1	Efficient Prediction of Vitamin B Deficiencies via
2	Machine-learning Using Routine Blood Test Results in Patients
3	With Intense Psychiatric Episode
4	
5	Hidetaka Tamune ^{1)2)3)5)*} , Jumpei Ukita ³⁾⁴⁾⁵⁾ , Yu Hamamoto ¹⁾²⁾ , Hiroko Tanaka ¹⁾²⁾ , Kenji
6	Narushima ¹⁾ , Naoki Yamamoto ¹⁾
7	1) Department of Neuropsychiatry, Tokyo Metropolitan Tama Medical Center, Tokyo,
8	Japan
9	2) Department of Neuropsychiatry, Graduate School of Medicine, The University of
10	Tokyo, Tokyo, Japan
11	3) Mental Health Research Course, Faculty of Medicine, The University of Tokyo,
12	Tokyo, Japan
13	4) Department of Physiology, Graduate School of Medicine, The University of Tokyo,
14	Tokyo, Japan
15	5) H. Tamune and JU contributed equally to this work
16	
17	* Correspondence:
18	Hidetaka Tamune, M.D., tamune-tky@umin.ac.jp
19	
20	Abstract: 294 words
21	Main text: 1878 words + 4 Tables + 4 Figures + 24 references
22	Keywords: Machine Learning; Random Forest Classifier; Vitamin B Deficiency; Folic
23	Acid; Early Diagnosis; Decision support techniques or decision making.
24	

Vitamin deficiency predicted by ML

25 Abstract

26**Background:** Vitamin B deficiency is common worldwide and may lead to psychiatric 27symptoms; however, vitamin B deficiency epidemiology in patients with intense 28psychiatric episode has rarely been examined. Moreover, vitamin deficiency testing is costly and time-consuming. It hampered to effectively rule out vitamin 2930 deficiency-induced intense psychiatric symptoms. In this study, we aimed to clarify the 31epidemiology of these deficiencies and efficiently predict them using machine-learning 32models from patient characteristics and routine blood test results that can be obtained 33 within one hour. Methods: We reviewed 497 consecutive patients deemed to be at imminent risk of 3435seriously harming themselves or others over 2 years. Machine-learning models were trained to predict each deficiency from age, sex, and 29 routine blood test results. 36 **Results:** We found that 112 (22.5%), 80 (16.1%), and 72 (14.5%) patients had vitamin 3738 B_1 , vitamin B_{12} , and folate (vitamin B_9) deficiency, respectively. Also, the 39 machine-learning models well generalized to predict the deficiency in the future unseen 40data; areas under the receiver operating characteristic curves for the validation dataset (i.e. dataset not used for training the models) were 0.716, 0.599, and 0.796, respectively. 4142The Gini importance of these vitamins provided further evidence of a relationship 43between these vitamins and the complete blood count, while also indicating a hitherto 44rarely considered, potential association between these vitamins and alkaline phosphatase (ALP) or thyroid stimulating hormone (TSH). 4546**Discussion:** This study demonstrates that machine-learning can efficiently predict some vitamin deficiencies in patients with active psychiatric symptoms, based on the largest 47

48 cohort to date with intense psychiatric episode. The prediction method may expedite 49 risk stratification and clinical decision-making regarding whether replacement therapy 50 should be prescribed. Further research includes validating its external generalizability in 51 other clinical situations and clarify whether interventions based on this method can

52 improve patient care and cost-effectiveness.

54 **1. Introduction**

Vitamin B deficiency is common worldwide and may lead to psychiatric symptoms^{1–4}. For example, meta-analyses have shown that patients with schizophrenia or first-episode psychosis have lower folate (vitamin B₉) levels than their healthy counterparts^{4,5}. Moreover, vitamin therapy can effectively alleviate symptoms in a subgroup of patients with schizophrenia^{3,6–8}. However, the epidemiology of vitamin B deficiency in patients with active mental symptoms requiring immediate hospitalization has rarely been examined.

