

Article

Screening of Mood Symptoms Using MMPI-2-RF Scales: An Application of Machine Learning Techniques

Sunhae Kim ¹, Hye-Kyung Lee ² and Kounseok Lee ^{1,*}

¹ Department of Psychiatry, Hanyang University Medical Center, Seoul 04763, Korea; sunhk0906@hanyang.ac.kr

² Department of Nursing, College of Nursing and Health, Kongju National University, Kognju 32588, Korea; hkleee@kongju.ac.kr

* Correspondence: dual@hanyang.ac.kr; Tel.: +82-2-2290-8423; Fax: +82-2-2298-2055

Abstract: (1) Background: The MMPI-2-RF is the most widely used and most researched test among the tools for assessing psychopathology, and previous studies have established its validity. Mood disorders are the most common mental disorders worldwide; they present difficulties in early detection, go undiagnosed in many cases, and have a poor prognosis. (2) Methods: We analyzed a total of 8645 participants. We used the PHQ-9 to evaluate depressive symptoms and the MDQ to evaluate hypomanic symptoms. We used the 10 MMPI-2 Restructured Form scales and 23 Specific Problems scales for the MMPI-2-RF as predictors. We performed machine learning analysis using the k-nearest neighbor classification, linear discriminant analysis, and random forest classification. (3) Results: Through the machine learning technique, depressive symptoms were predicted with an AUC of 0.634–0.767, and the corresponding value range for hypomanic symptoms was 0.770–0.840. When using RCd to predict depressive symptoms, the AUC was 0.807, but this value was 0.840 when using linear discriminant classification. When predicting hypomanic symptoms with RC9, the AUC was 0.704, but this value was 0.767 when using the linear discriminant method. (4) Conclusions: Using machine learning analysis, we defined that participants' mood symptoms could be classified and predicted better than when using the Restructured Clinical scales.

Keywords: Minnesota Multiphasic Personality Inventory; depression; bipolar disorders; risk factors; machine learning; Patient Health Questionnaire; mood disorder questionnaire



Citation: Kim, S.; Lee, H.-K.; Lee, K. Screening of Mood Symptoms Using MMPI-2-RF Scales: An Application of Machine Learning Techniques. *J. Pers. Med.* **2021**, *11*, 812. <https://doi.org/10.3390/jpm11080812>

Academic Editor: Moon-Soo Lee

Received: 17 July 2021

Accepted: 19 August 2021

Published: 20 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) was developed to discriminate and diagnose commonly occurring mental disorders. The MMPI-2 is the most widely used tool in personnel screening for occupations that require appropriate psychological coordination and responsibility, such as police officers, firefighters, air traffic controllers, and flight attendants [1–6]. The reason for the widespread use of the MMPI-2 in personnel review is its relatively easy management, objective scoring and interpretation procedures, and validity scales of clinically relevant symptomatic behaviors obtained from diverse populations [7].

As such, the MMPI-2 has been the subject of various validity studies and has the advantage of being able to use validity scales to detect response bias not only in psychiatry, but also in forensic evaluation and daily screening tests [8]. Additionally, the use of the MMPI-2 in the differential diagnosis of psychiatric disorders contributes to large-scale research-based evidence of its validity [9].

Mood disorders, including major depressive disorder (MDD) and bipolar disorder, are among the most common mental disorders [10]. They cause social dysfunction due to frequent relapses and represent major causes of suicide [11–13].

Many who suffer from mood disorders make repeated primary care visits without receiving an accurate diagnosis. Among low-income individuals, patients with mood

disorders are evaluated by an average of eight physicians before receiving a correct diagnosis [14]. Moreover, symptoms began at least 10 years before the average patient was diagnosed with a mood disorder [15]. Because primary care detects only 30–50% of mood disorders, patients with these disorders often remain untreated [16–20], which can lead to detrimental outcomes such as mood instability or deterioration, and even suicide [19,21].

Prevention of mood disorders may be possible through early screening and intervention [22–25]; therefore, accurate diagnosis of mood disorders is vital in providing appropriate treatment [26–29]. Many studies have demonstrated that early diagnosis and treatment of mood disorders improves outcomes dramatically [30–32], and efficient screening tools allow both mental health care providers and researchers to identify and assist those most likely to benefit from a full psychiatric assessment. This can also help streamline treatment resources.

However, the early detection of mood disorders also entails the necessity of discriminating between depressive and bipolar disorders. Early detection of bipolar patients is difficult because they spend much more time in the depressive phase than in the manic or hypomanic phase [33]. Individuals with bipolar disorder are more likely to seek help for depression than for hypomania or mania because the former's symptoms tend to be more distressing [20]. This tendency can lead to misdiagnosis of bipolar disorder as depressive disorder. Such misdiagnosis and subsequent inappropriate treatment (e.g., antidepressant-induced manic switch) can lead to detrimental outcomes such as increased suicide risk and mood instability [19,21]. These outcomes mirror symptoms of depressive disorder, as depression often also entails great personal suffering, increased risk of suicide, and social dysfunction [14,34–39].

Late diagnosis of these mood disorders can worsen the prognosis [40]. As treatment options for mood disorders are currently available in primary care settings but rarely used in a timely manner due to late diagnosis, new strategies to improve mood disorder detection in primary care settings are urgently needed [41]. Because accurate diagnosis of mood disorders is important for providing adequate treatment and early diagnosis and treatment yield much better outcomes [30–32], efficient screening tools, such as full psychiatric assessments, can benefit both mental health care providers and researchers [26–29]. Efficient screening tools can identify and assist those who need treatment most and streamline treatment resources.

