



ISSN: 0067-2904

COVID-19 Detection via Blood Tests using an Automated Machine Learning Tool (Auto-Sklearn)

Noor Maher Moatz*, Suhad A.Yousif

Department of Computer Science, Al-Nahrain University, Baghdad, Iraq

Received: 3/10/2022

Accepted: 1/1/2023

Published: 30/11/2023

Abstract

Widespread COVID-19 infections have sparked global attempts to contain the virus and eradicate it. Most researchers utilize machine learning (ML) algorithms to predict this virus. However, researchers face challenges, such as selecting the appropriate parameters and the best algorithm to achieve an accurate prediction. Therefore, an expert data scientist is needed. To overcome the need for data scientists and because some researchers have limited professionalism in data analysis, this study concerns developing a COVID-19 detection system using automated ML (AutoML) tools to detect infected patients. A blood test dataset that has 111 variables and 5644 cases was used. The model is built with three experiments using Python's Auto-Sklearn tool. First, an analysis of the Auto-Sklearn process is done by studying the impact of several learning settings and parameters on the COVID-19 dataset using different classification methods, namely meta-learning, ensemble learning, and a combination of ensemble learning and meta-learning. The results show that using Auto-Sklearn with a meta-learning and ensemble learning parameter model predicts the patients infected with COVID-19 with high accuracy, reaching 96%. Furthermore, the best algorithm selected is the Random Forest Classifier (RF), which outperforms other classification methods. Finally, AutoML can assist those new to data sciences or programming skills in selecting the appropriate algorithm and hyperparameters and reducing the number of steps required to achieve the best results.

Keywords. COVID-19, AutoML, Auto-Sklearn, Bayesian optimization, meta-learning, ensemble Learning.

الكشف عن Covid-19 عن طريق اختبارات الدم باستعمال أداة التعلم الآلي الذاتية (Auto-Sklearn)

نور ماهر معتز*، سهاد عبدالرحمن يوسف

قسم علوم الحاسوب، كلية العلوم، جامعة النهرين، بغداد، العراق

الخلاصة

أشعل انتشار عدوى COVID-19 محاولات عالمية لاحتواء الفيروس والقضاء عليه بشكل مثالي. يستعمل معظم الباحثين خوارزميات التعلم الآلي (ML) للتنبؤ بالاصابة من عدمها بهذا الفيروس. ومع ذلك، يواجه الباحثون تحديات كثيرة، مثل اختيار المعاملات المناسبة وأفضل خوارزمية لتحقيق التنبؤ بصورة دقيقة. لذلك، هناك حاجة إلى علماء البيانات. لغرض تقليل الاعتماد على علماء البيانات، تهتم هذه الدراسة بتطوير

*Email: noor.maher.cs2020@ced.nahrainuniv.edu

نظام الكشف عن COVID-19 باستخدام أدوات التعلم الآلي (AutoML) للكشف عن المرضى المصابين بناءً على اختبارات الدم الخاصة بهم المتكونه من 111 عمود و 5644 سطر. تم تطوير هذا النموذج باستخدام أداة Auto-Sklearn المتوفرة في Python. يتم إجراء تحليل لعملية التعلم التلقائي من خلال دراسة تأثير العديد من إعدادات ومعاملات التعلم على مجموعة بيانات COVID-19 باستخدام طرق تصنيف مختلفة ، وهي **meta-learning, ensemble learning** ومزيج من **meta-learning** و **ensemble learning**. أظهرت النتائج أن استعمال Auto-Sklearn مع دمج معلمة Meta-Learning و **ensemble learning** يتنبأ بالمرضى المصابين بـ COVID-19 بدقة عالية تصل إلى 96 % وأن أفضل خوارزمية مختارة هي (RF) الذي يتفوق على الطرق الأخرى المقترحة. أخيرًا ، يمكن لـ AutoML مساعدة المبتدئين في علوم البيانات أو مهارات البرمجة في اختيار الخوارزمية والمعاملات المناسبة وتقليل عدد الخطوات المطلوبة لتحقيق أفضل النتائج.

1. Introduction

On March 11th, 2020, the COVID-19 virus was classified as a pandemic by the World Health Organization (WHO) [1]. Many people who have the novel SARS-COVID-19 symptoms can be difficult to distinguish from those of other common diseases [2] and [3]. Therefore, reverse transcription polymerase chain reaction (RT-PCR) is still extensively used in COVID-19 screening [4]. Moreover, the missed detection rates for RT-PCR range between 15% and 20%. Additionally, the time it takes to get test results after collecting patient samples can be many hours or even days. Therefore, a quicker and more precise COVID-19 detection method is required [5]. ML algorithms helped find this pandemic because of their ability, and they were used in many studies to classify patients into those who were infected or not [6], [7], and [8]. However, ML requires several steps to make a good model, such as choosing the best preprocessing steps, tuning the hyperparameters, and choosing the suitable algorithm [9] and [10]. Furthermore, because most healthcare professionals lack sufficient programming experience, AutoML solutions help build and enhance ML pipelines [11] and [12]. Furthermore, for AutoML, numerous frameworks are available [13] to tackle the above difficulties.

