PE-Predictor

STEM High School for Boys – 6th of October

Grade 12 | 2023-2024

Mohammed Ragab Saad

Zyad Ahmed Mahmoud

Ziad Ahmed Saad

Keywords:

Venous Thromboembolism (VTE)

Deep Vein Thrombosis (DVT)

Pulmonary Embolism (PE)

Wi-Fi - Random Forest - Chronic Diseases

Table of Contents

Abstract

According to The Egyptian Society of Thrombosis and Hemostasis, around 0.2% of the Egyptian society has Venous Thromboembolism disease (VTE), a medical condition that occurs when a blood clot forms in a deep vein. These clots usually develop in the lower leg, thigh, or pelvis, but they can also occur in the arm. The specific problem lies in the early prediction and prevention of the complications of Deep Vein Thrombosis (DVT) and Pulmonary Embolism (PE). The proposed approach aims to facilitate early prediction of PE complications through the implementation of an AI model. This model employs a Random Forest algorithm that considers modifiable risk factors, such as age, gender, blood oxygen concentration, temperature, and heart rate. After being used, the project achieved all the design requirements needed from it. The model achieved an accuracy of about 92.3% in predicting the complications, satisfying the determined design requirements. By integrating wearable technology and artificial intelligence, the project presents a promising advancement in the early detection of VTE, potentially mitigating the risks associated with DVT and PE.

Introduction

Public health is one of the most prominent issues around the world, including Egypt, which particularly faces challenges due to the weak technological and industrial bases. Chronic diseases are especially developing as a grueling public health challenge in Egypt. In 2016 alone, chronic diseases accounted for about 84% of the total number of deaths (El-Saadani et al., 2021b).

Venous Thromboembolism is one of the chronic diseases that proved to be fatal as well as abundant in Egypt. A study that was published in the Journal of Vascular Surgery indicates that 11,000 Egyptians are estimated to get VTE annually. **Figure (1)** shows the estimated annual incidence of VTE in Egypt by age group. Information and communication technology (ICT) has already revolutionized public health and presents room for more promising advancements. Therefore, a semester long study was conducted to explore the potential of ICT in facilitating the lives of VTE patients.

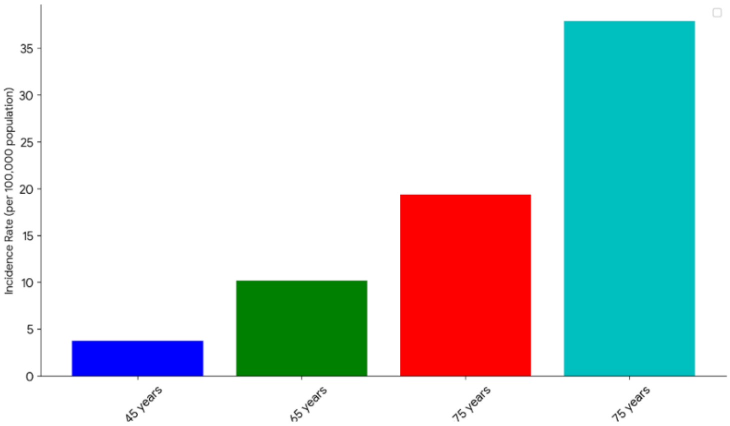


Figure (1) Annual incidence of VTE in Egypt by age group

During the background research, various prior attempts were found. Two of which were the most prominent. First is the artificial neural network with particle swarm optimization, which is a project with the goal of predicting five prevalent chronic diseases: breast cancer, diabetes, heart attack, hepatitis, and kidney disease. The method proved to be very accurate, achieving an accuracy of 98.67%. However, patients found it challenging to use this AI-based project effectively because of its complexity and lack of user-friendliness, which decreased its impact. Another attempt is DeepMind's Ai for Diabetic retinopathy. The project could be implemented in primary care settings, expanding access to DR detection for patients who might not otherwise have access to specialist care. However, integrating AI screening systems into existing healthcare workflows can be challenging, which limited the applicability of the project.

The selected solution is an Ai model with the purpose of predicting the risk of complications of Pulmonary embolism patients. The required data is gathered using an apparatus made up of a pulse oximeter and temperature sensor. The model is accessed through a custom-created mobile application, where the patient can track their health and be warned in the case of a complication. In the process of selecting the solution, two main design requirements were specified. First of all, the Ai model should have an accuracy not less than 80%. Second, it should achieve specificity and sensitivity not less than 80%.

