ic Infection & Chemotherapy

Original Article

Check for updates

Machine-Learning-Based Approach to Differential Diagnosis in Tuberculous and Viral Meningitis

Young-Seob Jeong (iD) [1], **Minjun Jeon** (iD) [2], **Joung Ha Park** (iD) [3], **Min-Chul Kim** (iD) [3,4], **Eunyoung Lee** (iD) [5,6], **Se Yoon Park** (iD) [5], **Yu-Mi Lee** (iD) [7], **Sungim Choi** (iD) [8], **Seong Yeon Park** (iD) [8], **Ki-Ho Park** (iD) [7], **Sung-Han Kim** (iD) [3], **Min Huok Jeon** (iD) [9], **Eun Ju Choo** (iD) [10], **Tae Hyong Kim** (iD) [5], **Mi Suk Lee** (iD) [7], and **Tark Kim** (iD) [10]

[1]Big Data Engineering department, Soonchunhyang University, Asan, Korea
[2]R&D department, PharmCADD, Busan, Korea
[3]Department of Infectious Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea
[4]Division of Infectious Diseases, Department of Internal Medicine, Chung-Ang University Hospital, Seoul, Korea
[5]Department of Internal Medicine, Soonchunhyang University Seoul Hospital, Seoul, Korea
[6]Division of Infectious Diseases, Department of Internal Medicine, Korea Institute of Radiological & Medical Sciences, Seoul, Korea
[7]Department of Internal Medicine, Kyung Hee University Hospital, Kyung Hee University School of Medicine, Seoul, Korea
[8]Division of Infectious Diseases, Dongguk University Ilsan Hospital, Goyang, Korea
[9]Department of Internal Medicine, Soonchunhyang University Cheonan Hospital, Cheonan, Korea
[10]Department of Internal Medicine, Soonchunhyang University Bucheon Hospital, Bucheon, Korea

OPEN ACCESS

Corresponding Author:
Tark Kim, MD, PhD
Department of Internal Medicine,
Soonchunhyang University Bucheon Hospital,
170 Jomaru-ro, Bucheon 14584, Gyeonggi-do,
Korea.
E-mail: ktocc2@naver.com

ORCID iDs
Young-Seob Jeong (iD)
https://orcid.org/0000-0002-9441-2940
Minjun Jeon (iD)
https://orcid.org/0000-0002-5964-0104
Joung Ha Park (iD)
https://orcid.org/0000-0003-0480-3903
Min-Chul Kim (iD)
https://orcid.org/0000-0003-4410-5608
Eunyoung Lee (iD)
https://orcid.org/0000-0001-8280-3605
Se Yoon Park (iD)
https://orcid.org/0000-0002-4538-7371
Yu-Mi Lee (iD)
https://orcid.org/0000-0002-0486-6608
Sungim Choi (iD)
https://orcid.org/0000-0001-8692-5570

ABSTRACT

Background: Tuberculous meningitis (TBM) is the most severe form of tuberculosis, but differentiating between the diagnosis of TBM and viral meningitis (VM) is difficult. Thus, we have developed machine-learning modules for differentiating TBM from VM.

Material and Methods: For the training data, confirmed or probable TBM and confirmed VM cases were retrospectively collected from five teaching hospitals in Korea between January 2000 - July 2018. Various machine-learning algorithms were used for training. The machine-learning algorithms were tested by the leave-one-out cross-validation. Four residents and two infectious disease specialists were tested using the summarized medical information.

Results: The training study comprised data from 60 patients with confirmed or probable TBM and 143 patients with confirmed VM. Older age, longer symptom duration before the visit, lower serum sodium, lower cerebrospinal fluid (CSF) glucose, higher CSF protein, and CSF adenosine deaminase were found in the TBM patients. Among the various machine-learning algorithms, the area under the curve (AUC) of the receiver operating characteristics of artificial neural network (ANN) with ImperativeImputer for matrix completion (0.85; 95% confidence interval 0.79 - 0.89) was found to be the highest. The AUC of the ANN model was statistically higher than those of all the residents (range 0.67 - 0.72, $P < 0.001$) and an infectious disease specialist (AUC 0.76; $P = 0.03$).

Conclusion: The machine-learning techniques may play a role in differentiating between TBM and VM. Specifically, the ANN model seems to have better diagnostic performance than the non-expert clinician.

Keywords: Tuberculosis; Virus; Meningitis; Machine learning; Diagnosis

Generated by xmlinkpress

Infection &
Chemotherapy

Seong Yeon Park
https://orcid.org/0000-0001-8762-7006
Ki-Ho Park
https://orcid.org/0000-0003-0000-9016
Sung-Han Kim
https://orcid.org/0000-0002-6596-8253
Min Huok Jeon
https://orcid.org/0000-0002-9452-4379
Eun Ju Choo
https://orcid.org/0000-0003-2842-7041
Tae Hyong Kim
https://orcid.org/0000-0003-2920-9038
Mi Suk Lee
https://orcid.org/0000-0001-8951-5032
Tark Kim
https://orcid.org/0000-0002-8829-4183

Conflict of Interest
No conflicts of interest.

Author Contributions
Conceptualization: JYS, KT. Data curation:
PJH, KMC, LE, PSY, LYM, CS, PSY, PKH, KSH,
JMH, CEJ, KTH, LMS, KT. Formal analysis: JYS,
JM, KT. Methodology: JYS, KT, KSH. Writing-
original draft: JYS, KT. Writing-review, and
editing: KSH, KTH.