In a psychiatric emergency, psychiatrists should promptly distinguish treatable 6263 patients with altered mental status due to a physical disease from patients with an 64authentic mental disorder (international statistical classification of diseases and related health problems-10, ICD-10 code: F2-9). However, vitamin deficiency testing is very 65costly (around 60 dollars for each measurement of vitamin B_1 (vit B_1), vitamin B_{12} 66 $(vitB_{12})$, or folate in the U.S.; 15–25 dollars for each test in Japan) and usually requires 67at least two days. Therefore, an efficient, cost-effective method of predicting vitamin B 68 69 deficiency is needed.

Although several studies have applied machine-learning to the prediction of diagnosis or treatment outcomes^{9–11}, no study using machine-learning has focused on vitamin B deficiencies. We herein explore whether vitB₁, vitB₁₂, and folate deficiencies can be predicted using a machine-learning classifier from patient characteristics and routine blood test results obtained within one hour based on a large cohort of patients requiring urgent psychiatric hospitalization.

77 **2. Methods**

78 2.1. Medical chart review

We reviewed consecutive patients admitted to the Department of Neuropsychiatry at Tokyo Metropolitan Tama Medical Center between September 2015 and August 2017 under the urgent involuntary hospitalization law, which requires the immediate psychiatric hospitalization of patients at imminent risk of seriously harming themselves or others. The necessity of hospitalization was judged by designated mental health specialists. The patient characteristics, ICD-10 codes, and laboratory data were gathered retrospectively.

Since the reference ranges for vitB₁, vitB₁₂, and folate are 70–180 nmol/L (30–77 ng/mL), 180–914 ng/L, and > 4.0 μ g/L, respectively¹², a deficiency of the nutrients was defined as < 30 ng/mL, < 180 ng/L, and < 4.0 μ g/L, respectively, unless otherwise stated.

90

91 **2.2. Random forest classifier and statistics**

92A random forest classifier was trained to predict the deficiency of each substance from age, sex, and 29 routine blood variables (described in the Result section 93 94with values). The random forest classifier was trained using the dataset populated in the 95period from September 2015 to December 2016 (the "Training set"). First, we 96 optimized the hyperparameters of the classifier by selecting the best combination of 97hyperparameters that maximized the "5-fold cross validation" accuracy, among many 98 combinations within appropriate ranges. The cross-validation accuracy was computed as follows; in one session, the classifier was trained using 80% of the training set and 99 100 evaluated on the withheld 20% of the training set. This session was performed five 101 times so that every data would be withheld once. The accuracies were finally averaged 102across sessions to yield the cross-validation accuracy. By incorporating this process, the 103 classifier was generalized to unseen data (Graphical method is shown in Figure 1).

Using the optimized hyperparameters, the classifiers were then validated using data collected from January 2017 through August 2017 (the "Validation set"). We report the classification performance on the validation set in the results section unless otherwise stated. We quantified the sensitivity, specificity, and accuracy (defined as the average of the sensitivity and the specificity on the optimal operating point) using receiver operating characteristic curves (ROCs). We also quantified the 95% confidence interval of the accuracy using 1000-times bootstrapping.

111 When investigating the Gini importance and the partial dependency¹³, we 112 retrained the classifiers using all datasets. All data analyses were performed using 113 Python (2.7.10) with the Scikit-learn package (0.19.0) and R (3.4.2) with the edarf 114 package (1.1.1).

115

116 **2.3. Ethical considerations**

Informed consent was obtained from participants using an opt-out form on the website. The study protocol was approved by the Research Ethics Committee, Tokyo Metropolitan Tama Medical Center (Approval number: 28-8). The study complied with the Declaration of Helsinki and the STROBE statement.

