Deep Learning Hybrid Models for COVID-19 Prediction

Ziyue Yu, Faculty of Applied Sciences, Macao Polytechnic University, China
Lihua He, Faculty of Applied Sciences, Macao Polytechnic University, China
Wuman Luo, Faculty of Applied Sciences, Macao Polytechnic University, China
Rita Tse, Faculty of Applied Sciences, Macao Polytechnic University, China
Giovanni Pau, Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

ABSTRACT

COVID-19 is a highly contagious virus. Blood test is one of effective methods for COVID-19 diagnosis. However, the issues of blood test are time-consuming and lack of medical staff. In this paper, four deep learning hybrid models are proposed to address these issues (i.e., CNN+GRU, CNN+Bi-RNN, CNN+Bi-LSTM, CNN+Bi-GRU). In addition, two best models, CNN and CNN+LSTM, from Turabieh et al. and Alakus et al., are implemented, respectively. Blood test data from Hospital Israelita Albert Einstein is used to train and test six models. The proposed best model, CNN+Bi-GRU, is accuracy of 0.9415, precision of 0.9417, recall of 0.9417, F1-score of 0.9417, AUC of 0.91, which outperforms the best models from Turabieh et al. and Alakus et al. Furthermore, the proposed model can help patients to get blood test results faster than traditional manual tests without errors caused by fatigue. The authors can envisage a wide deployment of proposed model in hospitals to alleviate the testing pressure from medical workers, especially in developing and underdeveloped countries.

KEYWORDS

Blood Test, CNN+Bi-GRU, COVID-19 Infection, Deep Learning Hybrid Models

1. INTRODUCTION

The coronavirus disease (COVID-19) broke out and spread rapidly in the world (Tai et al., 2021). This disease has not only become one of the deadliest epidemics in human history, but also causes damage to the society and economy (Van et al., 2020; Alankar et al., 2021). Furthermore, different variants of COVID-19 virus such as Delta and Omicron are more infectious than the original virus (Mallapaty et al., 2021; World Health Organization, 2021). To make matters worse, vaccines may be less effective against these new variants. And even if people are vaccinated, they are still at risk of being infected and can spread the virus to others (World Health Organization, 2021). Facts show that the use of accurate test methods and timely test results can isolate positive patients as early as possibly.
possible and reduce the risk of transmission (World Health Organization, 2021). As a result, regular routine testing is essential in the fight against COVID-19 (World Health Organization, 2021).

Typically, computed tomography (CT) scan of lungs, nucleic acid test and blood test are effective ways of COVID-19 tests (Brinati et al., 2020; Dhiman et al., 2021). CT scan is used to detect abnormalities in the lungs, but this method is not used as a primary test method because of the expensive equipment. The nucleic acid test is widely used because of its simplicity (Chang et al., 2021). However, research proves that the false-negative rate of nucleic acid test can be as high as 21.4% (Xiao et al., 2020), and the materials used for the nucleic acid test in underdeveloped countries are limited (Abdel-Basset et al., 2021). Blood test has the lower error of false positive and false negative rate than nucleic acid test, and the materials it required are more readily available (Ferrari et al., 2020). Besides, blood test can detect seasonal coronavirus patients, so it can avoid unnecessary isolation (Peeling et al., 2020). However, there are two major issues of blood test. The first is time-consuming, patients usually have to wait several days to get the test result (World Health Organization, 2021; Amanda, 2020). The second is the shortage of medical staff for COVID-19 blood test (Wynants et al., 2020). These issues are common in developing and underdeveloped countries (Giri et al., 2020).

To address these issues, it is necessary to predict the infection of COVID-19 via artificial intelligence. Currently, machine learning and deep learning models are applied in medical diagnosis (Jin et al., 2020). For example, deep multilayer perceptron was applied in diabetes risk prediction (Zhu et al., 2020). However, the research of deep learning hybrid model for COVID-19 prediction based on blood test is still in the preliminary stage. In 2020, Alakus et al. adopted four single deep learning models and two deep learning hybrid models to predict COVID-19 infection based on the blood test data from Hospital Israelita Albert Einstein (Alakus et al., 2020). Their proposed deep learning hybrid model convolutional neural network (CNN)+Long Short-Term Memory (LSTM) achieved the best accuracy of 92.3%. Turabieh et al. selected the same data resource from the work of Alakus et al. and applied CNN to predict the infection with the best accuracy of 80% (Turabieh et al., 2021). There are other applied machine learning models for predicting COVID-19 infection such as logistic regression (LR), support vector machine (SVM), decision tree (DR) and random forest (RF), and the accuracy of these machine learning models is between 80% to 88% (Jiang et al., 2020; Batista et al., 2020; Cabitza et al., 2021). However, the prediction accuracies of these machine learning models and single deep learning models are suboptimal compared with the deep learning hybrid models.