This study focuses explicitly on Auto-Sklearn, built upon Scikit-Learn, a well-known Python ML library. Sequential Model-based Algorithm Configuration (SMAC) is a Bayesian optimization method used by Auto-Sklearn to solve algorithm selection and hyperparameter optimization problems. The standard SMAC optimization integrates both meta-learning and ensemble techniques [14]. Effective AutoML pipelines are made up of preprocessing steps and ML classifiers, chosen by using Auto-Sklearn, which employs meta-learning, Bayesian optimization, and ensemble selection [11-13]. In these studies, the authors used a dataset of blood tests to predict COVID-19. However, their research has some limitations, like using a small data set or having low accuracy. In order to address the above restrictions and attempt to identify the COVID-19 virus using a blood test dataset. The purpose of this paper is to present a model for early detection of this virus using one of the AutoML tools (Auto-Sklearn) and to try to find answers to the following questions:

- Is it feasible to use AutoML to achieve higher accuracy than traditional ML?
- Can AutoML replace the position of data scientists?

The structure of this paper is organized as follows: Related works are presented in Section 2. The definitions are described in Section 3. The methodology is introduced in Section 4, and the results and discussion are presented in Section 5. Finally, Section 6 illustrates the conclusion.

2. Related work

Using blood tests to diagnose COVID-19 is uncommon; however, some studies have covered similar ideas using the same dataset. In this paper, some of these studies will be mentioned. The authors in [15] proposed an “AutoML” method for classifying raw spectra data generated by the well-established UV-VIS spectroscopy methodology. They compared the performance of four AutoML models, TPOT, Auto-Sklearn, Hyperopt-Sklearn, and Auto-Keras, to determine the best model for sickle cell disease and sickle cell trait (SCT/SCD) screening using spectroscopy data. The proposed approach has 100% sensitivity and 93.84% specificity for detecting sickle hemoglobin. The work shows how the ML-based absorption spectroscopy test could be used in screening programs for large groups of people in places with limited resources.

The study in [16] suggested logistic regression (LR) and six ML algorithms (Adaboost, Stochastic Gradient Boosting, Support Vector Machine (SVM), Multinomial Naive Bayes (MNB), Decision Tree (DT), and RF) for COVID-19 detection and classification, utilizing 212 clinical reports. LR performed well in the simulations, achieving 96 percent precision, 96 percent sensitivity, 96.20 percent accuracy, and a 95 percent F1 score.

The study [17] used six classifiers (Bayesian Networks (BN)), LR, a lazy-classifier (IBk), a meta-classifier (Classification via Regression (CR)), a rule-learner (PART), and a decision tree (J48) based on 14 clinical variables to build six prediction models for COVID-19 diagnosis. The study reviewed 114 instances from the Chinese province of Zhejiang's Taizhou hospital. The findings indicated that, with an accuracy of 84.21 percent, the CR meta-classifier is the most reliable classifier for predicting positive and negative COVID-19 cases. The findings may aid in the early identification of COVID-19. The limitations of this study are as follows: first, it was challenging to gather larger samples during the pandemic, and the sample size is relatively small. Second, it is too bad that this study did not look at COVID-19 symptoms to predict viral infection because there was insufficient data.

The authors in [18] used a variety of ML classifiers, including Multilayer Layer Perceptron (MLP), SVM, Random Tree (RT), RF, BN, and Naïve Base (NB). They developed a low-cost COVID-19 identification system using standard blood samples. The Brazilian Albert Einstein Hospital provided the authors with a dataset that included 5644 data samples and 559 confirmed COVID-19 cases. Because there were so few positive samples in the dataset, SMOTE oversampling was employed to enhance their models' performance. The results showed that the classification was done well, with an overall accuracy of 95.159 percent. The sensitivity was 96.8 percent, the precision was 93.8 percent, and the specificity was 93.6 percent. The results of the experiments showed that BN performed better than other models. In [19], the study addresses the issue of COVID-19 screening in regular blood tests and attempts to give intelligible solutions based on ML approaches. They evaluated various ML classifiers using a public dataset from the Albert Einstein Hospital. The RF classifier obtained the best performance (accuracy 0.88, specificity 0.91, sensitivity 0.66, F1-score 0.76, and AUROC 0.86).

