

Convolutional Neural Network Based Identification of Respiratory Disease (CNN-IRD)

1st Qasim Umer

Department of Computer Science
Hanyang University
Seoul, South Korea.
qasimumer@hanyang.ac.kr

2nd Zunaira Naveed

Department of Computer Science
Institute of Southern Punjab
Multan, Pakistan.
zunairanaveed85@gmail.com

3rd Choonhwa Lee*

Department of Computer Science
Hanyang University
Seoul, South Korea.
lee@hanyang.ac.kr

4th Asif Ali

School of Automation
Beijing Institute of Technology
Beijing, China
3820192003@bit.edu.cn

5th Malik Khizar Saeed

Department of Computer Science
COMSATS University Islamabad
Vehari, Pakistan
malikkhizar1@gmail.com

Abstract—Coronavirus is a type of virus that can cause Respiratory Disease (RD) in people. The World Health Organization (WHO) states that signs and symptoms in mild cases include dry throat, fever, nasal secretions, shortness of breath, fever, and malaise. The disease is more dangerous than viruses and can cause serious illness. Although many researchers have tried various techniques for classifying RD patients, it is essential to identify the critical features before applying machine learning methods for classification to save time and cost. To this end, this paper proposes a Convolutional Neural Network (CNN) based Identification of RD patients (CNN-IRD) caused by Coronavirus and divides them into two classes, i.e., $C19_{+ve}$ and $C19_{-ve}$. First, we apply the binarization technique to preprocess data into useful information. Second, we identify the significant features using Linear Discriminant Analysis (LDA). Finally, we train a deep learning classifier (CNN) with two publicly available datasets. The evaluation results suggest that CNN yields other classifiers in predicting RD patients. The performance improvements of CNN-IRD in accuracy, precision, recall, and f-measure with both datasets are (7.92%, 5.35%, 16.92%, and 11.22%) and (4.04%, 9.08%, 25.18%, and 17.25%), respectively.

Index Terms—Convolutional Neural Network, Classification, Decision Tree, Linear Regression, COVID-19, Respiratory

I. INTRODUCTION

In Hubei Province, Wuhan, China, several cases of pneumonia, fever, cough, and shortness of breath due to medical symptoms have been diagnosed since December 2019 [1]. Acute Respiratory Disease (RD) is a new strain of the coronavirus family. The virus is most often seen when coughing, sneezing, or talking to

each other for short periods. Drops are then moved to the surface where they breathe or come in contact with others, which can be contaminated when they connect with the eyes, mouth, or nose [2]. The results show that the COVID-19 virus is transmitted from individual to individual. Complications of the disease include respiratory problems, heart damage, and secondary infections. Most of the time, the affected people should seek medical treatment. The victim suffers from severe shortness of breath. Computed Tomography (CT) scan of the affected person expresses that coronavirus has features. Therefore, medical professionals need CT images for insanity to work in the initial phases of coronavirus. Coronavirus remains a significant public health emergency in the United States regarding the novel coronavirus SARS-Co-2. The increasing awareness of drug therapy for coronavirus has established an effective patient triage and a lack of precautionary resources [3].

The current study shows that a huge proportion of affected patients have comorbidities. Most patients are at the threat stage for serious conditions that may be needed and, as a result, may be admitted to an intensive care unit [4]. The capability to classify RD patients in danger of exacerbating their disease is essential for patient referral and effective treatment [5]. Without an effective trajectory, the collapse of the medical system is essential in several ways to strengthen national affairs [6]. Implementing multiple machine learning models can be accessed by selecting and classifying multivariate contrast data functions and promises to conclude medical practice [7]. It is a priority recommended for predicting trends [8] and risk stability

*Corresponding Author: Choonhwa Lee (lee@hanyang.ac.kr)

in patients through coronavirus [9]. Some researchers find a relationship between the coronavirus image profile and clinical classification. By March 12, 2020, 124,922 confirmed cases will be recorded worldwide (e.g., USA, Vietnam, and Germany) [10]–[13].

