

# Application of machine learning methods for the prediction of distress in patients with oncological diseases

Ginka Marinova<sup>1</sup>, Todor Ganchev<sup>1</sup> and Nedyalko Nikolov<sup>2</sup>

1 – Technical University of Varna, Department of Computer Science and Engineering, 9010, 1 Studentska Str., Varna, Bulgaria

2 – Technical University of Varna, Department of Software and Internet Technologies, 9010, 1 Studentska Str., Varna, Bulgaria

Corresponding author contact: [gmarinova@tu-varna.bg](mailto:gmarinova@tu-varna.bg)

**Abstract.** *Distress management is of particular importance in all disease treatment strategies that aim to cope with medical conditions, which require prolonged therapy. Here, we present results obtained in a comparative study of various classification methods for automated distress detection. For the purposes of the present study, use was made of a common experimental protocol that relies on a dataset of approximately 6 000 oncological patients at different stages of therapy. The dataset consists of the binary responses to specific questions in a purposefully-designed self-evaluation questionnaire on the degree of distress. Conducted, within such a framework, was a performance assessment of three distress detectors based on Multilayer Perceptron Neural Network (MLP NN), boosting and bagging meta-classification methods and evaluated, further, was the performance of nine characteristic descriptors (KR1-KR9) representing the informative content of the dataset in different ways. The results obtained in the experiments prove conclusively that one of the characteristic descriptors, KR8 and KR9, significantly outperform the other descriptors in terms of classification accuracy, precision, recall, and F-measure.*

**Keywords:** distress management, oncological dataset, classification, boosting, bagging, Multilayer Perceptron Neural Network

## 1 Introduction

The development of information technologies and data storage in clinical psychological diagnostics prove to be extremely beneficial for tracing patients with oncological diseases, based on the use of psychological tests. It is recommended to apply tests to find ways and screening tools to provide the necessary interventions according to the specific needs of the patient. As the volume of data increases, there is a need to use different methods to obtain the necessary and useful information. Data extraction is a powerful tool for analysis, synthesis, knowledge retrieval, and distress prediction.

Distress is a base factor influencing the clinical and mental state of cancer patients. Early psychodiagnostic evaluation of distress and appropriate intervention is crucial in their antitumor treatment, therapy and to improve the quality of life of patients and their families (VanHoose L. et al., 2015). The "Screening Tool for Measurement of Distress" (Riba, Michelle B. et al., 2019; Власаков, В., и колектив 2015) is an extensive approach and a rapid method of identifying more accurately the mental health of patients with tumours, assessing the individual and psychosocial needs through a validated questionnaire that seem more appropriate for determining the level and factors of distress (Riba, Michelle B. et al., 2019). Recognizing the importance of addressing the emotional and social concerns of oncology patients, the National Comprehensive Cancer Network (NCCN) strongly recommends distress screening and management as a standard of care within oncology health services delivery.

The screening tool is embedded as a therapeutic link with health professionals (doctors, psychologists- oncologists, etc.) to initiate interventions related to the management of disastrous events. As the volume of stored data increases, techniques of machine learning turn out to play a crucial role in finding the patterns and extracting the knowledge to provide better patient care and effective diagnostic capabilities. Knowledge discovery from this data can greatly benefit patients and psycho oncologists for better distress management. Machine learning is connected with the implementation of computer software that learns autonomously and is an approach for extracting features from information sources (Mitchell, T., 1997).

The primary focus of the present paper is the critical evaluation of three methods and nine characteristic descriptors on the distress detection task (Section 2). Full advantage is, therefore, taken of the prevalent well-established classifiers and purposefully-developed characteristic descriptors (Marinova et al., 2020) and present their performance in terms of classification accuracy, precision, recall, and F-measure (Section 4).

