Predicting compassion fatigue among psychological hotline counselors using machine learning techniques

Lin Zhang[1] · Tao Zhang[1] · Zhihong Ren[1] · Guangrong Jiang[1]

Abstract

During the outbreak of coronavirus disease 2019, psychological hotline counselors frequently address help-seekers' traumatic experiences from time to time, which possibly causes counselors' compassion fatigue. The present study aimed to explore the predictors of compassion fatigue among a high-risk population of psychological hotline counselors. Seven hundred and twelve psychological hotline counselors were recruited from the Mental Health Service Platform at Central China Normal University, Ministry of Education, then were asked to complete the questionnaires measuring compassion fatigue, trait empathy, social support, trait mindfulness, counselor's self-efficacy, humor, life meaning, and post-traumatic growth. A chi-square test was utilized to filter for the top-20 predictive variables. Machine learning techniques, including logistic regression, decision tree, random forest, k-nearest neighbor, support vector machine, and naïve Bayes were employed to predict compassion fatigue. The results showed that the most important predictors of compassion fatigue were meaning in life, counselors' self-efficacy, mindfulness, and empathy. Except for the decision tree, the rest machine learning techniques obtained good performance. Naïve Bayes presented the highest area under the receiver operating characteristic curve of 0.803. Random forest achieved the least classification error of 23.64, followed by Naïve Bayes with a classification error of 23.85. These findings support the potential application of machine learning techniques in the prediction of compassion fatigue.

Keywords Compassion fatigue · Hotline psychological counselor · Machine learning · COVID-19

Introduction

Coronavirus disease 2019 (COVID-19), which was declared a pandemic on 11 March 2020 by World Health Organization, has led to 2.85 million deaths as of 5 April 2021 (World Health Organization, 2021). Besides, the pandemic has generated huge psychological problems for the general population worldwide. Until the end of May 2020, the pooled prevalence of psychological stress, anxiety, depression, and posttraumatic stress symptoms caused by COVID-19 was 29.6%, 31.9%, 33.7%, and 23.9%, respectively (Salari et al., 2020). In addition to the general population, mental health professionals have faced unprecedented challenges as well. They need to deal with diverse and significant psychological problems of an increasing number of people each day. The heightened level of work pressure may affect their mental health and wellbeing (Zhang et al., 2021). In China, the pandemic also leads to physical and psychological consequences significantly with more than 90-thousand infections and more than 20% Chinese public suffering from mental distress by end of April 2020 (Ren et al., 2020).

Compassion fatigue (CF) is the negative aspect of providing care, which manifests exhaustion, frustration, anger, and depression typical of burnout, or a negative feeling driven by fear and work-related trauma (Stamm, 2010). According to Stamm (2010), compassion fatigue includes two components, burnout and secondary trauma, respectively. Professional helpers, such as nurses, social workers, psychotherapists who frequently or repeatedly witness or listen to a person's traumatic experience, are susceptible to CF (Figley, 2002). As an example, during the outbreak of COVID-19, psychological hotline counselors may experience CF when helping clients from time to time, as a majority of the clients were under severe stress and suffering from traumatic distress (Zhao et al., 2020). According to previous surveys, CF is prevalent

✉ Zhihong Ren
  ren@ccnu.edu.cn

[1] School of Psychology, Central China Normal University, Key
   Laboratory of Adolescent Cyberpsychology and Behavior, Ministry
   of Education, Key Laboratory of Human Development and Mental
   Health of Hubei Province, Wuhan, China

among mental health professionals, the prevalence of which can reach nearly 50% (Turgoose & Maddox, 2017). CF produces many consequences, including professional helpers' emotional and physical issues, high turnover and absenteeism of professional helpers, poor quality of patient/client care, and a decrease of patients' trust and confidence in their help providers (Sorenson et al., 2016).

The prevalence and adverse consequences of CF among mental health professionals raise important questions about the risk and protective factors of CF, which have been discussed in the etiological and multi-factor model of CF theoretically and investigated in many previous empirical studies (Figley, 2002; Turgoose & Maddox, 2017). The etiological and multi-factor model is the theoretical framework of the present study. According to this model, CF began with the empathy of psychological counselors to clients. Moreover, some factors, like disengagement and satisfaction with the efforts to help clients, would decrease the risk of CF, while other factors, like prolonged exposure, traumatic memories, and the degree of life disruptions, would increase the risk of CF (Figley, 2002).

Previous empirical studies not only provided evidence for the hypothesis of the theory but also shed light on some new influencing factors which were not considered in the theory. The influencing factors of CF can be summarized into three categories from the results of empirical studies, organizational, demographic, and personal factors. With regard to organizational factors, firstly, the caseload, which refers to the total number of clients handled in a period, is a common influence factor of CF. Previous studies found that caseload size had a positive relationship with emotional exhaustion (Kim et al., 2018). Besides, Boscarino et al. (2004) found caseloads with high percentages of trauma clients were correlated with an increased incidence of CF. Second, supervision, as a kind of support from clinical supervisors, has been proven to play a critical role in impeding the development of CF in mental health professionals (Bell et al., 2019). Last, one study found that work settings can influence the level of CF, for example, mental health professionals who worked in inpatient care settings experienced more CF than their counterparts in independent practice settings (Craig & Sprang, 2010).

Considering demographic factors, gender, age, marriage status, and education level have been found to associate with CF. Specifically, female mental health professionals were found to have a significantly higher risk of CF than males (Mangoulia et al., 2015). One study found that younger mental health professionals were more likely to report CF than older ones (Sprang et al., 2011). A survey after the September 11 terrorist attacks found that unmarried social workers had a higher incidence of CF compared to their married counterparts (Boscarino et al., 2004). Rudolph et al. (1997) found that mental health professionals with a master's degree had a higher risk of CF than their counterparts with a bachelor or doctoral degree.

