International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 4.988

| Volume 4, Issue 1, January 2021 |

Prediction of Pesticides Induced Diabetes using Machine Learning

Nithya P[1], Dr.A.M.Kalpana[2]

Assistant Professor, Department of Computer Engineering, Government College of Engineering, Salem, India[1]

Professor and Head, Department of Computer Engineering, Government College of Engineering, Salem, India[2]

ABSTRACT: Diabetes is a chronic disease which is also considered as a deadliest public illness caused due to increase in blood sugar as well as the usage of pesticides. The use of pestiicdes is also one of the reasons to promote or increase diabetes disease. The main aim of this work is to predict the diabetes disease where the usage of pesticides is high. To solve this tedious problem artificial intelligence has more impact in introducing learning techniques where this approach solve critical problem by diagnosing the disease. This paper, proposed a novel architecture for predicting the diabetes mellitus. The process is done through various phases like data collection, data preprocessing, pattern generation and prediction. This paper, explains the various architecture with various phases in detail. Future, performance of the proposed architecture and analyzed result will evaluate.

I. INTRODUCTION

Diabetes is a dangerous disorder that affects approximately 425 million people living in the world at 2017, and this number may be increased to 629 million by 2045[1]. According to the American Diabetes Association [2], diabetes is a chronic disease that occur when high blood glucose levels and deficiencies in the body's ability to produce and / or use insulin.Commonly, diabetes has categorized by three types that are type 1, type 2 and gestational[3]. Type 1 diabetes is mostly found in adults while the body does not produce insulin properly. Type 2 diabetes is the famous type among the peoples and diagnosed any type of people. This type occurredbecause the body does not produce sufficient insulin or cells avoid the insulin. Pesticides is one of the reason to environmental factors are almost always linked to type 2 diabetes[4]. Third one is gestational diabetes, mostly the pregnant ladies affected by this type . Diabetes comes due to the reason of glucose level, that is come from the food which taken by individual. But some experts told that is pesticides are considered to be one of the major factors involved in today's global environmental pollution[5,6].These are an agricultural chemical that is suspected to be a important contributor to the worldwide diabetes epidemic[7]. In agricultural field[8] , different pesticides such like insecticides, fungicides, bactericides or rodenticides which are used to kill weeds, insect, bacteria etc., In the current scenario, the production of chemicals is greatly increased as well as the impact of diabetes is high, which indicating a positive relationship between chemical exposure and the prevalence of diabetes [9]. Further , the some experts concluded use of different chemicals in the environment lead to cause of diabetes , therefore the prevention in diabetes is very important to peoples who living with this environment. The aim of this study is to confirm previous findings of the connection among diabetes and some pesticides used in agricultural field as well as the risk of diabetes.

II. RELATED WORK

A. Alexandru et al. [10] presented a survey on big data in cloud computing with healthcare services. This survey explained the various tools that are used to store and analyze the patients'data in the cloud environment and also proposed a novel architecture of big data in healthcare. But there was lack of explanation about the proposed architecture and its tools.

Firouzi et al. [11] examined the role of big data in health care and discovered that body sensors provide huge amount of health-related data. Two challenges were analyzed in this area namely, integrating large volume of data with electronic health records (EHR) and presenting it to the doctors in real time. Based on these observations, a sensor integration framework was proposed for scalable cloud architecture to provide a complete solution to the EHR sensor system. Apache Kafka and Spark were used to process massive data in a real-time manner. Although visualizing patients' health in real-time manner can aid physicians, the proposed model lacks in security aspects.

S. Manogaran et al. [12], proposed a novel framework to analyze and process massive amount of healthcare data on cloud context based on Hadoop. This paper described the importance of big data to take correct

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 4.988

| Volume 4, Issue 1, January 2021 |

decisionsfor patients by selecting the right care. Various cryptography techniques were used to secure the framework and MapReduce technique was used for Health data to improve the performance. Hadoop was used to analyze the massive health data moreappropriately.

J. Hanen and et al. [13], recommended a novel medical cloud multi-agent system (MCMAS) to aidphysiciansfor detecting heart disease of the patients remotely. The recommended MCMAS was applied into a Google's Android operating system and CloudSim were used to provide the solution for traditional problem in diagnosis of cardiovascular infection. The proposed system diagnosed cardiovascular infection with the help of mobile cloud computing. Improvement was needed to enhance the architecture and deploy prototype not only polyclinic.

Sharma et al. [14], presented an application system for big healthcare based on IoT and big data analytics. The proposed approach provided better solution for the growing chronic diseases (GCD), the increasing medical expenses (IME) and uneven process distribution of medical resources (UDMR). Furthermore the architecture, technology challenges of big healthcare system and several typical applications were discussed,

Elhoseny et al. [15], proposed a new method to review huge volume of patients' data in real-time by using mobile cloud computing. The proposed model was applied to Google's Android operating system. Moreover the roles of cloud computing and big data analytics in its activation were also discussed. The techniques, applications and tools of big data analytics were described. Furthermore, it states that more complex methods were needed to process such large data of healthcare and the cost has to be considered when working on a real time analysis.

