IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11346

A Survey on celiac disease prediction using AI Techniques

Mayura D Tapkire[1] , Vanishri Arun[2]

[1] Department of Information Science and Engineering, National Institute of Engineering, Mysuru.

[2] Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering,

Constituent College of JSS Science and Technology University, Mysuru.

Abstract: In recent times, health and wellness of the one has taken the centre stage and many programs have implemented in order to cater the lifestyle diseases, virus infections and many other health disorders. The disease fighting and improvement of oneself in the form of diet in order to develop immunity as well as prolong or cure the diseases are some of many areas that provides a scope for going in detailed analysis and research using Artificial Intelligence (AI) based on the data sets of different food formulations as well as patients' history. It is proposed to study and research the data in the field of disease prediction and healthcare for prediction and early detection of celiac diseases (CD) by applying the artificial intelligence techniques.

Machine Learning (ML) is one of the AI techniques that find high relevance in the medicine and health care. CD, a rare malabsorption syndrome of childhood and was limited to individuals of European ancestry. However, it is now identified as a common condition that may be diagnosed at any age and affects many organ systems in around 1% of the population. Currently symptomatic cases are diagnosed by physicians; however, this is not being effective in asymptomatic cases and it may remain undiagnosed. AI Techniques including Machine Learning can help in overcoming the existing limitations of diagnosis of CD.

This review provides an overall peek of current practices and research outcomes of the application of the ML techniques in the healthcare and medical practices. It further describes ML techniques in CD prediction and its relevance is summarized.

Keywords: Artificial Intelligence, celiac disease, disease prediction, machine learning

INTRODUCTION

CD affects the small intestine and is caused by an autoimmune reaction to gluten, which is found in roughly 1% of the world's population [1]. The symptoms of the patient are multisystemic, affecting the nervous, endocrine, reproductive, hematologic, cardiovascular, and other systems. Celiac disease's subsequent immune response frequently destroys the small intestine mucosa, particularly in the proximal parts, such as the duodenum and jejunum. The villi that absorb nutrients in the small intestine might become blunted, leading in inadequate nutrient absorption. As a result, fissuring, a mosaic look, and scalloping of mucosal folds can be seen in the small intestinal mucosa.

The current gold standard for CD diagnosis is an upper gastrointestinal endoscopy with duodenal biopsy, followed by serological testing for antibodies to tissue transglutaminase. However, because this method is invasive, it may overlook mucosal pathology in the distal small intestine, as well as pathology that are difficult to detect visually. CD is a long-term inflammatory illness of the small intestine characterized by an inappropriate immune response to gluten in genetically predisposed people. The easy treatment for celiac disease is a gluten free diet [2].

In Western countries, CD is most frequent chronic enteropathy and is prevalent to about 1% in the European continent. According to data obtained from screenings of population, it is reported that a considerable number of celiac cases still go undiagnosed. The vast range of CD symptoms makes diagnosis difficult: the classic expression of malabsorption syndrome, weight loss, and diarrhea account for very limited percentage of cases nowadays, and large number of patients show minor gastrointestinal symptoms or single-nutrient shortage. Due to this, doctors screening patients with nonspecific gastrointestinal symptoms which are present in different cases of functional and organic illnesses have a risk of missing a CD diagnosis.

LITERATURE REVIEW

Wireless capsule endoscopy (WCE) images were used by Wang et al [3] for development of modified module based on the idea of human virtual perception for extraction of important features and comparison with conventional modules. Hu et al. [4] in their studies on Squeeze-and-Excitation networks embedded these modules in ResNet50 and

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11346

Inception-v3. Output of these modules was taken as input for 3 classifiers – Support Vector Machine. K-nearest Neighbour (kNN) and Linear Discriminant Analysis (LDA) and concluded that ResNet50 with SVM performed best in detecting subtle villous changes in CD with accuracy of 97.2. However, the limitation was sample size and longer computation time.

Syed et al. [5] studied whether ML can make out the differences among healthy and diseases tissue using 3118 biopsy images of duodenum from unaffected, CD affected and environmental enteropathy cases. They developed AlexNet based modified CNN version and reported 93.4% accuracy with 2.4% false negative rate. Microscopic features could be automatically learnt by combination of CNN with deconvolutional network model from the input images eg. Epithelial secretory cells which are predictive of diagnosis.

