JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

Effective Diagnosing of Covid-19 from CXR Images Using Deep Learning Approaches and Optimized XG Boost Model

Sadeer Alaa Thamer[1,a)], Dr. Mshari A Alshmmri[2,b)]

[1,2]Department of computer science, computer science and mathematics college, Tikrit University, Iraq
[a)]sadeer.a.thamer35536@st.tu.edu.iq
[b)]Dr.mshary.alshmmry@tu.edu.iq

Abstract
Since the first confirmed incidence of the novel coronavirus Covid-19 in China, it has spread fast around the world, reaching a population of 442,602,593(at the start of 2022), according to World Health Organization figures.Therefore, the diagnosis of the virus is crucial to prevent its separation. However, the tools available for Covid-19 diagnosis are limited compared to the pressure at the increasing number of infected people. Therefore, to prevent the virus thread, it is necessary to find a quick automated system that can handle a bulk amount of data with high accuracy and a lower amount of false positive or false negative. This research presents a hybrid machine learning-based system that uses a pre-trained MobilNet model for feature extraction from chest X-ray images, followed by a dimensionality reduction technique to speed up the classification process and an XGBoost classifier to complete the task.
Furthermore, the Bayesian algorithm is used to choose the optimum hyperparameters for the XGBoost classifier. The suggested approach was evaluated on two datasets of X-ray images and produced both high and near results.The results for the first dataset were 97.65% accuracy, 97.63% F1-score, 97.65% recall, and 97.69% precision,and for the second dataset were 96.35% accuracy, 95.82% F1-score, 98.35% recall, and 96.38 precision.
Keywords: Covid-19, Machine Learning, Chest X-ray, MobilNet, Dimensionality Reduction, XGBoost, Bayesian Optimization.

1. Introduction

Covid-19 is a novel type of Coronaviruses (CoV), a group of viruses associated with the lungs that include the middle east respiratory syndrome (MERS) and severe acute respiratory syndrome (SARS)[1]. The first Covid-19 case was discovered in China in December 2019, and ever since, Covid-19 has aggressively spread over the world, posing a severe worldwide threat to millions of people.[2]. However, the respiratory disease caused by Covid-19 may be recovered for most people without the need for special treatment, while people with weak immune systems (like elders, people with chronic, other respiratory diseases, or cancer) are at high risk [3]. As of Mar 2022, There are 442,602,593 Covid-19 cases in totalglobally, with 6,003,695 deaths[4]. Till now, the whole world is struggling with Covid-19,and the number of people dying from pneumonia caused by Covid-19 is rising every day[5]. On the other hand, early diagnosis can effectively reduce Covid-19 spreading[6].

X-ray is widely used worldwide for pneumonia diagnosis and many other lung abnormalities due to its cheapness, speed, and non-invasive nature [7].However, an expert is required to diagnose the X-ray images correctly [8].Furthermore, even in normal circumstances, specialized doctors' number is lower than the number of general practitioners, and the need for radiologists is growing as the Covid-19 epidemic and the number of people positive cases grows.

Deep learning (DL) has recently been used to develop computer-aided diagnostic (CAD) systems to aid in the identification of various illnesses and abnormalities.such as breast cancer[9], melanoma [10], lung cancer [11], and skin cancer [12]. In the same way, it is critical to developing a rapid and accurate CAD system that can identify Covid-19 and pneumonia using chest X-ray images to prevent the health system from collapsing and limit the number of infections among healthcare personnel.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

Due to the high ability of DL in feature extraction and handling a massive amount of data, it is highly used in building CAD systems in the medical field [13, 14]. Recently, DL-based CADs have gained researchers' interest in reducing physicians' workload and providing reliable and high-quality analysis [9]. However, the process of reading and analyzing X-ray images requires a long time and a massive amount of data for training the model [7]. This article describes a novel hybrid end-to-end DL-based network called CovidConvXGB to identify Covid-19 and pneumonia using X-ray images.The suggested network includes three steps: the first step includes using the MobileNetV2 model [15] for feature extraction; choosing this model is due to its high speed and efficiency, making it suitable for working in low computational power devices.

Furthermore, since the MobileNetV2 is pre-trained using the ImageNet dataset [16] in this work, there is no need to train the model, and the MobileNetV2 is used for feature extraction directly using the transfer learning technique. The second step is using Linear Discriminant Analysis (LDA) [17]as a feature reduction technique in order to speed up the third step, which is using the extreme gradient boosting (XGBoost) [18] classifier for the final prediction.XGBoostuses the ensembling learning method and is widely used in machine learning due to its robustness and high performance since it ensembles many weak classifiers to produce a strong one and can handle even imbalanced datasets efficiently. However, the XGBoost contains many hyperparameters, and to achieve high results, these hyperparameters need to be optimized. For this purpose, Bayesian optimization (BO) is usedto solve the best hyperparameters selecting problem as an optimization problem.

The contribution of this study can be summarized as follow:

1- The proposed method utilizes transfer learning for feature extraction and XGBoost and BO to categorize X-ray images into normal, pneumonia automatically, and Covid-19 classes.
2- The suggested technique is an end-to-end framework that eliminates the need for time-consuming hand-crafted feature extraction methods that need extensive processing.
3- A set of two datasets were used to evaluate the method, and the results showed that it outperformed current best practices.
4- The proposed model is quick and lightweight, allowing it to be implemented on low-power computers or portable devices, which may subsequently be utilized in hospitals to relieve the load on the health system caused by the increasing number of infections.

The rest of this paper may be broken down into the following sections: Section 2 describes related works, Section 3 examines the methods used to develop the suggested model, Section 4 highlights the experimental findings, Section 5 discusses the acquired results, and Section 6 concludes and offers further study.

