Mathematical Biosciences and Engineering

MBE, 21(5): 6150–6166.
DOI: 10.3934/mbe.2024270
Received: 14 January 2024
Revised: 21 May 2024
Accepted: 29 May 2024
Published: 03 June 2024

http://www.aimspress.com/journal/MBE

Research article

Predicting the transmission trends of COVID-19: an interpretable machine learning approach based on daily, death, and imported cases

Hyeonjeong Ahn and Hyojung Lee*

Department of Statistics, Kyungpook National University, Daegu 41566, Republic of Korea

* **Correspondence:** Email: hjlee@knu.ac.kr.

Supplementary

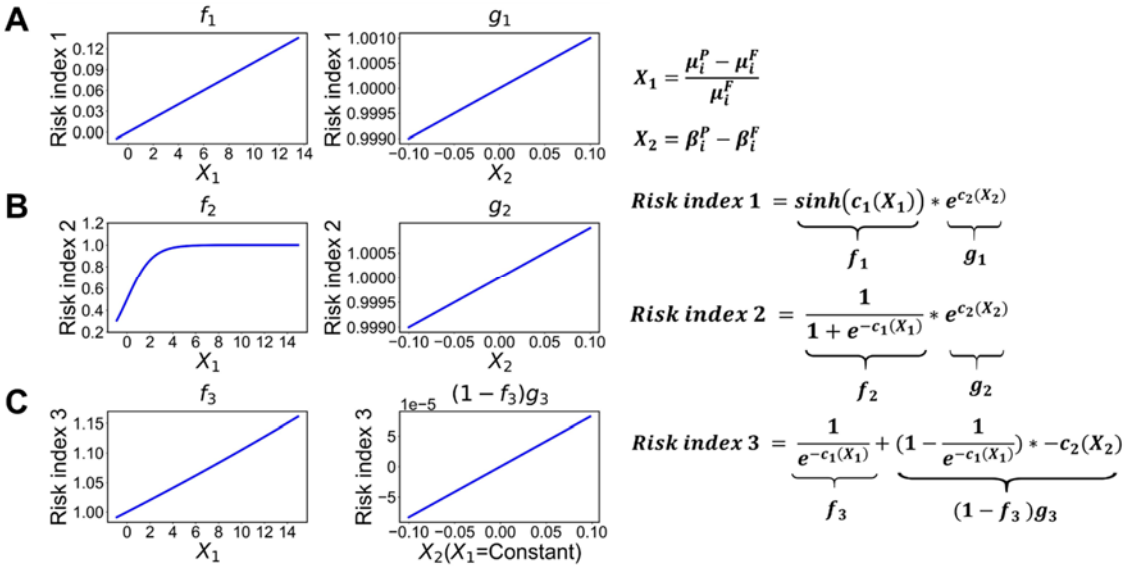

Figure S1. Graphs for the two functions, namely, f and g, for identifying transmission patterns. f and g are functions for the average of the cases and the slope of the linear regression, respectively, and consist of the risk index. In $(1-f_3)g_3$, the variable X_1, which is a constant, was set to be eight.

Table S1. Range of hyperparameters by grid search according to the 10-fold cross validation.

		SVM			RF		XGB	
		C	Gamma	Kernel	Number of trees	Maximum depth	Number of trees	Maximum depth
Confirmed cases	**Risk index 1**	160 (10–200)	0.1 (0.1–1)	rbf (rbf, poly, sigmoid)	60 (50–100)	9 (1–20)	75 (5–100)	3 (1–20)
	Risk index 2	190 (10–200)	0.1 (0.1–1)	rbf (rbf, poly, sigmoid)	50 (50–100)	15 (1–20)	75 (5–100)	11 (1–20)
	Risk index 3	160 (10–200)	0.1 (0.1–1)	rbf (rbf, poly, sigmoid)	70 (50–100)	9 (1–20)	80 (5–100)	13 (1–20)
Death cases	**Risk index 1**	50 (10–200)	0.1 (0.1–1)	rbf (rbf, poly, sigmoid)	95 (50–100)	14 (1–20)	60 (5–100)	5 (1–20)
	Risk index 2	80 (10–200)	0.1 (0.1–1)	rbf (rbf, poly, sigmoid)	50 (50–100)	17 (1–20)	55 (5–100)	17 (1–20)
	Risk index 3	50 (10–200)	0.1 (0.1–1)	rbf (rbf, poly, sigmoid)	95 (50–100)	14 (1–20)	60 (5–100)	5 (1–20)
Imported cases	**Risk index 1**	130 (10–200)	0.1 (0.1–1)	rbf (rbf, poly, sigmoid)	90 (50–100)	18 (1–20)	20 (5–100)	15 (1–20)
	Risk index 2	100 (10–200)	0.3 (0.1–1)	rbf (rbf, poly, sigmoid)	75 (50–100)	15 (1–20)	80 (5–100)	7 (1–20)
	Risk index 3	190 (10–200)	0.2 (0.1–1)	rbf (rbf, poly, sigmoid)	85 (50–100)	15 (1–20)	90 (5–100)	5 (1–20)

Table S2. The values of c_1 and c_2 according to respective risk index and data.

Data	**Risk index**	c_1	c_2	**Correlation between the value of risk index and labels**
Confirmed cases	Risk index 1	0.01	0.01	0.582
	Risk index 2	0.91	0.01	0.847
	Risk index 3	0.01	0.01	0.570
Death cases	Risk index 1	0.01	0.01	0.758
	Risk index 2	0.91	0.01	0.888
	Risk index 3	0.01	0.01	0.755
Imported cases	Risk index 1	0.01	0.01	0.415
	Risk index 2	0.91	0.01	0.852
	Risk index 3	0.01	0.01	0.395

