ORIGINAL ARTICLE

Eastern Journal of Medicine

East J Med 29(1): 118-128, 2024
DOI: 10.5505/ejm.2024.54775

Detection of Right Ventricular Dysfunction Using LogNNet Neural Network Model Based on Pulmonary Embolism Data Set

Mehmet Tahir Huyut[1*], Andrei Velichko[2], Maksim Belyaev[2], Şebnem Karaoğlanoğlu[3], Bunyamin Sertogullarindan[3], Abdussamed Yasin Demir[4]

[1]Erzincan Binali Yıldırım University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, [2]4000 Erzincan, Turkey
[2]Petrozavodsk State University, Institute of Physics And Technology, 185910 Petrozavodsk, Russia
[3]İzmir Katip Çelebi University, Medical Faculty, Department of Pulmonary Medicine, Izmir, Turkey
[4]Erzincan Binali Yıldırım University, Faculty of Medicine, Department of Genetics, 24000 Erzincan, Turkey

ABSTRACT

The high association of right ventricular dysfunction (RVD) with mortality in patients with acute pulmonary embolism (PE) remains an important health problem. In this respect, rapid, economical and highly-accurate detection of risk factors for early diagnosis of RVD in patients with PE is expected to greatly benefit the diagnosis and treatment of the disease and contribute significantly to the reduction of mortality. The aim of this study is to identify the most effective features from the PE dataset for RVD diagnosis, using a special-algorithm for the LogNNet reservoir neural-network. The cohort of patients diagnosed with acute PE in the last five years in our hospital was retrospectively analyzed and the data in accordance with our criteria were recorded. A total of 163 patients' data were accessed and the patients had 20 characteristics. RVD was diagnosed in 27 of these patients. 78-79 years of age was found to be an important threshold for the diagnosis of RVD. The LogNNet model revealed that older age, comorbidities and coronary-heart disease greatly increased the risk of RVD. The model also found that individuals with diabetes and COPD were at higher risk of RVD, while individuals with malignancies were at lower risk of RVD. However, the model found that unilateral-thrombus increased the risk of RVD more than bilateral-thrombus. The risk of RVD is high in PE patients with unilateral-thrombus. In addition, PE patients with comorbidities such as coronary heart disease, diabetes and COPD are at high-risk for RVD and should be followed closely.

Keywords: Right ventricular dysfunction, pulmonary embolism, thrombosis, LogNNet, artificial intelligence, supervised machine learning models

Introduction

Pulmonary embolism (PE) has been recognized as one of the leading causes of death due to cardiovascular diseases (1). Kasper et al. found that PE increased the risk of death in patients presenting with hemodynamic deterioration (2).

Cires-Drouet et al. stated that the degree of pulmonary artery occlusion caused by PE and the rise in pulmonary vascular resistance indicate the response of the right ventricle during acute PE (3). Furthermore, they reported that increased wall tension may cause myocardial ischemia by decreasing coronary artery perfusion. Lualdi and Goldhaber noted that these changes may lead to hemodynamic instability and hypotension, which in severe cases may lead to syncope and collapse (4).

Heit et al. and Raskob et al. emphasized that PE is characterized by obstruction of the pulmonary arteries by a blood clot, usually originating from the lower extremities, and this can lead to serious cardiopulmonary problems (5,6). Heit et al. stated that right ventricular dysfunction (RVD) can occur as a major complication of PE and is associated with increased morbidity and mortality due to hemodynamic instability (5). Similarly, Cho et al. reported that RVD increased the short-term mortality rate even in hemodynamically stable patients with acute PE (7).

Furthermore, some studies have noted that routine interventions performed in emergency care, such as fluid administration and endotracheal intubation, can cause rapid clinical deterioration in patients with RVD and pulmonary hypertension (8, 9). Therefore, Crager

*Corresponding Author: Mehmet Tahir Huyut, Erzincan Binali Yıldırım Universitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, 2400, Erzincan, Türkiye
E-Mail: mehmettahirhuyut@gmail.com, Phone: 0553 865 65 99
Mehmet Tahir Huyut: 0000-0002-2564-991X, **Andrei Velichko:** 0000-0002-9341-1831, **Maksim Belyaev:** 0000-0002-1771-6502, **Şebnem Karaoğlanoğlu:** 0000-0002-5740-0326, **Bunyamin Sertogullarindan:** 0000-0002-1478-1990, **Abdussamed Yasin Demir:** 0000-0002-0420-5017
Received: 12.12.2023, **Accepted:** 30.12.2023

and Humphreys stated that understanding the risk factors affecting acute decompensated RVD is important to avoid pitfalls in diagnosis and treatment that can trigger significant morbidity and mortality (8).

In accordance with the 2019 European Society of Cardiology (ESC) guidelines, the detection of RVD in patients with PE can be determined by the following methods: Cardiac biomarkers, hemodynamic stability, echocardiography or computed tomography and clinical prediction scores (10). Lang et al. reported that patients with thrombus in both the main pulmonary artery and bilateral pulmonary arteries had a high risk of RVD (11).

It has been reported that studies with larger samples are needed to confirm these findings regarding the detection of RVD in patients with PE and to investigate additional risk factors (10). Cires-Drouet et al. noted that despite all the commentary on the detection of RVD, the risk factors affecting RVD and changes in acute PE are not well defined (3). Similarly, Piazza and Goldhaber noted that not all patients with acute PE and RVD had a history of coronary artery disease or right-left heart failure (12). Therefore, they emphasized the continued need to understand the mechanism of how RVD may occur in patients with PE. In the light of these approaches, the importance of early and accurate identification of RVD is understood in order to intervene in a timely and effective manner in the patient population with acute PE and to reduce the mortality of the disease. However, detection of RVD is costly and time-consuming in line with the current practices mentioned above. Furthermore, current practices require specialized equipment and specialized knowledge on the part of health systems and providers. Detection of diseases and prediction of possible complications without the use of advanced high-tech devices and techniques can help to address various problems such as patient comfort as well as health system providers and economic shortages (from our sensor publication). With this perspective, rapid, economical and highly accurate detection of risk factors for early detection of RVD is expected to be of tremendous benefit to the diagnosis and treatment process of the disease and will contribute significantly to reducing its mortality.

