COVID-19 PREDICTION SYSTEM USING RANDOM FOREST

Presented by:
NURUL SYAZWINA SYAHIRAH BINTI MOHD AZIZ
(2023168543)

Supervised by:
DR. SITI KHATIJAH NOR BINTI ABDUL RAHIM

01. INTRODUCTION
BACKGROUND OF THE STUDY

COVID-19 is an example of how global changes have increased the likelihood of an outbreak.

In Malaysia, many underestimate COVID-19 symptoms such as fever, cough, fatigue making early detection crucial. Immunization has reduced the spread, but high-risk groups still face mortality, emphasizing the need for early diagnosis and treatment.

A predictive modeling application can help users detect COVID-19 early by analyzing symptoms, enabling timely precautions and clinical tests.

PROBLEM STATEMENT

Challenges in COVID-19 Detection

The challenge lies in the current methods' inability to identify subtle infection patterns, leading to high false negatives and poor COVID-19 detection rates. (Khanna et al., 2023)

Need for Accurate Machine Learning in COVID-19 Prediction

There is a need for a system that uses more accurate machine learning techniques to predict COVID-19 more reliably. (Nayak et al., 2021)

Urgent Need for Effective COVID-19 Detection Tools

The high infection rate and emergence of more dangerous COVID-19 variants underscore the urgent need for effective tools to detect the virus in the human body. (Vianello et al., 2021)

OBJECTIVES OF THE STUDY

1

To identify a suitable machine learning approach in predicting COVID-19

2

To design and develop a COVID-19 prediction system that can be used to analyse COVID-19 based on patient's symptoms using the identified approach

3

To evaluate the performance and accuracy of the COVID-19 prediction system using the COVID-19 dataset