The study in [20] aimed to use and evaluate ML-based models to predict COVID-19 diagnosis and disease severity. Using an open data set from the Albert Einstein Hospital in So Paulo, Brazil, all of the ML-based models artificial neural networks (ANN), partial least squares discriminant analysis (PLS-DA), DT, and k-nearest neighbor algorithm (k-NN) were able to accurately predict the diagnosis of COVID-19 and the severity of the disease with a rate of accuracy above 84 percent, which is comparable to the results of RT-PCR and the

minimal threshold recommended for diagnostic procedures. The ANN model did the best (94 to 98 percent), meaning that healthcare professionals in the field could use it to help them make decisions.

In this study [21], four deep learning techniques—Deep Neural Network (DNN), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM)—along with seven ML techniques, including LR, K-NN, DT, SVM, NB, RF, and Extremely Randomized Trees (ET), were presented to diagnose positive cases of COVID-19 from three routine laboratory blood test datasets. The suggested (DNN) model had the most outstanding accuracy scores across all three datasets. In the first dataset, an accuracy of 92.11% was attained; in the second dataset, an accuracy of 93.16% was attained. Finally, the accuracy value for the third dataset was 92.5%.

[22]. Auto-Sklearn and XGBoost are advanced models used in this paper. The dataset used for cardiovascular disease contains data from 70,000 patients. The results of the XGBoost model are 73% accurate. At the same time, the outcome of the Auto-sklearn model's accuracy reaches 83%.

3. Background

This section will provide a general overview of Auto-Sklearn, its optimization technique (the Bayesian optimization technique) for choosing the best model, the best feature selection method, and finally, the best hyperparameters for tuning the model. Also, a general idea of meta-learning and ensemble learning will be clarified.

3.1 Auto-Sklearn Model Overview

Recent developments in Bayesian optimization, meta-learning, and ensemble learning are utilized in selecting algorithms and hyperparameters in Auto-Sklearn [23]. SMAC as a Bayesian optimization technique [24]. The two main components of the iterative Bayesian optimization algorithm are a probabilistic surrogate model and an acquisition function that determines which point to evaluate next [25]. The surrogate model is fitted to all observations of the target function made so far in each iteration. The acquisition function uses the probabilistic model's prediction distribution to determine the utility of various candidate points. The acquisition function is cheap to compute and may thus be fully optimized instead of assessing the pricey black box function. Although there are many acquisition functions, the expected improvement (EI) is frequently used since it can be calculated in the closed form if the model prediction y at configuration λ follows a normal distribution.

$$E[I(\lambda)] = (f_{min} - \mu(\lambda))\varphi\left(\frac{f_{min}-\mu(\lambda)}{\sigma}\right) + \sigma\varphi\left(\frac{f_{min}-\mu(\lambda)}{\sigma}\right) \quad (1)$$

f_{min} It is the best-observed value so far, (φ) is the standard normal density, and (Φ) is the standard normal distribution function.

Auto-Sklearn includes 4 data preparation methods, 14 preprocessing techniques, and 15 classification algorithms. Each was parameterized, creating a space of 110 hyperparameters. Most are conditional hyperparameters, meaning they only take effect when the appropriate component is chosen. We point out that SMAC has native support for this conditionality [26]. One of its key benefits is that Auto-Sklearn provides a variety of execution strategies. For instance, the AutoML optimization process uses just SMAC optimization techniques in its basic vanilla version (Auto-Sklearn-v). However, Auto-Sklearn also gives the end-users the option to enable or disable the various optimization options, such as the use of meta-learning (Auto-Sklearn-m) and ensemble learning (Auto-Sklearn-e), in addition to the complete

version (Auto-Sklearn), where all options are enabled. Figure 1 shows the mechanism of the Auto-SKlearn process.