However, despite recognizing the importance of early diagnosis, diagnosis is still unclear. Furthermore, despite significant progress over the last 100 years in clinical psychology and psychiatry [42–44], prognosis is still uncertain [45]. In psychiatry, machine learning (ML) research has been conducted for more than 20 years [46], and the advances in algorithms and computing power over the past decade have enabled even more precise research. ML is useful when analyzing multiple predictors, especially when there are nonlinear observations and interaction terms that are difficult to conceptualize theoretically and model in practice. One Danish study used ML to determine predictors of suicide using patient datasets from national medical registries [47]. This study highlights the importance of mental illness as a risk factor for suicide and reveals that ML-based predictive models can improve the accuracy and usefulness of tools developed for specific diagnostic groups.

ML-based predictive models are becoming increasingly popular due to their ability to combine large amounts of data into a single model and their self-evaluation of predictive value for previously unseen patients. Specifically, researchers have used ML methods successfully to predict the persistence, duration, and severity of major depressive disorder [48], as well as treatment responses [49], suicide attempts [50], and first onset of major depressive episodes in US soldiers [51,52]. ML methods are especially useful in fields where evaluation is complex, time-consuming, and expensive, such as when a differential diagnosis is required because the initial diagnosis is unclear [53]. In this study, we explore the MMPI-2's effectiveness in screening for mood disorders and whether it can be improved through ML techniques with predictive efficiency.

2. Materials and Methods

2.1. Participants

This study used a dataset from a survey conducted by Kongju National University across South Korea [54]. The researchers collected data from 8882 total participants, excluding from analysis 237 responses from those who did not complete the questionnaire or did not meet the requirements for the MMPI-2 validity scale. Additionally, participants who completed the full MMPI-2-RF scales, PHQ, and MDQ scales were included (response rate = 97.33%). The participants were healthy university students, 18.94 years old (SD = 1.64), and 4274 of them were female. All participants provided written informed consent, and the researchers assured the anonymity of the results and that the data would be used strictly for research purposes. This study was approved by the Research Ethics Committee of Kongju National University.

2.2. Psychopathology Assessment Method

2.2.1. MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form)

The MMPI-2 Restructured Form (MMPI-2-RF) [55,56] emerged after the development of the MMPI-2. It is shorter than the MMPI-2 (338 items instead of 567) and does not include existing clinical and content measures from the MMPI-2. Instead, the Restructured Clinical (RC) scale constitutes the MMPI-2-RF's main content [57]. The MMPI-2-RF contains one new scale, seven revised Validity scales, three Higher Order scales based on the RC scale, 23 Specific Problem scales, two Interest scales, and five revised Personality Psychology Five (PSY-5) scales. Because the MMPI-2-RF is shorter and demonstrates greater internal consistency than the original clinical scale, it is better at identifying differences and similarities between cultures and has more advantages considering the large number of available translations. The internal consistency of the MMPI-2-RF differs across groups and items; nevertheless, it has good overall internal consistency (Cronbach's $\alpha = 0.76\text{--}0.93$) [58,59], and many studies have utilized the measure. In this study, we used the measure's 10 Restructured Clinical (RC) scales and 23 Specific Problems (SP) scales as predictors.

The RC scale was developed by Tellegen (2003) et al. to solve the problems of the clinical scale of MMPI-2-RF and the difficulty in interpretation [57]. Ben-Proath and Tellegen recommend interpreting factors by RC scales [55], and by fixing the higher-order structure in the historical outline of personality structure studies, the analysis of RC scales is further confirmed [60]. Within this framework, interpretive information from each of the RC scales is organized, along with the specific problems (SP) scales associated with each RC scale. The RC scale separates the variable of Demoralization (RCd) that is generally reflected in the clinical scale, and extracts and scales the unique key factors measured in each clinical scale. As the convergent validity and discriminant validity of the RC scale improved compared to the clinical scale, it contributed greatly to resolving the ambiguity in the interpretation of the clinical scale. Additionally, RC scales are a more discriminating and predictive measure that can be used independently as empirical studies are accumulated [61].

The 23 Specific Problems scales are stand-alone scales constructed to either highlight important characteristics contained by or associated with the RC scales or to measure unique constructs not assessed by those mid-level scales [56]. The SP scales are divided into four subsets: Somatic/Cognitive, Internalizing, Externalizing, and Interpersonal Problems scales. The five Somatic/Cognitive Problems scales were derived primarily from Somatic Complaints (RC1). The nine Internalizing Problems scales were designed to measure components of Demoralization (RCd) and Dysfunctional Negative Emotions (RC7). Four scales (i.e., Suicidal/Death Ideation, Helplessness/Hopelessness, Self-Doubt, and Inefficacy) assess facets of RCd, whereas five scales (i.e., Stress/Worry, Anxiety, Anger Proneness, Behavior-Restricting Fears, and Multiple Specific Fears) assess aspects of RC7 [62].

2.2.2. MDQ (Mood Disorder Questionnaire)

The Mood Disorder Questionnaire (MDQ) is a self-report that can be quickly and easily scored by a doctor, nurse, or trained medical assistant. The MDQ screens for a

lifetime history of manic or hypomanic syndrome, including 13 yes/no items derived from DSM-IV criteria and clinical experience. The yes/no question also asks whether several of the reported manic or hypomanic symptoms or behaviors were experienced during the same period [63].

The MDQ comprises 13 items (Criterion 1) asking whether one has experienced symptoms of mania or hypomania in one's lifetime. Next, there are items evaluating how many of these symptoms occurred simultaneously (Criterion 2) and how many functional problems resulted from the symptoms (Criterion 3). All Criterion 1 items utilize a yes/no response format, with the sum of the "yes" answers constituting the total score. For Criterion 3, responses are measured on a 4-point scale ranging from "no problem" to "serious problem". The original authors state that if the respondent answers "yes" to more than seven of the 13 items and these symptoms appeared at the same time and caused more than moderate problems and dysfunction, then Criteria 1, 2, and 3 were all met. In such cases, there is a strong possibility of bipolar spectrum disorder. In the original development study, the MDQ showed good internal consistency (Cronbach's $\alpha = 0.90$, and the item-total correlation coefficients ranged from 0.50–0.75 [63]. The validation study for the Korean version of the measure also indicated good internal consistency (Cronbach's $\alpha = 0.88$) [64].