Concerning personal factors, the influences of trait mindfulness, work experience, traumatic history, and empathy on CF have been examined. Trait mindfulness is considered as a strong protective factor against CF, and a strong negative association between the two was found among mental health professionals (Thompson et al., 2014). Regarding the association between work experience and CF, a small and negative correlation was found (Thompson et al., 2014), while one study found that CF increased with work years (Birck, 2001). These inconsistent findings can be explained by that more experienced mental health professionals are more likely to shoulder larger caseloads and challenging cases, but they have learned how to handle these challenges effectively with their experience (Turgoose & Maddox, 2017).

Traumatic history is one of the commonly studied risk factors for CF. Mental health professionals with previous trauma reported higher CF (Rossi et al., 2012). Empathy is considered as a prerequisite for CF according to the etiological model of CF (Figley, 2002). This model suggests that empathy is a risk factor for CF, which has been proven by other researchers as well. One study found that among mental health professionals, the level of CF increased as the level of empathy increased (MacRitchie & Leibowitz, 2010). Moreover, some studies further investigated the association between different dimensions of empathy and CF and found that personal distress was most strongly associated with CF compared to the other three dimensions, namely, empathic concern, fantasy, and perspective-taking (Thomas, 2013).

Additionally, other factors, such as social support, coping strategy, and posttraumatic growth (PTG), have been found to associate with CF among mental health professionals. One study on trauma workers found that a higher level of perceived social support predicted a lower level of CF (MacRitchie & Leibowitz, 2010). Concerning the coping strategy, a previous study found that the use of lighthearted humor was related to lower CF, while the use of maladaptive coping styles, such as self-criticism and giving up, were related to higher CF (Craun & Bourke, 2014). PTG showed an influence on CF, although the potential causal mechanisms, including the direction of the effects and the underlying moderators and mediators, were largely uncertain (Cosden et al., 2016).

Besides mental health professionals, many studies exploring the predictors of CF were conducted among other groups, such as nurses and clinicians. Meaning in life was found to correlate negatively with CF among nurses, which indicated that meaning in life could serve as a protective factor against the onset of CF (Mason, 2013). Additionally, context-specific self-efficacy, which refers to the belief about the ability to deal with job-related challenges, was found to correlate with CF among nurses (Wahlberg et al., 2016).

Taken together, previous research has demonstrated that it is possible to predict CF (Singh et al., 2020; Turgoose & Maddox, 2017), but it is unclear how important these predictors are and how well the non-linear predictive models perform. Moreover, due to the limitation of conventional statistical methods, most studies investigated only one or several predictors of CF and have not obtained the patterns of CF. Furthermore, although scientific psychology is aiming at explaining and predicting human behaviors, explanation (e.g., investigation of moderating or mediating variables) has been studied widely while prediction or classification (e.g., what's the accuracy for predicting or classifying the specific behavior) has been ignored for a long time (Yarkoni & Westfall, 2017). Machine learning is a good way to solve these problems.

Machine learning refers to a computational technique that enables computers to automatically learn from multidimensional data sets and deliver a solution. As a subset of machine learning, the classification algorithm identifies patterns that can differ among categories to assess various measurements and assign observations to these categories (Dwyer et al., 2018). The classification algorithm has several advantages in prediction compared with conventional statistical analysis. First, the classification algorithm is better at processing a large number of predictors and selecting variables that have more predictive power. Second, more reliable models can be obtained by using k-fold cross-validation. Moreover, non-linear models can be built by using classification algorithms, such as random forest and support vector machine (Yarkoni & Westfall, 2017).

With respect to the application in psychology, classification algorithms have been proven to be highly effective to predict participants' mental status. A recent systematic review summarized the application of machine learning in the area of mental health and suggested that the performance of classification algorithms was satisfied in the prediction of depression, suicide, job-related stress, bipolar disorder, mood, posttraumatic stress disorder, anxiety, substance abuse, and schizophrenia (Thieme et al., 2020). According to the "No Free Lunch Theorem", no single algorithm is consistently superior to the others (Wolpert & Macready, 1997). The performance of the classification algorithm is mainly dependent on the type of data and the features of the particular dataset. Therefore, in the present study, we implemented six classification algorithms and compared their performance based on the current dataset.

Overall, the research question of the present study was to find the profile (i.e., patterns) of psychological hotline counselors who were more likely to experience CF during the pandemic of COVID-19. Specifically, the primary purpose of the present study was to develop a model that can predict CF among psychological hotline counselors during the COVID-19 outbreak. Additionally, we aimed to compare several classification algorithms and identify which one could perform best using questionnaire data. Last, we aimed to investigate which feature patterns could show more predictive power for CF.

Methods

Participants

A total of 712 psychological hotline counselors were recruited from the Mental Health Service Platform at Central China Normal University, Ministry of Education (MOE-CCNU-MHSP). Their data were collected via online self-report questionnaires during the COVID-19 pandemic, between April 10th and 15th, 2020.

The present study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards, and it was approved by the Life Science Ethics Committee of Central China Normal University. Participation in the study was voluntary and anonymous, which means that participants could quit the study at any time without any disadvantage. Their data would be used only for research. The results of the research will be disseminated through publication in academic peer-reviewed journals and scientific conferences. After receiving a brief introduction to the study and the above explanation, all participants gave their consent to participate.

Predictors and Outcome

The outcome measure of CF among psychological hotline counselors was collected via the Chinese version of the Professional Quality of Life Scale (ProQoL) version 5 (Stamm, 2010; Zheng et al., 2013). The ProQoL is a 30-item, self-report scale, including three subscales, namely *compassion satisfaction* (e.g., "I get satisfaction from being able to help people"), *burnout* (e.g., "I am not as productive at work because I am losing sleep over traumatic experiences of a person I help"), and *secondary traumatic stress* (e.g., "I jump or am startled by unexpected sounds"). CF can be reflected by the burnout subscale and the secondary traumatic stress subscale. Participants were asked to indicate how frequently they experience the things describing in items on a 5-point Likert scale ("1 = Never", "5 = Very often"). Following the recommendation of Stamm (2010), a t-score of the burnout subscale ≥57 or t-score of the secondary traumatic stress subscale ≥57 was used to define CF in this study, while both subscale t-scores <57 can be defined as non-CF. In the present study, the internal consistencies for compassion satisfaction, burnout, and secondary traumatic stress were acceptable (Cronbach's α = 0.88, 0.72, 0.81, respectively).