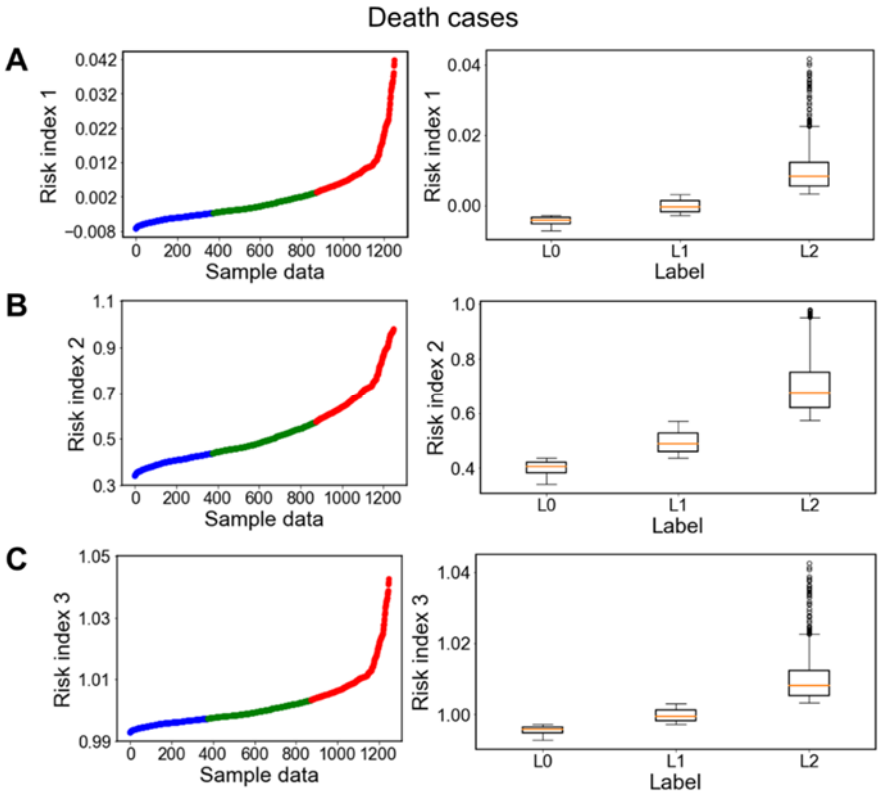

Figure S2. Range and distribution of the values of the three types of risk index for death cases.

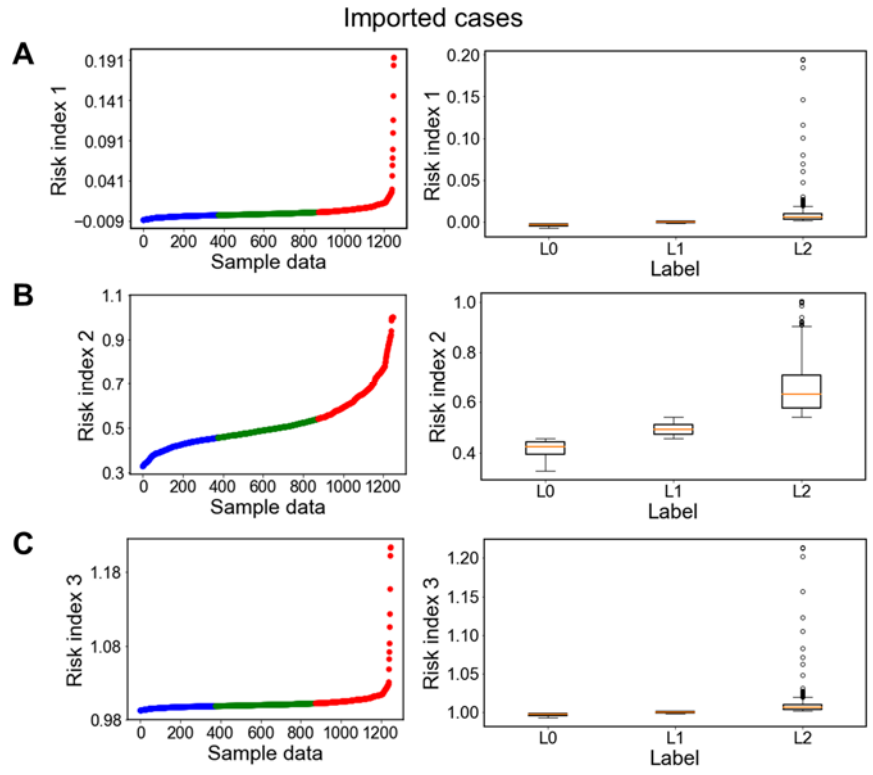

Figure S3. Range and distribution of the values of the three types of risk index for imported cases.

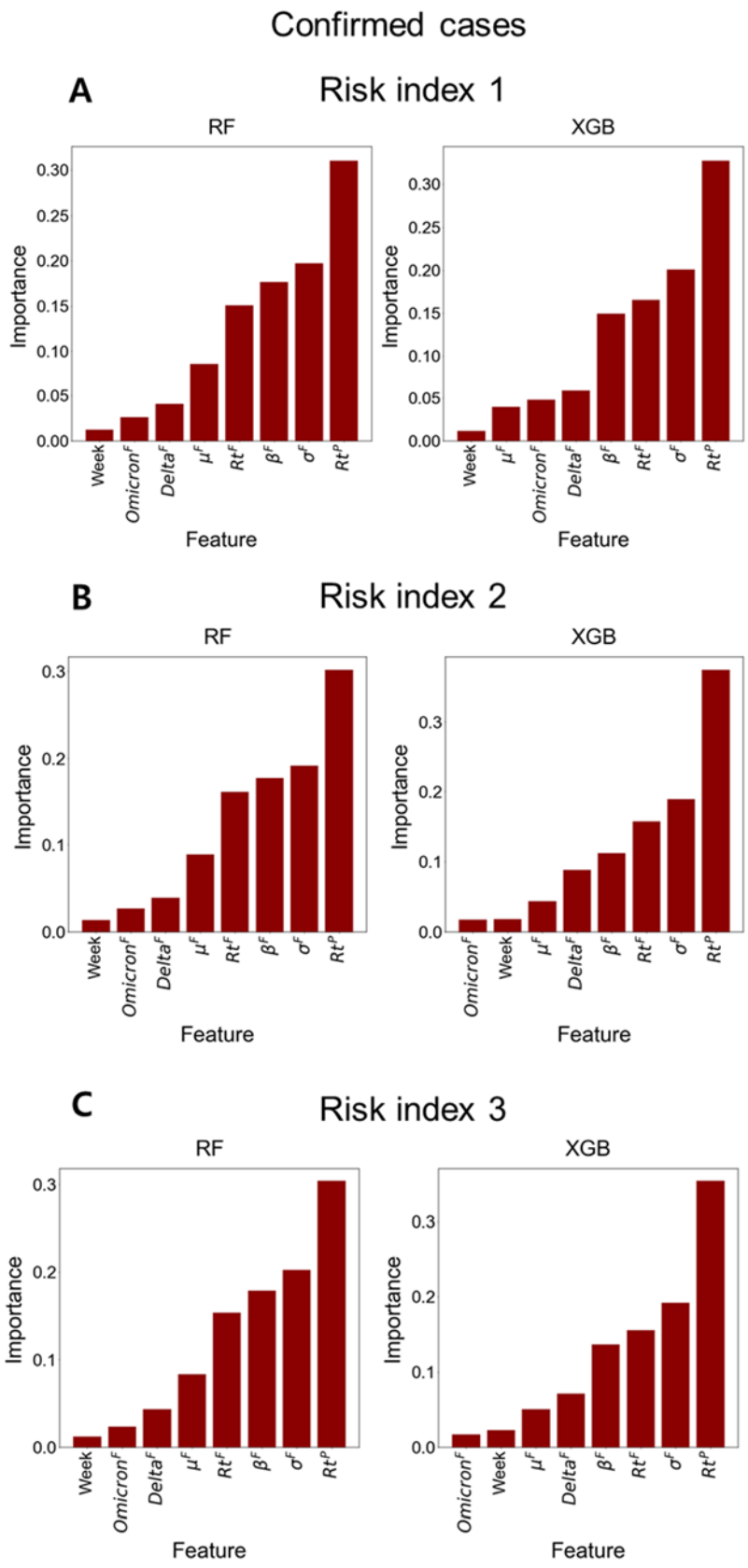

Figure S4. Feature importance of RF and XGB for confirmed cases.

