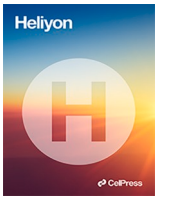
Heliyon 9 (2023) e18328

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Proof of concept: Predicting distress in cancer patients using back propagation neural network (BPNN)

Schulze Jan Ben *, Marc Dörner, Moritz Philipp Günther, Roland von Känel, Sebastian Euler

Department of Consultation-Liaison-Psychiatry and Psychosomatic Medicine, University Hospital Zurich, University of Zurich, Zurich, Switzerland

ARTICLE INFO

Keywords:
Prediction of distress
Distress thermometer
Risk factors for distress
Back propagation neural network (BPNN)
Artificial neural network (ANN)
Psycho-oncology
Cancer

ABSTRACT

Background: Research findings suggest that a significant proportion of individuals diagnosed with cancer, ranging from 25% to 60%, experience distress and require access to psycho-oncological services. Until now, only contemporary approaches, such as logistic regression, have been used to determine predictors of distress in oncological patients. To improve individual prediction accuracy, novel approaches are required. We aimed to establish a prediction model for distress in cancer patients based on a back propagation neural network (BPNN).

Methods: Retrospective data was gathered from a cohort of 3063 oncological patients who received diagnoses and treatment spanning the years 2011–2019. The distress thermometer (DT) has been used as screening instrument. Potential predictors of distress were identified using logistic regression. Subsequently, a prediction model for distress was developed using BPNN.

Results: Logistic regression identified 13 significant independent variables as predictors of distress, including emotional, physical and practical problems. Through repetitive data simulation processes, it was determined that a 3-layer BPNN with 8 neurons in the hidden layer demonstrates the highest level of accuracy as a prediction model. This model exhibits a sensitivity of 79.0%, specificity of 71.8%, positive predictive value of 78.9%, negative predictive value of 71.9%, and an overall coincidence rate of 75.9%.

Conclusion: The final BPNN model serves as a compelling proof of concept for leveraging artificial intelligence in predicting distress and its associated risk factors in cancer patients. The final model exhibits a remarkable level of discrimination and feasibility, underscoring its potential for identifying patients vulnerable to distress.

1. Introduction

Distress among patients with cancer is characterized as a complex and unpleasant encounter encompassing physical, social, psychological (cognitive, behavioral, emotional), and/or spiritual aspects, which can impede effective coping with the disease, its associated manifestations, and treatment [1]. Studies reveal that a significant amount of patients with cancer (25%–60%) state to be distressed when evaluated, emphasizing the need for psycho-oncological services [2–4]. Consequently, all patients with cancer should be screened for distress [1,5,6]. A widely used instrument for distress screening is the distress thermometer (DT), combined with a problem list. Nurses regularly administer the DT during hospitalization and outpatient care as part of the standard procedure [7]. It is

* Corresponding author.
E-mail address: jan.schulze@usz.ch (S. Jan Ben).

https://doi.org/10.1016/j.heliyon.2023.e18328
Received 13 August 2022; Received in revised form 12 July 2023; Accepted 13 July 2023
Available online 15 July 2023

important to identify sources and predictors of distress, so that hospital staff has the opportunity to implement specific interventions, which reduce psychosocial burden. Previous studies have emphasized several factors associated with higher levels of distress. These include female sex [8,9], younger age [10,11], unmarried patients [10], patients diagnosed with specific types of cancer (e.g. of the breast, lung, colon, pancreas, brain, or head and neck) [12–14], low social support and increased fear of recurrence [15], as well as a lower level of life satisfaction [16]. Moreover, recent research identified several items of the DT's problem list as potential sources of distress. On the one hand, there are nonphysical predictors from the emotional domain [11,17,18], such as nervousness [19,20] or depression [20] and items from the practical and family domain [11], such as financial strain [21] or dealing with children at home [17]. On the other, physical problems such as pain [20], appearance [17] and fatigue [16] also predicted distress in cancer patients.

The already extensive literature on factors predicting distress in patients with cancer provides an adequate basis for future research. While distress in cancer patients is influenced by various interconnected factors, previous research has predominantly relied on conventional prediction models, such as logistic regression. However, these statistical approaches often fall short in fully capturing the multifaceted nature and interdependencies underlying distress, leading to less accurate predictions.

To tackle this challenge, more precise estimation of distress predictors can be achieved by applying innovative approaches, such as Back Propagation Neural Network (BPNN). BPNN derives from Artificial Neural Network (ANN), which emulates the human brain. By leveraging its self-organizing, self-learning and adaptive nature, ANN can effectively discern intricate non-linear associations among variables. Therefore, utilizing BPNN allows for a more comprehensive understanding of the complex relationships involved in distress prediction [22]. As a widely used and relatively advanced method, introduced over 40 years ago, BPNN has already proven its ability to contribute to the diagnostic workup of psychiatric and psychosomatic disorders, to predict length of psychiatric hospitalization and to find an accurate prediction model for suicide attempts [23–28].

As a novelty, this study is the first, which aimed to establish a prediction model for distress in patients with cancer based on BPNN. This study might serve as a proof of concept, possibly encouraging other researchers in this field to use artificial intelligence. Our prediction model using BPNN is better suited for identifying cancer patients at risk of distress, especially in large samples with diverse participant characteristics, surpassing the limitations of conventional methods [22–28]. Moreover, the results might help improve individual prediction accuracy and develop more precise prevention and intervening strategies for cancer patients vulnerable to distress.