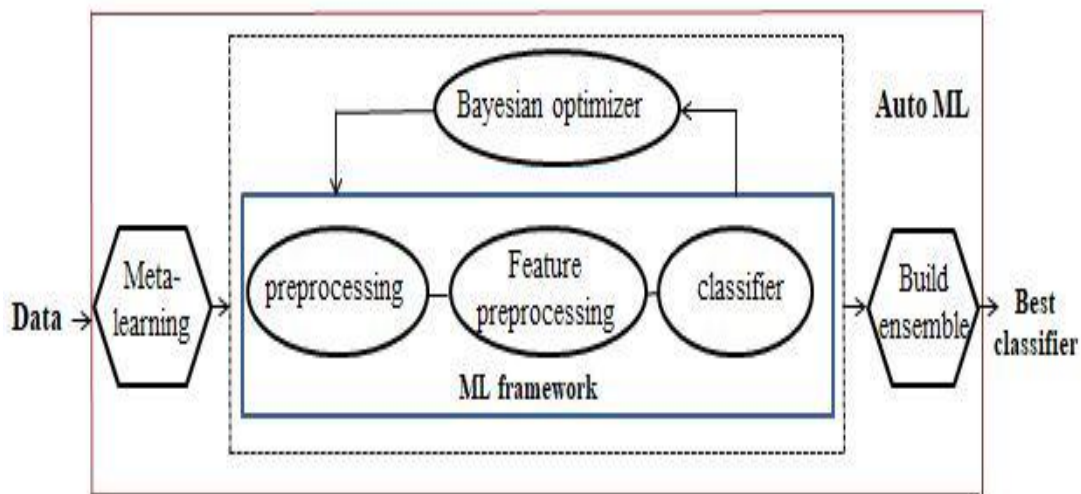


Figure 1: The Auto-SKlearn process[27].

3.1.1 Meta-learning

Meta-learning is the practice of using knowledge of the performance of learning methods to enhance the efficiency (or choice) of learning algorithms. As a result, building AutoML systems does not need to start from scratch because promising configurations are created for previously unexplored ML tasks, allowing for faster convergence and fewer trial-and-error iterations [28]. The real benefit of meta-learning techniques in AutoML is that they enable the automated replacement of manually created algorithms with data-driven procedures. A knowledge base that stores the meta-features of datasets and the top-performing pipelines on these datasets serves as the foundation for the meta-learning technique used by Auto-Sklearn. Thirty-eight meta-features from information theory and statistics are performed. For each dataset in their repository (140 datasets from the Open ML repository), the meta-features and the empirically best-performing pipelines are stored during the offline phase [29]. The system extracts any new dataset's meta-features during the online phase. It searches for the most similar datasets and returns the top k best-performing pipelines on those datasets. These k pipelines serve as a start for the framework's Bayesian optimization algorithm. Any meta-learning mechanism's primary objective is to improve the search process by allowing the optimization strategy to begin with the most promising pipelines rather than random ones.

3.1.2 Ensemble

In order to create a more accurate prediction model, multiple ML-based models trained on the same job are combined through the process of assembly. Various techniques can integrate these base models, such as weighted/straight-forward voting (averaging), bagging, boosting, and other methods [30]. In theory, the main benefit of assembling approaches is that they allow the base models to work together to make more general predictions than if only one base model were used.

3.1.3 Combination

Auto-Sklearn has a unique feature that allows the end-users to enable or disable the various optimization options, such as using meta-learning (Auto-Sklearn-m) and ensembling (Auto-Sklearn-e) together.

4. Proposed Methodology

The proposed methodology includes three main stages, as discussed in Figure 1, and then we applied three different experiments described in detail in Figure 2 and Algorithm 1. Three sections discuss the proposed methodology. The first section will describe the dataset used after downloading and cleaning it. The second section will explain the experiments performed on this dataset using Auto-Sklearn, which will choose the best algorithm based on Bayesian optimization. Finally, Section 3 gives the results and the evaluation of the models.

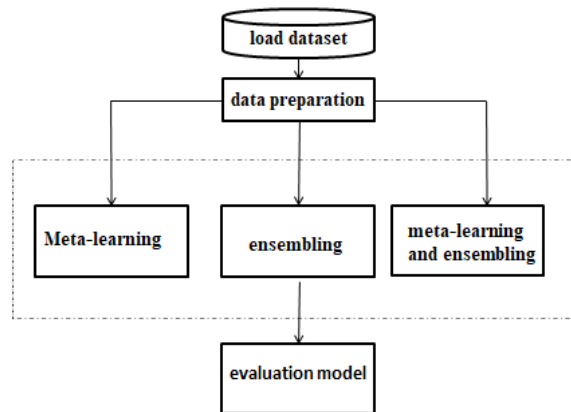


Figure 2: The workflow steps.

4.1 Datasets

In this study, the blood test data was gathered from Brazil's Albert Einstein Hospital. Unfortunately, there are 5644 instances and 111 variables, with the COVID-19 dependent variable being one of them (positive or negative). Because it consists of 558 positive and 5086 negative tests that make the dataset unbalanced [31], we will explain the steps to prepare this data to be entered into the proposed model.

4.2 Auto-Sklearn Model

When providing the input dataset to AutoML, users must set some parameters known as “learning settings.” The parameters used in this study are shown in Table 1. After splitting the dataset into 75% for training and 25% for testing, perform several experiments in Auto-Sklearn as described below.