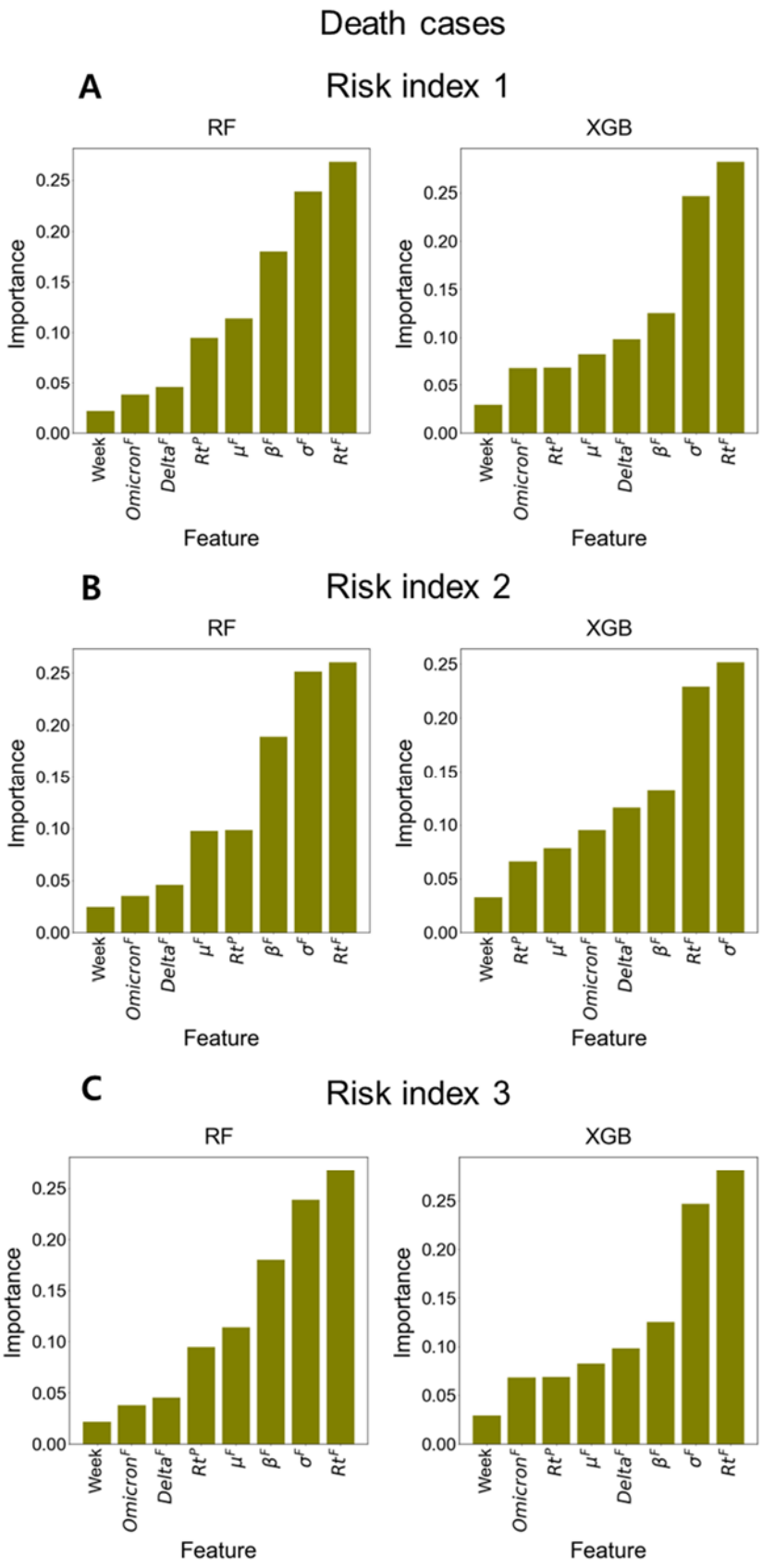

Figure S5. Feature importance of RF and XGB for death cases.