Portugal et al. [16] reviewed the use of eight data mining algorithms, namely SVM, KNN, linear discriminant analysis (LDA), NB, C4.5, artificial neural network (ANN), C5.0 and deep learning ANN (DLANN), for IoTdata. These algorithms were compared based on their confusion matrix, classification accuracy and execution time. Based on the classification accuracy, C4.5, C5.0, ANN, and DLANN performed well than SVM, KNN, NB, and LDA. However, C4.5, ANN, and C5.0 were very similar in classification accuracy. Meanwhile, NB and LDA have the fastest execution time but LDAhas slightly better processing time than NB. In this work, larger and diverse IoT datasets were not considered in detail.

Plachkinova et al [17] proposed a framework for supporting medical data in a cloud environment based on IoT and big data.Smart Sensors were utilized to collect huge data from patients, which assist to get healthy knowledge for rare diseases.However, further details of the proposed model and test results were required to evaluate in terms of its performance, especially in difficult applications.

II. METHODS AND MATERIALS

In this paper , a new architecture was proposed to analysis the relation that is the use of pesticides whether promotes diabetes or increases the risk of diabetes. The proposed architecture shown in figure 1, it include different phases such as data collection , data preprocessing , pattern generation and prediction. The phases of the proposed structure are explained in the following section.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 4.988