- the first experiment using Auto-Sklearn-m
- the second experiment using Auto-Sklearn-e
- The last experiment used meta-learning and ensemble learning together in (Auto-Sklearn m, and e).

Table 1: Auto-Sklearn parameter used

Parameters used	Number Used	Definition
meta-learning	25	Initialize the hyperparameter optimization algorithm with many configurations as were successful on prior seen datasets.
Ensembling	50	The number of models in an ensemble determines how many models are chosen from a library.
Time_left_for_this_task	1000	This parameter limits how long it can take to find suitable models in seconds. Auto-sklearn of discovering good models is increased by increasing this parameter.
Per_run_time_limit	10	The maximum amount of time for a single call to the ML model.
The resampling_strategy	holds out.	To handle overfitting, you might need 'cv': cross-validation.

The Auto-Sklearn model algorithm is presented in Algorithm 1.

Algorithm 1: Auto-Sklearn

Input: covid-19 dataset

Output: the best classification model

Begin:

#Load data of covid-19

Step 1: dataset $\leftarrow (x, y)$;

#Preprocessing Data

#filling the empty cell

Step 2: Dropping the dataset's target ['SARS'].

Step 3: Fill in all the NULL values using the mean and delete the empty columns.

#makes dataset number

Step 4: Label encoding involves replacing categorical values with numerical ones.

make dataset balance

Step 5: Balancing the dataset using the RandomOverSampler technique

#use the Auto-Sklearn model

Step 6: result model \leftarrow Auto-Sklearn (parameters);

#train the data

Step 7: classification_model \leftarrow result model.

Fit (x_train ,y_train);

Print the best classification model with the best parameter

#Check the model's performance by calculating the accuracy

Step 8: accuracy \leftarrow TP+TN/all samples.

End

4.3 Evaluation Metric

This paper used a confusion matrix to evaluate the model. Several evaluation metrics can be computed based on the confusion matrix, such as accuracy, recall, precision, and F-measure. The equation for them is mentioned in equations [2, 3, 4, 5]. This is based on the values:

- True Negative (TN) and True Positive (TP): The model adequately predicts the class (negative or positive).
- False Negative (FN) and False Positive (FP): The model wrongly predicts the class (switch negative and positive classes).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$F - \text{measure} = \frac{2 \times Precision \times Recall}{precision + Recall} \quad (5)$$

5. Results and discussion

This study uses Auto-Sklearn with different parameters in three experiences to detect if the patient suffers from COVID-19 and compares their results. After making some adjustments, as mentioned earlier. Auto-Sklearn (m,e) performed exceptionally well in classification and achieved the best accuracy.

5.1 Auto-sklearn results

In this paper, several Auto-Sklearn experiments were conducted on the dataset using three parameters. In each experiment, one of these three parameters is used. The classification accuracy of the pipelines was evaluated using the confusion matrix. Moreover, other metrics are used to evaluate the model, such as accuracy, recall, precision, and F-measure. Figure 3 shows the confusion matrix of the Auto-Sklearn experiments. When the results of these experiments were compared, the most accurate experiment was (Auto-Sklearn m.e.), with an accuracy of 96%, a recall of 94%, a precision of 99%, and an F-score of 97%. Figure 3 (Auto-Sklearn m,e) illustrates the confusion matrix of this experiment, where 1,238 true positives have been accurately predicted and the false positive number is 10. Hence, the number of positive cases in which the model predicts a mistake is 1,219, and of the 1,219 true negative cases, only 76 were false negatives.

In an experiment (Auto-Sklearn-e), this experiment achieved 94% accuracy, recall 90%, precision 99%, and F-score 94%. Figure (Auto-Sklearn-e) shows the confusion matrix of this experiment, where 1,235 true positive COVID-19 cases have been accurately predicted, and the false positive number is 13, with 1,170 true negative cases and 125 false negatives. Finally, the experiment (Auto-Sklearn-m) achieved the lowest accuracy (93%), recall (88%), precision (99%), and F-score (94%). Figure (Auto-Sklearn m) shows the confusion matrix that contained 1,240 true positive COVID-19 cases that have been accurately predicted out of 1,248 positive cases and 1,144 true negative cases out of 1,295 negative cases. All these results are shown in Table 2.

