aurum

JOURNAL OF HEALTH SCIENCES
A. J. Health Sci.

Volume 4, No 1 | Summer 2022, 45-55

RESEARCH ARTICLE

A NEW METHOD BASED CNN COMBINED WITH GENETIC ALGORITHM AND SUPPORT VECTOR MACHINE FOR COVID-19 DETECTION BY ANALYZING X-RAY IMAGES

Karam Mohammed Dawood
Institute of Graduate Studies, Altınbaş University, Turkey
karam.dawod@ogr.altinbas.edu.tr ORCID No: 0000-0002-0596-2546

RECEIVED DATE: 21.04.2022, ACCEPTED DATE: 26.04.2022

Abstract : COVID-19 is an infectious disease caused by the most recently discovered coronavirus. This new type of virus and disease was unknown before the outbreak began in Wuhan, China in December 2019. COVID-19 poses a serious public health threat. Older adults and people with pre-existing medical conditions such as diabetes, hypertension, heart disease, chronic lung diseases, obesity are at higher risk of experiencing complications and serious illness. The computer scientist applied several machine learning and deep learning techniques to detect COVID-19 in last year. In this study, efficient COVID-19 detection framework presented to detect COVID-19 by analyzing x-ray tests. The proposed framework based CNN combined with genetic algorithm and SVM classifier. The main contribution in this study is combining CNN with genetic algorithm and SVM to detect COVID-19 with accurate estimation and minimum execution time. Several scenarios are executed to validated the presented method. Finally, the obtained results compared with several studies presented to solve this problem.

Keywords: COVID-19, Genetic algorithm, SVM, CNN.

1. Introduction

COVID-19 is a serious epidemic disease that emerged in Wuhan, China in December 2019, passed from bats to humans, spread all over the world in a very short time, turned into a pandemic and collapsed the health systems of many countries. Severe acute respiratory syndrome, COVID-19, causes severe respiratory failure in infected organisms with fatal consequences as the disease progresses. The main symptoms and signs of the disease; known as fever, dry cough, sore throat, headache, weakness, weakness, diarrhea, shortness of breath. In more advanced cases, it can progress to severe pneumonia, which causes pneumonia due to oxygen imbalance and multiple organ failure. The disease has much more dangerous and tiring consequences, especially for people with chronic diseases, people with weak endurance and immunity, smokers and the elderly (WHO, 2020), (Chen et al., 2020), (Yin and Wunderink, 2018), (Wang et al., 2020).

COVID-19 is usually transmitted through physical contact eg. B. by respiratory, hand or mucus contact with a person carrying the virus (Diprose et al., 2017). Antibiotics, antimalarials, antipyretics, cough suppressants, and pain relievers are commonly used to treat illness after infection. Hospitalization of infected patients depends on the degree, condition and severity of the disease. The number of patients infected with the COVID-19 virus around the world is increasing day by day. Even powerful countries like the United States, Italy and Spain failed to protect themselves from the virus and were badly affected. With all of this information, given the health system, early diagnosis of the disease is essential to completely prevent a pandemic, or at least to minimize potential damage from the virus. In other words: at least suspected cases must be recognized without error, quickly and accurately (Li et al., 2020).

Currently, the real-time reverse transcription polymerase chain reaction (RT-PCR) is widely used in the diagnosis and diagnosis of COVID-19. Chest x-rays such as computed tomography (CT) and x-rays are preferred for the early detection of COVID-19. The rapid spread of the disease and increasing death rates in many countries point to the need for effective treatments. Therefore, disease management, including diagnosis, early quarantine, and follow-up care, is essential. For now, AI can contribute to the above perspective. Despite strong similarities and a severe shortage of skilled workers between COVID-19 and traditional pneumonia, artificial intelligence (AI) -based self-recognition models may be an important step in significantly reducing testing time There is sex (Li and Zhu, 2020). In this context, artificial intelligence (AI) solutions, effective in solving complex problems, are essential. The solution under study offers both inexpensive and more accurate diagnostic-diagnostic treatment for COVID-19 and similar diseases. Today, especially in the medical field, positive results are obtained with deep learning techniques using imaging data sets such as retinal imaging, chest x-ray and brain MRI with deep learning. It is used in many applications to extract, analyze and recognize data patterns (Huang et al., 2020).