| Volume 4, Issue 1, January 2021 |

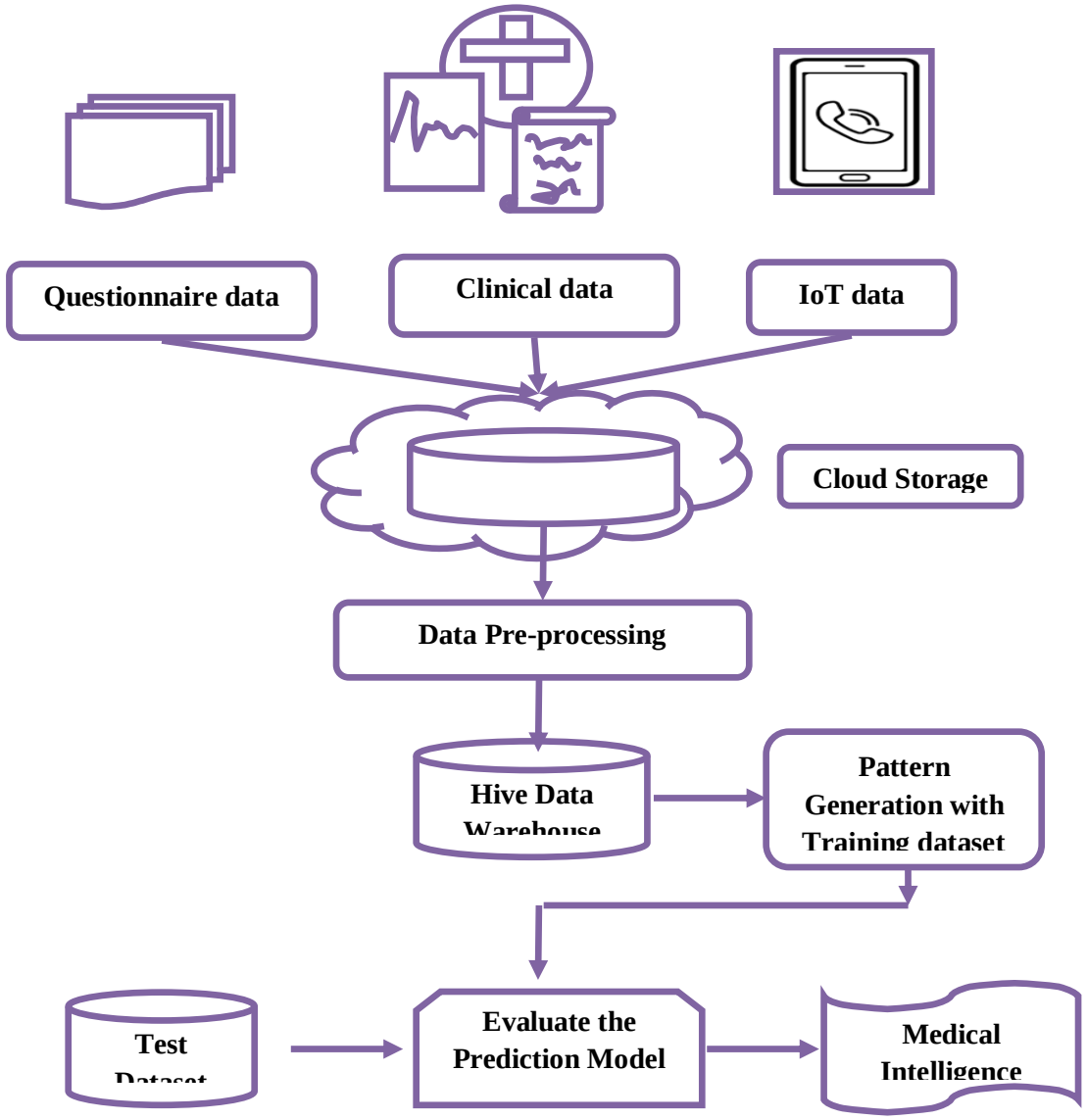

Figure 1 : Association among Pesticides and Diabetes prediction architecture

III. STUDY MATERIALS

The study was conducted using data from participants in tamilnadu and aims to clarify that is the pesticides are exposure to various diseases. The aim of this cohort study was to resolve the prevalence and common risk factors of chronic disease which was preventable such as diabetes, hypertension, metabolic syndrome and cardiovascular disease. The participants of this study were mostly farmers living in the rural areas. Each of the participantsis asked to answer all the questions whichwereprovided in the questionnaire sheet. The survey was conducted with 2292peoples between November 2018 and January 2019. Of them, 568 participants with information on pesticide exposure were included. Thus, 2290 records were added to the analysis part.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 4.988

| Volume 4, Issue 1, January 2021 |

Data collection

The first process in the proposed structure is data collection. Here data collection is performed in various ways to this experiment. The standardized model utilized to collect the pesticides usage details in a particular area. The questionnaire includes details about pesticides as well as the personal details of the particular's such like age, gender, BMI, alcohol consumption, smokingstatus, education qualification and monthly income.The clinical information collected from the hospital and the environmental status also assessed through the sensors.

These various data collections method explained in the following figure 2.

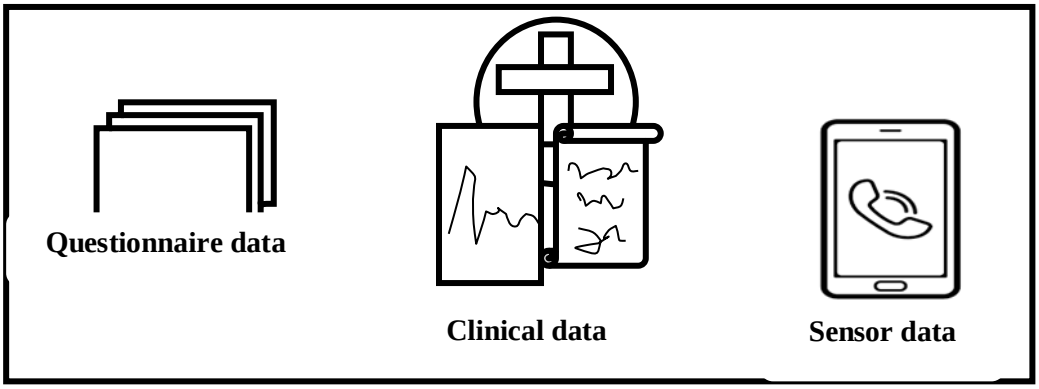

Figure 2: Data collection phases

Data preprocessing

The second process is data preprocessing, this is plays an important when creating models in data analysis. In which, different processing techniques were followed like data cleaning, dataintegration, data transformation and data reduction. Data cleaning method used to remove the noise and inconsistencies data in the data set. Data integration method combined the data which comes from various sources into a single data warehouse.The remove correlation among data features, redundant data in the dataset and clustering the data are done by implementing the data reduction method. Data transformation applied to transform the data from one order to another appropriate form. Normalization, attribute selection,discretization and concept hierarchy generation are involved in data transformation method. The characteristics of data preprocessing techniques as follow in table.

Training set and testing set

The dataset divided into two categories,training and testing. The one by fourth of the data that is 30% is defined as a testing set and the three by fourth of the data is training set that's is 70% of the data. The data is trained when model developing, data tested after the model developed and trained

Model building

The model has developed by using deep learning algorithms. Deep learning is the subset of machine learning algorithms that work based on ANN. Here learning can be supervised or unsupervised or semi supervised. In this study, CNN based model is developed, which is used to classify and predict the collected data. Basically, CNN is the combination of input,convolutional, pooling and more dense layers. The convolutional layer placed after pooling and dense layers, it consists of filters that have size, stride and padding of filters. This layer receives all inputs from the previous layer. After reviewing the existing related works, fixed this algorithm is suitable to this experiment. The model was created topredict whethera person affected by diabetes disease due to the use of pesticides or not. The developed model attains an accuracy of 93 % in the prediction.