2. Related works

The effectiveness of DL-based methods encouraged the researchers to apply themin many challenging problems in the medical field, like skin cancer[12], brain disease [19], breast cancer [9], pneumonia detection [20], and lunge tumor segmentation [21].

[22]were introduce a system utilizing a transfer learning approach to compare the performance of five pre-trained networks (Inception, Xception,VGG19, InceptionResNetV2, and MobileNet) in classifying X-ray images into three categories (normal, Covid-19, and pneumonia). To test their proposed system, they employed two datasets containing 1427 and 224 X-ray images. Their results revealed that Deep learning-based technique's ability to construct an automatic detecting system using X-ray images. In[23],they were suggested a methodologyusing X-ray images from 1078 patients to build a DL-based anomaly detection network. The approach was then tested on 1531 X-ray images and achieved great results, with a specificity of 70.65% and a sensitivity of 96%. The model, though, has a significant proportion of false positives.[24]proposed a strategy for classifying chest computed tomography (CT) images into two categories (infected and normal) using convolutional neural networks (CNNs). Furthermore, they used multi-objective differential evolution to initiate the CNN parameters.[14]suggested a pre-trained DenseNet201 network to classify CT images into infected and normal classes. The obtained results outperformed the

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

results obtained by using pre-trained VGG16 and ResNetV2 models.[25] proposed using U-Net [26]with residual attention mechanism in segmentation of infected area in the chest. For training their model, 110 axial CT images of 60 patients are used.[27]proposed using two-staged DensNet architecture with X-ray and CT images. However, The findings obtained using X-ray images were superior to those obtained with CT images, and the suggested method correctly classified a sample of 22 CT images for people infected with Covid-19.[28] used AOCTNet[29], ShuffleNet[30], and theMobilNet with transfer learning technique for feature extraction then, used a machine learning classifierfor classifying the extracted features. As a result, the Softmax and K-nearest neighbor classifiers achieved the highest results compared with the other classifiers for all feature extracting networks.In[31],CoroNet,an Xception architecture-based model, was presented for exploiting X-ray images to classify Covid-19 and other chest pneumonia disorders.The dataset used was small,with 1351 images, and achieved 87.02% accuracy. [32] introduced the Bayesian DL model with transfer learning to estimate the model's uncertainty while classifying X-ray images. The dataset used consists of 5941 images with four classes and 89% accuracy achieved. [33]introduced COVIDX-Net, a model, based on pre-trained networks like VGG19 and InceptionV3. The collection included 50 X-ray images, 25 of which were verified positive for Covid-19.The best results achieved by their method were 91% for F1-score.

A study of the literature on using X-ray imaging to identify or categorize Covid-19 reveals some limitations that must be considered. First, The majority of prior studies relied on insufficient and limited datasets,which affects its performance when applying real-world data due to its weak generalization ability. Second, most methods use transfer learning techniques in a way that requires re-training the models, which is a time-consuming process, while in our suggested method, A large dataset is used to train the algorithm, and another is used to demonstrate the method's generalizability.Furthermore, in our suggested method, no re-train process is required since the pre-trained model (MobilNetV2) is used for feature extraction only without the need for forward and backward propagation and the process of weights correction. Additionally, our proposed method can be successfully applied for small and big datasets.

3. Materials and methods

3.1. Dataset

For training and testing the method proposed, a dataset which compiled by a team of academics from the University of Qatar in Qatar, Dhaka University in Bangladesh, and its Malaysian and Pakistani partners.The dataset is publiclyavailable for academic use in the Kaggle repository called "Covid-19 radiography dataset" and the Figshare.com website. This dataset has been updated to include 3,616 X-ray images of cases diagnosed with Covid-19, 1,345 images of cases identified with viral pneumonia, and 10,192 images of normal X-rays.Each image is 299x299 pixels in size and the PNG format.For the training purpose, all images were resized into 128x128 pixels and scaled with a 1./255 factor for normalization

.

The chest X-ray images of Covid-19, viral pneumonia, and normal are shown in Figure 1.

