Machine Learning based Digital Contact Tracing for Covid-19 in Myanmar

Swe Swe Aung
Faculty of Computer Science
University of Computer Studies
Taunggyi, Shan state, Myanmar
swesweaung@ucstgi.edu.mm

Kyawt Yin Khaing
Faculty of Computer Science
University of Computer Studies
Taunggyi, Shan state, Myanmar
kyawtyinkhaing@ucstgi.edu.mm

May Phyo Aung
Faculty of Computer Science
University of Computer Studies
Taunggyi, Shan state, Myanmar
mayphyoaung@ucstgi.edu.mm

Nang Wai Wai Phyo
Faculty of Computer Science
University of Computer Studies
Taunggyi, Shan state, Myanmar
nangwaiwaiphyo@ucstgi.edu.mm

Darli Mying Aung
Faculty of Computer Science
University of Computer Studies
Taunggyi, Shan state, Myanmar
darlimyintaung@ucstgi.edu.mm

Thinn Thu Naing
Faculty of Computer Science
University of Computer Studies
Taunggyi, Shan state, Myanmar
thinnthunaing@ucstgi.edu.mm

***Abstract*—Digital contact tracing is one of supportive tools for the prevention and reduction of wide spread of the covid-19 virus. Machine learning is one of the most supportive approaches for contact tracing applications. Thus, this research aims to develop a web-based contact tracing system by applying machine learning approaches. The purposes of the web-based contact tracing system are to spy out the contact cases, to support a short-term prediction for Covid-19 transmission rate, to give a warning or alert message to a person, who is in outbreak zone, to give a real-time information about Covid-19 to public in time, to collect everyday patient data in the same data standardized format from anywhere in Myanmar, and provide a richest dataset to other researchers via this website. The system is mainly composed of two sub-systems, one is digital contact tracing and other one is a short-term covid-19 transmission rate prediction. Digital contact tracing will give a prediction of possible hidden cases infected by a confirmed case by applying Neural Network algorithm and using confirmed-case-id, contact-date, contact-time-duration, contact-frequency, and contact-place.**

Keywords— Machine Learning; Contacting Tracing for Covid-19; A short-term prediction for Covid-19 transmission rate

I. INTRODUCTION

Since last year December 2019, the Covid-19 has been spreading through the world with a high speed leaving much collapse in every sectors of our society. According to the live statistics and coronavirus news on 21th November, 2020, the covid-19 is effecting 218 countries including Myanmar, reaching over 75 million cases and over 40 million recovery. Thus, the statistic notifies us that the virus spreading is still continuing with a non-stop speed among us. In the case of Myanmar, Myanmar is one of the most virus effected countries in Asia with over 1000 cases since early August 2020.

That is why, World Health Organization (WHO), national authorities around the world, scientists, and clinicians combats the Covid-19 with many possible and effective ways [1]. Among those ways, contact tracing is one of an effective prevention approaches from the rapid distribution of Covid-19. The actual situation of the epidemiology of Covid-19, the transmission rate, the identification of case clusters could be known in time and could be distributed to public via contact tracing tools. ECDC [3] stated that contact tracing is a core public health intervention that plays an important role in the control of COVID-19. According to ECDC, contact tracing has three steps. These are contact identification, contact listing, and contact follow-up. ECDC describes contact tracing methods to scale up the contact tracing efficiency. As we know, the conventional method is by following up cases and contacts with health care staffs. Other methods stated in [3] are using non-public-health staff and volunteers, repurposing existing resources, reducing the intensity of follow up of contacts, and using technology. Technology-based softwares (web-based and mobile applications) are playing an important role to help the contact tracing process to go on in effective and better way.

These days are technologically developed eras. Therefore, using the advantages of technology and based on acquiring general concepts from the previous stored events or processes, the estimation of future events could be established with more precisely and higher computation speed than before. Thus, machine learning arrived to us to provide these necessary abilities for dramatic future event predictions. Machine learning is a pack of classification and prediction algorithms for learning problems. As we know, machine learning algorithms that seem like an astrology make a prediction of best fit category for every newcomer by modeling a learning problem based on stored dataset.