2.2.3. PHQ-9 (The Patient Health Questionnaire)

The PHQ-9 is a part of the PHQ (Patient Health Questionnaire), which is a self-reporting questionnaire developed in order to detect and diagnose mental illnesses in a primary clinical setting. The PHQ-9 is designed to be consistent with DSM-IV depression diagnostic criteria for the selection of major purposes; there is a total of nine questions, and the end of the daily life due to symptoms is added. We have considered that 9 points were used as a cut-off point in the evaluation using the total score of the PHQ-9 [65].

2.3. Procedure

We used 33 total scales (10 RC scales and SP scales) as predictors. To compare the discriminating ability with the existing clinical scale, we analyzed the receiver operating characteristic (ROC) curve with the demoralization RC scale (RCd) for depressive symptoms and the hypomanic activation T-score (RC9) of 65 for hypomanic symptoms as the cutoff points. The ROC curve is a plot that displays the full picture of trade-off between sensitivity and (1-Specificity) over a series of cut-off points. The area under the receiver operating characteristic curve (AUC) is considered as an effective measure of inherent validity of a diagnostic test [66]. Therefore, in this study, a diagnostic analysis was performed with the AUC value using the ROC curve. For ML methods, we used the k-nearest neighbor classification (KNN), linear discriminant analysis (LDA), and random forest classification. All data analysis took place within JASP v0.14.4 (Amsterdam, the Netherlands) and MedCalc v20.009 (MedCalc Software, Mariakerke, Belgium).

2.4. Computational Analyses

Random forest is an ensemble learning method composed of multiple decision trees [67]. Decision trees allow simple and fast learning and testing by simply dividing complex problems in a hierarchical structure, and they show high accuracy and generalization performance by constructing an arbitrary decision tree and performing ensemble learning. The ensemble learning method allows for learning new hypotheses by creating multiple classifiers and combining their predictions. The goal of ensemble learning is to obtain high reliability prediction values compared to a single classifier by synthesizing the prediction results of multiple classifiers. Random forest performs ensemble learning using bootstrap aggregating (bagging). Bagging constructs many sample sets by applying a bootstrapping sampling method that allows duplicates to the dataset to be used for analysis, using each sample set as a training dataset to create a classifier set and then creating a complex classifier.

The KNN is one nearest-neighbor search technique. It is an optimization method of unsupervised learning that finds a solution in given k-nearest neighbors. As a method of classifying new data values as a criterion, a majority vote determines the group. As an algorithm with a relatively simple structure, estimates use only local information, and “lazy learning” that can be classified after the calculation of the entire space is completed by the lazy learning method. This method has the advantage of being efficient and simple to implement, even when the training dataset is large or contains much noise. Additionally, then, the KNN algorithm has been presented as a good machine learning method in previous studies to discriminate mood disorders [68–70].

As a dimension-reduction algorithm, LDA finds a vector that maximizes the difference between the average values of each class and minimizes the variance [71]. Therefore, LDA aims to achieve maximum class discrimination by finding optimal discriminant vectors (transformation) in maximizing the ratio of between-class and within-class distances [72]. As LDA considers the correlations between categorical variables and already correlated variables [73], we selected the minimum set of weakly correlated and potentially discriminable variables according to the results. The study sample was divided into two subsets and used to test the model’s ability to discriminate among depressive, hypomanic, and symptom-free groups.

3. Results

3.1. Experimental Results

3.1.1. General Characteristics

Of the 8645 total responses, we used 5532 as training data, 1384 as validation data, and 1729 as test data. The average age of the sample was 18.92 years (± 1.64). Women accounted for 4273 responses (36.9%). A total of 1116 participants (9.6%) exhibited depressive symptoms, and 1068 (9.2%) exhibited hypomanic symptoms (Table 1).

Table 1. Basic characteristics of participants ($n = 8645$).

Variables		Values
age		18.92 \pm 1.64
gender	male	4371 (73.7%)
	female	4274 (36.9%)
PHQ (+)		1116 (9.6%)
MDQ (+)		1068 (9.2%)

PHQ (+): Patient Health Questionnaire screening positive; MDQ (+): Mood Disorder Questionnaire screening positive.

3.1.2. Diagnostics Characteristics of MMPI-2-RC Scales

When using RCd to predict depressive symptoms, the sensitivity was 90.43% (95% CI = 89.76–91.07), PPV was 95.64 (95% CI = 95.06–96.09), and the area under the curve (AUC) was 0.807; however, when using LDA, the AUC was 0.840. When predicting hypomanic symptoms using RC9, the sensitivity was 88.83% (95% CI = 88.13–89.5), PPV was 96.23 (95% CI = 95.77–96.64) and the AUC was 0.704; in contrast, when using KNN, the AUC was 0.767 (Table 2).

Table 2. Diagnostic Characteristics of Reconstructive Clinical Scales.

	AUC	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
RC9	0.704	88.83 (88.13–89.5)	34.55 (30.1–39.22)	96.23 (95.77–96.64)	14.14 (12.1–16.37)
RCd	0.807	90.43 (89.76–91.07)	51.91 (48.08–55.71)	95.64 (95.16–96.09)	31.72 (29.0–34.54)

AUC: Area Under the ROC Curve; PPV: Positive Predictive Value; NPV: Negative Predictive value.

3.1.3. Mood Disorders Item Combination by ML Technique

For both depressive and hypomanic symptoms, and for each of the three ML techniques, the accuracy was more than 87% (87.3–88.5%). Each ML algorithm showed high sensitivity (88.13–91.71%) and PPV (95.58–99.60%), with LDA showing a sensitivity of 91.71% (95% CI = 90.24–93.02) for depression and KNN showing sensitivity for hypomanic symptoms. LDA resulted in the best sensitivity of 89.29% (95% CI = 87.72–90.72; Table 3).

Table 3. Measurement characteristics according to machine learning techniques.

