Prediction Model of Post-TAVR Complication Using a Medical Twin Web Navigator

Se-Min Hyun and KangYoon Lee*

Department of Computer Engineering, Gachon University, Republic of Korea
E-mail: tpalsgus0@gachon.ac.kr; keylee@gachon.ac.kr
*Corresponding Author

Abstract

Transcatheter aortic valve replacement (TAVR) has been introduced as an alternative to surgical aortic valve replacement for patients with severe aortic valve disease and is expanding into a universal treatment. However, complications after TAVR can have devastating consequences for patients and must be predicted. By designing a TAVR medical twin architecture based on real-world data (RWD), we can minimize complications and achieve optimal clinical outcomes through analysis and simulation results in a virtual environment that can predict complications. The simulation phase utilizes machine learning algorithms for complication prediction to predict patients with conduction abnormalities, a complication of TAVR, and provides the prediction results through a web-based monitoring system. We also conduct research to identify factors that influence complications, so that complication prediction in a virtualized environment on a medical twin architecture can serve as a guide for personalized care design for patients undergoing TAVR.

Keywords: Transcatheter aortic valve replacement (TAVR), medical twin, web navigator, complication prediction model.

1 Introduction

Transcatheter aortic valve replacement (TAVR) is the latest surgical method used to treat aortic valve stenosis. It has been commonly used in large-scale clinical studies in the United States and Europe since Dr Alain Cribier of France attempted and succeeded in treating patients with severe aortic valve stenosis in 2002 [1]. However, while TAVR has a growing body of research on its efficacy and safety relative to surgical risk compared with surgical aortic valve replacement (SAVR), TAVR is associated with a variety of clinical outcomes and side effects depending on the individual patient's condition, the size of the valve used during the procedure, and the location of insertion.

In the medical system, personalized care refers to the provision of optimized treatment of patients by considering genetic, environmental, and lifestyle factors. Compared with conventional treatments, personalized care is more effective and has fewer side effects. For such a system, we propose a medical twin, which is a virtual representation of a real-world object. Medical twins and personalized medicine are complementary technologies. Medical twins can be used to monitor a patient's condition and make treatment plans, and are expected to contribute to improving the quality of life of patients in medical systems. For example, a patient's condition can be monitored in real time to detect and treat disease deterioration at an early stage, treatment can be optimized to increase treatment effectiveness, and the patient's course of treatment can be monitored to prevent side effects.

Traditional prediction models rely on statistical regression models [2]. However, machine learning, a subfield of computer science, has shown good predictive power in various areas, including Alzheimer's prediction [3]. The machine learning algorithm can take the already output results and learn from them to optimise the prediction for this outcome. The optimised prediction results can also be provided to the physician as a web dashboard so that he can see the patient's information and the prediction results for TAVR, and provide feedback on the results to provide a better prediction model.

In this paper, we study a medical twin-based simulated web navigator based on artificial intelligence machine learning algorithms using real world data (RWD).

2 Related Research

2.1 Transcatheter Aortic Valve Replacement (TAVR)

The aortic valve is located between the left ventricle and the aorta [4] and is responsible for preventing blood from the aorta from flowing back into
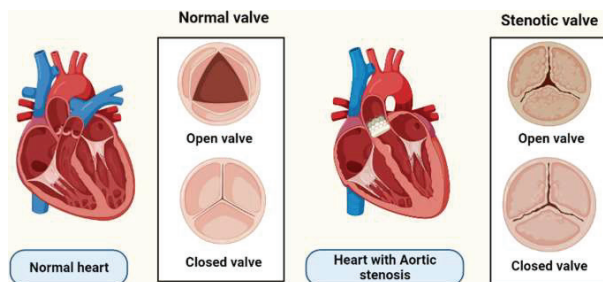
Figure 1   Diagrammatic representation of a normal aortic valve versus aortic stenosis [8].