Covid-19

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

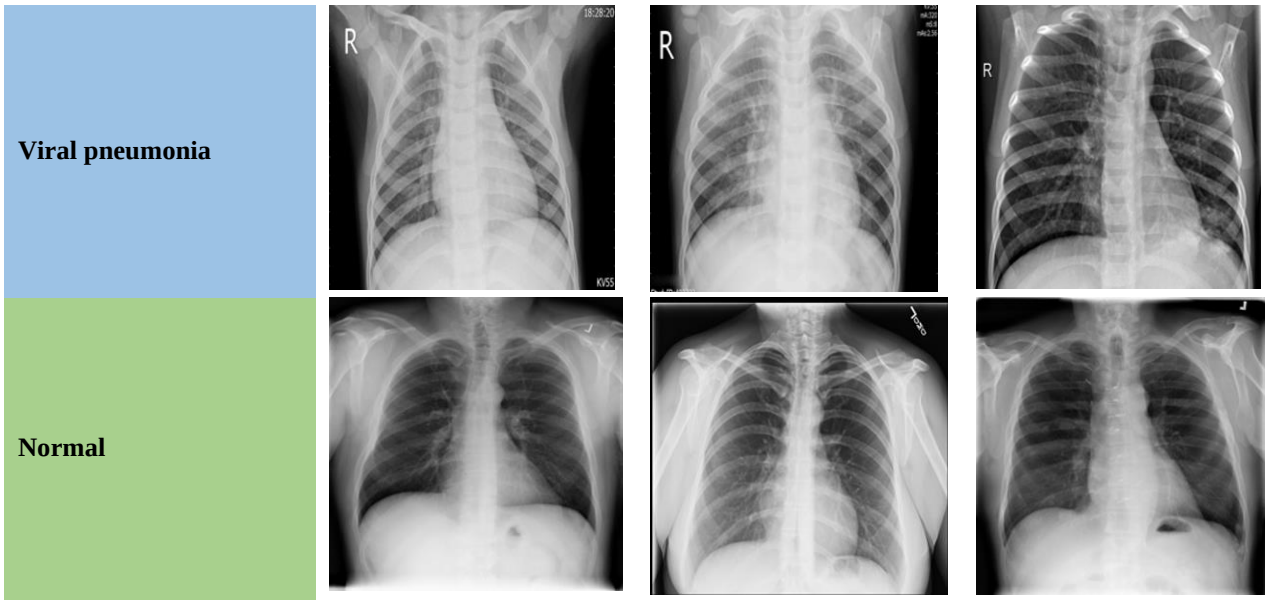

Figure 1: Chest X-ray images of patients with COVID-19 (first row), viral pneumonia (second row), and normal (healthy) individuals (third row).

The second dataset utilized has 6,432 X-ray images (756 infected, 4,273 viral pneumonia, and 1,583 normal). This dataset is collected from different resources,making it suitable for testing the generalization ability of the suggested method. In addition, the dataset is publicly available in the Kaggle repository called "Chest X-ray Covid-19 pneumonia".

3.2. MobilNetV2

MobilNetV2 is presented by Sandler et al.[15] as an efficient and light-weighted network and suitable for low computational power devices. MobilNetV2 outperformed state-of-the-art algorithms for classification, segmentation, and object recognition when utilized as a backbone for feature extraction.Furthermore, it is faster than the other deep convolutional networks due to replacingslow convolutional layers with depthwise separable convolution blocks (Dwise), which reduced the network complexity significantly with high performance and allowed to use of MobilNetV2 in many real-time applications [34].

The construction of the MobilNetV2 is illustrated in Figure 2, which employs two distinct types of blocks. The first is a residual block whose stride is equal to one. Additionally, the second block is for strideis equal to two for downsizing. Additionally, ReLU6 is employed as an activation function because of its resilience while doing low-precision computations.,as in Eq. (1):

$$ReLU6(x) = \min(max(x, 0), 6) \qquad (1)$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

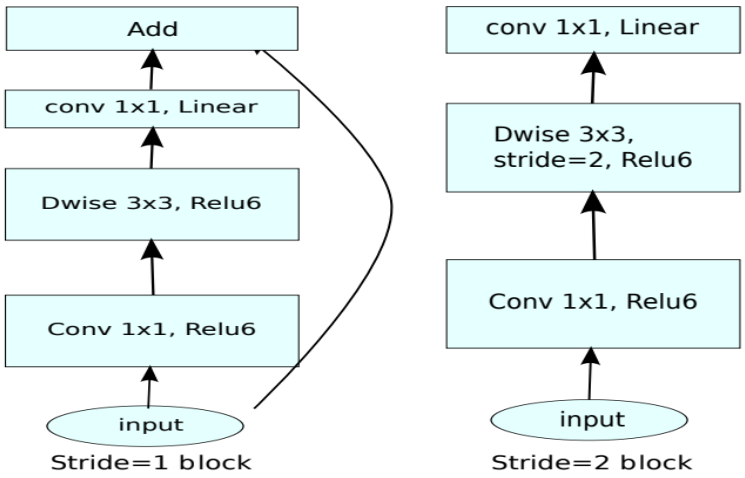

Figure 2: The MobilNetV2 architecture[15].

3.3. LDA

The primary goal of LDA is to minimize computing costs by projecting a large number of features onto a lower-dimensional space with strong class separability via supervised learning[35]. The main LDA are as fellow:

1- A d-dimension mean vector is calculated for each class in the dataset.
2- The scatter matrices are calculated.
3- For the scatter matrices, the eigenvectors and eigenvalues are calculated.
4- Descending order of eigenvalues is used to sort the eigenvectors, then the number of the required reduced dimension is selected to form a *d*i* matrix (where $i \leq x - 1$).
5- A new subspace is created by transforming the input samples using the *d*i* eigenvector matrix.

3.4. XGBoost

As a tree-boost-based machine learning algorithm, XGBoost is widely utilized due to its scalability. It has a more accurate loss function and performs the second-order Taylor expansion. Furthermore, adding regularization into the objective function reduces therisk of overfitting[18].

For a dataset $X_{data} = \{(x_i, y_i)\}$, where *i=1,2,…,n,* and (x_i, y_i) are the input and the ground truth label, respectively.

The predicted value $(\hat{y}_i)$ of a trained model with *K*trees can be calculated as in Eq. (2):

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in W \qquad (2)$$

Where $f(x)$ is a regression tree and *W*is classification and regression trees (CART) (also defined as the hypothesis space of the regression trees as in Eq. (3):

$$W = \{f(x) = \mu_{l(x)}\} \qquad (3)$$

Where μ is the leaf score, and $l(x)$ represents the leaf node.