When the literature was reviewed, we did not find a study on early diagnosis of RVD in patients with PE using artificial intelligence methods. On this basis, our study using the LogNNet artificial intelligence model will be an important resource for early diagnosis of RVD based on PE data. Artificial intelligence (AI) models are methods that solve complex relationship structures between variables in the detection of diseases and are widely used in various disciplines (13-16). In health sciences, AI approaches are increasingly used in real-time decision making to improve the quality of healthcare, reduce drug costs, increase patient confidence and comfort, etc. (17-20).

In Velichko presented the LogNNet neural network as a new classifier model from deep neural network models based on the MNIST database (21). Later, Velichko demonstrated the application of LogNNet to identify risk factors for disease diagnosis based on a patient dataset (22). Moreover, in some recent studies (15-18), LogNNet has been successfully applied to predict the diagnosis, prognosis and mortality of COVID-19 disease. The LogNNet deep neural network model approach can be summarized as follows; the system is a feed-forward network that maximizes the classification accuracy by passing the feature vector through a specially designed reservoir matrix and transforming it into a feature vector of different dimension (23). Izotov et al. showed that the higher the entropy of a chaotic map filling a reservoir matrix, the better the classification accuracy (24). Furthermore, the use of RAM is significantly reduced with the LogNNet neural network model, which is based on a procedure to optimize the chaotic map parameters (23). This discovery showed that LogNNet can be used in Internet of Things (IoT) mobile devices. Indeed, in a recent study Velichko et al. the IoT Application of LogNNet for the diagnosis of COVID-19 was successfully presented.

This study will investigate the factors related to RVD based on PE patient dataset. In addition, RVD diagnoses of patients with PE will be predicted with the LogNNet classifier model. Finally, the performance of the LogNNet deep neural network model in RVD prediction will be determined with various criteria. This study aims to provide a faster and cost-effective alternative tool for early detection of RVD based on PE data.

Materials and Methods

This retrospective single-center study was conducted in accordance with the 1989 Declaration of Helsinki. The records of the last five years in our hospital were retrospectively scanned digitally and patients diagnosed with PE were filtered. The characteristics of these patients were analyzed and 20 variables without missing data were identified (Table 1). Age, gender and comorbidities of the patients were also recorded. Approval for the use of the data set used in this study was obtained by the Ethics Committee of our center on October 21, 2021 (approval number: 0439).

Characteristic of Participants, Exclusion Criteria and Define Datasets: The patient cohort of this

study consisted of patients diagnosed by contrast-enhanced multidetector computed tomography (CT) with echocardiography and lower extremity venous Doppler ultrasound reports. We excluded pregnant patients, those for whom echocardiography and Doppler ultrasound reports were not available, individuals with chronic thromboembolic pulmonary hypertension, and those with known right heart dysfunction. Patients younger than 18 years were excluded. A total of 163 patients, 80 (49.1%) females and 83 (50.9%) males, were included in the study. RVD was observed in 27 patients (16%).

Standard transthoracic echocardiography was performed in all patients on admission, preferably within 48 hours of diagnosis. Thrombus location was determined according to official radiology reports. Echocardiography and lower extremity venous Doppler ultrasonography reports were analyzed for right ventricular dysfunction and deep vein thrombosis. Right ventricular enlargement was defined as baseline RV diameter greater than baseline LV diameter or baseline and mid-RV diameters >41 mm and >35 mm, respectively. We used the following criteria to diagnose right ventricular dysfunction: Tricuspid annular plane systolic excursion (TAPSE) < 16 mm, Tricuspid Lateral Annular Systolic Velocity < 9. Doppler tissue imaging showed a 5 cm/sec, flattened interventricular septum in parasternal short axis view, dilated inferior vena cava, decreased inspiratory collapse in subcostal view, pulmonary ejection acceleration time < 60 msec, and peak systolic gradient between the mid-systolic notch and the right ventricle and right atrium < 60 mmHg (10, 11).

The 20 characteristics of the dataset of patients diagnosed with PE in this study are given in Table 1. RVD diagnosis was coded as 1: Presence, 0: Absence. The database contains a total of 163 vectors distributed into two classes, Class_1 (0: Absence) and Class_2 (1: Presence). Class_1 contains 136 elements and Class_2 contains 27 elements. The class data is unbalanced as there is a large difference between the number of elements of class 1 and the number of elements of class 2. In such cases, the classification accuracy (ACC) metric does not make much sense. Therefore, we use the Matthews Correlation Coefficient (MCC) metric for classification accuracy (see Methods).

Feature 2 (Age) is real and is presented in the form of values in the range of 22-93 years; feature 12 (Appearance of DVT) has 3 gradations; the remaining features are presented in binary form. The RVD examination result was also presented in binary format (1: Presence, 0: Absence). DVT: Deep vein thrombosis.

Classification Metrics: It has been noted (25, 26) that the most appropriate metric for classification performance when class data is imbalanced is the Matthews Correlation Coefficient (MCC). The MCC is basically a correlation coefficient between observed and predicted binary classifications; it takes values between -1 and +1. A value of +1 represents a perfect prediction; 0 is no better than a random guess, and -1 indicates total disagreement between prediction and observation.