2. Methods

2.1. Subjects and data collection

For this study, we retrospectively collected the data from the case files of oncological outpatients and inpatients (diagnosed and treated at the Comprehensive Cancer Center Zurich at the University Hospital Zurich from 2011 to 2019). The study sample selection process is depicted in Fig. 1. There were 13174 cases with a primary diagnosis of cancer between 2011 and 2019. Before treatment began, more than half of patients agreed to the reuse of their data for research purposes (general consent; 55% of the 13174 cases with

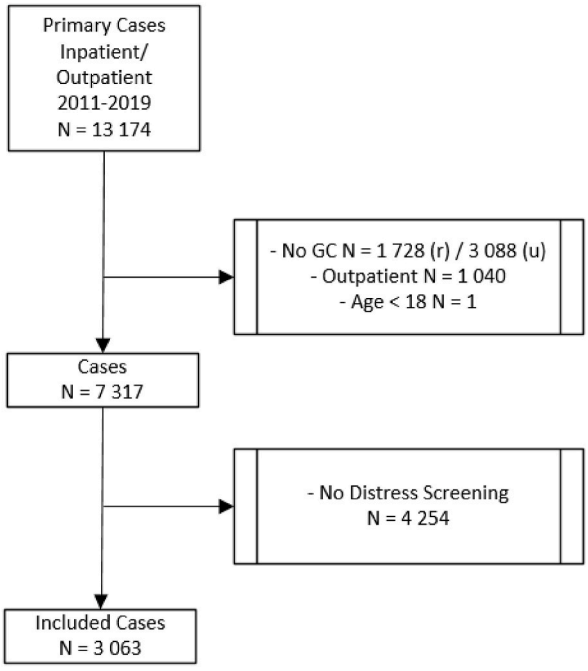


Fig. 1. Study sample selection process. r = rejected; u = unknown; GC = general consent.

a primary cancer diagnosis). Patients were informed on the utilization of their data for research projects (non-genetic) and on their right to object to their former consent without justification at all times. The patient's consent is saved in the hospital information system. Patients who objected to participate (n = 1728), with unknown general consent (n = 3088), and who were treated as outpatients (n = 1040) only, were also excluded from the study. The Ethics Committee of the State of Zurich, Switzerland, approved the study (BASEC NR. 2020-00977; June 2020). Of the remaining 7317 cases, 3063 were screened for distress and formed the final study sample. Distress screenings up to six months after the initial cancer diagnosis were included in the analysis.

Table 1 illustrates the absolute and relative distributions of cancer entities in the final study sample.

2.2. Measurements

Distress levels were evaluated using the Distress Thermometer (DT), which employs an 11-point visual scale (range 0 (no distress) to 10 (maximum distress)). Past studies have proposed a cut-off score of $\geq$5 as a threshold for considering referral to psycho-oncological services [29,30]. All patients with a DT score of $\geq$5 were asked if they wanted to call on psycho-oncological services. If those patients demand help, the treating physician refers them to psycho-oncological services, provided by a psychiatrist or psychologist at the hospital. This process is also described in Ref. [29].

As candidate predictors for distress, we considered socio-demographic variables, such as age, sex, marital status, mother tongue (German) and nationality, cancer entities, stage of cancer at initial diagnosis, psychiatric disorders, the length of hospital stay, medication administered during the first six months after initial cancer diagnosis and all items from the problem list. Cancer stage was categorized in accordance with the guidelines established by the Union for International Cancer Control (UICC) [31]. Furthermore, we analyzed the age-adjusted Charlson comorbidity index (CCI) as a potential distress predictor. The CCI illustrates the impact of chronic comorbid diseases on mortality [32–34].

All candidate predictor variables were dichotomized (0 = No and 1 = Yes (one encoding)). Age and length of hospital stay were measured on a metric scale. Nationality was categorized into 1 = Switzerland, 2 = Europe (including Switzerland) and 3 = non-European. All data were accessed via the clinical management software (KISIM) and the institutional cancer register (OncoStar).

2.3. Statistical analysis

Relative and absolute distributions, mean scores, and standard deviation were computed to describe qualitative and quantitative data. Preliminary screening of predictor variables was performed with logistic regression analysis. BPNN was utilized to build the final prediction model for distress in patients with cancer. Level of significance (two-sided p-value) was $p < 0.05$. IBM SPSS Statistics, version 27 (Chicago, IL), was used for statistical analysis, and Matlab (R2019a. Mathworks Inc.) for BPNN computations.

Information flows from the input layers, through the hidden layers, and eventually reaches the output layers in a feedforward neural network. This one-directional movement of information characterizes it as a feedforward network. The primary objective is to reduce disparities between the computed output and the desired output from the training sample. To achieve this, the network undergoes iterations by virtue of the error rate received in the preceding iterations. By continuously reducing the error rates, the reliability and generalizability of the backpropagation neural network (BPNN) model are enhanced. Consequently, BPNN serves as a widely adopted approach for managing and adaptively monitoring artificial neural networks [22,25].

Table 1
Cancer entities in the final study sample.

ICD-10 (3-digit diagnosis code)	Type of cancer	2011–2019	(%)
		N	
C00–C14	Malignant neoplasms of lip, oral cavity and pharynx	423	13.8
C15–C26	Malignant neoplasms of digestive organs	345	11.3
C30–C39	Malignant neoplasms of respiratory and intrathoracic organs	542	17.7
C40–C41	Malignant neoplasms of bone and articular cartilage	3	0.1
C43	Melanoma and other malignant neoplasms of skin (without basalioma)	135	4.4
C45–C49	Malignant neoplasms of mesothelial and soft tissue	44	1.4
C50	Malignant neoplasms of breast	30	1
C51–C58	Malignant neoplasms of female genital organs	42	1.4
C60–C63	Malignant neoplasms of male genital organs	384	12.5
C64–C68	Malignant neoplasms of urinary tract	104	3.4
C69–C72	Malignant neoplasms of eye, brain and other parts of central nervous system	374	12.2
C73–C75	Malignant neoplasms of thyroid and other endocrine glands	37	1.2
C76–C80	Malignant neoplasms of ill-defined, other secondary and unspecified sites	4	0.1
C81–C96	Malignant neoplasms of lymphoid, hematopoietic and related tissue	586	19.1
D00-D09	In situ neoplasms	10	0.3
Total		**3063**	**100**

Note: ICD-10: International Statistical Classification of Diseases and Related Health Problems, Version 10.

3. Results

3.1. Depiction of sociodemographic and further variables of the final study sample

Table 2 depicts sociodemographic and further characteristics of the final study sample. One third of the study sample is female. The average age (±SD) was 61.5 ± 13.9 years (range 18–95).

