International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 1284-1290 www.ijaem.net ISSN: 2395-5252

Design a Novel Disease Risk Prediction Framework Based on Environmental Factors

Rohan Benhal, Tanmayee Tushar Parbat, Honey Jain,
BBA IT, Pune
B.E IT, Pune
B.E IT, Pune

Submitted: 10-08-2021 Revised: 22-08-2021 Accepted: 25-08-2021

ABSTRACT: The classification of coronary heart disease can be useful for medical practitioners if it is automated with the purpose of obtaining a quick and accurate result. The ability to accurately predict the existence of heart disease can help patients live longer lives. The goal of this research is to dissect how AI devices are used to predict and order heart disease. Support Vector Machine (SVM) and Artificial Neural Network (ANN) are used to classify different types of heart disease (ANN). On the basis of accuracy and training duration, the examination is done using one of two ways. This paper provides a sane, purposeful, precise, and rapid medical decision backing framework for coronary disease characterization. The Cleveland Heart Database and Statlog Database, both from the UCI Machine Learning dataset repository, were used. We divide the data records into two classes in the suggested system model, using a Support Vector Machine and an Artificial Neural Network. Analyse the performance of both datasets as well.
Disease Risk Prediction, Support Vector Machines, and Artificial Neural Networks are some of the terms used in this paper.

I. INTRODUCTION

The impact of the environment on human health has always been complex. Even when the immediate causes were evident and particular, with limited health consequences, as in the case of many environmental hazards or chemical toxins, their origins were usually deeper and more far-reaching. However, the scope and complexity of environmental health issues have grown more apparent in recent years. This shift is due to a number of variables [1]. One is the current usage of more powerful technologies, each of which has the potential to alter the environment more extensively and dramatically. Another is increased globalisation and connectedness of societies, as a result of which impacts are felt more widely – in terms of economic and social effects as well as health, on people far removed from the source of the hazard, and even on future generations – rather than just locally and immediately. In addition, in reaction to shifting policy concepts, structures, and imperatives, policies have become more expansive. Many modern environmental health problems are thus examples of systemic risks [2, 3], which are complex health risks buried in larger environmental, social, economic, and political systems. Systemic risks, according to the International Risk Governance Council [4], necessitate more integrated and cautious approaches to risk governance. Because systemic risks often take a long time to manifest, spread widely, and have long-term consequences, early intervention is important to prevent the risks from being established. Policies must be increasingly integrated, both because the problems are complex and linked, and because they cut through traditional policy-making frameworks, necessitating coordination among various agencies, policy areas, levels of administration, and spatial scales. At the same time, the costs of policies, as well as the penalty of doing them wrong, are rising, resulting in increased demands for financial accountability and policy openness [5, 6]. Furthermore, because of the vast breadth of systemic risks, many diverse stakeholders are necessarily implicated – as risk purveyors, victims, or managers [7]. These people must be educated and involved, not only because they have a moral right to know about the risks they face, but also because they are critical risk responders. Many of these ideas were recognised quite early in environmental policies. Integration and prudence, for example, have been foundation values of the Environment Action Plan of the since

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 1284-1290 www.ijaem.net ISSN: 2395-5252

the 1970s. Health policies, on the other hand, have been slower to react. The EU only agreed to a formal commitment to assure (rather than merely contribute to) human health protection with the passage of the Amsterdam Treaty in 1997, and only in 2003 did the World Health Organization's effort to construct national environmental health action plans become a reality. [9] Nonetheless, the Environment Health Action Plan is a watershed moment in health policy, with far-reaching ramifications for the sciences on which it is based. It not only lays out action priorities in terms of health outcomes and important risk factors, but it also emphasises the need for better, timelier, and more integrated data to assist policy. As a result, work is being done to make "evaluation of the overall environmental impact on human health more efficient by taking into account impacts such as cocktail effects, combined exposure, and cumulative effects," as well as other factors. It also emphasises the necessity for "an environment and health cause-effect framework' that would give the required information for the formulation of Community policy dealing with health stressor sources and pathways of impact."