122 **3. Results**

123 **3.1. Eligible patients**

124 During the 2-year study period, 497 consecutive patients (496 were Asian) were enrolled. The mean age (standard deviation, SD) was 42.3 (± 15.4) years, and 228 125patients (45.9%) were women. F2 (Schizophrenia, schizotypal, delusional, and other 126127non-mood psychotic disorders) was diagnosed in over 60% of the patients. The ICD-10 codes of the patients and the number of deficiencies at several cut-off values for vitB₁, 128vitB₁₂, and folate are shown in **Table 1**. According to the predefined cut-off values¹², 129112 (22.5%), 80 (16.1%), and 72 (14.5%) patients exhibited a deficiency of vitB₁ (<30130ng/mL), vitB₁₂ (<180 ng/L), and folate (<4.0 μ g/L), respectively. Vitamin B deficiencies 131132in sub-groups are shown in **Table 2**. A summary of the full dataset is shown in **Table 3**. 133 Detailed information (sub-datasets) is shown in Supplementary Table 1, 2, and 3 134online. Histograms of vitB₁, vitB₁₂, and folate values are shown in Figure 2 A-C.

135

136 **3.2. Prediction via machine-learning using routine blood test results**

A random forest classifier was trained to predict the deficiency of each substance from patient characteristics and routine blood test results. The classifier was trained using the dataset gathered in the period from September 2015 to December 2016 (the "Training set", n = 373), which was then validated from January 2017 through August 2017 (the "Validation set", n = 124).

The area under the ROCs (AUCs) for the validation set were 0.716, 0.599, and 0.796, for vitB₁, vitB₁₂, and folate, respectively (**Figure 2 D-F** and **Table 4**). With some operative points on the ROC, the sensitivity, specificity and accuracy for the validation set were calculated (**Table 4**. See also **Supplementary Table 4** for training set and **Supplementary Table 5** for different operating points).

147 When the prediction performances were compared between the classifiers 148 trained using the dataset from the F2 population and the classifiers trained using the

149 dataset from the other population, the AUC was not statistically different (DeLong's 150 test), except in the case of vitB₁ (see **Supplementary Table 6**).

Figure 3 shows the Gini importance (a–c) and partial dependency plots (d–f) for the eight most important variables for each substance. The results provided further evidence of a relationship between the vitamin B levels and complete blood count while also indicating the hitherto rarely considered, potential association between these vitamins and alkaline phosphatase (ALP) or thyroid stimulating hormone (TSH).

156

157 **3.3. Robustness verification**

We verified the robustness of the results by two independent means. First, we used different cut-off values to define the deficiency^{14–16}. Although the AUC for the validation set, shown in **Supplementary Table 7**, tended to be higher when strict cut-off values were used, the obtained AUCs were not statistically significant (p > 0.05, DeLong's test with Bonferroni correction).

Second, we trained and evaluated random forest classifiers using a dataset split 163164in a different way; the classifier was trained using the dataset collected in the period from the 31st of January, 2016 to August 2017, which was then validated with data 165gathered from September 2015 to the 31st of January, 2016. Note that the sample sizes 166167of the training and validation sets were equal to those in the original setting. The AUCs for the validation set were 0.771, 0.621, and 0.745 for vitB₁, vitB₁₂, and folate, 168169respectively; none were statistically different from the AUC trained using the original 170 setting (DeLong's test), further demonstrating the robustness of the performance.

4. Discussion

173 **4.1. Relevance of the present study**

Based on the largest cohort to date of patients at imminent risk of seriously 174harming themselves or others, this study indicated that deficiency of certain vitamins 175can be predicted in an efficient manner via machine-learning using routine blood test 176177results. Given the large number of patients with vitamin B deficiencies, empirical therapy might be acceptable; however, risk stratification is preferred for personalized 178medicine and shared decision-making. The prediction method presented here may 179expedite clinical decision-making as to whether vitamins should be prescribed to a 180 patient (Graphical abstract is shown in Figure 4). 181

Remarkably, the AUC for folate deficiency was 0.796. Folate features the potential to maintain neuronal integrity and is one of the homocysteine-reducing B-vitamins⁵; homocysteine has been linked to the etiology of schizophrenia¹⁷, and vitamin B supplements have been reported to reduce psychiatric symptoms significantly in patients with schizophrenia⁷. As our study does not present longitudinal results, an intervention effect of folate supplementation in the cohort remains to be clarified.