Materials and methods

Materials

Material	Quantity	Usage	Image
Oxi-meter Sensor MAX300100	1 piece	Used to measure the oxygen concentration in the blood stream and heartbeats	
Body Temperature Sensor	1 piece	Used to measure the body temperature	

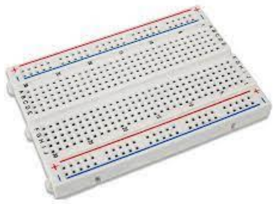
Jumpers	30 pieces	Used in the connection of bread board and the sensors	
Bread Board	2 pieces	Used as relay for the jumpers	
Firebase		Used in storing the patients' data to connect the application with the ESP32	Firebase App Distribution

Table (1) Materials used in the prototype construction

Methods

1- First, prior research papers were reviewed, the parameters that can be used for early prediction of VTE's complications were specified. The chosen parameters are heart rate, oxygen saturation in blood, temperature, and age.

2- A real-time Firebase database was created to store the data updates from the sensors, and the demographic data from the patients.

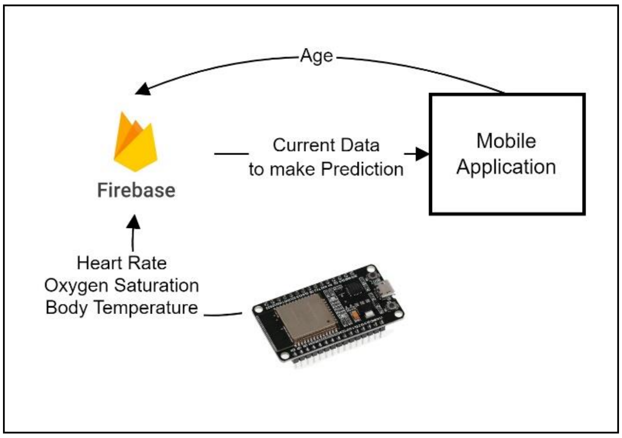


Figure (2) Communication flow chart

3- **Figure (2)** shows the flowchart of the communication between the application, Firebase, and ESP32.

4- The design of the circuit **figure (3)** was created by combining the connections of the MAX30100 heart rate and oximeter sensor and DS18B20 temperature sensor.

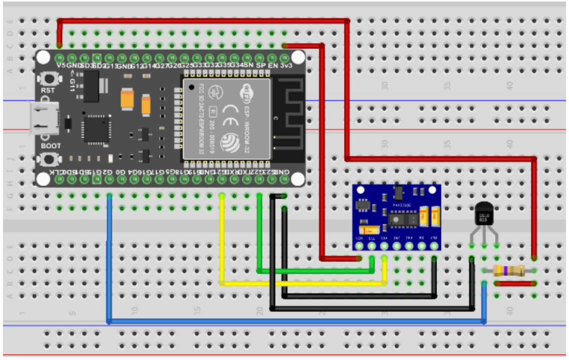


Figure (3) The circuit diagram

5- A dataset of 4000 cases containing the specified parameters as well as whether the corresponding cases developed complications of Pulmonary embolism was used.
6- KNN, Random Forest, ANN, and SVM machine learning models were implemented using Python to compare the performance of each model and use the model with highest accuracy.
7- Each model was trained on 80% of the data and testing them with the remaining 20%.
8- The model with highest accuracy was converted into TensorFlow Lite, which supports the implementation of mobile applications.
9- A mobile application was created using Flutter on Arduino Studio, allowing the patients to:
   a. Input their age and gender,
   b. View the output of the sensors,
   c. Get the results of the Random Forest model.

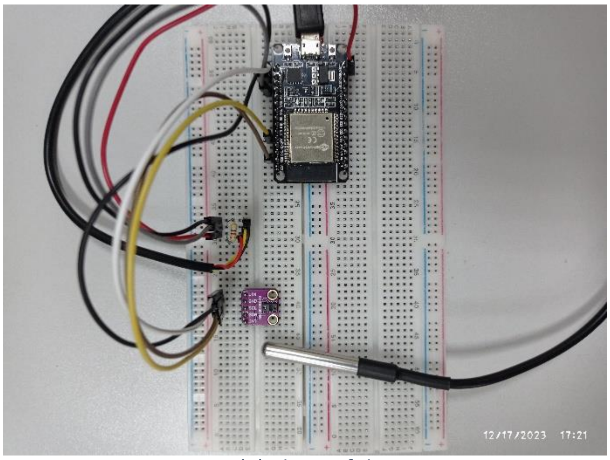


Figure (4) Photo of the circuit

Test plan

The Ai model was tested with the remaining 20% of the cases for the accuracy, sensitivity, and specificity of the model.