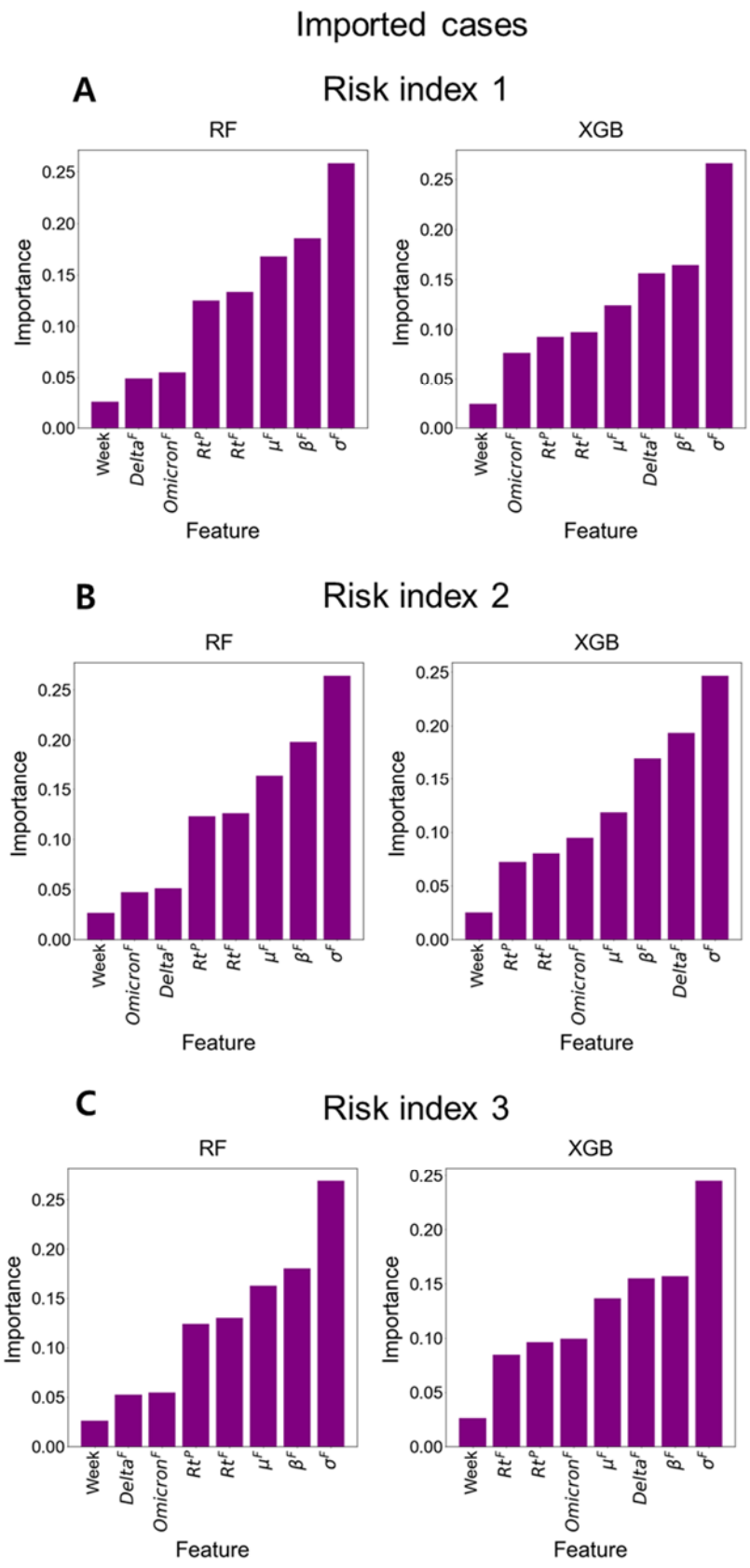

Figure S6. Feature importance of RF and XGB for imported cases.

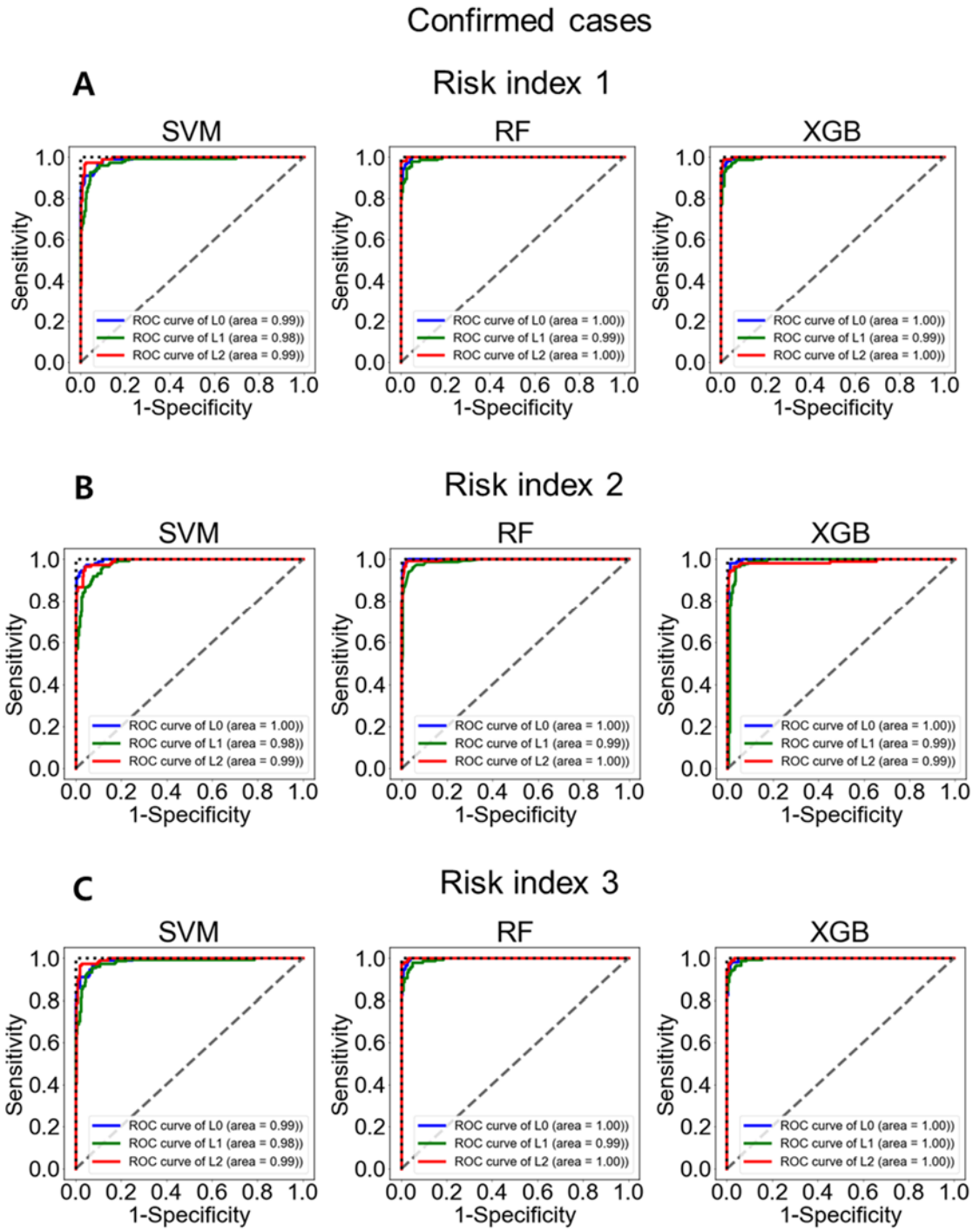

Figure S7. ROC curves for confirmed cases.

