IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Implementing Of Machine Learning Techniques For Severity Classification Of Chikungunya Disease

[1]Dr.P.M.S.S.Chandu, [2]N.Jhansi,[3]A.Neeraj Kumar, [4]T.Nikhil Sai, [5]B.Niranjan Reddy

[1]Professor, [2-5]Student

[1-5]Department of Computer Science and Engineering,

[1-5]Siddharth Institute of Engineering & Technology, Puttur, India

Abstract: As a result, AI methods are rapidly being applied to healthcare issues. In recent years, scientists have employed deep learning methods to categorize the severity of Chikungunya infection. Overfitting and hyper-parameter tweaking are problems that arise when using these methods, though. Methods: In this study, we propose a cyber-physical system (CPS) powered by artificial intelligence for determining how severe cases of Chikungunya are. Computational algorithms are combined with the CPS system's physical components to improve performance. Random forest is used to create a model for determining how severe cases of Chikungunya sickness are (RF). Overfitting and slow computing are problems plaguing RF, however, because of its complicated design and huge number of connection weights. Thus, we propose a genetic method with adaptive crossovers to be used in a developing RF model (ACGA). ACGA has the ability to effectively optimize RF design, leading to improved outcomes and increased computing speed. Multiple tests are run on the Chikungunya illness dataset. To sum up, the examination of performance reveals that ACGA-RF is superior to competing models in terms of F-measure, accuracy, sensitivity, and specificity. Patients who reside distant from hospitals will still have access to medical care according to the CPS proposal.

Index Terms - Machine Learning, Random Forest, Logistic Regression, Decision Tree, adaboost, ML techniques, e-learning, evaluation.

I INTRODUCTION

Healthcare systems are crucial to every country's progress. The government must create effective healthcare measures to safeguard its population against epidemics. Any new illness epidemic, such as the coronavirus, places a significant strain on the healthcare system. There are a number of viruses that may infect both humans and animals. The fast spread of Chikungunya is a potential disaster for the healthcare sector. This virus is transmitted to humans mostly via the bites of two mosquito species: Aedes albopictus and Aedesagypti.

Chikungunya is characterized by painful joints, a very high temperature, and a rash. Some infected people have conjunctivitis, headaches, tiredness, and gastrointestinal issues. Chikungunya and dengue are transmitted by the same vector and have comparable symptoms. Chikungunya, on the other hand, causes far more intense joint pain and eye redness. Dengue fever doesn't present with sore throat-like symptoms. The fatality rate from Chikungunya is debatable. Literature shows that full recovery from this illness may be expected in about a week. However, you could be dealing with joint discomfort for a while. Predicting the onset of chronic illnesses, their course, and the factors that contribute to their development are all within the realm of machine learning's capabilities. Unique and directly applicable to bettering clinical judgment and the layout of healthcare facilities, the results are promising.

To begin therapy, clinicians must identify the source of the patient's distress. It is not feasible to precisely quantify these symptoms, however. As a result, it's possible that people won't benefit from therapy. Chikungunya is diagnosed using a combination of a test called reverse transcription-polymerase chain reaction (RT-PCR) and a test called serological testing. Patient blood samples are needed for both examinations. But the performance of these tests for this condition is unreliable. Taking into account patient symptoms and laboratory testing, supervised learning methods like machine learning and deep learning can assess the severity of this condition.

However, it is still an ill-posed challenge to determine how seriously an individual has been infected with Chikungunya. In, we provide the first comprehensive fog-based framework for the rapid and accurate detection of Chikungunya illness. Patient cases of Chikungunya infection were categorized using J48. In, a fog-based classification and control system for Chikungunya illness was developed using wearable IoT devices. The virus Chikungunya was classified using a FuzzyC means (FCM) classifier. In contrast, overfitting and difficulties tweaking hyper-parameters plague J48 and FCM. In, an ANFIS (PANFIS) model based on particle swarm optimization was used to detect cases of Chikungunya fever.