Thus, it is not amazement that being able to help human's lives in different ways appeals to many researchers to dive into the field of machine learning. The authors [4] designed a digital contact tracing system to predict the trends and possible stopping time of the current COVID-19 outbreak in Canada and around the world applying deep learning method. T.Chakraborty and I.Ghosh [4] developed ARIMA and wavelet-based forecasting techniques-based real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases system that gives a short-term forecasts of the future COVID-19 cases for multiple countries and risk assessment (in terms of case fatality rate) of the novel COVID-19 using five univariate time-series datasets collected for the real-time prediction purpose of COVID-19 cases for India, Canada, France, South Korea, and the UK.

This research proposal aims to develop a web-based contact tracing system using ML algorithms to spy the contact cases, to support a short-term prediction for Covid-19

transmission rate, to give a warning message to a person, who is in outbreak zone, and to give a real-time information about Covid-19 to public in time. The web-based contact tracing system will be designed as a system that can be accessed anytime and anywhere via any device and any platform.

Another problem is that health care centers in Myanmar scatter in every state and division and they individually collect and store everyday data in their own way. That is why, it is very difficult to give real-time information to public in time and to forecast a short-time virus transmission rate because of scattering of everyday infected patient data. For this case, the contact tracing website will allow health centers to store everyday patient data in the same data standardized format from anywhere in Myanmar. Because of that, the public will get to know the current situation of the whole Myanmar in time. Besides, this method will solve the troubles in collecting everyday patient data in remote region. Moreover, for the short-term prediction, the system could support an accurate short-term forecast for the future Covid-19 transmission rate based on a richest and completed dataset.

The rest of this paper is organized as follows. Section II describes related works. Section III presents Covid-19 in Myanmar. Section IV discusses machine learning technology, and the proposed system is presented in Section V. Finally, Section VI is the conclusion.

II. Related Works

This session will discuss the machine learning area in health care systems. S.Lalmuanawma [1] reviewed the role of AI and ML as one significant method in the arena of screening, predicting, forecasting, contact tracing, and drug development for SARS-CoV-2 and its related epidemic. In his review, machine learning models are designed for Covid-19 screening with clinical and mammographic data and the models outperform the prediction process with over 90% accuracy. Another mention in his review paper is that machine learning-based models for predicting and forecasting Covid-19 contribute to Brazil with around 90% accuracy. The Canadian researcher designed a prediction model using time-series and deep learning algorithms for a long-short-term memory network.

S.S.Aung, I.Nagayama, and S.Tamaki [7] proposed a prediction model for healthcare experts using a dimensional tree based Dual-kNN. In that research work, the experimental results by conducting six different datasets (breast cancer, heart stalog, parkinsons, darmetology, thyroid, and abalone) shows an expected high classification accuracy with total average 94%.

S.S.Aung, I.Nagayama, and S.Tamaki [8] designed another detection model for healthcare experts using a new machine learning approach, Dual-kNN. As stated in their research works, the experimental results by conducting six different datasets (breast cancer, heart stalog, parkinsons, darmetology, thyroid, and abalone) proves that the detection model is helpful for healthcare workers.

Anhvinh Doanvo, et al.,[9] found out the key discrepancies between literature about COVID-19 and the expectation of research on other coronaviruses by analyzing the massive amount of information applying artificial intelligence (AI)/machine learning techniques. These discrepancies are the lack of basic microbiological research, and raise questions regarding the research community's ability to quickly respond to future crises. This research work intends to identify knowledge gaps and inform resource allocation decisions for research during future crises.

A.M.U.D.Khanday, et al.,[10] designed a Covid-19 prediction system with aid of AI technologies. According to the research paper, the system classifies textual clinical reports into four classes by using classical and ensemble machine learning algorithms. Then, feature engineering is performed using techniques like Term frequency/inverse document frequency (TF/IDF), Bag of words (BOW) and report length. These features are supplied to traditional and ensemble machine learning classifiers. As discussed in that research paper, logistic regression and Multinomial Naı¨ve Bayes achieve better results than other ML algorithms by having 96.2% testing accuracy.