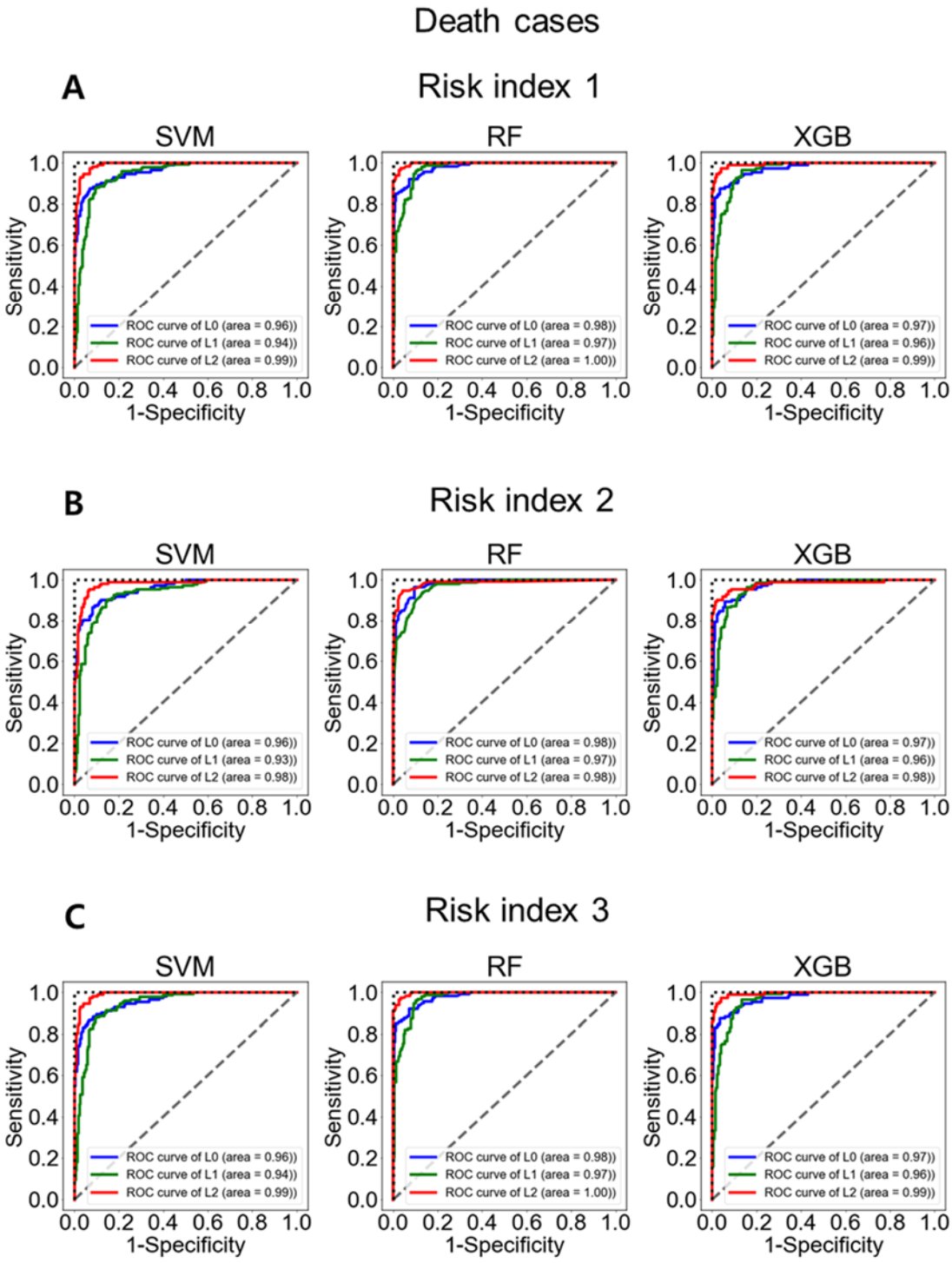

Figure S8. ROC curves for death cases.

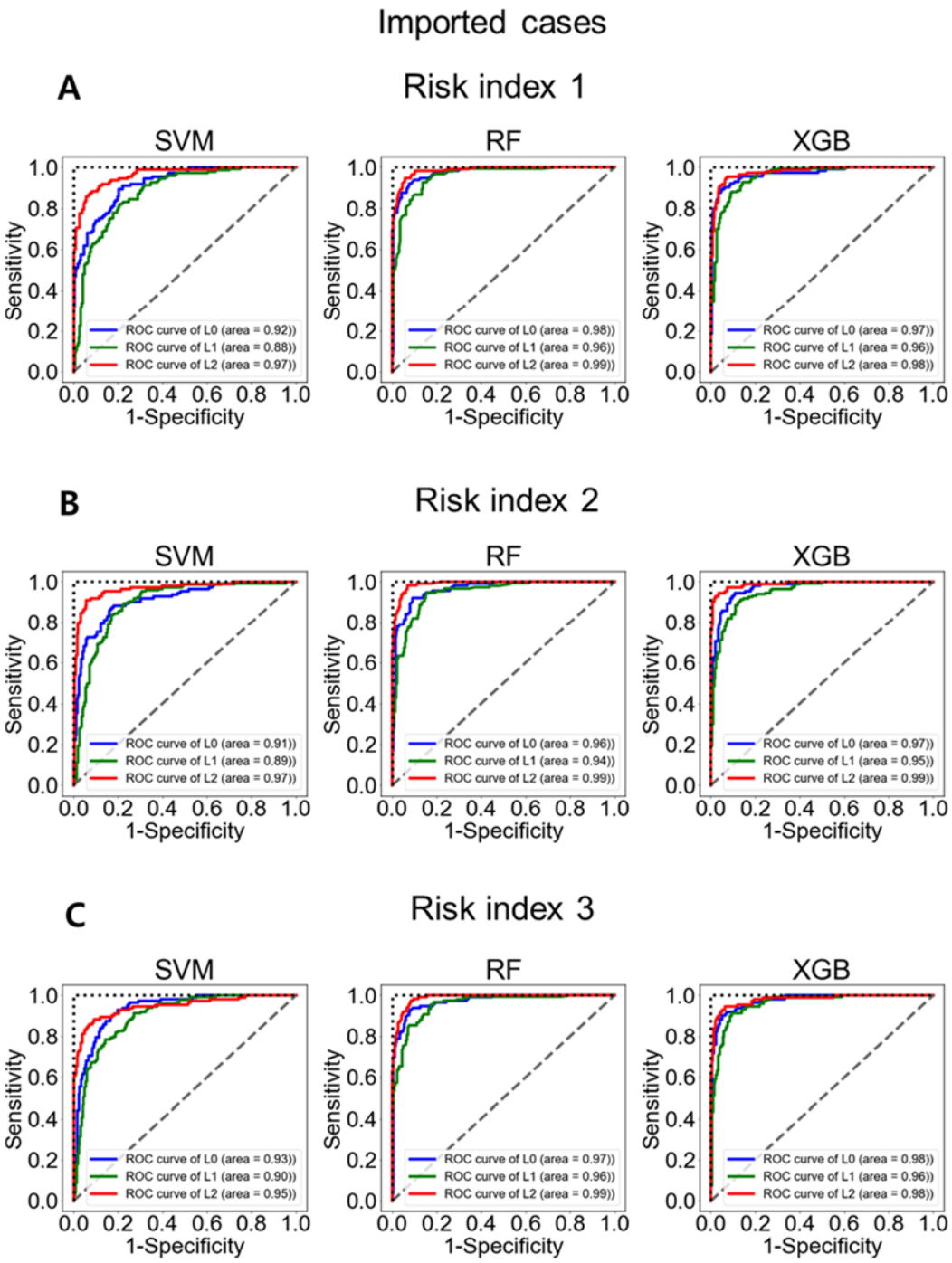

Figure S9. ROC curves for imported cases.

Table S3. Accuracy under different fitting and prediction periods.