For the *t* number of iterations, the predicted result can be estimated as in Eq. (4):

$$\hat{y}_i^t = \hat{y}_i^{t-1} + f_t(x_i) \qquad (4)$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

While Eq. (5) calculates the objective function for XGBoost:

$$J(f_t) = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{t-1} + f_t(x_i) + \omega(f_t)\right) \quad (5)$$

Where *L*represents the loss function and $\omega(f_t)$ is the model complexity and can be calculated as in Eq. (6):

$$\omega(f_t) = \beta.T_t + \delta\frac{1}{2}\sum_{j=1}^{T}\mu_j^2 \quad (6)$$

Here *T:* leaves number,
β *and* δ:penalty coefficients.
The second-order Tylor expansion can be applied to the objective function in Eq. (5) as follow in Eq. (7):

$$J(f_t) = \sum_{i=1}^{n}[L(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)] + \omega(f_t) \quad (7)$$

$$g_i = \frac{\partial L(y_i, \hat{y}_i^{t-1})}{\partial \hat{y}_i^{t-1}} \quad (8)$$

$$h_i = \frac{\partial^2 L(y_i, \hat{y}_i^{t-1})}{\partial \hat{y}_i^{t-1}} \quad (9)$$

Eq. (10) calculates the final objective function:

$$j(f_t) = \sum_{i=1}^{n}[g_i\mu_l(x_i) + \frac{1}{2}h_i\mu_l^2(x_i) + \beta T + \delta\frac{1}{2}\sum_{j=1}^{T}\mu_j^2 \quad (10)$$

3.5. BO

As mentioned earlier, XGBoost is used to classify X-ray images after extracting features and the dimensionality reduction processby LDA. However, XGBoost comes with many hyperparameters which need to be tuned to obtain the highest performance. Therefore, the BO algorithm is used to optimize XGBoost'shyperparameters, including n_estimators, max_depth,learning_rate, booster, gamma,subsample,colsample_bytree,colsample_bylevel,colsample_bynode, andreg_lambda. The optimum XGBoost model is considered using the maximum classification accuracy.
Most of the widely used optimization algorithms require the objective function $f(x)$ to be in a known mathematical form, which is not in the case of hyperparameters optimization. On the other hand, the BO algorithm proved its ability to solve optimization problems when the objective function is unknown and complex [36].
The selection of hyperparameters may be viewed as an optimization problem: anobjective function $f(x)$ with an independent variable as the optimum hyperparameter value. Therefore, we require an effective method for modeling the objective function's distribution to employ the BO methodology. For instance, if x includes continuous hyperparameters, an indefinite number of x can be used to simulate $f(x)$. The Gaussian process yields a multidimensional Gaussian distribution for this situation, a highly flexible high-dimensional normal distribution capable of modeling any objective function.
More clearly, BO assumes that the function to be optimized is$f: x \rightarrow \mathbb{R}$, Then, in each iteration ($t = 1,2,\cdots,T$),$f(x_t)$ is calculated using the acquisition function(α_t). Then, a noisy observation $y_t = f(x_t) + \varepsilon$is calculated, where ε follows the zero-mean Gaussian distribution $\varepsilon \sim N(0, \sigma^2)$, and σ is the noise variance. Then, new

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

observations (x_t, y_t) are added to the observation data, and then the next iteration is performed.By learning the objective function and determining the parameters that increase the outcome to the global optimum, BO makes the most of the knowledge from the previous sample point. Finally, the method evaluates the posterior distribution's most likely point.

3.6. The proposed Covid ConvXGB model

The proposed CovidConvXGB model consists of three steps: image preparation and feature extraction using the MobilNetV2, dimensionality reduction using LDA, and XGBoost with BO construction step, which includes evaluating the model and using the best XGBoosthyperparameters. The overall proposed method is shown in Figure 3.

In the first step, the data is split into training and testing sets (as mentioned in the dataset section). Then, using the MobilNetV2, 20480 features were extracted for each image. In order to reduce this high number of features into two components, LDA is used in the second step. Finally, in the last step, the XGBoost is constructed with initialization parameters as in Eq. (2)- Eq. (10).tHere, BO is utilized to obtain the optimum hyperparameters, and the measured accuracy of the XGBoost model is used as the objective function.