3.2. Preliminary screening of potential predictor variables of distress on the basis of logistic regression analysis

To preliminary screen possible predictor variables for distress, we applied binary and multivariable logistic regression analysis. In the regression analysis, a dichotomized distress score ≥5 versus <5 was the dependent variable. In total, there were more than 60 independent variables to be screened as candidate predictors for distress in patients with cancer. Scatter plots indicated that there was an approximately linear relationship between each independent variable and the dependent variable. Thus, logistic regression analysis is an adequate method to preliminary screen candidate predictor variables.

Table 3 highlights the 13 significant independent variables derived from the significant logistic regression model for distress in cancer patients ($X^2 = 1073.556$, $p < 0.001$). Cox & Snell R^2 and Nagelkerke R^2 indicate a high model quality (0.295 respectively .396).

Binary logistic regression indicated that among the large number of candidate predictors only certain items of the problem list predicted cancer patients' distress with statistical significance. The most significant predictor for distress was fear, followed by sadness. Patients reporting fear or sadness were about 206% and 308% more likely to feel distressed than those who did not indicate these emotional problems. Other emotional problems, such as depression, nervousness and worries predicted distress as well. Practical problems regarding work or school as well as physical problems, such as digestive problems, immobility, insomnia, fatigue and pain also predicted distress but to a lesser extent.

The receiver operating characteristic (ROC) curve of the binary logistic regression model showcases a sensitivity of 78.4% and a specificity of 72.9%, along with an area under the curve (AUC) of 0.825 (Fig. 2). These results indicate a high level of discrimination in the model.

3.3. Structure of the BP neural network

In our study, we employed a standard three-layer backpropagation neural network (BPNN) architecture comprising an input layer, a hidden layer, and an output layer. The input variables for the BPNN prediction model, aimed at assessing distress, were derived from 13 significant independent variables identified through logistic regression. The dependent variable is a dichotomized distress score (1 = distress score ≥5, 0 = distress score <5), that was set as output variable. We calculated the number of neurons in the hidden layer by equation $H = \sqrt{M + N} + \alpha$. α represents a constant from one to ten, N the number of input neurons, and M the number of output neurons. Fig. 3 displays the BPNN structure.

3.4. BP neural network training

The dataset was randomly partitioned into three subsets: the training data (70%), the validation data (15%), and the test data (15%). Varying the amount of neurons in the hidden layer yields distinct values for coincidence rate (π), positive and negative predictive value, sensitivity, specificity, and the number of iterations. A more accurate BPNN model is achieved by a higher coincidence rate and a lower number of iterations. In our BPNN model for assessing distress in cancer patients, the hidden layer neurons were explored within a range of four to thirteen. Through training of the BP neural network and performing repetitive data simulations for different numbers of hidden layer neurons using Matlab, it was determined, that the most precise evaluation indexes were obtained, when the number of neurons fell within a range of seven to nine (Table 4).

Table 2
Depiction of the study sample.

	total	women	men
Patients	**N (%)** 3063 (100)	1022 (33.3%)	2041 (66.6%)
Age (mean and SD; range)	61.5 (13.9); 18-95	60.6 (14.8); 18-92	62 (13.4); 18-95
Married	1750 (57%)	511 (50%)	1239 (60.7%)
Primary language German	3019 (98.6%)	1012 (99%)	2007 (98.3%)
Nationality			
ch	2538 (82.9%)	870 (85.1%)	1668 (81.7%)
eu	2797 (91.3%)	946 (92.6%)	1851 (90.6%)
non-eu	267 (8.7%)	76 (7.4%)	191 (9.4%)
Advanced cancer stage	1451 (47.4%)	526 (51.5%)	925 (45.3%)
CCI (≥5)	1865 (61.4%)	625 (61.2%)	1240 (60.7%)
Distress score ≥5	1710 (55.8%)	675 (66%)	1085 (53.1%)
Current psychiatric disorder	209 (6.8%)	72 (7%)	137 (6.7%)

Note: SD: standard deviation. ch: Switzerland, eu: Europe, non-eu: non-European. CCI: Age-adjusted Charlson comorbidity index.

Table 3
Binary logistic regression for distress in cancer patients.

Variables	Standard B	OR	P value
PLPhysProb Indigestion	0.343	1.409	0.018
PLPhysProb Eating	0.459	1.583	<.001
PLPhysProb Mouth sores	0.381	1.464	0.01
PLPhysProb Getting around	0.325	1.385	0.005
PLPhysProb Sleep	0.374	1.453	<.001
PLPhysProb Fatigue	0.472	1.603	<.001
PLPhysProb Pain	0.403	1.497	<.001
PLEmoProb Nervousness	0.477	1.612	<.001
PLEmoProb Depression	0.416	1.516	0.05
PLEmoProb Sadness	0.723	2.060	<.001
PLEmoProb Fears	1.124	3.076	<.001
PLEmoProb Worry	0.688	1.989	<.001
PLPractProb Work/school	0.505	1.658	0.008
Cox & Snell R^2	0.295		
Nagelkerke R^2	0.396		

Note: OR: odds ratio. PLPhysProb: Problem list physical problems. PLEmoProb: Problem list emotional problems. PLPractProb: Problem list practical problems.

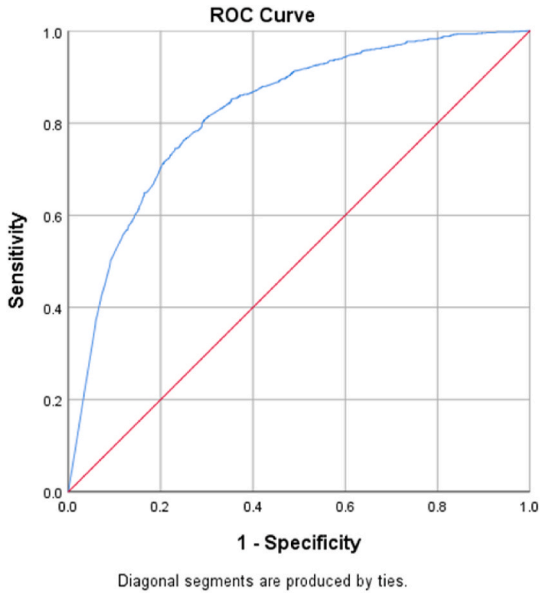


Fig. 2. The binary logistic regression model's receiver operating characteristic (ROC) curve (area under curve (AUC): 0.825).