Three stages of coronavirus can be categorized according to the type of biological contact the individual has with the virus. In the first stage, doctors may recommend undertaking plasma recovery therapy which involves transferring plasma from patients who have previously been cured to patients to reduce viral infection and hasten recovery. At the pulmonary (second) stage, doctors suggested using TPA blood (tissue plasminogen activator), which helps break blood clots. However, doctors recommend anti-inflammatory drugs to prevent the formation of blood clots in the blood vessels and capillaries at the hyper-inflammatory (last) stage. Although many researchers have tried various techniques for classifying patients with coronavirus, it is critical to save time.

To this end, this paper proposes a CNN based Identification of RD patients (CNN-IRD) caused by Coronavirus by exploiting feature selection techniques. First, it employs LDA as a feature selection method. Second, it creates a feature vector based on the selected features. Third, it trains a CNN classifier with two publicly available datasets. Notably, 10-fold cross-validation is exploited for the evaluation of CNN-IRD. The results suggest that CNN yields other classifiers in predicting RD patients.

II. RELATED WORKS

In this section, we discuss the state-of-the-art research in this area, categorized into machine learning-based prediction and deep learning-based prediction of COVID patients.

Machine Learning-based Prediction: Rishikesh Magar et al. [14] proposed a model for predicting the effectiveness of synthetic antibodies against the coronavirus. J.S. Cobb and M.A. Seale [15] employed statistical analysis and Random Forest (RF) classification to accurately classify COVID-19 patients, demonstrating that SIP (shelter-in-place) measures effectively reduced the growth rate of COVID-19 cases in the United States. Fu-Yuan Cheng et al. [16] also employed RF for predicting COVID-19 patients, achieving encouraging results.

Bogdan Georgescu et al. [17] utilized Logistic Regression (LoR) and RF classifiers to distinguish COVID-19 cases from pneumonia, PID, and normal scans using chest computed tomography. Frank S. Heldt et al. [18] utilized RF, Extreme Gradient Boosting (XGBoost), and LoR for early assessment of COVID-19 patients by leveraging electronic health record (EHR) data from patients' initial visits to the emergency department.

Nathan A. Brooks et al. [19] proposed an approach that utilized K-means Clustering (KC) and Logistic Regression (LR) classifiers to predict COVID-19 patients. Mucahid Barstugan et al. [20] employed a combination of feature extraction techniques and Support Vector Machine (SVM) classifiers. Jiangpeng Wu et al. [21] developed a model based on RF that utilized 11 clinically available blood indicators to predict COVID-19 patients. Li Yan et al. [22] employed XGBoost to identify severe cases of acutely ill patients at risk of death. Akhil Vaid et al. [23] proposed a model that analyzed SHAP (SHapley Additive exPlanations) scores to determine important features, including age, for predicting COVID-19 patients. Muhammad et al. [24] developed a model using multiple classifiers to predict high-risk patients who may not recover from COVID-19. Furqan Rusta et al. [25] proposed an approach using LASSO, SVM, and Exponential Smoothing (ES) to predict COVID-19 patients. Milind Yadav et al. [26] proposed a novel support vector regression method to investigate various functions associated with the novel coronavirus. Hoyt Burdick et al. [27] adopted an XGBoost classifier to predict the demand for mechanical ventilation within a 24-hour timeframe.

Deep Learning-based Prediction: Deep learning techniques have also been explored for predicting COVID patients. Rahul Kumar et al. [28] implemented ResNet152, RF, and XGBoost classifiers to predict outbreaks and control them effectively. Their model achieved a precision of 97% with RF and 97.7% with XGBoost. Sina F. Ardabili et al. [29] investigated the use of two machine learning models, Adaptive Network-based Fuzzy Inference System (ANFIS) and Multilayer Perceptron (MLP), for long-term prediction of COVID-19 outbreaks. They demonstrated the efficiency of machine learning in simulating epidemic timing.

Some studies employed a combination of techniques and models for COVID-19 prediction. Papaa et al. [30], [31] used multiple regression and machine learning models, including SVM, RF, and neural networks, for COVID patient prediction. They found that the proposed approach produced consistent results, with EARTH, PPR, SVM-radial, PLS, SVM-linear, and MLR providing the best statistical results.