188

189 **4.2. Trade-off of interpretability and generalizability using machine-learning**

Compared to the AUC of folate, AUCs of vitB₁ and vitB₁₂ were relatively low.
Using other parameters that were not incorporated into this model or using other models
including deep neural networks might increase the accuracy of prediction.

However, interpretability and completeness of machine-learning classifiers are subject to trade-off¹⁷. Although completeness and generalizability are desirable, interpretability is also indispensable, especially in clinical settings, since it provides meaningful and trustworthy findings for clinical physicians as well as new biological insights¹⁸. In this study we chose random forest classifiers since they provide expressive and interpretable data, with sufficient accuracy.

199

200 **4.3. Biological mechanism suggestion**

Using the random forest classifiers, as shown in **Figure 3**, we identified several items related to complete blood count as top hits. Notably, our classifier was blind to any biological knowledge, including the well-established association between anemia and vitamin B deficiency, including folate¹⁹. The results provide further evidence of a relationship between vitamin B levels and the complete blood count and support the use of machine-learning to investigate novel, underlying biological mechanisms²⁰.

ALP and its metabolites indicate the vitamin B_6 status²¹; low vitB₁₂ is 207potentially associated with low ALP²². More generally, ALP may have a close and 208complicated relationship with the overall vitamin B group. Autoimmune disorders, 209especially thyroid disease, are commonly associated with pernicious anaemia²³, but 210there has been no established hypothesis regarding the causal relationships between 211thyroid disease and vitamin B deficiencies. The potential association between the levels 212 of these vitamins and ALP or TSH awaits further study, both investigations of 213populations and basic research²⁴. 214

215

216 **4.4. Limitations**

This study is subject to several limitations. First, the findings of this 217single-center retrospective study may have limited generalizability. Second, the patients' 218219long-term prognosis was not investigated due to administrative restrictions; the extent to 220which this method can expedite clinical decision-making is therefore unclear. Further, we did not investigate the relationship between serological values and the need for 221222intervention. The lack of data for vitamin B deficiency in the Japanese general population hampered the comparison between the experimental cohort and their 223224counterparts who lacked psychiatric symptoms. Establishing appropriate reference 225values and an assessment method requires further investigation. Finally, we did not

assess the predictive value of other nutritional impairments, including vitamin B_6 and homocysteine deficiency, which were previously shown to have a close link with psychiatric symptoms^{3,5}; however, our study provides fundamental data on nutritional impairment based on the largest cohort of patients with intense psychiatric episode ever assembled for this purpose and presents a potential framework for predicting nutritional impairment using machine-learning.

232

4.5. Conclusion

The present report is, to the best of our knowledge, the first to demonstrate that machine-learning can efficiently predict nutritional impairment. Further research is needed to validate the external generalizability of the findings in other clinical situations and clarify whether interventions based on this method can improve patient care and cost-effectiveness.

5. Contribution to the Field Statement

Vitamin B deficiency is common worldwide and may lead to psychiatric 241symptoms; however, vitamin B deficiency epidemiology in patients with intense 242psychiatric symptoms has rarely been examined. Moreover, vitamin deficiency testing is 243244costly and time-consuming. Based on the largest cohort to date of patients at imminent 245risk of seriously harming themselves or others, this study demonstrated that the 246deficiency of certain vitamins can be predicted in an efficient manner via 247machine-learning models from patient characteristics and routine blood test results 248obtained within one hour.

