Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Integrating Models and Fusing Data in a Deep Ensemble Learning Method for Predicting Epidemic Diseases Outbreak

Nesrine Ben Yahia *, Mohamed Dhaeddine Kandara, Narjes Bellamine BenSaoud

RIADI laboratory, National School of Computer Sciences, University of Manouba, 2010, Tunisia

ABSTRACT

Due to the continuous and growing spread of the novel coronavirus (COVID-19) worldwide, it is urgent, especially in the data science era, to develop accurate data driven decision-aided methods to predict and early detect the outbreak of this epidemic disease and then to support healthcare decision makers. In this context, the main goal of this paper is to build an accurate and generic data driven method that can predict daily COVID-19 positive cases and therefore helps stakeholders to make and review their epidemic response plans. This method is based on the integration of three deep learning models: Long Short Term Memory (LSTM), Deep Neural Networks (DNN) and Convolutional Neural Networks (CNN) and takes advantage of their complementarity. The proposed method is validated on two experimental scenarios where the first one aims to validate the method on China and Tunisia case studies and the second one is based on data fusion and transfer learning process where China data and models will be reused to predict Tunisia COVID-19 outbreak. Experiment results indicate that, compared with individual learners, the stacked-DNN meta-learner, whose inputs are results of LSTM, DNN and CNN learners, achieved the best results in terms of accuracy as well as RMSE and it required the lowest time for training as well as prediction for the two scenarios. The main outcomes of this paper are i) to adopt deep learning models combined to stacking ensemble learning to accurately forecast COVID-19 positive cases and ii) to merge data and to adopt transfer learning for the prediction of confirmed cases by reusing China data, learners and meat-learners to make prediction of the epidemic trend for other countries, with less facilities of collecting data, when preventive and control measures are similar.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

COVID-19 [1], the official name announced by the World Health Organization (WHO) for the 2019 novel coronavirus with the current reference name SARS-CoV-2, has recently emerged as a severe acute respiratory syndrome. COVID-19 outbreak was originally reported in China, but it has subsequently been spread rapidly and continually across the world. In fact, this epidemic disease is persistently threatening public health worldwide, causing serious and critical concerns. Thus, it is urgent, especially in the data science era, to develop accurate data driven decision-aided methods to predict and early detect COVID-19 outbreak and then to support healthcare decision makers in controlling COVID-19 for a successful public health response plans. In this context, we mainly focus in this study on healthcare analytics, data analytics artificial intelligence (AI) solutions to contribute in the battle against COVID-19 outbreak. In fact, the growing healthcare industry is generating a large volume of clinical, medical and administrative data attracting the attention of practitioners and academicians alike. Thus, healthcare analytics, in the data science era, are introduced to provide accurate techniques and tools to properly manage this big and complex data and to support decision making in healthcare [2].

In the last decade, data analytics in healthcare field has grown rapidly which has prompted increasing interests in the generation of analytical and data driven methods based on machine and deep learning in health informatics [3]. In this context, we aim here to apply deep learning advances, a machine learning sub-field, in order to forecast COVID-19 daily confirmed or positive cases and to early detect its outbreak. Indeed, clinicians and practitioners need to predict and estimate this number in order to make the necessary measures such as treatment measures, quarantine protocols and hospitals preparedness.

The goal of this paper is to propose a generic and data driven method for the prediction of COVID-19 outbreak that is based on deep learning techniques. The proposed method can be used as a healthcare decision support tool for different countries. Our mo-
tivation is to exploit the confirmed advantages of deep learning techniques both in health decision aided processes and epidemic diseases outbreak [4]. And, our main contribution is to use firstly three chosen deep learning models, LSTM (Long Short-Term Memory), DNN (Deep Neural network) and CNN (Convolutional Neural Networks). Then, these models will be stacked in ensemble learning models in order to generate the most accurate results. The meta-learners use as inputs the predicted values of these three learners in order to generate the final COVID-19 outbreak predictions.

From another side, due to the highly complex nature of the COVID-19 outbreak, the high level of uncertainty and lack of essential data for some countries with less facilities of collecting data, the main issue is not only to consider the accuracy and robustness abilities of predictive models but also their generalization. In this context, this paper presents an attempt to address this double issue. Furthermore, as training deep neural networks and deep ensemble learning require a large amount of computational resources, time and especially big data, we reflect on the adoption of data fusion and transfer learning by reusing trained predictive models on one county dataset for another one when we consider the hypothesis that these two countries have comparable policies and establish similar measures to fight COVID-19. In fact, in this paper, we focus on data fusion, which is a technology that merges data to obtain more informative, consistent and accurate information than the original data that are mostly imprecise, uncertain, inconsistent and insufficient [5], to build accurate prediction for countries that can have some difficulties in collecting COVID-19 data.

More specifically, the novelties of this research consider two research questions (RQ):

- **RQ1**: In terms of accuracy, can deep ensemble learning, based on knowledge sharing, improve prediction of epidemic disease outbreak, COVID-19 example, and what can be the best unified method i.e. combination of deep learning models with highest accuracy and lowest time cost?
- **RQ2**: In terms of generalization, can we merge data and adopt transfer learning, based on knowledge transfer, to reuse models trained from one country data to predict COVID-19 trends for other countries when preventive policies are comparable or similar?