the left ventricle. The normal aortic valve is a trileaflet structure located at the junction between the left ventricular outflow tract and aortic root [5] as in Figure 1. However, degenerative aortic valve stenosis, caused by congenital abnormalities or calcification due to calcium deposition in the heart valve, gradually leads to thickening of the heart muscle, resulting in cardiac dysfunction and symptoms such as shortness of breath, chest pain, fainting, and a high mortality rate. Although cholesterol and high blood pressure may contribute to the issue, the primary cause is aging. As a result, the majority of patients are over 70 years old. In recent years, the aging population has led to an increase in the number of patients [6] with aortic stenosis, a condition in which the aortic valve ages and calcifies. In the past, the surgical treatment of older patients was either impossible or difficult; however, recent improvements in the outcomes of TAVR using a catheter have led to an increase in the number of patients undergoing the procedure [7].

Unlike traditional surgery, which involves replacing the valve through an incision in the chest, TAVR involves guiding a balloon along a blood vessel in the thigh to the valve, inflating the balloon between the narrowed valves, and securing a stent to the aortic valve, which can act as a valve. The introduction of TAVR as complementary technique to SAVR has revolutionized the treatment of severe aortic stenosis. TAVR has become the standard of care for patients with severe aortic stenosis who are at increased risk for surgery [9]. While TAVR has a numerically low surgical risk, studies have shown that pre- and post-procedural complications can affect postoperative survival, as in Figure 2, which shows moderate/severe paravalvular leakage (PVL) [10], major vascular and bleeding com-plications [11], disabling stroke [12], acute kidney injury (AKI) [13], and conduction abnormalities such as high-degree atrioventricular block (AVB) [14] requiring permanent pacemaker implantation. The rate of new-onset left bundle branch block (LBBB) after TAVR varies from approximately 4% to 65%, depending on the
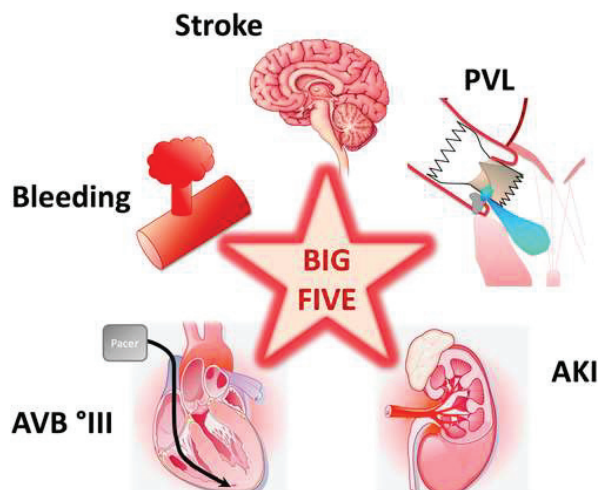
Figure 2   Big five periprocedural complications that significantly impact long-term outcomes after transcatheter aortic valve replacement [16].

type of valve used. The two most commonly used valves are self-expanding and balloon-expandable [15].

A conduction abnormality (CA) occurs when electrical signals from the heart do not travel properly, and conduction abnormalities can cause an irregular heartbeat or heart attack. It is not known exactly why conduction abnormalities occur after a TAVR procedure, but they are thought to be caused by damage to the nerves that carry the electrical signals of the heart during the TAVR procedure. A pacemaker is a device that can regulate the heart's electrical signals and normalize the heart's rhythm; therefore, if conduction abnormalities occur after a TAVR procedure, a pacemaker may need to be implanted. Therefore, to predict CA as a complication of TAVR, it is necessary to identify factors that affect CA and to study the possibility of CA in post-TAVR through machine learning prediction models.

2.2 Medical Twin Web Navigator Architecture

A digital twin is a technology that realizes twins that reflect the physical features of real objects in a three-dimensional model and uses them to make decisions about objects through simulation in synchronization with reality. They can be applied in various fields, such as manufacturing, construction, smart cities, and medical services [17], and are growing as a core technology for digital transformation. With recent advancements in data analysis

technology and the emergence of wearable devices, digital twins can be combined with medical data to enable personalized precision medicine and innovative healthcare services in the medical field. The key to precision medicine and prevention are personalized medical data and disease prediction, which are in contact with digital twins. A medical twin is a concept that applies digital twins to the medical field and provides services by applying digital twins at all stages, such as prevention, diagnosis, treatment, and management in healthcare. By leveraging individual patients' electronic medical records and patient-generated data, digital twin technology can also promote personalized medicine research by enabling better research results, such as drug interactions and treatment effects [18].