IV. CONCLUSION

Now a day's diabetes is one of the chronic diseases all over world and increasing tremendously rate among people. Moreover, the usage of pesticides also increased. Some researches discussed that is the pesticides is one of the reasons to promote various diseases especially diabetes diseases. So,this paper introduces an architecture using Deep learning to find out the association amongpesticides and diabetes. The data set was collected through a survey, clinical centers and devices such as sensors. The collected dataset was processed by various phases of the proposed architecture. The model attains an accuracy of 93 % in the prediction. In future, the performance will be analyzed using deep learning technique and the prediction results will obtain.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 4.988

| Volume 4, Issue 1, January 2021 |

REFERENCES

1. International Diabetes Federation. Idf Diabetes Atlas—7th edition.
2. International Diabetes Federation. 2015.
3. American Diabetes Association. "Diagnosis and classification of diabetes mellitus", Vol.37, Issue.1, 2014, pp : 81-90.
4. Olson, Brooke. "Applying medical anthropology:Developing diabetes education and prevention programs in American Indian cultures", American Indian Culture and Research Journal, Vol.23, Issue.3, 1999, pp: 185-203
5. Paul, Kimberly C., Michael Jerrett, and Beate Ritz. "Type 2 diabetes mellitus and Alzheimer's disease: overlapping biologic mechanisms and environmental risk factors." *Current environmental health reports* 5, no. 1 (2018): 44-58.
6. Kuo C-C, Moon K, Thayer KA, Navas-Acien A. Environmental chemicals and type 2 diabetes: an updated systematic review of the epidemiologic evidence. Curr Diab Rep. 2013;13(6):831–49.
7. Lee D-H, Lee I-K, Song K, Steffes M, Toscano W, Baker BA, et al. A strong dose-response relation between serum concentrations of persistent organic pollutants and diabetes: results from the National Health and examination survey 1999-2002. Diabetes Care. 2006;29(7):1638–44.
8. Vasiliu O, Cameron L, Gardiner J, DeGuire P, Karmaus W. Polybrominated biphenyls, polychlorinated biphenyls, body weight, and incidence of adultonset diabetes mellitus. Epidemiology. 2006;17(4):352–9.
9. Park, Sungjin, Sung-Kyung Kim, Jae-Yeop Kim, Kyungsuk Lee, Jung Ran Choi, Sei-Jin Chang, Choon Hee Chung, Kyu-Sang Park, Sung-Soo Oh, and Sang-Baek Koh. "Exposure to pesticides and the prevalence of diabetes in a rural population in Korea." *NeuroToxicology* 70 (2019): 12-18.
10. A.ALEXANDRU, C. A. ALEXANDRU, D. COARDOS and E. TUDORA" Healthcare, Big Data and Cloud Computing", WseasTransactions On Computer Research, 2016, Vol. 4, pp. 123- 131.
11. Firouzi, Farshad, Amir M. Rahmani, KunalMankodiya, Mustafa Badaroglu, Geoff V. Merrett, P. Wong, and BaharFarahani. "Internet-of-Things and big data for smarter healthcare: from device to architecture, applications and analytics." (2018): 583-586.
12. Manogaran, Gunasekaran, RamachandranVaratharajan, Daphne Lopez, PriyanMalarvizhi Kumar, RevathiSundarasekar, and ChanduThota. "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system." *Future Generation Computer Systems* 82 (2018): 375-387.
13. Hanen, Jemal, ZiedKechaou, and Mounir Ben Ayed. "An enhanced healthcare system in mobile cloud computing environment." *Vietnam Journal of Computer Science* 3, no. 4 (2016): 267-277.
14. Sharma, Sagar, Keke Chen, and AmitSheth. "Toward practical privacy-preserving analytics for iot and cloud-based healthcare systems." *IEEE Internet Computing* 22, no. 2 (2018): 42-51.
15. Elhoseny, Mohamed, Ahmed Abdelaziz, Ahmed S. Salama, Alaa Mohamed Riad, Khan Muhammad, and Arun Kumar Sangaiah. "A hybrid model of internet of things and cloud computing to manage big data in health services applications." *Future generation computer systems* 86 (2018): 1383-1394.
16. Portugal, Ivens, Paulo Alencar, and Donald Cowan. "The use of machine learning algorithms in recommender systems: A systematic review." *Expert Systems with Applications* 97 (2018): 205-227.
17. Plachkinova, Miloslava, Au Vo, Rahul Bhaskar, and Brian Hilton. "A conceptual framework for quality healthcare accessibility: a scalable approach for big data technologies." *Information Systems Frontiers* 20, no. 2 (2018): 289-302.