In this study, using the widely used deep learning approaches Convolutional Neural Networks (CNN) technique with new structure and SoftMax method, three different conditions (COVID-19, viral pneumonia, and normal) were used to diagnose COVID-19. The aim of this study is to diagnose and detect patients with COVID-19. In the second part of the study, literature review; In the third part, the details of the method and technique, in the fourth part, the experimental findings and finally in the fifth part, the results and suggestions obtained from the study are given.

2. Convolutional Neural Network Based on Support Vector Machine and Factor Analysis

In this chapter the proposed method of our research named convolutional neural network based on support vector machine and factor analysis was proposed. The proposed method consist from three stages, feature extraction using CNN, genetic algorithm applied as feature selector to reduce the size of output features of the CNN. Then, SVM applied for for classification of the extracted features.

In the first stage, AlexNet is a CNN channel hosted by Alex Krizhevsky. AlexNet was involved in ImageNet's core work on image recognition in 2012. The network was ranked in the top five errors with a 15.3% rate, down 10.8% from its completion rate.

AlexNet is the name of a convolutional network that has had a great impact on machine learning, especially in deep learning computer vision applications. As you know, we won the 2012 ImageNet LSVRC competition by a wide margin (26.2% (second place) with an error rate of 15.3%). Network - Yang LeCun et al. But LeNet was deeper in a convolutional layer that applies multiple filters layer by layer. Consists of 11x11, 5x5.3x3, line, max connection, interrupt, data increase, ReLU trigger, SGD with pulse. Activating Connected ReLU after fully connected layer with all stacking layers. AlexNet trained with two Nvidia Geforce GTX 580 GPUs simultaneously for six days, splitting the network into two channels.

AlexNet has eight levels. The convolutional layer was the first 5 geotags, some of which had the maximum level of clustering, the fc layer shown in the last three, and the activation function was not applied. Shown improved training accuracy for Thane and Sigmoid. The AlexNet structure presented in the Figure 1.

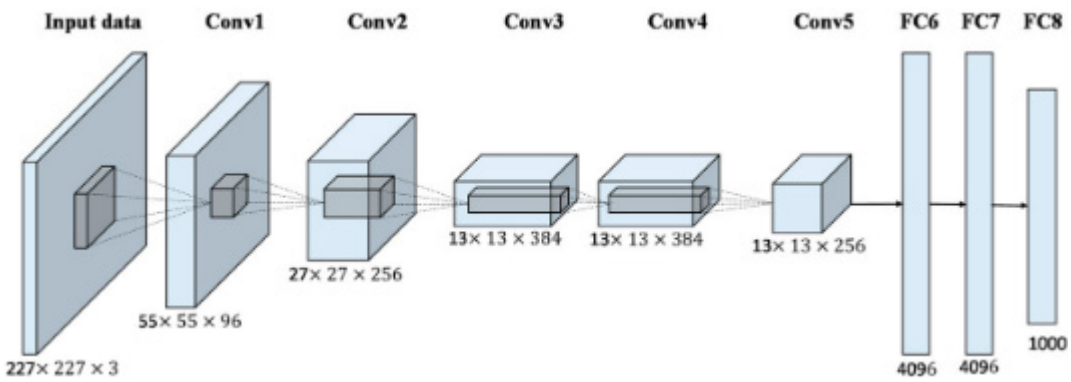

Figure 1: *AlexNet Structure*

Then, the Genetic Algorithm, used in the field of Artificial Intelligence, is a kind of algorithm that searches for the best spot. It's about finding a solution to a problem. If we can turn real problems (like containers on dry cargo ships, how to move from point to point, how to create optimal delivery routes, etc.) into research problems, we can solve this problem. Genetic algorithm's theme. A genetic algorithm (GA) is a function that performs permutation-based optimization and looks for a criterion for the likelihood of