Period (Fitting/Prediction)	Method	Risk index 1		
		Confirmed cases	**Death cases**	**Imported cases**
21days/14days	SVM	0.9229	0.8777	0.7766
	RF	0.9548	0.9176	0.8697
	XGB	0.9521	0.8936	0.8803
28days/14days	SVM	0.9225	0.9011	0.869
	RF	0.9251	0.9091	0.9037
	XGB	0.9332	0.8824	0.8984
21days/28days	SVM	0.9409	0.8952	0.8602
	RF	0.957	0.9355	0.9355
	XGB	0.9543	0.9274	0.914
21days/21days	SVM	0.9332	0.8663	0.8797
	RF	0.9118	0.9144	0.9144
	XGB	0.9037	0.9385	0.9037
Period (Fitting/Prediction)	**Method**	**Risk index 2**		
		Confirmed cases	**Death cases**	**Imported cases**
21days/14days	SVM	0.9149	0.8564	0.7872
	RF	0.9521	0.8777	0.859
	XGB	0.9548	0.891	0.883
28days/14days	SVM	0.9011	0.8824	0.8984
	RF	0.9439	0.9064	0.9225
	XGB	0.9385	0.8877	0.9251
21days/28days	SVM	0.922	0.8952	0.8522
	RF	0.9355	0.9301	0.9355
	XGB	0.9247	0.9328	0.9274
21days/21days	SVM	0.9064	0.8743	0.8663
	RF	0.9118	0.9305	0.9091
	XGB	0.8877	0.9332	0.9011
Period (Fitting/Prediction)	**Method**	**Risk index 3**		
		Confirmed cases	**Death cases**	**Imported cases**
21days/14days	SVM	0.9229	0.8777	0.7899
	RF	0.9548	0.9176	0.8617
	XGB	0.9521	0.8936	0.8936
28days/14days	SVM	0.9225	0.9011	0.7899
	RF	0.9251	0.9091	0.8617
	XGB	0.9332	0.8824	0.8936
21days/28days	SVM	0.9409	0.8952	0.8602
	RF	0.957	0.9355	0.9328
	XGB	0.9543	0.9274	0.922
21days/21days	SVM	0.9332	0.8663	0.8797
	RF	0.9118	0.9144	0.9144
	XGB	0.9037	0.9385	0.9037

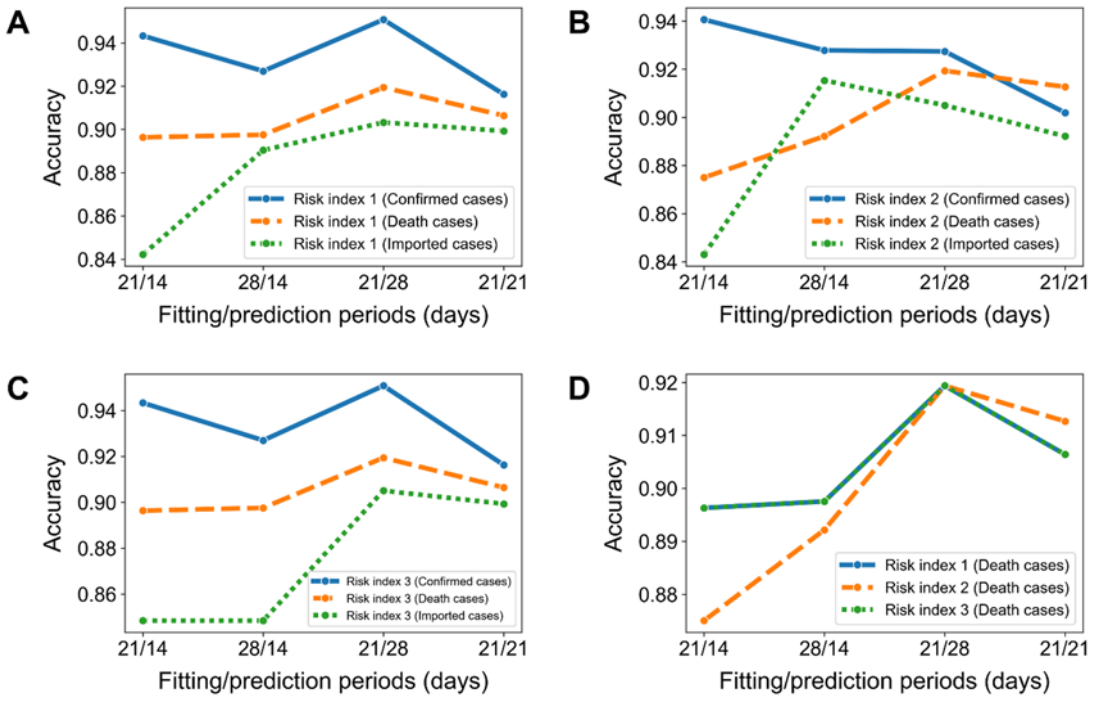

Figure S10. Prediction performance based on varying lengths of the fitting and prediction periods.

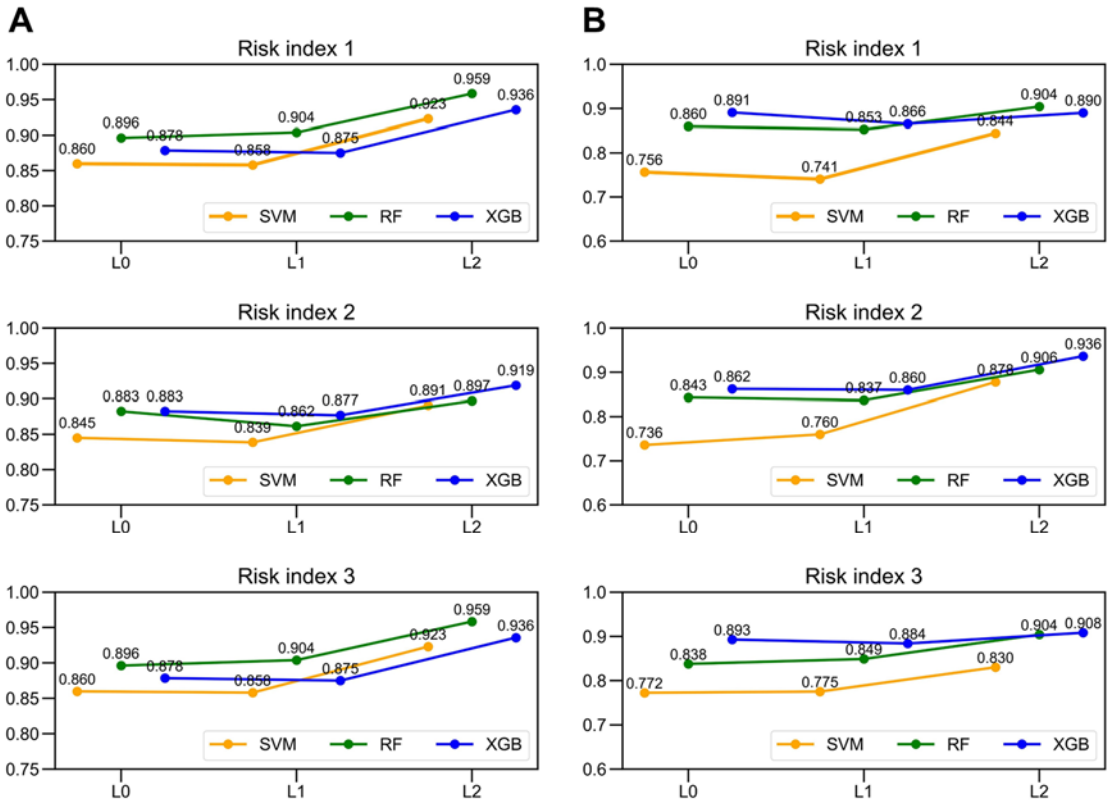

Figure S11. Accuracy of prediction of classification according to labels. A. Death cases. B. Imported cases.