Deep learning models have also been utilized for COVID patient classification using medical images. Aayush Jaiswal et al. [32] used Deep Transfer Learning (DTL) based on DenseNet201 to classify patients as COVID-19 positive or negative. Comparative analyses showed that their proposed DTL-based model outperformed other models. Dilbag Singh et al. [33] employed Convolutional Neural Network (CNN) for classifying COVID-19 infected patients as positive or negative, achieving high accuracy, sensitivity, and precision compared to other models.

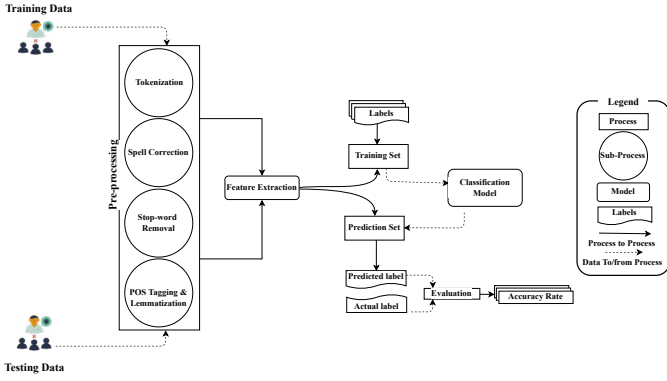


Fig. 1. Overview of CNN-IRD

Moreover, studies have explored the combination of machine learning and deep learning models. Amir Ahmad et al. and Narinder Singh Punn et al. [34], [35] used SVR, PR, DNN, LSTM, RMSE, and other models to analyze COVID-19 evolution and reduce the number of infected individuals. Lu Wang Li et al. [36] developed a model based on historical patterns and used RNN, LSTM, and supervised learning methods for accurate prediction and explanation of parameters. Dan Assaf et al. [37] employed neural networks, RF, and DT for predicting COVID patients based on various constraints and the APACHE II score.

III. PROPOSED APPROACH

A. Problem Definition

An overview of the proposed approach (CNN-IRD) is illustrated in Fig. 1. A RD patient record p from a set of RD patients' record \mathbb{P} can be formalized as

$$p = \langle f_i, l \rangle \quad (1)$$

where, f_i are the features of p and l is an assigned label to p .

CNN-IRD suggests the class of a new RD patient p as either C_{+ve} and C_{-ve} , where C_{+ve} represents a COVID-patient and C_{-ve} represents a patient that is having other medical problem, e.g., pneumonia. As a result, the mapping of f can be defined as the automatic prediction of a new RD patient p :

$$f : p \rightarrow c$$

$$c \in \{C_{+ve} \text{ or } C_{-ve}\}, \quad p \in \mathbb{P} \quad (2)$$

where, c is a suggested RD patient category from a classification set (C_{+ve} , C_{-ve}).

B. Preprocessing

During preprocessing, data is sorted, unnecessary data columns are removed, and data conversion is applied to convert string values into numerical data.

After the preprocessing, a RD patient record p can be formalized as

$$p = \langle f'_i, l \rangle \quad (3)$$

where, f'_i are the preprocessed features of p .

C. Key Feature Selection

In this step, CNN-IRD preprocesses the given data to find the key features to reduce its prediction time (cost). We exploit Decision Tree (DT), Linear Regression (LR), and Linear Discriminant Analysis (LDA) for the selection of key features. Notably, DT applies Gini Entropy Impurity, LR predicts the cost of dependents and LDA variables, and LDA reduces the dimensionality for selecting key features. Among them, LDA-based feature selection yields the performance of CNN-IRD in identifying COVID patients. After the key feature selection, a patient record p can be formalized as

$$p = \langle f''_i, l \rangle \quad (4)$$

where, f''_i are the key features of p and $f''_i < f'_i$.