convergence. It is a research and optimization method that works in the same way as the evolutionary process observed in nature. Genetic algorithms are described in the literature as follows: Genetic algorithms are powerful evolutionary strategies based on the basic principles of biological evolution. The researcher must first correctly identify the type of the variable and the problem with which he is dealing, and then code according to that definition. Then the fitness function is determined, which is one of the inputs to the algorithm, and this is the objective function that needs to be optimized. Genetic operators such as mating and mutation are applied stochastically at many stages of evolution, so it is necessary to determine the probability of their occurrence. Finally, we need to meet the convergence criteria and solve the problem at the lowest cost. If the problem is influenced by too many factors, the literature recommends using a genetic algorithm to solve it.. Briefly, the working principle of the Genetic Algorithm is as shown in figure 2.

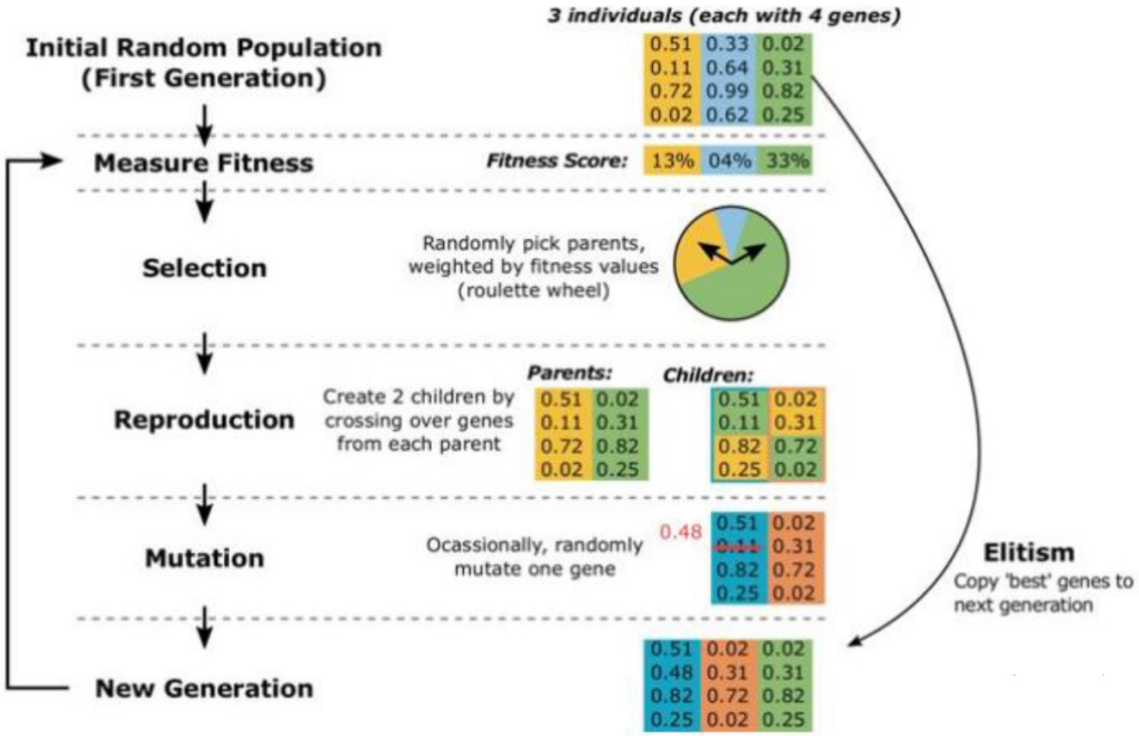

Figure 2: *GA Process.*

The flowchart of the algorithm presented in the figure 3.