In this study, we designed the TAVR navigator architecture as shown in Figure 3. The patient data (baseline, EKG, cardiac) collected from the hospital is entered into the eCRF of the hospital information system (HIS) and stored in the database for development. The patient's data is stored in the post-TAVR navigator DB through the API, data analysis and prediction models are developed in the post-TAVR simulation layer, and the result data is stored in the navigator result DB. Then, in the digital twin application layer, the results of the post-TAVR prediction model are delivered to the web navigator through the API, visualized through the web dashboard, and modified or improved through feedback from medical staff on the results of the prediction model.
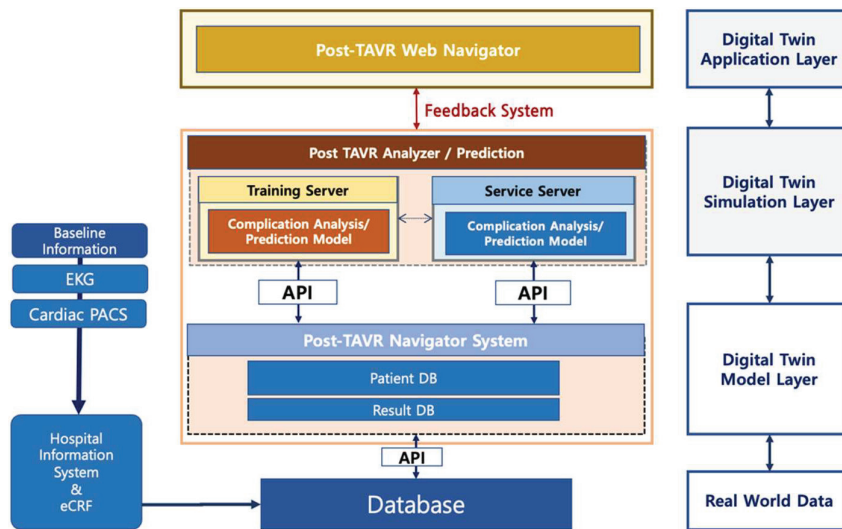

Figure 3   TAVR medical twin navigator.

After defining the data required to build a medical twin environment, we designed a database, synchronized RWD in real time to create a virtual environment framework, and created a virtual twin model that can reproduce the TAVR medical twin framework based on the input RWD to build a virtual environment that can reproduce TAVR. The virtual environment manages the patient's treatment decision process, aortic condition, etc. It utilizes deep learning and machine learning analysis techniques to provide feedback on patient status information and outcomes generated by the results of the clinical twin simulation environment. After surgery, the patient's RWD medical twin and surgical outcome data were precisely checked, utilized, and synchronized to continuously monitor the patient's condition.

To improve the post-TAVR complication risk prediction model based on the virtual object of the medical twin, we pre-processed patient data including RWD-based ECG data, trained a risk status classification machine learning/deep learning AI model, and applied and evaluated the classification prediction algorithm for complication risk management prediction. Finally, we performed patient-specific complication risk management simulations and implemented feedback-based prediction model improvement to improve the accuracy of the prediction model through feedback via a web-based dashboard. In addition, we provided a predictive model for conduction disturbances among complications in a virtual environment on a medical twin architecture, we went through data pre-processing and exploration, and trained the model using machine learning algorithms such as logistic regression, decision tree, random forest, and gradient boosting.

3 Research Method

The database included TAVR patient data from Gil Hospital in Incheon, Republic of Korea, where the study was conducted with institutional permission.