In this work, we proposed four deep learning hybrid models to predict COVID-19 infection, namely CNN+Gated Recurrent Unit (GRU), CNN+Bidirectional Recurrent Neural Network (Bi-RNN), CNN+Bidirectional Long Short-Term Memory (CNN+Bi-LSTM), CNN+Bidirectional GRU (CNN+Bi-GRU). For the sake of fairness, we use the same data resource from Alakus et al. and Turabieh et al., i.e., blood test data from Israelita Albert Einstein Hospital. We select the same five evaluation metrics from Alakus et al., namely accuracy, precision, recall, F1-score and AUC. The experiment results show that our proposed deep learning hybrid model CNN+Bi-GRU outperforms the best models from the work of Alakus et al. (CNN+LSTM) and Turabieh et al. (CNN) in terms of all five evaluation methods.

In summary, four contributions of this paper are presented as follows:

- We proposed four deep learning hybrid models include CNN+GRU, CNN+Bi-RNN, CNN+Bi-LSTM and CNN+Bi-GRU. To the best of our knowledge, we are the first to design these hybrid models for the prediction of COVID-19 infection. The data set of this work is the blood test data from Israelita Albert Einstein Hospital provided by Alakus et al.
- We implement CNN and CNN+LSTM proposed by Alakus et al. and Turabieh et al., respectively. This work also uses the data from Alakus et al. And evaluate the performance of these two models and our proposed hybrid models in terms of accuracy, precision, recall, F1-score and AUC.
- As a result, the proposed deep learning hybrid model CNN+Bi-GRU gives better performance (i.e., accuracy of 0.9415, precision of 0.9417, recall of 0.9417, F1-score of 0.9417, AUC of
0.9100) than the best models from the work of Alakus et al. and Turabieh et al. in terms of all evaluation methods.

- Compared with the traditional manual blood testing process, our best-proposed model CNN+Bi-GRU can help patients to get blood test reports faster. Meanwhile, this proposed model relieves the medical staff from the heavy testing work, especially in developing and underdeveloped countries.

The remainder of this paper is organized as follows. In Section 2, we review related works for the prediction of COVID-19 infection by machine learning and deep learning models. The methodology of our proposed models will be given in Section 3. Section 4 describes the blood test data, experiments environment, evaluation methods, result analysis and discussion. We summarize this paper in Section 5.

2. LITERATURE REVIEW

Nowadays, major test methods for COVID-19 infection include CT scan of lungs, nucleic acid test and blood test (Giri et al., 2021). The use of CT scan to detect COVID-19 infection depends on whether there are multiple mottling or ground-glass opacity in the lungs (Hu et al., 2020; He et al., 2020). However, patients are often already infected with the COVID-19 virus when CT results show that the lungs are inflamed or the tissues are diseased (Wu et al., 2020). Therefore, the method of CT scan of lungs may delay the patient’s isolation and treatment and cause the spread of the epidemic. Moreover, underdeveloped countries may not be able to purchase sufficient CT screening equipment because of expensive price (Li et al., 2021). The nucleic acid test serves as the gold standard method for COVID-19 infection (Wyllie et al., 2020). However, nucleic acid test proved to have the issues of false negatives and false positives, and the testing materials are in short supply (Esbin et al., 2020). Besides, underdeveloped countries and some developing countries lack medical staff who have received rigorous training in nucleic acid testing (Elhadi et al., 2020). The blood test has smaller false negative and false positive rates than nucleic acid test (Aktar et al., 2021). And blood test can identify the difference between seasonal coronavirus flu and COVID-19 infection, so it avoids unnecessary isolation of false positive patients (Zhang et al., 2021). Other benefits are that medical institutions can meet the equipment requirement of blood test if they have conventional medical collection and testing tools (e.g., test tube). (AlJame et al., 2020; An et al., 2020; Cabitza et al., 2021). However, blood test faces the limitation of time-consuming and lack of adequate medical staff (Woolf et al., 2021; Looi, 2020).

To address these issues, we need to apply artificial intelligence (AI) approaches for COVID-19 prediction based on blood test. The purpose is to free healthcare workers from excessive blood testing work and speed up the testing process. So far, existing AI techniques for COVID-19 prediction include machine learning and deep learning models. Major works on machine learning/deep learning prediction on COVID-19 blood test data are summarized as follows.