SCOPE OF THE STUDY

Focus for employee, students and elderly

5,861,480 COVID-19 patient symptom records from Kaggle

Use Random Forest algorithm

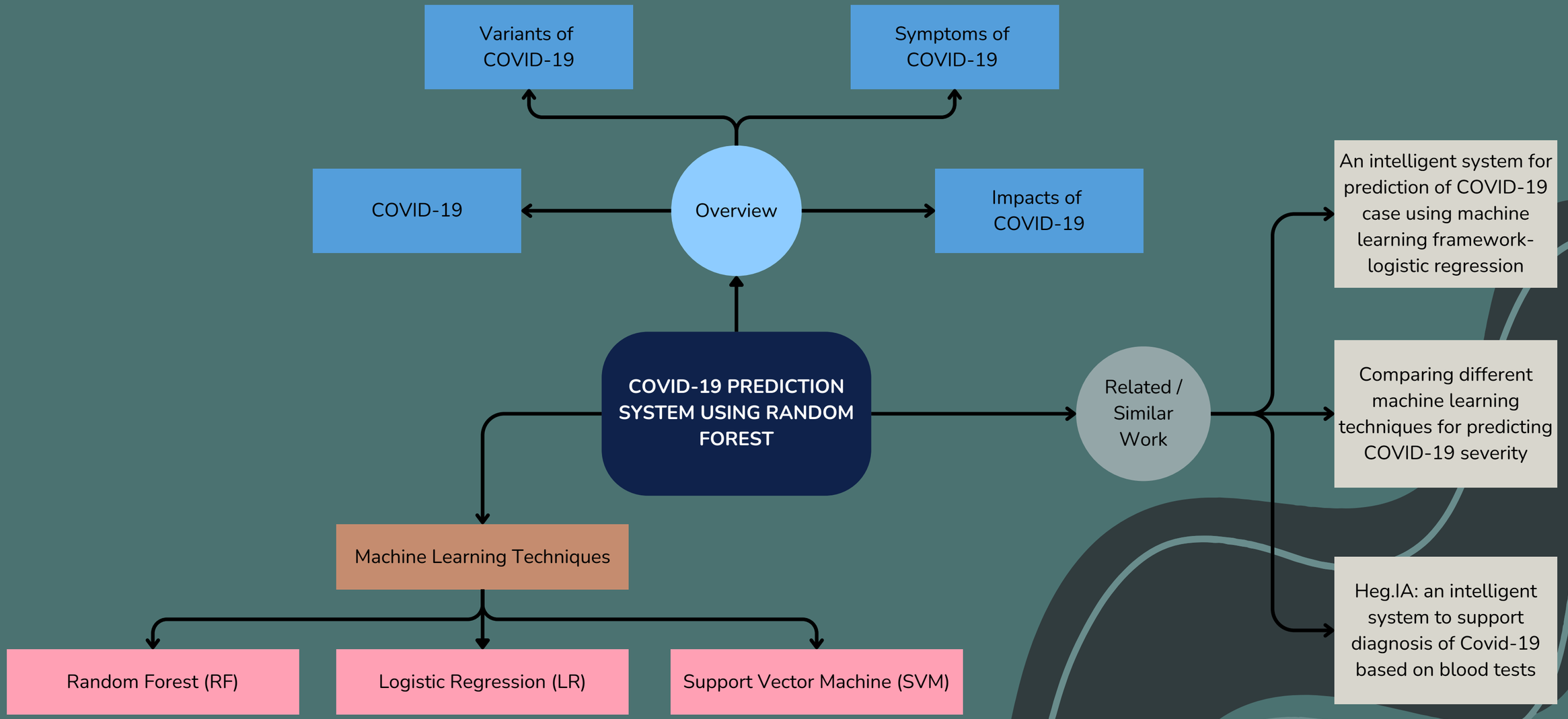
SIGNIFICANCE OF THE STUDY

Helps users take precautionary measures and reduce the risk of COVID-19

Allows users to identify and check their symptoms based on COVID-19 patterns

Provides a foundation for future development and improvements in disease detection

LITERATURE REVIEW



SIMILAR WORK ON COVID-19 PREDICTION

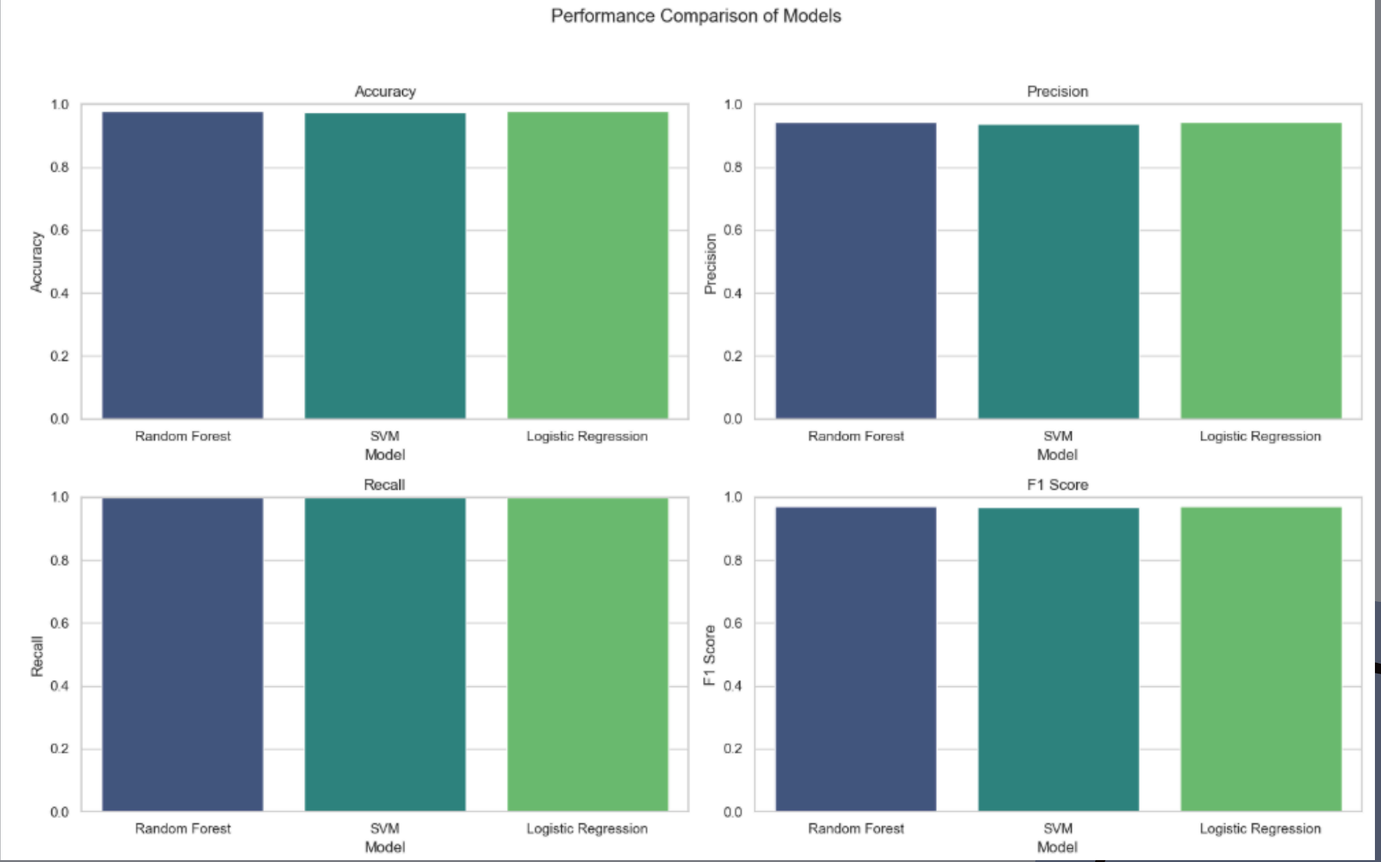
Reference	Classification Approach	Datasets	Quantity of Data	Quantity of Class	Machine Learning Approach	Result (Accuracy %/ AUC)
(Majumder et al., (2021)	Supervised	Mexican government (Kaggle)	563201	23	LR	92
(Xiong et al., (2022)	Supervised	Hospital information system	287	5	RF, SVM, LR	0.970 0.948 0.928
(de Freitas Barbosa et al., (2022)	Supervised	Hospital Israelita Albert Einstein in São Paulo, Brazil	5644	111	MLP, SVM, RF, Bayesian Network, Naive Bayes	95.159% ± 0.693