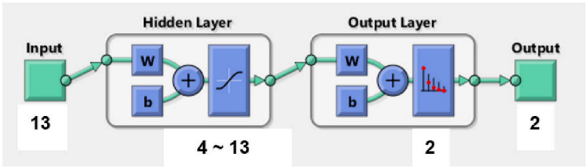


Fig. 3. BP neural network structure. w: weights coefficient (w). b: threshold value (b).

During the training of the BPNN, varying the weight coefficients (w) and threshold values (b) results in distinct learning models. As the primary values of (w) and (b) are randomly chosen from a range of −1 to 1, repetitive data simulations were conducted using Matlab, with distinct initial values of (w) and (b), along with a range of seven to nine neurons in the hidden layer. Evaluating the performance based on the evaluation indeces, Network 4, consisting of eight neurons in the hidden layer, emerges as the ideal BPNN model for predicting distress in cancer patients (Table 5).

Table 4
Different number of hidden layer neurons with different predicted evaluation indices.

N_{hidden}	Se	Sp	1-Sp	PV+	PV-	π	Number of iterations
7	76.6	73.7	0.26	79.5	70.2	75.4	21
8	79.0	71.8	0.28	78.9	71.9	75.9	20
9	79.2	71.4	0.29	78.7	71.9	75.8	21

Note: Nhidden: Number of hidden layer neurons. Se: Sensitivity. Sp: Specificity. PV+: Positive predictive value. PV-: Negative predictive value. π: Total coincidence rate.

Table 5
Different initial values of the weights coefficient (w) and the threshold value (b) predicting different evaluation indices.

		π			Total sample					Number of iterations
Network	Nhidden	Training sample	Verification sample	Test sample	Se	Sp	PV+	PV-	π	
1	7	76.6	73.4	74.7	79.0	71.6	78.8	71.8	75.8	22
2	7	76.1	75.8	71.2	76.6	73.7	79.5	70.2	75.4	21
3	8	76.3	72.8	75.6	77.7	72.9	79.3	71.0	75.7	20
4	**8**	**75.1**	**78.6**	**76.9**	**79.0**	**71.8**	**78.9**	**71.9**	**75.9**	**20**
5	9	76.1	74.9	75.4	79.2	71.4	78.7	71.9	75.8	21
6	9	76.6	75.2	72.8	79.9	70.4	78.3	72.4	75.8	21

Note: Nhidden: Number of hidden layer neurons. Se: Sensitivity. Sp: Specificity. PV+: Positive predictive value. PV-: Negative predictive value. π: Coincidence rate.

3.5. BP neural network verification

The error histogramm of our BPNN model is displayed in Fig. 4, which suggests a high BPNN model quality with an excellent discrimination efficiency and predominantly small errors by virtue of the errors' concentration around the zero error line in most cases.

Fig. 5. Confusion matrices of the BPNN model (a-d); (a) confusion matrix of the training sample; (b) confusion matrix of the validation sample; (c) confusion matrix of the test sample; (d) confusion matrix of the total sample. Bottom right corner of confusion matrices (a-d): coincidence rate. Bottom left corner of confusion matrices (a-d): sensitivity. Upper right corner of confusion matrices (a-d): positive predictive value. Bottom-middle of confusion matrices (a-d): specificity. Middle-right of confusion matrices (a-d): negative predictive value.

4. Discussion

In this study, using data from the DT and problem list, it was possible to build a precise prediction model for distress in cancer patients. To our best knowledge, this study is the first, which examines predictors of distress in cancer patients utilizing BP neural network analysis, thus serving as a proof of concept study for further research embracing advanced statistics (broadly referred to as machine learning or artificial intelligence).

The majority of our final study sample (more than 55%) reported to be distressed and in need of psycho-oncological services, which is in line with past studies that explored the prevalence of distress in oncological patients [2–4,35]. Contrary to prior studies [8–14],

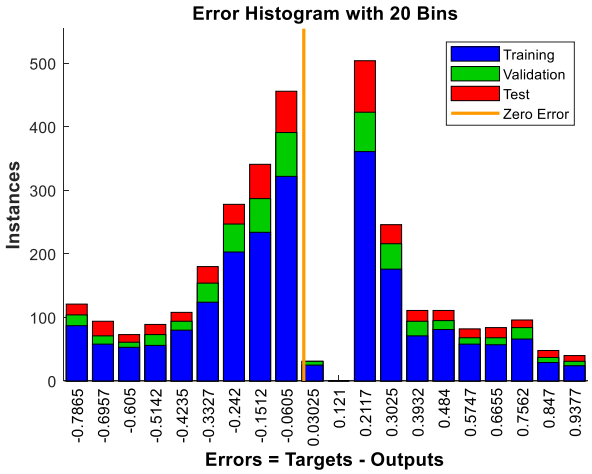


Fig. 4. BPNN model's error histogramm for distress in patients with cancer.