The paper proceeds as follows: In Section 2, we review related works on healthcare analytics and deep learning techniques used for the prediction of COVID-19 outbreak. In section 3, we briefly introduce the background knowledge of LSTM, DNN and CNN models that we use in this study. In Section 4, we detail our method and its salient contributions. Section 5 is dedicated to present the experimental results and section 6 is devoted to discuss our method and to present its threats of validity. Finally, we conclude and summarize our main contributions in section 7.

2. Related works

The main goal of this section is to review some of the main recent related works on applying machine and deep learning advances for the prediction of COVID-19 outbreak. We especially focus on recent works whose goals are to predict daily confirmed or confirmed cases. For example, a machine learning approach is proposed in [6] for predicting the daily numbers of cumulative confirmed cases, new confirmed cases and death cases of COVID-19 in China from Jan 20, 2020, to Mar 1, 2020 using the data of the National Health Committee of China. [7] presented also a comparative analysis of machine learning and soft computing models to predict the COVID-19 outbreak for five counties China, Italy, Ger-

many, Iran and USA. The results of two machine learning models (multi-layered perceptron and adaptive network-based fuzzy inference system) showed promising results and reported a high generalization ability for long-term prediction. Thus, they suggested that, due to the highly complex nature of the COVID-19 outbreak and variation in its behavior from nation-to-nation, machine learning is an effective tool to model the outbreak. In [8], Recurrent Neural Networks (RNN), that can model temporal (sequential) data prediction, are used for predicting COVID-19 confirmed (positive), negative, released and death cases. They proposed three models, a Long short-term memory (LSTM) model, a Gated Recurrent Unit (GRU) model and a combined LSTM-GRU model. Experimental results on COVID-19 dataset of South Korea from January 20, 2020 to March 12, 2020 show that the highest accuracy is obtained by the combined model.

In [9], a Convolutional Neural Network (CNN) model is proposed to predict the number COVID-19 confirmed cases in China using dataset from January 23, 2020 to March 2, 2020 provided by Surging News Network and WHO. Experiment results indicated that the proposed CNN model is the best performing algorithm compared with MLP (Multilayer Perceptron), LSTM and GRU. [10] used equally deep Learning-based models, specifically, LSTM variants such as deep LSTM, convolutional LSTM and bi-directional LSTM models for predicting the number of novel coronavirus (COVID-19) positive reported cases for 32 states and union territories of India and concluded that bi-directional LSTM gives the best results, and convolutional LSTM gives the worst. [11] presented a comparative study of five deep learning models (simple RNN, LSTM, Bidirectional-LSTM, Gated recurrent units (GRUs) and Variational AutoEncoder (VAE)) to forecast the number of new cases and recovered cases for six countries namely Italy, Spain, France, China, USA, and Australia. Their results demonstrated the promising potential of the deep learning models in forecasting COVID-19 cases and highlight the superior performance of the VAE compared to the other algorithms. [12] proposed an algorithm to perform and evaluate the ARIMA model for 145 countries, which are distributed into 6 geographic regions (continents) where they attempted to create a relationship between the countries, which are in the same geographical area to predict the advance of the virus. In [13], DNN and LSTM models were already compared with the auto-regressive integrated moving average (ARIMA) for the infectious diseases prediction where results showed that the DNN and LSTM models perform better than ARIMA. This point justifies also our selection for deep learning techniques that have proven their accuracy and performance in predicting COVID-19 outbreak. In this context, the goal of this paper is not only to find the best accurate predictive model but also to integrate different models in order to take advantage of their complementarity. Thus, the salient contribution of this study is to propose a unified data driven accurate and generic predictive method for the prediction of COVID-19 outbreak. In fact, in this method, we aim to test the ability of LSTM, DNN and CNN models to accurately forecast daily COVID-19 positive cases. The selection of these three predictive models is motivated by the confirmed advantages of LSTM and DNN to predict epidemic and infectious diseases outbreak [13] and also the confirmed practicality of CNN to deal with this same issue [9]. Then, we will reflect on the integration of these models in a unified method based on ensemble learning to combine them in deep meta-learners that fuse the predicted values of LSTM, DNN and CNN learners in order to improve the prediction accuracy and to give best accurate results. The background knowledge of these three models is presented in the next section.
3. Background knowledge of LSTM, DNN and CNN

3.1. Long Short Term Memory Networks (LSTM)

LSTM are an amelioration of recurrent neural networks (RNN) that are able to model sequential and temporal data and to predict times series [14]. More specifically, a cell state is added in LSTM to store long-term states and to build more stable RNN for time series prediction by detecting and memorizing the long-term dependencies existing in the time series. LSTM have recently attracted much interest in temporal data processing of infection and epidemic diseases such as in [15] where authors proposed a LSTM model for real-time influenza forecasting and to capture the seasonal flu temporal dynamics.