Jiang et al. presented five machine learning models for COVID-19 prediction in 2020, including K-Nearest-Neighbor (KNN), LR, DT, RF, and SVM (Jiang et al., 2020). They used data from Wenzhou Central Hospital and Cangnan People’s Hospital to train the models. LR is to utilize existing data to create a classification regression equation. DT is a model that is used to monitor and realize the underlying rules of data, as well as to categorize and forecast the outcomes of new data. RF is derived from the concept of DT. But it employs the ensemble learning technique, in which the RF votes on the classification results of many weak classifiers to create strong classifiers. KNN is to compute and compare the distances between a test point and points belonging to various categories within a specified interval. And the test point’s categorization is decided by its k closest neighbors. SVM is widely used for 0-1 classification problems. SVM calculates the greatest margin between two kinds of labeled data. The experiment findings indicate that the highest accuracy was achieved when SVM was used (80%).
Batista et al. used gradient boosted trees, SVM, RF, LR to predict COVID-19 infection. The data set was the blood samples from Sao Paulo’s Hospital Israelita Albert Einstein (Batista et al., 2020). Gradient boosted trees utilize DT as the basis function and build the model in the direction of the gradient drop for each iteration, and this model is robust to the outliers. Their experiment results showed that RF got the highest accuracy of 84.7%.

Cabitza et al. analyzed the data from San Raffaele Hospital between February to May 2020 (Cabitza et al., 2020). Their explored models included Naive Bayes (NB), LR, KNN, RF, SVM, and RF model achieved the highest accuracy rate of 88%.

Turabieh et al. presented four models, namely CNN, KNN, NB and DT (Turabieh et al., 2021). They used the data from Hospital Israelita Albert Einstein to train and test the models. CNN is utilized in a variety of applications, including image processing, text processing, and voice recognition (Li et al., 2021). It consists of three operations: convolutional operation, pooling operation and fully connected operation. The convolutional operation is used to extract features, so it can reduce the total number of parameters of the networks. Pooling operation accelerates computation and prevents overfitting. Fully connected operation can be regarded as classifier to classify the model’s output (Shin et al., 2016). Additionally, shared weights and biases were be used during convolution and pooling operation, which simplifies the optimization process. The best accuracy was 80% when using CNN.

Alakus et al. used four deep learning models, namely multi-layer artificial neural network (ANN), CNN, RNN and LSTM (Alakus et al., 2020). And proposed two deep learning hybrid models include CNN+RNN and CNN+LSTM. RNN is well-suited for time series data processing (Su et al., 2019). This model has a hidden state for storing historical information, and the activation function adopts “Tanh” to converge faster than “Sigmoid”. The LSTM is made up of four gates: a forgotten gate, an input gate, an output gate, and a memory cell (Jelodar et al., 2020). The forget gate is adopted to forget useless data, the input gate is used to refer to data that needs to be inserted into the memory cell, and the output gate is used to control whether the data can be inserted into the memory cell. Memory cells are utilized to remember the knowledge from the past. The best accuracy was 92.3% when CNN+LSTM was used, which is better than the accuracy of current machine learning models and single deep learning models. This study shows that combining CNN with other deep learning models can improve accuracy because CNN extracts features from the data.

Overall, these major works illustrate the importance of machine learning and deep learning models in COVID-19 prediction based on blood test. Specifically, machine learning and deep learning models can understand the potential relationship of different attributes in blood test data and provide information support for decision-making. Moreover, these models will not be fatigued, distracted, or affected by emotions. Therefore, the model will become effective when detecting large-scale COVID-19 infection, because the fatigue of medical staff is common at that time. Based on the works of Jiang et al., Batista et al., Cabitza et al., Turabieh et al. and Alakus et al., we concluded that deep learning hybrid models have higher accuracy than machine learning and single deep learning models. The higher accuracy of model prediction illustrates the stronger ability to accurately predict COVID-19 infection. Although deep learning is helpful in predicting COVID-19 infection based on blood test, if the model’s prediction accuracy is too low, it will accelerate the spread of the epidemic caused by false negative patients. Inspired by the feature acquisition capabilities of CNN, we aim to propose more CNN-based deep learning hybrid models with higher accuracy for COVID-19 prediction based on blood test.

3. METHODOLOGY

In this section, we present the relevant theory of four proposed hybrid models for COVID-19 infection prediction. We first describe the overall mechanism of proposed deep learning hybrid models. Then, we introduce the mathematics definition of each deep learning hybrid model in detail.
3.1 The Overall Mechanism of Deep Learning Hybrid Models

The proposed structure of each deep learning hybrid model is shown in Table 1. Each hybrid model includes three convolution operations (i.e., Conv1D), two pooling operations (i.e., MaxPool), one recurrent neural layer, one fully connected layer and one output layer. Specifically, the convolution operation kernel size is 256, 128 or 64, respectively. Dropout method with 0.15 is utilized to avoid overfitting after two convolution operations. In pooling operation, the kernel size is 3. The hidden state dimension of recurrent neural layer is 256, and we adopt Batch Normalization to speed up the training process. We also use dropout method with 0.15 in fully connected layer to prevent overfitting. Finally, using SoftMax function in output layer to classify the COVID-19 infection result.