K-folding was used for all metric results and calculated by averaging multiple blends. The basic algorithm for calculating the metrics is as follows:

1) The initial base was shuffled so that each layer contains vectors from different classes;

2) The base was divided into 10 parts (K-fold), i.e. 9 training and 1 test part, alternating 10 times between them. The MCC metric was calculated based on the training part of the data.

3) After mixing the samples on a basic basis, K-folding was performed 10 times.

4) As a result of multiple comparisons, mean values of the MCC metric were calculated.

Architecture of LogNNet and Feature Selection Method: LogNNet is a neural network based on the technology of "reservoir computing with auto-generation of weighting coefficients". The LogNNet diagram is shown in Figure 1.

As a first step, a reservoir in the form of a matrix W is filled with a sequence Xn/L_1, where Xn is a chaotic sequence and is generated by a coherent generator according to formula (1).

$$\begin{cases} x_1 = L_4 \\ x_{n+1} = (L_2 - L_3 \cdot x_n) \bmod L_1 \\ N = \lceil L_5 \cdot 21 \rceil \\ z_1 = L_9 \\ z_{n+1} = (L_7 - L_8 \cdot z_n) \bmod L_6 \end{cases} \quad (1)$$

The parameters of the first generator are L_1, L_2, L_3, L_4. Individual features can be removed by zeroing the corresponding columns of the W matrix. The total number (N) of features used in the model is determined by the optimized hyperparameter L_5. The method of feature selection and the use of the second adaptive generator with parameters L_6, L_7, L_8, L_9 is presented in section 2.4.1.

Inputs called feature vector V are fed to the LogNNet N:P:H:M classifier (stage 2). This vector contains N=21 coordinates v1, v2, ..., vN, taking into account the features in Table 1. The output of the LogNNet classifier is the class of the input object from the M

possible classes. In this study, we have two output classes Right Heart Failure (1: Present, 0: Absent). The vector v is transformed into an N-dimensional vector Y (step 3) and each of its components is normalized (step 4). Then, the N.P-dimensional matrix W is multiplied by the vector Y (step 5) to obtain a P-dimensional vector S (step 6). This vector is normalized and transformed into a vector Sh of size (P +1) with a bias element Sh[0] = 1 (stage 7). The vector S_h is fed into a two-layer linear classifier, while the hidden layer contains H neurons, and the output layer S_{out} contains M=2 neurons, which is equal to the number of possible classes (stage 8).

LogNNet is trained on the training database (step 9) and tested on the test database (step 10) according to the number of epochs Ep=L_{10}. The test database is used to calculate the measurement criteria using the K-folding method described in section 2.2 (stage 11). The reservoir size P=L_{11} and the number of neurons in the hidden layer H=L_{12} were optimized using the hyperparameters L_{11} and L_{12}.

The addition of a chaotic component to the training data leads to the creation of a more generalized model that gives better results for classification after mixing the sample data sets. Therefore, a slight chaotic deviation d_{vi} was added to the training data at each vector coordinate, with the hyperparameter L_{13} in the range 0-1 according to the following formula:

$$dv_i = (v\max_i - v\min_i) \cdot L_{13} \cdot random(-1...1) \quad (2)$$

where ($vmax_i$-$vmin_i$) represents the difference between the maximum and minimum values of the elements of the input feature vector and *random* (-1…1) represents a chaotic deviation generator ranging from -1 to 1. No chaotic variance was added to the test data. We also duplicated the number of small classes in the training set for the balancing analog. For balancing, an additional hyperparameter L_{14} in the range 0-1 was introduced, which regulates the number of replicates. Thus, the LogNNet model has 14 (L_1-L_{14}) hyperparameters.

Selection of LogNNet Hyperparameters, Feature Selection Technique: LogNNet hyperparameters were selected in the training database by the particle swarm method. The four selected hyperparameters (L_1, L_2, L_3, L_4) correspond to the coefficients of the linear adaptive generator specified in equation 2, where the first filling of the reservoir is realized row by row with x_n/L_1 elements. The feature vector was optimized with the other five hyperparameters (L_5, L_6, L_7, L_8, L_9). Determining the number of key features extracted from the input vector and taken into account in subsequent computations formed the basic approach of the optimization. Features that are not considered in the calculation are eliminated by setting the coefficients of the reservoir matrix to zero. In particular, all coefficients in the column with a number matching the number of the deleted feature are set to zero. According to Equation (1), the number of important features N is determined by the hyperparameter L_5, where L_5 is a number between 0 and 1. Under the condition that N cannot be less than 1 and greater than the maximum number of features, N_{max}=21. The hyperparameters L_6, L_7, L_8, L_9 are responsible for selecting the number of key features using a special algorithm. These hyperparameters determine the order of the chaotic coherent generator $\tilde{x}_n$ (1), whose values are translated into the number of basic features. Their optimization allows to identify the best features and also to identify a certain number of features *(N)* remaining in the model by adjusting the L_5 hyperparameter.

The particle swarm optimization method allows the selection of any number of hyperparameters by determining in advance the fitness function and the limits of variation of the hyperparameters. We used *MCC* as a fitness function. Method constants (inertia coefficient =0.5, learning rates local =2, global =2), the number of particles and iterations varied from 50 to 300.

Threshold Approach For One Feature: The single feature threshold approach involves a simple model in which there is a single threshold V_{th} separating the two classes. The separation algorithm can be represented by a formula in which there are two types of dependencies (Type 1 and Type 2).

$$\begin{cases} Type\ 1: if\ feature\ value > V_{th}\ then\ RVD\ (1: presence)\ else\ RVD\ (0: absence) \\ Type\ 2: if\ feature\ value > V_{th}\ then\ RVD\ (0: presence)\ else\ RVD\ (1: absence) \end{cases} \quad (3)$$

The search for V_{th} was carried out by sequential search within the limits of changes in the feature, with the determination of the maximum MCC value.