D. Proposed Model

Assuming a one-dimensional feature vector \mathbf{x} and a RD patient p with a maximal key feature (kf) length of n , we can calculate n by finding the patient p in P with the longest kf length and padding the remaining patients in P to match that length. This representation of \mathbf{x} can be expressed as follows:

$$\mathbf{x} = \langle kf_1, kf_2, \dots, kf_n \rangle \quad (5)$$

$$\mathbf{x}' = \langle vkf_1, vkf_2, vkf_3, \dots, vkf_n \rangle \quad (6)$$

where, vkf_i is the vector representation of p . The CNN algorithm applies a filter $\mathbf{ft} \in \mathbb{P}^{dk}$ to a window of d features, generating a new feature. For example, a new feature nf_i can be generated from a window of words $v_{i,i+d-1}$. This process can be formalized as follows:

$$nf_i = f(\mathbf{ft}.v_{i,i+d-1} + b) \quad (7)$$

In the above equation, b represents a bias and f represents a hyperbolic tangent non-linear function. The filter generates a feature map using each window of features $\langle v_{1:d}, v_{2:d+1}, \dots, v_{n-d+1:n} \rangle$. A feature map \mathbf{nf} generated from this process belongs to \mathbb{P}^{n-d+1} and can be formalized as:

$$\mathbf{nf} = \langle nf_1, nf_2, \dots, nf_{n-d+1} \rangle \quad (8)$$

A max-pooling operation is applied to obtain the maximum value nf from the feature map defined in Eq. 8. This operation is used to identify the most significant features in the feature map, i.e., the features with the highest value. In the CNN-IRD, multiple filters of varying sizes are applied, and one feature is extracted from each filter based on its maximum value. These features construct the penultimate layer, which is then passed to a fully connected softmax layer. The output

TABLE I
DESCRIPTION OF DATASETS

	No. of Patients	C _{+ve}	C _{-ve}
Dataset-1	83,176	23,472	59,704
Dataset-2	499,693	220,658	278,981

TABLE II
PERFORMANCE OF CNN-IRD

	Approach	A	P	R	FM
Dataset-1	CNN-IRD	95.13%	96.84%	93.93%	95.36%
	CO19-Analysis	88.15%	91.92%	80.34%	85.74%
	CO19-Restro	91.44%	88.78%	75.03%	81.33%
Dataset-2	CNN-IRD	94.46%	96.11%	91.71%	93.86%
	CO19-Analysis	87.90%	91.26%	76.45%	83.20%
	CO19-Restro	80.94%	88.12%	62.59%	73.19%

of this layer is a probability distribution of COVID patients.

IV. EVALUATION

A. Dataset

We reuse the two publicly available datasets from Kaggle. The statistics of the datasets are presented in Table I. The dataset-1 has 83,176 patient records consist of two classes as C_{+ve} and C_{-ve}, where 23,472 and 59,704 patient records belong to C_{+ve} and C_{-ve}, respectively. The dataset-2 contains 499,693 patient records, where 220,658 and 278,981 patient records belong to C_{+ve} and C_{-ve}, respectively. Dataset-1 and dataset-2 consist of 18 columns (gender, patient type, intubated, pneumonia, age, pregnancy, diabetes, chronic obstructive pulmonary disease, asthma, immunosuppression, hypertension, other decease, obesity, kidney failure, smoker, ICU, others, and class label) and 21 columns (patient id, gender, patient type, intubated, pneumonia, age, pregnancy, diabetes, chronic obstructive pulmonary disease, asthma, immunosuppression, hypertension, other decease, cardiovascular, obesity, kidney failure, smoker, contact with COVID, ICU, others, and class label), respectively.

B. Results

1) *Performance Comparison of CNN-IRD against Baseline Approaches:* We contrast the performance results of CNN-IRD, CO19-Analysis, and CO19-Restro for their performance comparison. We perform 10-fold cross-validation and present the average evaluation results of all approaches in Table II. We also visualize the distribution of f-measure across the cross-validation folds by plotting beanplots (Fig. 2 and Fig. 3) for the approaches. The analysis concludes that CNN-IRD outperforms baseline approaches. The improvement of CNN-IRD upon CO19-Analysis in Accuracy (A), Precision (P), recall (R), and f-Measure (FM) against dataset-1 is up to (7.92%, 5.35%, 16.92%,