3.2. Deep Neural Networks (DNN)

DNN, [16], are deep Artificial Neural Networks (ANN) with multiple (at least two) hidden layers where the “deep” refers to the number of hidden layers through which the data is transformed from the input to the output layers. In classical DNN, each layer is composed of a set of neurons and an activation function and is fully connected. A set of weights is affected to each neuron where each weight is multiplied by one input into the neuron. They are then summed to form the output from the neuron after it has been fed through the activation function. In order to be able to give performing results, deep neural networks often require huge numbers of training data and big data [17].

3.3. Convolutional Neural Networks (CNN)

CNN, [18], contain generally four types of layers in their structure: an input layer, convolutional layers, pooling layers, and fully connected layer (output). In the convolutional layer, which represents the most important CNN part, the input will be convoluted with different filters where each filter is considered as a smaller matrix. Then, corresponding feature maps will be generated after the convolution operation. The pooling operation consists in reducing the size, while preserving the important features. The efficiency of the network is thus improved and over-fitting is avoided. So the main role of convolutional and pooling layers is generally to extract features, and the main goal of fully connected layers is usually to output the information from feature maps together, and then provide them to later layers.

To conclude, LSTM, DNN and CNN are complementary in their modeling capabilities. In fact, LSTM are good at contextual and temporal times series modeling, DNN are appropriate for mapping different features to a more separable space with many hidden layers, and CNN are good at extracting features and reducing frequency variations. Thus, we mainly focus in this study on integrating these deep learning models in an ensemble method to take advantage of their complementarity and to improve the accuracy of COVID-19 outbreak prediction. Furthermore, as training deep learning models require a large amount of computational resources, time and especially big data, we aim also in this study to adopt data fusion and transfer learning by reusing trained predictive models on one country dataset for another one. These contributions will be presented in the next section.

4. Materials and methodology

In this study, we propose a generic and data-driven method that may be used as a healthcare decision support tool for forecasting COVID-19 epidemic trend and which can be exploited by several countries. The Fig. 1 illustrates the pipeline of the proposed data driven method where we aim in a first step to train the three chosen deep learning learners (LSTM, DNN and CNN). Then, in a second step, we will use these forecasted outputs learners to form a new time series dataset that we will use to train ensemble learning meta-learners (stacked LSTM, stacked DNN and stacked CNN). Our objective here is to find the best integration and combination by identifying the best accurate ensemble method in the end that improves the forecasting accuracy. Materials used in this study and the different steps of this method will be explained below.

4.1. Materials

Data used in this method are administrative medical data that include COVID-19 confirmed cases (CC) as well as the death cases (DC) and the recovered cases (RC) daily numbers across the world starting from January 22, 2020 until November 09,2020 provided by the verified sources of universal John Hopkins University [19]. These data include time series COVID-19 confirmed cases datasets worldwide where rows represent countries and columns represent cases number (confirmed, recovered and deaths) for each country. In terms of data size, for each country, we used 293 columns for confirmed cases number, 293 columns for recovered cases number and 293 columns for death cases in the original data set time series (number of days of the studied period). From these data, including time series datasets and global situation reports, we only extract the data of China and Tunisia. We also study the ongoing COVID-19 outbreak for Tunisia using the time series data provided.
by the official sources the Tunisian National Observatory of New and Emerging Diseases\(^1\) that represent official data published by the Tunisian ministry of the public health.

4.2. Grid searching process steps

Successful ANN applications usually depend on the appropriate choice of the best modeling hyperparameters. These modeling hyperparameters can concern the network architecture (hidden layers number and units or neurons number in each layer, etc.) or data preparation such as delays number or the window size that will be used in time series applications [20]. That's why, we will use in this study the grid searching technique whose role is to select the optimal and suitable hyperparameters modeling by performing optimization decisions based on several combinations and solid statistical criteria [21]. In fact, the role of this technique is to ensure that we choose the best and suitable hyperparameters values and it is helpful to mention that without grid search, these values are randomly selected and then used when training the models. However, the combination of the selected values may not be the best one i.e. the optimal one. According to [20], Grid searching technique searches exhaustively the different possible combinations of all hyperparameters values and evaluates models accuracy on these different combinations then it selects the optimal values. In our case, grid searching technique will be used for data preparation to find the best window size hyperparameter and for models architecture to select the best number of hidden layers and their neurons. We also rely on grid searching technique to select the best optimizer whose role is to minimize the loss function by updating weight parameters and the value of the dropout layer whose role is to select relevant neurons and to prevent over-fitting by ignoring irrelevant neurons during training. For the data preparation hyperparameter, i.e. the window size or lag, we remind here that our input data represent COVID-19 confirmed cases times series dataset. This dataset can be considered as a sequence of vectors, \(x(t)\), where \(x\) represents COVID-19 confirmed cases number and \(t\) represents elapsed time i.e. days where \(x\) varies continuously with \(t\). Furthermore, to find the different possible values of the window size, we will use the sliding window method that is commonly used in time series dataset based ANN predictors [22] and where the general principle is to use \(w\) previous times steps or \(w\) previous series values as input data or features and the next step as the output target value. So, the time steps value or the window size is part of our Grid dictionary (a grid of parameters with a discrete number of values for each one) and instead of randomly selecting its value, the used Grid search cross validation technique is used here to automatically find its best value. Following \(t\), we will check different values of the window size (from 2 to 6). So, we will test 2, 3, 4, 5 and 6, and according to Grid search, the best one will be returned. For architecture models hyperparameters, the selected values to be tested and that are commonly used in the literature are 3, 4, 5 and 6 for the number of hidden layers, and 16, 32, 64, 128, 256 and 512 for the number of neurons of each layer. Concerning the optimizer, we will try Adam optimizer, Root Mean Square Propagation (RMSprop) and Stochastic Gradient Descent (SGD). Finally, for the dropout layer value, we will test the values 0.1, 0.2, 0.3, 0.4 and 0.5. Then, by varying and combining these hyperparameters values, grid search technique will be applied, for the three learners in a first step and for the stacked meta-learners in a second step where it will select for each model the best combination that generates the most accurate model using the coefficient of determination \(R^2\) score that measures how good models might be constructed from these values combinations. This coefficient presents the proportion of the variance in the dependent variable that is predictable from the independent variable(s). In fact, it provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model. Next, the optimal value of the window size selected by grid search will be used for data preparation and the best network architecture hyperparameters selected by grid search will be used to build different learners and meta-learners.