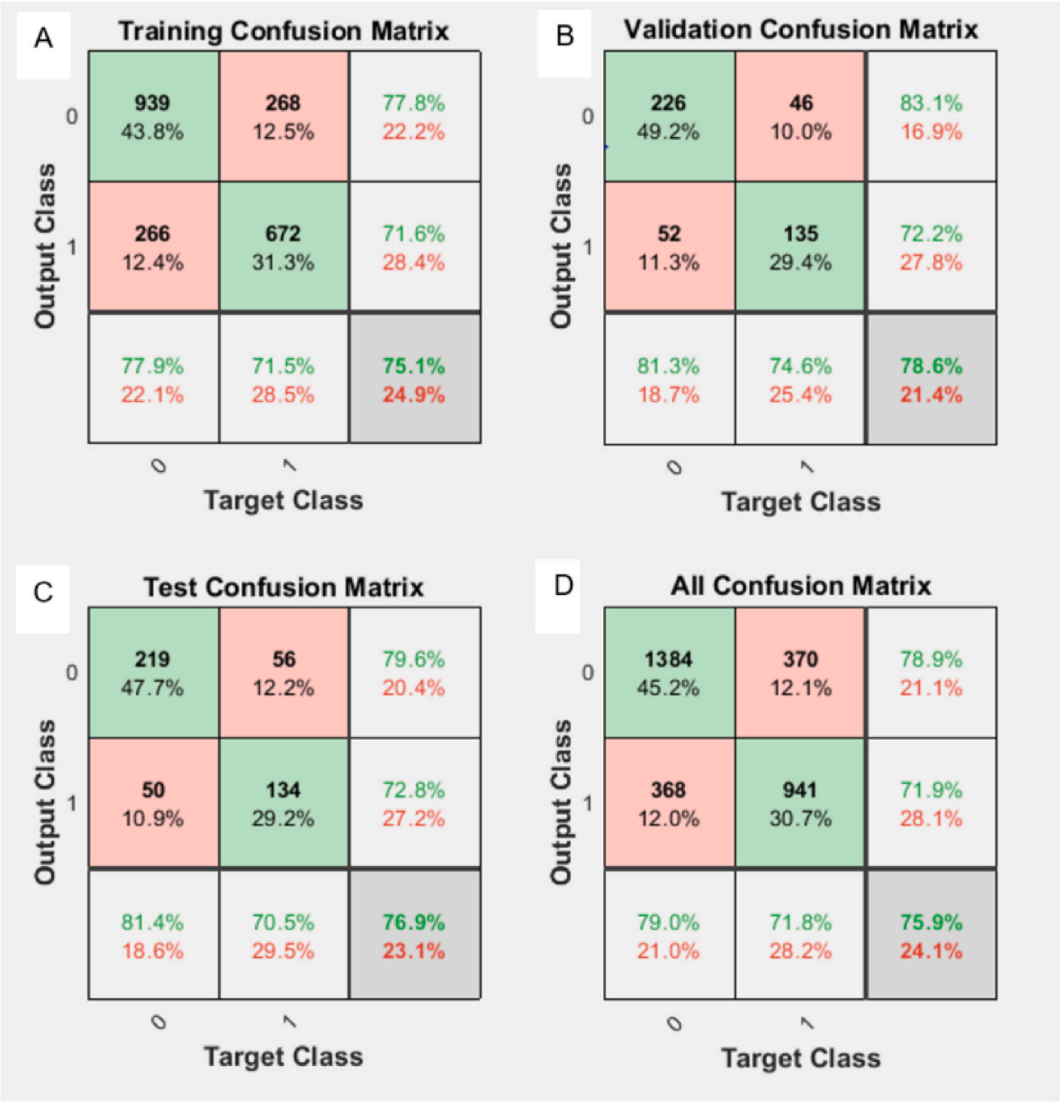


Fig. 5. Highlights confusion matrices with a coincidence rate (π) of more than 75% for each sample and a total coincidence rate of 75.9%. The ROC curve of the BPNN model is demonstrated in Fig. 6. The total AUC added up to 0.827 and is another indication of the high level of discrimination.

our study did not identify sociodemographic aspects, such as age, sex or marital status, or specific types of cancer predicting distress. There are multiple reasons possibly explaining this observation. First, our study uses data from Switzerland collected between 2011 and 2019. By contrast, prior studies on predictors of distress used more historic data collected in the United States of America or other parts of the globe, thus involving other treatment regimen and lifestyles. Second, some studies included a much smaller sample size with only a few hundred cases [8–11,14], or only specific types of cancer [10,11]. Our study comprises more than 3000 cases, all types of cancer and all currently available cancer treatments (including immunotherapy).

Other examined variables, such as mother tongue, nationality, psychiatric diagnoses, length of hospital stay, administered medication and the CCI did not predict distress in oncological patients. To our knowledge, there is no other published data considering these variables. Future studies should include these variables to validate our results. However, in accordance with the already existing literature [15–21], many items from the problem list (physical, emotional and practical problems) predicted distress in our study sample. The largest risk factors for distress in our study sample were fear, sadness and worry from the emotional domain, which is consistent with previous studies [7].

Proof of concept is achieved by meticulously describing statistical procedures in the results section: In summary, after applying logistic regression analysis to filter potential predictors of distress, the 13 significant independent variables from the problem list were utilized as input variables for the BPNN prediction model. The output variable was a dichotomized distress score of $\geq$5 versus $<$5. Subsequently, the data was randomly divided into a verification sample, a test sample and a training sample. The most precise prediction model for distress in oncological patients is network 4, with 8 neurons in the hidden layer, as was demonstrated by our 3-layer