3.1 Study Patient Population

The following 44 variables were collected for each patient as in Table 1: clinical data (age, gender, BMI (kg/m^2), smoking status, hypertension [HTN], diabetes melitus [DM], chronic kidney disease [CKD]), lab data (hemoglobin (g/dL), HbA1c(%), aspartate aminotransferase [AST] (U/L), alanine aminotransferase [ALT] (U/L), creatinine (mg/dL), Nt-proBNP (pg/mL)), vital data (systolic blood pressure [sBP], diastolic blood pressure [dBP], pulse rate

Table 1 The variables of TAVR patient data

Clinic Data	Lab Data	Vital	EKG	Cardiac CT	Cardiac Echo	Drug	Valve
Age	Hemoglobin	sBP	RBBB (Y/N)	LVOT perimeter	AV max	ASA	Valve code
Gender	HbA1c	dBP	PR interval	LVOT Diameter (mean)	AV mgPG	Antiplatelet other than ASA	Valve size
BMI	AST	Pulse rate	First degree AV block		LVEDD	Anticoagulants	
Smoking status	ALT	Respiratory rate	QRS interval		LVEF	Beta blocker	
HTN	Creatinine		Baseline QRS interval > 120 ms (Y/N)		Peak AV Vmax	CCB	
DM	Nt-proBNP		Delta PR interval		AV mean pressure gradient	ACEi ARB	
CKD			Delta QRS interval			Diuretics	
			New-onset LBBB (Y/N)				
			Pacemaker implantation (Y/N)			Aldosterone antagonist	

(per/min), respiratory rate (per/min)), drug (angiotensin-converting enzyme inhibitors or angiotension II receptor blocker [ACEi or ARB], beta blocker, aldosteroneantagonist, antiplatelet other than ASA, calcium-channel blockers [CCB], diuretics, acetylsalicylic acid [ASA]), EKG (RBBB, LBBB, first degree AV block, PR interval (ms), QRS interval (ms), Baseline QRS > 120 ms, DeltaPR, DeltaQRS, new onset RBBB, new onset LBBB, pacemaker imputation), TAVR cardiac CT (left ventricular outflow tract [LVOT] perimeter (mm), LVOT diameter (mean) (mm)), TAVR cardiac echo (aortic valve (AV) Vmax (m/s), AV mgPG (mmHg), left ventricular end diastolic diameter [LVEDD] (mm), left ventricular ejection fraction [LVEF] (%), peak AV velocity [Vmax] (m/s), AV mean pressure gradient (mmHg)), valve (valve size, valve code).

Baseline characteristics of clinic data, vital data, ECG data, and valve data for the raw data in the study population are shown in Table 2. Categorical variables expressed as 'Yes' and 'No' resulted in only 'Yes' outcomes. Continuous variables are expressed as mean $\pm$ standard deviation for normal distribution and median and interquartile range for non-normal distribution. Categorical variables are expressed as N (%). Table 2 shows the results of the raw data prior to data pre-processing, missing values (PR_interval, Delta_PR) were addressed in the study design of the prediction model.

The baseline characteristics of the patients are as in Table 2. The mean age was 81.6 years, and approximately 43% (43 patients) were female. New onset LBBB was present in approximately 13.8% (11 patients).

3.2 Prediction Model Research Design

The prediction model was trained using machine learning algorithms, including logistic regression, decision tree, random forest, and XGBoost (extreme gradient boosting). Logistic regression was chosen due to its common use in clinical studies.

The data was randomly divided into two subsets: a training set (70%) and a test set (30%) to ensure fair evaluation of the model. Missing data values were imputed using a random forest algorithm. To address the imbalance in patient representation within the existing training set, the synthetic minority oversampling technique (SMOTE) was applied prior to model development [19]. SMOTE is a technique that generates new samples by adding random values to existing samples from classes with fewer data points. Although randomized oversampling has a higher probability of overfitting, SMOTE has the advantage of a lower probability of overfitting, less information loss compared to under-sampling, and no data reduction.