INTRODUCTION

Tuberculous meningitis (TBM) is the most severe form of tuberculosis and causes inflammation of the meninges. TBM accounts for approximately 1% of all cases of TB and 5% of all extrapulmonary disease in immunocompetent individuals [1]. More than 100,000 new TBM cases are estimated to occur globally yearly [1]. Patients with TBM initially complain of an insidious onset of malaise, lassitude, headache, and low-grade fever. Nausea, vomiting, and confused mental state occur as the disease progresses, thereby leading to coma, seizures, and neurological damage. Despite the advancements in medicine, the case-fatality ratio is high and early diagnosis and anti-tuberculous therapy are critical for patients with TBM [2].

However, the diagnosis of TBM is markedly challenging. In the early phase, it may be difficult to differentiate between TBM and viral meningitis (VM) because of similar clinical manifestations. Additionally, the diagnostic tools for TBM and VM show low sensitivity. The isolation of acid-fast bacilli in the cerebrospinal fluid (CSF) as a rapid and specific method for diagnosing TBM has poor sensitivity, reported as low as 30% [3]. The CSF mycobacterial culture that is regarded as a definitive diagnostic tool for TBM also has low sensitivity and requires incubation of up to two months. Nucleic acid amplification tests (NAATs) and Xpert MTB/Rif polymerase chain reaction (PCR)-based assays are also not sensitive tools for the diagnosis of TBM, although they are highly specific [4, 5]. For these reasons, the differential diagnosis of TBM and VM depends on the judgment of the clinician.

Machine-learning techniques are very useful for resolving problems of discrimination and the development of diagnostic modules using machine-learning is an active research topic in the field of medicine. The differentiation between TBM and VM is also a matter of discrimination, so machine-learning techniques are expected to play auxiliary roles in guiding the diagnosis in situations where quick judgment is required. Deep learning is especially attractive because of its superior performance (*e.g.*, accuracy) compared to many existing machine-learning models. Nevertheless, to the best of our knowledge, there exist no studies on the diagnostic role of deep learning techniques to differentiate between TBM and VM.

Thus, we investigated various machine-learning techniques, including deep learning models, for differentiating TBM from VM and compared the results with diagnoses made by clinicians.

MATERIALS AND METHODS

1. Study design & population

A retrospective cohort study was conducted at five teaching hospitals in the Korea (range of bed numbers, 642 – 2,705). The medical records of the cases between January 2000 - July 2018 that met the following conditions were reviewed: 1) *Mycobacterim tuberculosis* growth from CSF, 2) positive nucleic acid amplification tests for *M. tuberculosis*, 3) diagnostic code of tuberculous meningitis (KCD A170), and 4) positive nucleic acid amplification tests for herpes simplex virus (HSV), varicella-zoster virus (VZV), or enterovirus. The exclusion criteria were incomplete medical records for review, lack of CSF fluid analysis, and patients younger than 18 years of age. This study was approved by the Institutional Review Board of Soonchunhyang University Bucheon Hospital (2018-09-026). The requirement for informed consent was waived because of the retrospective nature of this study, no patient interventions, and no additional specimen collections. All procedures involving human

participants were performed according to the ethical standards of the institutional and/or national research committees and in accordance with the 2013 Declaration of Helsinki and its later amendments or comparable ethical standards.

2. Definition and data collection

Meningitis was defined as CSF white blood cell (WBC) count >5 cells/mm^3 with two or more of the following findings: headache, nausea/vomiting, photophobia, neck stiffness, and fever >38ºC. Patients whose clinical presentation was indicative of meningitis and with positive CSF PCR results for HSV, VZV, or enterovirus PCR had confirmed viral meningitis. Patients whose clinical presentation was indicative of CNS infection had confirmed TBM if the CSF specimens were positive for *M. tuberculosis* based on the culture or the PCR assay. Patients whose clinical presentation was indicative of CNS infection plus a culture of other body fluids was positive for *M. Tuberculosis*, and without other known etiologies of meningitis, had probable TBM.

Since the number of cases was not large enough, only a limited number of features known as useful characteristics for differentiating between TBM and VM were searched and included in the analysis because of the overfitting concern. Based on the previous studies, data on age [6], duration of illness from the appearance of symptoms and onset of signs to hospital visit [7], vomiting [7], neurologic symptoms and signs [6], serum sodium [6], CSF glucose [8], CSF protein[8], and CSF adenosine deaminase (ADA) [9] were collected as discriminative features for machine-learning. Neurologic symptoms and signs were defined according to one of the following symptoms or signs: lethargy, confusion, cranial nerve palsy, hemiparesis, delirium, stupor, coma, seizures, hemiplegia, or paraparesis [6].