Furthermore, in order to verify the generalization capacity of the proposed method and to avoid the overfitting caused by models learned and tested on the same data, we are based on the classical cross validation approach by splitting the available data into two sets (80% for the training set and 20% for the test set). Then, we are based on the K-folds cross-validation process to ensure that the performance of the selected hyper-parameters will be measured on a dedicated validation set that was not used during the model selection step. The idea of the K-folds cross-validation process is to divide the data into K folds. Out of these K folds, K-1 parts are used for training while the remaining part is used for testing. The procedure is repeated K times, each time a new part is used as testing set while remaining parts are used for training. Finally, the result of the K-folds cross-validation process is the average of the results obtained on each part. In our case, we have 80% of the dataset dedicated to the train set on which we perform 5-folds cross-validation as illustrated in Fig. 2. In fact, the training-set was randomly divided into five folds: one fold was used as validation set and the other 4 folds as training set. We repeat this procedure 5 times where in each of the iterations, a different part was taken as the validation set and finally the average prediction error was obtained by assessing the average errors in the 5 iterations performed on each 5 validation sets. In fact, to do this, we used the cross_val_score helper function of Scikit-learn to estimate the accuracy of each model by splitting the data, fitting the model and computing the score 5 consecutive times (with different splits each time) as illustrated in Fig. 2.

4.3. Data preparation

In this step, our input data, dataset 0, which represents COVID-19 daily confirmed, recovered and death cases time series will be reframed using the optimal value of the window size selected by the grid searching technique. The output is the new time series dataset 1 which will be used for the training of our deep learning models or learners.

4.4. Training of DNN, LSTM and CNN learners

In this first training step, we separately build the three selected learners (LSTM, DNN and CNN) where they will learn their best parameters i.e. weights of neurons of different layers basing on the optimal hyperparameters selected by grid searching technique. They will use dataset 1, the result of data preparation, as input dataset. Then, their forecasted outputs will be collected in a new dataset 2.

4.5. Training of stacked meta-learners

In this phase, we aim to combine and to gather the knowledge of the different trained learners and to take advantage of their complementarity basing on an ensemble learning method. According to [23], there are three techniques of ensemble learning: bagging, boosting and stacking. In bagging, homogeneous (weak) learners are trained independently from each other in parallel and then combined following a deterministic averaging process. In boosting, homogeneous weak learners are sequentially

\(^{1}\) http://onligne.tn/fr/index.php.
4.6. Algorithm of the proposed method

In order to illustrate the process of the proposed unified method in detail, we expose in Fig. 3 an algorithm that formally presents the input, output, parameters and steps of our proposal. In fact, our main goal is to test, compare and combine the three deep learning learners i.e. DNN, LSTM and CNN that are recommended in the literature for epidemic outbreak as explained previously. The input of this algorithm is a time series \( Y_1 \) that presents the total confirmed, recovered and death COVID-19 cases per day. Then, the time series \( Y_1 \) is separated into training time series (\( X_{	ext{Train1}}, Y_{	ext{Train1}} \)) using 80% of the data from the original time series \( Y_1 \) and testing time series \( (X_{	ext{Test1}}, Y_{	ext{Test1}}) \) presenting the other 20%. In the first step, the three learners are trained and then tested and compared using Accuracy (Acc) and Root Mean Square Error (RMSE) that measures the stability between the original data and forecast data. In the next step, and in order to improve the prediction accuracy, the predicted values \( \text{Pred1}, \text{Pred2} \) and \( \text{Pred3} \) of the learners DNN, LSTM and CNN are stacked to form a new unified input \( X_2 \) for the same period \( t \). The output is always \( Y_1 \), so, for a given date \( t \), the COVID-19 cases number will be forecasted using the predicted values of these learners for the same date \( t \). Then, this new dataset is divided into training time series \( (X_{	ext{Train2}}, Y_{	ext{Train2}}) \) and testing time series \( (X_{	ext{Test2}}, Y_{	ext{Test2}}) \) forming respectively 80% and 20% of this input. These unified training and testing time series will be then used to train deep learning meta-learners (S-DNN, S-LSTM and S-CNN) and to output a unified final prediction based on the different learners’ predictions. Finally, the algorithm returns the best candidate unified method i.e. the best performing deep ensemble learning meta-learner with maximum accuracy and minimum RMSE.