## 2 Methods

The Ensemble algorithm - Bagging is a modern method for increasing the efficiency of the classification models used. In this technology, all classifiers are implemented in parallel based on different methods (Breiman, L., 1996). The accuracy of prediction of the combined classifiers constructed with the help of bagging is higher than the accuracy of the individual models. This algorithm allows the development of combined models, which usually work better than a single model. The Bagging algorithm is described in the following steps:

Training extract is represented as:

$$D = \{(x_1, y_1), \dots, (x_n, y_n)\} \quad (1)$$

Selection of  $t$  elements from training data set  $D$ :  $D_1, \dots, D_s$

Training in  $D_1, \dots, D_s$  and obtaining a sequence of  $s$  outputs:  $f_1(x), \dots, f_s(x)$

The resultant end classifier is:

$$f(x) = \sum_{i=1}^s \text{sign}(f_i(x)) \quad (2)$$

The Boosting method is for the sequential development of a composition of ensemble algorithms in machine learning (Кашницкий, Ю.С., & Игнатов, Д.И., 2015). It is an iterative algorithm that is applied to classification in an effort to increase the accuracy of analytical models. The steps required to execute the algorithm are:

Initialization of the weights of training examples:

$$v_i = 1/N \quad (3)$$

$N$  - number of training parameters.

Learning through the target function:

$$\text{err} = \sum_i v_i [y_i \neq t_i] \quad (4)$$

$v_i$  weight of  $i$ -th training example;

$y_i$  and  $t_i$  output signals;

$[y_i \neq t_i]$  indicator function;

Calculation of classifier weights:

$$\alpha_m = \ln \left( \frac{1 - e_m}{e_m} \right) \quad (5)$$

Updating weights:

$$v_i^{(m+1)} = v_i^{(m)} \exp(\alpha_m [y_i \neq t_i]) \quad (6)$$

Decision making is based on the calculation:

$$y(x) = \sum_{m=1}^M \alpha_m y_m(x), \quad (7)$$

$y_m(x)$  the output signal of the  $m$ -th classifier for the input vector.

Artificial neural networks are an effective tool for solving problems requiring analytical calculations, similar to the human brain, implemented in the used to study and modelling of various data from psychological diagnostics. It is a dynamic system applied to increase the accuracy and objectivity of various psychological tests. The Neural network (NN) is a method for parallel processing of information that

has the quality to store the user data. The training process in neural networks is appropriate to change the weights as a means to achieve the necessary properties. The information in the neural network is superimposed during the training process, as the strength of the connections between the individual nodes is formed by the weights of the respective connections, which are used to store information for a particular study (Mitchell, T., 1997).

### 3 Classification metrics

The choice of classification metrics is an important step in conducting a study. Assessed, thus, in machine learning to solve problems for classification in medicine, is the quality of the models to be subsequently compared via different algorithms. The confusion matrix is presented in tabular form to visualize the effectiveness of the studied classification algorithms, comparing the predicted value of the target variable with its actual value. This is a way to divide objects into four categories with connectivity between the combination of the correct answer and the answer of the algorithm.

**Table 1.** Confusion matrix

Category i		Expert evaluation	
		Positive +	Negative -
System evaluation	Positive +	(True Positive) TP	(False Positive) FP
	Negative -	(False Negative) FP	(True Negative) TN

Major calculated indicators are Accuracy, Precision and Recall, and F1- measure, (Kantardzic, M., 2002; Hossin, M., Sulaiman, M.N. 2015; Калчева-Арабаджиева, Н., & Николов, Н., 2017).

A popular indicator of classification is accuracy. It is the ratio between the correctly classified examples to the total number of examples in the extract (Kantardzic, M., 2002; Hossin, M., Sulaiman, M.N. 2015).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

The Precision is the ratio of correctly identified positive results (Kantardzic, M., 2002; Hossin, M., Sulaiman, M.N. 2015).

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

The Recall is the evaluation of the ratio of all correctly positively predicted objects in a given class to the total number of correctly classified ones (Kantardzic, M., 2002; Hossin, M., Sulaiman, M.N. 2015; Zhu W., Zeng, N, & Wang, N., 2010; Powers, D.M.W., 2011; Muhamedyev, R., 2015).