The circuit **figure (4)** and the application were tested to make sure of the responsiveness of the sensors, as follows:

a. The ESP32 was connected to a power source.
b. The axillary temperature was measured by the sensor.
c. The user's finger was placed on the MAX30100 sensor to measure the heart rate and oxygen saturation.
d. The changes in the parameters along with the predictions were viewed on the application.

Results

Negative results

Prior to the Random Forest, other machine learning models were used. The models used include ANN, KNN, and SVM. However, those models achieved accuracies less than the specified design requirements, and thus, the Random Forest model was chosen as it achieved the required efficiency.

Positive results

The project satisfied all the specified requirements. First of all, the Ai model achieved an accuracy of 92.25%, correctly predicting 738 out of 800 cases. The confusion matrix for the Ai model is shown in **figure (5)**. Second, the project achieved a specificity of 92.42% and a sensitivity of 92.06%. **Figure (6)** shows the performance graph of the model, comparing the accuracy, specificity, and sensitivity.

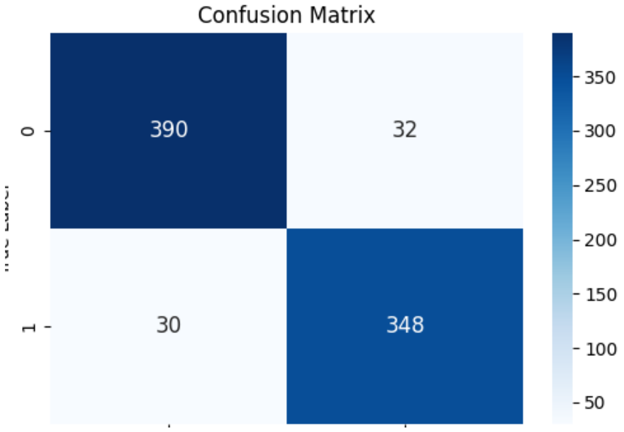


Figure (5) confusion matrix

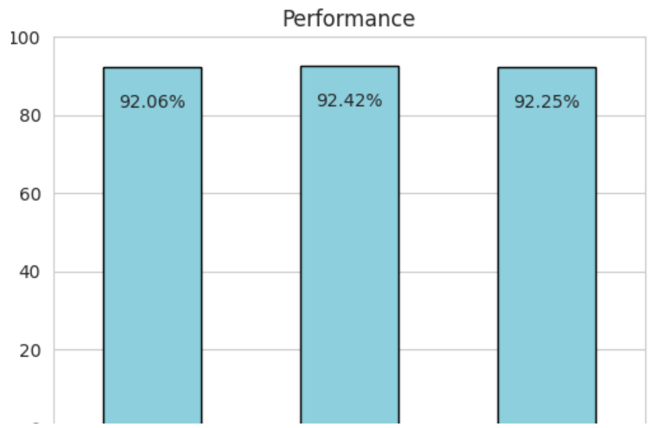


Figure (6) Performance graph

Analysis

Venous clot formation in pulmonary embolism

Pulmonary Embolism (PE) describes a blood clot that blocks and stops blood flow to an artery in the lung. In most cases, the blood clot starts in a deep vein in the leg and travels to the lung. (Lavorini, 2013) Pulmonary Embolism can lead to myocardial infarction, stroke, and other respiratory risk factors that may lead to death. (BI.3.01) A blood clot contains a mixture of platelets and fibrin and in some cases red blood cells. Venous clots form under low shear stress on the surface of a largely intact endothelium. They are fibrin-rich (so called “red clots” because they also contain red blood cells) and are treated with anticoagulant drugs.

The process begins with vasoconstriction, narrowing blood vessels to minimize blood loss. In response, platelets adhere to the exposed collagen at the injury site (BI.3.05). Adhered platelets release chemicals that make nearby platelets sticky. This initiates the platelet release reaction shown in **figure (7)**, attracting more platelets to the injury site.