M.K. Siddiqui, et al. [11] found out the correlation between temperature and different cases situation (suspected, confirmed, and death cases) with help of k-means clustering-based machine learning method by conducting the data set from different regions of China, which has been obtained from the WHO. The authors had a deep research in the temperature field in the original WHO data set and further an exploration of the effect of temperature on each region in three different perspectives of COVID-19 – suspected, confirmed and death.

Y.Zoabi, et al. [17] proposed the machine-learning model that predicts a positive SARS-CoV2 infection in a RT-PCR test. They wanted to assist the health staff worldwide in triaging patients and that model can be implemented globally for effective screening and prioritization of testing for the virus in the general population. In that paper, the COVID-19 test results were predicted with high accuracy using eight binary features.

In paper [18], they focused the contributions of deep learning techniques at several scales including medical imaging, disease tracing, analysis of protein structure, drug discovery and virus severity and infectivity to control the ongoing outbreak. They tried to explore and discuss the overall applications of deep learning on multiple dimensions to control novel coronavirus (COVID-19).

III. 2020 Covid-19 in Myanmar

As stated in [1], the whole world is struggling with the biggest health problem COVID-19 name coined by the World Health Organization (WHO), raised from China in December 2019. End of March 2020, the Covid-19 arrives with a slow spreading in Myanmar. The first and second Covid-19 cases were confirmed in Chin state. As a good preparation and strategy of the Health and Sports Ministry, Myanmar successfully fought against the first Covid-19 wave, though a huge amount of people returned from Thailand, UK, etc. Until August 2020, the wide spread of the virus could be controlled well with the aid of local communities without having a large number of confirmed Covid-19 cases. In the mid of August 2020, the virus have already reached among Myanmar people in the regions of Yangon and Rakhine. From that time, the virus has been spreading with a high transmission rate around the country.

Fig.1 demonstrates the situation of the Covid-19 spreading around the country wide. According to the figure, it can be

guessed that north, south and east of Myanmar are not serious as west and middle of Myanmar concerned with the virus spread. Table I illustrates the total number of cases, death, and recovery. As stated in the table, the highest number of cases can be found in Yangon, and the

TABLE I. CASES, RECOVERY, AND DEATH, [12] ON THE DATE (27 NOVEMBER 2020)

Region	Confirm Cases	Deaths
Yangon	40103	1166
Rakhine	2877	16
Bago Region	2669	3
Mandalay Region	1841	7
Ayeyarwady	1146	6
Magway Region	695	1
Mon State	623	9
Sagaing Region	383	1
Kayin State	304	5
Naypyidaw Union Territory	297	3
Tanintharyi Region	226	0
Kachin	140	0
Shan	122	1
Chin State	66	1
Kayah State	4	0

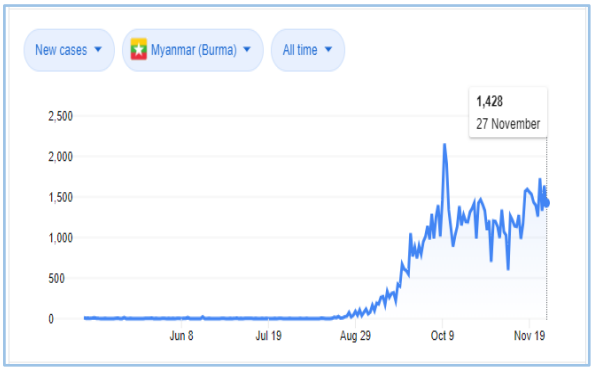

Fig. 2. Each day new cases reports on the date (27 November 2020) [12]

lowest region is Chin state, where the first case was found. The highest death rate can also be found in Yangon. The non-death rate regions are Kachin and Tanintharyi regions. As shown in Fig.1, the everyday new cases are increasing with some fluctuations. It means that the transmission rate is still going with a high speed in Myanmar. According to the data shown in Fig. 1, the total numbers of tested people are 1,114,697 cases, the total PUI 15,581 people, lab confirmed 86,633, 1,856 death, and 65,972 recovered people. This data shows that there is no big gap between confirmed cases and recover.