Based on the investigations of previous studies on the risk or protective factors of compassion fatigue, a wide range of

predictors was considered in the present study. Specifically, predictors encompassed demographic variables (e.g., gender, age, educational level, marital status, workplace, professional qualification or certification, work years, prior work experience, prior traumatic experience, overall caseloads, number of traumatic cases, and frequency of supervision), trait empathy, social support, trait mindfulness, counselor's self-efficacy, humor, life meaning, and post-traumatic growth. A total of 185 predictors were initially collected and included in the research, and a full list can be found in Online source 1.

Trait empathy was measured using the Interpersonal Reactivity Index-Chineses Version (IRI-C) (Zhang et al., 2010). The IRI-C, which is adapted from the Interpersonal Reactivity Index (Davis, 1980), consists of 22 items on a 5-point Likert scale ("0= does not describe me well", "4= describes me very well"), measuring trait empathy on four separate subscales, namely, *perspective taking* (PT) (e.g., "I try to look at everybody's side of a disagreement before I make a decision"), *fantasy* (FS) (e.g., "I really get involved with the feelings of the characters in a novel"), *personal distress* (PD) (e.g., "In emergency situations, I feel apprehensive and ill-at-ease"), and *empathic concern* (EC) (e.g., "I often have tender, concerned feelings for people less fortunate than me"). High PT scores indicate a high tendency to adopt other person's perspective. High FS scores suggest a high tendency to identify emotionally with fictional characters in books, movies, and plays. High PD scores indicate a high tendency to experience personal distress when observing the suffering of others. High EC scores suggest a high tendency to have feelings of sympathy and concern for others. In the present study, the internal consistencies of the IRI-C and subscales were acceptable, with Cronbach alphas of 0.763 (overall), 0.710 (PT), 0.626 (FS), 0.828 (PD), and 0.569 (EC), respectively.

Social support was assessed by the Chinese version of the Multidimensional Scale of Perceived Social Support (MSPSS) (Huang et al., 1996). The MSPSS is a 12-item, 7-point Likert scale ranging from 1 (very strongly disagree) to 7 (very strongly agree). It consists of three subscales, namely, social *support from family* (e.g., "My family really tries to help me"), *friends* (e.g., "My friends really try to help me"), and *significant other* (e.g., "There is a special person who is around when I am in need"). Higher total scores indicate stronger social support. In the present study, the internal consistencies of the MSPSS and subscales were good, with Cronbach alphas of 0.933 (overall), 0.915 (family), 0.916 (friends), and 0.856 (significant other).

Trait mindfulness was measured using the Chinese version of the Mindful Attention Awareness Scale-Chinese version (MAAS) (Chen et al., 2012). The MAAS is a 15-item, one-dimension scale. Respondents were asked to rate how frequently they currently have experience from 1 (almost always) to 6 (almost never). High total scores indicate a higher level of trait mindfulness. In the present study, the internal consistency was good (Cronbach's $\alpha = 0.85$).

Counselor's self-efficacy was measured using the Chinese version of the Counselor Self-Efficacy Scale (CSES) (Gao, 2013; Melchert et al., 1996). The CSES is a 20-item, 5-point Likert scale ("1= agree strongly", "5= disagree strongly") assessing knowledge and skill competencies in the practice of counseling or psychotherapy. High total scores indicate a high degree of respondents' confidence in their counseling or psychotherapy abilities. In the present study, the internal consistency was good (Cronbach's $\alpha = 0.89$).

Humor was measured using the Chinese version of the six-item revision of the Sense of Humor Questionnaire (SHQ-6) (Chen et al., 2009; Svebak, 1996). The SHQ-6, which consists of three items from the dimension of meta-message sensitivity (M-dimension) and three items from the dimension of liking of humorous situations (L-dimension), is a 4-point Likert scale ranging from 1 (total agreement) to 4 (total disagreement). The Chinese version of SHQ-6 only supports the two-factor model, but not the one-factor model, which means that the total score of SHQ-6 is not applicable to Chinese participants. The scores of M-dimension and L-dimension have to be calculated separately (Chen et al., 2009). High scores of M-dimension indicate high sensitivity to humorous content and meta-messages. High scores of L-dimension suggest a positive attitude toward humorous people and situations. In this study, the internal consistencies of the M-dimension and the L-dimension were acceptable, with Cronbach alphas of 0.640 and 0.643, respectively.

Life meaning was measured using the Chinese version of the Meaning in Life Questionnaire (MLQ) (Liu & Gan, 2010; Steger et al., 2006). The MLQ is a 10-item, 7-point Likert scale ("1= absolutely untrue", "7= absolutely true"), measuring the *presence of meaning in life* (e.g., "I have a good sense of what makes my life meaningful"), and the *search for meaning in life* (e.g., "I am seeking a purpose or mission for my life"). High total scores indicate that participant has more meaning in life. In the present study, the internal consistencies of the MLQ, Presence subscale, and Search subscale were acceptable, with Cronbach alphas of 0.764, 0.768, and 0.919, respectively.

PTG was measured using the Chinese version of the Posttraumatic Growth Inventory (PTGI) (Tedeschi & Calhoun, 1996; Wang et al., 2011). Item 18 "I have a stronger religious faith" of the PTGI was deleted in the Chinese version because of the cultural incompatibility. The Chinese version of PTGI, which consists of five subscales including *relating to others* (e.g., "Knowing that I can count on people in times of trouble"), *new possibilities* (e.g., "I developed new interests"), *personal strength* (e.g., "A feeling of self-reliance"), *spiritual change* (e.g., "A better understanding of spiritual matters"), and *appreciation of life* (e.g., "An appreciation for the value of my own life"), is a 20-item, 6-point Likert scale ranging from 0 (I did not experience this change as a result of my crisis) to 5 (I experienced this change to a very great degree as a result of

my crisis). Higher total scores indicate more PTG. In the present study, the internal consistencies of the Chinese version of PTGI and its subscales were good, with Cronbach alphas of 0.974 (overall), 0.932 (relating to others), 0.897 (new possibilities), 0.889 (personal strength), 0.889 (spiritual change), and 0.878 (appreciation of life), respectively.

