training of cerebral palsy.

The performance of the interactive software module is evaluated taking into account the methods of explainable artificial intelligence. The findings point out better recognition of the smile exercise in comparison to the blink and muscle nose one.

Acknowledgement

This research is supported by Bulgarian National Science Fund in the scope of the project "Exploration the application of statistics and machine learning in electronics" under contract number КП-06-Н42/1.

References

[1] Sarker, I. H., Machine Learning: Algorithms, Real-World Applications and Research Directions, *SN Computer Science* 2, 160, 2021, https://doi.org/10.1007/s42979-021-00592-x.

[2] Linardatos, P., Papastefanopoulos. V., Kotsiantis, S., Explainable AI: A Review of Machine Learning Interpretability Methods, *Entropy*, 23(1):18, 2021, https://doi.org/10.3390/e23010018.

[3] Vilone, G. and Longo, L., Explainable Artificial Intelligence: a Systematic Review, 2020, https://arxiv.org/pdf/2006.00093.pdf.

[4] Rasheed, S., A Review of the Role of Machine Learning Techniques towards Brain–Computer Interface Applications, *Machine Learning and Knowledge Extraction*, 3(4), 835-862, 2021, https://doi.org/10.3390/make3040042.

[5] Pee, L. G., Pan, S. L., Cui, L., Artificial intelligence in healthcare robots: A social informatics study of knowledge embodiment, *Journal of the association for Information science and technology*, 70(4), Special Issue on Social Informatics of Knowledge, 2019, 351-369, https://doi.org/10.1002/asi.24145.

[6] Vélez-Guerrero, M. A., Callejas-Cuervo, M., Mazzoleni, S., Artificial Intelligence-Based Wearable Robotic Exoskeletons for Upper Limb Rehabilitation: A Review, *Sensors*, 21(6):2146, 2021, https://doi.org/10.3390/s21062146.

[7] Weber, P., Bolli, P., Heimgartner, N., Merlo, P., Zehnder, T., Katterer, C., Behavioral and emotional problems in children and adults with cerebral palsy, *Official Journal of the European Paediatric Neurology Society*, 2015, http://dx.doi.org/10.1016/j.ejpn.2015.12.003.

[8] Ahmadi, M. N., O'Neil, M. E., Baque, E., Boyd, R. N., Trost, S.G., Machine Learning to Quantify Physical Activity in Children with Cerebral Palsy: Comparison of Group, Group-Personalized, and Fully-Personalized Activity Classification Models, *Sensors*, 20(14):3976, 2020, https://doi.org/10.3390/s20143976.

[9] Bedla, M. et al., Estimation of Gross Motor Functions in Children with Cerebral Palsy Using Zebris FDM-T Treadmill, *Journal of clinical medicine*, vol. 11(4):954, 2022, doi:10.3390/jcm11040954.

[10] Illavarason, P., Arokia, R. J., Mohan, K. P., Medical Diagnosis of Cerebral Palsy Rehabilitation Using Eye Images

in Machine Learning Techniques, *J Med Syst.* 9, 43(8):278, 2019, doi: 10.1007/s10916-019-1410-6. PMID: 31289923.
[11] Martín, A., Pulido, J. C., González, J. C., García-Olaya, Á., Suárez, C., A Framework for User Adaptation and Profiling for Social Robotics in Rehabilitation, *Sensors*, 20(17):4792, 2020, https://doi.org/10.3390/s20174792.
[12] Usama, N., Niazi, I. K., Dremstrup, K., Jochumsen, M., Single-Trial Classification of Error-Related Potentials in People with Motor Disabilities: A Study in Cerebral Palsy, Stroke, and Amputees. Sensors, 22(4), 2022, https://doi.org/10.3390/s22041676.
[13] Faria, B. M., Reis, L. P., Lau, N., Cerebral Palsy EEG Signals Classification: Facial Expressions and Thoughts for Driving an Intelligent Wheelchair, *2012 IEEE 12th International Conference on Data Mining Workshops*, 2012, 33-40, doi: 10.1109/ICDMW.2012.89.

40

Research paper ISSN 2603-4697 (Online) Complex Control Systems Vol. 4, No. 1, 2022, 41-45

Stochastic processes with applications in supply chain management of electronic industry

Meglena Lazarova, Fatima Sapundzhi

Technical University of Sofia, 1000 Sofia, Bulgaria

South-West University "Neofit Rilski", 2700 Blagoevgrad, Bulgaria

e-mail: meglena.laz@tu-sofia.bg, sapundzhi@swu.bg

***Abstract* — A stochastic process which is a sum of a Polya-Aeppli process and a homogeneous Poisson process, called a Non-central Polya-Aeppli process and applied as a counting process in the electronic industry is considered. Some nice properties of the process and some useful applications in electronic process management are given.**

***Keywords* — electronic process management, Non-central Polya-Aeppli process, Polya-Aeppli process, pure birth process, stochastic processes**

I. INTRODUCTION

In this paper a Non-central Polya-Aeppli process, introduced by Lazarova and Minkova (2019), [1] is considered and applied in supply chain management of the electronic industry. It is a sum of Polya-Aeppli process, introduced by Minkova (2004), [2] and characterized by Chukova and Minkova (2013), [3] and a homogeneous Poisson process. The first process included in the construction of the Non-central Polya-Aeppli process is the Polya-Aeppli process (PAP) which is a compound Poisson process with geometric compounding distribution and rate $\lambda_1 > 0$.

The notation of this stochastic process is given by $N_1(t) \in PA(\lambda_1 \mathrm{t}, \rho)$ where $\rho \in [0,1)$ is an additional parameter is known as correlation coefficient, see [4].