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

F1-measure is the average harmonic value of the precision and sensitivity of the model (Zhu W., Zeng, N, & Wang, N., 2010; Powers, D.M.W., 2011).

$$F \text{ measure} = \frac{2}{1/Recall+1/Precision} \quad (11)$$

### 4 Experimental setup

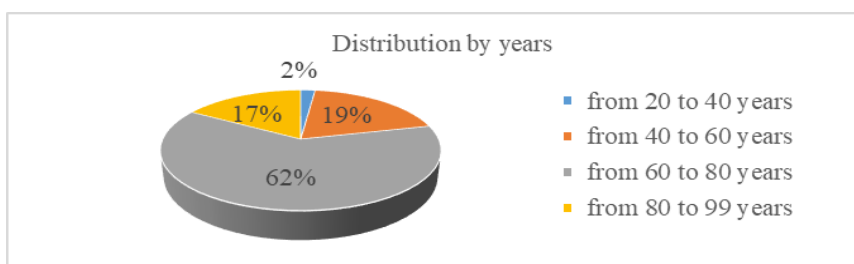
Applied in the proposed paper is an array of data consisting of responses for 6,000 cancer patients through purposefully developed "Screening Tools for Distress Measurement" software, collected from the medical visits to the respective patients, with the answers themselves comprising forty-three features. Since the answers are binary further transformations should be made. The transformations of the data are called the synthesis of characteristic descriptors. The characteristic descriptor is a vector by which machine learning methods are applicable (Marinova et al., 2020). The questionnaire is systematized into

five groups: practical problems, family problems, emotional problems, spiritual/religious concerns, physical problems related to the disease, using a combination of data extraction and classification approach is used. Tags are set for high and low levels of distress upon the opinion of specialist psycho-oncologists on the effect of therapeutic care upon the quality of life. Three algorithms for the distress detection task have been investigated: Bagging, Boosting, and Multilayer Perceptron Neural Network (MLP NN). For building machine learning models, it's important to measure the results of our models. Metrics are accuracy, F measure, precision and recall were considered relevant to compare the performance of three different algorithms on a dataset.

## 5 Experimental results

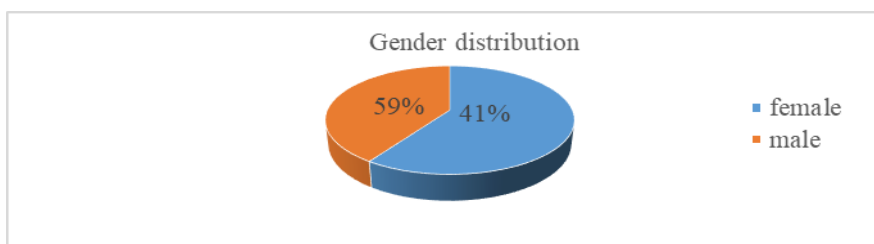
The subject area of research is the clinical psychological diagnostics, for the extraction of characteristic descriptors for the detection of high levels of distress. The information in the database is for patients with a specific oncological disease.

There was step-by-step processing of data containing the patient's age, gender, list of signs divided into five categories, practical problems related to the disease, family problems, emotional problems, spiritual/religious problems, physical problems. Presented in **Fig. 1** is the distribution of the cancer patients data set by years.



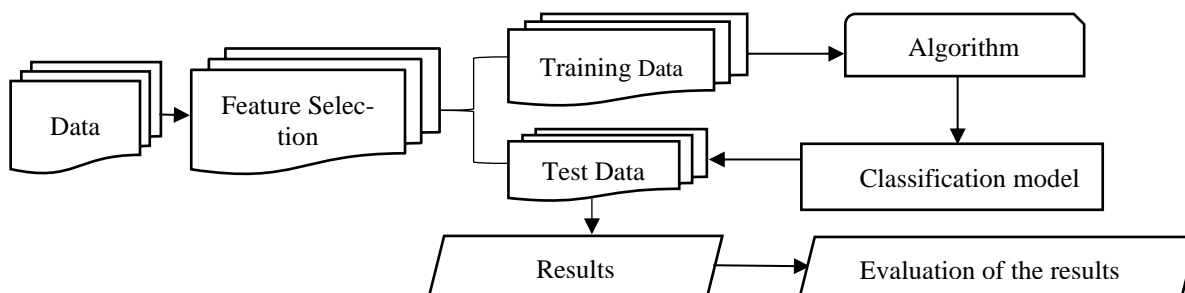
**Fig. 1.** Distribution of patients with cancer by years

In **Fig.2**, we show the distribution of patients by gender.