Figure 1 is the flow diagram of deep learning hybrid models. First, convolution layer is applied in blood test data. Then, we use five different variants of recurrent neural layer to process the output from convolution layer, respectively, i.e., LSTM, GRU, Bi-RNN, Bi-LSTM, Bi-GRU. Finally, SoftMax function is used to predict whether to be infected with the COVID-19 virus.

As discussed in the literature review, CNN is used for feature extraction. Specifically, we apply convolution and pooling operations for blood test data. Each patient’s blood test data can be represented as \( A = \{a_1, a_2, \ldots, a_n\} \), where \( a_i \) denotes the attribute value of data \( i \) is the index. \( i = 1, 2, \ldots, 18 \) because the number of attributes is 18). First, the convolution operation for \( A \) is defined as follows:

\[
H = K \times A
\]  

where \( K \) denote kernel (learnable weight vector); \( H \) is feature representation of \( A \). Second, the pooling operation is applied for reducing the dimension of \( H \) but still maintaining the invariance of

<table>
<thead>
<tr>
<th>Layer</th>
<th>Type</th>
<th>Kernel</th>
<th>Stride</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Conv1D</td>
<td>256</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>MaxPool</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Conv1D</td>
<td>128</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>MaxPool</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>Conv1D</td>
<td>64</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>Recurrent Neural Layer</td>
<td>256</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>Fully Connected Layer</td>
<td>256</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>SoftMax</td>
<td>2</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 1. The flow diagram of deep learning hybrid models
feature. Feature representation $H$ will be divided into many area $R$ under pooling operation. In this work, we adopt max pooling method based on the following equation:

$$X = \max_{i \in R} h_i$$

(2)

where $h_i$ denote the values of area $R$. $X$ denote the feature representation matrix after using the max value of each $h_i$. Based on equation (1) and (2), the final feature representations $X$ of blood test data can be used as new input to the recurrent layer. RNN has the ability of short-term memory, so it can further capture the correlation from $X$. In the recurrent layer, we explore the performance of five different variants of RNN. The detail of each variant is introduced in 3.2 to 3.5. In the classification, the output from recurrent layer is flattened as a 1-dimensional vector $v$. Then, the fully connected layer is utilized for the flatted vector $v$:

$$y = w_y v + b_y$$

(3)

where $w_y$ denote learnable weight parameter and $b_y$ is learnable bias vector. Finally, $y$ is used in SoftMax function for predicting COVID-19 infection based on blood test:

$$S = \frac{e^{y_i}}{e^{y_1} + e^{y_2}}$$

(4)

where $S$ denote the label of COVID-19 prediction result. $i$ is index, $i = 1, 2$ because our work is binary classification.

Overall, the structure of proposed deep learning hybrid model can learn the latent feature representation in blood test data by convolution and pooling operations. And the use of recurrent layer enables the learning of correlations between feature representations.

### 3.2 Proposed Model 1: CNN+GRU

CNN+GRU is the first proposed deep learning hybrid model. GRU is an RNN variant. The benefit of GRU is to address the issue of gradient explosion or disappearance from traditional RNN (Zhang et al., 2018). Figure 2 shows that the reset gate determines whether forget the historical information. The update gate decides whether information from past step should be used to compute current hidden state.

The GRU model is defined using following equations:

$$R_t = \sigma \left( w_r \left[ h_{t-1}, x_t \right] + b_r \right)$$

(5)

$$\tilde{h}_t = \tanh \left( w_h \left[ R_t \odot h_{t-1}, x_t \right] + b_h \right)$$

(6)

$$U_t = \sigma \left( w_u \left[ h_{t-1}, x_t \right] + b_u \right)$$

(7)

$$h_t = U_t \odot h_{t-1} + (1-U_t) \odot \tilde{h}_t$$

(8)
where reset gate $R_t$, update gate $U_t$ and candidate hidden state $\tilde{h}_t$ are calculated with different learnable parameters $w_r$, $w_h$ and $w_u$. Hidden state $h_t$ is updated under the control of update gate, and $h_{t-1}$ is the previous hidden state. $X = \{x_1, x_2, \ldots, x_t\}$ as inputs, and $b_r$, $b_h$, $b_u$ are bias. Vector dot product is represented by $\odot$, and $\sigma$ is activation function “Sigmoid”. The historical information is forgotten if the value of reset gate $R_t$ close to 0, and the candidate hidden state $\tilde{h}_t$ only contains current information $x_t$. This reset mechanism can remove unnecessary information to minimize computational complexity. The current hidden state $h_t$ only use $\tilde{h}_t$ if the value of update gate $U_t$ close to 0; when its value close to 1, the current hidden state $h_t$ is equal to the previous state (i.e., $h_{t-1}$).