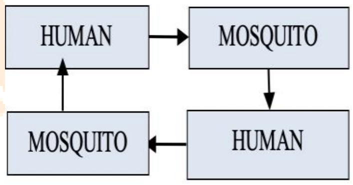

Fig 1 Chikungunya Transmission Cycle

It is possible that the virus may be transmitted from one person to another if they were to be bitten by an infected mosquito. Transmission occurs in two stages, first from humans to mosquitoes and then back to humans, as seen in Fig. 1. Patients with the infection were first classified using an ANFIS classifier. The difficulty in properly adjusting ANFIS's parameters was then tackled by use of particle swarm optimization (PSO). In comparison to other types of AI, it performed very well (ANN). But PANFIS has the over-fitting issue. There is the possibility of PSO being caught in local optimums and experiencing premature convergence.

It is common practice for doctors to manage this condition by monitoring their patients for symptoms, although these observations are not always reliable. Therefore, it becomes difficult to accurately diagnose or treat this illness. Taking into consideration the indications and symptoms of this condition, an expert system, which is used to help the human decision-making process, might be looked at as an alternative to the traditional method of assessment. Chikungunya risk assessment in this setting may make use of rule-based expert systems.

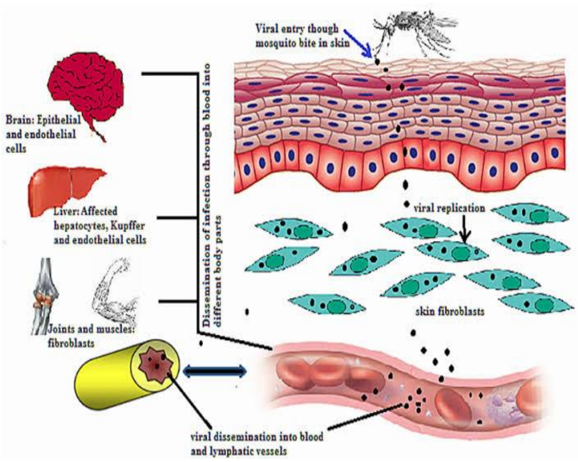

Fig 2 Schematic representation

The over-fitting and hyperparameters tuning issues are addressed by proposing an efficient evolving Random Forest (RF) model for Chikungunya illness severity classification. The following are the most significant results of this study: 1) We present a Chikungunya illness severity classification model that uses data from cyber-physical systems (CPS). Combining physical parts with computer techniques, as in a CPS system, allows for improved performance. Chikungunya illness severity is to be classified using an evolving RF model, as stated in the second hypothesis. The RF model is evolved using an ACGA, a genetic algorithm with adaptive cross-over. Comparison of the ACGA-RF with a deep learning model for Chikungunya illness severity classification.

II LITERATURE SURVEY

During the course of our research, we looked at the following articles and projects on the topic of machine learning techniques for chikungunya disease severity classification.Majhi, S. K., [1] (2019). An automotive insurance fraud detection system using a fuzzy clustering method based on a modified while optimization algorithm. Intelligent Evolution: A 1–12 Page Overview. The clustering process is carried out using the fuzzy c-means (FCM) approach. Out of all the other clustering methods, this one is by far the most used. But it has a bad habit of staying at the local best. When searching for global optimum solutions in a given dataset, a stochastic global optimization technique known as the whale optimization algorithm (WOA) might be quite useful. Improvements are made to the WOA so that it can find a better global optimum. By combining the benefits of the MWOA with the FCM, this research suggests a new approach to fuzzy clustering. Popular already-existing measures are taken into account to determine how effectively the suggested clustering method performs. In the proposed auto insurance fraud detection system (AIFDS), the hybrid clustering approach based on MWOA is used as an under sampling technique to optimize the cluster centroids.