Table 2 Baseline characteristics of TAVR patients

Variable	N	Total
age	80	81.0 ± 6.1
Female	43	43 (53.8%)
BMI	80	24.8 ± 3.7
Smoking	80	72 (90.0%)
HTN	80	10 (12.5%)
sBP	80	133.2 ± 15.5
dBP	80	72.6 ± 10.2
pulse_rate	80	70.0 ± 10.9
respiratory_rate (=20)	80	64 (80.0%)
RBBB	80	11 (13.8%)
LBBB	80	2 (2.5%)
PR_interval	70	180.7 ± 43.4
first_degree_AVblock	80	11 (13.8%)
QRS_interval	80	99.85 ± 23.2
Delta_PR	65	3.8 ± 40.6
Delta_QRS	80	16.6 ± 26.2
BaselineQRS > 120ms	80	14 (17.5%)
New_onset_RBBB	80	7 (8.8%)
New_onset_LBBB	80	11 (13.8%)
Pacemaker	80	5 (6.2%)

The prediction performance of the machine learning model was evaluated on the test set using accuracy, F1-score, precision, recall, and the area under the curve (AUC). The F1 score, which is the harmonic mean of precision and recall [20], was used to measure performance. Precision represents the proportion of positive results that are truly positive, while recall represents the ability of a test to correctly identify positive results and obtain the true positive rate. The F1 score achieves its optimal value at 1, indicating perfect precision and recall, and its worst value at 0 [21]. To evaluate the importance of each feature in the prediction model using RF, we applied the permutation-based feature importance measurement. Because RF classification model leads to a reduced correlation between each decision tree and can alleviate the overfitting issue, which is identified as a weakness of the decision tree [22], and permutation feature importance is defined as the decrease in a model's score when the feature's value is randomly permuted, breaking the relationship between the feature and the true outcome [23]. The significance of a feature is determined by computing the rise in the model's prediction error when the feature is permuted [24]. The measurement of

Table 3   Comparison of machine learning algorithm performance evaluation for conduction abnormalities (CA) post-TAVR

	Logistic Regression	Decision Tree	Random Forest	XGBoost
Accuracy	0.81	0.87	0.93	0.83
Precision	0.72	0.84	0.88	0.73
Recall	0.84	0.83	1.00	0.84
F1-score	0.77	0.83	0.94	0.76
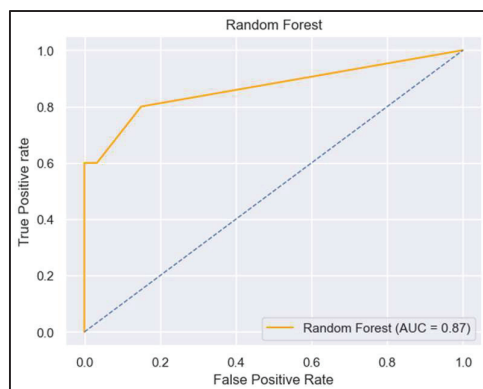
Figure 4   ROC curve and ROC AUC of random forest.

feature importance based on permutation was applied and explained in detail in a study on plasma biomarkers for risk stratification in heart failure patients with preserved ejection fraction [25].

The statistical analysis was performed using R software, version 4.3.1 (The R Foundation, Korea Yeungnam University), and machine learning techniques were implemented in python(v 3.9.13) using Jupyter Notebook.

3.3 Results of Prediction Model

Table 3 shows the results of the performance evaluation of the CA prediction classification model using four machine learning algorithms: logistic regression, decision tree, random forest, and XGBoost. Figure 4 shows the receiver operating characteristic (ROC) curve and ROC-AUC (area under the curve) results for the random forest model. Figure 5 lists the top 30 variables that predict CA after TAVR. The results of the feature importance analysis indicate that EKG data was heavily represented in the TAVR patient data.
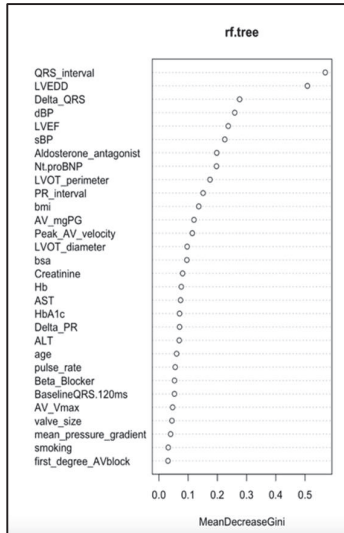
Figure 5   Visualization of random forest features importance.