For binary signals, the classification result does not depend on the threshold value if it is in the range 0< V_{th} <1. Therefore, for binary signals it is convenient to introduce a truth table for Type 1 and Type 2.

Results

Correlation dependencies: The correlation distribution of the traits with each other is presented in Figure 2a. The levels of correlation between the binary variables are presented in Table 2. The highest Pearson correlation coefficient is observed in the feature pairs presented in Table 2. Traits 9 and 11 have a very high correlation with each other ~0.93. Then, correlations of ~0.88 and ~0.86 were found between traits 9 and 12 and 11 and 12 respectively.

Table 1: List of Features Indicating The Range of Changes In Values

№	Feature name	Range of changes in features values (Minimum - Maximum)
1	Sex (1:Male, 0:Female)	(0-1)
2	Age	(22-93)
3	Thrombus Main Pulmonary (1: Presence, 0: Absence)	(0-1)
4	Thrombus Lobe Arteries (1: Presence, 0: Absence)	(0-1)
5	Thrombus Segment (1: Presence, 0: Absence)	(0-1)
6	Thrombus Subsegment (1: Presence, 0: Absence)	(0-1)
7	Thrombus (0: Unilateral, 1: Bilateral)	(0-1)
8	Right Ventricular Involvement (1: Presence, 0: Absence)	(0-1)
9	DVT (1: Presence, 0: Absence)	(0-1)
10	DVT Distal Ven (1: Presence, 0: Absence)	(0-1)
11	DVT Proksimal Ven (1: Presence, 0: Absence)	(0-1)
12	Appearance of DVT (Absence: 0 Unilateral :1 Bilateral :2)	(0-2)
13	Comorbid Disease (1: Presence, 0: Absence)	(0-1)
14	Malignancy (1: Presence, 0: Absence)	(0-1)
15	Diabetes Mellitus (1: Presence, 0: Absence)	(0-1)
16	Hypertension (1: Presence, 0: Absence)	(0-1)
17	Coronary Artery Disease (1: Presence, 0: Absence)	(0-1)
18	COPD (1: Presence, 0: Absence)	(0-1)
19	Asthma (1: Presence, 0: Absence)	(0-1)
20	Cerebrovascular Occlusion (1: Presence, 0: Absence)	(0-1)

The correlation levels between the features and the main diagnosis are shown in Figure 2b. The absolute values of the Pearson correlation coefficient are also given in the graph. The strongest correlations with the diagnosis of SVD were seen between features 2, 7, 13, 17. The sign of the correlation coefficient indicates whether the trait belongs to Type 1 (positive coefficient) or Type 2 (negative coefficient). Therefore, features 2, 13, 17 belong to Type 1 and feature 7 belongs to Type 2.

Efficiency Based On One Feature: The MCC values for the performance of the features in identifying type 1 and type 2 class labels according to the two models (Threshold Approach, LogNNet) are shown in Figure 3.

Based on the results of the two models, features can be divided into three group. Group 1 - the strongest single features, Group 2 - weak single features, Group 3 - very weak single features.

The relationships in Figure 3 correlate with the Pearson coefficient in Figure 2b. Strong features from group 1 have the highest Pearson correlation coefficient. In the tables, features of groups 1 and 2 will be denoted by the symbols ** and *, respectively.

Feature 2 (Age) is real and is presented as values in the range 22-93 years. The distribution of vectors in the database by age value in two groups is shown in Figure 4a. The threshold curve constructed using two models is shown in Figure 4b-c. Threshold Approach and LogNNet demonstrate the presence of one threshold in the region of Vth = 78 and 79 years, respectively (Table 4). Both models belong to Type 1 and, if age is more than Vth years, patients tend to have the disease. The Threshold Approach model and LogNNet have the same mask and threshold value, however, the MCC for LogNNet is lower, since the calculation involves the K-fold method with mixing, and Threshold Approach gives maximum accuracy on all data, so its values are overestimated. Thus, the LogNNet model, in our opinion, gives MCC values close to reality without over-tuning to the nuances of the base. Feature 12 has an extremely small MCC ~ 0.03 according to two models, Vth lies in the range of 0.01-0.99 (Type 2). The performance of all features in the detection of SVD compared to the two models is given in Table 5.

Different combinations of features may lead to different MCC performance in SVD detection. With this in mind, three sets of optimizations were performed for LogNNet parameters (L1-L14) and the

Table 2: A list of Pairs of Features That Have The Strongest Correlation With Each Other

Feature 1	Feature 2	Pearson correlation Coefficient
9	11	0,9276
9	12	0,88856
11	12	0,86029
2	16	0,40938
10	12	0,39528
9	10	0,39121
13	16	0,36987
3	7	0,36202

Table 3: Grouping Features According To Their Performance (MCC) in SVD Detection

Features of Group 1	Features of Group 2	Features of Group 3
2**	1*	3
7**	6*	4
13**	8*	5
17**	10*	9
	11*	12
	14*	19
	15*	20
	16*	
	18*	

Table 4: Performance of The Second Characteristic (age) in Detecting SVD and Associated Cut-Off Value According To The Two Models

	Threshold Approach			LogNNet		
Feature	MCC	V_{th}	Type	MCC	V_{th}	Type
2**	0,30696	78	Type 1	0,229	79	Type 1

results for the classification performance of the model using varying number of feature combinations are presented in Figure 5 and Table 6.

From these findings, the following conclusions can be drawn:

1) When N=1 (the number of features used) to detect SVD, LogNNet detects the most important feature as 17 according to all optimization results.

2) When N=2, LogNNet reveals the most important feature pair as 2-17 according to all optimization results.

3) When N=3, LogNNet identified the most important feature triplet as 2-7-17 for all optimization results.