For the subject of education in Myanmar during the pandemic, as we know, all school has been closed to avoid and prevent the spread of virus around the country wide. But, the education sector has been utilizing this pandemic time as a tool for revising and updating the old curriculum, the preparation for online lecture teaching, a basic e-Learning management system training course from UCSY, holding online conferences around the country (MURC, ICAIT, etc.), 2019-dissemination seminars including teaching/learning for tomorrow, digitalization in higher education, university internationalization, comprehensive university, university industry link, internal quality assurance and research capacity building, and internal and nationwide online programming contests through digital platforms(Microsoft team, Zoom, etc). For university capacity building, staffs are given a chance of attending courses, for example Coursera, and internal and external research collaborations.

IV. MACHINE LEARNING TECHNOLOGY

Here let us to discuss briefly the mysterious natural processes of machine learning. Machine learning is a pack of classification and prediction algorithms for learning problems. As we know, machine learning algorithms that seem like an astrology make a prediction of best fit category for every newcomer by modeling a learning problem based on stored datasets. Another key thing to remember is that machine learning algorithms are powerful tools for a variety of application domains giving widely divergent dimensions such as reliable, precision, robustness, high speed solution, etc. By utilizing the advantages of machine learning and technology, the estimation of natural disaster including rainfall estimation, etc., classification of disease based medical records, prediction of traffic congestion, crime detection using videos surveillance, speech recognition including text-speech and speech-to-text for enable person, detecting virus, e-mail spam and malware filtering in social media services [14] can be successfully supported for helping out in making human being's lives better with the great assistance of machine learning. The integral tasks of machine learning can be naturally categorized into five groups as follows:

1) Classification
2) Clustering
3) Regression
4) Density estimation
5) Dimensionality reduction

Classification is a process of discovering and categorizing objects from large data storage that have similar characteristics, properties, and patterns. Classification learning is sometimes called supervised learning because, in a sense, the scheme operates under supervision by being provided with the actual outcome for each of the training examples. Some classification algorithms are Naïve Bayes, k-Nearest Neighbor, Decision Tree, Support Vector Machine, Logistic Regression, and Neural Network.

Clustering is basically defined as an unsupervised learning method. An unsupervised learning method is a method that explores the instance throughout unlabeled dataset by inferring their characteristics and similarities via distance matric. The function of clustering is the collection of objects into the same group based on their similarity (or distance) among them. Thus, the objects that are in the same group are more likely to be different from the objects of other groups [15].

Regression is one of the most widely used prediction method in machine learning for analyzing categorical

dependent variables by estimating the probability of dependent variable based on one or more independent variables. In other words, regression finds out the best appropriate model to define the relationship between dependent variables and independent variables. The commonly used regression methods are linear regression and logistic regression.

In machine learning, dimension reduction is the function of reduction the attributes, also known as variable, in dataset aid to upgrade the data quality. Feature selection and feature extraction are the primary schemes of dimension reduction. Some part of this research mainly applied feature selection technique for the purpose of removing noisy instances, reducing storage and improving computation time. Feature selection, by identifying the most salient features for learning, focuses a learning algorithm on those aspects of the data most useful for analysis and future prediction [16]. A technique for selecting features is to evaluate the correlation among features.

V. DIGITAL CONTACT TRACING AND SHORT-TERM TRANSMISSION RATE WITH ML ALGORITHMS

This research aims to develop a web-based contact tracing and short-term transmission rate using ML algorithms. The web-based contact tracing system will be designed as a system that can be accessed anytime and anywhere via any device and any platform. Health care centers in Myanmar scatter in every state and division and they individually collect and store everyday data in their own way. Consequently, it is very difficult to give real-time information to public in time and to forecast a short-time virus transmission rate. For these cases, the web-based contact tracing system will allow health centers to entry and store everyday patient data in the same data standardized format from anywhere in Myanmar. Besides, the system will support the public to know the current situation of the whole Myanmar in time. Furthermost, this method will cover the troubles facing in collecting everyday patient data in remote region.