5. Experimental results

In order to validate and evaluate the proposed data driven method for the prediction of COVID-19 outbreak, two experimental scenarios will be presented in this section. In the first scenario, we aim to prove the genericity of our method and to answer our RQ1 by dealing with models knowledge sharing and checking if the integration of models in a deep ensemble learning method can really improve the prediction accuracy by choosing two cases studies: Tunisia and China. China has been chosen as it represents the first country where COVID-19 outbreaks and especially for which a huge amount of open data is available. We have also chosen to apply our method on Tunisia, our country, where containment preventive and control measures are applied early similarly to China, and where we are following the dynamics of the propagation day by day. In the second experimental scenario, we aim to answer our RQ2 by dealing with models knowledge transfer and reusing models trained from one country data to predict COVID-19 trends for other countries when preventive policies are comparable or similar (such as quarantine policies). Furthermore, root mean squared error (RMSE) metric which is commonly used for the evaluation of infectious diseases spreading predictors [13] and that measures the difference between forecasted and real values is used in this study to evaluate learners and meta-learners. We equally use the
Fig. 3. Algorithm of the proposed method.

accuracy that identifies the overall effectiveness of the models. Regarding consumed time and in order to estimate run-time costs, we also measure the required time (in seconds) for training as well as prediction for each learner and meta-learner.

5.1. First experimental scenario: method application on Tunisia and China case studies

In order to validate the proposed method, we start with applying the grid searching technique on the COVID-19 time series datasets of the two case studies. We note that grid searching cross validation technique selected the same hyperparameters for the two case studies. In fact, according to the used grid searching cross validation technique, the optimal hyperparameters for the two datasets are as follows. The best value of the window size or lag for the three learners (LSTM, DNN and CNN) is 3. So, the time series dataset will be reframed and prepared using three times steps as input to predict the next step. For instance, the number of COVID-19 confirmed cases of Day1, Day 2 and Day 3 consider the input of the target value of Day 4. So, to predict \( x(t) \) we need \( x(t-3), x(t-2) \) and \( x(t-1) \). LSTM learner encompasses three hidden layers with 256 neurons where “relu” activation function is used for each layer. Its input layer contains 3 neurons that represents the three time steps. Its output layer contains one neuron that represents the next time step with the “linear” activation function and it is preceded by a dropout layer with the value 0.3.

Regarding the optimizer, the Adam function has been selected by the grid searching cross validation technique. DNN learner also encompasses three hidden layers that contain respectively 256, 64 and 16 neurons. The “relu” activation function is equally used for each one. Its input layer contains the three neurons presenting the three time steps and its output contains the next step presented by a single neuron with the “linear” activation function. Regarding the optimizer, Adam function is equally selected by the grid searching technique and a dropout value of 0.2. CNN learner input layer has the same three neurons. Then, it encompasses a convolutional layer containing 128 units, a max-pooling layer containing 128 units, a convolutional layer containing 32 units, a convolutional layer containing 64 units then an average-pooling layer containing 64 units, a dropout layer containing 64 units with the value 0.3 and finally a fully connected output layer containing a single unit. The “relu” activation function is equally used for all these layers except the output layer that is based on the “linear” activation function. Then, during the training step, these three learners, for each case study, will learn in parallel their parameters i.e. the weights of their neural networks using as input the reframed dataset 1 that contains times series of Tunisia COVID-19 confirmed cases. The forecasted outputs of these learners will be collected in the new dataset 2 to be used for the training of the stacked meta-learners. Finally, during the stacking method, three meta-learners (stacked-LSTM, -DNN and stacked-CNN) will be evaluated and compared for each case study in order to identify the
best meta-learner with the highest accuracy. These meta-learners will be trained using the predicted values of individual learners (LSTM, DNN and CNN). Experimental results of Tunisia and China case studies learners and meta-learners are summarized in Table 1. Evaluation of these models is based on accuracy, RMSE, time (in seconds) for training (t-time) and prediction (p-time) using Tunisia and China time series.