3.3 Proposed Model 2: CNN+Bi-RNN

CNN+Bi-RNN is the second proposed deep learning hybrid model. Figure 3 shows that Bi-RNN has forward network and backward network, and the output layer connects these two networks. The bidirectional structure enables the model to comprehend the correlation between current information, historical information and future information.

Bi-RNN model is formulated as follow.

$$\tilde{h}_t = \sigma \left( U x_t + w \tilde{h}_{t-1} + b \right)$$  \hspace{1cm} (9)

$$\tilde{h}_t = \sigma \left( U' x_t + w' \tilde{h}_{t+1} + b' \right)$$  \hspace{1cm} (10)

$$y_t = (\tilde{h}_t \oplus \tilde{h}_t)$$  \hspace{1cm} (11)
where $\vec{h}_t$ is the hidden state of forward layer, and $\overrightarrow{h}_t$ is the hidden state of backward layer. $\vec{h}_{t-1}$ denote the previous hidden state in forward layer, and $\overrightarrow{h}_{t+1}$ denote the next hidden state in forward layer. In the forward layer, $U$ represents the learnable weight for $x_t$. $w$ denote the learnable weight for $\vec{h}_{t-1}$, and $b$ is the bias. $U'$, $w'$ and $b'$ are used for backward layer. $\oplus$ represents the vector stitching, and the output $y_t$ is based on the result from $\vec{h}_t$ and $\overrightarrow{h}_t$.

### 3.4 Proposed Model 3: CNN+Bi-LSTM

The third deep learning hybrid model is CNN+Bi-LSTM. Bi-LSTM is based on LSTM, the difference is that Bi-LSTM add forward network and backward network to understand the previous, current and future information from hidden state. We only need to introduce the calculation method of LSTM because calculation process of forward and backward network is the same. Figure 4 shows that LSTM includes the forget gate, input gate, output gate and the memory cell. Forget gate controls whether information from the long-time memory unit can forget the historical information from the last moment. Input gate controls whether the new inputs can be inserted into long-time memory unit. Output gate controls whether the information from long-time memory unit can be outputted. Cell is used for storing the information.

The LSTM equations are defined as follows:

\[
F_t = \sigma \left( w_f \left[ h_{t-1}, x_t \right] + b_f \right) \\
I_t = \sigma \left( w_i \left[ h_{t-1}, x_t \right] + b_i \right) \\
o_t = \sigma \left( w_o \left[ h_{t-1}, x_t \right] + b_o \right) \\
\tilde{c}_t = \tanh \left( w_c \left[ h_{t-1}, x_t \right] + b_c \right) \\
c_t = I_t \odot \tilde{c}_t + F_t \odot c_{t-1} \\
h_t = O_t \odot \tanh \left( c_t \right)
\]
where $F_t$ is forget gate, $I_t$ is input gate, $O_t$ is output gate. $w_f$, $w_c$, $w_i$, $w_o$ are learnable weight parameters, and $b_f$, $b_c$, $b_i$, $b_o$ are learnable bias parameters. $h_t$ is the hidden state, and $h_{t-1}$ is the hidden state of $t-1$. $\tilde{c}_t$ is candidate cell memory state and $c_t$ is cell memory state. The forget mechanism can discard the useless information from previous cell memory state $c_{t-1}$. Specifically, when the value of forget gate $F_t$ close to 0, the current cell memory state $c_t$ will not use the information from previous cell memory state $c_{t-1}$. Otherwise, the current cell memory state $c_t$ contain the information from input gate $I_t$, candidate cell memory and previous cell memory state $c_{t-1}$.

### 3.5 Proposed Model 4: CNN+Bi-GRU

CNN+Bi-GRU is the fourth proposed hybrid model. Specifically, we adopt CNN (i.e., convolution and pooling operation) to process the blood test data. The different kernel sizes from convolution operation can extract richer blood feature representations. Pooling operation reduce the calculation complexity during the training process. GRU use the memory cell to store the information, and the gated mechanism used to decide whether the information can be retained in the cell. Therefore, GRU can guarantee the useful feature representation will not be lost. Compare to the GRU model, Bi-GRU adds forward layer and backward layer so that it can capture historical information and future information. The structure of Bi-GRU is shown in Figure 5. For the purpose of better understanding, we will not repeat the internal structure of GRU (the structure of GRU model is described in Figure 2).