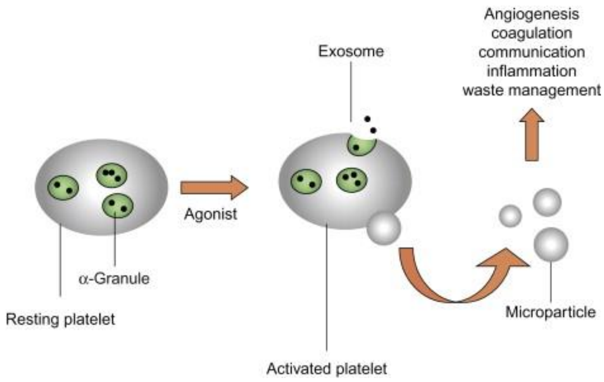


Figure (7) illustrates the Thrombocyte Release Reaction

The intrinsic and extrinsic pathways (Sanjeev Palta, 2014) converge at Factor X, triggering the coagulation cascade. This cascade involves a series of enzymatic reactions that ultimately convert prothrombin to thrombin.

Thrombin plays a central role by converting fibrinogen into fibrin threads. Fibrin threads form a mesh that traps blood cells shown in **figure (8)**, creating a stable blood clot.

The clot contracts (BI.3.03), pulling torn areas together in a process known as clot retraction, forming the venous clots.

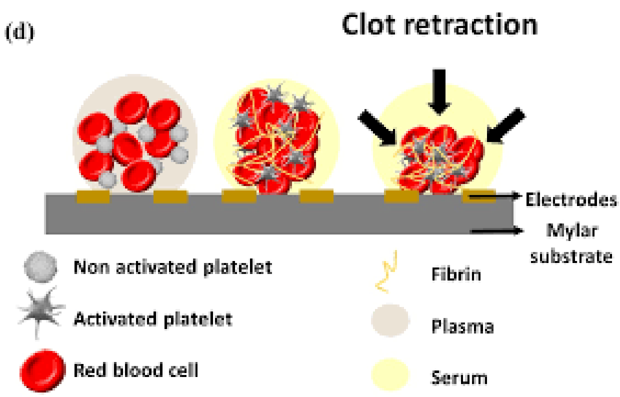


Figure (8) illustrates the clot retraction.

Thus, after the formation of venous clots along the whole body, the symptoms, swelling Pain, tenderness, redness of the skin, difficulty breathing, faster than normal or irregular heartbeat, chest pain or discomfort, which usually worsens with a deep breath or coughing, coughing up blood, very low blood pressure, lightheadedness, and fainting start to appear. Afterward, the clot starts to move until it reaches the lungs, forming Pulmonary Embolism (PE).

System overview

As shown in the flow chart of the system in **figure (2)**, there are two separate types of data inputted through the Ai model: health parameters – including heartrate, oxygen concentration, and temperature – and demographics – including age and gender. The health parameters are measured using the corresponding sensors, while demographics are entered through the mobile application. The data is then stored in a fire base until requested by the mobile application. In the application, patients can track their health parameters and check for the possibility of complications using the prediction feature. When a prediction is made, the data is requested by the application and is transmitted from the fire base. It is then analyzed by the Ai model implemented in the application to predict the risk for complications.

Circuit components and communication

Max30100 oximeter

The MAX30100 is made up of two high-intensity LEDs (RED and IR, both of which have different wavelengths) and a photodetector. These LEDs have wavelengths of 660 nm and 880 nm, respectively. The sensor uses a technique called a photoplethysmogram, which measures the amount of reflected light using a photodetector after shining both lights onto the finger or wherever the skin isn't too thick, and both lights can easily penetrate the tissue.

The higher the hemoglobin content of the blood, the redder the blood and the more IR light is absorbed. The amount of reflected light varies as the blood is pumped through the finger with each heartbeat, causing a changing waveform at the photodetector's output and displaying the heartrate reading. Similarly, blood that has been deoxygenated absorbs more red light and more infrared light. The oxygen concentration in the blood is determined by calculating the photodetector's ratio of IR to RED light.

Ds18b20

A silicon band gap sensor, such as the ds18b20, is reliant on the proportional relationship between temperature and bandgap voltage. The sensor is composed of a basic silicon diode. In a diode, the bandgap voltage at different current densities is proportional to the diode's temperature. As a result, a distinct signal is produced when the bandgap voltage varies due to changes in the diode's temperature.

ESP 32

As studied in (Ph.03.04), certain components can be used to transfer data from a sender to a receiver. In this case, the circuit containing data about the heartrate, blood oxygen concentration, and temperature is the sender, while the application, containing the AI model, is the receiver. To achieve such a task, the ESP32 development board was used as the microcontroller in the circuit. The ESP can accomplish the regular tasks of a microcontroller, but it also contains a Wi-Fi module, which is crucial for communication.