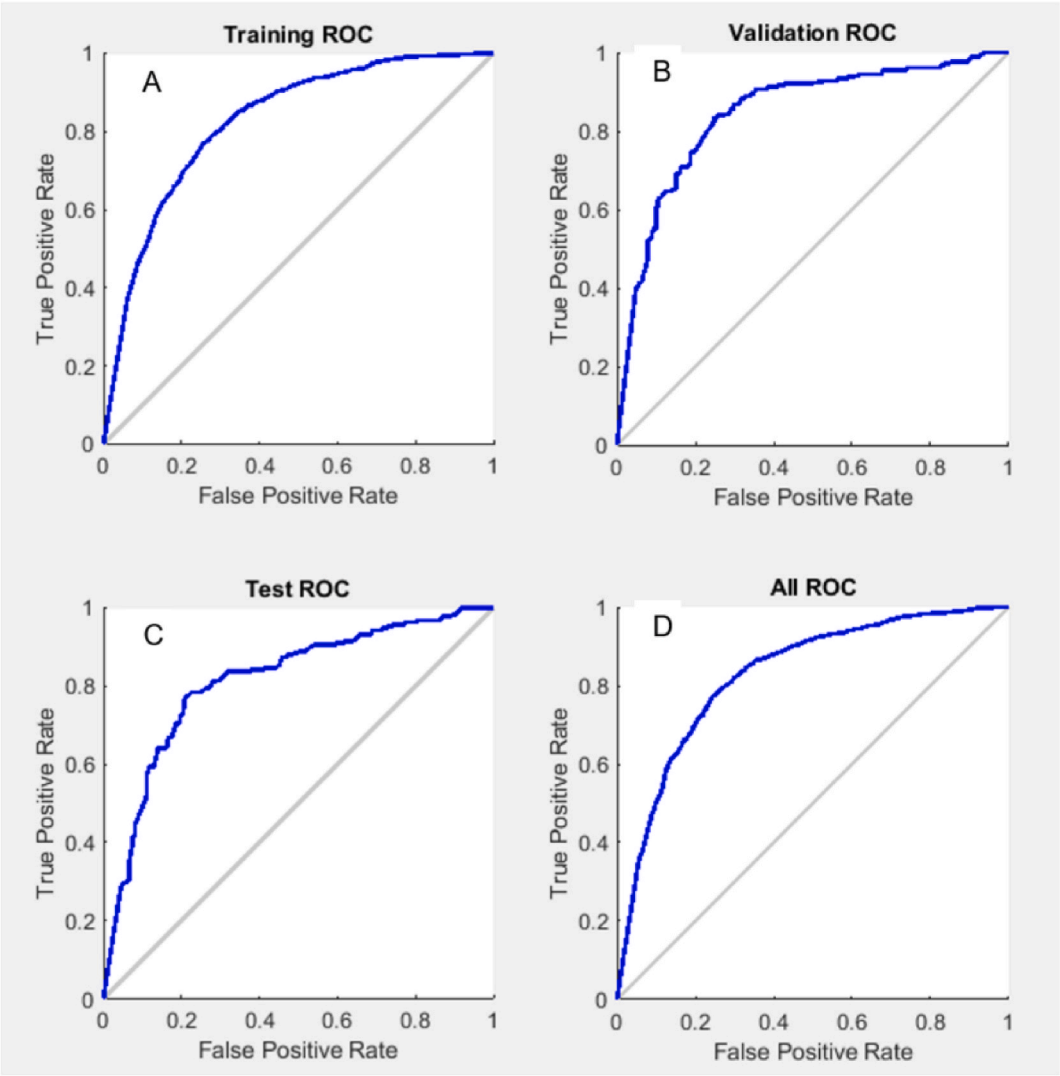


Fig. 6. BPNN model's ROC curve for each sample (a-d; total AUC: 0.827); (a) ROC of the training sample; (b) ROC of the validation sample; (c) ROC of the test sample; (d) ROC of the final sample.

BP neural network training and verification. The model featured a relatively high sensitivity (79.0%), a relatively high positive predictive value (78.9%), a total coincidence rate (π) of 75.9% and a specificity and negative predictive value above the 70% threshold. Thus, these results indicate a clinical significance of the constructed BPNN model. The model displays a high level of discrimination and could help users to easily screen, if cancer patients are vulnerable to distress. In sum, the BPNN model is slightly superior to our logistic regression model. However, our BPNN model with a relatively low number of 13 independent variables cannot reach its full potential. A higher quantity of independent variables would probably lead to more significant superiority compared to the conventional regression analysis.

Criticisms of BPNN point out, that it might not be suitable to manage real-world data and is best used in combination with other methods [25,28]. The present study provides evidence to the contrary. While logistic regression analysis is able to preliminary screen a great number of potential risk factors, BPNN has the advantage, by virtue of its simulation of the human brain and its flexibility, of a high level of discrimination and precise data fitting. Thus, BPNN is more suitable to accurately predict outcomes for the individual patient [36]. The integration of both methods might simplify the detection of risk factors for distress in patients with cancer.

Several limitations exist within our study. Firstly, the data utilized was derived from a single cancer care center (CCC) located in Switzerland, which may restrict the generalizability of our findings to other settings. However, it is worth noting, that our study encompassed various types of cancer and treatment options. Additionally, the procedures and instruments employed for distress screening, exhibit similarities across different countries worldwide [7]. Furthermore, record keeping of patient-related information in our study adhered to stringent institutional standards. As a result, the data's quality and the reliability of the obtained results are notably high. However, it is essential to acknowledge a significant limitation of our study, namely the underrepresentation of certain

types of cancer, including gynaecological, urological, and endocrinological malignant neoplasms (see Table 1). This is a possible explanation, why female sex did not predict distress in cancer patients. Interinstitutional differences might be responsible for low screening rates for certain cancer entities, as was already suspected by Ref. [29]. Staff, time and skilled worker shortage may be reasons for these differences [37,38]. Since the largest part of our patients were male (two-third), prospective studies may want to include a greater share of female patients. Furthermore, different DT cut-off scores for specific patient populations were assumed by previous studies, such as lower cut-off scores for newly diagnosed patients [39] or higher cut-off scores for women recently diagnosed with breast cancer [40], which complicates comparability between study findings. However, only one specific DT cut-off score is used in the hospital, where our study was conducted. Lastly, there does not exist any standard process regarding the construction of the BP neural network. The network desgin is solely based on individual expert knowledge and repetetive data simulation processes. Dealing with a great number of input variables is another issue, which can be solved by combining BPNN with traditional methods.

5. Conclusions

In this study, the largest part of patients with cancer reported to be distressed. Overall, we identified 13 distress predictors in cancer patients: emotional problems, such as fear and sadness, physical problems, such as pain and fatigue, and practical problems regarding work or school. Furthermore, this study provided proof of concept for a BPNN distress prediction model in cancer patients, which improves individual prediction accuracy compared to conventional methods. Our work sets the theoretical foundation for artificial intelligence assisted prediction of risk factors, which are associated with distress. The integration of BPNN distress prediction algorithms in the hospital information system might be one example. Because this study was conducted in an institution similar in structure, treatment processes and patient population to those of other national and international CCCs, the results are generalizable. Prospective studies should try to replicate and extend our results. Inspired by our results, finding more precise prevention and intervention strategies for cancer patients vulnerable to distress may be another desirable goal of studies.