Fig. 3 is a proposed system for digital contact tracing for Covid-19 with ML approach in Myanmar. The proposed system is categorized into four sub-systems. These are data standardization, contact tracing, transmission rate prediction, the provision of up-to-date Covid-19 data to other researchers, and location-based messaging awareness of outbreak zone via the website.

A. Covid-19 Data Standardization

Data standardization means that the data comes from various sources are mapped into a defined standard format for further analysis processes. In this research work, the Covid-19 data will go under data standardization processes aiming for gaining accurate prediction results. Data standardization is composed of data cleaning, data transformation, data integration, discretization, attribute construction, generalization, and normalization components.

For data transformation case, this work will generate a patient id for each entry instead of using patient's name. For example, for patient U Ba, the system will automatically parse patient's name into patient id (c_21_01_00001) with the defined format (c_yy_mm_00000). In this case, c means covid-19 patient, yy and mm are related year and month.

Discretization method is for transforming the continue data into discrete format. Patient age is in the form of continuous data format. Thus, for the case of age group, this research will utilize discretization method to categorize the patient's age. For example, (0-2,3-39,40-59,>60) is mapped into (baby, young adults, middle-age adults, and old adults).

TABLE II. SAMPLE COVID-19 STANDARD DATA FORMAT

Component of Covid-19 Patient	**Source Data Format**	**Targeted Data Format**
Name	U Mg	c_yy_mm_id (eg. C_21_01_19)
Address		1030714
State/Division	Yangon	01
Township	Hlaing	03
Ward	HninSiGone	07
Street	street_3	14
Travel History		0101
Country, State/Division	Myanmar, Yangon	0101
Contact Status	Primary/Secondary	p/s
Confirm Status	Positive/Negative/Dead	p/n/d
DOB	DD/MM/YYYY	DD/MM/YYYY
Age		middle -age adults
	Value (eg. 56)	baby
		young adults
		middle -age adults
		old-age adults

For patient's address, the system will map patient's source data format (Hninsigone, Hlaing Township, Yangon) into the targeted data format (State/Divisio_Township_Ward_Street, 03_07_01_00). In this format, 00 is unknown value. Table II. demonstrates the sample covid-19 standard data format by mapping source data into the defined standard format.

B. ANN-based Digital Contact Tracing for Covid-19

The following figure illustrates the neural network model for digital contact tracing. The model takes fime input attributes (contact-id, contact-date, contact-timeduration, contact-frequency, and contact-place) at input layer and provides the probability of infected people by a confirm case as shown in Fig. 2.

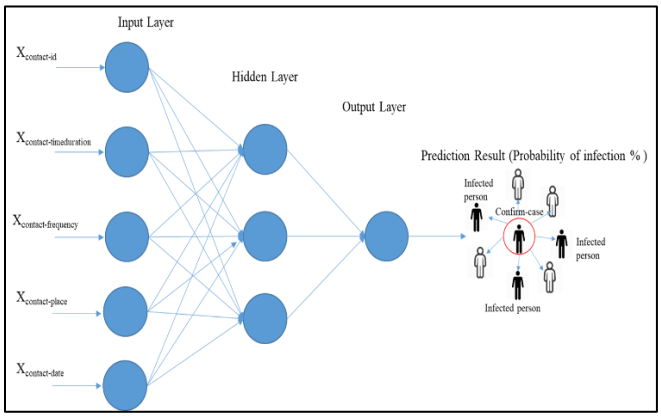

Fig. 2. Neural Netork Model for Digital Contact Tracing

Covid-19 patient data is critical to trace and reveal the clue of root cause of virus and to prepare some preventions of spreading the virus. It is very important to make sure that data collection of a healthcare center is in harmony with other healthcare centers around the country [13]. Thus, the provision of accessing the right and efficient patient data is key factor for researchers to conduct experimentations accurately and timely. For this case, in this research works normalization methodologies will be considered to contribute the formation of the Covid-19 patient data characterizations in the development of hypothesis regarding Covid-19 disease.