In detail, among the 497 patients investigated (over 60% was diagnosed with 249250schizophrenia or related psychotic disorders), 22.5%, 16.1%, and 14.5% patients had a deficiency of vitamin B_1 , B_{12} , and folate, respectively, by direct measurement. Also, the 251machine-learning models well generalized to predict the deficiency in unseen datasets; 252253areas under the receiver operating characteristic curves for the validation dataset were 0.716, 0.599, and 0.796, respectively. The prediction method presented in this study 254255may expedite risk stratification and clinical decision-making regarding whether replacement therapy should be prescribed. The results also provided further evidence for 256257a well-known relationship between these vitamins and the complete blood count and 258supported the application of machine-learning to investigate novel, underlying 259biological mechanisms.

261 **6. Acknowledgements**

262	We thank Mr. James Robert Valera for his assistance in editing this manuscript
263	and all the staff for their care of the patients and their contributions to this study.
264	
265	7. Author Contributions Statement
266	H. Tamune has full access to all data and takes responsibility for the integrity of
267	the data. H. Tamune, JU, KN, and NY conceived the study. H. Tamune, YH, and H.
268	Tanaka collected the data. JU performed the statistical analyses. H. Tamune and JU
269	drafted the first version of the manuscript. All authors critically revised the manuscript
270	for intellectual content and approved the final version.
271	
272	8. Data Availability Statements
273	The datasets and source code utilized in the current study are available from the
274	corresponding author upon reasonable request.
275	
276	9. Conflict of Interest Statement
277	The authors declare no conflict of interest, except for a scholarship grant
278	awarded to JU from Takeda Science Foundation and Masayoshi Son Foundation.
279	
280	

281 **References**

- 1. Harper C. Thiamine (vitamin B1) deficiency and associated brain damage is still
- common throughout the world and prevention is simple and safe! *Eur J Neurol.* (2006)
- 284 13: 1078–1082.
- 285 2. Reynolds E. Vitamin B12, folic acid, and the nervous system. *Lancet Neurol.* (2006)
 5: 949–960.
- 287 3. Arai M, Yuzawa H, Nohara I, Ohnishi T, Obata N, Iwayama Y et al. Enhanced
- carbonyl stress in a subpopulation of schizophrenia. Arch Gen Psychiatry (2010) 67:
- 289 **589–597**.
- 4. Cao B, Wang DF, Xu MY, Liu YQ, Yan LL, Wang JY et al. Lower folate levels in
- schizophrenia: A meta-analysis. *Psychiatry Res.* (2016) 245: 1–7.
- 5. Firth J, Carney R, Stubbs B, Teasdale SB, Vancampfort D, Ward PB, et al. Nutritional
- 293 deficiencies and clinical correlates in first-episode psychosis: A systematic review and
- 294 meta-analysis. *Schizophr Bull.* (2018) 44: 1275–1292.
- 6. Levine J, Stahl Z, Sela BA, Ruderman V, Shumaico O, Babushkin I et al.
- 296 Homocysteine-reducing strategies improve symptoms in chronic schizophrenic patients
- with hyperhomocysteinemia. *Biol Psychiatry* (2006) 60: 265–269.
- 298 7. Firth J, Stubbs B, Sarris J, Rosenbaum S, Teasdale S, Berk M et al. The effects of
- vitamin and mineral supplementation on symptoms of schizophrenia: A systematic
- 300 review and meta-analysis. *Psychol Med.* (2017) 47: 1515–1527.
- 8. Itokawa M, Miyashita M, Arai M, Dan T, Takahashi K, Tokunaga T et al.
- 302 Pyridoxamine: A novel treatment for schizophrenia with enhanced carbonyl stress.
- 303 *Psychiatry Clin Neurosci.* (2018) 72: 35–44.
- 304 9. Koutsouleris N, Kahn RS, Chekroud AM, Leucht S, Falkai P, Wobrock T et al.
- 305 Multisite prediction of 4-week and 52-week treatment outcomes in patients with
- 306 first-episode psychosis: A machine learning approach. Lancet Psychiatry (2016) 3:
- 307 **935–946**.