Caetano et al [6] deployed ML using serology as base for CD diagnosis using model based on 2579 sample images of IgA-class endomysium antibody in AdaBoost with SVM, which is a meta-algorithm that can boost the performance of weak classifier on combination [7]. Another model was used to address multiclassification error and was based on random under-sampling. Both the models reported similar performance when tested using MATLAB®. Model 1 reported accuracy of 96.8, SN 82.84 and SP 99.4, whereas, Model 2 reported 98.5, SN 98.91 and SP 98.81, a net reduction of 2.05% error.

Choung et al [8] used serum samples from CD patients and control as input data in their work towards development of new biomarkers for diagnosis & monitoring of CD. Unreliable peptides were removed by developing and processing the synthetic biomarkers in rapmad R-package [9]. Study reported an accuracy of 99%, SN 99% and SP 100% after removing background noise and running in SVM model using Python package. It concludes that new biomarkers can identify new CD cases accurately and will help in diagnosis and mucosal healing. However, the major limitation was that the biomarkers were synthetic and not biological.

CELIAC DISEASE PREDICTION MODEL

In healthcare, ML is an emerging technique which is a rich field of AI that helps in prediction and diagnosis of various diseases. Nowadays accurate prediction of disease is very challenging. ML will help to build the model that will help to predict disease, based on symptoms possessed by patients. We can divide the process in four parts as follows in fig 1.

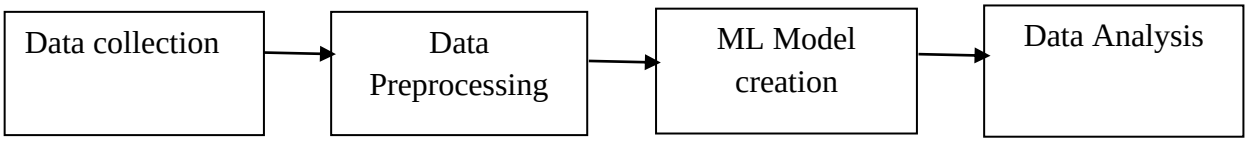

Fig 1: Basic steps followed in ML process for prediction of disease.

Data collection or data gathering is the primary step in any ML application. While building any ML model, it is not always the case that we come across the properly clean and formatted data. Data preprocessing is the process of raw data preparation and making it suitable for ML Model. After data collection and data preprocessing, data is ready for training of ML algorithms. After training we can predict the disease for the input symptoms by combining the predictions of all algorithms.

ML algorithms are useful in CD and similar diseases which requires extensive and continuous testing for diagnosis and treatment of patients. A basic ML model is illustrated in fig 2[2].

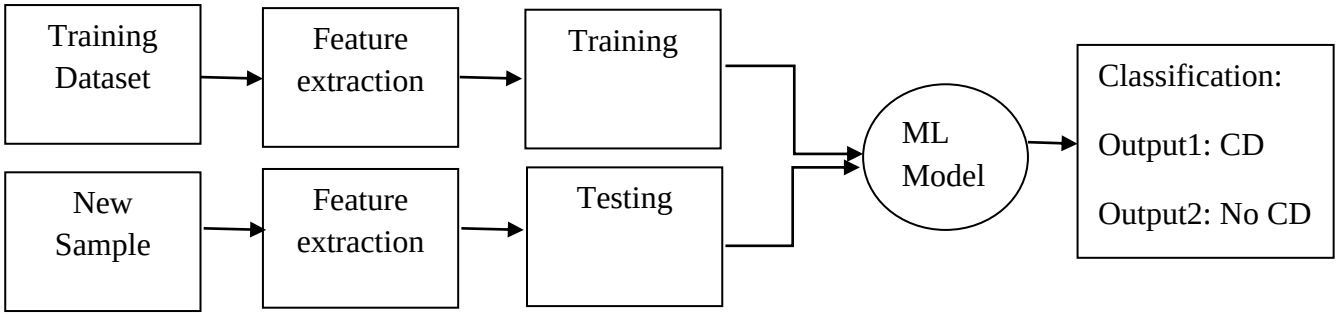

Fig. 2. Schematics of a classification model.