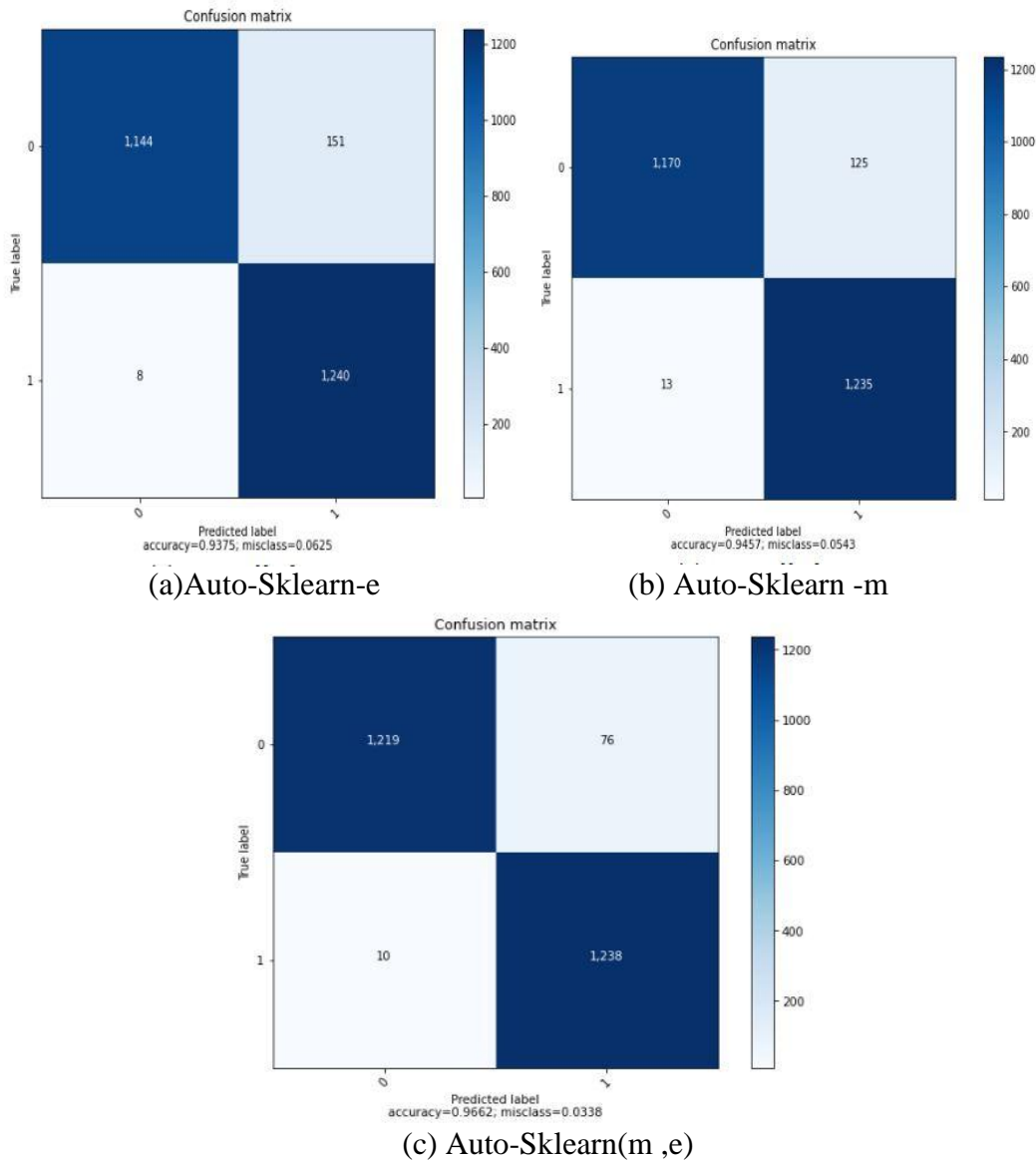


Figure 3: The confusion matrix of Auto-Sklearn experiments.

Table 2: The result of all experiments

	Auto-Sklearn-m	Auto-Sklearn-e	Auto-Sklearn(m,e)
Accuracy	93%	94%	96%
Recall	88%	90%	94%
Precision	99%	99%	99%
F-1	94%	94%	97%

5.2 Comparisons with previous work

In this section, the Auto-Sklearn model is evaluated by comparison with another classification model from earlier studies [7] [18] [19] [20]. All these previous studies used the same dataset [31]. Furthermore, these studies used traditional ML. Moreover, Auto-Sklearn outperforms traditional ML methods due to the comparisons illustrated in Table 3.

Table 3: comparison of the proposed method with previous studies

Study	Method	Accuracy
[7]	AdaBoost	85%
[18]	Multilayer perceptron	95%
[19]	RF	88%
[20]	ANN	84%
Proposed method	Auto-Sklearn-(m,e)	96%

Thus, AutoML will shorten many of the steps of ML and reduce the trials and errors that the researcher needs to find the best results in the fewest steps based on Bayesian optimization for choosing the best parameters.