3. Model development and validation

We conducted experiments with different machine-learning models that included naïve Bayes (NB), logistic regression (LR), random forest (RF), support vector machine (SVM), and artificial neural network (ANN) models. The parameter settings of the machine-learning models are summarized in **Table 1**. Note that the ANN model had a simple hierarchical structure with two layers to avoid overfitting. The ANN model is trained with L2 regularization with alpha 0.0001. We applied the take leave-one-out cross-validation (LOO-CV) to evaluate the models, wherein each data plays the role of test data, while the other remaining data are used for training, equivalent to 203-fold cross-validation. The experiments with all machine-learning models, except for the ANN, were conducted using the Weka tool, whereas the experiments with the ANN were performed using Tensorflow 1.12

Table 1. Parameter settings of the machine-learning models

Model	Setting
Random forest	- Maximum number of trees: 100
Naïve Bayes	- No kernel estimator, so it uses a normal distribution
Logistic regression	- Ridge: 1.0×10^{-8} - Training algorithm: Broyden–Fletcher–Goldfarb–Shanno
Support vector machine	- Training algorithm: Sequential minimal optimization - C: 1.0 - Epsilon: 1.0×10^{-12} - Kernel: PolyKernel (exponent: 1.0)
Artificial neural network	- Hidden layers: [20, 5] - Activation function: ReLU - Number of epoch: 250 - Training algorithm: Adam (initial learning rate: 0.001)

(Google, San Francisco, CA, USA). The experiments were conducted employing a computer of eight CPUs of i7-7700 3.6 GHz and two NVIDIA Geforce 1080 Ti.

Besides the machine-learning models, we collected results from six human clinicians: four residents in their fourth year of internal medicine and two board-certified infectious disease (ID) specialists with more than 10 years of experience. The summary of non-imputed medical information of each patient was shown to the the clinicians and they were asked to estimate the diagnosis (**Supplementary Fig. 1**).

4. Statistics

All statistical analyses were calculated using the SPSS Statistics version 25.0 (SPSS, Chicago, IL, USA) and the MedCalc version 19.3 (MedCalc Software Ltd., Ostend, Belgium). The categorical variables were compared using the Chi-squared test or the Fisher's exact test. The continuous variables were analyzed using the Mann-Whitney U test. A non-significant Little's missing complexly at random (MCAR) test, $\chi^2 = 27.244$, $df = 22$, $P = 0.20$, indicated an MCAR pattern. As the number of data was small, we retained all data by applying imputation algorithms. When we adopt an imputation algorithm to fill the missing values, the machine learning model performance (*e.g.*, accuracy) depends on the imputation algorithm. So, we applied three imputation algorithms: 1) an imputing strategy of iterative round-robin fashion (IterativeImputer), 2) the iterative soft thresholding of Singular Vector Decomposition (SoftImputer), and 3) imputing by K-nearest neighbor (KnnImputer). The Delong method was used to calculate the area under the curve (AUC) of the receiver operator characteristic. The machine-learning model with the highest AUC value was chosen for comparison with human judgment, although the difference in the AUC values between machine learning models was statistically insignificant. Bootstrapping was used to compare the AUC between machine-learning and human judgment. Cohen's kappa statistics were used to analyze the diagnostic agreement. All tests were two-tailed and differences were significant at P <0.05.

RESULTS

As shown in **Figure 1**, 234 patients were excluded and a total of 60 patients with confirmed or probable TBM and 143 patients with confirmed VM were included for training. Of 39 patients with confirmed TBM, *the M. tuberculous* complex was cultured in the CSF in only five patients. The median annual number of TBM was 3 patients (interquartile range [IQR], 1 - 7 patients). Of 144 VM, viral etiologies were as follows: HSV in 49 (34.0%) patients, VZV in 65 (45.1%) patients, and enterovirus in 29 (20.1%) patients. As shown in **Table 2**, the median age of the patients was 37 years (IQR, age 29 - 58 years), and the median duration of illness before the visit was five days (IQR, 3 - 7 days). Typically, 80% of patients complained of vomiting and 70% of patients had neurologic symptoms and signs. Older age, longer duration of illness before hospital visits, frequent neurologic symptoms and signs, lower serum sodium, lower CSF glucose, higher CSF protein, and higher CSF ADA were reported in TBM patients compared to VM patients.