Table S4. Accuracy of prediction using three labels and risk index for COVID-19 data.

Data	Risk index	Classification method	Accuracy	F1-score		
				L0 (Decrease)	L1 (Maintain)	L2 (Increase)
Confirmed cases	Risk index 1	SVM	0.9229	0.9132	0.9172	0.9406
		RF	0.9548	0.9545	0.9457	0.968
		XGB	0.9521	0.9493	0.9427	0.9683
	Risk index 2	SVM	0.9149	0.955	0.8961	0.9009
		RF	0.9521	0.9649	0.9408	0.9545
		XGB	0.9548	0.9689	0.945	0.9541
	Risk index 3	SVM	0.9229	0.9132	0.9172	0.9406
		RF	0.9548	0.9593	0.9457	0.9633
		XGB	0.9521	0.9593	0.9419	0.9593
Death cases	Risk index 1	SVM	0.8777	0.8598	0.858	0.9231
		RF	0.9176	0.8962	0.904	0.9585
		XGB	0.8936	0.8785	0.875	0.9358
	Risk index 2	SVM	0.8564	0.8451	0.8387	0.8908
		RF	0.8777	0.8826	0.8615	0.8972
		XGB	0.891	0.8826	0.877	0.9189
	Risk index 3	SVM	0.8777	0.8598	0.858	0.9231
		RF	0.9176	0.8962	0.904	0.9585
		XGB	0.8936	0.8785	0.875	0.9358
Imported cases	Risk index 1	SVM	0.7766	0.7565	0.7407	0.8444
		RF	0.8697	0.8597	0.8526	0.9041
		XGB	0.8803	0.8908	0.8656	0.8899
	Risk index 2	SVM	0.7872	0.7359	0.76	0.8778
		RF	0.859	0.843	0.8366	0.9058
		XGB	0.883	0.8622	0.8599	0.9364
	Risk index 3	SVM	0.7899	0.7725	0.7752	0.8302
		RF	0.8617	0.8378	0.8489	0.9041
		XGB	0.8936	0.8929	0.8839	0.9083

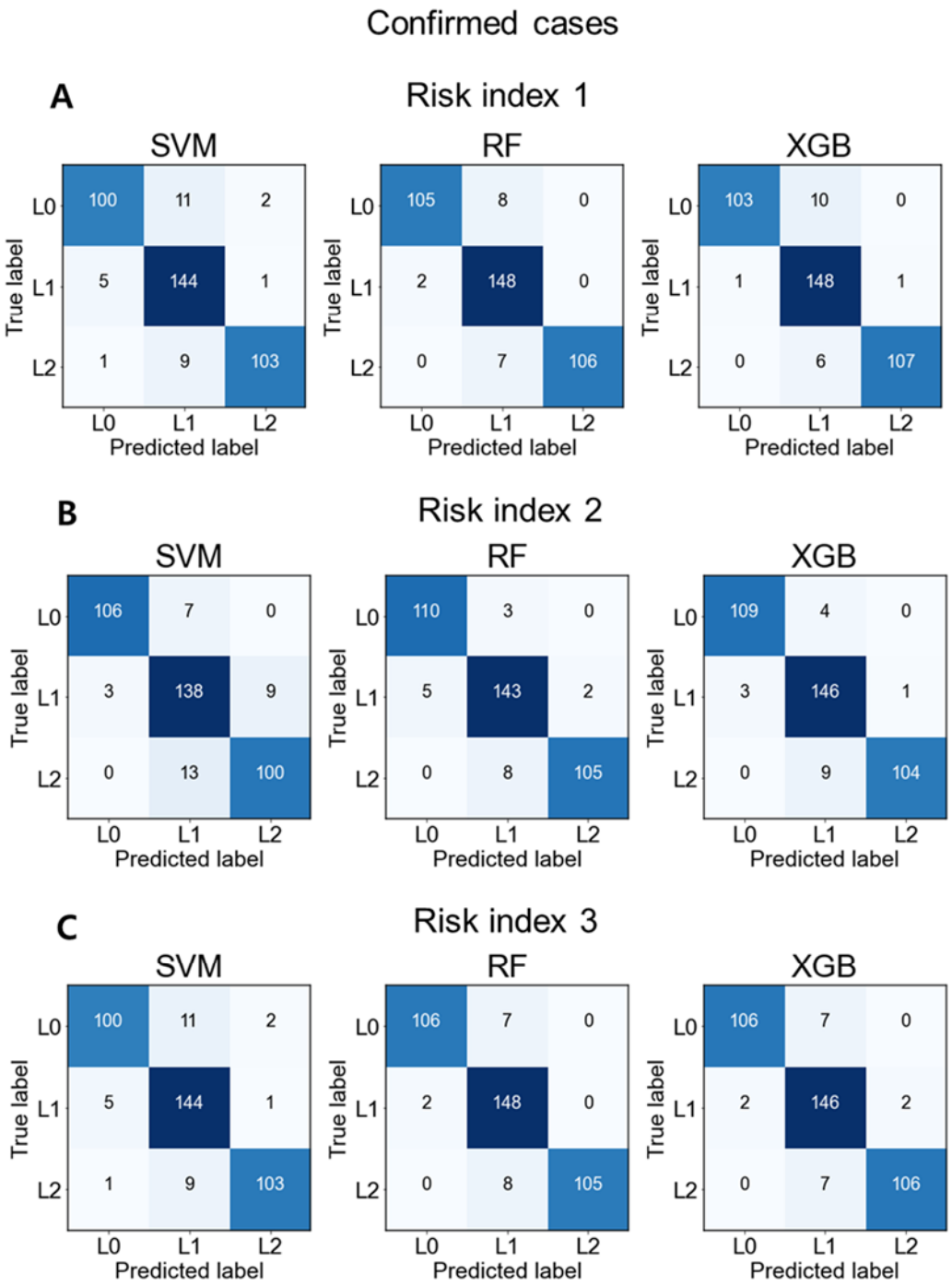

Figure S12. Confusion matrix for confirmed cases.