6. Conclusion & Future work

In this study, COVID-19 disease was identified from blood testing using Auto-Sklearn after data preparation, dataset balancing, and input into the Auto-Sklearn model. Three separate Auto-Sklearn experiments were performed, but the Auto-Sklearn (m,e) achieved the best result Among the three experiments. These results were compared to earlier research that used several traditional ML methods. We answered most of the questions posed in this paper and concluded that AutoML is superior to traditional ML. It shortens most programming steps, making it a valuable tool for beginners and those without sufficient programming experience. However, AutoML will only replace data scientists sometime soon. It is a tool to help data scientists and is a great way to clarify this complex field for non-experts so they can learn from the ML experience. Also, in some situations, there is no need for an expert data scientist to handle the hyperparameter tuning and choose the best preprocessing steps and feature engineering to achieve higher accuracy. The future work focuses on using a large and real dataset from the Iraqi hospital dataset and deep learning or other AutoML tools to get the highest accuracy.

References

- [1] N. Alballa and I. Al-Turaiki, "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: a review," *Informatics in Medicine Unlocked*, vol. 24, p. 100564, 2021. <https://doi.org/10.1016/j.imu.2021.100564>.
- [2] A. Banerjee *et al.*, "Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population," *International immunopharmacology*, vol. 86, p. 106705, 2020. <https://doi.org/10.1016/j.intimp.2020.106705>.
- [3] E. A. Al-Zubaidi and M. M. Mijwil, "Medical Image Classification for Coronavirus Disease (COVID-19) Using Convolutional Neural Networks," *Iraqi Journal of Science*, pp. 2740-2747, 2021. DOI: 10.24996/ij.s.2021.62.8.27.
- [4] R. Barza, P. Patel, L. Sabatini, and K. Singh, "Use of a simplified sample processing step without RNA extraction for direct SARS-CoV-2 RT-PCR detection," *Journal of Clinical Virology*, vol. 132, p. 104587, 2020. doi: 10.1016/j.jcv.2020.104587.
- [5] J. Wu, J. Shen, M. Xu, and M. Shao, "A novel combined dynamic ensemble selection model for imbalanced data to detect COVID-19 from complete blood count," *Computer Methods and Programs in Biomedicine*, vol. 211, p. 106444, 2021. <https://doi.org/10.1016/j.cmpb.2021.106444>.
- [6] J. Wu *et al.*, "Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results," *MedRxiv*, 2020. doi: <https://doi.org/10.1101/2020.04.02.20051136>.
- [7] M. Almansoor and N. M. Hewahi, "Exploring the Relation between Blood Tests and Covid-19 Using Machine Learning," in *2020 International Conference on Data Analytics for Business and*