Through the use of a Wi-Fi module, devices can establish wireless network connections, access the internet, share files, and interact with other devices in the Wi-Fi network region. The connection of the wi-fi module to the mobile application is facilitated by a device known as the Access point, a device that links a wireless network to a wired network. It serves as a hub for the wireless network, enabling connections to the internet and inter-device communication for gadgets like laptops, cellphones, and tablets (Harahap et al., 2023). **Figure (9)** displays the function of the access point.

Figure (9) the connection of devices to the internet through the AP

Wi-fi can readily transmit data through radio waves. The data is stored in the microcontroller's memory in the form of a binary signal. The wi-fi module converts the binary signal to a radio wave. Through the AP, the waves are transmitted from the module to wi-fi chips in laptops or smartphones, where they are stored in the fire base along with the demographic data. For a prediction to be made, the data is requested from the fire base, and it is transmitted through wi-fi to the mobile application, where they are analyzed by the Ai model to make the prediction.

Random Forest model

Random Forest is a classification machine learning model. The main purpose of classification models is to divide the data into two or more groups (classes). Used by Random Forest, ensemble classification is a type of classification that utilizes numerous classifiers to improve the performance of the model. After each classifier outputs its own results, the final result of a new record is made by getting

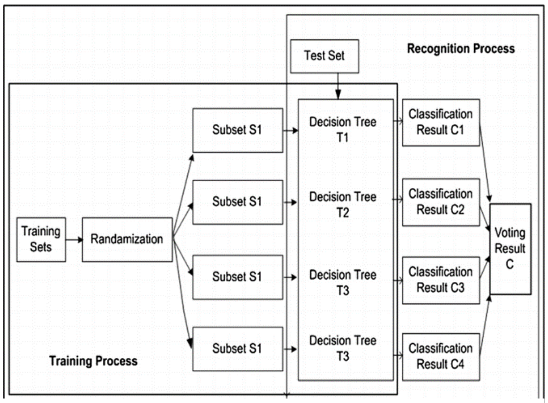


Figure (10) Diagram of Random Forest classifier the most common result of all single classifiers: a simple voting process. **Figure (10)** shows a diagram of the process of classification (Parmar et al., 2018). Each single classifier in the Random Forest is a Decision Tree. Decision Trees operate like flow charts, reaching a decision, classification result, after answering several binary tests (Kotsiantis, 2011).

The Random Forest achieved the highest accuracy after compared with other machine learning models in paper by Wang it (Wang et al., 2019). In the training process, the decision trees are constructed using 80% of the data, while the remaining 20% is used in the recognition process to test the model performance. For the evaluation of the Random Forest model, 20% of the records were reserved to test the model. The model predicted if each record patient will have complications. Those predictions, then, were compared to the real values to implement a confusion matrix (Ting, 2017). From the confusion matrix, the accuracy, sensitivity, and specificity can be calculated using equations shown in **figure (11)**.

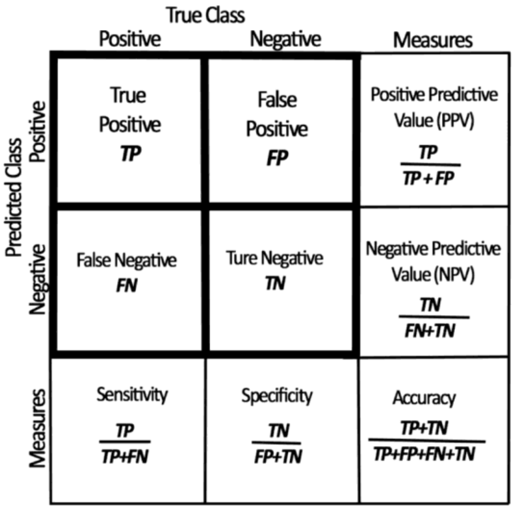


Figure (11) Confusion matrix, evaluation metrics equations

Conclusion

After a thorough study of the grand challenge of eradicating public health issues in Egypt, it was concluded that ICT can be used to facilitate the lives of venous thromboembolism patients. An Ai model was programmed with the intent of predicting complications of pulmonary embolism in patients. After running tests, it achieved the chosen design requirements for accuracy, sensitivity, and specificity, surpassing the 80% threshold, and proved its ability in precisely predicting complications of the disease. Comparing it with prior solutions, the project proved to pose a more efficient and sustainable solution. Therefore, it can be implemented in real life to combat the challenges in the public health sector as well as the industrial and technological one.