II. RELATED WORK

G. -C. Lan et al.,[1] method outperformed traditional mechanism in terms of accuracy, precision and sensitivity for predicting the risk of diabetes. In particular, insightful observations show that the consideration of life-style information can effectively enhance whole performance for risk prediction. Moreover, classification rules produced by our mechanism which integrates C4.5 and CBA provide physicians disease related health risk patterns such that appropriate treatments could be given to people for disease prevention.

M. Pak et al [2] to handle this problem, we regenerated the training data by using the SMOTE approach and used them for disease risk prediction modelling. For model evaluation, the proposed method was employed to predict the risk of Type-2 diabetes disease. The experiment results showed that our SVM classifiers based on selective environmental factors could produce very comparable results to the prediction model with genetic factors in forecasting the risk of specific disease.

A. Mohawish, et al [3] enhance our analysis with the National Health and Nutrition Examination Survey (NHANES) data from the Centres for Disease Control (CDC), the United States public health agency. Our preliminary findings indicate that HRA data can be successfully used as input for the Framingham Risk Model in predicting risk of CHD utilizing NHANES data to predict missing attributes, thus extending the use of HRAs for disease risk prediction.

S. Ambedkar et al [4] To overcome the problem of missing medical data, we perform data cleaning and imputation to transform the incomplete data to complete data. We are working on heart disease prediction on the basis of the dataset with help of Naïve bayes and KNN algorithm. To extend this work, we propose the disease risk prediction using structured data. We use convolutional neural network based unimodel disease risk prediction algorithm. The prediction accuracy of CNN-UDRP algorithm reaches more than 65%. Moreover, this system answers the question related to disease which people face in their life.

C. Zhu, et al [5] The risk assessment model includes three parts: risk prediction, clustering analysis and regression analysis of risk factors, which can automatically predicate the risk level and risk factors for the discharged patients in thirty days. The model was accurate 90.62% of the time. Combined the model assessment results with risk control knowledge base, a personalized health management and health guidance given by care workers can be put forward intelligently, which can not only help medical personnel in the rational allocation but also guide patients to carry out self-management better, resulting in the decrease of readmission rate.

T. D. Pham et al [6] While the statistical and geostatistical techniques provide better results than those obtained from some other methods, the geostatistical approach yields superior results in terms of sensitivity and specificity in various designs of the data set for validation, training, and testing. The proposed computational strategies are very promising for predicting major adverse cardiac events within six months.

R. A. Vinarti et al [7] This paper introduces these tools and evaluates the effectiveness of the resulting BN for a single infectious disease, Anthrax. We have compared conditional probabilities predicted by our BN against incidence estimated from real patient visit records. Experiments explored the role of different context data in prediction accuracy. The results suggest that building a BN from an ontology is feasible. The experiments also show that more context results in better risk prediction.

III. PROPOSED METHODOLOGY

As previously stated, the numerous risk and health effect assessment approaches that have been created in recent years have resulted in a

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 1284-1290 www.ijaem.net ISSN: 2395-5252