The number of missing values was as follows: one for the duration of illness, one for vomiting, two for serum sodium, one for CSF glucose, one for protein, and 16 for CSF ADA. As shown in **Table 3**, all machine-learning models except SVM achieved the highest accuracy with ItrerativeImputer for matrix completion. Among the machine-learning models, the NB with SoftImputer had the highest sensitivity (80.0%; 95% confidence interval [CI], 67.7 -
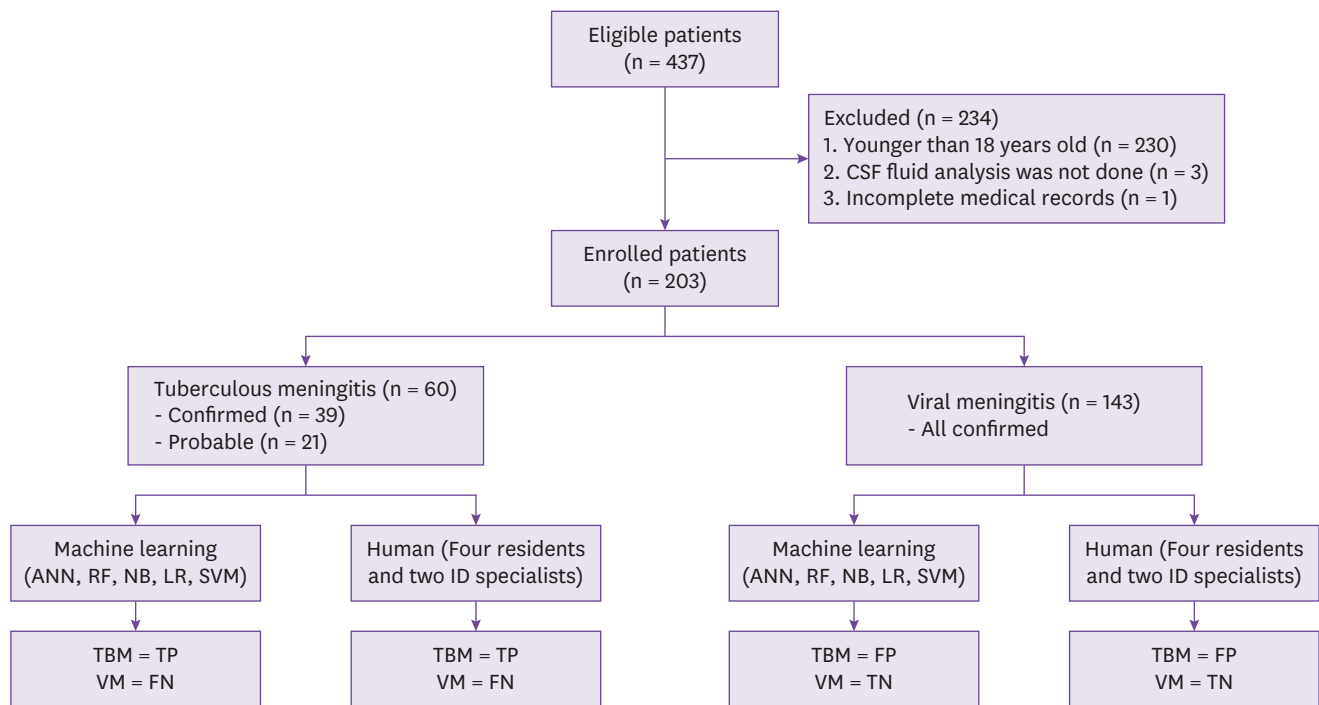
Figure 1. Flow chart of the study.
Patients whose clinical presentation was indicative of meningitis and with a positive CSF PCR result for HSV, VZV, or enterovirus PCR had confirmed viral meningitis. Patients whose clinical presentation was indicative of CNS infection had confirmed TBM if the CSF specimens were positive for *Mycobacterim tuberculosis* by culture or PCR assay. Patients whose clinical presentation was indicative of CNS infection plus a culture of other body fluids was positive for *M. Tuberculosis*, without other known etiologies of meningitis, had probable TBM. True positive means a correct diagnosis of tuberculous meningitis and true negative means a correct diagnosis of viral meningitis.
ANN, artificial neural network; RF, random forest; NB, naïve Bayes; LR, logistic regression; SVM, support vector machine; ID, infectious diseases; TBM, tuberculous meningitis; TP, true positive; VM, viral meningitis; FN, false negative; FP, false positive; TN, true negative.

Table 2. Comparison of features between tuberculous and viral meningitis

Features	All (N = 203)	Tuberculous (N = 60)	Viral (N = 143)	*P*-value
Median age, years (IQR)	37 (29 - 58)	49 (33 - 64)	34 (29 - 55)	<0.001
Median symptom duration before the visit, days (IQR)	5 (3 - 7)	9 (6 - 15)	4 (2 - 6)	<0.001
Vomiting (%)	80 (39.4)	28 (46.7)	52 (36.4)	0.21
Neurologic symptoms and signs (%)	70 (34.5)	39 (65.0)	31 (21.5)	<0.001
Median serum sodium, mg/dl (IQR)	137 (134 - 139)	133 (128 - 136)	138 (136 - 140)	<0.001
Median CSF glucose, mg/dl (IQR)	53.3 (45.1 - 66.0)	41.6 (28.8 - 61.5)	57.6 (49.0 - 67.0)	<0.001
Median CSF protein, mg/dl (IQR)	117.0 (67.9 - 169.6)	175.5 (118.7 - 317.1)	101 (56.3 - 141.3)	<0.001
Median CSF ADA, IU/L (IQR)	7 (3 - 12)	14 (8 - 21)	5 (3 - 8)	<0.001

IQR, interquartile range; CSF, cerebrospinal fluid; ADA, adenosine deaminase.

89.2%) and SVM had the highest specificity (97.2%; 95% CI, 93.0 - 99.2%). However, ANN with IterativeImputer, LR with IterativeImputer, LR with KnnImputer (K = 2), and LR with KnnImputer (K = 3) had the highest accuracy (87.7%; 95% CI, 82.4 - 91.9%). The highest AUC value for differentiating TBM from VM was found in ANN with IterativeImputer among machine-learning models (0.85; 95% CI, 0.79 - 0.89).