PROJECT FRAMEWORK

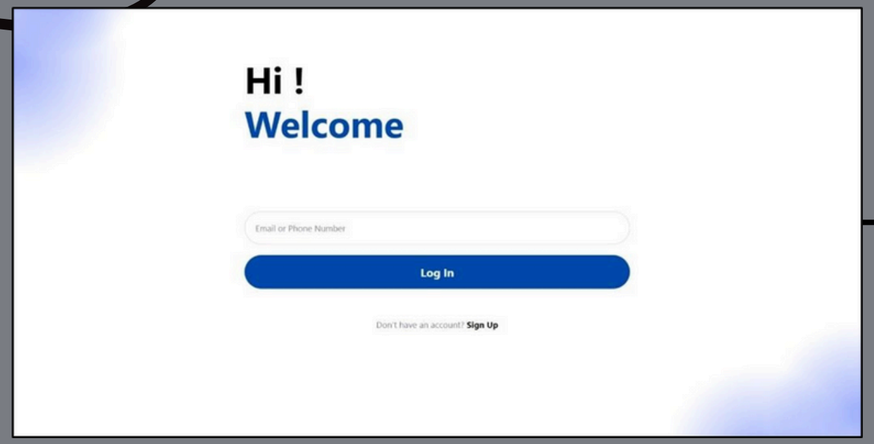
Objective	Phase	Activity	Deliverables / Outcome
Objective 1: To identify a suitable machine learning approach in predicting COVID-19.	Requirement Analysis	1. Do a preliminary study on: • Infectious disease • COVID-19 disease • Machine learning techniques • Random Forest • Similar studies	1. Problem statements 2. Objectives 3. Scope of study 4. Significance of study 5. Comparison table for the machine learning techniques 6. Comparison table of machine learning in similar studies 7. Comparison table of similar application
		2. Data Collection • Search for a suitable COVID-19 dataset that can be used to train and test the application	One set of COVID-19 symptoms datasets from Kaggle have been obtained
		3. Data Preparation • Perform data preparation on the collected dataset to make it ready for analysis or model training.	1. Load and Read Dataset 2. Data Pre-processing 3. Data Splitting 4. Comparison of Different Algorithms 5. Build and Train Random Forest Model
Objective 2: To design and develop a COVID-19 prediction systemthat can be used to analyse COVID-19 based on patient's symptoms using the identified approach.	System Design	1. Design System Architecture 2. Design Use Case diagram 3. Design Flowchart Diagram 4. Design User Interface design 5. Design Algorithm	1. System Architecture 2. Use Case Diagram 3. Flowchart Diagram 4. User Interface Design 5. Algorithm Design
	Implementation	1. Develop frontend 2. Develop backend 3. Determine hardware and software requirements	Functional COVID-19 Prediction System Application
Objective 3: To evaluate the performance and accuracy of the COVID-19 prediction systemusing the COVID-19 dataset.	Testing and Evaluation	1. Test the accuracy of the application 2. Create and evaluate test cases	1. Completed system 2. Test Cases 3. Accuracy result 4. User Evaluation report