An interesting fact for the Polya-Aeppli process is that it is over-dispersed process related to the standard Poisson process. This property gives a good flexibility of this process especially when it is used for a construction of data with big dispersion. A good indicator for this is the Fisher index of dispersion which is introduced by Fisher in 1934 [5]. It expresses the ratio of the random variable's variance to its mathematical expectation, see [6] and is given by the following formula:

$$\text{(1)} \qquad \mathrm{FI}(N) = \frac{\mathrm{Var}(N)}{\mathrm{E}(N)}.$$

The Fisher index of dispersion for the Poisson distribution is given by $\mathrm{FI}(N) = 1$. When the Fisher index is $\mathrm{FI}(N) > 1$ then it is an indication that the given distribution is over-dispersed related to the Poisson distribution. When the Fisher index is $\mathrm{FI}(N) < 1$ then the distribution is under-dispersed related to the Poisson distribution. The motivation is to construct some flexible distributions which are corresponding distributions of stochastic processes which could be applicable in models with count data.

The second process is the homogeneous Poisson process (HPP) with rate $\lambda_2 > 0$ and the notation is given by $N_2(t) \in P_oP(\lambda_2 \mathrm{t})$. The homogeneous Poisson process has exponentially independent identically distributed inter-arrival times and it has an associated counting process which represents the numbers of the events up to time t, see Arnold et al (2020), [7]. The stochastic processes $N_1(t)$ and $N_2(t)$ are independent and the Non-central Polya-Aeppli process is constructed as a sum of these two processes i.e. $N(t) = N_1(t) + N_2(t)$ where the stochastic process's rate is $\lambda_1 + \lambda_2$, see [1]. Often in the literature the expression process's rate is found as an intensity function of the process, see [7].

The paper is organized to show the importance of the chain management in the business of the electronic industry. A new application of using a stochastic counting process in the electronic process and management is given. For the construction of the theoretical profit model in the terms of the stochastic processes in section IV, a Non-central Polya-Aeppli counting process with probability mass function given in section II is used. Section III is devoted to a description of the company's electronic management and the importance of using the supply chain management in business. The significant part of this paper is given in section IV where the application of the electronic process management is shown.

II. THE PROBABILITY GENERATING FUNCTION AND THE PROBABILITY MASS FUNCTION OF THE NON-CENTRAL POLYA-AEPPLI PROCESS

The probability generating function and the probability mass function of the Non-central Polya-Aeppli process $N(t)$, see [5] are given by:

$$\text{(2)} \quad \psi_{N(t)}(s) = e^{-\lambda_1 t[1-\psi_1(s)]} . e^{-\lambda_2 t(1-s)},$$

where $\psi_1(s)=\frac{(1-\rho)s}{1-\rho s}$ is the probability generating function of the shifted geometric distribution with success probability $1-\rho<1$, denoted by $Ge_1(1-\rho)$.

(3) $P(N(t)=i)=$

$$\begin{cases} e^{-(\lambda_1+\lambda_2)t}, & i=0\\ e^{-(\lambda_1+\lambda_2)t}\left[\frac{(\lambda_2 t)^i}{i!}+\sum_{j=1}^{i}\frac{(\lambda_2 t)^{i-j}}{(i-j)!}\sum_{k=1}^{j}\binom{j-1}{k-1}\frac{[\lambda_1(1-\rho)t]^k}{k!}\rho^{j-k}\right] & i=1,2,\ldots\end{cases}$$

So this is a counting process which starts at zero and for each $t>0$ the corresponding distribution of the process is the Non-central Polya-Aeppli distribution, see [8]. Some useful and nice properties of this distribution are presented in [8] where the authors construct some graphics based on the idea of changing the different values of the parameters λ_1, λ_2 and ρ.

It is shown that the distribution's sensitivity depends on the two parameters λ_1 and λ_2 in a way that when ρ is fixed and $\lambda_1<\lambda_2$ then more events happen in the beginning and in this case the distribution forms a right tail. The Non-central Polya-Aeppli distribution is a right tailed distribution and it could be a suitable distribution for data in the industry applicable for such kind of distribution.

A good characteristic for the Non-central Polya-Aeppli is that this stochastic process is **a pure birth** process, [1]

III. ELECTRONIC MANAGEMENT AND INDUSTRY

The description of the company's management from some certain information, product materials supplying and finally to the company's production during the supply chain form the component producers and suppliers to the final step - the client's delivery is called a supply chain management, see [8]. For shortly in the literature the notation SCM is used. As there is a strong connection among all the chain's members this fact brings the importance of this term in the business because one member of the chain is capable to influence the profit and the wealth of the other business partners. The increasing of the electronic request flow and thus the business growth is dependent on the fact that the company should ensure the electronic requests to its customers on time. Also important for the company policy is to get a feedback on its products and customer service, also to have the ability to determine the future customer's needs and etc. All these steps are needed and significant for the increasing of the company's profit. Of course there will be losses too.

In 2004 Van der Vorst, [9] gives a definition of the supply chain as a sequence of decision making processes and information and material flows that aim to meet the final customer requirements and needs. The supply chain includes the manufacturers, the suppliers, the transporters, the retailers and finally the customers. The retailers and the manufacturers work together to fulfil the customers' needs. The manufacturers and the supplies work together to supply the final products in such way that there would be a minimal stocks left. Also an important element of the chain is the logistic planning and service and the transport coordination too, [9].

Many authors search and want to create a good chain supply management in the business field. For example in the paper [10] the authors explore the SCM activities performed by electronic manufacturing organizations in Malaysia. They have been conducted on different industrial sectors as grocery, computers, telecommunications etc. And the question that have to be answered is how extensively is the supply chain management adopted in Malaysia? To answer this question 25 SCM activities were presented and pointed out the adoption level of their business. The result is given after one sample T-test for identifying the organizational adoption in practice of the SMC activities is done. The result showed that if the firms would like to improve the revenue growth then it have to deliver its own products on time and directly to the customers, to get a good feedback, to determine the future customers' needs and also to use a supply chain management specialists. The conclusion that the authors made [10] is that the spread of the idea for using the supply chain management in Malaysian manufacturing industry is not very popular but those firms that decided to took the chance of using the SCM practices had a good benefit in their revenue growth.