The diagnostic performance of humans for differentiating TBM from VM is shown in **Table 4** and **Fig. 2**. The ANN with IterativeImputer model was chosen as a machine-learning model for the comparison with humans because it showed the highest value of AUC among the machine-learning models. Residents tended to diagnose less sensitively as TBM than ID specialists. The sensitivity of the residents was less than 53.7%, while that of the ID

Table 3. Diagnostic performances of various machine-learning algorithms for differentiating tuberculous from viral meningitis

Machine-learning algorithm	Matrix completion	TP	FP	TN	FN	Sensitivity (% [95% CI])	Specificity (% [95% CI])	Accuracy (% [95% CI])	AUC (95% CI)
Artificial neural network	IterativeImputer	46	11	132	14	76.7 (63.9 - 86.6)	92.3 (86.7 - 96.1)	87.7 (82.4 - 91.9)	0.85 (0.79 - 0.89)
	SoftImputer	41	20	123	19	68.3 (55.0 - 79.7)	86.0 (79.2 - 91.2)	80.8 (74.7 - 86.0)	0.77 (0.71 - 0.83)
	KnnImputer (K = 1)	43	11	132	17	71.7 (58.6 - 82.5)	92.3 (86.7 - 96.1)	86.2 (80.7 - 90.6)	0.82 (0.76 - 0.87)
	KnnImputer (K = 2)	43	10	133	17	71.7 (58.6 - 82.5)	93.0 (87.5 - 96.6)	86.7 (81.2 - 91.0)	0.82 (0.76 - 0.87)
	KnnImputer (K = 3)	42	12	131	18	70.0 (56.8 - 81.2)	91.6 (85.8 - 95.6)	85.2 (79.6 - 89.8)	0.81 (0.75 - 0.86)
	KnnImputer (K = 4)	42	12	131	18	70.0 (56.8 - 81.2)	91.6 (85.8 - 95.6)	85.2 (79.6 - 89.8)	0.81 (0.75 - 0.86)
Random forest	IterativeImputer	42	11	132	18	70.0 (56.8 - 81.2)	92.3 (86.7 - 96.1)	85.7 (80.1 - 90.2)	0.81 (0.75 - 0.86)
	SoftImputer	38	9	134	22	63.3 (49.9 - 75.4)	93.7 (88.4 - 97.1)	84.7 (79.0 - 89.4)	0.79 (0.72 - 0.84)
	KnnImputer (K = 1)	40	13	130	20	66.7 (53.3 - 78.3)	90.9 (85.0 - 95.1)	83.7 (77.9 - 88.5)	0.79 (0.73 - 0.84)
	KnnImputer (K = 2)	41	11	132	19	67.8 (54.4 - 79.4)	91.3 (86.7 - 96.1)	85.2 (79.5 - 89.8)	0.80 (0.74 - 0.85)
	KnnImputer (K = 3)	42	14	129	18	70.0 (56.8 - 81.2)	90.2 (84.1 - 94.5)	84.2 (78.5 - 89.0)	0.80 (0.74 - 0.85)
	KnnImputer (K = 4)	40	11	132	20	66.7 (53.3 - 78.3)	92.3 (86.7 - 96.1)	84.7 (79.0 - 89.4)	0.80 (0.73 - 0.85)
Naïve Bayes	IterativeImputer	36	11	132	24	60.0 (46.5 - 72.4)	92.3 (86.7 - 96.1)	82.8 (76.8 - 87.7)	0.76 (0.70 - 0.82)
	SoftImputer	48	24	119	12	80.0 (67.7 - 89.2)	83.2 (76.1 - 88.9)	82.3 (76.3 - 87.3)	0.82 (0.76 - 0.87)
	KnnImputer (K = 1)	38	13	130	22	63.3 (49.9 - 75.4)	90.9 (85.0 - 95.1)	82.8 (76.9 - 87.7)	0.77 (0.71 - 0.83)
	KnnImputer (K = 2)	39	13	130	21	65.0 (51.6 - 76.9)	90.9 (85.0 - 95.1)	83.3 (77.4 - 88.1)	0.78 (0.72 - 0.84)
	KnnImputer (K = 3)	38	13	130	22	63.3 (49.9 - 75.4)	90.9 (85.0 - 95.1)	82.8 (76.9 - 87.7)	0.77 (0.71 - 0.83)
	KnnImputer (K = 4)	38	13	130	22	63.3 (49.9 - 75.4)	90.9 (85.0 - 95.1)	82.8 (76.9 - 87.7)	0.77 (0.71 - 0.83)
Logistic regression	IterativeImputer	44	9	134	16	73.3 (60.3 - 83.9)	93.7 (88.4 - 97.1)	87.7 (82.4 - 91.9)	0.84 (0.78 - 0.88)
	SoftImputer	43	15	128	17	71.7 (58.6 - 82.5)	89.5 (83.3 - 94.0)	84.2 (78.5 - 89.0)	0.81 (0.75 - 0.86)
	KnnImputer (K = 1)	42	9	134	18	70.0 (56.8 - 81.2)	93.7 (88.4 - 97.1)	86.7 (81.2 - 91.0)	0.82 (0.76 - 0.87)
	KnnImputer (K = 2)	42	7	136	18	70.0 (56.8 - 81.2)	95.1 (90.2 - 98.0)	87.7 (82.4 - 91.9)	0.83 (0.77 - 0.88)
	KnnImputer (K = 3)	42	7	136	18	70.0 (56.8 - 81.2)	95.1 (90.2 - 98.0)	87.7 (82.4 - 91.9)	0.83 (0.77 - 0.88)
	KnnImputer (K = 4)	41	7	136	19	68.3 (55.0 - 79.7)	95.1 (90.2 - 96.1)	87.2 (81.8 - 91.5)	0.82 (0.76 - 0.87)
Support vector machine	IterativeImputer	34	6	137	26	56.7 (43.2 - 69.4)	95.8 (91.1 - 98.5)	84.2 (76.5 - 89.0)	0.76 (0.70 - 0.82)
	SoftImputer	45	17	126	15	75.0 (62.1 - 85.3)	88.1 (81.6 - 92.9)	84.2 (78.5 - 89.0)	0.82 (0.76 - 0.87)
	KnnImputer (K = 1)	33	4	139	27	55.0 (41.6 - 67.9)	97.2 (93.0 - 99.2)	84.7 (79.0 - 89.4)	0.76 (0.70 - 0.82)
	KnnImputer (K = 2)	34	8	135	26	56.7 (43.2 - 69.4)	94.4 (89.3 - 97.6)	83.3 (77.4 - 88.1)	0.76 (0.69 - 0.81)
	KnnImputer (K = 3)	34	8	135	26	56.7 (43.2 - 69.4)	94.4 (89.3 - 97.6)	83.3 (77.4 - 88.1)	0.76 (0.69 - 0.81)
	KnnImputer (K = 4)	35	7	136	25	58.3 (44.9 - 70.9)	95.1 (90.2 - 98.0)	84.3 (78.5 - 89.0)	0.77 (0.70 - 0.82)