Carneiro, E. M., Dias, L. A. V., et al.,[2] (2015, April). Examining the effectiveness of cluster analysis and artificial neural networks in identifying credit card fraud. To be held in 2015, the 12th Annual International Conference on Information Technology: New Generations (pp. 122-126). IEEE. With the rapid expansion and widespread use of internet-based transactional events in the financial industry comes the increase of fraudulent actions like fraud that cause monetary loss. There is no consistent pattern to the manner in which criminals operate; instead, their methods, strategies, and even their very personalities evolve with the rapid development of new technologies. Every time a new piece of technology hits the market, the hoaxer community adapts to it by learning about it and then using that knowledge to spread their harmful wares online. A con artist will study consumer trends in order to devise an effective strategy for deceiving their victims and making off with their money. With so much money being lost to fraud, hoaxes, and the like in the financial sector, it is imperative that a fraud detection system be developed to detect such behaviors in online monetary transactions via the use of machine learning.

Duan, L., Xu, L., et al [3] (2009). Identification of anomalies using a clustering algorithm. 168(1), 151-168 in the journal Annals of Operations Research. Important data mining applications include fraud detection, consumer behavior research, and intrusion detection, all of which rely on the ability to spot outliers. An outlier is an item in a data collection that is extremely dissimilar to, or inconsistent with, the rest of the data. While it is common to think of an outlier as a single instance, it is an important fact that many anomalous occurrences have temporal and geographical localization, which may create tiny clusters that must also be regarded outliers. That is to say, it's possible for a cluster of data points, rather than just a single one, to constitute an anomaly. In this study, we provide a revised concept of anomalies: cluster-based outlier, which is significant and gives attention to local data behavior, and how to find outliers using a clustering method. LDBSCAN.

Minastireanu, E. A., & Mesnita, G. [4] (2019). A lightweight gbm machine learning method for identifying click fraud in the digital realm. To cite this article: J. Inform. Assur. Cybersecur., 2019. There are already enough of dangers associated with internet marketing, the advertising sector, and e-business without adding fraud to the mix. When it comes to marketing online, click fraud is seen as a major problem. Online marketers are always seeking for new and better ways to protect themselves against click fraud, despite ongoing attempts to enhance traffic filtering algorithms. Therefore, a reliable fraud detection algorithm is crucial for marketing firms operating in the digital sphere. Our paper's goal is to evaluate the effectiveness of a contemporary machine learning algorithm for detecting click fraud in an online setting. Click patterns on a dataset with 200 million clicks over four days have been the focus of our investigation. The primary objective was to analyze click paths throughout their portfolio and identify IP addresses that generate many clicks but do not convert to app installs.

Y. Lin, X. Lu,et al.,[5] ‘‘Proc. 1st International Symp. Future Information Common Technology Ubiquitous HealthCare (Ubi-Health Tech), July 2013, pp. 1-5. Personal health care monitoring and emergency response methods. There were instances of a novel coronavirus (2019-nCoV) of pneumonia in Wuhan, Hubei Province, in December 2019, and on 12 January 2020, the WHO officially called it "2019 novel coronavirus (2019-nCoV)". Since this new respiratory illness has the potential to spread rapidly and extensively, it has captured the world's interest, despite the lack of a comprehensive guide to its prevention, diagnosis, and treatment. An urgent need exists for the development of an evidence-based guideline for the care of 2019-nCoV infected pneumonia, as requested by frontline doctors and public health experts. We drew on the fast advise guidelines technique and basic WHO guideline creation criteria to create this document, and we supplemented it with management information from Zhongnan Hospital at Wuhan University.

M. Kaur et al., [6]‘‘ Joint Infection with Chikungunya and Dengue Viruses in Northwestern Punjab, India: A Serological Study, J. Laboratory Physicians, vol. 10, no. 4, pages 443-447, October 2018. It would seem that cases of dengue and chikungunya (CHIK) are on the rise over the whole of India. Common vectors for dengue virus (DENV) and chikungunya virus (CHIKV) include Aedes aegypti mosquitoes (CHIKV). Both viruses may be spread simultaneously in geographic regions where they cocirculate. A very few research have addressed dengue-chik coinfection in the Punjab area of India. The purpose of the current investigation was to compare dengue-CHIK coinfection to single infections and to characterize their clinical manifestations. Serum samples were tested using an ELISA for nonstructural protein 1 antigen, a MATERIALS AND IgM antibody capture (MAC) ELISA for dengue IgM and CHIK IgM, and an ELISA for IgM antibodies. Positive results for DENV were seen in 2178 of 3160 samples from patients suspected of having dengue fever, or 68.92%, whereas CHIK IgM antibodies were seen in 127 of 373 individuals, or 34.04%.