The SCM software is an useful tool for delivering goods on time and the benefit of this process is to determine the customers' added value at minimal cost. Nowadays the companies need to invest in the information and communication technologies (ICT) and the implementation of SCM software is essential because it facilitates the operation of the entire supply chain. The use of such applications provides an accurate information about the processes in the company from the very beginning to the very end of the entire process. Of course each SCM process requires different software but often the IT developers combine and adapt these elements to the needs of each individual client.

The main components of the SCM are as follows:

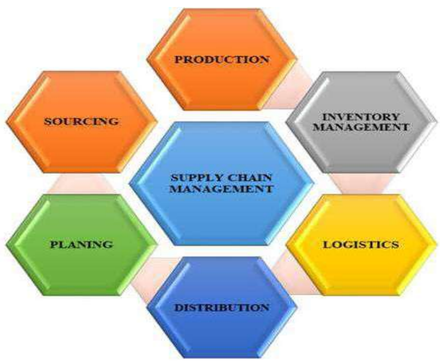


Fig. 1: Main components of SCM.

42

The applications that are integrated into the SCM can be divided into three essential groups - transaction management, communication management and relationship management [11-16].

The transactional management includes:

- Enterprise Resource Planning (ERP) which is a software that manages the company's finance resources, the supply chain and operations, commerce, reporting, manufacturing and human resource activities.
- Warehouse Management Systems (WMS) which is a software that allows the organizations to control and administer the warehousing operations from the moment that the goods or materials enter the warehouse until they are exported.
- Transport Management Systems (TMS) which is a software for planning, executing and optimizing the shipment of the goods.

The communication management includes:

- Supply Chain Event Management (SCEM) which is a software that is used to manage events that occur between organizations or partners in the supply chain.
- Radio Frequency Identification (RFID) which is a software that enables close cooperation of the supply chain partners through real-time information visibility.
- Collaborative Planning, Forecasting and Replenishment (CPFR) is an approach that aims to enhance the SCM integration by supporting and assisting joint practices.

The relationship management includes:

- Customer Relationship Management (CRM) which a software that allows companies to develop their customer's experience using complete real-time information.
- Advanced Planning and Scheduling (APS) which is a software that offers various functions of SCM at the level of strategic, tactical and operational planning.

Nowadays the development of the business and the ICT lead to the popularity of ERP systems consisting of many software components and each of them is the main functional area of the organization [11], [15], [16].
The most common components of ERP are as follows:

Fig.2: ERP components model.

As it is seen in Fig. 2 the main components of the ERP system, see [15-18] are the following:

- Supply Chain Management (SCM);
- Customer Relationship Management (CSM) – this is a strategy for managing interactions with the customers. This component helps the organizations to build the customer relationships i.e. it improves the increasing sales and profitability and also it can improve the customer service;
- Manufacturing Resource Planning (MRP) - this component provides the information about the manufacturing process and methods from which the organization can choose the best for them;
- Financial Management Component (FRM) – this component is the core of the ERP system and here some financial data from different functional departments and generates valuable financial reports are collected;
- HRM (Human Resource Management) - manages human resources and capital and maintains a database with information about employees;

Two years after the beginning of the pandemic Covid-19 and the related restrictions for the business, the cloud ERP systems have already been established as a standard system on the Bulgarian market. The cloud ERP systems are preferred by many managers in our country because of their flexibility which allows the opportunity many companies easily to meet the market changes while maintaining their efficiency.

According to CBN Pannoff, Stoytcheff & Co. prognosis, (http://web.cbn-bulgaria.com/) at about 71% of all new ERP projects in Bulgaria for 2021 are cloud solutions. Nowadays the teams are much more mobile and so the business is in a great process of digitalization. This leads to the need of using cloud ERP systems which are much better prepared to answer these challenges. These systems are more preferred in times of crisis which is very important fact.

According to CBN data by the middle of 2021 a total of 27 ERP vendors offer cloud ERP solutions through their partner network or directly in Bulgaria. The number of integrators who have announced that they are implementing such systems for Bulgaria is 56. The data show that in Bulgaria 11 out of 19 industries have installed and working cloud ERP systems. Among them there are some leading industries such as the ICT industry, manufacturing, financial and insurance activities, transport and storage. Thanks to these ERP systems it is much easier to increase or decrease the number of the users at the end of the chain, the functionalities and the parameters according to the needs of each company. Adding some new functionalities and modules is much easier and it is preferred because the customer pays only for what he uses and always works with the latest version of the software.

During the implementation of the ERP cloud system, the company team does not need to take care of the hardware,

43

the system software and the data storage. Also there is no need to think for the protector hiring and the training of a large number IT professionals.

In the following Table 1 we can see a list of some popular international SCM and ERP software, some of which have been developed by IT specialists in Bulgaria. The ranking of the best suppliers of SCM software for 2022 in the world see [18-21].

Table 1: A list of popular cloud SCM and ERP software, part of which was developed by the Bulgarian IT industry.