- 308 10. Mechelli A, Lin A, Wood S, McGorry P, Amminger P, Tognin S et al. Using clinical
- 309 information to make individualized prognostic predictions in people at ultra-high risk
- 310 for psychosis. *Schizophr Res.* (2017) 184: 32-38.
- 311 11. Vieira S, Pinaya WH, Mechelli A. Using deep learning to investigate the
- 312 neuroimaging correlates of psychiatric and neurological disorders: Methods and
- applications. *Neurosci Biobehav Rev.* (2017) 74: 58–75.
- 12. Mayo Foundation for Medical Education and Research, Rochester Test Catalog.
- 315 https://www.mayomedicallaboratories.com/test-catalog/ (2018).
- 13. Friedman, JH. Greedy function approximation: A gradient boosting machine. *Ann.*
- 317 *Stat.* (2001) 29: 1189–1232.
- 14. Sasaki T, Yukizane T, Atsuta H, Ishikawa H, Yoshiike T, Takeuchi T et al. A case of
- thiamine deficiency with psychotic symptoms: Blood concentration of thiamine and
- 320 response to therapy. Seishin Shinkeigaku Zasshi (2010) 112: 97–110.
- 321 15. Clarke R, Refsum H, Birks J, Evans JG, Johnston C, Sherliker P et al. Screening for
- vitamin B-12 and folate deficiency in older persons. *Am J Clin Nutr.* (2003) 77:
 1241–1247.
- 16. Goff DC, Bottiglieri T, Arning E, Shih V, Freudenreich O, Evins AE et al. Folate,
- homocysteine, and negative symptoms in schizophrenia. Am J Psychiatry (2004) 161:
- 326 1705–1708.
- 327 17. Muntjewerff JW, Kahn RS, Blom HJ, den Heijer M. Homocysteine,
- 328 methylenetetrahydrofolate reductase and risk of schizophrenia: A meta-analysis. *Mol*
- 329 *Psychiatry* (2006) 11: 143–149.
- 18. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining
- explanations: An overview of interpretability of machine learning. In 2018 IEEE 5th
- 332 International Conference on Data Science and Advanced Analytics (DSAA): 80–89.
- 19. Evans TC, Jehle D. The red blood cell distribution width. J Emerg Med. (1991) 9:
- 334 **71–74**.

Vitamin deficiency predicted by ML

- 335 20. So HC, Chau CK, Chiu WT, Ho KS, Lo CP, Yim SH et al. Analysis of genome-wide
- association data highlights candidates for drug repositioning in psychiatry. *Nat Neurosci.*
- 337 (2017) 20: 1342–1349.
- 338 21. Ueland PM, Ulvik A, Rios-Avila L, Midttun Ø, Gregory JF. Direct and functional
- biomarkers of vitamin B6 status. *Annu Rev Nutr.* (2015) 35: 33–70.
- 340 22. Carmel R, Lau KH, Baylink DJ, Saxena S, Singer FR. Cobalamin and
- 341 osteoblast-specific proteins. *N Engl J Med.* (1977) 319: 70–75.
- 342 23. Stabler SP. Vitamin B12 deficiency. *N Engl J Med.* (2013) 368: 149–160.
- 343 24. Zheng Y, Cantley LC. Toward a better understanding of folate metabolism in health
- and disease. J Exp Med. (2018) 216: 253–266.

346 Supporting Material List

- Supplementary Table 1 (related to Table 1). Divided patient distribution data (n = 497)
- 348 Supplementary Table 2 (related to Table 2). Divided data of vitamin B deficiencies in
- 349 sub-groups
- 350 Supplementary Table 3 (related to Table 3). Divided dataset of age, sex, and 29
- 351 parameters
- 352 Supplementary Table 4 (related to Table 4). Summary of sensitivity, specificity, and
- accuracy for the training set
- 354 Supplementary Table 5 (related to Table 4). Sensitivities and specificities at other
- 355 operating points
- 356 Supplementary Table 6. Subgroup analyses
- 357 Supplementary Table 7. AUC with different cut-off values