4) In other combinations with N>3, there was no significant increase in the classification performance (MCC) of the LogNNEt model for SVD detection. Therefore, we can say that N=3 for the saturation and maximum number of effective features of the LogNNet model for SVD detection.

Discussion

PE patients diagnosed with RVD have been associated with premature death, even in conditions of likely low risk based on available clinical severity scores (27). Beckman et al. recognized PE as one of the leading causes of death due to cardiovascular disease (1). Similarly, other studies have reported that RVD as a hemodynamic impairment increases short-term mortality in a cohort of acute PE patients (3, 4, 7). Kobayashi et al. reported that RVD was seen in approximately 25% of acute PE cases (28).

Therefore, identifying factors associated with RVD in the PE population may assist health services in the

Table 5: Performance of all The Features In Detecting SVD According To The Two Models

	Threshold Approach	LogNNet	
Feature	MCC	MCC	Type
1*	0,04131	0,048	Type 1
3	0,01353	0	Type 2
4	0,00508	0	Type 1
5	0,0042	0	Type 2
6*	0,07283	0,061	Type 2
7**	0,2313	0,216	Type 2
8*	0,05881	0,065	Type 1
9	0,03624	0,023	Type 2
10*	0,10627	0,088	Type 1
11*	0,07037	0,053	Type 2
13**	0,21593	0,199	Type 1
14*	0,17252	0,171	Type 2
15*	0,16059	0,119	Type 1
16*	0,07053	0,049	Type 1
17**	0,34674	0,323	Type 1
18*	0,20784	0,11	Type 1
19	0,02363	0,01	Type 1
20	0,02484	0,022	Type 2

Table 6: Combinations of Features With Variations In Their Number

LogNNet: optimization 1			LogNNet: optimization 2			LogNNet : optimization 3		
N	MCC	List of features	N	MCC	List of features	N	MCC	List of features
1	0.317	17	1	0.323	17	1	0,323	17
2	0.364	2-17	2	0.359	2-17	2	0,325	2-17
3	0.430	2-7-17	3	0.400	2-7-17	3	0,429	2-7-17
4	0.409	1-2-7-17	4	0.372	1-2-13-17	4	0,351	7-13-14-18
5	0.358	10-11-12-13-14	5	0.378	1-2-3-7-17	5	0,359	1-2-3-5-17