Product	Features	Website
Acumatica	It is an intelligent industry which is focused on cloud-based ERP platform for oversees, finances and manufacturing distribution. Also it can perform other roles in the companies.	https://www.acumatica.com
AVAMB LOGICIEL ERP (BG)	It is a cloud system that allows you to track all business processes and it is suitable in the field of online sales, marketing, transport companies and more.	https://avamb-logiciel.com/
bgERP (BG)	It is a free web-based on business management software that can track, manage and automate all important processes in a company.	https://bgerp.com/
DELTEK ERP	It offers a suite of products designed for key work environments and provides a centralized multi-functional management system.	https://www.deltek.com/
Dynamics 365 Business	It is the next generation of business solutions that provides business - ready mobile apps for finance, SCM, Retail and other ERP functionalities.	https://www.isystems.bg/
ECount ERP	It is a web-based ERP system which provides many functions to help the optimization and improvement of the business workflow.	https://www.ecount.com/
E2Open	It is a leading cloud-based, mission-critical, end-to-end supply chain management software.	https://www.e2open.com/
Odoo	It is an open-source solution that features applications, such as inventory management, financial tracking and project organization	https://www.odoo.com/
Oracle NetSuite	It is a cloud based on ERP system and it provides a full set of business solutions in one place: ERP, CRM, BI (Business Intelligence), PSA (Professional Services Automation), SRP (Services Resource Planning), HRMS (Human Resource Management Software) HR, e-commerce, etc.	https://www.balkanservices.com/
Prim (BG)	Ii is an integrated ERP system with CRM and BI module for sales, purchases, logistics management and financial operations.	https://prim.bg/
Priority Software	Priority is a module-based system offering financial supply chain, customer service and etc.	https://www.priority-software.com/
SAP S4/HANA	The Bulgarian branch of the global software leader SAP SE offers an entire portfolio in the following areas- applications, analysis, mobility, database and technology and the cloud,	https://www.sap.com/
TECHNO CLASS (BG)	It is a cloud-based integrated system designed for process management for companies with different organizational and territorial structure, subject of activity and volume of operations.	https://techno-class.com/
Tonagen ERP (BG)	Module based–integrated software management system with logical division of business processes	https://tonegan.bg
Zeron V/4 (BG)	Zeron V/4 is a software system for business management - ERP, CRM and BI.	https://zeron.bg/

IV. AN APPLICATION IN ELECTRONIC PROCESS MANAGEMENT

In this theoretical model is considered that a business company produces two types of electronic components on the market. The first type of electronic components is a board which contains motion sensors and the second one is a board which contains heat sensors. The flow of the customers' requests comes electronically to the company per unit time. The company processes the requests for a period of three days and after that the ordered components are sent back to the customers. Also the company has a market policy which includes a clause of a company penalty to the client in case of an unprocessed and returned on time request.

The state of a business company at a time t where u is an initial capital can be described by the following stochastic process:

$$P(t) = u + \pi(t) - S_{N(t)}, \quad t \geq 0,$$

where $\pi(t)$ is a non-negative function with finite variation.

The sum $u + \pi(t)$ describes the company's revenue growth and the accumulated sum $S_{N(t)}$ determines the company's expenditures. Then the process $P(t)$ defined on the complete probability space $(\Omega, \Im, P)$ represents the profit of the company's business in the time interval $[0, t]$.

The accumulated sum up to time t of the company's penalties for unprocessed and returned on time requests is given by:

$$S_{N(t)} = \sum_{i=1}^{N(t)} Z_i, \quad t \geq 0$$

The process $\{S_{N(t)}, t \geq 0\}$ is a random sum of random variables with the condition that $S(t) = 0$ when the counting process $N(t) = 0$.

The counting process $N(t) = \sup\{n : \sigma_n \le t\},$, $t \ge 0$ where $0 \le \sigma_1 \le \sigma_2 \le \dots$ are the times of requests arrivals is the Non-central Polya-Aeppli process, [1].

The interpretation of the counting process is that there are two types of unprocessed and return on time requests to the customers. The first type is counted by the Polya-Aeppli process and the second type is counted by the homogeneous Poisson process. The intensity functions of the two stochastic processes give the processing speed of the two types of the requests that the company offers to the market. They can be measured by the orders flows that the business company couldn't send on time to its customers.

The relations between the times of the requests arrivals and the counting process $\{N(t), t \ge 0\}$ are given by:

$$\{N(t) = n\} = \{\sigma_n \le t < \sigma_{n+1}\},\ \ n = 0, 1, \dots$$

The sequence $\{Z_n, n = 1, 2, \dots\}$ of independent identically distributed random variables with common distribution function F, $F(0) = 0$ and expectation μ is independent of the counting process $N(t)$.

V Conclusion

This paper gives a new approach of using a counting stochastic process in the electronic management and industry. The main aim is to determine the firm's profit and the essential key for constructing a good model with a good application is to use the chain supply management in the business field. In the theoretical part is shown that the company's loses are represented by a sum of the accumulated and not sent on time to the clients requests. The cloud ERP systems have already been discussed in this paper because they are established as a standard system on the Bulgarian market. They are preferred in our country because their flexibility allows the opportunity that the companies could easily meet the market changes.

References

[1] Lazarova, M. D., Minkova, L. D. Non-central Polya-Aeppli process and ruin probability. Ann. Acad. Rom. Sci., Ser. Math. Appl., 11, 2, 312-321, 2019.

[2] Minkova, L.D. The Polya-Aeppli process and ruin problems. J. Appl. Math. Stoch. Analysis, 3, 221-234, 2004.

[3] Chukova, S., Minkova, L. D. Characterization of the Polya-Aeppli process. Stochastic Analysis and Applications, 31, 590-599, 2013.

[4] Minkova, L. A generalization of the classical discrete distributions. Commun. Statist. Theory and Methods, 31, 871-888, 2002

[5] Fisher, R. A. The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics,* 6, 13-25, 1934.

[6] Xekalaki, E. Under and over dispersion. *Encyclopedia of actuarial science,* 3, (Eds., Teugels J. L., Sundt B.), John Wiley&Sons, Hoboken, New Jersey, 1700-1705, 2006.

[7] Arnold, R., Chukova S., Hayakawa, Y., Marshall, S. Geometric-Like Processes: An Overview and Some Reliability Applications. Reliability Engineering and System Safety 201, 1-12, 2020.

[8] Lazarova, M. D., Minkova, L. D. Non-central Polya-Aeppli distribution, *44-th International Conference on Applications of Mathematics in Engineering and Economics (AMEE'2018)*, AIP Conference Proceedings, 2048, USA, 020022-1-020022-5, 2018.

[9] Van Der Vorst, Jack GAJ. Supply Chain Management: theory and practices. Reed Business, 105-128, 2004.

[10] Zolat A.H., Ibrahim A. R., Chandran V.G.R., Sundram V. P. K. Supply Chain Management Practices and Firm Performance: An Empirical Study of the Electronic Industry in Malaysia. International Journal of Technology Diffusion (IJTD), vol. 12 Iss: 3, 210-221, 2010.