Data Preprocessing

First, multiple categorical variables were converted to binary code. Continuous and ordinal variables were not manipulated. In our dataset, categorical variables had no missing values, and each continuous variable had no more than 10% missing values which were imputed by means. Then, variables with no variance were deleted because they could not distinguish different categories. We kept only one variable among the highly correlated ones (i.e., their correlation coefficients were above or equal to 0.7, or below or equal to −0.7), in order to avoid collinearity problems (Dormann et al., 2013).

Statistical Analysis

Descriptive analysis was conducted by using IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp, Armonk, NY, USA). We also performed independent t-tests for continuous variables and chi-square tests for categorical variables to compare the difference in these variables between CF and non-CF groups.

Machine learning analyses were performed using Rapidminer 9.8. First, data were split into a training set (60% of the sample), a tuning set (10% of the sample), and a testing set (30% of the sample). Splitting data can help evaluate how the model will perform in a new dataset and alleviate overfitting. Stratified random sampling was used to ensure equal distribution of the CF between sets.

Then, we conducted several analyses in the training set. Specifically, continuous variables in the training set were standardized, followed by conducting a chi-square test to filter for the 20 most predictive variables for further model building. The rule of standardization and the selected variables were saved to be applied later to the tuning and the testing sets. Moreover, the smaller subset (CF) was oversampled employing a Synthetic Minority Over-sampling Technique (SMOTE) because of the imbalanced distribution of the CF/ non-CF (Chawla et al., 2002).

The performance of models was calculated and assessed in terms of the area under the receiver operating characteristic curve (AUC). Moreover, balanced accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated as well. Several machine learning techniques, including logistic regression, decision tree, random forest, k-nearest neighbor, support vector machine, and naïve Bayes, were utilized to build models and

finally to determine which one is the best fitting model for a testing set. A description of these machine learning techniques could be seen in Online source 2. Each technique trained models using 10-fold cross-validation. The model with the lowest classification error was tested using the tuning set for tuning parameters with an evolutionary computation approach. Once the performance of the tuning set was satisfactory after parameters tuning, the final models were applied to the testing set. 95% confidence intervals were calculated by bootstrapping 10,000 times.

Results

Descriptive Statistics

Table 1 shows the descriptive statistics of demographic and social characteristics. Among the current study sample, 230 psychological hotline counselors (32.3%) reported CF while the rest 482 (67.7%) did not. The p values from independent t-tests or chi-square tests for these variables were Bonferroni corrected for identifying significant differences between CF and non-CF groups. Significant differences were found in gender ($p < 0.001$), marital status ($p = 0.023$), work years ($p = 0.002$), academic qualification ($p = 0.001$), other qualifications ($p = 0.039$), experience on working as crisis psychological hotline counselors ($p = 0.046$), experience on working as not only general but also crisis psychological hotline counselors ($p = 0.001$), minutes per week they received for individual supervisions ($p = 0.032$).

Model Training

The performance measures (balanced accuracy, classification error, AUC, sensitivity, specificity, NPV, and PPV) on the training set (60% of the sample) without additional data nor parameter modifications were reported as the mean and standard deviation of the 10 folds, which can be seen in Table 2. The learning curve revealed that all models performed well on the training set. Random forest performed the best with the highest balanced accuracy (82.59 ± 5.23), AUC (17.41 ± 5.23), sensitivity (82.07 ± 5.09), specificity (83.10 ± 8.20), NPV (82.28 ± 5.00), and PPV (83.37 ± 7.09), and the least classification error (17.41 ± 5.23). The 20 most indicative variables determined by the chi-square statistic are presented in Table 3.

Model Tuning

The performance from the tuning set of the final models for each machine learning technique can be found in Table 2. The optimal parameters and the ranges we tried can be found in

Table 1 Descriptive statistics of demographic and social characteristics, with *p* values. Independent t-test for numeric variables, and chi-square test for categorical variables

Total (*N*=712)	Non-Compassion Fatigue	Compassion Fatigue	*p*-value
Gender (male/ female)	93/ 389	42/ 188	0.742
Age (mean±std.)	43.38±7.71	41.07±8.05	< 0.001
Marital status (married/ unmarried)	448/ 32	202/ 28	0.023
Work years	12.95±5.97	11.52±5.56	0.002
Education level			
High school or below	1 (0.2%)	0 (0.0%)	0.305
Junior college	8 (1.7%)	1 (0.4%)	
Bachelor	109 (22.6%)	44 (19.1%)	
Master or PhD	364 (75.5%)	185 (80.4%)	
Work unit			
Specialized psychiatric hospital	7 (1.5%)	5 (2.2%)	0.698
General hospital/other specialized hospital	7 (1.5%)	5 (2.2%)	0.698
College/research institute	366 (75.9%)	178(77.4%)	0.668
Primary and secondary school	45 (9.3%)	18 (7.8%)	0.507
Government organization	1 (0.2%)	1 (0.4%)	0.542
Company	4 (0.8%)	3 (1.3%)	0.687
Public security/justice sector	7 (1.5%)	3 (1.3%)	1.000
Troops	1 (0.2%)	0 (0.0%)	1.000
Non-profit civil society organization	10 (2.1%)	1 (0.4%)	0.182
Private psychological counselling organization	21 (4.4%)	13 (5.7%)	0.448
Others	13 (2.7%)	3 (1.3%)	0.241
Qualification			
Licenced psychiatrists	6 (1.2%)	5 (2.2%)	0.538
Licenced psychotherapists from Ministry of Health	20 (4.1%)	17 (7.4%)	0.068
Certificated psychological counselors by Chinese Psychological Society	84 (17.4%)	30 (13.0%)	0.136
Certificated psychological teachers	270 (56.0%)	111 (48.3%)	0.052
Licenced psychological counselors from Ministry of Human Resource	407 (84.4%)	198 (86.1%)	0.565
Overseas licenced psychiatrists/psychological counselors	16 (3.3%)	6 (2.6%)	0.608
Academic and systematic education and training on psychological counselling and treatment	282 (58.5%)	104 (45.2%)	0.001
Others	27 (5.6%)	5 (2.2%)	0.039
Experience in hotline psychological counseling			
Have worked as general psychological hotline counselors	152 (31.5%)	87 (37.8%)	0.096
Have worked as crisis psychological hotline counselors	10 (2.1%)	11 (4.8%)	0.046
Both	224 (46.5%)	76 (33.0%)	0.001
Neither	96 (19.9%)	56 (11.6%)	0.001
Experience in individual psychological counseling (hour)			
0–99	2 (0.4%)	5 (2.2%)	0.189
100–300	16 (3.3%)	8 (3.5%)	
301–500	56 (11.6%)	28 (12.2%)	
more than 500	408 (84.6%)	189 (82.2%)	
Have experienced trauma	196 (40.7%)	95 (41.3%)	0.871
Received total cases since worked on the platform	11.04±13.921	11.24±18.160	0.875
Received traumatic cases since worked on the platform	1.79±3.275	1.78±4.924	0.978
Received individual supervision since worked on the platform (minute/ week)			
0–29	275 (57.1%)	130 (56.5%)	0.032
30–59	76 (15.8%)	39 (17.0%)	
60–119	74 (15.4%)	33 (14.3%)	
120–179	20 (4.1%)	5 (2.2%)	