The calculation of Bi-GRU is shown as follows:

$$\bar{h}_t = GRU(x_t, h_{t-1})$$  \hspace{1cm} (18)
where $GRU\left(\right)$ represent the calculation process between equations (5) to (8), $h_t$ is the hidden state of forward layer and $h_t$ is the backward layer hidden state. In the forward layer, the hidden state $h_t$ is determined by $x_t$ and $t-1$ hidden state $h_{t-1}$. In the backward layer, the hidden state $h_t$ is decided by $x_t$ and $t+1$ hidden state $h_{t+1}$. Then the output $y_t$ is obtained through vector splicing of $h_t$ and $h_t$.

### 4. EXPERIMENTS AND EVALUATION

This section evaluates the performance of deep learning models by using the blood test data from Hospital Israelita Albert Einstein. We first introduce the dataset, experiment setup and evaluation methods. Then, we present experimental results analysis.

#### 4.1 Dataset

We conduct our experiments on a real dataset containing 600 blood test records for COVID-19 infection testing. These blood test records are collected in 2020 from Hospital Israelita Albert Einstein (Alakus et al., 2020). In the dataset, all patients’ personal privacy data is hidden, i.e., name, gender. Every row reflects a specific blood test result for a person. According to the work of Alakus et al. and Turabieh et al., 18 laboratory findings were identified as important factors in COVID-19 infection diagnosis (Alakus et al., 2020; Turabieh et al., 2021). Table 2 shows the name of 18 laboratory findings. In our paper, selected 18 laboratory findings as feature for our deep learning models.
4.2 Experiment Setup

We perform all evaluations using a desktop. It has an Intel Core i7 CPU (2.9GHz) with 8GB memory and NVIDIA GeForce 940FX GPU with 4GB. The development language and package are python3 and Keras, respectively. In the experiments, 80% of data is used for model training, 20% of data is used for model testing. We set the learning rate is 0.001, epoch is 200 and the batch size is 30.

4.3 Model Evaluation

In this section, we use the same model evaluation methods as Alakus et al., namely accuracy, precision, recall, F1-Score and AUC. Figure 6 is the Confusion matrix, which is used for calculating the accuracy, precision, recall and F1-Score (Alakus et al., 2020).

The matrix is composed of four types of values. First, True Positive (TP) is an outcome where the model correctly predicts the positive class. Second, True Negative (TN) is defined as the outcome where the model correctly predicts the negative class. The third value False Positive (FP) means the model incorrectly predicts the positive class. The final value is False Negative (FN), which is the number of incorrect predictions for the negative class (Visa et al., 2011).

Accuracy refers to the proportion of correct results that our model predicts. Precision is the ratio of positive identifications which is correct prediction. Recall measures the amount of positive class predictions made of all positive example from the dataset. F1-score is the harmonic average of the accuracy and recall. Mathematically, the definition of these four evaluation methods as follow.

Table 2. The name of 18 laboratory findings

<table>
<thead>
<tr>
<th>Name</th>
<th>Name</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hematocrit</td>
<td>Hemoglobin</td>
<td>Monocytes</td>
</tr>
<tr>
<td>Serum Glucose</td>
<td>Neutrophils</td>
<td>Platelets</td>
</tr>
<tr>
<td>Red blood Cells</td>
<td>Lymphocytes</td>
<td>Leukocytes</td>
</tr>
<tr>
<td>Basophils</td>
<td>Eosinophils</td>
<td>Urea</td>
</tr>
<tr>
<td>Sodium</td>
<td>Creatinine</td>
<td>Potassium</td>
</tr>
<tr>
<td>Proteina C reativa mg/dL</td>
<td>Alanine transaminase</td>
<td>Aspartate transaminase</td>
</tr>
</tbody>
</table>

Figure 6. The confusion matrix
Accuracy = \frac{TP + TN}{TN + FP + FN + TP} \quad (21)

Precision = \frac{TP}{TP + FP} \quad (22)

Recall = \frac{TP}{TP + FN} \quad (23)

F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (24)

AUC is used to evaluate the performance of classifier. If the AUC value is greater than 0.8, the classifier can be regarded as excellent classifier. If the AUC value is exceeding 0.9, the classifier can be regarded as outstanding (Mandrekar et al., 2010).

4.4 Experimental Results Analysis

We present the experimental performance in detail to prove the effectivity of our proposed model. To make comprehensive comparisons, we implement the four proposed model as well as the two models (CNN and CNN+LSTM) proposed in the work of Turabieh et al. and Alakus et al. (Turabieh et al., 2021; Alakus et al., 2020). Additionally, we evaluate the six deep learning models with five evaluation metrics, which are accuracy, F1-score, precision, recall and AUC.