[11] Gerhard G., Ghoshal R. 6 - Enterprise planning and supply-chain interaction, Practical E-Manufacturing and Supply Chain Management, Newnes, 146-184, 2004.

[12] Stadtler, H. Supply chain management and advanced planning- basics, overview and challenges. European Journal of Operational Research, 163(3), 575-58, 2005.

[13] Heizer J., Render B., Munson C. Operations Management: Sustainability and Supply Chain Management (12th Edition) Pearson; 2020.

[13] Stanton D. Supply Chain Management For Dummies, For Dummies; 1st edition, 2020.

[14] Report: Global Supply Chain Management Software Market 2021-2025: https://www.researchandmarkets.com/

[15] https://www.techtarget.com/

[16] Sagegg O., Alfnes E., ERP Systems for Manufacturing Supply Chains: Applications, Configuration, and Performance Auerbach Publications; 1st edition, 2020.

[17] Business Dictionary: Supply Chain Definition, 2019. Wayback Machine: https://link.springer.com/content/pdf/bbm%3A978-3-030-15058-7%2F1.pdf

[18] Report: Enterprise Resource Planning (ERP) Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2021-2026: https://www.researchandmarkets.com/reports/5441989/enterprise-resource-planning-erp-market-global#src-pos-29

[19] http://web.cbn-bulgaria.com/

[20] https://technologyadvice.com/erp/

[21] Griswold M., The Gartner Supply Chain Top 25 for 2022: https://www.gartner.com/en/articles/the-gartner-supply-chain-top-25-for-2022

45

Research paper ISSN 2603-4697 (Online) Complex Control Systems Vol. 4, No. 1, 2022, 46-50

Vendor's Quality Management Assurance in Automotive Electronic Products

Petya Petkova[1], Milena Petkova[2]

[1] Technical University of Sofia, Sofia, Bulgaria
[2] University of Library Studies and Information Technologies, Sofia, Bulgaria

***Abstract* — The top factors of economic growth in the Electronic manufacturing are focusing on the "vendor-customer" relationship and identify it as a key factor for sustainable business development. In order to apply the method in practice and establish effective relationships with their customers, companies automate sales and customer service processes with the help of Customer Relationship Management systems, which favor the follow-up of the established processes and their improvement. This research paper proposes methodical tool - System for Monitoring and Assessment in Real Time Database. For facilitating the vendors' risk assessment level process, a Machine Learning Model has been developed in accordance with customized requirements of SMART Database. Random Forest Algorithms is applied for training and testing the proposed Machine Learning Model and a high level of model accuracy has been achieved – to 97% in predicting the three risk levels for vendors – high, medium and low. All the study experiments in this research are made with data, derived from worldwide electronic producer company for automotive industry.**

***Keywords*—Automotive Vendors, Electronic Products, Machine Learning, Risk Assessment, Random Forest Algorithm**

I. INTRODUCTION

In recent years, there is a change in the overall business model and business process of organizations in the Electronics field. Manufacturers of electronic components or electro-mechanical products provide a valuable competitive advantage for sustainable development, but standing in front of world material crisis. In order to apply the method in practice and establish effective relationships with their customers, companies automate sales and customer service processes with the help of Customer Relationship Management (CRM) systems, which favor the follow-up of the established processes and their improvement. This leads to optimization of resources and increase of efficiency.

CRM solutions allow management to track the activities of the sales department, to provide better customer service, to manage marketing campaigns, contracts, pricing policies, discounts and more.

The top factors of economic growth in the Electronic manufacturing are focusing on the "vendor-customer" relationship and identify it as a key factor for sustainable business development. Customer-oriented organizations in the Electronic field need a flexible "vendor-customer" relationship management system. That should allow the storage and analysis of various traceability, information, tracking all stages of profitability, quality or delivery performance for the region and market. To ensure a smooth supply and rhythm of production, no less important aspect of the activity is to build long-term stable relationships with suppliers of raw materials. In electronics manufacturing, the required quality level of produced electronic modules as electronic devices and microchips are validated through qualification testing based on standards and user-defined requirements. The challenge for the global electronics industry is that product validation is time-consuming and costly. On other hand the increasing demand and the current lack of materials for mass production in electronics lead to stock shortages, delayed deliveries, and following difficulties to achieve the targets, required for a high level of competitiveness and reliability in Automotive Industry. This paper proposes a methodical tool for overcoming the milestone of the vendor's evaluation procedure in worldwide electronics manufacturing company for automotive and household industry, called in this paper "ComEX". The proposed tool is called SMART database - System for Monitoring and Assessment in Real Time. Machine learning model for solving classification task is developed and applied in the process of vendors' risk assessment level.

II. VENDORS' EVALUATION PROCEDURE AND SMART DATABASE

Automotive industry, especially the electronic components producers are highly dependently of their vendors. In Quality Management Assessment System, the vendors' evaluation and selection is a crucial milestone which overcoming will race the manufacturer into leading automotive positions. All automotive companies work in accordance with IATF 16949:2016 [1] Automotive QMS Standard which replaces ISO/TS 16949 standard. It applies to the design/development, production and, when relevant, installation and servicing of automotive-related products [2]. Nevertheless, each automotive company has the right to implement its own procedure in compliance with the standard for vendors' selection and evaluation.