**Fig. 2.** Distribution of patients with oncological diseases by gender

In **Fig.3**, we show the division of the data into training and test sets.



**Fig. 3.** The overall workflow of experimental work.

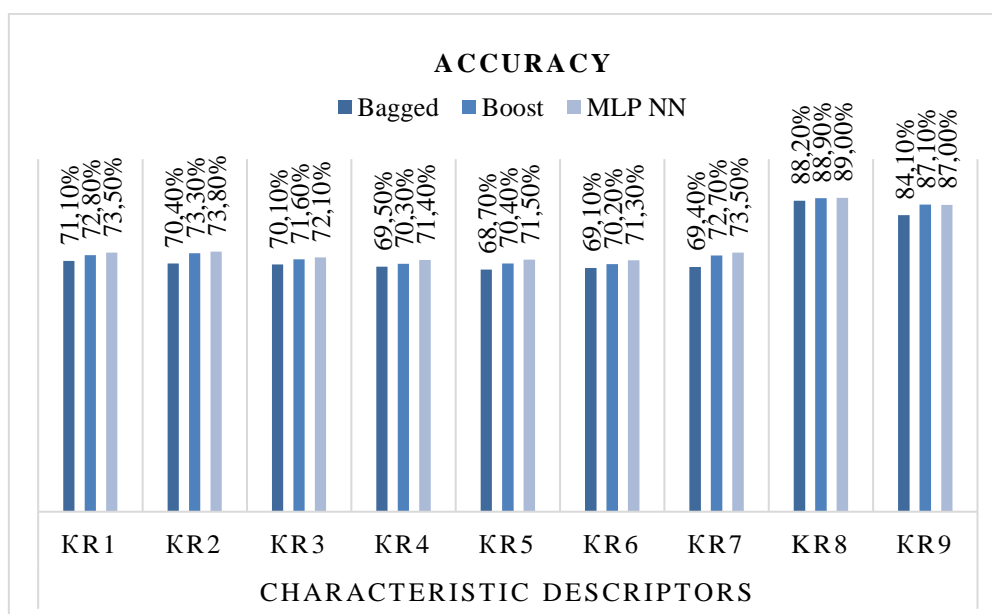
Nine characteristic descriptors were synthesized, including seven (KR1, KR2, KR3, KR4, KR5, KR6, KR7) obtained by summing up different weights for certain indications, and KR8, KR9 after a selection of features converted into a digit number by raising it to the power of two (Marinova et al., 2020).

The aim is to train a selected classifier and to calculate the accuracy, precision, and recall of its work. The size of the set consists of 9240 different records, divided into training 80% and test set 20%. The number of classes is two: small and big distress. The algorithms used for the study are Boosting, Bagging, and MLP NN. Ten times cross-validation is used to evaluate the selected method.

**Table 2.** Evaluation results for the three classifiers and the nine features

Classifier	Characteristic descriptors								
	KR1	KR2	KR3	KR4	KR5	KR6	KR7	KR8	KR9
<b>Bagged</b>	71.1%	70.4%	70.1%	69.5%	68.7%	69.1%	69.4%	88.2%	84.1%
<b>Boost</b>	72.8%	73.3%	71.6%	70.3%	70.4%	70.2%	72.7%	88.9%	87.1%
<b>MLP NN</b>	73.5%	73.8%	72.1%	71.4%	71.5%	71.3%	73.5%	89.0%	87.0%

In **Table 2**, we show that an MLP NN produces a relatively high accuracy of 89% for characteristic descriptor KR8, which is higher than the accuracy of other selected methods and characteristic descriptors. These results are summarized in **Fig. 4**.