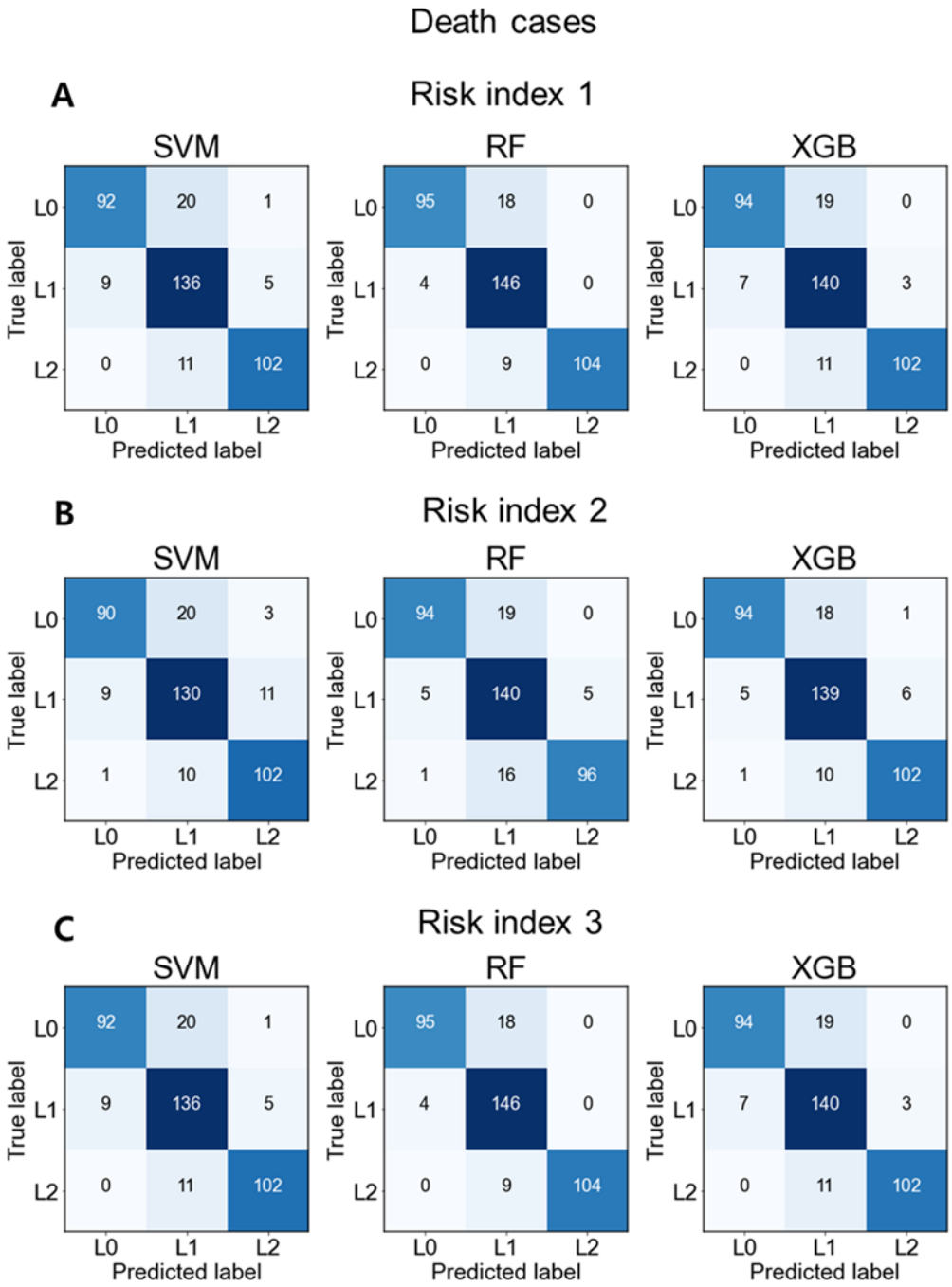

Figure S13. Confusion matrix for death cases.

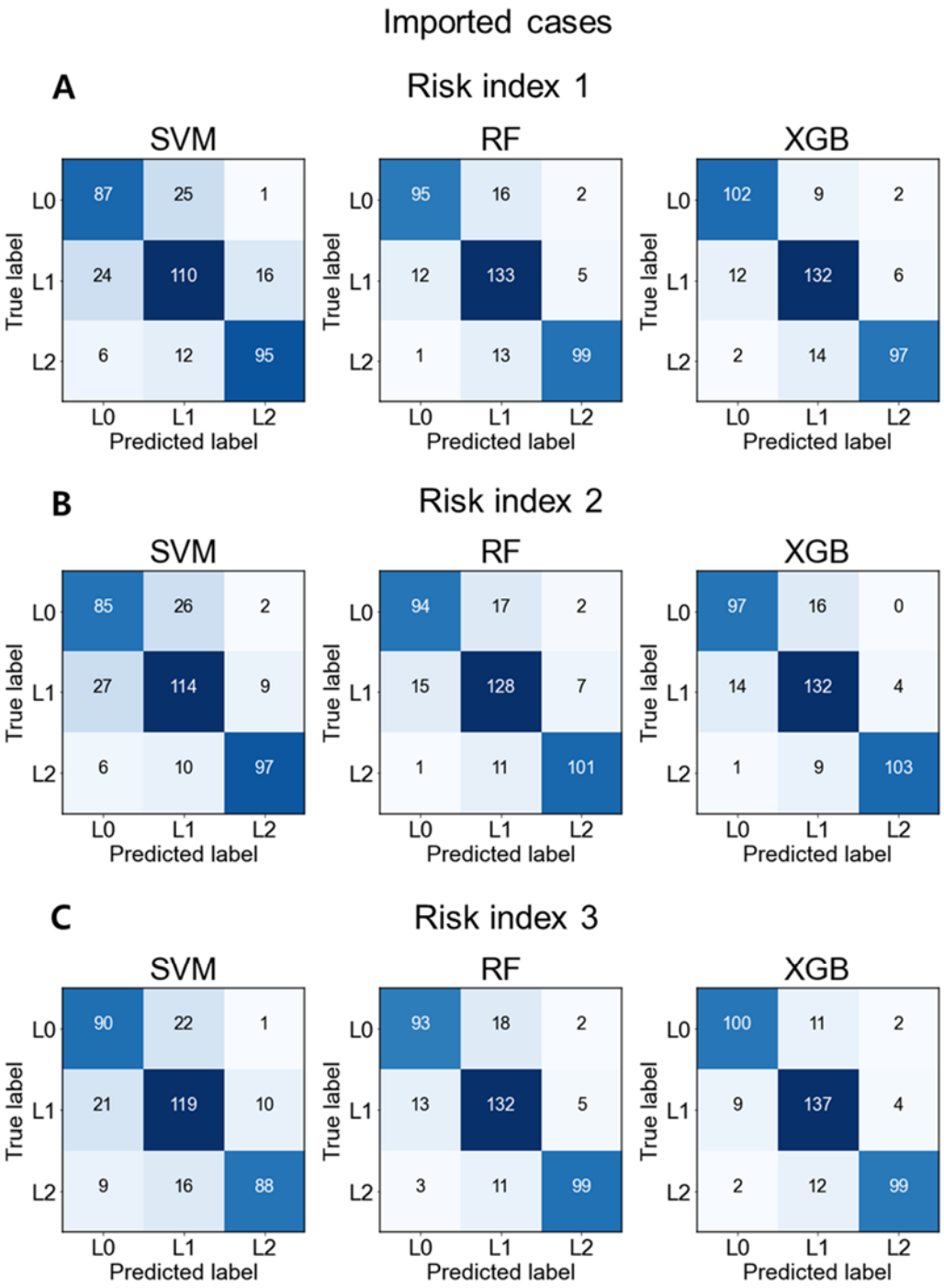

Figure S14. Confusion matrix for imported cases.