Jan Ben Schulze: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Marc Dörner: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Moritz Philipp Günther: Conceived and designed the experiments; Analyzed and interpreted the data.

Roland von Känel: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Sebastian Euler: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data. </p>

Data availability statement

Data will be made available on request.

Additional information

No additional information is available for this paper.

Ethics approval and consent to participate

The study underwent a thorough review and received approval from the Ethics Committee of the State of Zurich, Switzerland (Ref.-No. BASEC-NR. 2020-00977). It is important to note that this study is retrospective in nature, and therefore, formal consent from participants is not required. The Ethics Committee that approved the study also waived the need for informed consent (Ethics Committee of the State of Zurich, Switzerland, BASEC NR. 2020-00977; June 2020).

The authors affirm that all procedures conducted in this study adhere to the ethical standards set by the relevant national and institutional committees for human experimentation. Furthermore, the study aligns with the principles outlined in the Helsinki Declaration of 1975, with its 2008 revision.

Consent for publication

Not applicable.

Data and Materials Availability

The datasets utilized and/or analyzed during the present study can be obtained from the corresponding author upon reasonable request.

Funding

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Declaration of competing interest

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

References

[1] National Comprehensive Cancer Network, NCCN Clinical Practice Guidelines in Oncology: Distress Management. V2, 2019.
[2] J. Zabora, K. Brintzenhofe Szoc, B. Curbow, C. Hooker, S. Piantadosi, The prevalence of psychological distress by cancer site, Psycho Oncol. 10 (1) (2001) 19–28.
[3] A. Mehnert, T.J. Hartung, M. Friedrich, S. Vehling, E. Brähler, M. Härter, M. Keller, H. Schulz, K. Wegscheider, J. Weis, U. Koch, H. Faller, One in two cancer patients is significantly distressed: prevalence and indicators of distress, Psycho Oncol. 27 (1) (2017) 75–82.
[4] G.-L. Wang, C.-T. Cheng, A.-C. Feng, S.-H. Hsu, Y.-C. Hou, C.-Y. Chiu, Prevalence, risk factors, and the desire for help of distressed newly diagnosed cancer patients: a large-sample study, Palliat. Support Care 15 (3) (2017) 295–304.
[5] A.J. Mitchell, Screening for cancer-related distress: when is implementation successful and when is it unsuccessful? Acta Oncol. 52 (2) (2013) 216–224.
[6] A. Meijer, M. Roseman, V.C. Delisle, K. Milette, B. Levis, A. Syamchandra, M.E. Stefanek, D.E. Stewart, P. de Jonge, J.C. Coyne, B.D. Thombs, Effects of screening for psychological distress on patient outcomes in cancer: a systematic review, J. Psychosom. Res. 75 (1) (2013) 1–17.
[7] K.K. Ownby, Use of the distress thermometer in clinical practice, J Adv Pract Oncol 10 (2) (2019) 175–179.
[8] B. Jacobsen Paul, A. Donovan Kristine, C. Trask Peter, B. Fleishman Stewart, Zabora James, Baker Frank, C. Holland Jimmie, Screening for psychologic distress in ambulatory cancer patients, Cancer 103 (2005) 1494–1502.
[9] Shim Eun-Jung, Yong-Wook Shin, Hong Jin Jeon, Bong-Jin Hahm, Distress and its correlates in Korean cancer patients: pilot use of the distress thermometer and the problem list, Psycho Oncol. 17 (2008) 548–555.
[10] T. Hegel Mark, Collins E. Dale, Kearing Stephen, L. Gillock Karen, P. Moore Caroline, A. Ahles Tim, Sensitivity and specificity of the Distress Thermometer for depression in newly diagnosed breast cancer patients, Psycho Oncol. 17 (2008) 556–560.
[11] C. Acquati, K. Kayser, Predictors of psychological distress among cancer patients receiving care at a safety-net institution: the role of younger age and psychosocial problems, Support. Care Cancer 25 (7) (2017) 2305–2312.
[12] J. Zabora, K. Brintzenhofe Szoc, B. Curbow, C. Hooker, S. Piantadosi, The prevalence of psychological distress by cancer site, Psycho Oncol. 10 (1) (2001) 19–28.
[13] L.E. Carlson, M. Angen, J. Cullum, E. Goodey, J. Koopmans, L. Lamont, J.H. MacRae, M. Martin, G. Pelletier, J. Robinson, J.S.A. Simpson, M. Speca, L. Tillotson, B.D. Bultz, High levels of untreated distress and fatigue in cancer patients, Br. J. Cancer 90 (2004) 2297–2304.
[14] Hurria Arti, Daneng Li, Kurt Hansen, Sujata Patil, Ravi Gupta, Nelson Christian, M. Lichtman Stuart, P. Tew William, Hamlin Paul, Zuckerman Enid, Gardes Jonathan, Limaye Sewanti, Lachs Mark, Eva Kelly, Distress in older patients with cancer, J. Clin. Oncol.: official journal of the American Society of Clinical Oncology 27 (2009) 4346–4351.
[15] D. Raphael, R. Frey, M. Gott, Distress in post-treatment hematological cancer survivors: prevalence and predictors, J. Psychosoc. Oncol. 38 (3) (2020) 328–342.
[16] D.N.N. Lo-Fo-Wong, H.C.J.M. de Haes, N.K. Aaronson, D.L. van Abbema, M.D. den Boer, M. van Hezewijk, M. Immin, A.A. Kaptein, M.B.E. Menke-Pluijmers, A. K.L. Reyners, N.S. Russell, M. Schriek, S. Sijtsema, G. van Tienhoven, M.A.G. Sprangers, Predictors of enduring clinical distress in women with breast cancer, Breast Cancer Res. Treat. 158 (3) (2016) 563–572.