**Fig. 4.** Comparative analysis in terms of detection accuracy

In order to evaluate the quality indicators of a given algorithm, introduced are Precision and Recall, and the aim is to obtain a description of the ability to classify the selected classifier. The results in **Table 3** show that in the MLP NN for characteristic descriptors KR8, KR9, obtained is a precision rate of 94%.

**Table 3.** Evaluation results in terms of precision

Classifier	Characteristic descriptors								
	KR1	KR2	KR3	KR4	KR5	KR6	KR7	KR8	KR9
<b>Bagged</b>	67.33%	66.22%	66.97%	65.52%	64.41%	64.83%	63.93%	81.53%	77.68%
<b>Boost</b>	69.58%	66.58%	83.71%	71.39%	69.55%	71.90%	70.99%	84.40%	89.15%
<b>MLP NN</b>	75.52%	75.42%	78.03%	79.23%	75.64%	74.27%	76.29%	93.80%	94.24%

Displayed in Fig. 5 are the results for the precision of test methods.

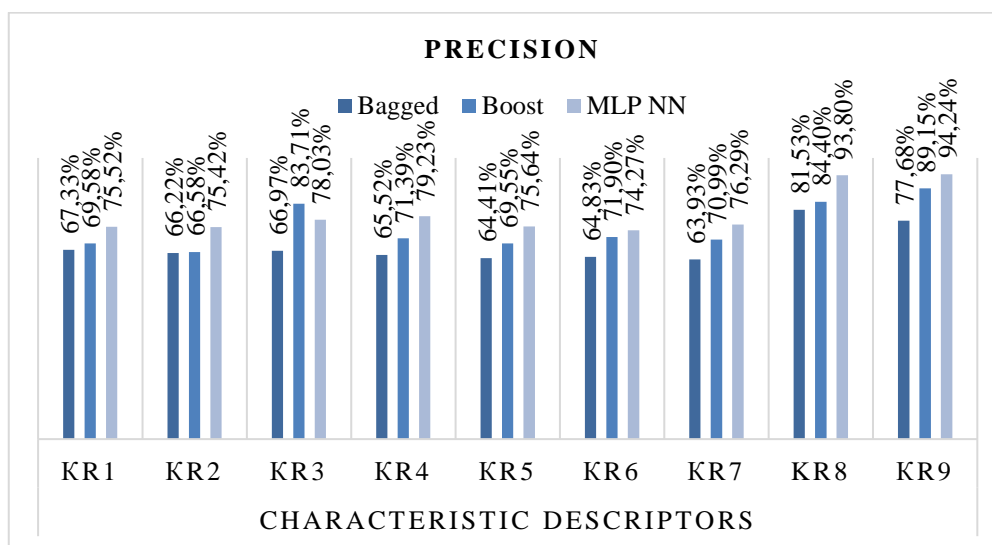


Fig. 5. Evaluation results in terms of precision

The Recall is the metric used to evaluate the results attained. The Recall is represented as the proportions of examples, of all positive examples that are classified correctly (Powers, D.M.W., 2011). From the analysis of the results in Table 4, it can be concluded that for a characteristic descriptor KR8 for the Boosting method obtained is 91.52%.

Table 4. Tested Recall methods

Classifier	Characteristic descriptors								
	KR1	KR2	KR3	KR4	KR5	KR6	KR7	KR8	KR9
<b>Bagged</b>	74.68%	74.03%	72.97%	73.28%	72.45%	72.65%	73.13%	90.98%	87.62%
<b>Boost</b>	76.39%	76.62%	75.65%	74.10%	74.15%	74.03%	76.04%	91.52%	90.24%
<b>MLP NN</b>	73.25%	73.28%	74.28%	74.68%	72.68%	71.69%	73.72%	85.92%	84.48%

The results are shown in Fig. 6.