Testing was conducted using the leave-one-out cross-validation.
True positive means a correct diagnosis of tuberculous meningitis and true negative means a correct diagnosis of viral meningitis.
TP, true positive; FP, false positive; TN, true negative; FN, false negative; AUC, area under the receiver operating characteristics curve; 95% CI, 95% confidence interval.

Table 4. Diagnostic performance of humans for differentiating tuberculous from viral meningitis

	TP	FP	TN	FN	Sensitivity (% [95% CI])	Specificity (% [95% CI])	Accuracy (% [95% CI])	AUC (95% CI)	Artificial neural network with IterativeImputer	
									P1[a]	P2[b]
Resident #1	32	20	123	28	53.3 (40.0 - 66.3)	86.0 (79.2 - 91.2)	76.4 (69.9 - 82.0)	0.70 (0.63 - 0.76)	<0.001	0.0002
Resident #2	23	7	136	37	38.3 (26.1 - 51.8)	95.1 (90.2 - 96.0)	78.3 (72.0 - 83.8)	0.67 (0.60 - 0.73)	<0.001	<0.001
Resident #3	31	20	123	29	51.7 (38.4 - 64.8)	86.0 (79.2 - 91.2)	75.9 (69.4 - 81.6)	0.69 (0.62 - 0.75)	<0.001	0.0001
Resident #4	30	9	134	30	50.0 (36.8 - 63.2)	93.7 (88.4 - 97.1)	80.8 (74.7 - 86.0)	0.72 (0.65 - 0.78)	<0.001	0.0004
ID specialist #1	39	18	125	21	65.0 (51.6 - 76.9)	87.4 (80.8 - 92.4)	80.8 (74.7 - 86.0)	0.76 (0.70 - 0.82)	<0.001	0.03
ID specialist #2	46	26	117	14	76.7 (64.0 - 86.6)	81.8 (74.5 - 87.8)	80.3 (74.2 - 85.6)	0.79 (0.73 - 0.85)	<0.001	0.16

[a]Cohen's kappa statistic was used to test the diagnostic agreement between machine-learning and human judgment.
[b]Comparison of the AUC of the machine-learning with that of human judgment.
True positive means a correct diagnosis of tuberculous meningitis and true negative means a correct diagnosis of viral meningitis.
TP, true positive; FP, false positive; TN, true negative; FN, false negative; AUC, area under the receiver operating characteristics curve; 95% CI, 95% confidence interval; ID, infectious disease.

specialists was more than 65%. The AUCs between the residents were not statistically different. Also, the value of AUC was not statistically different between ID specialist #1 and the ID specialist #2 (P = 0.38) (**Supplementary Table 1**). The higher AUCs of the ID specialists were found, although these differences were only statistically significant between the ID specialist #1 and the resident #2 (P = 0.01), the ID specialist #2 and resident #1 (P = 0.02), the ID specialist #2 and resident #2 (P = 0.003), and the ID specialist #2 and resident #3
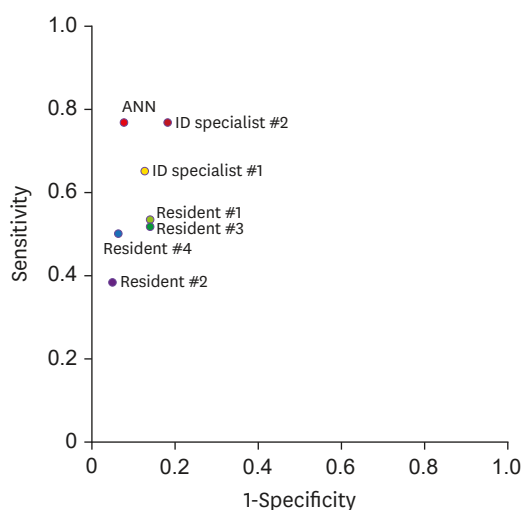
Figure 2. A plot of the diagnostic performance of the machine learning and clinicians for differentiating tuberculous from viral meningitis.
The area under curves (AUC) of receiver operating characteristics between the residents was not statistically different. Also, the value of the AUC was not statistically different between the ID specialist #1 and the ID specialist #2 (P = 0.38). The higher AUCs of the ID specialists were found, although the differences were only statistically significant between the ID specialist #1 and resident #2 (P = 0.01), the ID specialist #2 and resident #1 (P = 0.02), the ID specialist #2 and resident #2 (P = 0.003), and the ID specialist #2 and resident #3 (P = 0.01). The AUC of the ANN model was statistically higher than those of all the residents. Also, the diagnostic performance of the ANN model was statistically higher than the ID specialist #2 (P = 0.03) and comparable to the ID specialist #1 (P = 0.16).
ANN, artificial neural network; LR, logistic regression; ID, infectious diseases.