Table 1 (continued)

Total ($N = 712$)	Non-Compassion Fatigue	Compassion Fatigue	p-value
180–239	3 (0.6%)	9 (3.9%)	
more than 240	34 (7.1%)	14 (6.1%)	
Received group supervision since worked on the platform (minute/ week)			
0–29	14 (2.9%)	5 (2.2%)	0.851
30–59	18 (3.7%)	11 (4.8%)	
60–119	213 (44.2%)	108 (47.0%)	
120–179	132 (27.4%)	60 (26.1%)	
180–239	61 (12.7%)	30 (13.0%)	
more than 240	44 (9.1%)	16 (7.0%)	

online source 3. Except for the decision tree, the models performed similarly with an AUC above 0.7.

The logistic regression model performed with a classification error of 33.78 and an AUC of 0.715. Decision tree had a classification error of 28.06 and an AUC of 0.5. Random forest had a classification error of 24.65 and an AUC of 0.731. K-nearest neighbor had a classification error of 33.85 and an AUC of 0.757. Support vector machine had a classification error of 25.38 and an AUC of 0.762. Naïve Bayes had a classification error of 32.43 and an AUC of 0.789.

Table 2 Importance order of variables predicting compassion fatigue selected by chi-square test

Predictor	Importance
LM_9	1
M_8	0.947
M_10	0.937
M_12	0.918
SES_T	0.908
M_7	0.865
SEC_11	0.737
M_2	0.717
M_3	0.676
SEC_13	0.676
SEC_9	0.653
M_14	0.618
SEC_6	0.605
SEC_4	0.595
M_15	0.579
M_1	0.567
LM_5	0.551
Em_8	0.547
Em_21	0.541
Em_18	0.539

LM: meaning in life; M: mindfulness; SES: counselor's self-efficacy; Em: empathy

Model Testing

The final models were subsequently utilized to predict the CF in the test set. The performance from the testing set for each machine learning technique can be seen in Table 2. The learning curve revealed that final models performed well on the testing set except for the decision tree. Naïve Bayes had the highest AUC of 0.803. Random forest had the least classification error of 23.64, followed by Naïve Bayes (classification error of 23.85).

Discussion

Based on the questionnaire data from psychological hotline counselors working at MOE-CCNU-MHSP during the COVID-19 outbreak, the present study explored the machine learning techniques used to classify the CF among psychological hotline counselors, with a maximum AUC of 0.803 (using the Naïve Bayes) and a minimum classification error of 23.64 (using random forest model) on the testing set. When conducting classification algorithms, there are several potential underlying issues, such as category imbalance and model overfitting. For solving the category imbalance, SMOTE and stratified random sampling were utilized. For avoiding model overfitting, we utilized 10-fold cross-validation to train models and split the dataset into the training, tuning, and testing sets. Our results demonstrated that except for the decision tree, the performance (i.e., balanced accuracy, classification error, AUC, sensitivity, specificity, NPV, and PPV) of the other classification algorithms on the training, tuning, and testing sets was similar, which indicates that models are appropriate to the dataset without overfitting or underfitting.

Concerning the predictors of CF, the present study revealed some valuable findings. We found top-20 predictors out of a set of 185 variables, which included three empathy-related predictors, nine mindfulness-related predictors, six counselor's self-efficacy-related predictors, and two meaning-

Table 3 Classification performance of machine learning techniques on the training, tuning, and testing set

Performance on the training-set

Learner	Balanced accurancy (%)	Classification error (%)	AUC	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)
Logistic Regression	76.21±5.56	23.79±5.56	0.854±0.047	75.86±4.60	75.86±4.60	76.76±7.03	75.94±4.65
Decision Tree	77.76±3.30	22.24±3.30	0.785±0.047	79.31±6.08	79.31±6.08	78.91±4.71	79.31±6.08
Random Forest	82.59±5.23	17.41±5.23	0.906±0.050	82.07±5.09	83.10±8.20	82.28±5.00	83.37±7.09
k-Nearest Neighbor	77.07±5.86	22.93±5.86	0.876±0.047	91.38±4.67	62.76±13.09	88.59±5.94	71.67±6.48
Support Vector Machine	76.38±7.09	23.62±7.09	0.854±0.048	78.28±7.63	74.48±8.63	77.48±7.31	75.59±7.12
Naive Bayes	75.00±5.41	25.00±5.41	0.841±0.053	72.76±6.39	77.24±9.78	74.03±5.00	76.86±7.81