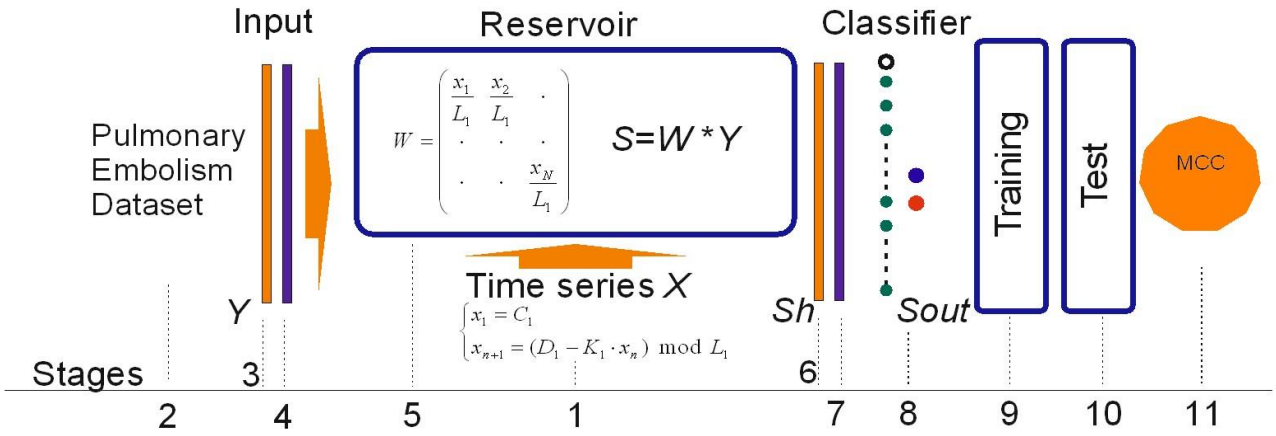

Fig. 1. LogNNet architecture and main stages of calculation

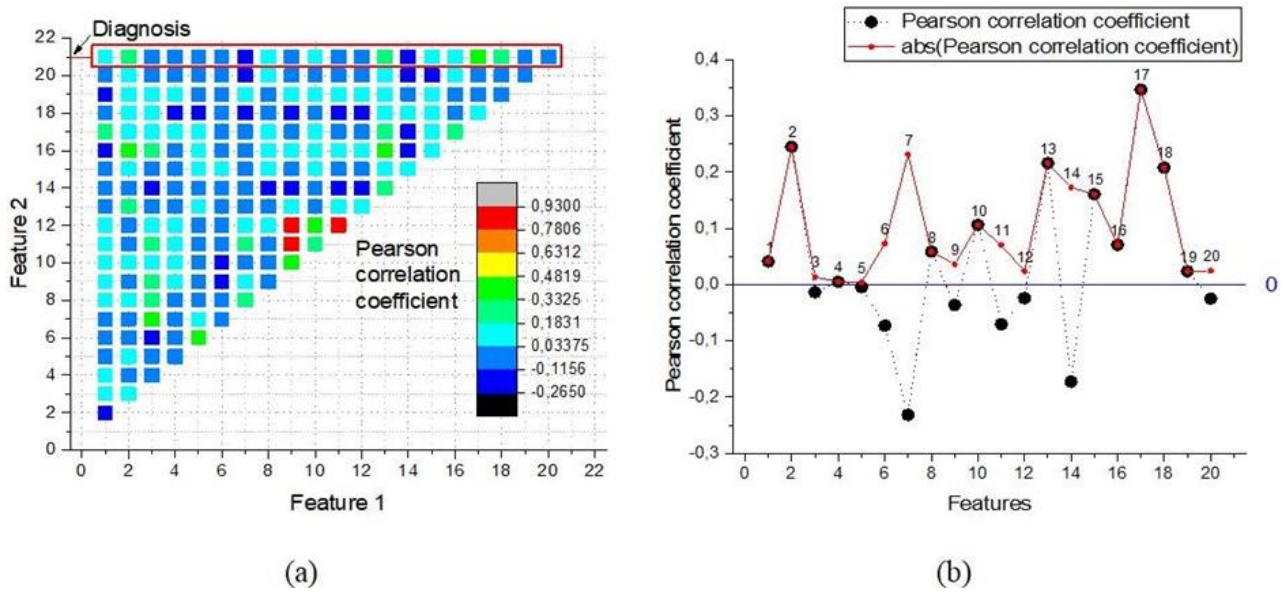

Fig. 2. The dependence of the correlation of features with each other and with the main diagnosis (a), The correlation between features and the main diagnosis (b)

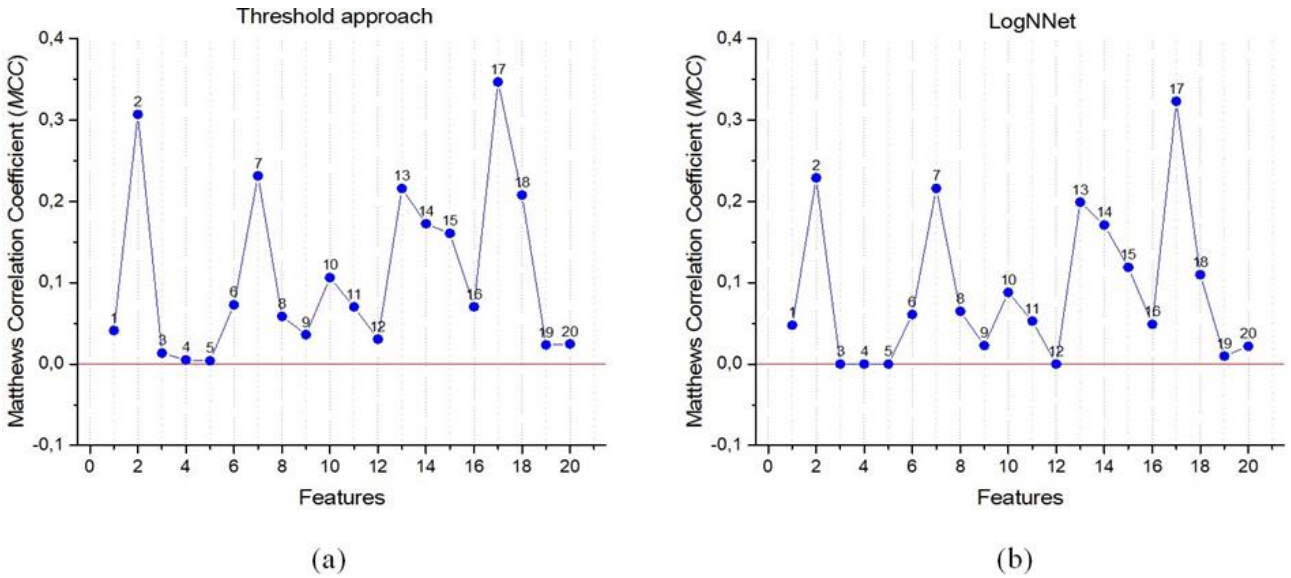

Fig. 3. Dependence of Matthews Correlation Coefficient (*MCC*) on the feature for Threshold Approach (a), LogNNet (b)

early detection and monitoring of such high-risk patients. Indeed, Becattini et al. stated that understanding the risk factors affecting decompensated RVD is important to avoid pitfalls in diagnosis and treatment that can trigger morbidity and mortality (29).

Under these approaches, early and accurate identification of RVD is the main concept of this study in order to intervene in the acute PE patient population in a timely and effective manner and to reduce the mortality of the disease. In our study, important features for the diagnosis of RVD were extracted from the PE patient database with the LogNNet neural network model. The mystery of these features in the diagnosis of RVD in acute PE could be explained to a significant extent.

In this study, a high positive correlation was found between proximal DVT, DVT type and the presence of deep vein thrombosis (DVT) (Table 2, Figure 2). The most important characteristics for the diagnosis of RVD were age, unilateral or bilateral thrombus, presence of comorbidities and coronary heart disease (Table 3, Figure 3).

In particular, 78-79 years of age was an important threshold for the diagnosis of RVD (Table 4). The LogNNet model revealed that older age,

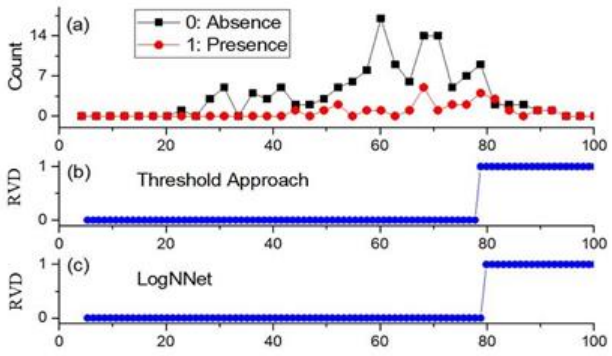

Fig. 4. Distribution of vectors by age and threshold curve for feature 2

comorbidities and coronary heart disease greatly increased the risk of RVD (Table 5). Furthermore, the model identified that individuals with diabetes and COPD were at higher risk for RVD, while individuals with malignancy were at lower risk for RVD (Table 3 and Table 5). However, when analyzed in terms of localization, the model sold that unilateral thrombus increased the risk of RVD more than bilateral thrombus.