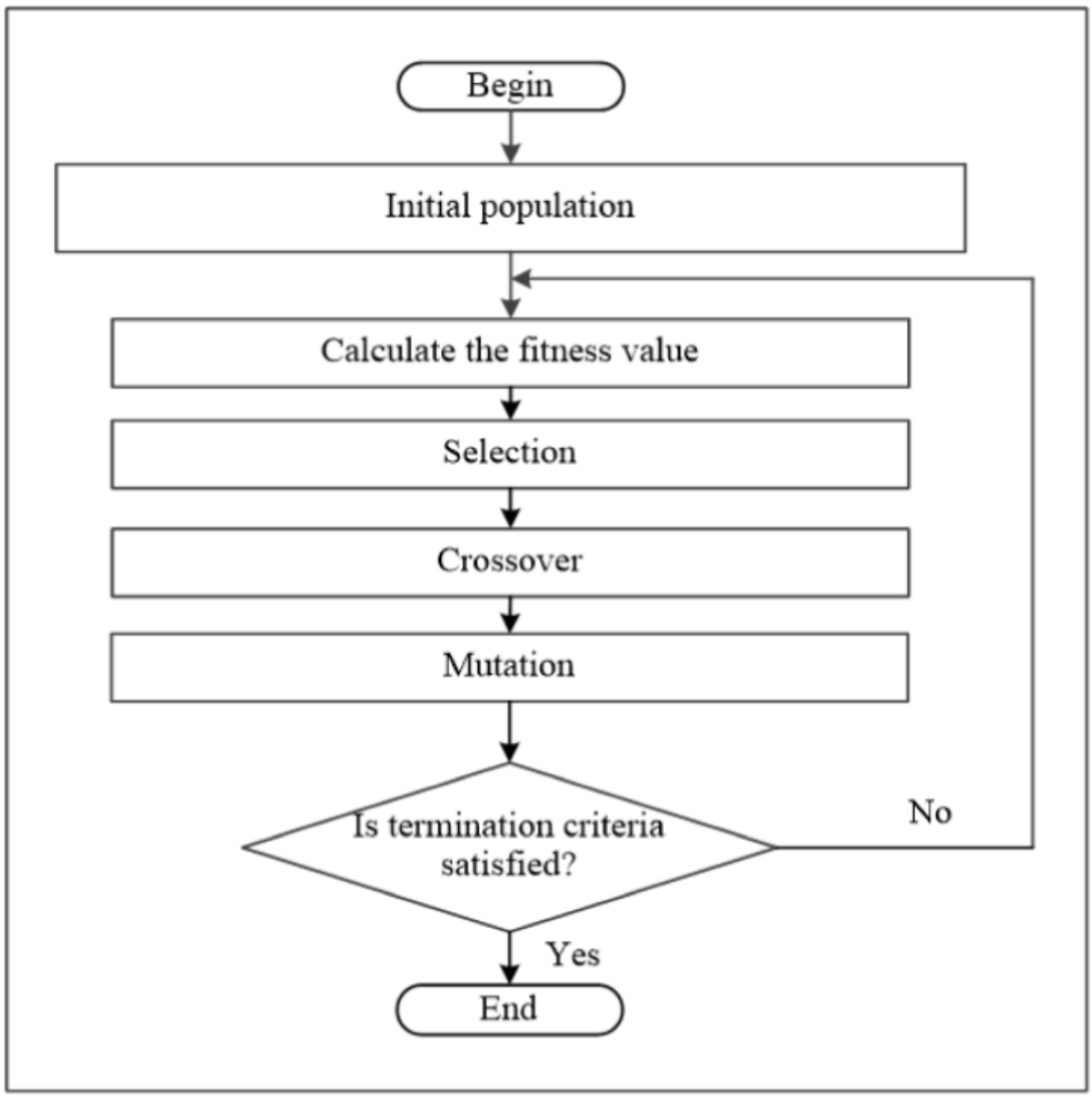

Figure 3: *GA flowchart.*

How people understand a problem depends on the problem. The most important factor that determines the success of the genetic algorithm in solving a problem is the representation of the person who offers the solution to the problem. Everyone in the population has a fitness function to decide whether

there is a solution to the problem. Individuals who score higher according to the evaluation obtained by the fitness function are capable of producing offspring with other individuals in the population. At the end of the mating process, these humans give birth to new humans called babies. The child has the characteristics of the parents (mother, father) who created it. Individuals with low fitness scores are less likely to be selected when creating new individuals, so they will be removed from the population after a while. A new population is created by increasing the number of healthy people in the old population. At the same time, this population contains most of the characteristics of well-educated individuals from the previous population. Therefore, good traits are passed down from generation to generation and combined with other good traits as a result of a genetic process. The more individuals with a high fitness value come together and create new individuals, the better a working area is obtained in the search space. In order to find the best solution to the problem; The concepts used in the genetic algorithm are used in a similar sense to the theory of evolution in biology. In natural life, populations are formed by the coexistence of individuals. The population created for the GA algorithm is formed by the gathering of many individuals, in other words, by the gathering of many possible solution candidates. Candidate solutions are kept in sequences coded according to the problem. Each element that makes up this array is called an individual, and each individual represents a specific region in the search space.