4 TAVR Web Navigator

Predictive model results and patient data can be provided as a web dashboard via API and doctors can check predictive results and receive feedback.

4.1 Medical Twin Navigator's Web Dashboard Design

The medical twin navigator's web dashboard provides baseline information about the patient's personal information and eCRF, and provides personalized prediction accuracy for CA, as shown in Figures 6 and 7. The patient information and data displayed in the web dashboard is fictitious patient data to anonymize patient information.

   The main page of the web dashboard displays patient information (patient name, TAVR procedure date, age, gender, height, weight), eCRF information (vital data, underlying condition, lab data, valve data), prediction result ('normal' or 'abnormal'), probability (%) for the prediction result, explanation and feedback for the prediction result, ROC curve, features importance, and interpretation and feedback for features importance.

   Figure 6 displays the predicted outcome and its corresponding probability, indicating to the doctor the need for an in-depth patient test based on the
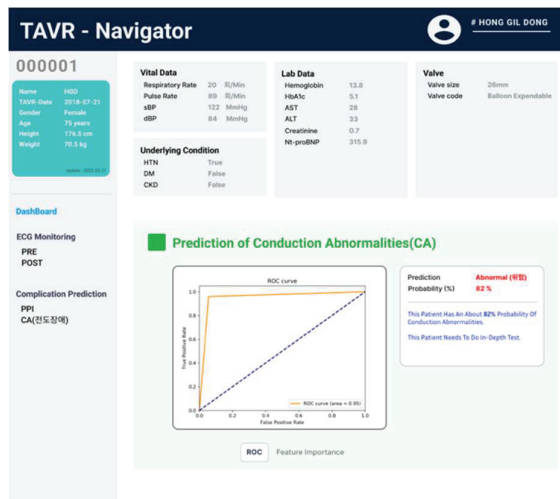
Figure 6   TAVR navigator dashboard (ROC curve).

prediction. Additionally, the ROC curve results are provided to assess the performance of the prediction model. Figure 7 displays the feature importance that influenced the prediction of the patient's complications. It suggests to the physician which patient features should be examined in-depth based on their importance. A separate menu is available to view the ECG data due to the feature importance results indicating an association between CA and ECG.

TAVR Web provides information and an explanation of the results to increase confidence in the accuracy of the prediction model. It also provides interoperability and the ability to predict complications for patients, and provides physicians with feedback on prediction results to improve clinical decision-making and help patients make informed procedural decisions [26].

5 Conclusions

In this study, based on the medical twin architecture layer, a complication prediction model based on machine learning is used at the simulation layer stage to predict complications. Factors affecting CA include clinic data, echocardiograms, and valves, although first degree AV block, PR interval, QRS interval, BaselineQRS > 120 ms, DeltaPR, and DeltaQRS items are the most numerous. While previous research has focused on simulations such as analysis and prediction in the real world, this study develops a RWD-based
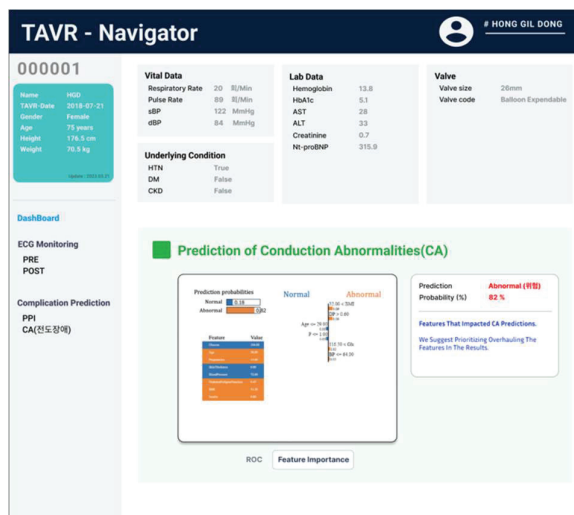
Figure 7 TAVR Navigator Dashboard (Feature Importance).

medical twin architecture to collect and analyse TAVR procedure data, predict complications, and provide feedback in a virtual environment.