Figure 7 to Figure 12 show the receiver operating characteristic (ROC) curves for the six models (Yu et al., 2021). In general, the false positive rate (FPR) is the x-value of the ROC curve, and the true positive rate (TPR) is the y-value of the ROC curve. The successive trapezoid area enclosed by ROC curve and coordinate axis represents the overall performance of the classification model, and we use the AUC value to quantify this area. The closer the AUC is equal to 1 means the stronger classification ability of the model is, and vice versa. Figure 12 shows that CNN+Bi-GRU model gets the best AUC value 0.91 among all the six models. This result illustrates that the CNN+ Bi-GRU has the best classification ability.

Figure 7. The ROC curve of CNN model

![Figure 7. The ROC curve of CNN model](image)
Figure 8. The ROC curve of CNN+LSTM model

![ROC curve of CNN+LSTM model](image)

Figure 9. The ROC curve of CNN+GRU model

![ROC curve of CNN+GRU model](image)

Figure 10. The ROC curve of CNN+Bi-RNN model

![ROC curve of CNN+Bi-RNN model](image)
The performances of six models in terms of five metrics is shown in Table 4. The AUC of all hybrid models is higher than the CNN model in the diagnosis of COVID-19. It shows the superiority of hybrid model. Specifically, the three proposed hybrid models CNN+GRU, CNN+Bi-LSTM, CNN+Bi-GRU outperform the single deep learning model CNN in terms of the five evaluation metrics. CNN+GRU model perform better than CNN+LSTM model. Although GRU and LSTM use the gating mechanisms to address the issue of gradient disappearance, GRU has a simper network structure than LSTM, so GRU will reduce the number of parameters during the training process. Therefore, GRU can accelerate the training and convergence of the network than LSTM. The experiment results show that the CNN+Bi-GRU model significantly improved the accuracy, F1-score, precision, recall and AUC over CNN+GRU. The reason is the bidirectional structure of Bi-GRU can better capture the correlations between the feature representation. Therefore, CNN+Bi-GRU has better evaluation performances than CNN+GRU.
In Figure 13, we provide the comparison results of CNN, CNN+LSTM and CNN+Bi-GRU in terms of four metrics, i.e., accuracy, F1-score, precision, recall. Specifically, CNN and CNN+LSTM are the best models proposed in the work of Turabieh et al. and Alakus et al. (Turabieh et al., 2021; Alakus et al., 2020), respectively. The comparison results show that our proposed model CNN+Bi-GRU outperforms CNN and CNN+LSTM with all four evaluation metrics.

Table 5 shows the detailed comparisons with the work of Batista et al., Turabieh et al. and Alakus et al. All these work uses the same blood test data set, which from the Hospital Israelita Albert Einstein. Batista et al. conduct SVM and RF achieves the best results are 0.87 of AUC and 0.72 of F1-score. Turabieh et al. conduct CNN model achieves the best results are 0.80 of accuracy and 0.78 of F1-score. Alakus et al. proposed CNN+LSTM model achieves the best accuracy, AUC and F1-score are 0.9230, 0.90 and 0.93, respectively. Our proposed CNN+Bi-GRU model provides the best accuracy, AUC and F1-score are 0.9415, 0.91 and 0.9417, respectively. Therefore, we can conclude that CNN+Bi-GRU outperforms the work of Batista et al., Turabieh et al. and Alakus et al.

4.5 Discussion

In this work, we propose four deep learning hybrid models and implement two models from the work of Turabieh et al. and Alakus et al. for COVID-19 prediction based on blood test. This work could be categorized as binary classification problem. In binary classification scenarios, we aim to classify the positive and negative patient. The objective of this study is speeding up the blood testing process, alleviating the pressure of heavy blood testing work from medical staff, and providing blood test prediction results with higher accuracy than previous studies. The proposed model CNN+Bi-GRU achieves the best performance in terms of accuracy, F1-score, precision, recall and AUC. This is because CNN+Bi-GRU first uses the convolution and pooling operations in CNN to establish the feature representations of blood test data, and then uses Bi-GRU to explore the correlation of different
feature representations. Based on the benefits and characteristics from bidirectional structure, Bi-GRU has a simpler network structure, and can store long-term information and use past, current, and future information during the model training process. A direct implication is that the proposed CNN+Bi-GRU in study can be utilized to develop an automated blood test system for COVID-19 infection. This system will allow patients to get results faster without waiting for several days or even over a week like traditional manual blood tests. We strongly recommend the application of the automated blood test system in underdeveloped countries and regions that require large-scale COVID-19 testing. Medical staff in these countries/regions often undertake more arduous blood testing tasks.