DOI: 10.17148/IJARCCE.2022.11346

Usually, a training dataset consists of text, video recordings or images. Few data features are having higher importance than others. Feature extraction techniques plays important role in getting those features for training or testing purpose. Once ML model is trained, different samples are fed to the model which results in binary or continuously varying values between 0 and 1 depending upon the type of model used [2].

CONCLUSION

AI is rapidly used in various disease predictions. These includes detecting diabetes, coronary heart disease, cancer detection, skin infections, lung infections and lot more. It also facilitates researchers in development of effective healthcare policies, and different models to prevent different types of disease. Early detection of diseases can reduce the risk factor [10].

REFERENCES:

1. Mehandiratta, A., Vij, N., Khanna, A., Gupta, P., Gupta, D., & Gupta, A. K. (2020). Prediction of Celiac Disease Using Machine-Learning Techniques. In International Conference on Innovative Computing and Communications (pp. 663-673). Springer, Singapore.
2. Sana MK, Hussain ZM, Shah PA, Maqsood MH. Artificial intelligence in celiac disease. Comput Biol Med. 2020 Oct;125:103996. doi: 10.1016/j.compbiomed.2020.103996. Epub 2020 Sep 17. PMID: 32979542.
3. X. Wang, H. Qian, E.J. Ciaccio, S.K. Lewis, G. Bhagat, P.H. Green, S. Xu, L. Huang, R. Gao, Y. Liu, Celiac disease diagnosis from videocapsule endoscopy images with residual learning and deep feature extraction, Comput. Methods Progr. Biomed. 187 (2020) 105236, https://doi.org/10.1016/j.cmpb.2019.105236. PMID: 31786452.
4. J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-Excitation networks, IEEE Trans. Pattern Anal. Mach. Intell. (2019), https://doi.org/10.1109/ tpami.2019.2913372. PMID: 31034408.
5. S. Syed, M. Al-Boni, M.N. Khan, K. Sadiq, N.T. Iqbal, C.A. Moskaluk, P. Kelly, B. Amadi, S.A. Ali, S.R. Moore, D.E. Brown, Assessment of machine learning detection of environmental enteropathy and celiac disease in children, JAMA network open 2 (6) (2019), e195822, https://doi.org/10.1001/ jamanetworkopen.2019.5822. PMID: 31199451 PMCID: PMC6575155.
6. F.L. Caetano Dos Santos, I.M. Michalek, K. Laurila, K. Kaukinen, J. Hyttinen, K. Lindfors, Automatic classification of IgA endomysial antibody test for celiac disease: a new method deploying machine learning, Sci. Rep. 9 (1) (2019) 9217, https://doi.org/10.1038/s41598-019-45679-x. PMID: 31239486 PMCID: PMC6592927.
7. S. Chen, B. Shen, X. Wang, S.-J. Yoo, A strong machine learning classifier and decision stumps based hybrid AdaBoost classification algorithm for cognitive radios, Sensors 19 (23) (2019) 5077, https://doi.org/10.3390/s19235077. PMID: 31757117.
8. R.S. Choung, S. Khaleghi Rostamkolaei, J.M. Ju, E.V. Marietta, C.T. Van Dyke, J. J. Rajasekaran, V. Jayaraman, T. Wang, K. Bei, K.E. Rajasekaran, K. Krishna, H. K. Krishnamurthy, J.A. Murray, Synthetic neoepitopes of the transglutaminase- deamidated gliadin complex as biomarkers for diagnosing and monitoring celiac disease, Gastroenterology 156 (3) (2019) 582–591, https://doi.org/10.1053/j. gastro.2018.10.025, e581, PMID: 30342033.
9. B.Y. Renard, M. Lower, Y. Kuhne, U. Reimer, A. Rothermel, O. Tureci, J.C. Castle, U. Sahin, rapmad: robust analysis of peptide microarray data, BMC Bioinf. 12 (2011) 324, https://doi.org/10.1186/1471-2105-12-324. PMID: 21816082 PMCID: PMC3174949.
10. Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. The Journal of Supercomputing, 77(5), 5198-5219.

 This work is licensed under a Creative Commons Attribution 4.0 International License