359 Legends

360 Figure 1: Graphical illustration of method of machine-learning

361

Figure 2: Histogram and ROC curves of each vitamin B value

- 363 (A-C) The histograms for vitamin B_1 , vitamin B_{12} , and folate (vitamin B_9).
- 364 Their medians (1st-3rd quartile) are 35 (30-42) ng/mL, 285 (206-431) ng/L, and 7.2
- 365 (4.9–10.8) µg/L, respectively.
- 366 (D-F) ROC curves for vitamin B₁, vitamin B₁₂, and folate. Operating points
 367 used in **Table 4** and **Supplementary Table 5** are depicted in blue.
- 368 Abbreviations: Vit B_1 , vitamin B_1 ; Vit B_{12} , vitamin B_{12} .
- 369

370 Figure 3: Gini importance and partial dependence plots of vitamin B deficiencies

- The Gini importance (A-C) and partial dependency plots of the probability of 371deficiency (D-F) are shown for the eight most important variables for vitamin B_1 , 372vitamin B_{12} , and folate (vitamin B_9). Combined with these, this machine-learning 373374classifier without hypothesis also provided further evidence of a relationship between 375vitamin B levels and the complete blood count while also indicating a potential 376association between these vitamins and alkaline phosphatase (ALP) or thyroid-stimulating hormone (TSH). 377
- Abbreviations: Vit B_1 , vitamin B_1 ; Vit B_{12} , vitamin B_{12} ; Hb, hemoglobin; Hct, 378 hematocrit; WBC, white blood cell count; CK, creatine kinase; RDW.CV, red blood cell 379 distribution width-coefficient variation; Plt, platelet; ALT, alanine transaminase; Lym, 380 Cre. creatinine; Neu, 381 lymphocyte fraction; neutrocyte fraction; γGTP , γ -glutamyltransferase; MCV, mean corpuscular volume; glu, plasma glucose. 382

383

Figure 4: Graphical abstract

Table 1. Patient distribution data (n = 497)

387

	ICD-10 code							VitB ₁		VitB ₁₂		Folate							
	F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	<20	<28	<30*	<150	<180*	<200	<3.0	<4.0*	⁴ <5.0
N	28	21	300	58	16	0	29	20	24	1	15	81	112	37	80	107	29	72	134
%	5.6	4.2	60.4	11.7	3.2	0.0	5.8	4.0	4.8	0.2	3.0	16.3	22.5	7.4	16.1	21.5	5.8	14.5	27.0

388

389 Asterisks show the predefined cut-off values for vitamin B_1 , vitamin B_{12} , and folate

390 (vitamin B₉) based on a reference¹²; different cut-off values based on other papers^{14–16}

are also presented for further investigation.

392

393 ICD-10 codes. F0, Mental disorders due to known physiological conditions; F1, Mental 394 and behavioral disorders due to psychoactive substance use; F2, Schizophrenia, 395schizotypal, delusional, and other non-mood psychotic disorders; F3, Mood disorders; 396 F4, Anxiety, dissociative, stress-related, somatoform, and other non-psychotic mental 397disorders; F5, Behavioral syndromes associated with physiological disturbances and 398physical factors; F6, Disorders of adult personality and behavior; F7, Intellectual 399 disabilities; F8, Pervasive and specific developmental disorders; F9, Behavioral and 400 emotional disorders with onset usually occurring in childhood and adolescence. 401

402 **Table 2. Vitamin B deficiencies in sub-groups**

403	3
-----	---

	F0	F1	F2	F3	F4	F6	F7	F8	F9
vitB ₁ < 30	9	4	70 (23%)	11 (19%)	3	7	5	3	0
	(32%)	(19%)			(19%)	(24%)	(25%)	(13%)	
$vitB_{12} < 180$	5	4	53 (18%)	7 (12%)	3	1 (3%)	4	3	0
	(18%)	(19%)			(19%)		(20%)	(13%)	
Folate < 4.0	5	7	38 (13%)	6(10%)	5	3	4	4	0
	(18%)	(33%)			(31%)	(10%)	(20%)	(17%)	