Similarly, in a similar study, Cires-Drouet al. reported that the risk of RVD was higher in those with a history of heart disease compared to those without (3). They also stated that RVD may occur as a result of pulmonary embolism and comorbidities, supporting the findings in this study. Similarly, in a study by Kosmala et al. diabetes mellitus was found to be associated with impaired RVD independently of comorbid hypertension (30).

In this study, three features (age, coronary heart disease and thrombus localization) were found to be sufficient for the saturation of the LogNNet model and the maximum number of effective features for RVD detection (Table 6 and Figure 5). Taken together, these features can be used with high accuracy for the detection of RVD. We believe that the main reason for the decrease in the MCC values of the LogNNet model compared to the cut-off value approach in the detection of RVD is due to the random division of the data sets into layers and the optimization process. However, this was used as a solution to the overfitting problem of our model and we believe that the results that we think is closer to the truth.

Similarly, Martin et al. and Barco et al. reported that acute PE is an important cause of cardiovascular-related mortality and found that age and gender tended to increase mortality (31, 32).

Similarly, several studies have also revealed that the localization of the thrombus in the artery is an independent risk factor for RVD (7, 10, 11). However, in contrast to the findings in this study, one

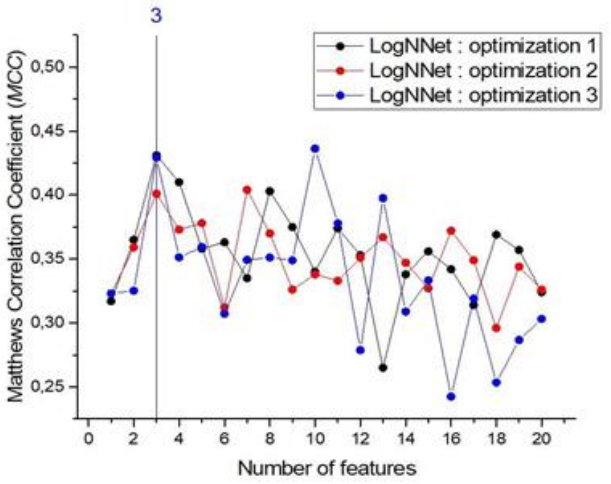

Fig. 5. MCC distribution when the number of features varies from 1 to 20

study found a higher proportion of bilateral thrombi in patients with RVD compared to those without RVD. Kurnicka et al. identified the presence of bilateral deep vein thrombosis as a risk factor for RVD in PE patients. This finding does not overlap with the findings in our study (37). We think that this is due to the limited number of patients with deep vein thrombosis in our study.

In another study Sanchez et al. the presence of malignancy was found to be lower in RVD cases (33). This result is consistent with the findings in our study. Similarly, one study confirmed that the presence of malignancy reduced the risk of developing RVD by 0.245-fold (34). A possible explanation for this finding is that malignancy may trigger incidental local thrombosis, which is more common in segmental and subsegmental pulmonary arteries (35).

In some other studies (5, 28) supporting the findings of this study, the frequency of deep vein thrombosis did not differ significantly between the groups with and without right ventricular dysfunction. Furthermore, in one study, the presence of distal or proximal deep vein thrombosis was not significantly associated with right ventricular dysfunction (36). These findings are consistent with the results of our study.

It can be said that especially 78-79 years of age may be an important threshold value to be considered for the diagnosis of RVD in acute PE cases. The risk of RVD is high in PE cases with unilateral thrombus. In addition, PE patients with comorbidities such as coronary heart disease, diabetes mellitus and COPD have a high risk of RVD and should be followed up closely. Cohort studies with larger samples are needed to confirm these findings and to investigate additional risk factors.

Acknowledgments: We would like to thank the management of Izmir Training and Research Hospital for providing access to the data used in this study.

Conflict of Interest: The authors declare no interests. There is no conflict of interest related to this article.

Funding: There was no funding for this study.