Performance on the tuning-set

Learner	Balanced accurancy (%)	Classification error (%)	AUC	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)
Logistic Regression	66.22 (56.34–74.65)	33.78 (25.35–43.66)	0.715 (0.591–0.827)	54.96 (35.71–73.33)	70.63 (60.00–80.77)	80.08 (70.00–89.58)	42.18 (26.09–58.62)
Decision Tree	71.94 (63.38–80.28)	28.06 (19.72–36.62)	0.5	0	100	71.94 (63.38–80.28)	–
Random Forest	75.35 (66.20–83.10)	24.65 (16.90–33.80)	0.731 (0.610–0.840)	47.54 (28.57–66.67)	86.23 (77.78–94.00)	80.79 (71.43–89.09)	57.45 (36.36–77.78)
k-Nearest Neighbor	66.15 (56.34–74.65)	33.85 (25.35–43.66)	0.757 (0.649–0.855)	69.88 (52.38–86.36)	64.70 (53.19–75.51)	84.63 (75.00–93.75)	43.58 (29.03–58.33)
Support Vector Machine	74.62 (66.20–83.10)	25.38 (16.90–33.80)	0.762 (0.656–0.856)	59.93 (41.18–77.78)	80.36 (70.83–89.13)	83.72 (74.51–92.00)	54.34 (36.36–72.22)
Naive Bayes	67.57 (57.75–76.06)	32.43 (23.94–42.25)	0.789 (0.694–0.873)	54.77 (35.71–73.33)	72.58 (61.90–82.69)	80.45 (70.59–89.80)	43.80 (27.27–60.87)

Performance on the testing-set

Learner	Balanced accurancy (%)	Classification error (%)	AUC	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)
Logistic Regression	73.81 (68.69–78.5)	26.19 (21.50–31.31)	0.794 (0.735–0.848)	65.73 (56.34–74.71)	77.99 (72.08–83.67)	81.47 (75.89–86.86)	60.72 (51.47–69.62)
Decision Tree	65.90 (60.75–71.03)	34.10 (28.97–39.25)	0.5	0	100	65.90 (60.75–71.03)	–
Random Forest	76.36 (71.50–81.31)	23.64 (18.69–28.50)	0.769 (0.706–0.829)	60.24 (50.72–69.57)	84.70 (79.58–89.58)	80.45 (75.00–85.71)	67.09 (57.41–76.62)
k-Nearest Neighbor	66.79 (61.68–71.96)	33.21 (28.04–38.32)	0.777 (0.716–0.837)	75.35 (67.07–83.33)	62.36 (55.56–68.97)	83.01 (76.70–88.89)	50.88 (42.99–58.62)
Support Vector Machine	73.34 (68.22–78.50)	26.66 (21.50–31.78)	0.796 (0.739–0.851)	67.10 (57.89–76.06)	76.57 (70.54–82.39)	81.81 (76.06–87.22)	59.71 (50.70–68.48)
Naive Bayes	76.15 (71.50–80.84)	23.85 (19.16–28.50)	0.803 (0.744–0.857)	71.24 (62.34–79.75)	78.68 (72.80–84.25)	84.09 (78.69–89.23)	63.37 (54.55–71.95)

AUC: area under the curve (level of discrimination); NPV: negative predictive value; PPV: positive predictive value; CI: confidence interval

related predictors. In line with previous studies, empathy and mindfulness are strong predictors of CF among mental health professionals (Turgoose & Maddox, 2017). To the best of our knowledge, the association between CF and meaning in life was only examined among nursing students (Mason, 2013), and the association between CF and job-related self-efficacy was only examined among nurses (Wahlberg et al., 2016). The present study examined these associations in a new sample, namely, mental health professionals, which increases the generalisability of the findings.

In the present study, the highest-ranking predictor of CF was the meaning in life. The relationship between CF and meaning in life was strong, which is consistent with the finding of a previous study (Mason, 2013). Mason (2013) stated that meaning was a pervasive potential, which can bring about a paradigm shift in the face of the darkness of life. Participants who know the meaning in life are more able to cope with different stressors and work without CF even in a stressful environment.

The present study found that the counselor's self-efficacy was a strong predictor of CF among mental health professionals, which could be supported by previous studies investigating other professional samples (Wahlberg et al., 2016). One implication of the finding is that psychological counselors need to know the limitations of their professional boundaries, besides, it is necessary to enhance their professional self-efficacy by training and supervision.

Consistent with the previous study (Thompson et al., 2014), mindfulness was a strong predictor of CF among mental health professionals in this study. Mindfulness is the ability to direct and maintain attention to the present moment without judgment. A less mindful psychological counselor would be less aware of his/her own negative emotions and be more easily to repress his/her emotional problems, which may more easily lead to CF.

The present study found that empathy was a strong predictor of CF, which can be supported by the etiological model of CF (Figley, 2002). This model assumes the centrality of empathy. Specifically, on the basis of the empathic ability of psychological counselors, CF first begins with the counselors' exposures to their clients/patients, then is motivated by their responses to clients/patients in need (i.e., empathic concern) and their efforts to reduce the suffering of clients/patients (i.e., empathic response). Additionally, we found all three empathy-related predictors belonged to the personal distress dimension, which indicates a strong association between personal distress and CF.

The meanings of the findings can be summarized as follows. First, this research found the predictive patterns for CF of psychological hotline counselors during the pandemic. Specifically, a counselor who has low levels of life meaning, professional self-efficacy, trait mindfulness but a high level of self-oriented empathy is more likely to be CF during the

pandemic. Second, the present study compared the performance of algorithms for the classification of CF and found that Naïve Bayes and random forest performed better than the conventional algorithm (i.e., logistic regression), which can enlighten psychologists to address conventional but unsolved problems using innovative techniques.

The present study has the following strengths. First, we considered a wide variety of potential predictors associated with CF and investigated not only the predictors that were examined among mental health professionals but also those which were associated with CF among other samples. Second, to the best of our knowledge, this is the first study to classify CF by using machine learning techniques. Machine learning techniques usually obtain better accuracies compared to conventional statistical analyses and can fit the data by using non-linear models, which breaks the limitations of the linear model. Third, the predictors in the present study were measured by questionnaires, which could be easily obtained by self-report at a low cost. This finding indicates that it is not necessary to collect data at a high cost for successfully classifying a CF model. Last but not the least, the present study emphasized classification and prediction rather than explanation in scientific psychology. That is, the present study was to obtain a compassion fatigue profile (patterns) of psychological hotline counselors during the pandemic, rather than explain the underlying mechanism (e.g., investigation of moderating or mediating variables).