	ML Methods	Accuracy	F1 Score	AUC	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
MDQ	KNN	0.880	0.843	0.634	89.29 (87.72–90.72)	33.33 (19.09–50.22)	98.31 (97.53–98.89)	6.70 (3.62–11.19)
	Linear Discriminant	0.873	0.743	0.767	89.06 (87.46–90.52)	43.94 (31.74–56.7)	97.56 (96.66–98.28)	13.74 (9.4–19.14)
	Random Forest	0.879	0.832	0.727	88.13 (86.5–89.62)	68.42 (43.45–87.42)	99.60 (99.14–99.85)	6.02 (3.24–10.07)
PHQ	KNN	0.883	0.848	0.770	88.66 (87.05–90.13)	75.56 (60.46–87.12)	99.27 (98.7–99.63)	15.11 (10.7–20.47)
	Linear Discriminant	0.885	0.877	0.840	91.71 (90.24–93.02)	55.03 (46.68–63.18)	95.58 (94.42–96.56)	38.5 (31.93–45.39)
	Random Forest	0.879	0.855	0.828	89.67 (88.09–91.1)	52.38 (41.19–63.4)	97.36 (96.42–98.11)	20.56 (15.35–26.6)

4. Discussion

In this study, we evaluated the MMPI-2's effectiveness in screening for mood disorder symptoms. The instrument yielded high accuracy for both depressive and hypomanic symptoms (87.3–88.5%) when using ML techniques, and the AUC was 0.634–0.840, indicating excellent predictive diagnostics.

The MMPI-2-RF consists of 50 scales, including 10 RC scales that address the limitations of the MMPI-2 clinical scale, and has 229 fewer items than the MMPI-2. The RC scale has significantly improved discriminative validity compared to the original clinical scale. Among them, researchers have studied RCd and RC9 to measure mood symptoms more specifically [74,75]. Typically, the RCd scale measures depressive symptoms and the RC9 scale measures hypomanic symptoms. The RCd and RC9 scales have predicted depressive and hypomanic symptoms of mood disorders with a high score for AUC and Sensitivity, respectively (RCd = 81% AUC, 90.43 sensitivity; RC9 = 70% AUC, 88.83 sensitivity).

In this study, we used ML techniques to examine the relationship between these RC scales and mood symptoms. Compared with the RCd of the MMPI-2-RF (AUC = 0.81), ML-based analysis yielded more accurate predictive values, with the exception of KNN (KNN = 77%, LDA = 74%, RF = 83%). In the case of hypomanic symptoms, when analyzing the RC9 items (AUC = 0.70), the ML techniques showed superior results compared to the traditional statistical techniques, except again for KNN (KNN = 63%, LDA = 77%, RF = 73%). This means that ML-based analysis can facilitate more sophisticated clinical interpretation of the MMPI-2 than analysis using existing clinical scales or content scales and scale pairs. ML algorithms are designed to maximize clinical significance and generalizability, and are well equipped to process a large number of variables as potential predictors [76]. Herein, using ML techniques, we evaluated the predictive value of screening which was analyzed through RCd and PHQ in the MMPI-2-RF scale for depressive symptoms among mood disorders and through RC9 and MDQ for hypomanic symptoms.

The introduction of ML into statistical analysis in psychiatric research represents a paradigm shift beyond simply adding analytical tools for combining and exploring larger datasets [77]. For many years, classical statistical approaches have been used to confirm or refute certain hypotheses, but modern ML research focuses on the overall predictive power of models, especially on how accurately they predict desired outcomes in new and unencountered datasets. Research in this field is primarily evaluated by its potential clinical impact and whether it can provide reliable information about the prognosis for future patients (especially when patients go unscreened and undiagnosed for many years),

leading to treatments and interventions [49]. It will be a useful predictive technique for detection-resistant mood disorders.

The advantage of using the MMPI-2-RF as a screening tool for mood disorders as performed in this study is that it may be a more efficient method of collecting various clinical data simultaneously [9]. Additionally, the MMPI-2-RF takes less than 30 min to complete on a computer, which is more efficient than unstructured interviews and some other self-report tools. This is true even considering that the most widely used screening tool for bipolar disorder, the MDQ, cannot adequately differentiate between bipolar disorder and borderline personality disorder [78]. Unlike most other measurement tools, the MMPI-2-RF can increase diagnostic reliability by including an effectiveness scale to detect underreporting and overreporting of psychiatric symptoms.

The MMPI-2 may also be useful for differential diagnosis of clinical patients with mood disorders. Five of the MMPI-2-RF scales showed significant differences between patients diagnosed with BD and those with MDD. The Activation (ACT) scale had the largest AUC (0.74) and effectively discriminated BD and MDD patients. Overall, 71% of patients were correctly classified according to the ACT scale, and 72% of patients currently experiencing depression were correctly classified. This is superior to the 66% classification rate found in a previous study of SAD-P using a tool other than the MMPI-2-RF [79]; further, because bipolar disorder uses a major depressive episode as a common diagnostic criterion, it can be compared with major depressive disorder. Because discrimination of mood disorders is important, the discriminating ability of the MMPI-2-RF scale suggests its strength as a measurement tool. This study examined the MMPI-2-RF's application as a screening tool, demonstrating that the existing RCd and RC9 also account for considerable AUC (0.807 and 0.704, respectively).