M. S. Hossain, Z. Sultana, et al.,[7] ‘‘Chikungunya Diagnosis in an Uncertain World: An Intelligent System, J. Wireless Mobile Netw., Ubiquitous Computing, Dependable Applications, Volume 10, Issue 2, Pages 37-54, 2019. Chikungunya is an illness caused by the CHIKV virus, which is transmitted through the bites of infected mosquitoes. Blood samples from patients in Tanzania were used to discover this virus. Chikungunya symptoms often include high body temperature, joint and muscle discomfort, and a severe headache. Preliminary diagnosis is usually made based on the doctor's evaluation of these symptoms. The doctor, however, is unable to provide precise measurements. As a result, the majority of patients may have an inaccurate first diagnosis, resulting in improper therapy.

V. Lakshmi et al., [8] Chikungunya fever: clinical presentation and molecular diagnosis in southern India This article first appeared in May 2008 in the journal Clinical Infectious Diseases, volume 46, issue 9, pages 1436-1442. In early 2006, after a 33-year lull, Chikungunya fever erupted in various regions of India, with certain areas of South India seeing a recurrence as recently as June 2007. A variety of molecular assays have been employed to detect Chikungunya virus infections, and their clinical symptoms are discussed in this article. The real-time loop-mediated isothermal amplification (RT LAMP) test stands out because it is easy to implement even in poorly equipped labs, is inexpensive, and can be performed in a short amount of time. The clinical signs were high body temperature, skin rash, and severe rheumatic indications. Comparative analysis of RT LAMP and reverse-transcriptase polymerase chain reaction revealed 20 more instances of Chikungunya virus positivity. Twenty samples were chosen at random and tested positive for Chikungunya virus.

III PROPOSED METHODOLOGY

Due to the development of machine learning, there has been a split in the field of computing between more conventional approaches and machine learning. This section discusses the works that build upon Artificial Intelligence-Based Cyber- Physical System for Severity Classification of Chikungunya Disease and explains why machine learning approaches are superior to more conventional approaches. In this project, we follow an established procedure to create the models. Present systems use Logistic Regression (LR) methods. However, it needs a lot of RAM and the results aren't reliable.

Disadvantages:

- Accuracy low
- Requires more time
- Difficult to handle

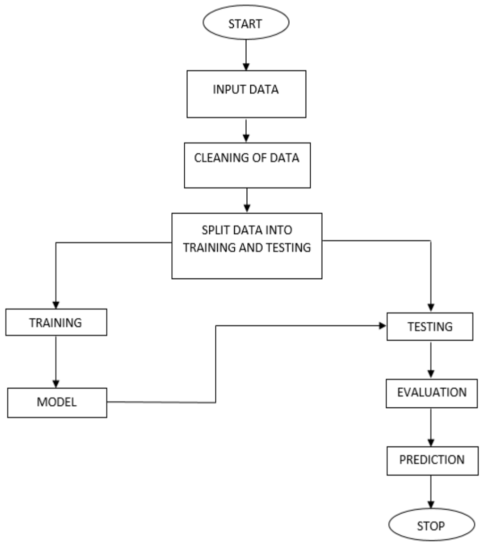

Fig 3 Block Diagram

There have been a number of attempts to use machine learning to identify diseases, but none of them have solved the misdiagnosis issue. Stevens Multi a disciplinary Artificial Intelligence-Based Cyber- Physical System for Severity Classification of Chikungunya Disease is one such tool. In addition, the variability and scale of the data are often ignored in research that have presented assessment models in a similar vein. In order to get rid of the bias and the deviation of instability, we offer a machine learning-based method that uses a novel technique of preprocessing the data for features transformation, a Decision Tree Classifier, and an XGBoost ML algorithm, all of which produce the best accuracy approaches.