References

1. Beckman MG, Hooper WC, Critchley SE, et al. Venous thromboembolism. A public health concern, Am J Prev Med. 2010;38(4):495–501.
2. Kasper W. Konstantinides S, Geibel A, et al. Management strategies and determinants of outcome in acute major pulmonary embolism: results of a multicenter registry, J Am Coll Cardiol. 1997;30(5):1165–1171.
3. Cires-Drouet R, LaRocco A, Soldin D, et al. Left ventricular systolic dysfunction during acute pulmonary embolism. Thrombosis Research. 2023;223:1–6.
4. Lualdi JC, Goldhaber SZ. Right ventricular dysfunction after acute pulmonary embolism: pathophysiologic factors, detection, and therapeutic implications, Am Heart J. 1995;130(6): 1276-1282.
5. Heit JA, Spencer FA, White RH. The epidemiology of venous thromboembolism. J Thromb Thrombolysis. 2016;41(1):3-14.
6. Raskob GE, Angchaisuksiri P, Blanco AN, et al. Thrombosis: a major contributor to global diseaseburden. Arterioscler Thromb Vasc Biol. 2014;34:2363-2371.
7. Cho JH, Kutti-Sridharan G, Kim SH, et al. Right ventricular dysfunction as an echocardiographicprognostic factor in hemodynamically stable patientswith acute pulmonary embolism: a meta-analysis. BMC Cardiovasc Disord. 2014;14(64): 2-9.
8. Crager SE, Humphreys C. Right Ventricular Failure and Pulmonary Hypertension. Emerg Med Clin N Am. 2022;40:519–537.
9. Vallabhajosyula S, Shankar A, Vojjini R, et al. Impact of right ventricular dysfunction on short-term and long-term mortality in sepsis: a meta-analysis of 1,373 patients. CHEST. 2021;159:2254–63.
10. Konstantinides SV, Meyer G, Becattini C, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS): The Task Force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). Eur Heart J. 2019;41:543–603.
11. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantificationby echocardiography in adults: an update from theAmerican Society of Echocardiography and the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging. 2016;17(4):412.
12. Piazza G, Goldhaber SZ. Pulmonary embolism in heart failure, Circulation. 2008;118(15):1598-1601.
13. Huyut MT, Huyut Z. Forecasting of Oxidant/Antioxidant levels of COVID-19 patients by using Expert models with biomarkers used in the Diagnosis/Prognosis of COVID-19. Int Immuno. 2021;100:108127.
14. Huyut MT, Üstündağ H. Prediction of diagnosis and prognosis of COVID-19 disease by blood gas parameters using decision trees machine learning model: A retrospective observational study. Med Gas Res. 2022;12:60–66.
15. Huyut MT, Velichko A. Diagnosis and Prognosis of COVID-19 Disease Using Routine Blood Values and LogNNet Neural Network. Sensors. 2022;22:4820.
16. Huyut MT, Velichko A, Belyaev M. Detection of Risk Predictors of COVID-19 Mortality with Classifier Machine Learning Models Operated with Routine Laboratory Biomarkers. Appl Sci. 2022; 12: 12180.
17. Huyut MT, Velichko A, Belyaev M, Izotov Y, Korzun D. Machine Learning Sensors for Diagnosis of COVID-19 Disease Using Routine Blood Values for Internet of Things Application. Sensors. 2022; 22: 7886.
18. Huyut MT, Velichko A. LogNNet model as a fast, simple and economical AI instrument in the diagnosis and prognosis of COVID-19. MethodsX. 2023;10:102194.
19. Huyut MT. Automatic Detection of Severely and Mildly Infected COVID-19 Patients with Supervised Machine Learning Models. IRBM. 2023;44(1):100725.
20. Huyut MT, Huyut Z. Effect of ferritin, INR, and D-dimer immunological parameters levels as predictors of COVID-19 mortality: A strong prediction with the decision trees. Heliyon. 2023;9(1):e14015.
21. Velichko A. Neural network for low-memory IoT devices and MNIST image recognition using kernels based on logistic map. Electronics. 2020;9:1432.
22. Velichko A. A method for medical data analysis using the lognnet for clinical decision support systems and edge computing in healthcare. Sensors. 2021;21:6209.
23. Velichko A, Heidari H. A Method for Estimating the Entropy of Time Series Using

Artificial Neural Networks. Entropy. 2021;23:1432.

24. Izotov Y, Velichko A, Boriskov PP. Method for fast classification of MNIST digits on Arduino UNO board using LogNNet and linear congruential generator. J Phys Conf Ser. 2021;2094:32055.
25. Matthews Correlation Coefficient | Encyclopedia MDPI Available online: https://encyclopedia.pub/entry/35211 (accessed on 15 October 2023).
26. Chicco D, Jurman G. The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. BMC Genomics. 2020;21(6):2-13.
27. Barco S, Mahmoudpour SH, Planquette B, et al. Prognostic value of right ventricular dysfunction orelevated cardiac biomarkers in patients with low-risk pulmonary embolism: a systematic review and meta-analysis. Eur Heart J. 2019;40(11):902-910.
28. Kobayashi S, Muto M, Yabe H, et al. A retrospectiveobservational study investigating the factorsassociated with right heart failure in patients withprimary acute pulmonary embolism and deep veinthrombosis. J Gen Fam Med. 2020;21(3):63-70.
29. Becattini C, Maraziti G, Vinson DR, et al. Right ventricle assessment in patients with pulmonary embolism at low risk for death based on clinical models: an individual patient data meta-analysis. Eur Heart J. 2021;42:3190–3199.
30. Kosmala W, Przewlocka-Kosmala M, Mazurek W. Subclinical right ventricular dysfunction in diabetes mellitus—an ultrasonic strain/strain rate study. Diabetic Medicine. 2007;24(1):656–663.
31. Martin KA, Molsberry R, Cuttica MJ, et al. Time trends in pulmonary embolism mortality rates in the United States, 1999 to 2018. J Am Heart Assoc. 2020;9:e016784.
32. Barco S, Valerio L, Ageno W, et al. Age–sex specific pulmonary embolism-related mortality in the USA and Canada, 2000–18: an analysis of the WHO Mortality Database and of the CDC Multiple Cause of Death database. Lancet Respir Med. 2021;9:33–42.
33. Sanchez O, Trinquart L, Colombet I, et al. Prognostic value of right ventricular dysfunction in patients with haemodynamically stable pulmonary embolism: a systematic review. Eur Heart J. 2008;29:1569–1577.
34. Coutance G, Cauderlier E, Ehtisham J, et al. The prognostic value of markers of right ventricular dysfunction in pulmonary embolism: a meta-analysis. Crit Care. 2011;15:R103.
35. Rennebaum S, Schneider SW, Henzler T, et al. Incidence of pulmonary embolism and impact on mortality in patients with malignant melanoma. Clinical Imaging. 2022;83:72–76.
36. Di-Minno MND, Ambrosino P, Ambrosini F, et al. Prevalence of deep vein thrombosis and pulmonary embolism in patients with superficial vein thrombosis: a systematic review and meta-analysis. J Tromb Haemost. 2016;14(5):964-72.
37. Kurnicka K, Lichodziejewska B, Goliszek S, et. al. Echocardiographic pattern of acute pulmonaryembolism: analysis of 511 consecutive patients. J Am Soc Echocardiogr. 2016;29(9):907-13.