In the genetic algorithm, the first starting individuals are usually randomly generated, but this is not a requirement. Especially in very constrained optimization problems, better candidates can be created by paying attention to some of the defined constraints to create the starting individuals. As a result of exposing individuals to the fitness function process, the fitness value, which evaluates how close the solution is to the optimal solution, is determined. The genetic algorithm with the initial population created works with three evolution operators. These; selection, crossover and mutation operators. In general, each of these operators is applied to each individual of the population that will form in the next generation.

Selection is the process of selecting parent individuals to create new individuals based on their fitness values in the population. The crossover operator is applied after the selection process and expresses the mutual replacement of certain parts of the chromosomes of the parent individuals and thus the formation of individuals with new characteristics. The mutation process is the process of changing a gene in any of the chromosomes of the newly formed individual depending on the probability of mutation.

There are various methods to terminate the genetic algorithm process. These methods are; When the desired solution is found during the operation of the algorithm, when the total number of iterations defined at the beginning of the GA is reached, or when the fitness value remains constant, the solution represented by the best individual found is presented as the most suitable solution found for the problem.

In the last stage, the SVM applied to classify the selected features by the genetic algorithm and the flowchart of the method presented in the Figure 4.

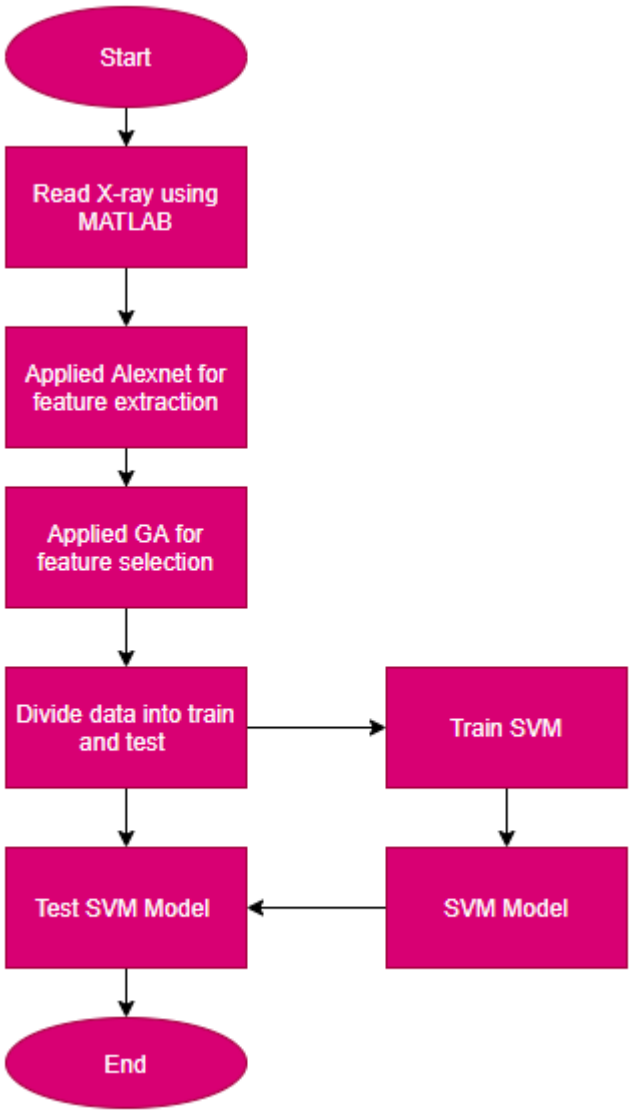

Figure 4: *Proposed Method flowchart.*

3. Experimental Results

In the chapter, the results of the proposed method presented and explained using confusion matrix to evaluate the results. The parameters that are calculated in this chapter lead to diagnosis the power and weaknesses of the proposed method.