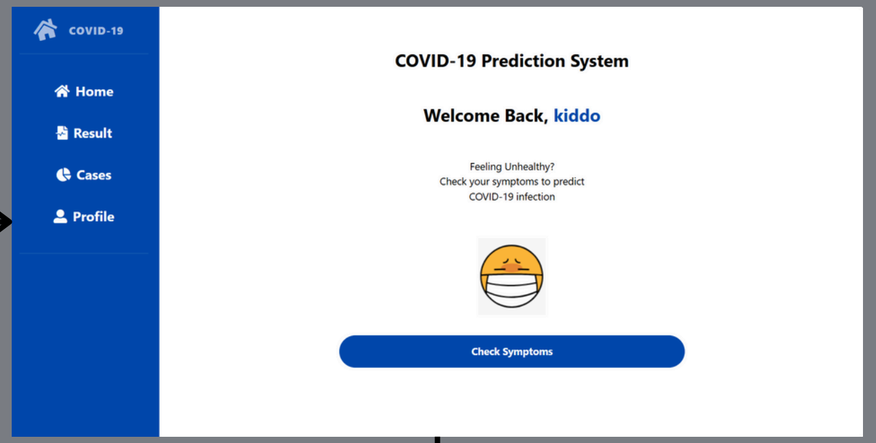
RESULTS

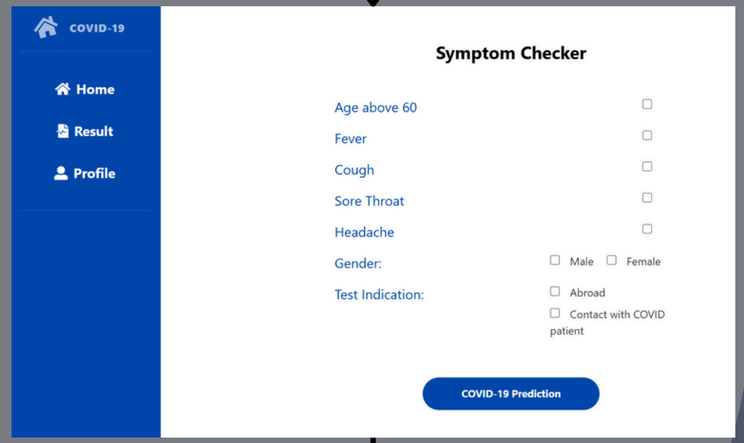
```
Performance Comparison:
                 Model  Accuracy  Precision    Recall  F1 Score
0        Random Forest  0.977272   0.943307  0.999662  0.970667
1                  SVM  0.974587   0.937275  0.999323  0.967305
2  Logistic Regression  0.976994   0.943500  0.998646  0.970290
Random Forest Cross-Validation Accuracy Scores: [0.97531658 0.97405361 0.97491467 0.97195142 0.97095488]
Random Forest Mean Cross-Validation Accuracy: 0.9734382290482907
```