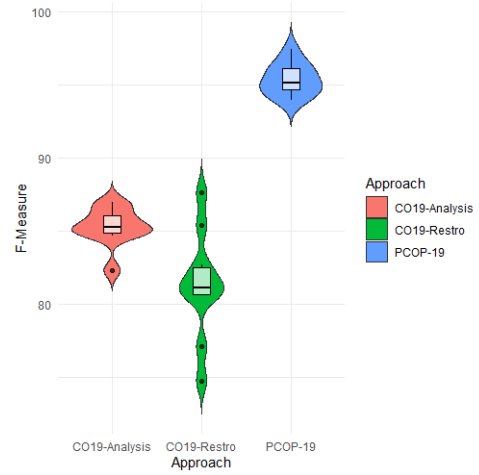


Fig. 2. Distribution of F-Measure against Dataset-1

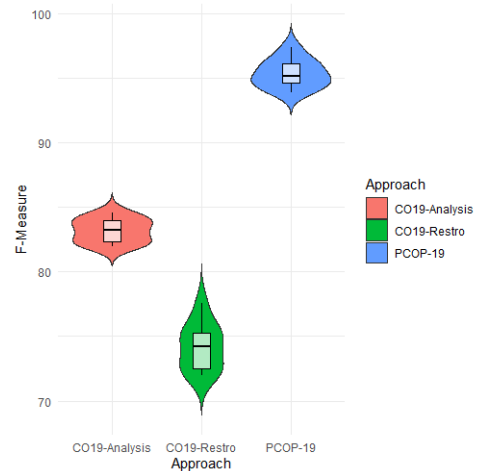


Fig. 3. Distribution of F-Measure against Dataset-2

and 11.22%) and (4.04%, 9.08%, 25.18%, and 17.25%), respectively. Similarly, The improvement of CNN-IRD upon CO19-Restro in A, P, R, and FM against dataset-2 is up to (4.84%, 4.71%, 17.85%, and 11.48%) and (13.85%, 8.44%, 43.95%, and 26.72%), respectively. Furthermore, CNN-IRD demonstrates a more consistent performance compared to baseline approaches, as evidenced by the relatively stable distribution of f-measure results across the 10 cross-validation folds.

Moreover, we conduct a one-way analysis of variance (ANOVA) to determine if there is a significant difference in performance between CNN-IRD, CO19-Analysis, and CO19-Restro. The ANOVA test was performed on two datasets, and the results show that the factor of using different approaches has a significant difference with p-values less than 0.05. The f-ratio for dataset-1 is p-value = $4.79E-13$, and the f-ratio for dataset-2 is p-value = $2.06E-23$.

TABLE III
INFLUENCE OF KEY FEATURES SELECTION

	Feature Selection	A	P	R	FM
Dataset-1	Enable	95.13%	96.84%	93.93%	95.36%
	Disable	90.73%	86.04%	74.65%	79.94%
Dataset-2	Enable	94.46%	96.11%	91.71%	93.86%
	Disable	83.55%	88.37%	73.49%	80.14%



Fig. 4. Influence of Machine Learning Techniques against Dataset-1

2) *Impact of Feature Selection Method:* We contrast the performance of CNN-IRD with and without different inputs (complete features and best features) to check the impact of the feature selection method on CNN-IRD. The performance results against different settings are presented in Table III. The results suggest that disabling key feature selection (i.e., input complete feature set) significantly reduces the performance of CNN-IRD against dataset-1. It reduces the A, P, R, and FM up to 4.85%, 12.55%, 25.83%, and 19.29%, respectively. Similarly, disabling key feature selection (i.e., input complete feature set) significantly reduces the performance of CNN-IRD against dataset-2. It reduces the A, P, R, and FM up to 13.06%, 8.75%, 24.79%, and 17.12%, respectively.

3) *Influence of Machine Learning Techniques:* We also compared the performance of CNN-IRD with that of DT, LR, SVM, and MNB. Fig. 4 and Fig. 5 show the evaluation results of CNN-IRD under different ML techniques against both datasets, respectively. The figures conclude that CNN outperforms LSTM, RF, DT, LR, SVM, and MNB against both datasets. The possible reason for the performance significance is that CNN handles long-term dependencies and does not suffer from the exploding gradient problem. Note that the performance of DT is very close to RF. DT may outperform RF on other datasets. Moreover, the Bayesian is outperformed by LR and SVM, possibly due to the high number of dependent features in patient data. According to [38], when there is a large number of training feature sizes, discriminative models, i.e., logistic regression, tend to perform better than generative models, i.e., Bayesian.