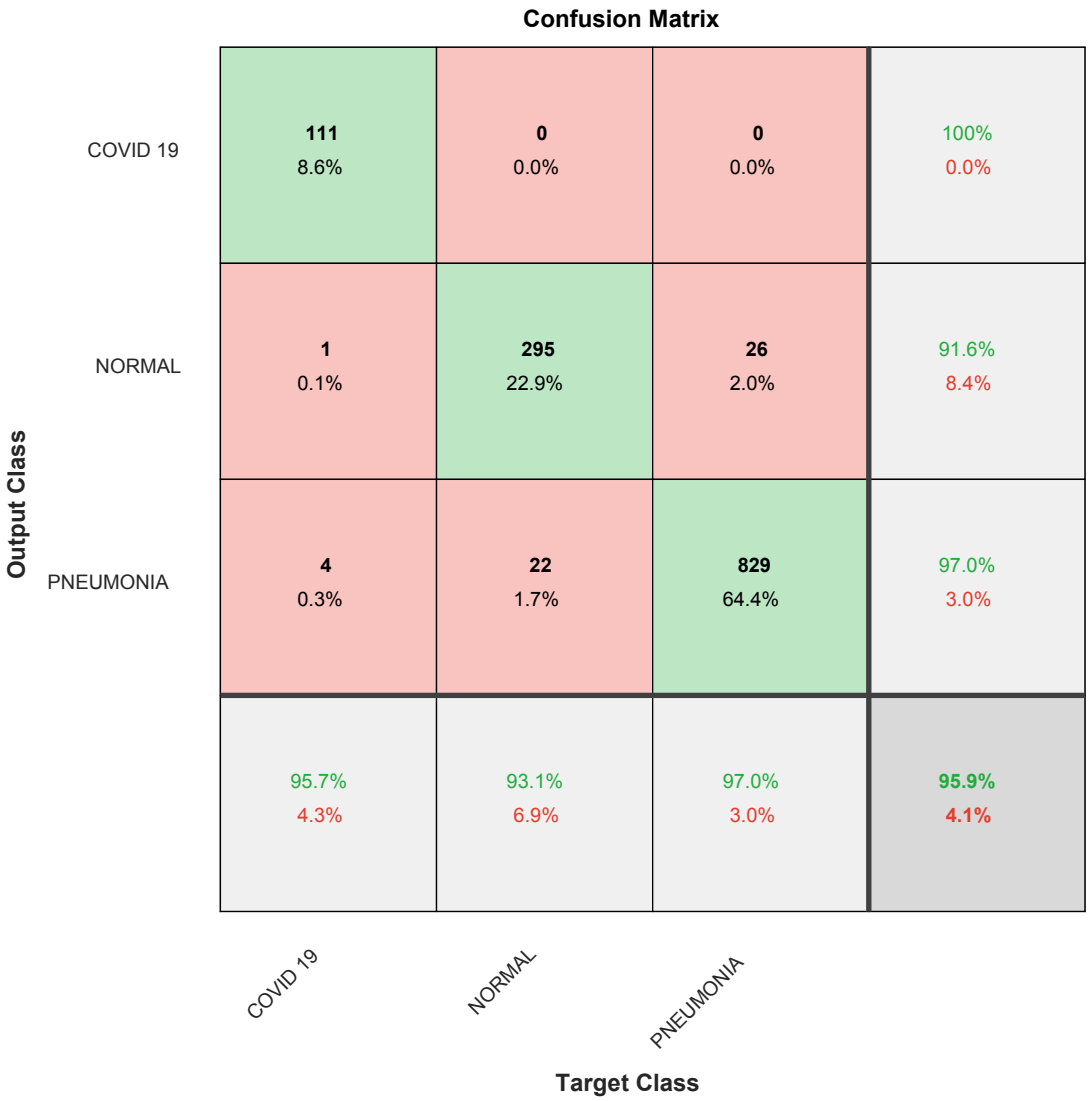

Figure 5: *Confusion matrix of our method.*

The presented framework is also compared with previous studies the proposed in related researches and is shown in Table 1.