somewhat perplexing scenario. Figure 1 depicts the issue as well as a possible solution. The issue is that there are a number of overlapping approaches that fail to meet the needs of policymakers for an integrated methodology for assessment– owing in part to differences in scientific perspective, conceptual inconsistencies, and a sloppy use of terminology that have characterised much of previous assessment research. Figure 1 shows a solution that provides a framework for what is known as integrated environmental health impact assessment. The goal is twofold: first, to integrate existing approaches together into a more coherent system so that users can make more informed decisions about which ones to use; and second, to extend these methods to give a more comprehensive framework for assessing complex, systemic risks and policies. The following is the rationale that underpins this framework. Integrated environmental health impact assessment is defined as a method of assessing environmental-related health problems, as well as the health-related effects of policies and other interventions that affect the environment, while taking into account the complexities, interdependencies, and uncertainties of the real world. As a result, both the environment and health are seen in a broad and comprehensive manner. In terms of the environment, for example, it encompasses not just the traditional emphasis of risk assessment, such as chemical dangers or environmental toxins, but also any other part of the ambient and living environment that may have a detrimental or beneficial impact on health. Health, on the other hand, is defined not just in terms of morbidity and mortality, but also in terms of overall human well-being. Human exposure to environmental risks or human access to and exploitation of environmental capital and services are hence responsible for health effects. Both are mediated by human behaviours and perceptions, and are thus a function of where people live and spend their time, as well as the populations' personal and societal characteristics (age, gender, socioeconomic status, culture, belief systems, and so on), as well as the associated susceptibilities, attitudes, and values. Exogenous factors, in turn, affect the entire environmental health system. These are the result of a variety of causes, including not only legislation, but also other actions and developments, such as technical, socioeconomic, and demographic changes. These external factors act as forces for change within the system, for example, by altering the state of environmental capital or hazards, influencing population distribution, characteristics, and behaviours, and affecting health care and other factors that condition their impacts (e.g., awareness raising, insurance systems). As a result, an integrated environmental health impact assessment examines the effects of environmental capital and dangers in the context of these shifting external pressures. Similarly, the effects of change can have a cascading effect on these drivers, altering social and demographic structures, economic situations, and policy initiatives. When heart illnesses worsen, they spiral out of control. Heart illnesses are complex, and they claim the lives of many people each year. If the early signs and symptoms of heart disease are ignored, the patient may face serious repercussions in a short period of time. In today's environment, sedentary lifestyles and excessive stress have exacerbated the problem. It is possible to keep the disease under control if it is discovered early. However, it is always essential to exercise on a daily basis and to break bad habits as soon as possible. Tobacco use and poor eating habits raise the risk of stroke and heart disease. It's a good idea to eat at least 5 servings of fruits and vegetables each day. It is recommended that heart disease patients limit their salt intake to one teaspoon per day. One of the key flaws in these studies is that the emphasis has been on the application of classification algorithms for heart disease prediction rather than on the numerous data cleaning and pruning strategies that prepare and prepare a dataset for mining. It has been discovered that a dataset that has been adequately cleaned and pruned delivers substantially greater accuracy than one that has missing values. The development of prediction systems with improved accuracy will be aided by the use of appropriate data cleaning procedures and effective categorization algorithms. In the future, an intelligent system may be developed that may help a patient diagnosed with heart disease choose the best treatment options. There has already been a lot of effort put into developing models that can predict whether or not a patient would acquire heart disease. Once a patient has been identified with a specific type of heart disease, there are numerous therapy options available. By collecting knowledge from such appropriate databases, data mining can be highly useful in determining the course of treatment to be taken.