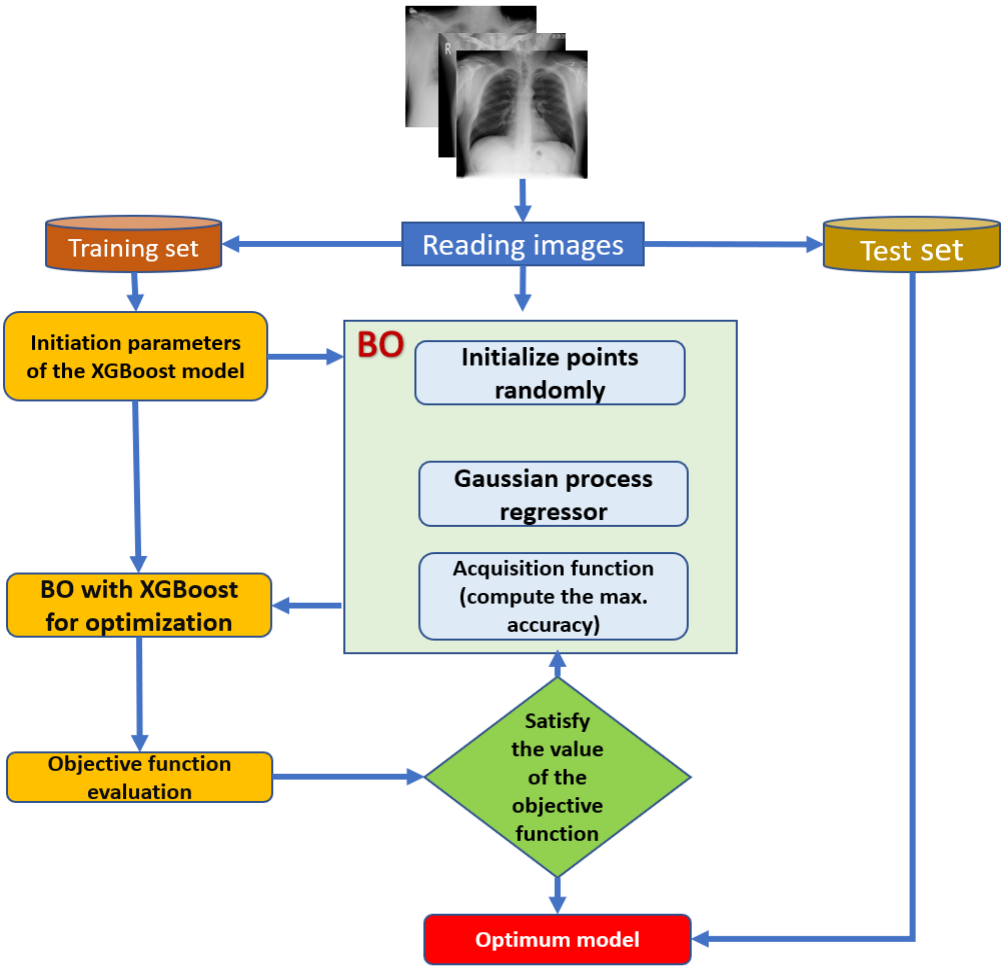

Figure 3: The overall proposed method.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

3.7. **Performance metrics**

In this work, four typically used evaluation metrics are employed to evaluate the performance of the recommended methods: accuracy Eq. (11), recall Eq. (12), precision Eq. (13), and AUC[37].

$$A_{cc} = \frac{True_p + True_n}{True_p + False_p + True_n + False_n} \quad (11)$$

$$R_{call} = \frac{True_p}{True_p + False_n} \quad (12)$$

$$\text{PRE} = \frac{True_p}{True_p + False_p} \quad (13)$$

Where $True_p$ denotes the true positive, $True_n$ denotes the true negative, $False_p$ is the false positive, and $False_n$ is the false negative.
AUC is the performance aggregation measure across all the possible thresholds for classification. The AUC value is [0,1], where a higher value means better model performance.

3.8. Computing environment

The trials were performed on an Intel Core i7 CPU running at 1.7 GHz, with 8 GB of RAM, 64-bit Windows 10 Pro, and an NVIDIA GeForce GT 2 GB display card. Python was used as the programming language.

4. Results and discussion

In this research, the XGBoostclassifier is combined with the BO algorithm to classify chest X-ray images into (normal, Covid-19, and viral pneumonia). The hyperparameters which enabled the model to obtain the highest performance were estimated depending on the optimization results. The hyperparameters search space and the selection of their adjustment of the XGBoost classifier in the BO process are shown in Table 1 from the search space X. The hyperparameters were randomly initialized, then the BO iterated 20 times with evaluating the model's accuracy for each iteration. Therefore, the optimum hyperparameters of the XGBoost classifier can be calculated when the classifier has the maximum accuracy. Additionally, the fitness values with the number of iterations for the process of the BO hyperparametersoptimization are shown in Figure 4.

Table 1: Search space and the optimum XGBoosthyperparameters.

Hyperparameter	Lower limit	Upper limit	Optimum value
n_estimators	200	300	299
max_depth	1	20	18
learning_rate	0.001	0.099	0.0987
Booster	'gbtree' or 'dart'		'gbtree'
gamma	0.01	299	0.7854
subsample	0.50	18	0.770
colsample_bytree	0.50	0.0987	0.5426
colsample_bylevel	0.50	0.90	0.864
colsample_bynode	0.50	0.90	0.7321
reg_lambda	1	10	3

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

It is noticeable that there is no improvement for the first 11 iterations, while after that, the fitness function was increased and converged for the end of the iteration process.

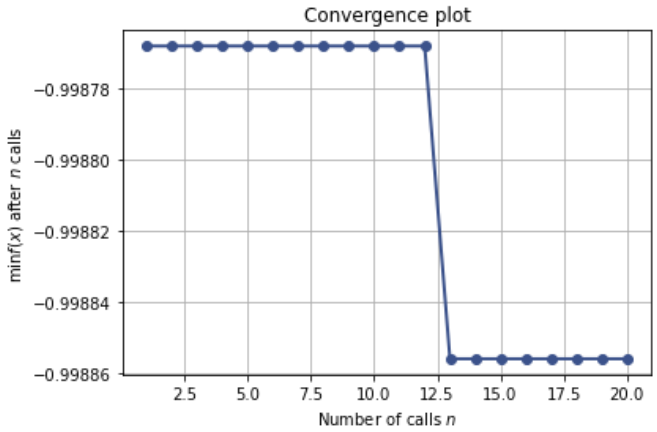

Figure 4: The fitness function value vs. the number of calls (iterations).

On the other hand, Figure 5 shows the dependency plot for the hyperparameters variation during the optimization process, marking the optimum value for each hyperparameter in red. According to this plot, the "n_estimators", "max_depth", and "learning_rate" are the most influencing hyperparameters on the fitness function.