5. CONCLUSION

In this paper, we proposed four deep learning hybrid models for COVID-19 prediction based on blood test, namely CNN+GRU, CNN+Bi-RNN, CNN+Bi-LSTM, CNN+Bi-GRU. Additionally, we implement two best models CNN and CNN+LSTM from the work of Turabieh et al. and Alakus et al., respectively. We used the blood test data from Israelita Albert Einstein Hospital to train and test these six models. The models’ performance is evaluated by five methods, i.e., accuracy, precision, recall, F1-Score and AUC. The experiment results show that our proposed hybrid model CNN+Bi-GRU outperforms the best models from Alakus et al. and Turabieh et al. In summary, our proposed model CNN+Bi-GRU achieves four desirable properties: 1) it provides better performance than previous works can prevent false-negative patients to spread the epidemic, 2) it speeds up the blood testing process can isolate and treat positive patients as early as possible, 3) it relieves the healthcare workers from demanding testing work and 4) it can operate unmanned 24/7. Therefore, CNN+Bi-GRU has shown significant practical value to be widely applied in the prevention and cure of this global pandemic. And we can envisage to deploy CNN+Bi-GRU in automated blood test system, especially in underdeveloped countries and large-scale COVID-19 infection detection regions. However, the main limitation is that we cannot avoid problems caused by contamination of blood data samples during blood collection. This limitation will lead to errors in the prediction results. Therefore, one of future work is that we plan to investigate a model for testing whether blood data samples are contaminated. On the other hand, Chang et al. suggested that we cannot sacrifice the patient privacy from deep learning on COVID-19 diagnosis (Chang et al., 2021). And Generative Adversarial Networks with Federated Learning is a realizable privacy protection method during the training process of model. Therefore, we also plan to apply it for COVID-19 prediction in our future works.

ACKNOWLEDGMENT

This work was supported in part by the Macao Polytechnic University – Edge Sensing and Computing: Enabling Human-centric (Sustainable) Smart Cities (RP/ESCA-01/2020). Lihua He, affiliated with Macao Polytechnic University, is the corresponding author and co-first author of this paper.
REFERENCES


Ziyue Yu is currently a Ph.D. student of Computer Applied Technology at Macao Polytechnic University (MPU), advised by Prof. Tansim Tse, Dr. Wuman Luo from MPU, and Prof. Giovanni Pau from the University of Bologna and University of California, Los Angeles. Before that, he received his M.Sc. degree in Advanced Information Systems from Hong Kong Baptist University. His research interests are graph representation learning, healthcare analysis. He won the Best Student Paper Award in IoTBDS 2021.

Lihua He received the M.Sc. degree in Mathematics from Yunnan Minzu University, Kunming, China, in 2020. She is currently pursuing the Ph.D. degree at the Macao Polytechnic University (MPU), advised by Dr. Wuman Luo from MPU. Her research interests focus on deep learning and data mining, especially their applications in spatial-temporal data mining and urban computing.

Wuman Luo is currently an Associate Professor at the Faculty of Applied Sciences at Macao Polytechnic University (MPU). She received the Ph.D degree in Computer Science and Engineering from Hong Kong University of Science and Technology (HKUST) in 2013. Her research interests include Artificial Intelligence and Data Analysis in Big Data, Database System, Parallel Computing. Her research works have been published in the international well-known conferences and journals, including ACM SIGMOD, ACM CIKM, IEEE ICDCS, IEEE ICPADS, IEEE MDM, Wireless Communications and Mobile Computing, etc. She has served as Principal Investigator or Principal Participant for many research and industrial projects, such as National Basic Research Program of China (973 Program), National Natural Science Foundation of Youth Program, Demonstration Research and Application of Expressway Electronic Toll Collection System in Sichuan Province, Real-Time Toll System for Sichuan Province Mobile Telecommunication, etc.

Rita Tse is the Dean of the Faculty of Applied Sciences at the Macao Polytechnic University. Her research interests include ubiquitous computing and urban sensing.

Giovanni Pau is a full professor at the University of Bologna. Previously he served as ATOS/Renault smart mobility Chair-Professor at the Sorbonne Université Paris, France. Before joining Sorbonne he was senior research Scientist at UCLA Computer Science Department where he still retains the title of Adjunct Faculty. Dr. Pau holds the “Habilitation à diriger des recherches” (HDR) granted by the Sorbonne Université in 2018. He was awarded the Italian Laura degree in Computer Science, the PhD in Computer Engineering by the University of Bologna in 1998 and 2002 respectively. Dr Pau published over 100 peer reviewed papers in scientific journals and international conferences.