Theoretical and Empirical Implications

It is worth noting that this study has some important implications. First, the theoretical implication of this study is to expand and deepen the understanding of the etiological model of CF. The model hypothesized that empathy was the prerequisite for CF (Figley, 2002). Our finding indicated that rather than stating empathy as the prerequisite for CF, it would be better to claim that self-oriented empathy, namely personal distress, is the prerequisite for CF. Second, this study reveals that meaning in life, trait mindfulness, trait empathy, and counselor's self-efficacy have more predictive power for CF. As a result, the empirical implication is that the intervention programs aimed to decrease CF should consider the potential roles of the four aspects in the future.

Limitations

Although the study has successfully demonstrated that machine learning methods are powerful tools for CF classification, we have to admit that there are several limitations in this study. First, the cross-sectional design of the study cannot acquire the time sequence of these variables. Although some

variables, such as trait empathy and trait mindfulness, can be assumed as predictors logically, it is hard to ascertain the causal relationship from a statistical perspective, therefore, longitudinal studies are needed to further examine the causality of these variables in future research. Second, we did not measure the impact of the pandemic on the psychological hotline counselors directly although all data were collected during the pandemic. It is necessary to consider in future studies that participants might have been personally affected by coronavirus or might have lost a closed one due to the pandemic. Moreover, data from other sources, such as neurobiological data, were not considered in the present study. Besides, we did not explore some potential predictors, such as self-compassion, which should be involved in analysis in future studies. Therefore, we recommend integrating different data sources with the application of technologies and computing methods and involving some potential predictors into the model of CF classification in order to develop a better classification accuracy in future study. Last, machine learning values classification accuracy over interpretability (Yarkoni & Westfall, 2017). Therefore, it is difficult to explore the underlying mechanisms of CF through this method.

Conclusion

This study highlights the application of machine learning techniques in CF prediction. The most important predictors of CF, namely meaning in life, trait mindfulness, trait empathy, and counselor's self-efficacy were found by using machine learning techniques. The finding contributes to our understanding of the etiological model of CF and may help us improve this theory. Moreover, the finding can help develop interventions for decreasing CF and further improving psychological hotline counselors' professional quality of life.

Declarations

References

Bell, S., Hopkin, G., & Forrester, A. (2019). Exposure to traumatic events and the experience of burnout, compassion fatigue and compassion satisfaction among prison mental health staff: An exploratory survey. *Issues in Mental Health Nursing, 40*(4), 304–309. https://doi.org/10.1080/01612840.2018.1534911.

Birck, A. (2001). Secondary traumatization and burnout in professionals working with torture survivors. *Traumatology, 7*(2), 85–90. https://doi.org/10.1177/153476560100700203.

Boscarino, J. A., Figley, C. R., & Adams, R. E. (2004). Compassion fatigue following the September 11 terrorist attacks: A study of secondary trauma among new York City social workers. *International Journal of Emergency Mental Health, 6*(2), 57–66.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*(1), 321–357. https://doi.org/10.1613/jair.953.

Chen, Wu, W., & Liu, Y. (2009). Preliminary test of the sense of humor questionnaire (SHQ-6) in China. *Journal of Campus Life & Mental Health, 7*(05), 305–307.

Chen, C., Zhou, H. R., & Jia, Y. (2012). Revision of mindful attention awareness scale (MAAS). *Chinese Journal of Clinical Psychology, 20*(02), 148–151. https://doi.org/10.16128/j.cnki.1005-3611.2012.02.024.

Cosden, M., Sanford, A., Koch, L. M., & Lepore, C. E. (2016). Vicarious trauma and vicarious posttraumatic growth among substance abuse treatment providers. *Substance Abuse, 37*(4), 619–624.

Craig, C. D., & Sprang, G. (2010). Compassion satisfaction, compassion fatigue, and burnout in a national sample of trauma treatment therapists. *Anxiety Stress and Coping, 23*(3), 319–339. https://doi.org/10.1080/10615800903085818.

Craun, S. W., & Bourke, M. L. (2014). The use of humor to cope with secondary traumatic stress. [article]. *Journal of Child Sexual Abuse, 23*(7), 840–852. https://doi.org/10.1080/10538712.2014.949395.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Journal of Personality and Social Psychology, 10*(85). https://doi.org/10.4324/9780429493898-3.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology, 14*, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816045037.

Figley, C. R. (2002). Compassion fatigue: Psychotherapists' chronic lack of self care. *Journal of Clinical Psychology, 58*(11), 1433–1441. https://doi.org/10.1002/jclp.10090.

Gao, Z. (2013). The characteristics and relationship of college counselor's self-efficacy and job burnout. Master, Hangzhou Dianzi University,

Huang, L., Jiang, Q., & Ren, W. (1996). Coping style, social support, and psychosomatic symptoms in patients with cancer. *Chinese Mental Health Journal, 10*(04), 160–161.

Kim, J. J., Brookman-Frazee, L., Gellatly, R., Stadnick, N., Barnett, M. L., & Lau, A. S. (2018). Predictors of burnout among community therapists in the sustainment phase of a system-driven implementation of multiple evidence-based practices in children's mental health. *Professional Psychology-Research and Practice, 49*(2), 132–141. https://doi.org/10.1037/pro0000182.

Liu, S., & Gan, Y. (2010). Reliability and validity of the Chinese version of the meaning in life questionnaire. *Chinese Mental Health Journal, 24*(06), 478–482, 1000-6729(2010)006-0478-05.

MacRitchie, V., & Leibowitz, S. (2010). Secondary traumatic stress, level of exposure, empathy and social support in trauma workers. *South Africa Journal of Psychology, 40*(2), 149–158. https://doi.org/10.1177/008124631004000204.

Mangoulia, P., Koukia, E., Alevizopoulos, G., Fildissis, G., & Katostaras, T. (2015). Prevalence of secondary traumatic stress among psychiatric nurses in Greece. *Archives of Psychiatric Nursing, 29*(5), 333–338. https://doi.org/10.1016/j.apnu.2015.06.001.