Patient RWD data and simulation data before and after the procedure are synchronized, stored, managed, and utilized within the medical twin architecture to ensure traceability for follow-up. In addition, a feedback system compares post-TAVR results with the composite predictive model results provided during the simulation modelling phase to improve simulation accuracy, and analytical reports are provided to physicians and patients via a web dashboard to predict outcomes and improve clinical outcomes through feedback, allowing for continuous development of the complication model. The medical twin navigator architecture also gives you access to a variety of post-TAVR complication predictions, not just CA, and allows you to track changes in features across complication types.

The medical twin navigator simulates the entire process of TAVR in a virtual environment, it replicates a real medical system or body organ in a virtual space, enabling doctors to predict treatment effectiveness, visualize hard-to-see organ parts based on image data (such as CT or X-ray), prepare for unexpected incidents during the procedure, and examine organs in detail that cannot be seen in real life. The medical twin navigator can aid in medical education, help make optimal procedural decisions, predict possible complications for patients, and facilitate post-procedure follow-up.

Therefore, the medical twin navigator is expected to facilitate the entire TAVR process, from data collection to feedback, in a virtual environment. This will enable doctors to predict surgical outcomes and improve clinical outcomes based on objective and specific data during pre-procedure diagnosis and evaluation. The use of this technology will increase the convenience and safety of treatment for patients who require TAVR, minimize complications, and reduce the burden of medical costs.

Acknowledgements

Ethics

The need for consent was waived by the institutional review board of Gachon University Gil Medical Center (ethics approval code: GAIRB 2022-345).

References

[1] Dowling, C., Kondapally Seshasai, S. R., Firoozi, S., and Brecker, S. J. (2020). Transcatheter aortic valve replacement versus surgery for symptomatic severe aortic stenosis: A reconstructed individual patient data meta-analysis. Catheterization and Cardiovascular Interventions, 96(1), 158–166.

[2] van Mourik M.S.J., Abdelghani M., et al. Guideline-defined futility or patient-reported outcomes to assess treatment success after TAVI: what to use? Results from a prospective cohort study with long-term follow-up. *Open Heart* 2018, 5(2), e000879.

[3] Lebedev, A.V., et al. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clinical* 2014, 6, 115–125.

[4] Blaser, M.C., Atkins, S.K., Aikawa, E. Tissue Engineering to Study and Treat Cardiovascular Calcification. *Tissue-Engineered Vascular Grafts* 2020, 429–468.

[5] Czarny, M.J., Resar, J.R. Diagnosis and Management of Valvular Aortic Stenosis. Clinical Medicine Insights: *Cardiology* 2014, CMC-S15716. [https://doi.org/10.4137/CMC.S15716].

[6] Al-Ghonaim, M., Pannu, N. Prevention and treatment of contrast-induced nephropathy. *Techniques in Vascular and Interventional* Radiology 2006, 9(2), 42–49.

[7] Hanyu F., Takasaki K. Pancreatoduodenectomy. Springer: Tokyo, Japan 1997.

[8] Wal, P., Rathore, S., Aziz, N., Singh, Y.K., Gupta, A. Aortic stenosis: a review on acquired pathogenesis and ominous combination with diabetes mellitus. *The Egyptian Heart Journal* 2023, 75(1), 1–17.

[9] Baumgartner, H., Falk, V., Bax, J.J., De Bonis, M., Hamm, C., Holm, P.J., et al. 2017 ESC/EACTS Guidelines for the management of valvular heart disease. *Kardiologia Polska* (Polish Heart Journal) 2018, 76(1), 1–62.

[10] Sinning, J.M., Vasa-Nicotera, M., Chin, D., Hammerstingl, C., Ghanem, A., Bence, J.,Werner, N. Evaluation and management of paravalvular aortic regurgitation after transcatheter aortic valve replacement. *Journal of the American College of Cardiology* 2013, 62(1), 11–20.