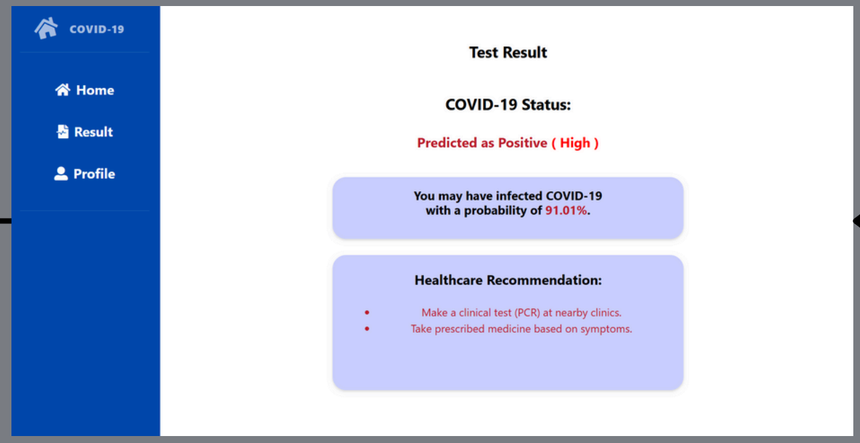


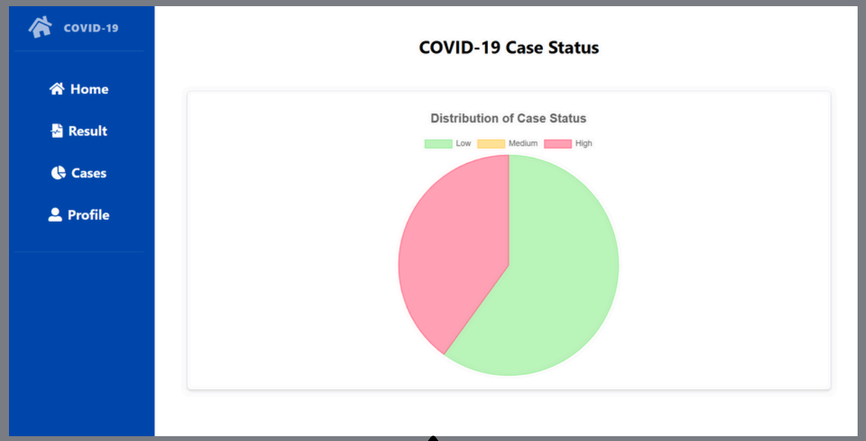
RESULTS

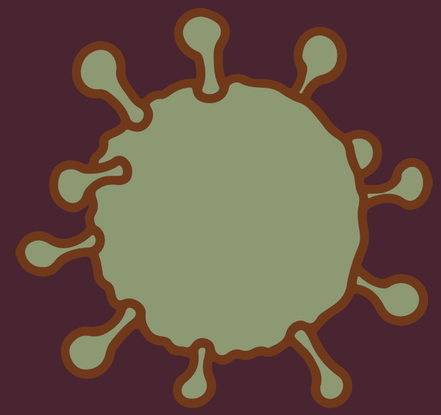

Hi !
Welcome
Email or Phone Number
Log In
Don't have an account? Sign Up


COVID-19
Home
Result
Cases
Profile
COVID-19 Prediction System
Welcome Back, kiddo
Feeling Unhealthy?
Check your symptoms to predict
COVID-19 infection
Check Symptoms


COVID-19
Home
Result
Profile
Symptom Checker
Age above 60
Fever
Cough
Sore Throat
Headache
Gender:
Male
Female
Test Indication:
Abroad
Contact with COVID patient
COVID-19 Prediction


COVID-19
Home
Result
Profile
Test Result
COVID-19 Status:
Predicted as Positive (High)
You may have infected COVID-19
with a probability of 91.01%.
Healthcare Recommendation:
Make a clinical test (PCR) at nearby clinics.
Take prescribed medicine based on symptoms.


COVID-19
Home
Result
Cases
Profile
COVID-19 Case Status
Distribution of Case Status
Low
Medium
High



SYSTEM DEMO

CONCLUSION

Contribution

- Implementation of random forest algorithm in predicting COVID-19 infection.
- Functionality to grade predictions into three levels of risks that are low, medium, and high.
- The system includes a pie chart to visually display risk categories, making test results easy to understand.

Limitation

- The dataset used to train the model for prediction included only a few different symptoms.
- The lack of a current symptom dataset.
- The dataset lacked representation from other countries, particularly Malaysia.

Future Works

- Focused on growing the dataset by working with healthcare institutions all over the world.
- Include more symptoms into the dataset.
- Wearable technology or smartphone apps use machine learning for real-time COVID-19 prediction.


THANK YOU!