Warning technologies is playing an important role in preventing and delaying the virus transmission. Many countries around the world have developed many mobile and web applications to support accurate and right information to public, to alert a warning message to a person, who is in outbreak zone, and to wear safeguards.

To reduce and delay onward virus transmission, several techniques are utilizing in several countries. The popular techniques are full-lockdown, semi-lockdown, contact tracing, digital contacting, and social distancing. Among these, full-lockdown and semi-lockdown are required to be able to control the virus spread, but for long duration, these strategies are harmful to mental health, physical health, local economy, and disruption to education. To deal with these

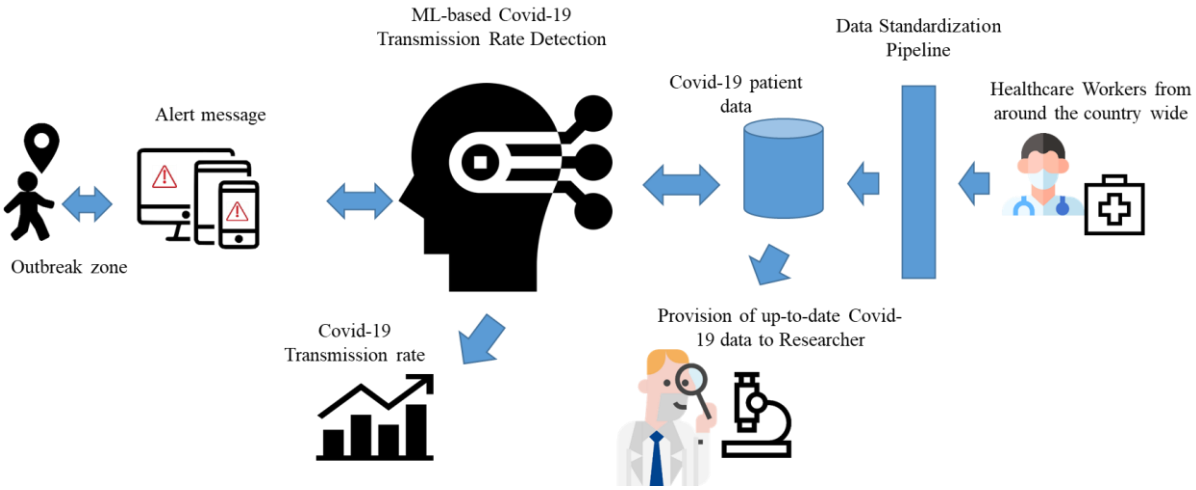

Fig. 3. Digital Contacting Tracing for Covid-19 with ML approach in Myanmar

issues, digital contact tracing is a more relevant application for delaying virus transmission rate. Based on the contact case pattern, probability of public transportation usage, geographical structure of a city, population, temperature are included in the elaboration of hypothesis for the prediction of Covid-19 transmission rate. In this research, the current virus transmission rate and the previous described properties of historical datasets will be combined with machine learning models for the future virus transmission rate prediction. Providing the probabilities of onward virus transmission rate to public in time and on time is the primary key for the preparations of health supplements, self-testing, communities testing, awareness of social distancing, and setting stay at home policy.

VI. Conclusion

This paper proposes the web-based digital contact tracing system that will give hidden infected cases based on the current up-to-date dataset applying one of machine learning technologies, ANN algorithms. Another module of this research work is to send a warning message to a person who is in serious region via the contact tracing website. The third module is a data standardization module to map the data from various health care center into a defined data format. This standard covid-19 dataset will be available for other researchers after implementing this system. The fourth module is to offer the up-to-date and accurate covid-19 information to public on time during the pandemic. This paper is ongoing process and as a future work, we are going to implement this system to be able to help people, researchers, and healthcare workers by establishing university internal collaboration soon.