A. *Vendors' evaluation procedure*

The process of selection and evaluation of suppliers varies depending on the product on offer. There are several steps to be passed in preliminary checking before their inclusion in the vendors' approved list. The preliminary selection contains *phase of research* and *phase of pre – selection.* Define abbreviations and acronyms the first time they are used in the text, even after they have already been defined in the abstract

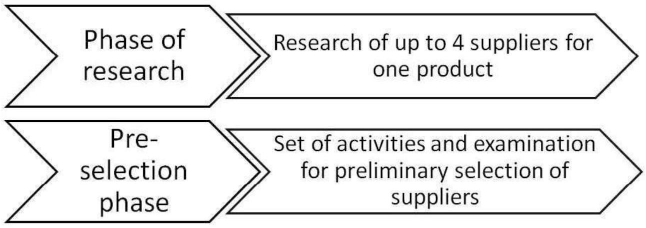


Figure 1. Preliminary Vendors' selection phases of ComEX

The pre – selection phase is applied to all the researched suppliers in the 1st phase. The activities and examinations in this phase vary based on the product and service they are providing. This set includes:
1) Verification of the proposed products and services for their compliance with company's requirements;
2) Purchase of product's sample quantity for verification through operational process;
3) Filling out questioners;
4) In house vendor's audit in order to establish the actual state of the processes and its ability to fulfil the requests;
5) Financial assessment for vendors' reliability and sustainability

In order to assure a sustainable quality delivers of electronic products or services is necessary a periodic reassessment of suppliers to be applied. The criteria for periodic revaluation should be related to the quality of the delivered products, compliance with delivery deadlines, flexible payment schemes, etc. and concern the validity of the cooperation agreement. The reassessment of suppliers implies to the same list of approved vendors, each of which shall be assigned the appropriate assessment according to the specified criteria. This re-evaluation process requires a management system implemented for every category of company's vendors. This Category Management cycle is composed of 5 process steps presented in figure 2.

Figure 2. Re-assessment category management cycle.

The category management perspective of improvement is one of the steps during the vendors' evaluation process in the ComEX company. Together with the development of proposed SMART database system

B. SMART Database

The implementation of developing SMART database is a tool for prediction using already validated key performance indicators for evaluation as the data constructed them is mined in real time process of testing the electronic products. The most common type of testing in the electronics industry—sequentially run electrical multi-parameter tests on the Device-under-Test (DUT). Data mining framework can identify the tests that have strong correlation to pending failure of the device in the qualification (tests sensitive to pending failure) as well as to evaluate the similarity in test measurements, thus generating knowledge on potentially redundant tests. SMART database aims to identify top worst vs. top best vendor's category of the Category Management Cycle. It is the linkage between the following aspects as a part of *Monitoring & Assessment of the Supplier performance:*
1) Supplier appraisal when the incumbent supplier is competing for the renewal of an existing contract.
2) Management of approved supplier lists.
3) Quality management & Evaluation.
4) Scrap & Cost of Non-Quality Reduction .
5) Turn-around incident treatment time management.
6) Risk assessment and live overview.

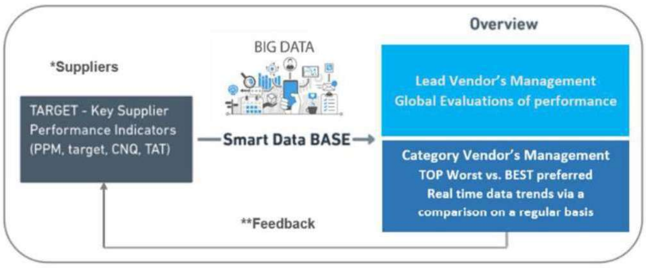


Figure 3. Re-assessment category management cycle

*For the purposes of this practice database the term 'Suppliers' includes contractors for EMS raw material's distributors and manufacturers.

**Feedback for 'performance monitoring' aims measuring, analyzing and managing Key Performance Indicators (KSPIs) as a supplier's ability to comply with, and preferably exceed, their contractual obligations, in accordance with the SQA Manual.

There are six basic steps for implementation of SMART database. Some of them need to be taken in cooperation with other departments (IT and SQA) of the company before put in use, other need to be required as additional information from the vendors. These are the following:

- ✓ Access – online via intranet
- ✓ Interface – home page with menu box
- ✓ Reporting - export a statistic data in excel
- ✓ Criteria and targets – based on the overall performance for last evaluated period
- ✓ Results – options for various visualizations by diagrams
- ✓ Feedback – regular email notification requesting:
    - Corrective actions
    - Continues improvement plan

On the figure 4 the referent architecture of the SMART Database System is presented. It is still a prototype one, but used as a referent one for the Machine Learning Model development.

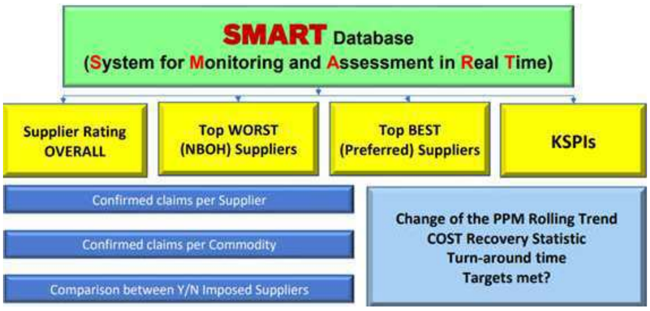


Figure 3. Referent Architecture of SMART Database System

This system represents an intelligent tool for quality assurance management. The idea is to extract data from real-time testing and combining with predefined indicators to perform vendors' overall rating, top worst and top best suppliers and most important KPIs.

As a part of classifying the top worst and top best vendors a machine learning method has been applied by using Random Forest algorithm for classification.

III. Random Forest Algorithm for Classification of Electronic Products Vendors

As the main goal of implementation SMART database is to enhance the vendors' evaluation procedure by integrating intelligent tool for prediction of new vendor and classified between the rate of three levels – Top Risk, Medium and Low Risk, vendors. For the purpose of prediction model development, a classification task has been solved by using machine learning algorithm Random Forest.

A. Feature Extraction

The features selected for the development of ML model are taken from the vendor's evaluation criteria and could be divided into two groups – quantitative and qualitative features. The table below gives detailed information about the evaluation criteria and the methods how the features are compound and selected.