[17] L. Jorgensen, B.S. Laursen, J.P. Garne, K.A. Sherman, M. Sogaard, Prevalence and predictors of distress in women taking part in surgical continuity of care for breast cancer: a cohort study, Eur. J. Oncol. Nurs. 22 (2016) 30–36.
[18] R.L. Johnson, C. Larson, L.L. Black, K.G. Doty, L. VanHoose, Significance of nonphysical predictors of distress in cancer survivors, Clin. J. Oncol. Nurs. 20 (5) (2016) E112–E117.
[19] Clover Kerrie Ann, Oldmeadow Christopher, Louise Nelson, Rogers Kerry, Alex J. Mitchell, Carter Gregory, Which items on the distress thermometer problem list are the most distressing? Support. Care Cancer: official journal of the Multinational Association of Supportive Care in Cancer 24 (2016) 4549–4557.
[20] J.-H. Park, M. Chun, Y.-S. Jung, S.H. Bae, Predictors of psychological distress trajectories in the first year after a breast cancer diagnosis, Asian Nurs. Res. 11 (4) (2017) 268–275.
[21] VanHoose Lisa, L. Black Lora, Doty Kimberly, Sabata Dory, Philip Twumasi-Ankrah, Sarah Taylor, Rhonda Johnson, An analysis of the distress thermometer problem list and distress in patients with cancer, Support. Care Cancer: official journal of the Multinational Association of Supportive Care in Cancer 23 (2015) 1225–1232.
[22] N. Kriegeskorte, T. Golan, Neural network models and deep learning, Curr. Biol. 29 (7) (2019) 231–236.
[23] P. Aruna, N. Puviarasan, B. Palaniappan, An investigation of neurofuzzy system in psychosomatic dirsorders, Expert Syst. Appl. 28 (4) (2005) 673–679.
[24] Y. Zou, Y. Shen, L. Shu, Y. Wang, F. Feng, K. Xu, Y. Ou, Y. Song, Y. Zhong, M. Wang, W. Liu, Artificial neural network to assist psychiatric diagnosis, Br. J. Psychol. 169 (1996) 64–67.
[25] S. Chattopadhyay, P. Kaur, F. Rabhi, U.R. Acharya, Neural network approaches to grade adult depression, J. Med. Syst. 36 (2012) 2803–2815.
[26] R. Fan, T. Hua, T. Shen, Z. Jiao, Q. Yue, B. Chen, Z. Xu, Identifying patients with major depressive disorder based on tryptophan hydroxylase-2 methylation using machine learning algorithms, Psychiatr. Res. 306 (2021), 114258.
[27] G.E. Davis, W.E. Lowell, G.L. Davis, A neural network that predicts psychiatric length of stay, M Comput. 10 (2) (1993) 87–92.
[28] J. Lyu, J. Zhan, BP neural network prediction model for suicide attempt among Chinese rural residents, J. Affect. Disord. (2019) 465–473.
[29] M.P. Günther, J. Kirchebner, J.B. Schulze, A. Götz, R. von Känel, S. Euler, Uncovering barriers to screening for distress in patients with cancer via machine learning, Journal of the Academy of Consultation-Liaison Psychiatry 63 (2022) 163–169.
[30] J.C. Holland, B.D. Bultz, The NCNN guideline for distress management: a case for making distress the sixth vital sign, J. Natl. Compr. Cancer Netw. 5 (2007) 3–7.
[31] L. Bertero, F. Massa, J. Metovic, R. Zanetti, I. Castellano, U. Ricardi, M. Papotti, P. Cassoni, Eighth edition of the UICC classification of malignant tumours: an overview of the changes in the pathological TNM classification criteria - what has changed and why? Virchows Arch. 472 (4) (2018) 519–531.
[32] M.E. Charlson, P. Pompei, K.L. Ales, C.R. MacKenzie, A new method of classifying prognostic comorbidity in longitudinal studies: development and validation, J. Chron. Dis. 40 (5) (1987) 373–383.
[33] H. Quan, B. Li, C.M. Couris, K. Fushimi, P. Graham, P. Hider, J.-M. Januel, V. Sundararajan, Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries, Am. J. Epidemiol. 173 (6) (2011) 676–682.
[34] D.H. Kim, H.C. Park, A. Cho, J. Kim, K.-S. Yun, J. Kim, Y.-K. Lee, Age-adjusted Charlson comorbidity index score is the best predictor for severe clinical outcome in the hospitalized patients with COVID-19 infection, Medicine (Baltim.) 100 (18) (2021), e25900.
[35] J. Xie, S. Ding, S. He, Y. Duan, K. Yi, J. Zhou, A prevalence study of psychosocial distress in adolescents and young adults with cancer, Cancer Nurs. 40 (2017) 217–223.
[36] J. Huan, J. Chen, BP neural networrk model for early diagnosis of Kawasaki disease, Biomed Eng Res 30 (4) (2011) 207–210.
[37] A.K. Knies, D.R. Jutagir, E. Ercolano, N. Pasacreta, M. Lazenby, R. McCorkle, Barriers and facilitators to implementing the commission on cancer's distress screening program standard, Palliat. Support Care 17 (2019) 253.

[38] E.A. Fradgley, E. Byrnes, K. McCarter, N. Rankin, B. Britton, K. Clover, G. Carter, D. Bellamy, C.L. Paul, A cross-sectional audit of current practices and areas for improvement of distress screening and management in Australian cancer services: is there a will and a way to improve? Support. Care Cancer 28 (2020) 249–259.
[39] A. Cutillo, E. O'Hea, S. Person, D. Lessard, T. Harralson, E. Boudreaux, The distress thermometer: cutoff points and clinical use, Oncol. Nurs. Forum 44 (2017) 329–336.
[40] F. Ploos van Amstel, J. Tol, K. Sessink, W.T. van der Graaf, J. Prins, P. Ottevanger, A specific distress cutoff score shortly after breast cancer diagnosis, Cancer Nurs. 40 (3) (2017) E35–E40.