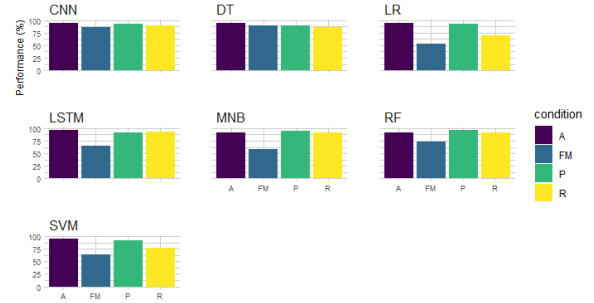


Fig. 5. Influence of Machine Learning Techniques against Dataset-2

V. CONCLUSION

Although multiple studies have tried various techniques to classify COVID patients, it is essential to identify the critical features before applying machine learning methods for classification to save time and cost. This paper proposes a deep learning (CNN) based approach to predict COVID patients and divide them into two classes, i.e., C_{+ve} and C_{-ve} . First, we apply the binarization technique to preprocess data into useful information. Second, we use the feature selection technique (LDA) to identify critical features. Finally, we train CNN classifier. The evaluation results suggest that CNN outperforms the other classifiers in predicting COVID patients. Notably, two datasets are reused for the evaluation of CNN-IRD. The improvements of CNN-IRD in accuracy, precision, recall, and f-measure with both datasets are (7.92%, 5.35%, 16.92%, and 11.22%) and (4.04%, 9.08%, 25.18%, 17.25%), respectively.

ACKNOWLEDGMENT

This research was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-00590, Decentralized High Performance Consensus for Large-Scale Blockchains).

REFERENCES

- [1] K. Li, Y. Fang, W. Li, C. Pan, P. Qin, Y. Zhong, X. Liu, M. Huang, Y. Liao, and S. Li, "Ct image visual quantitative evaluation and clinical classification of coronavirus disease (covid-19)," *European radiology*, vol. 30, no. 8, pp. 4407–4416, 2020.
- [2] M. Yadav, M. Perumal, and M. Srinivas, "Analysis on novel coronavirus (covid-19) using machine learning methods," *Chaos, Solitons & Fractals*, vol. 139, p. 110050, 2020.
- [3] H. Burdick, C. Lam, S. Mataraso, A. Siefkas, G. Braden, R. P. Dellinger, A. McCoy, J.-L. Vincent, A. Green-Saxena, G. Barnes *et al.*, "Prediction of respiratory decompensation in covid-19 patients using machine learning: The ready trial," *Computers in biology and medicine*, vol. 124, p. 103949, 2020.
- [4] A. A. Osi, H. G. Dikko, M. Abdu, A. Ibrahim, L. A. Isma'il, H. Sarki, U. Muhammad, A. A. Suleiman, S. S. Sani, and M. Z. Ringim, "A classification approach for predicting covid-19 patient survival outcome with machine learning techniques," *medRxiv*, 2020.