This study has several limitations. First, a study using the results of student health examinations lacks clinical diagnoses. The gold standard is analyzing diagnoses using structured interviews or clinical evaluations, but we used the results of the PHQ-9 and MDQ. The symptom evaluation was performed by scale; even if the scores are the same, it does not mean that they have the same symptoms. In future studies, it may be helpful to subclassify the symptom groups and analyze the symptom features. However, as a tool for screening mood disorder symptoms, our study still suggests the appropriateness of this measure's application. Second, there was no information on participants' clinical histories. Third, our sample of students from a single university may not represent the entire population. However, by securing a sufficiently large number of participants, we argue that there is a certain degree of generalizability with respect to the college student population. Fourth, while the machine learning techniques used in this study have high AUC and sensitivity, they show low specificity, so in future research, theoretical implications or mood disorder tools with high all statistical values should be verified. In addition, future studies use the statistical strength of this paper to recommend empirical studies that include two subgroups for screening for mood disorders (e.g., healthy individuals versus individuals with clinical diagnosis of an affective disorder). Additionally, through other ML techniques (e.g., unsupervised ML techniques, reinforcement learning), future studies could present which features were important based on the machine learning techniques in mood disorders.

In this study, we determined the reliability of the test using the RC scale of the MMPI-2-RF. The advantage of the self-report test is that it may be easier to communicate emotional difficulties compared to in interviews, but it is difficult to ensure reliability. This study was able to reasonably ensure reliability by using the clinically reliable MMPI-2-RF assessment. The MMPI-2 is widely used at selection and interview sites, and analysis using ML techniques suggests the appropriateness of using it for screening high-risk groups as well as for existing clinical and content measures. Additionally, because it is possible to identify emotional states without directly asking about emotional difficulties, it will be helpful in screening clinical symptoms in non-clinical sites (e.g., screening tests). In addition

to the existing MMPI interpretation methods used in the clinical field, prediction-based ML techniques can help screen for mood disorders.

5. Conclusions

We have confirmed that ML techniques exhibit excellent predictive power compared with the RC scales of the MMPI-2-RF. If such techniques are applied to the interpretation of the MMPI-2-RF, it will be possible to detect various psychopathologies early, which is crucial for effective treatment.

Author Contributions: K.L. and S.K. contributed to the design of the study and preparation of the manuscript. K.L. and H.-K.L. coordinated the data acquisition. K.L. and S.K. wrote the draft manuscript, interpretation of the results. K.L., H.-K.L. and S.K. contributed to the critical revision of the article for important intellectual content. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07050245), and the Technology Innovation Program (20012931) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea). This funding source had no role in the design of this study and will not have any role during its execution, analyses, interpretation of the data, or decision to submit results.

Institutional Review Board Statement: This study was approved by the Kongju National University Institutional Review Board (approval No. KNU2015-38).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethical restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Murphy, J.J. Current practices in the use of psychological testing by police agencies. *J. Crim. Law Criminol. Police Sci.* **1972**, *63*, 570–576. [[CrossRef](#)]
2. Schoenfeld, L.S.; Kobos, J.C.; Phinney, I.R. Screening Police Applicants—A Study of Reliability with the Mmpi. *Psychol. Rep.* **1980**, *47*, 419–425. [[CrossRef](#)]
3. Bernstein, I.H.; Schoenfeld, L.S.; Costello, R.M. Truncated component regression, multicollinearity and the MMPI's use in a police officer selection setting. *Multivar. Behav. Res.* **1982**, *17*, 99–116. [[CrossRef](#)]
4. Costello, R.M.; Schoenfeld, L.S.; Kobos, J. Police Applicant Screening—An Analog Study. *J. Clin. Psychol.* **1982**, *38*, 216–221. [[CrossRef](#)]
5. Beutler, L.E.; Storm, A.; Kirkish, P.; Scogin, F.; Gaines, J.A. Parameters in the Prediction of Police Officer Performance. *Prof. Psychol. Res. Pract.* **1985**, *16*, 324–335. [[CrossRef](#)]
6. Beutler, L.E.; Nussbaum, P.D.; Meredith, K.E. Changing personality patterns of police officers. *Prof. Psychol. Res. Pract.* **1988**, *19*, 503. [[CrossRef](#)]
7. Butcher, J.N. Psychological assessment of airline pilot applicants with the MMPI-2. *J. Pers. Assess.* **1994**, *62*, 31–44. [[CrossRef](#)]
8. Greene, R.L. *The MMPI-2/MMPI: An Interpretive Manual*; Allyn & Bacon: Boston, MA, USA, 1991.
9. Ben-Porath, Y.S.; Butcher, J.N.; Graham, J.R. Contribution of the MMPI-2 content scales to the differential diagnosis of psychopathology. *Psychol. Assess. A J. Consult. Clin. Psychol.* **2001**, *3*, 634–640.
10. Kessler, R.C.; Demler, O.; Frank, R.G.; Olfson, M.; Pincus, H.A.; Walters, E.E.; Wang, P.; Wells, K.B.; Zaslavsky, A.M. Prevalence and treatment of mental disorders, 1990 to 2003. *N. Engl. J. Med.* **2005**, *352*, 2515–2523. [[CrossRef](#)] [[PubMed](#)]
11. Harris, E.C.; Barraclough, B.M. Suicide as an outcome for medical disorders. *Medicine* **1994**, *73*, 281–296. [[CrossRef](#)] [[PubMed](#)]
12. MacQueen, G.M.; Young, L.T.; Joffe, R.T. A review of psychosocial outcome in patients with bipolar disorder. *Acta Psychiatr. Scand.* **2001**, *103*, 163–170. [[CrossRef](#)]
13. Calabrese, J.R.; Hirschfeld, R.M.; Reed, M.; Davies, M.A.; Frye, M.A.; Keck, P.E.; Lewis, L.; McElroy, S.L.; Wagner, K.D. Impact of bipolar disorder on a US community sample. *J. Clin. Psychiatry* **2003**, *64*, 425–432. [[CrossRef](#)]
14. Ormel, J.; VonKorff, M.; Ustun, T.B.; Pini, S.; Korten, A.; Oldehinkel, T. Common mental disorders and disability across cultures. Results from the WHO Collaborative Study on Psychological Problems in General Health Care. *JAMA* **1994**, *272*, 1741–1748. [[CrossRef](#)]
15. Araya, R.; Rojas, G.; Fritsch, R.; Acuña, J.; Lewis, G. Common mental disorders in Santiago, Chile: Prevalence and socio-demographic correlates. *Br. J. Psychiatry* **2001**, *178*, 228–233. [[CrossRef](#)]