- Industry: Way Towards a Sustainable Economy (ICDABI)*, 2020: IEEE, pp. 1-6. DOI: 10.1109/ICDABI51230.2020.9325673.
- [8] M. Kukar *et al.*, "COVID-19 diagnosis by routine blood tests using machine learning," *Scientific reports*, vol. 11, no. 1, pp. 1-9, 2021. <https://doi.org/10.1038/s41598-021-90265-9>.
- [9] M. M. Mijwil, "Implementation of Machine Learning Techniques for the Classification of Lung X-Ray Images Used to Detect COVID-19 in Humans," *Iraqi Journal of Science*, pp. 2099-2109, 2021. DOI: 10.24996/ij.s.2021.62.6.35.
- [10] J. C. Xavier-Júnior, A. A. Freitas, T. B. Ludermir, A. Feitosa-Neto, and C. A. Barreto, "An evolutionary algorithm for automated machine learning focusing on classifier ensembles: An improved algorithm and extended results," *Theoretical Computer Science*, vol. 805, pp. 1-18, 2020. <https://doi.org/10.1016/j.tcs.2019.12.002>.
- [11] S. A. Yousif, V. W. Samawi, and N. M. Al-Saidi, "Automatic Machine Learning Classification Algorithms for Stability Detection of Smart Grid," in *2022 IEEE 5th International Conference on Big Data and Artificial Intelligence (BD AI)*, 2022: IEEE, pp. 34-39. DOI: 10.1109/BD AI56143.2022.9862710.
- [12] V. W. Samawi, S. A. Yousif, and N. M. Al-Saidi, "Intrusion Detection System: An Automatic Machine Learning Algorithms Using Auto-WEKA," in *2022 IEEE 13th Control and System Graduate Research Colloquium (ICSGRC)*, 2022: IEEE, pp. 42-46. DOI: 10.1109/ICSGRC55096.2022.9845166.
- [13] K. Chauhan *et al.*, "Automated machine learning: The new wave of machine learning," in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2020: IEEE, pp. 205-212. 10.1109/ICIMIA48430.2020.9074859.
- [14] H. Eldeeb, S. Amashukeli, and R. E. Shawi, "An Empirical Analysis of Integrating Feature Extraction to Automated Machine Learning Pipeline," in *International Conference on Pattern Recognition*, 2021: Springer, pp. 336-344.
- [15] S. Srivastava, R. Srinivasan, N. K. Nambison, and S. S. Gorthi, "Diagnosis of sickle cell anemia using AutoML on UV-Vis absorbance spectroscopy data," *arXiv preprint arXiv:2111.12711*, 2021. <https://doi.org/10.48550/arXiv.2111.12711>.
- [16] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. Mohi Ud Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," *International Journal of Information Technology*, vol. 12, no. 3, pp. 731-739, 2020. <https://doi.org/10.1007/s41870-020-00495-9>.
- [17] I. Arpaci, S. Huang, M. Al-Emran, M. N. Al-Kabi, and M. Peng, "Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11943-11957, 2021. <https://doi.org/10.1007/s11042-020-10340-7>.
- [18] V. A. de Freitas Barbosa *et al.*, "Heg. IA: an intelligent system to support diagnosis of Covid-19 based on blood tests," *Research on Biomedical Engineering*, vol. 38, no. 1, pp. 99-116, 2022. <https://doi.org/10.1007/s42600-020-00112-5>.
- [19] M. A. Alves *et al.*, "Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs," *Computers in Biology and Medicine*, vol. 132, p. 104335, 2021. <https://doi.org/10.1016/j.compbiomed.2021.104335>.
- [20] A. de Fátima Cobre *et al.*, "Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators?," *Computers in biology and medicine*, vol. 134, p. 104531, 2021. <https://doi.org/10.1016/j.compbiomed.2021.104531>.
- [21] S. B. Rikan, A. S. Azar, A. Ghafari, J. B. Mohasefi, and H. Pirnejad, "COVID-19 diagnosis from routine blood tests using artificial intelligence techniques," *Biomedical Signal Processing and Control*, vol. 72, p. 103263, 2022. <https://doi.org/10.1016/j.bspc.2021.103263>.
- [22] M. Fayez and S. Kurnaz, "Novel method for diagnosis diseases using advanced high-performance machine learning system," *Applied Nanoscience*, pp. 1-7, 2021. <https://doi.org/10.1007/s13204-021-01990-6>.
- [23] L. Ferreira, A. Pilastri, C. M. Martins, P. M. Pires, and P. Cortez, "A comparison of AutoML tools for machine learning, deep learning and XGBoost," in *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021: IEEE, pp. 1-8. <https://doi.org/10.1109/IJCNN52387.2021.9534091>.

- [24] V. Werner, L. Maingault, and M.-L. Potet, "Fast Calibration of Fault Injection Equipment with Hyperparameter Optimization Techniques," in *International Conference on Smart Card Research and Advanced Applications*, 2021: Springer, pp. 121-138.
- [25] S. Shekhar, A. Bansode, and A. Salim, "A Comparative study of Hyper-Parameter Optimization Tools," in *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 2021: IEEE, pp. 1-6. <https://doi.org/10.1109/CSDE53843.2021.9718485>.
- [26] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," *Artificial intelligence in medicine*, vol. 104, p. 101822, 2020. <https://doi.org/10.1016/j.artmed.2020.101822>.
- [27] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019. DOI 10.1007/978-3-030-05318-5.
- [28] M. Bahri, F. Salutari, A. Putina, and M. Sozio, "AutoML: state of the art with a focus on anomaly detection, challenges, and research directions," *International Journal of Data Science and Analytics*, pp. 1-14, 2022. <https://doi.org/10.1007/s41060-022-00309-0>.
- [29] M. Feurer, K. Eggensperger, S. Falkner, M. Lindauer, and F. Hutter, "Auto-sklearn 2.0: Hands-free automl via meta-learning," *arXiv preprint arXiv:2007.04074*, 2020. <https://doi.org/10.48550/arXiv.2007.04074>.
- [30] S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Quasim, and G. R. Sinha, "An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach," *IEEE Access*, vol. 9, pp. 67488-67500, 2021. <https://doi.org/10.1109/ACCESS.2021.3075140>.
- [31] E. Data4u, "Diagnosis of COVID-19 and its clinical spectrum," retrieves from <https://www.kaggle.com/einsteindata4u/covid19>, 2020.