Mason, H. D. (2013). Relationship between meaning and professional quality of life among nursing students. *International Forum for Logotherapy, 36*(1), 25–31.

Melchert, T. P., Hays, V. L., Wiljanen, L. M., & Kolocek, A. K. (1996). Testing models of counselor development with a measure of counseling self-efficacy. *Journal of Counseling & Development, 74*(6), 640–644. https://doi.org/10.1002/j.1556-6676.1996.tb02304.x.

Ren, X., Huang, W., Pan, H., Huang, T., Wang, X., & Ma, Y. (2020). Mental health during the COVID-19 outbreak in China: A meta-analysis. *Psychiatric Quarterly, 91*(4), 1033–1045. https://doi.org/10.1007/s11126-020-09796-5.

Rossi, A., Cetrano, G., Pertile, R., Rabbi, L., Donisi, V., Grigoletti, L., Curtolo, C., Tansella, M., Thornicroft, G., & Amaddeo, F. (2012). Burnout, compassion fatigue, and compassion satisfaction among staff in community-based mental health services. *Psychiatry Research, 200*(2–3), 933–938. https://doi.org/10.1016/j.psychres.2012.07.029.

Rudolph, J., Stamm, B. H., & Stamm, H. (1997) Compassion fatigue: A concern for mental health policy, providers, and administration. In *Poster at the 13th Annual Meeting of the International Society for Traumatic Stress Studies, Montreal, PQ, CA.*

Salari, N., Hosseinian-Far, A., Jalali, R., Vaisi-Raygani, A., Rasoulpoor, S., Mohammadi, M., Rasoulpoor, S., & Khaledi-Paveh, B. (2020). Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: A systematic review and meta-analysis. *Globalization and Health, 16*(1), 57. https://doi.org/10.1186/s12992-020-00589-w.

Singh, J., Karanika-Murray, M., Baguley, T., & Hudson, J. (2020). A systematic review of job demands and resources associated with compassion fatigue in mental health professionals. *International Journal of Environmental Research and Public Health, 17*(19). https://doi.org/10.3390/ijerph17196987.

Sorenson, C., Bolick, B., Wright, K., & Hamilton, R. (2016). Understanding compassion fatigue in healthcare providers: A review of current literature. *Journal of Nursing Scholarship, 48*(5), 456–465. https://doi.org/10.1111/jnu.12229.

Sprang, G., Craig, C., & Clark, J. (2011). Secondary traumatic stress and burnout in child welfare workers: A comparative analysis of occupational distress across professional groups. *Child Welfare, 90*(6), 149–168.

Stamm, B. H. (2010). The Concise ProQOL Manual: The concise manual for the Professional Quality of Life Scale , 2 nd Edition.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53*(1), 80–93. https://doi.org/10.1037/0022-0167.53.1.80.

Svebak, S. (1996). The development of the sense of humor questionnaire: From SHQ to SHQ-6. *Humor, 9*(3–4), 341.

Tedeschi, R. G., & Calhoun, L. G. (1996). The posttraumatic growth inventory: Measuring the positive legacy of trauma. *Journal of Traumatic Stress, 9*(3), 455–471. https://doi.org/10.1007/BF02103658.

Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. *Acm Transactions on Computer-Human Interaction, 27*(5), 1–53. https://doi.org/10.1145/3398069.

Thomas, J. (2013). Association of personal distress with burnout, compassion fatigue, and compassion satisfaction among clinical social workers. *Journal of Social Service Research, 39*(3), 365–379. https://doi.org/10.1080/01488376.2013.771596.

Thompson, I. A., Amatea, E. S., & Thompson, E. S. (2014). Personal and contextual predictors of mental health counselors' compassion fatigue and burnout. *Journal of Mental Health Counseling, 36*(1), 58–77. https://doi.org/10.17744/mehc.36.1.p61m73373m4617r3.

Turgoose, D., & Maddox, L. (2017). Predictors of compassion fatigue in mental health professionals: A narrative review. *Traumatology, 23*(2), 172–185. https://doi.org/10.1037/trm0000116.

Wahlberg, L., Nirenberg, A., & Capezuti, E. (2016). Distress and coping self-efficacy in inpatient oncology nurses. *Oncology Nursing Forum, 43*(6), 738–746. https://doi.org/10.1188/16.Onf.738-746.

Wang, J., Chen, Y., Wang, Y., & Liu, X. (2011). Revision of the post-traumatic growth inventory and testing its reliability and validity. *Journal of Nursing Science, 26*(14), 26–28.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*(1), 67–82. https://doi.org/10.1109/4235.585893.

World Health Organization, W. H. (2021). WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int/. Accessed Apr 6, 2021.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393.

Zhang, F., Dong, Y., Wang, K., Zhan, Z., & Xie, L. (2010). Reliability and validity of the Chinese version of the interpersonal reactivity

index-C. *Chinese Journal of Clinical Psychology, 18*(02), 155–157. https://doi.org/10.16128/j.cnki.1005-3611.2010.02.019.

Zhang, L., Ren, Z., Jiang, G., Hazer-Rau, D., Zhao, C., Shi, C., Lai L., Yan Y. (2021). Self-oriented empathy and compassion fatigue: The serial mediation of dispositional mindfulness and counselor's self-efficacy. *Frontiers in Psychology,* 11, https://doi.org/10.3389/fpsyg.2020.613908.

Zhao, C., Shi, C., Zhang, L., Zhai, Z., Ren, Z., Lin, X., & Jiang, G. (2020). Establishment of online platform for psychological assistance during a public health emergency. *International Journal of Mental Health Promotion, 22*(3), 123–132. https://doi.org/10.32604/IJMHP.2020.011077.

Zheng, X., Yang, M., Gao, W., & Chen, F. (2013). The Chinese version professional quality of life scale: Testing of reliability and validity in nurses. *Journal of Nursing Science, 28*(5), 13–15. https://doi.org/10.3870/hlxzz.2013.05.013.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.