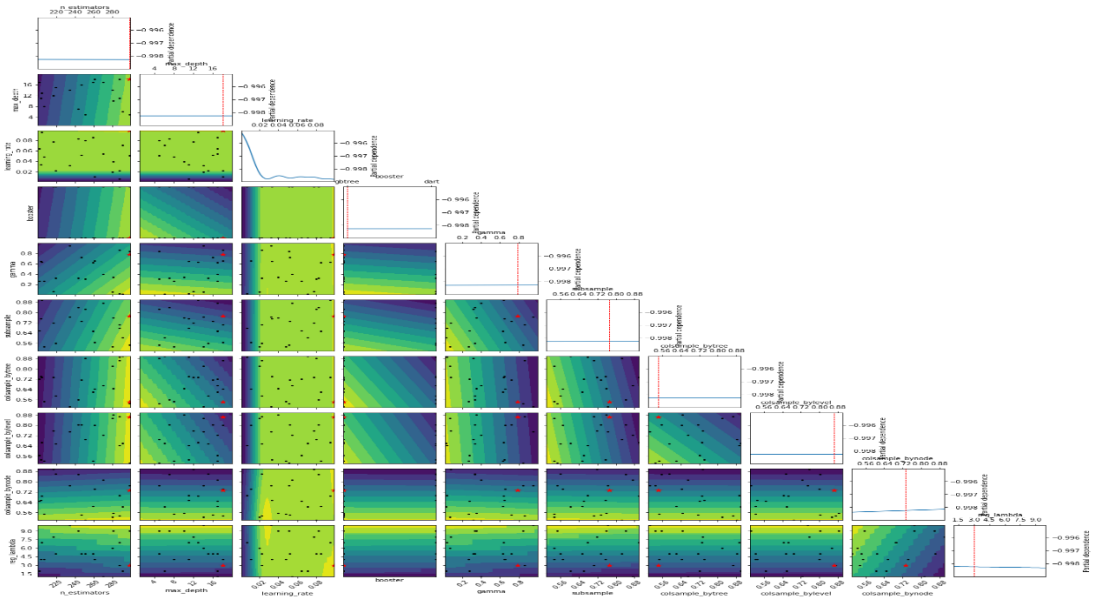

Figure 5: Hyper parameterre adjustment during the optimization process.

For comprehending the BO impact on the XGBoost classification results, Table 2 shows the accuracy obtained when using the default XGBoosthyperparameters and when using the BO for hyperparameters optimization. After the hyperparameters were optimized using the BO, the accuracy increased for both datasets. The results in Table 2prove that using the BO algorithm can help in improving the model performance and its generalization ability.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

Table 2: XGBoost model accuracy comparison with and without using the BO.

Dataset	Accuracy without using the BO	Accuracy with using the BO
Covid-19 radiography	0.9680	0.9765
Chest X-ray Covid-19 pneumonia	0.9526	0.9635

The performance metrics for the XGBoost model, which was trained using the optimum hyperparameters on the Covid-19 radiography dataset (using the test set), are shown in Table 3. While Figures 6 and 7 show the confusion matrix and the ROC curve, respectively.

Table 3: The performance metrics for the XGBoost model and the BO algorithm using the Covid-19 radiography dataset.

Class	Precision	Recall	F1-score
COVID19	0.9841	0.9434	0.9633
NORMAL	0.9710	0.9964	0.9836
PNEUMONIA	1.00	0.9198	0.9582
weighted avg	0.9769	0.9765	0.9763

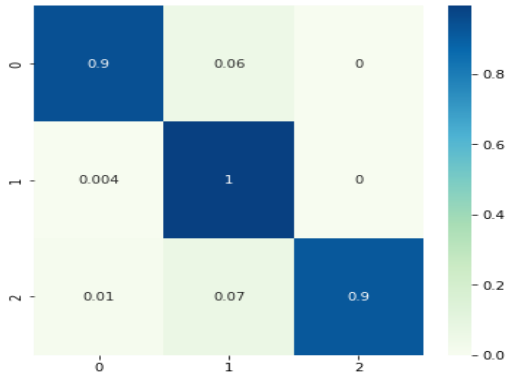

Figure 6: The confusion matrix for the XGBoost model and the BO algorithm using the Covid-19 radiography dataset. 1,2 and 3 represent Covid-19, Normal, and Pneumonia classes, respectively.

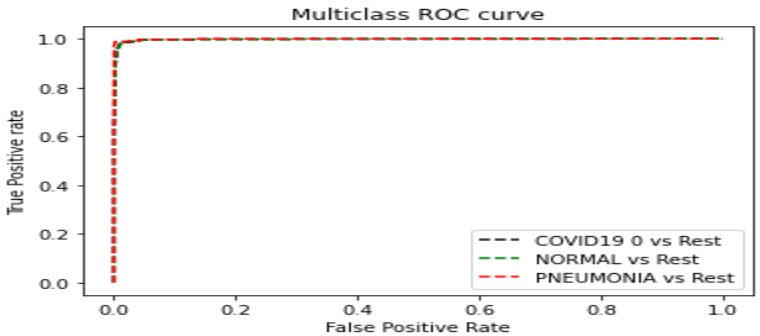

Figure 7: The ROC curve for the XGBoost model and the BO algorithm using the Covid-19 radiography dataset.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

For testing the proposed method's generalization ability, the same trained model was tested using another dataset "Chest X-ray Covid-19 pneumonia" and the results were so close to the previous results, as shown in Table 4. Moreover, the confusion matrix and the Roc curve are shown in Figures 8 and 9.

Table 4: The performance metrics for the XGBoost model and the BO algorithm using the "Chest X-ray Covid-19 pneumonia" dataset.

Class	Precision	Recall	F1-score
COVID19	1.00	0.9310	0.9643
NORMAL	0.9342	0.9401	0.9371
PNEUMONIA	0.9698	0.9766	0.9732
weighted avg	0.9638	0.9635	0.9635

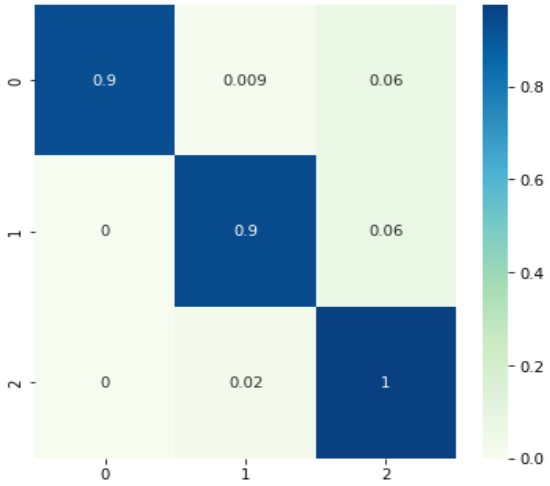

Figure 8: The confusion matrix for the XGBoost model and the BO algorithm using the "Chest X-ray Covid-19 pneumonia". 1,2 and 3 represent Covid-19, Normal, and Pneumonia classes, respectively.