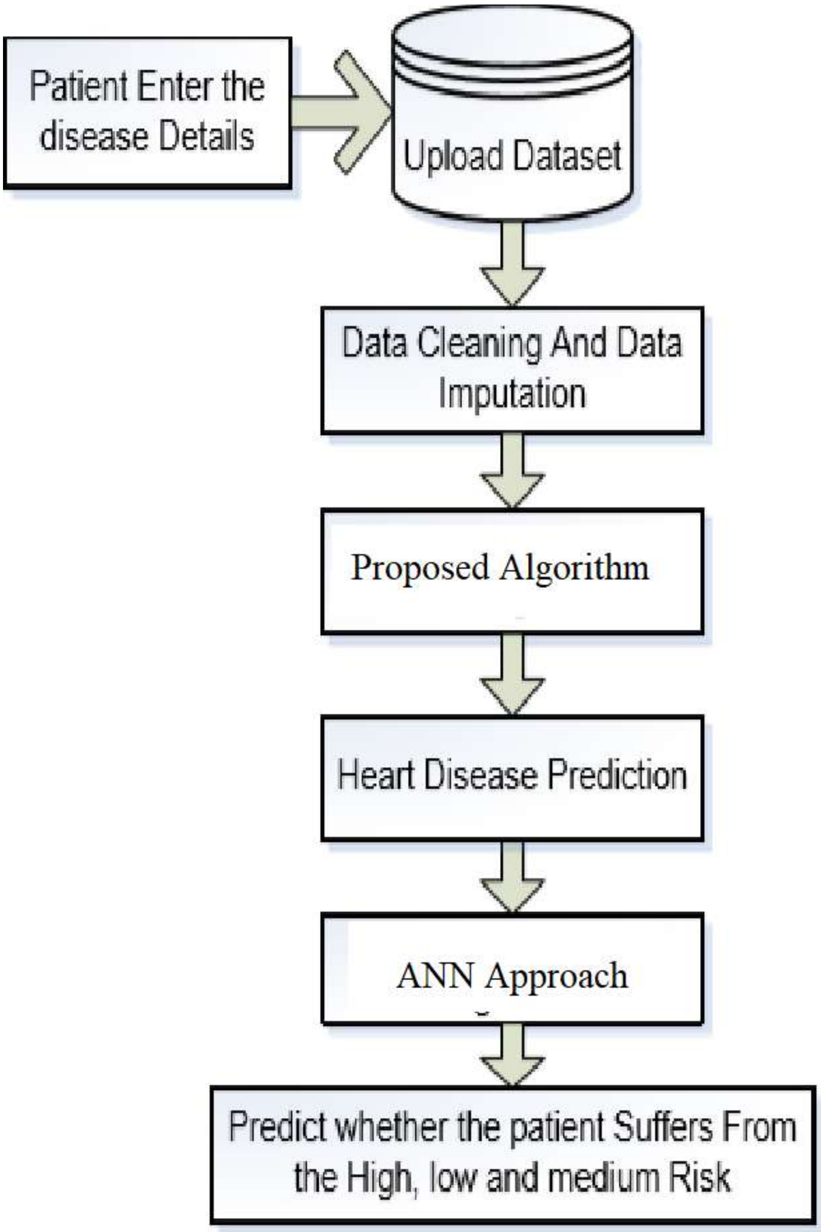

Figure 1: Proposed approach

It is a supervised learning method which classifies data into two classes over a hyper plane. Support vector machine performs a similar task like C4.5 except that it doesn't use Decision trees at all. Support vector machine attempts to maximize the margin (distance between the hyper plane and the two closest data points from each respective class) to decrease any chance of misclassification. Some popular implementations of support vector machine are scikit-learn, MATLAB and of LIBSVM. An artificial neural network (ANN) is a computational model based on the structure and functions of biological neural networks. Information which flows through the network affects the structure of the artificial neural network because a neural network changes or learns in a sense-based on input and output, for that particular stage and consequently for each stage. ANN's are considered nonlinear statistical data modelling tools where the complex relationships between inputs and outputs are modelled or patterns are found. ANNs have layers that are interconnected. Artificial neural networks are fairly simple mathematical models to enhance existing data analysis technologies

IV. RESULTS ANALYSIS

This paper summarises some of the recent works done in data mining related to cardiovascular diseases. Data mining algorithms can be effectively used to 'mine' relevant information from the huge amounts of data generated by the healthcare industry. These works show that rather than applying a single mining technique on a data set, results are far better if a collection of mining techniques are used. Java is chosen in most of the research work for practical execution of the project. WEKA, Tanagra, MATLAB etc. are some of the other popular tools used for data analysis. Careful selection of the combination of mining techniques and accurate implementation of those techniques on the data set yields a fast and effective implementation of a system for heart disease management. The required dataset is divided into two parts, one is used for mining and the smaller partition is used for verifying. Most of the time, 10-fold cross validation technique is used. Some of the works are about the comparison of different classification techniques on a dataset to correctly classify if a given patient has any probability of a heart disease or not. Others have worked on 'mining' the causes that lead to heart diseases from a given dataset. Commonly used classification techniques are Decision tree, Naïve Bayes, artificial neural network, association rule mining and fuzzy logic. Apart from analysing these commonly used techniques, some of the recent works have studied about "hybrid models". The idea of a hybrid model is to incorporate several known classification and selection techniques in a single model to give better results. It is observed that hybrid models give very high accuracy if proper combinations of different algorithms are chosen.