- [5] D. Assaf, Y. Gutman, Y. Neuman, G. Segal, S. Amit, S. Gefen-Halevi, N. Shilo, A. Epstein, R. Mor-Cohen, A. Biber *et al.*, "Utilization of machine-learning models to accurately predict the risk for critical covid-19," *Internal and emergency medicine*, vol. 15, no. 8, pp. 1435–1443, 2020.
- [6] P. Erika, V. Andrea, M. G. Cillis, E. Ioannilli, T. Iannicelli, and M. Andrea, "Triage decision-making at the time of covid-19 infection: the piacenza strategy," *Internal and emergency medicine*, vol. 15, no. 5, pp. 879–882, 2020.
- [7] H.-C. Lee, S. B. Yoon, S.-M. Yang, W. H. Kim, H.-G. Ryu, C.-W. Jung, K.-S. Suh, and K. H. Lee, "Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model," *Journal of clinical medicine*, vol. 7, no. 11, p. 428, 2018.
- [8] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai *et al.*, "Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions," *Journal of thoracic disease*, vol. 12, no. 3, p. 165, 2020.
- [9] G. Chassagnon, M. Vakalopoulou, E. Battistella, S. Christodoulidis, T.-N. Hoang-Thi, S. Dangeard, E. Deutsch, F. Andre, E. Guillo, N. Halm *et al.*, "Ai-driven ct-based quantification, staging and short-term outcome prediction of covid-19 pneumonia," *arXiv preprint arXiv:2004.12852*, 2020.
- [10] Y. F. L. P. P. Q. Z. X. L. K. Li, "Find the link between the imaging profile and the clinical classification of covid-19," 2020.
- [11] L. T. Phan, T. V. Nguyen, Q. C. Luong, T. V. Nguyen, H. T. Nguyen, H. Q. Le, T. T. Nguyen, T. M. Cao, and Q. D. Pham, "Importation and human-to-human transmission of a novel coronavirus in vietnam," *New England Journal of Medicine*, vol. 382, no. 9, pp. 872–874, 2020.
- [12] C. Rothe, M. Schunk, P. Sothmann, G. Bretzel, G. Froeschl, C. Wallrauch, T. Zimmer, V. Thiel, C. Janke, W. Guggemos *et al.*, "Transmission of 2019-ncov infection from an asymptomatic contact in germany," *New England journal of medicine*, vol. 382, no. 10, pp. 970–971, 2020.
- [13] M. L. Holshue, C. DeBolt, S. Lindquist, K. H. Lofy, J. Wiesman, H. Bruce, C. Spitters, K. Ericson, S. Wilkerson, A. Tural *et al.*, "First case of 2019 novel coronavirus in the united states," *New England Journal of Medicine*, 2020.
- [14] R. Magar, P. Yadav, and A. B. Farimani, "Potential neutralizing antibodies discovered for novel corona virus using machine learning," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.
- [15] J. Cobb and M. Seale, "Examining the effect of social distancing on the compound growth rate of covid-19 at the county level (united states) using statistical analyses and a random forest machine learning model," *Public health*, vol. 185, pp. 27–29, 2020.
- [16] F.-Y. Cheng, H. Joshi, P. Tandon, R. Freeman, D. L. Reich, M. Mazumdar, R. Kohli-Seth, M. A. Levin, P. Timsina, and A. Kia, "Using machine learning to predict icu transfer in hospitalized covid-19 patients," *Journal of clinical medicine*, vol. 9, no. 6, p. 1668, 2020.
- [17] E. J. M. Barbosa, B. Georgescu, S. Chaganti, G. B. Aleman, J. B. Cabrero, G. Chabin, T. Flohr, P. Grenier, S. Grbic, N. Gupta *et al.*, "Machine learning automatically detects covid-19 using chest cts in a large multicenter cohort," *European radiology*, pp. 1–11, 2021.
- [18] F. S. Heldt, M. P. Vizcaychipi, S. Peacock, M. Cinelli, L. McLachlan, F. Andreotti, S. Jovanović, R. Dürichen, N. Lipunova, R. A. Fletcher *et al.*, "Early risk assessment for covid-19 patients from emergency department data using machine learning," *Scientific reports*, vol. 11, no. 1, pp. 1–13, 2021.
- [19] N. A. Brooks, A. Puri, S. Garg, S. Nag, J. Corbo, A. El Turabi, N. Kaka, R. W. Zempel, P. K. Hegarty, and A. M. Kamat, "Covid-19 mortality and bcg vaccination: defining the link using machine learning," 2020.
- [20] M. Barstugan, U. Ozkaya, and S. Ozturk, "Coronavirus (covid-19) classification using ct images by machine learning methods," *arXiv preprint arXiv:2003.09424*, 2020.