404

405 Abbreviations; see **Table 1**.

406

408 Table 3. Summary of full dataset of age, sex, and 29 parameters for

409 machine-learning

410

Parameters	Units	Mean	SD	UN	mg/dL	12.9	6.7
Age	years	42.3	15.4	Cre	mg/dL	0.7	0.2
Sex	Woman	n = 228		T.bil	mg/dL	0.7	0.4
WBC	$\times 10^3 / \mu L$	8.2	2.8	Na	mmol/L	139	3
Hb	g/dL	13.7	1.7	Cl	mmol/L	105	4
Hct	%	40.3	4.5	K	mmol/L	3.7	0.4
MCV	fL	89	6.6	cor.Ca	mg/dL	9.1	0.5
Plt	$\times 10^4\!/\mu L$	24.9	6.3	СК	IU/L	514	1230
RDW.CV	%	13.5	1.3	AST	IU/L	31	34
Neu	%	70	11	ALT	IU/L	27	24
Lym	%	23	10	LDH	IU/L	239	91
Mono	%	6	2	ALP	IU/L	224	81
Eo	%	1	2	γGTP	IU/L	37	63
Baso	%	0	0	Glu	mg/dL	112	40
TP	g/dL	7.2	0.6	CRP	mg/dL	0.4	0.9
Alb	g/dL	4.4	0.4	TSH	µIU/mL	1.7	2.4

411

412 Two patients lacked age data (no photo ID was available), and one patient lacked
413 biochemistry data (inappropriate sample processing). For machine-learning, the missing
414 values were replaced using the mean.

415

Abbreviations: WBC, white blood cell count; Hb, hemoglobin; Hct, hematocrit; MCV,
mean corpuscular volume; RDW.CV, red blood cell distribution width-coefficient
variation; Plt, platelet; Neu, neutrocyte fraction; Lym, lymphocyte fraction; Mono,
monocyte fraction; Eo, eosinocyte fraction; Baso, basocyte fraction; TP, total protein;
Alb, albumin; UN, urea nitrogen; Cre, creatinine; T.bil, total bilirubin; Na, sodium; Cl,
Vitamin deficiency predicted by ML

- 421 chloride; K, potassium; cor.Ca, corrected calcium; CK, creatine kinase; AST, aspartate
- 422 transaminase; ALT, alanine transaminase; LDH, lactate dehydrogenase; ALP, alkaline
- 423 phosphatase; γGTP, γ-glutamyltransferase; Glu, plasma glucose; CRP, C-reactive
- 424 protein; TSH, thyroid-stimulating hormone.
- 425

Table 4. Summary of AUC, sensitivity, specificity, and accuracy for the validation

427 **set**

428

	vitB ₁	vitB ₁₂	Folate
AUC	0.716	0.599	0.796
Sensitivity	0.594	0.316	0.667
Specificity	0.783	0.943	0.917
Accuracy	0.688 [0.597–0.787]	0.629 [0.523–0.746]	0.792 [0.665–0.909]

429

Generalization performance of the classifiers was evaluated using AUC of the validation set. Sensitivity, specificity, and accuracy of the classification at the optimal operating points that maximized accuracy on the receiver operating characteristic curve of the validation set are also shown (see also **Figure 2 D-F**). Accuracy was defined as the average of the sensitivity and specificity. Square brackets indicate the 95% CI. Note that the 95% CI of each accuracy does not include 0.5, which demonstrates statistical significance. For further information, see **Figure 2** and **Supplementary Table 5**.

437

438 Abbreviations: AUC, area under the receiver operating characteristic curve; CI,439 confidence interval.



(A) Vit B₁









(D) Vit B₁

1.0 True Positive Rate 0.2 0.0 0.2 0.4 0.6 0. False Positive Rate 0.8 1.0

(E) Vit B₁₂

180

0



(F) Folate