Advantages:

– Requires less time
– Good Accuracy
– Easy to handle

In Fig 3, A system has to start out by taking an input data and preprocessing that data in a couple of ways such as cleaning it and splitting it. .It is intended that the splitting process will be divided between training and testing stages. This phase is after the system has been trained and understands the model and is now ready to go into the testing phase. During the testing phase, the analysis of data will take place and then these results will be considered for the prediction level, where the severity of the Chikungunya virus will be determined. Thus, this is the overall flow of the system that we are determining.

IV.MODULES:

Here, we have developed a system that includes two modules in order to determine the severity of the Chikungunya virus. The first is system related, followed by the user related module.

1. System:

In the system module, the storage of datasets, the training of models, the prediction of models, the splitting of data will occur.

- Store Dataset: The system will store the dataset given by the user, and when the user requests the data, it will provide it to them as soon as they request it. Following the data request, it will go through a couple of steps for the desired results.

- Model Training: As soon as the user requests data, it will go through model training where it will feed that data to the chosen model which will then identify patterns.

- Model Predictions: After feeding the data and we found patterns, based on this we can predict the desired output in this step.
- Data Splitting: After predicting now the system will split data into two categories, training and testing.

2. User:

In user module, we have designed with five steps. Registration, login, loading of the dataset, viewing the dataset and finally the results are the steps involved

- Registration: Every new user needs to register themselves in the system with some basic information about them like a unique name and email.
- Login: Registered new users can login anytime with the name and email they gave during registration.
- Load Dataset: After they login, they can upload the dataset for which they are expecting to work on.
- View Dataset: User can view the dataset which they have uploaded to work.
- Select model: User have an option to apply the model to the dataset which they uploaded for accuracy
- View results: Finally, user can view the predicted results whether the system is disease or not.

V.MODEL DESCRIPTION:

To develop this system, we have used following algorithms:

1. Decision Tree:

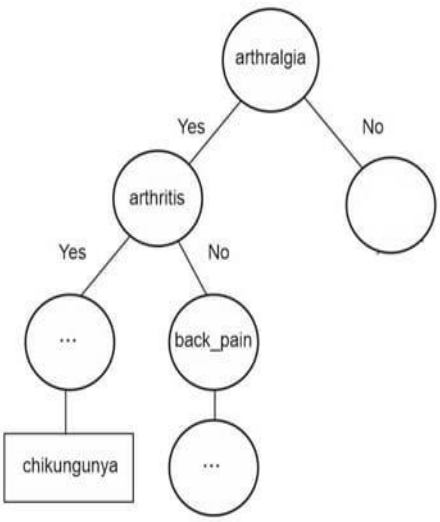

Fig 4 Model decision tree for chikungunya virus

2. Logistic Regression:

Fig 5 Logistic regression model

3. Random Forest Classifier:

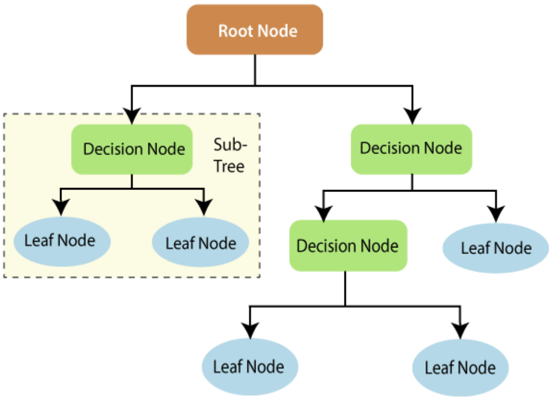

Fig 6 Model diagram for RFC

4. Support Vector Machine:

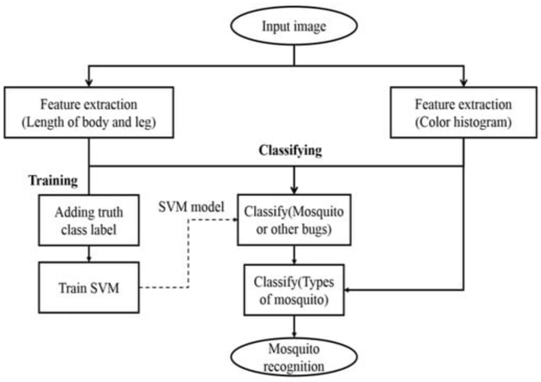

Fig 7 Model diagram for SVM

The model will start by allowing new user to register to the system with a unique name and email id. Once they register, they can login any time they want. Immediately after logging in the user will upload the dataset in which they are expecting to work on. Once they upload the dataset, they system will come into picture. System will take the dataset and it will preprocess it by cleaning and splitting. Meanwhile user can view the dataset whenever they and they have to split the size. After system splits the dataset into training and testing

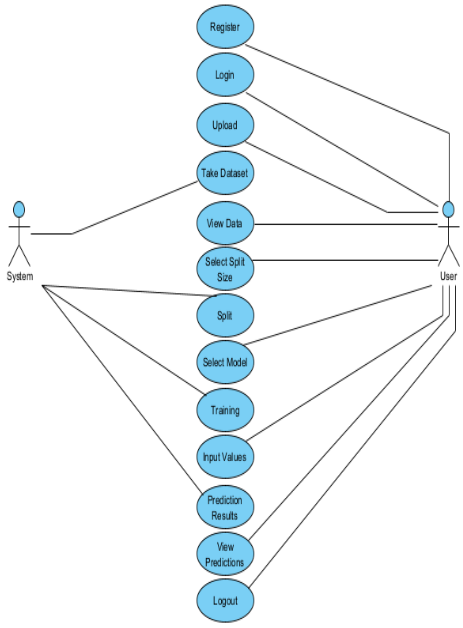

Fig 8 Use case diagram

User will have an option to choose model for accuracy. After selecting model by user, system will start evaluating the dataset. Finally, it will give predicted results to user which user can view. Here the user will know the severity of chikungunya virus. After User get desired results, they can logout from the system.

IV CONCLUSION

In this study, a cloud-based CPS is created and used to detect the Chikungunya virus. Physical and logical components make up the two main categories of the suggested system. Comparatively, the F-measure was 1.4145%, the accuracy was 1.3822%, the specificity was 1.3972%, and the sensitivity was 1.4145%. hence, the suggested approach for detecting Chikungunya disease may be used in emergency medical settings. In the near future, the deep transfer learning models may be used to more efficiently generate outcomes.

Additional cutting-edge metaheuristic methods can be created to effectively adjust the architectures for deep learning. Additionally, the suggested model can applied to different dataset types. cyberspace and space. After being gathered, the user-health data is kept in the cloud sub-system layer. Using ACGA, a developing RF model for categorising the severity of the Chikungunya virus was proposed. With more effective RF architecture optimization, ACGA can speed up computation and produce better results. Comparative research shows that ACGA-RF performs testing substantially better than the current models.

The development of a system that can reliably and quickly estimate the risk of Chikungunya infection is a major result of this study. The research describes the development of a belief rule based expert system (BRBES) that can handle ambiguous clinical data associated with Chikungunya symptoms and indications.

In light of this, the BRBES might be seen as a useful instrument for doing the first Chikungunya research. It has been shown that BRBES performs admirably, outperforming FLBES, human experts, and other machine learning methods like ANN and SVM. These days, particularly in Bangladesh, Chikungunya is a very real threat, and this technique may be utilized to acquire a quick, low-cost read on whether or not a patient may be infected. This technology will aid in quicker identification of this dangerous illness and provide faster, more timely treatment to those who are afflicted since it is both user-friendly and inexpensive. Medical professionals would benefit greatly from this BREBS, and many lives will be spared that would otherwise be lost owing to misdiagnosis or lack of detection of this illness.

REFERENCES

[1] Majhi, S. K. (2019). Fuzzy clustering algorithm based on modified whale optimization algorithm for automobile insurance fraud detection. Evolutionary Intelligence, 1-12.

[2] Carneiro, E. M., Dias, L. A. V., da Cunha, A. M., & Mialaret, L. F. S. (2015, April). Cluster analysis and artificial neural networks: A case study in credit card fraud detection. In 2015 12th International Conference on Information Technology-New Generations (pp. 122-126). IEEE.