Table 1: Classification accuracy and time complexity of Naïve Bayes, SVM and Proposed approach based on ANN

Algorithm	Time taken (Ms)	Accuracy
Naïve Bayes	550	88%
SVM	400	91%
MNN	340	93%
Proposed approach based on ANN	200	96%

V. CONCLUSION AND FUTURE WORK

Heart disease prediction system (HDPS) employing data mining and Support Vector Machine (SVM) artificial neural network (ANN) approaches is provided in this study paper. The system is built using a multilayer perceptron neural network and a back propagation method derived from the ANN. Because the MLPNN model produces superior results and assists domain experts and others in the field in planning for a better diagnosis and providing early diagnosis results to patients, it functions reasonably well even when not retrained. The trial results demonstrate that the algorithm predicts heart illness with about 96 percent accuracy using neural networks.

REFERENCE

[1]. G. -C. Lan et al., "Disease Risk Prediction by Mining Personalized Health Trend Patterns: A Case Study on Diabetes," 2012 Conference on Technologies and Applications of Artificial Intelligence, 2012, pp. 27-32, doi: 10.1109/TAAI.2012.53.

[2]. M. Pak and M. Shin, "Developing disease risk prediction model based on environmental factors," The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014), 2014, pp. 1-2, doi: 10.1109/ISCE.2014.6884338.

[3]. A. Mohawish, R. Rathi, V. Abhishek, T. Lauritzen and R. Padman, "Predicting Coronary Heart Disease risk using health risk assessment data," 2015 17th International Conference on E-health Networking, Application & Services (HealthCom), 2015, pp. 91-96, doi: 10.1109/HealthCom.2015.7454479.

[4]. S. Ambekar and R. Phalnikar, "Disease Risk Prediction by Using Convolutional Neural Network," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697423.

[5]. C. Zhu, S. Chi, R. Li, D. Tong, Y. Tian and J. Li, "Design and Development of a Readmission Risk Assessment System for Patients with Cardiovascular Disease," 2016 8th International Conference on Information Technology in Medicine and Education (ITME), 2016, pp. 121-124, doi: 10.1109/ITME.2016.0036.

[6]. T. D. Pham et al., "Computational Prediction Models for Early Detection of Risk of Cardiovascular Events Using Mass Spectrometry Data," in IEEE Transactions on Information Technology in Biomedicine,

vol. 12, no. 5, pp. 636-643, Sept. 2008, doi: 10.1109/TITB.2007.908756.

[7]. R. A. Vinarti and L. Hederman, "Personalization of Infectious Disease Risk Prediction: Towards Automatic Generation of a Bayesian Network," 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), 2017, pp. 594-599, doi: 10.1109/CBMS.2017.24.

[8]. Ambekar, S., & Phalnikar, R. (2018). Disease Risk Prediction by Using Convolutional Neural Network. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 1-5.

[9]. Purushottam, K. Saxena and R. Sharma, "Efficient heart disease prediction system using decision tree," International Conference on Computing, Communication & Automation, 2015, pp. 72-77, doi: 10.1109/CCAA.2015.7148346.

[10]. Rajawat A.S., Barhanpurkar K., Shaw R.N., Ghosh A. (2021) Risk Detection in Wireless Body Sensor Networks for Health Monitoring Using Hybrid Deep Learning. In: Mekhilef S., Favorskaya M., Pandey R.K., Shaw R.N. (eds) Innovations in Electrical and Electronic Engineering. Lecture Notes in Electrical Engineering, vol 756. Springer, Singapore. https://doi.org/10.1007/978-981-16-0749-3_54

[11]. D. Dahiwade, G. Patle and E. Meshram, "Designing Disease Prediction Model Using Machine Learning Approach," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019, pp. 1211-1215, doi: 10.1109/ICCMC.2019.8819782.