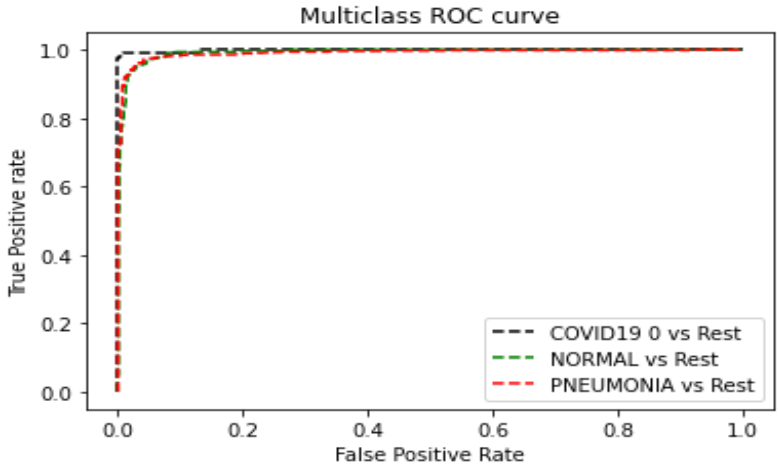

Figure 9: The ROC curve for the XGBoost model and the BO algorithm using the "Chest X-ray Covid-19 pneumonia".

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

4.1. Comparison with the literature works

Many methods were suggested in the literature for Covid-19 prediction using chest X-ray images. Nevertheless, the dataset size and the number of classes differ for each study, which should be considered when comparing the performance. Table 5 compares the proposed method with some related approaches. The findings show that our proposed method achieved the highest accuracy and used the highest number of X-ray images.

Table 5: Comparison of the proposed method with some related literature work.

Work	Dataset (# X-ray images)	# Classes	Method	Accuracy
[38]	127 Covid-19 500 Normal 500 Pneumonia	3	Xception	0.974
[39]	5,863	2	CNN	0.9465
[40]	13,975	4	Capsule Networks	0.957
[31]	1300	4	Xception	0.8960
[41]	140	2	Transfer learning	0.80
[23]	100 Covid-19 1431 Pneumonia	2	residual convolutional neural network	0.9518
This paper	3,616 Covid-19 10,192 Normal 1,345 Pneumonia	3	CovidConvXGB	0.9765
This paper	756 Covid-19 1,583 Normal 4,273 Pneumonia	3	CovidConvXGB	0.9635

5. Conclusion

It is critical to predict COVID-19 individuals early in order to prevent the disease from spreading to others.Therefore, This research suggests a hybrid technique for automatically predicting COVID19 cases utilizing chest X-ray images from normal, COVID-19, and viral pneumonia patients.The proposed method is based on feature extracting using the MobilNetV2 in the first step, then reducingthe extracted features using LDA. Finally, using the XGBoost classifier with the BO algorithm for selecting the best hyperparameters to achieve the highest accuracy.The proposed technique provided both high and near results on two datasets of X-ray images. The accuracy, F1-score, recall, and precision for the first dataset were 97.65 %, 97.63%, and 97.69%, respectively, while the second dataset had 96.35% accuracy, 95.82% F1-score, 98.35% recall, and 96.38% precision.With the help of this proposed method, we think radiologists will be able to make better decisions in the field because COVID-19 can be detected early.Furthermore, this study gives insight into how the BO algorithm can enhance the performance of the XGBoost classifier. For future work, we suggest trying other classifiers such as the AdaBoost and Gradient Tree Boosting with the BO algorithm and adding more classes that belong to different chest X-ray abnormalities such as pleural effusion, embolisms, emphysema, and lung scarring.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1236 - 1250
https://publishoa.com
ISSN: 1309-345

References

1. S. B. Stoecklin*et al.*, "First cases of coronavirus disease 2019 (COVID-19) in France: Surveillance, investigations and control measures, January 2020," *Eurosurveillance*, vol. 25, no. 6. 2020. doi: 10.2807/1560-7917.ES.2020.25.6.2000094.
2. K. Roosa*et al.*, "Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020," *Infectious Disease Modelling*, vol. 5, 2020, doi: 10.1016/j.idm.2020.02.002.
3. N. Chen *et al.*, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," *The Lancet*, vol. 395, no. 10223, 2020, doi: 10.1016/S0140-6736(20)30211-7.
4. Dadax, "HTTPS://WWW.WORLDOMETERS.INFO/CORONAVIRUS/," *COVID-19 Coronavirus Pandemic*, 2020.
5. A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, vol. 24, no. 3, 2021, doi: 10.1007/s10044-021-00984-y.
6. Q. Sun, H. Qiu, M. Huang, and Y. Yang, "Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province," *Annals of Intensive Care*, vol. 10, no. 1. 2020. doi: 10.1186/s13613-020-00650-2.
7. B. Antin, J. Kravitz, and E. Martayan, "Detecting Pneumonia in Chest X-Rays with Supervised Learning," *Semanticscholar. org*, 2017.
8. A. K. Jaiswal, P. Tiwari, S. Kumar, D. Gupta, A. Khanna, and J. J. P. C. Rodrigues, "Identifying pneumonia in chest X-rays: A deep learning approach," *Measurement: Journal of the International Measurement Confederation*, vol. 145, 2019, doi: 10.1016/j.measurement.2019.05.076.
9. A. Cruz-Roa*et al.*, "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," in *Medical Imaging 2014: Digital Pathology*, 2014, vol. 9041. doi: 10.1117/12.2043872.
10. M. F. Jojoa Acosta, L. Y. Caballero Tovar, M. B. Garcia-Zapirain, and W. S. Percybrooks, "Melanoma diagnosis using deep learning techniques on dermatoscopic images," *BMC Medical Imaging*, vol. 21, no. 1, 2021, doi: 10.1186/s12880-020-00534-8.
11. S. T. Jünger*et al.*, "Fully Automated MR Detection and Segmentation of Brain Metastases in Non-small Cell Lung Cancer Using Deep Learning," *Journal of Magnetic Resonance Imaging*, vol. 54, no. 5, 2021, doi: 10.1002/jmri.27741.
12. K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin Cancer Classification using Deep Learning and Transfer Learning," 2019. doi: 10.1109/CIBEC.2018.8641762.
13. G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Computer Methods and Programs in Biomedicine*, vol. 187, 2020, doi: 10.1016/j.cmpb.2019.06.023.
14. A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 15, 2021, doi: 10.1080/07391102.2020.1788642.
15. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018. doi: 10.1109/CVPR.2018.00474.
16. Stanford Vision Lab, "ImageNet Dataset," *Stanford Vision Lab, Stanford University*, 2016.
17. J. and J. R. and L. Q. Ye, "Two-Dimensional Linear Discriminant Analysis," *MIT Press*, vol. 17, 2004.
18. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, vol. 13-17-August-2016. doi: 10.1145/2939672.2939785.