16. Nielsen, A.C.; Williams, T.A. Depression in ambulatory medical patients: Prevalence by self-report questionnaire and recognition by nonpsychiatric physicians. *Arch. Gen. Psychiatry* **1980**, *37*, 999–1004. [[CrossRef](#)]
17. Pérez-Stable, E.J.; Miranda, J.; Muñoz, R.F.; Ying, Y.-W. Depression in medical outpatients: Underrecognition and misdiagnosis. *Arch. Intern. Med.* **1990**, *150*, 1083–1088. [[CrossRef](#)]
18. Ormel, J.; Koeter, M.W.; Van den Brink, W.; Van de Willige, G. Recognition, management, and course of anxiety and depression in general practice. *Arch. Gen. Psychiatry* **1991**, *48*, 700–706. [[CrossRef](#)]
19. Ghaemi, S.N.; Boiman, E.E.; Goodwin, F.K. Diagnosing bipolar disorder and the effect of antidepressants: A naturalistic study. *J. Clin. Psychiatry* **2000**, *61*, 804–808; quiz 809. [[CrossRef](#)] [[PubMed](#)]
20. Ghaemi, S.N.; Ko, J.Y.; Goodwin, F.K. “Cade’s disease” and beyond: Misdiagnosis, antidepressant use, and a proposed definition for bipolar spectrum disorder. *Can. J. Psychiatry* **2002**, *47*, 125–134. [[CrossRef](#)]
21. Zimmerman, M.; Galione, J.N. Screening for bipolar disorder with the Mood Disorders Questionnaire: A review. *Harv. Rev. Psychiatry* **2011**, *19*, 219–228. [[CrossRef](#)]
22. Hegadoren, K.; Norris, C.; Lasiuk, G.; da Silva, D.G.V.; Chivers-Wilson, K. The Many Faces of Depression in Primary Care. *Texto Contexto Enferm.* **2009**, *18*, 155–164. [[CrossRef](#)]
23. Almeida, O.P. Prevention of depression in older age. *Maturitas* **2014**, *79*, 136–141. [[CrossRef](#)] [[PubMed](#)]
24. Hall, C.A.; Reynolds-Iii, C.F. Late-life depression in the primary care setting: Challenges, collaborative care, and prevention. *Maturitas* **2014**, *79*, 147–152. [[CrossRef](#)] [[PubMed](#)]
25. Park, L.T.; Zarate, C.A., Jr. Depression in the Primary Care Setting. *N. Engl. J. Med.* **2019**, *380*, 559–568. [[CrossRef](#)] [[PubMed](#)]
26. Lish, J.D.; Dime-Meenan, S.; Whybrow, P.C.; Price, R.A.; Hirschfeld, R.M. The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. *J. Affect. Disord.* **1994**, *31*, 281–294. [[CrossRef](#)]
27. Dunner, D.L. Clinical consequences of under-recognized bipolar spectrum disorder. *Bipolar Disord.* **2003**, *5*, 456–463. [[CrossRef](#)] [[PubMed](#)]
28. Ghaemi, S.N.; Wingo, A.P.; Filkowski, M.A.; Baldessarini, R.J. Long-term antidepressant treatment in bipolar disorder: Meta-analyses of benefits and risks. *Acta Psychiatr. Scand.* **2008**, *118*, 347–356. [[CrossRef](#)]
29. Smith, D.V.; Huettel, S.A. Decision neuroscience: Neuroeconomics. *Wiley Interdiscip. Rev. Cogn. Sci.* **2010**, *1*, 854–871. [[CrossRef](#)]
30. Franchini, L.; Zanardi, R.; Smeraldi, E.; Gasperini, M. Early onset of lithium prophylaxis as a predictor of good long-term outcome. *Eur. Arch. Psychiatry Clin. Neurosci.* **1999**, *249*, 227–230. [[CrossRef](#)]
31. Scott, J.; Paykel, E.; Morriss, R.; Bentall, R.; Kinderman, P.; Johnson, T.; Abbott, R.; Hayhurst, H. Cognitive-behavioural therapy for severe and recurrent bipolar disorders: Randomised controlled trial. *Br. J. Psychiatry* **2006**, *188*, 313–320. [[CrossRef](#)]
32. Berk, M.; Dodd, S.; Callaly, P.; Berk, L.; Fitzgerald, P.; de Castella, A.R.; Folia, S.; Folia, K.; Tahtalian, S.; Biffin, F.; et al. History of illness prior to a diagnosis of bipolar disorder or schizoaffective disorder. *J. Affect. Disord.* **2007**, *103*, 181–186. [[CrossRef](#)]
33. Judd, L.L.; Akiskal, H.S.; Schettler, P.J.; Endicott, J.; Maser, J.; Solomon, D.A.; Leon, A.C.; Rice, J.A.; Keller, M.B. The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch. Gen. Psychiatry* **2002**, *59*, 530–537. [[CrossRef](#)]
34. Wells, K.B.; Stewart, A.; Hays, R.D.; Burnam, M.A.; Rogers, W.; Daniels, M.; Berry, S.; Greenfield, S.; Ware, J. The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. *JAMA* **1989**, *262*, 914–919. [[CrossRef](#)]
35. Broadhead, W.E.; Blazer, D.G.; George, L.K.; Tse, C.K. Depression, disability days, and days lost from work in a prospective epidemiologic survey. *JAMA* **1990**, *264*, 2524–2528. [[CrossRef](#)]
36. Simon, G.; Ormel, J.; VonKorff, M.; Barlow, W. Health care costs associated with depressive and anxiety disorders in primary care. *Am. J. Psychiatry* **1995**, *152*, 352–357. [[CrossRef](#)] [[PubMed](#)]
37. Frasure-Smith, N.; Lesperance, F.; Gravel, G.; Masson, A.; Juneau, M.; Talajic, M.; Bourassa, M.G. Social support, depression, and mortality during the first year after myocardial infarction. *Circulation* **2000**, *101*, 1919–1924. [[CrossRef](#)]
38. Penninx, B.W.; Beekman, A.T.; Honig, A.; Deeg, D.J.; Schoevers, R.A.; van Eijk, J.T.; van Tilburg, W. Depression and cardiac mortality: Results from a community-based longitudinal study. *Arch. Gen. Psychiatry* **2001**, *58*, 221–227. [[CrossRef](#)]
39. Posternak, M.A.; Miller, I. Untreated short-term course of major depression: A meta-analysis of outcomes from studies using wait-list control groups. *J. Affect. Disord.* **2001**, *66*, 139–146. [[CrossRef](#)]
40. Pearson, S.D.; Katzelnick, D.J.; Simon, G.E.; Manning, W.G.; Helstad, C.P.; Henk, H.J. Depression among high utilizers of medical care. *J. Gen. Intern. Med.* **1999**, *14*, 461–468. [[CrossRef](#)]
41. Vohringer, P.A.; Jimenez, M.I.; Igor, M.A.; Fores, G.A.; Correa, M.O.; Sullivan, M.C.; Holtzman, N.S.; Whitham, E.A.; Barroilhet, S.A.; Alvear, K.; et al. Detecting mood disorder in resource-limited primary care settings: Comparison of a self-administered screening tool to general practitioner assessment. *J. Med. Screen.* **2013**, *20*, 118–124. [[CrossRef](#)] [[PubMed](#)]
42. Hirschfeld, R.M.; Holzer, C.; Calabrese, J.R.; Weissman, M.; Reed, M.; Davies, M.; Frye, M.A.; Keck, P.; McElroy, S.; Lewis, L.; et al. Validity of the mood disorder questionnaire: A general population study. *Am. J. Psychiatry* **2003**, *160*, 178–180. [[CrossRef](#)]
43. Hirschfeld, R.M.; Lewis, L.; Vornik, L.A. Perceptions and impact of bipolar disorder: How far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. *J. Clin. Psychiatry* **2003**, *64*, 161–174. [[CrossRef](#)] [[PubMed](#)]
44. Freedman, R.; Lewis, D.A.; Michels, R.; Pine, D.S.; Schultz, S.K.; Tamminga, C.A.; Gabbard, G.O.; Gau, S.S.; Javitt, D.C.; Oquendo, M.A.; et al. The initial field trials of DSM-5: New blooms and old thorns. *Am. J. Psychiatry* **2013**, *170*, 1–5. [[CrossRef](#)]
45. Wunderink, L.; Sytema, S.; Nienhuis, F.J.; Wiersma, D. Clinical recovery in first-episode psychosis. *Schizophr. Bull.* **2009**, *35*, 362–369. [[CrossRef](#)] [[PubMed](#)]