(P = 0.01) (**Supplementary Table 1**). The AUC of the ANN model was statistically higher than those of all the residents. Also, the diagnostic performance of the ANN model was statistically higher than the ID specialist #2 (P = 0.03) and comparable to the ID specialist #1 (P = 0.16), although the diagnostic agreement was statistically different between the machine-learning and ID specialists.

DISCUSSION

Our findings showed the potential of the machine-learning models, especially the ANN model, to distinguish TBM from VM. Specifically, the diagnostic performance of the machine-learning models was more accurate than that of the non-expert clinicians and was comparable to the judgment of the experts. To the best of our knowledge, this is the first study to differentiate between TBM and VM using various machine-learning models.

There have been several studies on the differential diagnosis between TBM and VM. In a retrospective study by Hristea et al. [6], symptom duration, advanced neurologic status, CSF glucose ratio <0.5, and CSF protein >100 mg/dl were identified to be associated with TBM and not VM. A model using these significant variables showed excellent sensitivity (92%, 95% CI: 87 - 97), specificity (94%, 95% CI: 92 - 97), and the AUC value (0.977, 95% CI: 0.964 - 0.990) for the diagnosis of TBM. Lee also reported that a grade scoring system including the variables of hyponatremia, CSF lactate dehydrogenase >70 IU/L, CSF protein 160 mg/dl, cranial nerve palsy, voiding difficulty, and confusion had 89.4% sensitivity, 90.4% specificity, and 0.901 accuracy (95% CI: 0.839 - 0.963) to differentiate TBM from VM [7]. Despite the excellent diagnostic performance of the previous models, there were some critical

limitations. First, there is a concern of overfitting, because the previous studies comprised of too many variables, despite the small number of cases. Second, the diagnostic performance of these models developed in one center cannot be assured in other centers. Third, almost all the reported studies included cases of possible TBM. Marais et al. proposed uniform definitions of probable and possible TBM [10]. However, the definition of possible TBM showed low specificity [11]. Since the definition of possible TBM involves the same variables included in the diagnostic models, the diagnostic performance of the previous models can be overestimated. Thus, we included only cases with positive culture or PCR results. Our study design increased the reliability of our results, although the diagnostic performance was weaker compared with the outcomes of previous studies. Additionally, the variables included in our model were easily and autonomously accessible in data acquired from the electronic health records. Thus, it should not be difficult to develop a program for application in clinical practice.

As shown in **Table 4**, the ID specialists tended to more sensitively diagnose TBM than the residents, although their sensitivity was not satisfactory. The higher sensitivity demonstrated by the ID specialists may be the result of knowledge and experience with possible TBM cases. Interestingly, this tendency of the ANN model was somewhat similar to the ID specialists, so the ANN model's behavior can be associated with the ID specialists.

There is no statistically significant difference in the AUC values of each machine learning model. Considering the cost and effort required for machine learning, a statistical analysis such as logistic regression may be more favorable than ANN. Statistical analysis is a good method of research, but it is markedly different from the studies of the machine learning field. As described in the report by Bradley A. Fritz [12], the statistical analysis may be practically limited because it renders the knowledge to clinicians but is not suitable for developing real-world applications (*e.g.*, forecasting outcomes). As the purpose of this study was to investigate the feasibility of machine learning models for 'classification' between TBM and VM, we chose the method of 'machine-learning'.

This study had limitations because of the small number of data used for learning. This limited the number of features for the analysis and accuracy of the models. It was found that deeper structures usually have better performance (*e.g.*, accuracy), but it does not mean that deeper structures are 'always' better. It depends on the data size and complexity of the problem. We found that the ANN with 2 layers is enough to solve our problem. Also, the validation data set was not separated, although the LOOCV must be preferable if it is computationally feasible in the machine learning fields [13, 14]. Because of the unbalanced sample size, it is not difficult to achieve 70% accuracy. Previous studies also included only small numbers of confirmed cases, as well as probable or possible cases that were variously defined, because only a few confirmed cases can be collected from a center over a decade. To collect maximal cases possible, the probable TBM was included in our study. It is unlikely that VM will develop simultaneously with the onset of another site of tuberculosis. It will be not easy to reproduce similar studies by collecting more confirmed cases, although probable TBM was included in our study. Using a cloud-based system for globally sharing the data from experts may be a solution to these limitations. Additionally, further information such as medical records of present illness and brain radiologic findings was provided to the clinicians. Nevertheless, the machine-learning had a similar diagnostic performance with the ID specialists. Also, only a limited number of clinicians participated in the study. In future research, a proper study design is necessary to provide the same data to humans and machines.

In conclusion, there is a possibility that machine-learning could play a role in differentiating TBM from VM. Further studies should be conducted to improve the performance of the machine-learning algorithms and to assess their safety and usefulness in real clinical practice.