Table 1
Performances and time costs of learners and meta-learners for China and Tunisia case studies.

| Models               | Accuracy | RMSE   | t-time | p-time |
|----------------------|----------|--------|--------|--------|
| Tunisia LSTM         | 0.62     | 80,9212 | 38,372 | 0.0234 |
| Tunisia DNN          | 0.80     | 10,3384 | 2,3584 | 0.0078 |
| Tunisia CNN          | 0.79     | 7,8214  | 2,9859 | 0.0097 |
| Tunisia Stacked-LSTM | 0.82     | 75,3678 | 15,7361| 0.0330 |
| Tunisia Stacked-DNN  | 0.97     | 2,396   | 2,0545 | 0.0075 |
| Tunisia Stacked-CNN  | 0.91     | 10,5528 | 2,8836 | 0.009  |
| China LSTM           | 0.53     | 130,92  | 62,13  | 0.4687 |
| China DNN            | 0.71     | 14,17   | 7,55   | 0.0302 |
| China CNN            | 0.69     | 15,14   | 4,84   | 0.0846 |
| China Stacked-LSTM   | 0.72     | 65,24   | 135,142| 0.6065 |
| China Stacked-DNN    | 0.92     | 5,43    | 6,32   | 0.0235 |
| China Stacked-CNN    | 0.84     | 18,13   | 6,2681 | 0.0329 |

5.2. Second experimental scenario: data fusion and transfer learning from China to Tunisia

In this second experimental scenario, we aim to adopt transfer learning technique to build stacked meta-learners from the learners of China case study to predict COVID-19 epidemic outbreak in Tunisia and then across the world. In fact, our main goal here is to reuse the models already trained and built from China time series dataset to predict COVID-19 trend in Tunisia. Our motivations are to increase the size of training dataset as ANN often require a good volume of data [17] and to fight lack of data problems for some countries. In fact, some countries have better facilities of collecting COVID-19 data. This data can be used for countries with less facilities of collecting data to control and predict COVID-19 cases and to augment their dataset. Thus, data fusion can be used here to deal with this issue. In fact, the most important elements of data fusion are data sources (single or multiple data sources), operation (operation of combination of data and refinement of information, which can be described as transformation) and purpose (gaining improved information with superior reliability and less error possibility in detection or prediction and as the goal of fusion) [5]. In our cases, we have two data sources i.e. two time series datasets where the first one represents the COVID-19 trend for China and the second one for Tunisia. Our operation consists on fusing them into one dataset where its first part i.e. China data will be used for as training set for models learning and its second part i.e. Tunisia data will be used as test set to evaluate the trained models, to check their accuracy and their ability of generalization for new data. Finally, our purpose is to augment the data and to build a necessary and sufficient volume of data for our deep learning models in order to generate accurate prediction. Moreover, we propose also the use of transfer learning for the prediction of COVID-19 confirmed cases by using models trained on data from one country with the same preventive and control measures. In such cases, knowledge transfer or transfer learning would greatly improve the performance of learning by avoiding rebuilding from scratch predictive models using newly collected training data with much expensive data-labeling efforts [24]. So, in this scenario, China data are used as training data and Tunisia data are considered as Test data. In this case, trained models on China data will transfer their knowledge to accurately and rapidly predict COVID-19 confirmed cases for Tunisia or other countries. Thus, learners are not trained again from scratch. In fact, we used the models already trained for China case study which allows us to reduce computational time costs. The performances of the stacked meta-learners are precisely summarized in Table 2 where the three options (stacked-LSTM, stacked-DNN and stacked-CNN) will be tested again to find the best candidate.

6. Discussion

Recently, deep learning models demonstrated important improvements when handling and forecasting time-series data. In this study, we examined carefully whether deep ensemble learning could build accurate methods for the prediction of epidemic disease outbreak. In fact, we proposed a generic and accurate data-driven method for forecasting COVID-19 epidemic outbreak across the world by integrating and stacking three complementary deep learners (LSTM, DNN and CNN). The proposed predictive method is validated on two experimental scenarios where results demonstrate that the proposed stacking method can indeed improve the prediction accuracy. In fact, experimental results show that compared with individual learners, the stacked meta-learners that are based on an ensemble learning technique achieved higher accuracy by fusing the predicted values of LSTM, DNN and CNN learners. More specifically, the experimental results illustrated in Table 1 show that, for the two case studies China and Tunisia, the stacked-DNN whose inputs are predicted values of LSTM, DNN and CNN perform better than the stacked-LSTM and the stacked-CNN and they have the greatest prediction performance with 0.97 of accuracy and 2.396 of RMSE for Tunisia case study and with 0.92 of accuracy and 5.43 of RMSE for China case study. It is helpful to mention here that one of the main objectives of this research is to make a comparative study of deep learning models and deep ensemble learning models to select the best candidate for an accurate prediction of epidemic disease: COVID-19 case study. We aim specifically to find the best combination or unified method in terms of accuracy, RMSE and time costs. That’s why we tested the three selected learners (LSTM, DNN and CNN) and the three combination i.e. meta-learners (Stacked-LSTM, Stacked-DNN and Stacked CNN). By the end, our findings demonstrate that the unified method based on Stacked-DNN is the best performing unified method compared with its counterparts tested. In fact, [25] empirically evaluated several state-of-the-art methods for constructing ensembles of heterogeneous classifiers with stacking and showed that combining classifiers with stacking method is better than selecting the best one from the ensemble. Furthermore, in terms of consumed time and run-time costs, experimental results show also that compared to the different combinations, the stacked-DNN converges faster and requires less time for training as well as prediction.