46. Modai, I.; Israel, A.; Mendel, S.; Hines, E.L.; Weizman, R. Neural network based on adaptive resonance theory as compared to experts in suggesting treatment for schizophrenic and unipolar depressed in-patients. *J. Med. Syst.* **1996**, *20*, 403–412. [[CrossRef](#)]
47. Gradus, J.L.; Rosellini, A.J.; Horvath-Puho, E.; Street, A.E.; Galatzer-Levy, I.; Jiang, T.; Lash, T.L.; Sorensen, H.T. Prediction of Sex-Specific Suicide Risk Using Machine Learning and Single-Payer Health Care Registry Data From Denmark. *JAMA Psychiatry* **2020**, *77*, 25–34. [[CrossRef](#)] [[PubMed](#)]
48. Kessler, R.C.; van Loo, H.M.; Wardenaar, K.J.; Bossarte, R.M.; Brenner, L.A.; Cai, T.; Ebert, D.D.; Hwang, I.; Li, J.; de Jonge, P. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol. Psychiatry* **2016**, *21*, 1366–1371. [[CrossRef](#)] [[PubMed](#)]
49. Chekroud, A.M.; Bondar, J.; Delgadillo, J.; Doherty, G.; Wasil, A.; Fokkema, M.; Cohen, Z.; Belgrave, D.; DeRubeis, R.; Iniesta, R.; et al. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry* **2021**, *20*, 154–170. [[CrossRef](#)]
50. Kessler, R.C.; Warner, C.H.; Ivany, C.; Petukhova, M.V.; Rose, S.; Bromet, E.J.; Brown, M.; Cai, T.; Colpe, L.J.; Cox, K.L. Predicting suicides after psychiatric hospitalization in US Army soldiers: The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry* **2015**, *72*, 49–57. [[CrossRef](#)]
51. King, M.; Walker, C.; Levy, G.; Bottomley, C.; Royston, P.; Weich, S.; Bellon-Saameno, J.A.; Moreno, B.; Švab, I.; Rotar, D. Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: The PredictD study. *Arch. Gen. Psychiatry* **2008**, *65*, 1368–1376. [[CrossRef](#)]
52. Wang, J.L.; Patten, S.; Sareen, J.; Bolton, J.; Schmitz, N.; MacQueen, G. Development and validation of a prediction algorithm for use by health professionals in prediction of recurrence of major depression. *Depress. Anxiety* **2014**, *31*, 451–457. [[CrossRef](#)]
53. Koutsouleris, N.; Meisenzahl, E.M.; Davatzikos, C.; Bottlender, R.; Frodl, T.; Scheuerecker, J.; Schmitt, G.; Zetsche, T.; Decker, P.; Reiser, M. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch. Gen. Psychiatry* **2009**, *66*, 700–712. [[CrossRef](#)] [[PubMed](#)]
54. Kim, S.; Lee, H.-K.; Lee, K. Which PHQ-9 items can effectively screen for suicide? Machine learning approaches. *Int. J. Environ. Res. Public Health* **2021**, *18*, 3339. [[CrossRef](#)]
55. Ben-Porath, Y.S.; Tellegen, A. *Minnesota Multiphasic Personality Inventory-2 Restructured Form*; University of Minnesota Press: Minneapolis, MN, USA, 2008.
56. Tellegen, A.; Ben-Porath, Y.S. *MMPI-2-RF: Technical Manual*; University of Minnesota Press: Minneapolis, MN, USA, 2008.
57. Tellegen, A.; Ben-Porath, Y.; McNulty, J.; Arbisi, P.; Graham, J.; Kaemmer, B. *MMPI-2 Restructured Clinical (RC) Scales: Development, Validation, and Interpretation*; University of Minnesota Press: Minneapolis, MN, USA, 2003.
58. Van der Heijden, P.; Egger, J.; Derksen, J. Comparability of scores on the MMPI-2-RF scales generated with the MMPI-2 and MMPI-2-RF booklets. *J. Personal. Assess.* **2010**, *92*, 254–259. [[CrossRef](#)] [[PubMed](#)]
59. Lee, T.T.; Graham, J.R.; Arbisi, P.A. The utility of MMPI-2-RF scale scores in the differential diagnosis of Schizophrenia and Major Depressive Disorder. *J. Personal. Assess.* **2018**, *100*, 305–312. [[CrossRef](#)] [[PubMed](#)]
60. Ben-Porath, Y.S. *Interpreting the mmpi-2-rf*; University of Minnesota Press: Minneapolis, MN, USA, 2012.
61. Butcher, J.N. *Minnesota Multiphasic Personality Inventory-2: Manual for Administration, Scoring, and Interpretation*; University of Minnesota Press: Minneapolis, MN, USA, 2001.
62. Forbey, J.D.; Lee, T.T.; Handel, R.W. Correlates of the MMPI-2-RF in a college setting. *Psychol. Assess.* **2010**, *22*, 737. [[CrossRef](#)] [[PubMed](#)]
63. Hirschfeld, R.M.; Williams, J.B.; Spitzer, R.L.; Calabrese, J.R.; Flynn, L.; Keck, P.E., Jr.; Lewis, L.; McElroy, S.L.; Post, R.M.; Rapport, D.J.; et al. Development and validation of a screening instrument for bipolar spectrum disorder: The Mood Disorder Questionnaire. *Am. J. Psychiatry* **2000**, *157*, 1873–1875. [[CrossRef](#)]
64. Jon, D.I.; Yoon, B.H.; Jung, H.Y.; Ha, K.S.; Shin, Y.C.; Bahk, W.M. A validation study of the Korean version Mood Disorder Questionnaire (K-MDQ). *J. Korean Neuropsychiatr. Assoc.* **2005**, *44*, 583–590.
65. Kroenke, K.; Spitzer, R.L.; Williams, J.B. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* **2001**, *16*, 606–613. [[CrossRef](#)]
66. Kumar, R.; Indrayan, A. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics* **2011**, *48*, 277–287. [[CrossRef](#)]
67. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
68. Ansari, A.; Khalili, M. A High Performance Scheme to Detect Major Depressive Disorder Using MLP, RBF and KNN Neural Network Models. *Int. J. Artif. Intell. Mechatron.* **2014**, *3*, 123–126.
69. Ghasemi, M.M.; Khalili, M. A Joint KNN-RBF Based Algorithm to Diagnose the Bipolar Disorder. *Int. J. Artif. Intell. Mechatron.* **2014**, *3*, 239–242.
70. Mantri, S.; Chavan, P.; Kadam, P.; Patil, D.; Wadhai, V. Classifying mood disordered patients and normal subjects using various machine learning techniques. *Int. J. Adv. Res. Comput. Sci. Manag. Stud.* **2013**, *1*, 90–97.
71. Khan, A.; Farooq, H. Principal component analysis-linear discriminant analysis feature extractor for pattern recognition. *arXiv* **2012**, arXiv:1204.1177.
72. Fukunaga, K. *Introduction to Statistical Pattern Recognition*; Elsevier: Amsterdam, The Netherlands, 2013.
73. McLachlan, G.J. *Discriminant Analysis and Statistical Pattern Recognition*; John Wiley & Sons: Hoboken, NJ, USA, 2004; Volume 544.