ACKNOWLEDGMENT

SUPPLEMENTARY MATERIALS

Supplementary Table 1
Interpersonal comparison of diagnostic performance for differentiating tuberculous meningitis from viral meningitis

Click here to view

Supplementary Figure 1
The summary document form of non-imputed medical information.

Click here to view

REFERENCES

1. World Health Organization (WHO). Global tuberculosis report 2018. Geneva: WHO; 2018.

2. Wilkinson RJ, Rohlwink U, Misra UK, van Crevel R, Mai NTH, Dooley KE, Caws M, Figaji A, Savic R, Solomons R, Thwaites GE; Tuberculous Meningitis International Research Consortium. Tuberculous meningitis. Nat Rev Neurol 2017;13:581-98.
   PUBMED | CROSSREF

3. Erdem H, Ozturk-Engin D, Elaldi N, Gulsun S, Sengoz G, Crisan A, Johansen IS, Inan A, Nechifor M, Al-Mahdawi A, Civljak R, Ozguler M, Savic B, Ceran N, Cacopardo B, Inal AS, Namiduru M, Dayan S, Kayabas U, Parlak E, Khalifa A, Kursun E, Sipahi OR, Yemisen M, Akbulut A, Bitirgen M, Dulovic O, Kandemir B, Luca C, Parlak M, Stahl JP, Pehlivanoglu F, Simeon S, Ulu-Kilic A, Yasar K, Yilmaz G, Yilmaz E, Beovic B, Catroux M, Lakatos B, Sunbul M, Oncul O, Alabay S, Sahin-Horasan E, Kose S, Shehata G, Andre K, Alp A, Cosić G, Cem Gul H, Karakas A, Chadapaud S, Hansmann Y, Harxhi A, Kirova V, Masse-Chabredier I, Oncu S, Sener A, Tekin R, Deveci O, Karabay O, Agalar C. The microbiological diagnosis of tuberculous meningitis: results of Haydarpasa-1 study. Clin Microbiol Infect 2014;20:O600-8.
   PUBMED | CROSSREF

4. Solomons RS, van Elsland SL, Visser DH, Hoek KG, Marais BJ, Schoeman JF, van Furth AM. Commercial nucleic acid amplification tests in tuberculous meningitis--a meta-analysis. Diagn Microbiol Infect Dis 2014;78:398-403.
   PUBMED | CROSSREF

5. Nhu NT, Heemskerk D, Thu DA, Chau TT, Mai NT, Nghia HD, Loc PP, Ha DT, Merson L, Thinh TT, Day J, Chau N, Wolbers M, Farrar J, Caws M. Evaluation of GeneXpert MTB/RIF for diagnosis of tuberculous meningitis. J Clin Microbiol 2014;52:226-33.
   PUBMED | CROSSREF

6. Hristea A, Olaru ID, Baicus C, Moroti R, Arama V, Ion M. Clinical prediction rule for differentiating tuberculous from viral meningitis. Int J Tuberc Lung Dis 2012;16:793-8.
   PUBMED | CROSSREF

7. Lee SA, Kim SW, Chang HH, Jung H, Kim Y, Hwang S, Kim S, Park HK, Lee JM. A new scoring system for the differential diagnosis between tuberculous meningitis and viral meningitis. J Korean Med Sci 2018;33:e201.
   PUBMED | **CROSSREF**

8. Hong SI, Kim T, Jung J, Park SY, Chong YP, Lee SO, Choi SH, Kim YS, Woo JH, Lee SA, Kim SH. Tuberculous meningitis-mimicking varicella-zoster meningitis. Infect Chemother 2017;49:123-9.
   PUBMED | **CROSSREF**

9. Cho BH, Kim BC, Yoon GJ, Choi SM, Chang J, Lee SH, Park MS, Shin JH, Kim MK, Cho KH. Adenosine deaminase activity in cerebrospinal fluid and serum for the diagnosis of tuberculous meningitis. Clin Neurol Neurosurg 2013;115:1831-6.
   PUBMED | **CROSSREF**

10. Marais S, Thwaites G, Schoeman JF, Török ME, Misra UK, Prasad K, Donald PR, Wilkinson RJ, Marais BJ. Tuberculous meningitis: a uniform case definition for use in clinical research. Lancet Infect Dis 2010;10:803-12.
    PUBMED | **CROSSREF**

11. Solomons RS, Wessels M, Visser DH, Donald PR, Marais BJ, Schoeman JF, van Furth AM. Uniform research case definition criteria differentiate tuberculous and bacterial meningitis in children. Clin Infect Dis 2014;59:1574-8.
    PUBMED | **CROSSREF**

12. Fritz BA, Chen Y, Murray-Torres TM, Gregory S, Ben Abdallah A, Kronzer A, McKinnon SL, Budelier T, Helsten DL, Wildes TS, Sharma A, Avidan MS. Using machine learning techniques to develop forecasting algorithms for postoperative complications: protocol for a retrospective study. BMJ Open 2018;8:e020124.
    PUBMED | **CROSSREF**

13. Hawkins DM, Basak SC, Mills D. Assessing model fit by cross-validation. J Chem Inf Comput Sci 2003;43:579-86.
    PUBMED | **CROSSREF**

14. Raschka S. Model evaluation, model selection, and algorithm selection in machine learning. ArXiv 2018;1811.12808.