Table. 1 Evaluation Criteria and scoring methods of the features +customer (in the second row)

	EVALUATION CRITERIA	CRITERIA ON SCORING			
Quantitative Features	Delivered quantity	The total quantity delivered by the vendor for the accounted period			
	Quantity of rejection	The amount of rejected quantity for particular vendor as a sum of rejected by incoming inspection and during production process			
	Peace Per Million PPM (Maximum of 25 point) PPM Value = (Number of confirmed NOK parts for the evaluation period) / (Number of delivered quantities for the evaluation period) X1000000	0 DPPM			25
		1-50 DPPM			20
		51-100 DPPM			15
		101-200 DPPM			10
		201-300 DPPM			8
		301-500 DPPM			5
		>500 DPPM			0
	Number & Type of complaints (Maximum of 25 points) Decreasing of points in case of confirmed supplier claims. The given score is calculated by specific formula as a weighted fraction of numbers of different types of complains. NOTE: No negative score is allowed	C1 (customer complaint)			10
		C2 (ComEX production line)			2
		C3 (incoming inspection)			1
	Cost of Non-Quality (Maximum 15 points)	100% CNQ Recovery			15
		< 100% CNQ recovery			0
	Response Time (Maximum of 20points)	<10 working days to submit 8D report			20
		10< X <15 working days to submit 8D report			15
		16< X <20 working days to submit 8D report			5
		>21 working days to submit 8D report			0
Qualitative Features	Accuracy of documentation/communication (max 15 points)	The score is based on the vendor's analysis report quality.			
	Availability of Product or Service	Risk Level			
		High	Medium	Low	
		5	3	1	
	Supplier Performance Rating	Risk Level			
		High	Medium	Low	
		5	3	1	
	Strategic Partnership	Risk Level			
		High	Medium	Low	
		5	3	1	

The quantitative features consist of a) delivered by the vendor quantity for the accounted period; b) rejected from delivery quantity both from incoming inspection and production lines process; c) PPM (piece per million) – the fraction of total scrapped quantity and total delivered quantity as a million parts; d) number of complains scored by specific know-how company's formula based on three different types of complains; e) cost of non-quality; and f) response time. There are four qualitative indicators used as features for classification dataset, which scoring criteria could be considered as a subjective factor at first sight, but also is calculated very precisely via confidential know-how formula. These features are: a) accuracy of documentation / communication; b) availability of product and service; c) supplier performance rating and d) strategic partnership.

The categorization of the vendors in three levels of risk (high, medium and low) is calculated again using a complicated method that is under company's confidentiality. For the purpose of our study and the needs of this paper it was kindly provided information of full list vendors' classification for the first six months of 2021.

B. Machine learning model methodology

The development of Machine Learning model has passed through four consequence steps, shown on the fig. 4.

48

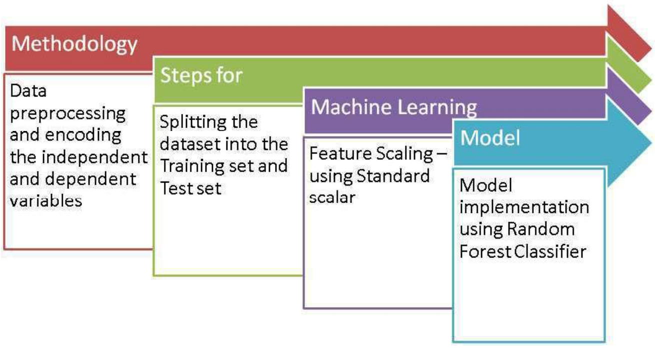


Figure 4. Methodology for ML model implementation

The dataset of preprocessing module is constructed of 10 features (independent variables) described in the previous section, one label feature (dependent variable) which represents our class – high, medium and low level risks and 311 examples of vendors of all categorical electronic products for the first semester of 2021.

Index			DELIVERED QUANTITY										
0	1	S1	164955	0	0	25	20	15	15	3	1	1	3
1	2	S2	513400	30	58.434	23	0	15	15	1	1	1	3
2	3	S3	906500	3000	3309.43	24	20	15	10	5	1	3	2
3	4	S4	3.8625e+07	0	0	25	20	15	15	1	1	1	3
4	5	S5	2.25e+06	0	0	25	20	15	15	3	1	1	3
5	6	S6	36000	0	0	25	20	15	15	3	1	1	3
6	7	S7	588520	0	0	25	20	15	15	5	1	1	2
7	8	S8	3.5025e+07	0	0	25	20	15	15	1	1	1	3
8	9	S9	4.10485e+06	0	0	25	20	15	15	5	1	1	2
9	10	S10	2.57642e+06	0	0	25	20	15	15	1	1	3	3
10	11	S11	2721.2	0	0	25	20	15	15	5	1	1	2

Figure 5. Classification Dataset

The ML model is trained with 80% of the dataset and test with the rest 20 %. After the splitting process a feature scaling is applied to the data features, using the method of standardization:

$$\mathrm{Fe}_{\mathrm{stand}} = \frac{Fe - mean(Fe)}{Standard\ Deviation\ (Fe)}$$

where Fe is the value of the every feature.

```
In [25]:print(sc.transform(X_test[:,2:]))
        [[-1.06767706 -1.06767706 -1.06767706 ... -0.28988045 -2.41361855
  -1.77881442]
        [-1.06767706 -1.06767706 -1.06767706 ... -0.28988045 -2.41361855
  -1.77881442]
        [-1.06767706 -1.06767706 -1.06767706 ... -1.0895684  -2.41361855
   0.14330572]
 ...
        [-1.06767706 -1.06767706 -1.06767706 ... -1.88925635 -2.41361855
  -1.77881442]
        [-1.06767706 -1.06767706 -1.06767706 ... -0.28988045 -2.41361855
  -1.77881442]
        [-1.06767706 -1.06767706 -1.06767706 ... -1.88925635 -2.41361855
  -1.77881442]]
```

Figure 6. Test data values after standardization process

C. Random Forest implementation and results

Random Forest (RF) is a popular and powerful ensemble supervised classification method [3]. Ensemble algorithms have achieved success in machine learning by combining multiple weak learners to form one strong learner [4].