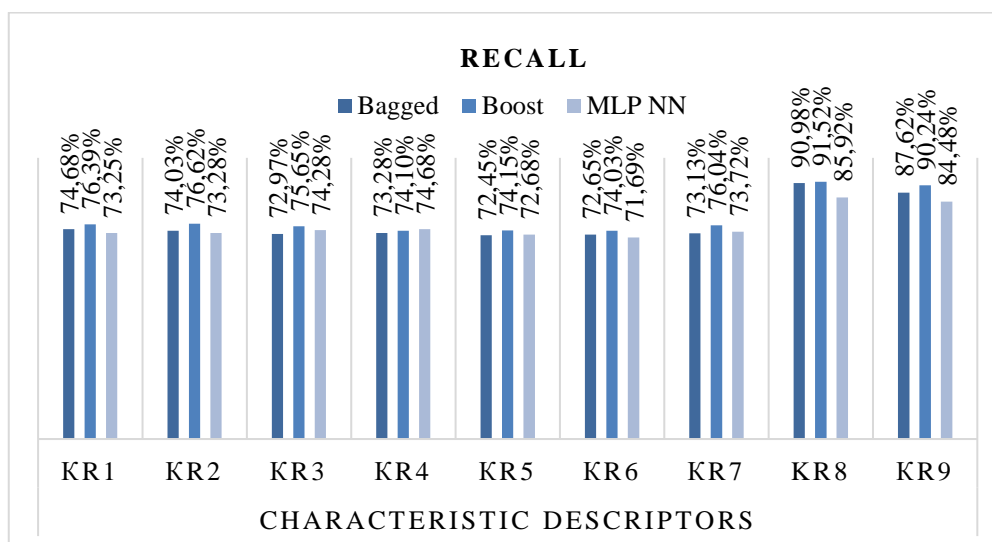


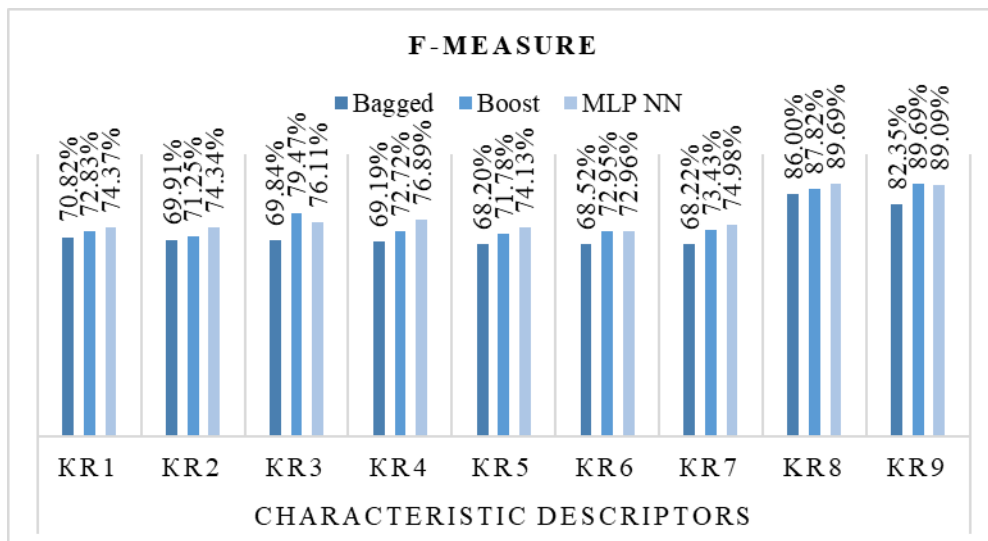
Fig. 6. Comparative analysis - recall

F-measure is represented as a combined measure of the weighted average harmonic value of recall and accuracy. This measure is used to explain and understand number of cases the model is likely to predict correctly and the amount of true specimens the model will not miss.