[3] Duan, L., Xu, L., Liu, Y., & Lee, J. (2009). Cluster-based outlier detection. Annals of Operations Research, 168(1), 151-168.

[4] Minastireanu, E. A., & Mesnita, G. (2019). Light gbm machine learning algorithm to online click fraud detection. J. Inform. Assur. Cybersecur, 2019.

[5] Fang, Y., Zhang, Y., & Huang, C. Credit Card Fraud Detection Based on Machine Learning.

[6] Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002, January). Credit card fraud detection using Bayesian and neural networks. In Proceedings of the 1st international naiso congress on neuro fuzzy technologies (pp. 261-270).

[7] Wang, M., Yu, J., & Ji, Z. (2018). Credit Fraud Risk Detection Based on XGBoost-LR Hybrid Model.

[8] Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. Decision Support Systems, 95, 91-101.

[9] West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. Computers & security, 57, 47-66.

[10] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

[11] C.-S. Lee and M.-H. Wang, "A fuzzy expert system for diabetes decision support application," IEEE Trans-actions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 41, no. 1, pp. 139–153, May 2011.

[12] S. Lam, K. Chua, P. Hooi, M. Rahimah, S. Kumari, M. Tharmaratnam, S. Chuah, D. Smith, and I. Samp-son, "Chikungunya infection- an emerging disease in Malaysia," The Southeast Asian Journal of Tropical Medicine and Public Health, vol. 32, no. 3, pp. 447–451, September 2001.

[13] V. K. Ganesan, B. Duan, and S. Reid, "Chikungunya virus: Pathophysiology, mechanism, and modeling,"
Viruses, vol. 9, no. 12, p. 368, December 2017.

[14] S. Rahman, S.Suchana, S. Rashid, and O. Pave, "A review article on chikungunya virus," World Journal of Pharmaceutical Research, vol. 6, no. 13, pp. 100–107, November 2017.

[15] R. V. da Cunha and K. S. Trinta, "Chikungunya virus: clinical aspects and treatment - a review," Memo´rias do
Instituto Oswaldo Cruz, vol. 112, no. 8, pp. 523–531, August 2017.

[16] J.-B. Yang, "Rule and utility based evidential reasoning approach for multiattribute decision analysis underuncertainties," European journal of operational research, vol. 131, no. 1, pp. 31–61, May 2011.

[17] J. Liu, J.-B. Yang, J. Wang, H.-S. SII, and Y.-M. Wang, "Fuzzy rule-based evidential reasoning approach forsafety analysis," International Journal of General Systems, vol. 33, no. 2-3, pp. 183–204, 2004.

[18] P. P. Shenoy and G. Shafer, “Axioms for probability and belief-function propagation,” in Classic Works of the Dempster-Shafer Theory of Belief Functions, ser. Studies in Fuzziness and Soft Computing. Springer, Berlin, Heidelberg, 2008, vol. 219, pp. 499–528.

[19] V. Kecman, Learning and soft computing: support vector machines, neural networks, and fuzzy logic models. MIT press, June 2001.

[20] X. Yan and X. Su, Linear regression analysis: theory and computing. World Scientific, June 2009.

[21] P.-S. Yu, S.-T. Chen, and I.-F. Chang, “Support vector regression for real-time flood stage forecasting,” Jour-nal of Hydrology, vol. 328, no. 3-4, pp. 704–716, September 2006.

[22] D. Han, L. Chan, and N. Zhu, “Flood forecasting using support vector machine,” Journal of Hydroinformat-icsg, vol. 9, no. 4, pp. 267–276, October 2007.

[23] Z. He, X. Wen, H. Liu, and J. Du, “A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region,” Journal of Hydrology, vol. 509, pp. 379–386, February 2014.

[24] M. S. Reis and P. M. Saraiva, “Integration of data uncertainty in linear regression and process optimization,”
AIChE journal, vol. 51, no. 11, pp. 3007–3019, November 2005.