[12]. Rajawat A.S., Rawat R., Barhanpurkar K., Shaw R.N., Ghosh A. (2021) Sleep Apnea Detection Using Contact-Based and Non-Contact-Based Using Deep Learning Methods. In: Bansal J.C., Paprzycki M., Bianchini M., Das S. (eds) Computationally Intelligent Systems and their Applications. Studies in Computational Intelligence, vol 950. Springer, Singapore. https://doi.org/10.1007/978-981-16-0407-2_7

[13]. T. Wang, Y. Tian and R. G. Qiu, "Long Short-Term Memory Recurrent Neural Networks for Multiple Diseases Risk Prediction by Leveraging Longitudinal Medical Records," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 8, pp. 2337-2346, Aug. 2020, doi: 10.1109/JBHI.2019.2962366.

[14]. Z. Huang, W. Dong, H. Duan and J. Liu, "A Regularized Deep Learning Approach for Clinical Risk Prediction of Acute Coronary Syndrome Using Electronic Health Records," in IEEE Transactions on Biomedical Engineering, vol. 65, no. 5, pp. 956-968, May 2018, doi: 10.1109/TBME.2017.2731158.

[15]. Rajawat A.S., Rawat R., Barhanpurkar K., Shaw R.N., Ghosh A. (2021) Blockchain-Based Model for Expanding IoT Device Data Security. In: Bansal J.C., Fung L.C.C., Simic M., Ghosh A. (eds) Advances in Applications of Data-Driven Computing. Advances in Intelligent Systems and Computing, vol 1319. Springer, Singapore. https://doi.org/10.1007/978-981-33-6919-1_5

[16]. M. Raihan et al., "Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design," 2016 19th International Conference on Computer and Information Technology (ICCIT), 2016, pp. 299-303, doi: 10.1109/ICCITECHN.2016.7860213.

[17]. Bedi P., Goyal S.B., Rajawat A.S., Shaw R.N., Ghosh A. (2022) A Framework for Personalizing Atypical Web Search Sessions with Concept-Based User Profiles Using Selective Machine Learning Techniques. In: Bianchini M., Piuri V., Das S., Shaw R.N. (eds) Advanced Computing and Intelligent Technologies. Lecture Notes in Networks and Systems, vol 218. Springer, Singapore. https://doi.org/10.1007/978-981-16-2164-2_23

[18]. P. Ramprakash, R. Sarumathi, R. Mowriya and S. Nithyavishnupriya, "Heart Disease Prediction Using Deep Neural Network," 2020 International Conference on Inventive Computation Technologies (ICICT), 2020, pp. 666-670, doi: 10.1109/ICICT48043.2020.9112443.

[19]. N. S. Rajliwall, R. Davey and G. Chetty, "Machine Learning Based Models for Cardiovascular Risk Prediction," 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), 2018, pp. 142-148, doi: 10.1109/iCMLDE.2018.00034.

[20]. V. Balasubramanian, R. Gouripeddi, S. Panchanathan, J. Vermillion, A. Bhaskaran and R. Siegel, "Support vector machine based conformal predictors for risk of complications following a coronary Drug Eluting Stent procedure," 2009 36th Annual Computers in Cardiology Conference (CinC), 2009, pp. 5-8.
[21]. M. D. Mantzaris et al., "A Multimodal Advanced Approach for the Stratification of Carotid Artery Disease," 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), 2019, pp. 706-709, doi: 10.1109/BIBE.2019.00133.
[22]. S. Hartati, "A Kohonen artificial neural network as a DSS model for predicting CAD," 2010 International Conference on Distributed Frameworks for Multimedia Applications, 2010, pp. 1-5.
[23]. E. Alharbi, F. Nadeem and A. Cherif, "Smart Healthcare Framework for Asthma Attack Prediction and Prevention," 2021 National Computing Colleges Conference (NCCC), 2021, pp. 1-6, doi: 10.1109/NCCC49330.2021.9428842.