74. Sellbom, M.; Bagby, R.M.; Kushner, S.; Quilty, L.C.; Ayeart, L.E. Diagnostic construct validity of MMPI-2 Restructured Form (MMPI-2-RF) scale scores. *Assessment* **2012**, *19*, 176–186. [[CrossRef](#)] [[PubMed](#)]
75. Haber, J.C.; Baum, L.J. Minnesota Multiphasic Personality Inventory-2 restructured form (MMPI-2-RF) Scales as predictors of psychiatric diagnoses. *S. Afr. J. Psychol.* **2014**, *44*, 439–453. [[CrossRef](#)]
76. Linthicum, K.P.; Schafer, K.M.; Ribeiro, J.D. Machine learning in suicide science: Applications and ethics. *Behav. Sci. Law* **2019**, *37*, 214–222. [[CrossRef](#)]
77. Breiman, L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Stat. Sci.* **2001**, *16*, 199–231. [[CrossRef](#)]
78. Zimmerman, M.; Galione, J.N.; Ruggero, C.J.; Chelminski, I.; Young, D.; Dalrymple, K.; McGlinchey, J.B. Screening for bipolar disorder and finding borderline personality disorder. *J. Clin. Psychiatry* **2010**, *71*, 1212–1217. [[CrossRef](#)] [[PubMed](#)]
79. Solomon, D.A.; Leon, A.C.; Maser, J.D.; Truman, C.J.; Coryell, W.; Endicott, J.; Teres, J.J.; Keller, M.B. Distinguishing bipolar major depression from unipolar major depression with the screening assessment of depression-polarity (SAD-P). *J. Clin. Psychiatry* **2006**, *67*, 434–442. [[CrossRef](#)] [[PubMed](#)]