So, we end up by recommending deep ensemble learning based on stacked-DNN combination meta-learner that achieved the highest accuracy and required the lowest time for training as well as prediction. Furthermore, we compare in Table 3 our experiment results with others recent studies results that aim to predict COVID-19 positive cases using machine, deep and statistical models. We specifically used the RMSE metric to compare them as it is mainly and commonly used in these works to evaluate models performance. We remark that our proposed unified method based on stacked-DNN performs better in terms of RMSE as it minimizes
Table 3

| References | Models        | Countries     | RMSE      |
|------------|---------------|---------------|-----------|
| [7]        | MLP           | China         | 2318.22   |
| [7]        | ANFIS         | China         | 2960.81   |
| [12]       | ARIMA         | China         | 1177.88   |
| [12]       | ARIMA         | Tunisia       | 8.77      |
| [11]       | LSTM          | China         | 784.579   |
| [9]        | LSTM          | Tunisia       | 2994.851  |
| [9]        | CNN           | China         | 109.439   |
| Our unified method | Stacked-DNN  | Tunisia       | 2.396     |
| Our unified method | Stacked-DNN  | China         | 5.43      |

its value RMSE compared to its counterparts (2.396 for Tunisia and 5.43 for China). Thus, our study demonstrates the feasibility and practicality of deep ensemble learning methods to deal with epidemic disease outbreak and to support healthcare decision process.

Our findings indicate also that it is possible to merge data and to reuse China learners and trained models to make transfer learning of COVID-19 epidemic trend in Tunisia. In fact, experimental results shown in Table 2 demonstrate that the stacked-DNN meta-learner, trained on China dataset and tested on Tunisia dataset, resulted 0.99 of accuracy and 2.1736 of RMSE. Thus this transfer learning scenario proves that, instead of training all networks from scratch, we can reuse models already trained from China case study which allows us to increase the size of training data and to converge faster. We also conclude that countries that adopted the same preventive measures and control policies are comparable in the outbreak of COVID-19.

Thus, our answers to the two research questions are as follows:

- Concerning our RQ1, we confirm the application of deep ensemble learning that takes advantage of the complementarity of DNN, CNN and LSTM and which significantly improves generalization accuracy. In fact, our experimental results for two cases studies (Tunisia and China) demonstrate that the stacked-DNN combination is the best as it is characterized by a higher accuracy and a lower time cost. In terms of future works, we aim to apply this accurate method that can predict daily positive cases of COVID-19 patient for other countries. We aim also to take into account during training other factors which are associated with the spread and outbreak of COVID-19 and that can provide more meaningful analysis and hopefully, more reasonable predictions. These factors include for instance, culture, politics, education, minimizing outdoor activities, enforcement policies of wearing masks, etc. We equally aim to test and validate this ensemble learning method for the prediction of other epidemic disease outbreak such as influenza.

- Concerning our RQ2, our findings from the transfer learning where models trained on China data and tested on Tunisia data, demonstrate that it is possible to reuse models trained from one country to predict COVID-19 for other countries when preventive policies are comparable. Furthermore, this scenario allows us to share data and to reduce computation costs (i.e. time training). In terms of future works, we aim also to apply the transfer learning for other countries and to study the real hypotheses to be considered and that can impact the similarity between countries (geographical position, weather data...)

At this stage, we assume that our research findings might have some threats of validity, and we try to self-assess them here in order to denote the trustworthiness of our experimental results, to what extent they are correct and not biased by our subjective point of view. In addition, we treat these potential threats according to the classification proposed in [26]. Concerning the construct validity, we assume that the provided measures could be biased regarding the researchers’ expected results. However, RMSE has been used in this study, to evaluate the adopted learners and meta-learners performance, which is usually and commonly used in the prediction of epidemic diseases spreading [13]. We also use accuracy which is considered among the standard metrics that reduce biases. Concerning the external validity, there might be some threats regarding the generalization of our proposed method. To overcome this issue, this method has been validated on two case studies (Tunisia and China) which can provide more consistent feedback about the relevance of our results. Finally, regarding reliability, there might be a potential issue that concerns the dependency of data and analysis on the specific researchers. However, in order to minimize this threat, we proposed the transfer learning experimental scenario where China data and models are used for the prediction of COVID-19 outbreak in Tunisia.