References

[1] S.Lalmuanawma, J.Hussain, L.Chhakchhuak, "Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review", Journal of Chaos, Solitons and Fractals 139 (2020), 23th June, 2020.

[2] C.E.Juneau, A.S. Briand, T.Pueyo,P.Collazzo,L.Potvin,"Effective Contact Tracing for COVID-19: A Systematic Review", the preprint server for health science, July 25, 2020.

[3] ECDC technical report,"Contact tracing for COVID-19: current evidence, options for scale-up and an assessment of resources needed", European Center for Disease Prevention and Control, April 2020.

[4] V.K.R.Chimmul, L.Zhang,"Time series forecasting of COVID-19 transmission in Canada using LSTM networks", Journal of Chaos, Solitons and Fractals , Vol.135, June 2020.

[5] A.A.Enughwure, and I.C.Febaide,"Applications of Artificial Intelligence in Combating Covid-19: A Systematic Review", Open Access Library Journal, Vol.7, August 2020.

[6] A.Bagaria, et al., "Application of artificial intelligence in prevention and detection of Covid- 19", International Journal of Applied Dental Sciences ,Vol.6, No.4, pg.285-288, September 5, 2020.

[7] S.S.Aung, I.Nagayama, and S.Tamaki, "A High-Performance Classifier by Dimensional Tree based Dual-kNN", IEIE

Transactions on Smart Processing and Computing, Vol. 1, No. 1, July 2012.

[8] S.S.Aung, I.Nagayama, S.Tamaki,"Dual-kNN for a Pattern Classification Approach", IEIE Transactions on Smart Processing and Computing, vol. 6, no. 5, October 2017.

[9] A.Doanvo, X.Qian,D.Ramjee, H.Piontkivska,A.Desai, M. Majumder, "Machine Learning Maps Research Needs in COVID-19 Literature", Journal of SciencDirect ,16 September 2020.

[10] A.M.U.D.Khanday, S.T.Rabani1, Q.R.Khan, N.Rouf,M.M.U.Din, "Machine learning based approaches for detecting COVID-19 using clinical text data", Vol.12(3), page-731–739, September 2020.

[11] M.K. Siddiqui, et al., "Correlation Between Temperature and COVID-19 (Suspected, Confirmed and Death) Cases based on Machine Learning Analysis)", Journal of Pure and Applied Microbiology, Vol.14,page-1017-1024, April 2020.

[12] "Coronavirus Disease 2019 (COVID-19) Surveillance Dashboard (Myanmar), the republic of the union of Myanmar, the ministry of health and sport."

https://mohs.gov.mm/Main/content/publication/2019-ncov

[13] "Global COVID-19 Clinical Data Platform for clinical characterization and management of hospitalized patients with suspected or confirmed COVID-19", World Health Organization, 2020.

https://www.who.int/teams/health-care-readiness-clinical-unit/covid-19/data-platform

[14] S.J.Delany, al etc., "SMS spam filtering: Methods and data", Elsevier Expert Systems with Applications, Volume 39, Issue 10, August 2012, Pages 9899-9908.

[15] Lan H. Witten, Eibe Frank, University of Waikato, New Zealang, "Data Mining : Practical Machine Learning Tools and Techniques with Java Implementations", 2000, ISBN 1-55860-552-5.

[16] M. A. Hall, Department of Computer Science, the University of Waikato, Mamilton, NewZealand, "Correlation-based Feature Selection for Machine Learning", thesis book submitted in April, 1999.

[17] Y.Zoabi, S.D.Rozov, and N.Shomrom, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms", npj Digital Medicine 4, Article number: 3 (2021)

[18] A. Asraf, et al., " Deep Learning Applications to Combat Novel Coronavirus (COVID-190 Pandemic", Springer Nature Computer Science (2020).