19. M. Talo, O. Yildirim, U. B. Baloglu, G. Aydin, and U. R. Acharya, "Convolutional neural networks for multi-class brain disease detection using MRI images," *Computerized Medical Imaging and Graphics*, vol. 78, 2019, doi: 10.1016/j.compmedimag.2019.101673.
20. S. H. Ahammad*et al.*, "Chexnet reimplementation for pneumonia detection using pytorch," *International Journal of Pharmaceutical Research*, vol. 12, no. 2, 2020, doi: 10.111/J.0975-2366.
21. C. le Van, V. Puri, N. T. Thao, and D. N. Le, "Detecting lumbar implant and diagnosing scoliosis from vietnamese x-ray imaging using the pre-trained api models and transfer learning," *Computers, Materials and Continua*, vol. 66, no. 1, 2021, doi: 10.32604/cmc.2020.013125.
22. I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, 2020, doi: 10.1007/s13246-020-00865-4.
23. J. Zhang, Y. Xie, Y. Li, and Y. Xia, "COVID-19 Screening on Chest X-ray Images Using Deep Learning based Anomaly Detection," *IEEE Transactions on Medical Imaging*, vol. 1, no. 3, 2020.
24. D. Singh, V. Kumar, Vaishali, and M. Kaur, "Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks," *European Journal of Clinical Microbiology and Infectious Diseases*, vol. 39, no. 7, 2020, doi: 10.1007/s10096-020-03901-z.
25. X. Chen, L. Yao, and Y. Zhang, "Residual Attention U-Net for Automated Multi-Class Segmentation of COVID-19 Chest CT Images," Apr. 2020.
26. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351. doi: 10.1007/978-3-319-24574-4_28.
27. N. Chand and D. Adhikari, "Infection Severity Detection of CoVID19 from X-Rays and CT Scans Using Artificial Intelligence," *International Journal of Computer (IJC)*, vol. 38, no. 1, 2020.
28. Q. S. A. A. Alqudah AM, "Automated systems for detection of COVID-19 using chest X-ray images and lightweight convolutional neural networks.," *https://doi.org/10.21203/ rs.3.rs-24305/v1*, 2020.
29. Mooney P, "Chest X-ray images (Pneumonia)," *Kaggle Repository. https://www.kaggle.com/paultimothymooney/chestxray-pnemonia*, 2018.
30. K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, 2016, doi: 10.1186/s40537-016-0043-6.
31. A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," *Computer Methods and Programs in Biomedicine*, vol. 196, 2020, doi: 10.1016/j.cmpb.2020.105581.
32. B. Ghoshal and A. Tucker, "Estimating Uncertainty and Interpretability in Deep Learning for Coronavirus (COVID-19) Detection," Mar. 2020.
33. E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," Mar. 2020.
34. A. Michele, V. Colin, and D. D. Santika, "Mobilenet convolutional neural networks and support vector machines for palmprint recognition," in *Procedia Computer Science*, 2019, vol. 157. doi: 10.1016/j.procs.2019.08.147.
35. G. T. Reddy *et al.*, "Analysis of Dimensionality Reduction Techniques on Big Data," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2980942.
36. J. Zhou, Y. Qiu, S. Zhu, D. J. Armaghani, M. Khandelwal, and E. T. Mohamad, "Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization," *Underground Space (China)*, vol. 6, no. 5, 2021, doi: 10.1016/j.undsp.2020.05.008.
37. W. Zhu, N. Zeng, and N. Wang, "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations.," *Northeast SAS Users Group 2010: Health Care and Life Sciences*, 2010.

38. N. Narayan Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, "Automated Deep Transfer Learning-Based Approach for Detection of COVID-19 Infection in Chest X-rays," *IRBM*, 2020, doi: 10.1016/j.irbm.2020.07.001.
39. D. Singh, V. Kumar, V. Yadav, and M. Kaur, "Deep Convolutional Neural Networks based Classification model for COVID-19 Infected Patients using Chest X-ray Images," *International Journal of …*. 2020.
40. P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images," *Pattern Recognition Letters*, vol. 138. 2020. doi: 10.1016/j.patrec.2020.09.010.
41. K. Sahinbas and F. O. Catak, "Transfer learning-based convolutional neural network for COVID-19 detection with X-ray images," in *Data Science for COVID-19*, 2021. doi: 10.1016/b978-0-12-824536-1.00003-4.