- [21] J. Wu, P. Zhang, L. Zhang, W. Meng, J. Li, C. Tong, Y. Li, J. Cai, Z. Yang, J. Zhu *et al.*, "Rapid and accurate identification of covid-19 infection through machine learning based on clinical available blood test results," *MedRxiv*, 2020.
- [22] L. Yan, H.-T. Zhang, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, S. Li, M. Zhang *et al.*, "Prediction of criticality in patients with severe covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in wuhan," *MedRxiv*, 2020.
- [23] A. Vaid, S. Somani, A. J. Russak, J. K. De Freitas, F. F. Chaudhry, I. Paranjpe, K. W. Johnson, S. J. Lee, R. Miotto, S. Zhao *et al.*, "Machine learning to predict mortality and critical events in covid-19 positive new york city patients," *medRxiv*, 2020.
- [24] L. Muhammad, M. M. Islam, S. S. Usman, and S. I. Ayon, "Predictive data mining models for novel coronavirus (covid-19) infected patients recovery," *SN Computer Science*, vol. 1, no. 4, pp. 1–7, 2020.
- [25] F. Rustam, A. A. Reshi, A. Mehmood, S. Ullah, B.-W. On, W. Aslam, and G. S. Choi, "Covid-19 future forecasting using supervised machine learning models," *IEEE access*, vol. 8, pp. 101 489–101 499, 2020.
- [26] M. Yadav, M. Perumal, and M. Srinivas, "Analysis on novel coronavirus (covid-19) using machine learning methods," *Chaos, Solitons & Fractals*, vol. 139, p. 110050, 2020.
- [27] H. Burdick, C. Lam, S. Mataraso, A. Siefkas, G. Braden, R. P. Dellinger, A. McCoy, J.-L. Vincent, A. Green-Saxena, G. Barnes *et al.*, "Prediction of respiratory decompensation in covid-19 patients using machine learning: The ready trial," *Computers in biology and medicine*, vol. 124, p. 103949, 2020.
- [28] R. Kumar, R. Arora, V. Bansal, V. J. Sahayashela, H. Buckchash, J. Imran, N. Narayanan, G. N. Pandian, and B. Raman, "Accurate prediction of covid-19 using chest x-ray images through deep feature learning model with smote and machine learning classifiers," *MedRxiv*, 2020.
- [29] S. F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A. R. Varkonyi-Koczy, U. Reuter, T. Rabczuk, and P. M. Atkinson, "Covid-19 outbreak prediction with machine learning," *Algorithms*, vol. 13, no. 10, p. 249, 2020.
- [30] E. Papa, J. Doucet, A. Sangion, and A. Doucet-Panaye, "Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches," *SAR and QSAR in Environmental Research*, vol. 27, no. 7, pp. 521–538, 2016.
- [31] M. Kukar, G. Gunčar, T. Vovko, S. Podnar, P. Černelč, M. Brvar, M. Zalaznik, M. Notar, S. Moškon, and M. Notar, "Covid-19 diagnosis by routine blood tests using machine learning," *Scientific reports*, vol. 11, no. 1, pp. 1–9, 2021.
- [32] 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*, pp. 1–8, 2020.
- [33] D. Singh, V. Kumar, M. Kaur *et al.*, "Classification of covid-19 patients from chest ct images using multi-objective differential evolution-based convolutional neural networks," *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 39, no. 7, pp. 1379–1389, 2020.
- [34] A. Ahmad, S. Garhwal, S. K. Ray, G. Kumar, S. J. Malebary, and O. M. Barukab, "The number of confirmed cases of covid-19 by using machine learning: Methods and challenges," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2645–2653, 2021.
- [35] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Covid-19 epidemic analysis using machine learning and deep learning algorithms," *MedRxiv*, 2020.
- [36] L. Wang, D.-w. Han, K. Li, X. Yang, X.-l. Yin, J. Qiu, D.-x. Liang, and Z.-y. Ma, "A universal model for prediction of covid-19 pandemic based on machine learning," 2020.
- [37] D. Assaf, Y. Gutman, Y. Neuman, G. Segal, S. Amit, S. Gefen-Halevi, N. Shilo, A. Epstein, R. Mor-Cohen, A. Biber *et al.*, "Utilization of machine-learning models to accurately predict the risk for critical covid-19," *Internal and emergency medicine*, vol. 15, no. 8, pp. 1435–1443, 2020.
- [38] A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. MIT Press, 2002, pp. 841–848.