Recommendations

Educational Initiatives

Implementing and developing educational programs to increase awareness of people about DVT symptoms, risk factors, and the importance of early diagnosis.

Also, creating informative materials, such as brochures, posters, and online resources, that includes how to use the project are crucial to disseminate information to the public. Moreover, launching targeted public awareness campaigns through various channels, including social media and community events to reach a wide audience.

Using TR250Z 25% Oxygen sensor

The TR250Z **figure (12)** is a control board that provides multiple analog linear outputs of oxygen concentration. It is designed to be rugged, maintenance-free and can be calibrated in any known oxygen concentration. The Fujikura 95% oxygen sensor used with the TR250Z is a zirconium based, current limiting sensor. This type of sensor has many advantages over other ways of measuring oxygen. For example, it does not require periodic calibration and will last at least 3 or more years under normal conditions. Also, unlike common sensors, which have a maximum accuracy of 80%, this sensor has a higher accuracy of 98.67%.

Figure (12) the TR250Z 25% sensor

Doppler Ultrasonic Velocity Sensor

The Doppler ultrasound sensor, shown in **figure ((13),** emits high-frequency sound waves into the body, which bounce off moving blood cells. By analyzing the frequency shift in the echoed waves, the sensor detects blood flow velocity and direction. This non-invasive technique helps identify abnormalities, including potential blood clots, in deep veins like those affected by Pulmonary embolism (PE).


Figure (13) Doppler Sensor

Real-life application

The project's practical application entails developing a wearable or portable medical device for continuous monitoring of people who are susceptible to Pulmonary Embolism (PE), such as patients recovering from surgery or people with clotting disorders. With its temperature and oxygen concentration sensors, the device allows for continuous, non-invasive blood flow monitoring and provides real-time data. The application receives timely alerts from the system in the event that a PE develops. The goal of this early detection and treatment strategy is to improve patient outcomes and care quality by reducing the likelihood of serious complications.

Literature Cited

1. El-Saadani, S., Saleh, M., & Ibrahim, S. A. (2021). Quantifying non-communicable diseases' burden in Egypt using State-Space model. PLOS ONE, 16(8), e0245642. https://doi.org/10.1371/journal.pone.0245642

2. Harahap, B. S. D., Silalahi, J., Saragih, A., & Singarimbun, E. R. (2023). Wireless networking How Wi-Fi works and the different types of wireless networks. ResearchGate. https://www.researchgate.net/publication/367128043_Wireless_Networking_How_Wi-Fi_Works_and_the_Different_Types_of_Wireless_Networks

3. Kotsiantis, S. B. (2011). Decision trees: A recent overview. Artificial Intelligence Review, 39(4), 261–283. https://doi.org/10.1007/s10462-011-9272-4

4. Lavorini, F., Di Bello, V., De Rimini, M. L., Lucignani, G., Marconi, L., Palareti, G., Pesavento, R., Prisco, D., Santini, M., Sverzellati, N., Palla, A., & Pistolesi, M. (2013). Diagnosis and treatment of pulmonary embolism: a multidisciplinary approach. Multidisciplinary Respiratory Medicine, 8(1), 75. https://doi.org/10.1186/2049-6958-8-75

5. Mackman, N. (2012). New insights into the mechanisms of venous thrombosis. Journal of Clinical Investigation, 122(7), 2331–2336. https://doi.org/10.1172/jci60229

6. Palta, S., Saroa, R., & Palta, A. (2014). Overview of the coagulation system. Indian Journal of Anaesthesia, 58(5), 515. https://doi.org/10.4103/0019-5049.144643

7. Parmar, A., Katariya, R., & Patel, V. (2018). A review on Random Forest: An ensemble classifier. International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018, 758–763. https://doi.org/10.1007/978-3-030-03146-6_86

8. Ting, K. M. (2017). Confusion matrix. Encyclopedia of Machine Learning and Data Mining, 260–260. https://doi.org/10.1007/978-1-4899-7687-1_50

9. Wang, X., Yang, Y., Liu, S., Hong, X., Sun, X., & Shi, J. (2019). Comparing different venous thromboembolism risk assessment machine learning models in Chinese patients. Journal of Evaluation in Clinical Practice, 26(1), 26–34. https://doi.org/10.1111/jep.13324