7. Conclusion and learned lessons

In this study, we proposed a generic and accurate COVID-19 outbreak predictive method that may be used as a decision support tool for improving its surveillance, controlling and managing epidemics. This method has been validated on two case studies (Tunisia and China) to predict the daily COVID-19 positive cases number. We also merged data and performed learners from China case study and reused them for Tunisia case study. Our findings demonstrated that the stacked deep ensemble meta-learners perform better than individual deep learning learners and they improve the prediction accuracy. More specifically, we found that the stacked-DNN meta-learner resulted in the best accuracy as well as RMSE and it consumed the least time for training and prediction. Our findings indicated also that it is possible to reuse China data and learners to make prediction of COVID-19 epidemic trend in Tunisia and then across the world when preventive and control measures are similar. In a nutshell, our learned lessons confirmed that (i) daily COVID-19 confirmed cases time series of countries, where preventive and control measures are comparable, can be fused to obtain more informative and sufficient data, (ii) deep learning models (LSTM, DNN and CNN) can be integrated in a unified stacked ensemble learning method to accurately forecast COVID-19 confirmed cases and (iii) transfer learning can be adopted to reuse models trained from one country data to predict COVID-19 trends for other countries and to reduce run-time costs.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Who – novel coronavirus – China, who, http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/, (Accessed 13 May 2020).
[2] M. Islam, et al., A systematic review on healthcare analytics: application and theoretical perspective of data mining, Healthcare 6 (2) (2018) 54, https://doi.org/10.3390/healthcare6020054.
[3] D. Ravi, et al., Deep learning for health informatics, IEEE J. Biomed. Health Inform. 21 (1) (2017) 4–21, https://doi.org/10.1109/JBHI.2016.2636655.
[4] R. Miotto, et al., Deep learning for healthcare: review, opportunities and challenges, Brief. Bioinform. 19 (6) (2018) 1236–1246, https://doi.org/10.1093/bib/bby044.
[5] T. Meng, et al., A survey on machine learning for data fusion, Inf. Fusion 57 (2020) 115–129, https://doi.org/10.1016/j.inffus.2019.12.001.
[6] M. Li, et al., Predicting the epidemic trend of Covid-19 in China and across the world using the machine learning approach, medRxiv (2020) 1–20, https://doi.org/10.1101/2020.03.18.20031817, doi.
[7] S.F. Ardabili, et al., Covid-19 outbreak prediction with machine learning, medRxiv, https://doi.org/10.1101/2020.04.17.20070094.
[8] S. Bandyopadhyay, S. Dutta, Machine learning approach for confirmation of Covid-19 cases: positive, negative, death and release, medRxiv (2020) 1–10, https://doi.org/10.1101/2020.03.25.20043505.

[9] C.J. Huang, et al., Multiple-input deep convolutional neural network model for Covid-19 forecasting in China, medRxiv, https://doi.org/10.1101/2020.03.23.20041608.

[10] P.H. Kumar, B.K. Panigrahi, Prediction and analysis of Covid-19 positive cases using deep learning models: a descriptive case study of India, Chaos Solitons Fractals 139 (2020) 110017.

[11] A. Zeroual, et al., Deep learning methods for forecasting Covid-19 time-series data: a comparative study, Chaos Solitons Fractals 140 (2020) 110121.

[12] A. Hernandez-Matamoros, et al., Forecasting of covid19 per regions using arima models and polynomial functions, Appl. Soft Comput. J. 96 (2020) 106610, http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/.

[13] S. Chae, et al., Predicting infectious disease using deep learning and big data, Int. J. Environ. Res. Public Health 15 (8) (2018) 1508–1596, https://doi.org/10.3390/ijerph15081596.

[14] R. Pascual, et al., How to construct deep recurrent neural networks, arXiv: 1312.6026.

[15] S.R. Venna, et al., A novel data-driven model for real-time influenza forecasting, IEEE Access 7 (2019) 7691–7701, https://doi.org/10.1109/ACCESS.2018.2888585.

[16] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507, https://doi.org/10.1126/science.1127647.

[17] F.A.B. Hamzah, et al., Coronavirus: worldwide Covid-19 outbreak data analysis and prediction, Bull. World Health Organ. (2020), https://doi.org/10.2471/BILT.20.255655.

[18] J. Gu, et al., Recent advances in convolutional neural networks, Pattern Recognit. 77 (2018) 354–377, https://doi.org/10.1016/j.patcog.2017.10.013.

[19] GitHub, Csegsisanddata/Covid-19: novel coronavirus (Covid-19) cases provided by Jhu csse, https://github.com/CSEGISandData/COVID-19. (Accessed 30 November 2020).

[20] B. Jiménez, et al., Finding optimal model parameters by deterministic and annealed focused grid search, Neurocomputing 72 (13) (2009) 2824–2832, https://doi.org/10.1016/j.neucom.2008.09.024.

[21] FJ. Pontes, et al., Design of experiments and focused grid search for neural network parameter optimization, Neurocomputing 186 (2016) 22–34, https://doi.org/10.1016/j.neucom.2015.12.061.

[22] R.J. Frank, et al., Time series prediction and neural networks, J. Intell. Robot. Syst. 31 (2001) 91–103, https://doi.org/10.1023/A:1012074215150.

[23] M. Gracyk, et al., Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal, in: Intelligent Information and Database Systems, Springer, Berlin, Heidelberg, 2010, pp. 340–350.

[24] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.

[25] D. Saso, B. Zenko, Is combining classifiers with stacking better than selecting the best one?, Mach. Learn. 54 (2004) 255–273.

[26] P. Runeson, M. Höst, Guidelines for conducting and reporting case study research in software engineering, Empir. Softw. Eng. 14 (2) (2009) 131–164, https://doi.org/10.1007/s10664-008-9102-8.