Due to its superior accuracy and robustness, and some ability to offer insights by ranking of its features, RF has effectively been applied to various machine learning applications, including many in bioinformatics and medical imaging [5]. Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model [6]. The reason why this algorithm has been chosen for the building the proposed model is avoiding the model over fitting and no need of hyper parameters' optimization. Random Forest hyper parameters work perfectly both for increasing the predictive power and festers the model.

For building ML model with Random Forest algorithm initially it has been used 10 trees and activation function "entropy" for model quality measurement. These two parameters are very important for the general model evaluation performance. The entropy criterion computes the Shannon entropy of the possible classes. It takes the class frequencies of the training data points that reached a given leaf m as their probability. Using the Shannon entropy as tree node splitting criterion is equivalent to minimizing the log loss (also known as cross-entropy and multinomial deviance) between the true labels y_i and the probalistic predictions $T_k(x_i)$ of the tree model T for class k.[7] On the figures 7 and 8 are shown a single tree of the applied RF algorithm (fig.7) and the first 5 trees (fig. 8) of the same algorithm with <<n_estimators =10, criterion = 'entropy'>>.

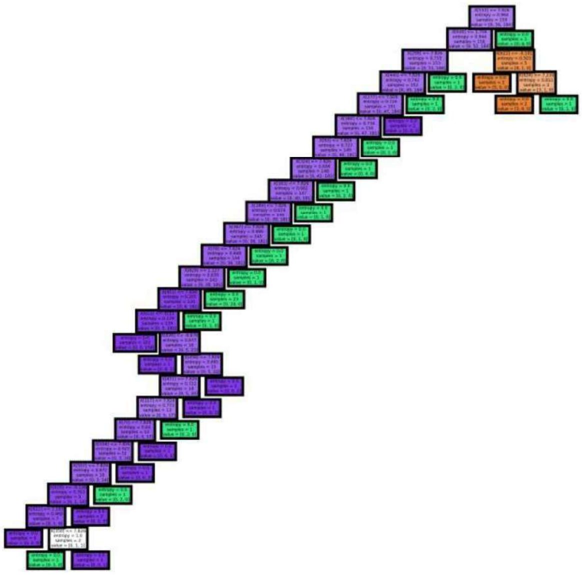


Figure 7. Individual tree of the created ML model with RF algorithm

49

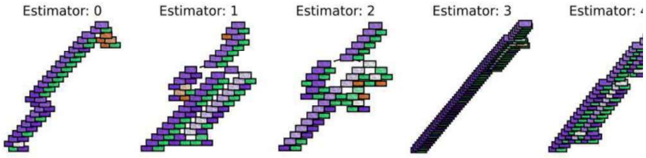


Figure 8. The first five trees of the created ML model with RF algorithm

The best accuracy has been achieved with 100 trees. The initially created model with 10 trees and activation function “entropy” achieved evaluation performance accuracy of 92%. In table 2 are presented the RF model accuracies with different numbers of trees with the one and the same activation function “entropy”:

Table 2 ML model accuracy achieved by different numbers of decision trees in the RF algorithm

No of Trees	Accuracy %
10	92
50	94
100	97
250	95
500	94

As there is no rule of the thumb regarding the optimal number of the trees in the Random Forest algorithm, in general the more trees are used the better results, but from the statistics which resulted in table 2 is visible that for every specific case the tuning of the hyper parameters in algorithm is individual. In the fig. 9 are shown the first 10 correctly predicted examples of the model.

Figure 9. Model’s correctly predicted examples

CONCLUSION AND FUTURE WORK

This research work proposes a System for Monitoring and Assessment in Real Time – SMART Database for evaluation the vendors’ quality performance in the automotive industry. It will help both Vendor’s Quality Management Assurance and Customer Relationship Management CRM systems. Its implementation in automotive electronic producers’ attempts to increase the customer satisfaction and company’s performance coefficient by real-time analysis, predicting, preventing shortages, target’s follow up, investment planning data overview with its negative and positive trends. Although its implementation carries a lot of technological and economic challenges with the standard adoption and big amount of investments for the companies, the benefits for all the involved parties are uncountable. A Machine Learning model has been developed especially for the customized requirements of the SMART Database by applying a Random Forest algorithm. As the system attempts to grow in time and the amount of data will be continuously pouring out the algorithm need to be enhanced which is a discussion for future work. Proposed in this paper Random Forest machine learning model is going to sophisticated and two more algorithms – ANN and Deep Learning will be implemented and tested in the ML model of SMART Database.

ACKNOWLEDGMENT

This research is supported by Bulgarian National Science Fund in the scope of the project “Exploration and application of statistics and machine learning in electronics” under contract number КП-06-Н42/1.

REFERENCES

[1] IATF 16949:2016 “International Automotive Task Force”, Derived by: https://www.iatfglobaloversight.org/iatf-169492016/about/
[2] ISO/TS 16949:2009, Quality management systems — Particular requirements for the application of ISO 9001:2008 for automotive production and relevant service part organizations “International Organization for Standardization”, 2009, derived by: https://www.iso.org/standard/52844.html
[3] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp.5–32, 2001.
[4] Jeremy D. Rogers and Steve R. Gunn, Ensemble Algorithms for Feature Selection, Rogers, J.D., Gunn, S.R., Ensemble Algorithms for Feature Selection. In: Winkler, J., Niranjan, M., Lawrence, N. (eds) Deterministic and Statistical Methods in Machine Learning. DSMML 2004. Lecture Notes in Computer Science, vol. 3635. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11559887_11
[5] Dragutin Petkovic, Russ Altman, Mike Wong, Arthur Vigil. Improving the explainability of Random Forest classifier – user centered approach, Pacific Symposium on Biocomputing 2018.
[6] Niklas Donges, “Random Forest Algorithm: A Complete Guide - All you need to know about the random forest model in machine learning”, built in Expert Contributer Network, July 22, 2021 Updated: April 14, 2022, derived by: https://builtin.com/data-science/random-forest-algorithm
[7] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, 2825-2830, 2011. Derived by https://scikit-learn.org/stable/modules/tree.html#tree-mathematical-formulation

Complex Control Systems

ISSN: 2603-4697 (Online)