Table1: Comparison with litreature

Ref	Acc (%)
Li and Zhu, 2020	92.3
Afshar et al., 2020	88.90
Farooq and Hafeez, 2020	95.7
Chowdhury et al., 2020	96.2
Batık et al., 2018	87.02
Gao et al., 2008	96.6
Karim and Mishra, 2022	96.67
Our Method	99.59

By analysing of the Table 1, the results show that the proposed method presented best results than other state-of- the-art studies. The presented method uses classical methods such as CNN and genetic algorithms, which can show remarkable results with a small number of datasets. On the other hand, deep learning methods require large amounts of data to perform better than other methods. Also, training is very expensive due to complex data models.

4. Conclusion

This thesis has developed a new COVID-19 detection framework that uses a SVM classifier based on CNN models and genetic algorithm. The main problem with machine learning methods is how to handle large functions with few samples. This results in model overfitting or poor performance when testing the model. Our goal was to avoid overfitting through the use of feature extraction and selection techniques. A key contribution to this research is the combination of the CNN feature extraction technique with the genetic algorithm of the feature selection technique. This combination results in active functions that are extracted from the input data related to the classifier. In addition, we have concluded that the SVM classifier gives fast and accurate results with fewer training datasets, as its modest technique is proposed to separate two discrete finite groups A and B into n-dimensional space. The combination of these methods then leads to a robust model that provides remarkable results when compared to other known studies.

In the future works, the researcher advises to apply other optimization algorithms that published in the last years instead of genetic algorithm to obtain new results which can be effective with COVID-19 detection.

5. References

Afshar, P., S. Heidarian, F. Naderkhani, A. Oikonomou, K.N. Plataniotis, A. Mohammadi et al., "-Caps: A Capsule Network-Based Framework for Identification of COVID- 19 Cases from X-Ray Images". ArXiv Prepr ArXiv200402696 2020:1–4.

Batık, Z., G. Atalı, D. Karayel, and S. S. Özkan, "Using heuristic distance functions in path planning of mobile robots," Electronics World, pp. 34–36, Oct. 2018. World Health Organization (WHO) - https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed at Feb 24, 2020).

Chen, N., M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han et al., "Epidemiological and Clinical Characteristics of 99 cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study". Lancet - 2020, 10.1016/S0140-6736(20)30211-7.

Chowdhury, M.E.H., T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub et al., "Can AI Help in Screening Viral and COVID-19 Pneumonia?" ArXiv 200313145 2020.

Diprose J, MacDonald B, Hosking J, Plimmer B. Designing an API at an appropriate abstraction level for programming social robot applications. J. Vis. Lang. Comput., 2017;39,22–40.

Farooq, M., A. Hafeez, "COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs", ArXiv:2003.14395, 2020.

Gao, Z. Zeng, X. Wang, J. Liu, J., 2008. FPGA implementation of adaptive IIR filters with particle swarm optimization algorithm. Communication Systems.

Huang, C., Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu et al., "Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan", China. Lancet - 2020.

Karim A.M., Mishra A. (2022) Novel COVID-19 Recognition Framework Based on Conic Functions Classifier. In: Garg L., Chakraborty C., Mahmoudi S., Sohmen V.S. (eds) Healthcare Informatics for Fighting COVID-19 and Future Epidemics. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-72752-9_1.

Li, Q., X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong et al., "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia". The New England Journal of Medicine. 2020.

Li, X., D. Zhu, "COVID-Xpert: An AI Powered Population Screening of COVID-19 Cases Using Chest Radiography Images". ArXiv:200403042 2020:1–6.

Yin, A.Y., R.G. Wunderink, "MERS, SARS and Other Coronaviruses as Causes of Pneumonia", Respirology, 23 (2018), pp. 130-137).

Wang, D., B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang et al., "Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan", China. Jama - 2020.