**Table 5.** The F-measure of test methods

Classifier	Characteristic descriptors								
	KR1	KR2	KR3	KR4	KR5	KR6	KR7	KR8	KR9
<b>Bagged</b>	70.82%	69.91%	69.84%	69.19%	68.20%	68.52%	68.22%	86.00%	82.35%
<b>Boost</b>	72.83%	71.25%	79.47%	72.72%	71.78%	72.95%	73.43%	87.82%	89.69%
<b>MLP NN</b>	74.37%	74.34%	76.11%	76.89%	74.13%	72.96%	74.98%	89.69%	89.09%

Fig. 7 reveals the results for the F-measure of the methods that were put to test.



**Fig. 7.** The F-measure for the various distress detectors

## 6 Conclusion

Solving classification problems is widely applied in various areas, but in the medical field, in particular, it is related to the data processing to support the decision-making process. The synthesized characteristic descriptors help to increase the accuracy of the classification and reduce the computational processes for creating classification models. The proposed study has employed three techniques of machine learning to classify the data set relative to the problem - distress in patients with tumors. The adopted machine learning methods present different results according to the processed data. The implemented tests include specific indications to summarize the mental health of the patients and the degree of their endangerment. To achieve good and reasonable results for assessment and control of distress, not only screening is mandatory, but also adequate guidance, follow-up, and treatment. More precisely, the findings of the study show that more accurate results are obtained for characteristic descriptors KR8, KR9 when compared with the other ones. This is an indicator of their adequacy in terms of data.

## Acknowledgement

The first author (G.M.) would like to acknowledge the support and funding of the project NP5 "Investigation into the possibilities of integrating machine learning and Blockchain technologies for the Internet of things".

## References

- VanHoose L. et al. (2015) An analysis of the distress thermometer problem list and distress in patients with cancer, *Support Care Cancer*, 23, 1225–1232. <https://doi.org/10.1007/s00520-014-2471-1>
- Riba, Michelle B. et al. (2019). Distress Management, Version 3, NCCN Clinical Practice Guidelines in Oncology, *JNCCN.org*, 17(10), 1229-1249. <https://doi.org/10.6004/jnccn.2019.0048>
- Власаков, В., и колектив (2015) Психосоциална подкрепа и рехабилитация в онкологията, национален експертен борд клинично ръководство, основано на доказателства, *Амп Трейсър*.
- Breiman, L. (1996) Bagging Predictors, *Machine Learning*, 24, 123-140. <https://doi.org/10.1007/BF00058655>
- Кашницкий, Ю.С., & Игнатов, Д.И. (2015) Ансамблевый метод машинного обучения, основанный на рекомендации классификаторов Интеллектуальные системы, Теория и приложения, *Интеллектуальные системы. Теория и приложения*, 19(4), 37-55.
- Mitchell, T. (1997) *Machine Learning, McGraw-Hill Science/Engineering/*.
- Kantardzic, M., (2002) *Data Mining Concepts, Models, Methods, and Algorithms*, (1st ed.) Wiley-IEEE Press.
- Hossin, M., Sulaiman, M.N. (2015) A Review on Evaluation Metrics for Data Classification Evaluations, *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 5(2), 1-11. <https://doi.org/10.5121/ijdkp.2015.5201>
- Калчева-Арабаджиева, Н., & Николов, Н. (2017) Сравнителен анализ между наивния байсов класификатор и метода на опорните вектори използващ оптимизация при класификация на български текст в машинното обучение, *Списание Компютърни науки и технологии*, ТУ – Варна, 97-105.
- Zhu W., Zeng, N., & Wang, N. (2010) Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations, *NESUG: Health Care and Life Sciences*, 19, 1-9.
- Powers, D.M.W., (2011) Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation, *Journal of Machine Learning Technologies*, 2(1), 37-63.
- Muhamedyev, R. (2015) Machine learning methods: An overview, *Computer Modeling & New Technologies*, 19(6), 14-29.
- Marinova, G., Ganchev, T., and Nikolov, N., (2020) Synthesis of characteristic descriptors for the detection of distress, *2020 International Conference on Biomedical Innovations and Applications (BIA)*, Varna, Bulgaria, 73-76, <https://doi.org/10.1109/BIA50171.2020.9244488>