[11] Généreux, P., et al. Incidence, predictors, and prognostic impact of late bleeding complications after transcatheter aortic valve replacement. *Journal of the American College of Cardiology* 2014, 64(24), 2605–2615.

[12] Kapadia, S.R., et al. Stroke after surgical versus transfemoral transcatheter aortic valve replacement in the PARTNER trial. *Journal of the American College of Cardiology* 2018, 72(20), 2415–2426.

[13] Sinning, J.M., et al. Renal function as predictor of mortality in patients after percutaneous transcatheter aortic valve implantation. JACC: *Cardiovascular Interventions* 2010, 3(11), 1141–1149.

[14] Auffret, V., et al. Conduction disturbances after transcatheter aortic valve replacement: current status and future perspectives. *Circulation* 2017, 136(11), 1049–1069.

[15] Grube, E., Sinning, J.M. The "big five" complications after transcatheter aortic valve replacement: do we still have to be afraid of them?. JACC: *Cardiovascular Interventions* 2019, 12(4), 370–372.

[16] Regueiro, A., Abdul-Jawad Altisent, O., Del Trigo, M., et al. Impact of new-onset left bundle branch block and periprocedural permanent pacemaker implantation on clinical outcomes in patients undergoing transcatheter aortic valve replacement: a systematic review and meta-analysis. *Circulation: Cardiovascular Interventions* 2016, 9(5), e003635.

[17] Kamel Boulos, Maged N., and Peng Zhang. "Digital twins: from personalised medicine to precision public health." Journal of personalized medicine 11.8 (2021): 745.

[18] Armeni, Patrizio, et al. "Digital Twins for Health: Opportunities, Barriers and a Path Forward." (2023).

[19] ISHAQ, Abid, et al. Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE access, 2021, 9, 39707–39716.

[20] Chicco, Davide, and Giuseppe Jurman. "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation." BMC genomics 21.1 (2020): 1–13.

[21] Grandini, Margherita, Enrico Bagli, and Giorgio Visani. "Metrics for multi-class classification: an overview." arXiv preprint arXiv:2008.05756 (2020).

[22] Jeong, Minsu, et al. "Ensemble Deep Learning Model using Random Forest for Patient Shock Detection." KSII Transactions on Internet & Information Systems 17.4 (2023).

[23] Truong, Vien T., et al. "Machine learning method for predicting pacemaker implantation following transcatheter aortic valve replacement." Pacing and Clinical Electrophysiology 44.2 (2021): 334–340.

[24] Breiman, L. RandomForests.MachLearn. 2001, 45, 5–32.

[25] Chirinos JA, Orlenko A, Zhao L, BassoMD, CvijicME, LiZ, et al. Multiple plasma biomarkers for risk stratification in patients with heart failure and preserved ejection fraction. J Am Coll Cardiol. 202, 75, 1281–1295.

[26] Shah, Adnan Muhammad, et al. "Examining different factors in web-based patients' decision-making process: systematic review on digital platforms for clinical decision support system." International Journal of Environmental Research and Public Health 18.21 (2021): 11226.

Biographies

Se-Min Hyun received her bachelor's degree in computer engineering from Gachon University in 2020, and she is currently working towards her Master's degree at the Gachon University, Republic of Korea. Her research areas include healthcare AI, machine learning, and data analysis.

KangYoon Lee received a B.Sc. in Electronics Engineering and an M.Sc. in Computer Science from Yonsei University, Seoul, South Korea in 1986 and 1996, respectively, and a Ph.D. in IT Policy Management from Soongsil University, Seoul in 2010.

From 2008 to 2014, he was the Director of the IBM Korea Laboratory for the Ubiquitous Computing and Software Solutions Lab, and he was promoted to the Leader of the IBM Watson Business Unit, South Korea in 2014. Since 2016, he has been a Professor with the Computer Engineering Department, Gachon University College of IT Convergence, Seongnam-Si, Republic of Korea. Dr Lee has been serving as the Director at the Gachon Institute of Artificial Intelligence since 2016. His research interests include cognitive computing, healthcare advising, IoT platforms, and industry transformation.