International Journal of Engineering & Technology, 7 (3.27) (2018) 56-61

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Improved Ant Colony on Feature Selection and Weighted Ensemble to Neural Network Based Multimodal Disease Risk Prediction (WENN-MDRP) Classifier for Disease Prediction Over Big Data

Gakwaya Nkundimana Joel[1*], S. Manju Priya[2]

[1]*Research Scholar, Dept. Computer Science, Karpagam Academy of Higher Education.*
[2]*Associate Professor, Dept. of CS, CA & IT, Karpagam Academy of Higher Education.*
E-mail:smanjupr@gmail.com
**Corresponding author E-mail:gajoel2000@gmail.com*

Abstract

As the big data is growing in biomedical and healthcare communities, so are precise analyses of medical data aids, premature disease identification, patient care as well as community services. On the other hand, the accuracy of the analysis decreases, if the medical data quality is imperfect. As a result, the choice of features from the dataset turns out to be an extremely significant task. Feature selection has exposed its efficiency in numerous applications by means of constructing modest and more comprehensive models, enlightening learning performance and preparing clean and clear data. The proposed method analyzes the difficulties of feature selection for big data analytics. Improved Ant Colony Optimization based Feature Selection (IACO) algorithm is presented for resolving this issue. The reconstruction of missing data before the incomplete data available was performed with help of latent factor mode. Therefore, it was not easy to choose the best features from the structured and unstructured data. the unheard technique which is called Weighted Ensemble Based Neural Network for multimodal disease risk prediction(WENN-MDRP) algorithm is implemented in order to provide the best features selection among structured as well as unstructured data. The research method provides improved prediction accuracy when matched with conventional techniques. In the MATLAB environment, the presented classifiers are implemented. The outcomes are computed in regard to recall, precision, accuracy, f-measure and error rate.

Keywords: *Disease prediction, machine learning, ensemble, neural network, prediction accuracy*.

1. Introduction

50% of Americans contain one or more chronic diseases, 80% of American medical care fee is expended on chronic disease treatment in keeping with a report by McKinsey [1]. The occurrence of chronic disease is rising with the enhancement of living standards. On chronic disease treatment, the US has expended an average of 2.7 trillion USD yearly. This amount encompasses 18% of the complete annual GDP of the US. The healthcare issue of chronic diseases is as well extremely significant in numerous countries. Chronic diseases are the foremost reason of death in China.86.6% of deaths are occurred by chronic diseases in keeping with a Chinese report on nutrition as well as chronic diseases in 2015. So, it is vital to carry out risk assessments for chronic diseases. Gathering Electronic Health Records (EHR) [3] is progressively expedient with the development in medical data [2].
More considerationis paid to disease prediction with the growth of big data analytics technology.With the aim of enhancing the accurateness of risk classification, numerous researches were carried out by means of choosing the characteristics robotically from a huge amount of data [4-5], instead of the formerlychosen characteristics. At the present time, more than a few machine learning techniques [6] are implemented for getting the precious knowledge from enormous data in disease prediction. On the other hand, they are unsuccessful while faces varied nature of feature values and that previous work typically took structured data.
As data is of massive as well as complexity naturally, taking out precious knowledge from big data is not simple with conventional data was processing as well as data handling methods. Henceforth, for dealing with big data to get the valuable model, the growth of progressive technologies is needed. Deep as well as machine learning are the finest solution fortaking out beneficial knowledge as well as make suitable decisions from big data, [7]. The multiple or ensemble classifier is the mixtures of a diverse classifiers called base classifiers [8]. Typically, ensemble classifier gives improved classification accurateness compared to the single classifiers. Present years, use of deep learning quickly rising in numerous areas. In deep learning models, the neural network acts as a big role. This workpresents the mixture of neural network as well as ensemble learning techniques in the big data environment.
Previous data processing techniques are not suitable to deal with, examine and process. It is because of the complexity, unlabeled data and Un categorized. Therefore, in examining such enormous volumes of unsupervised data, the deep learning techniques acts as a vital role, which are suitable for exploiting huge amount of data and for examining raw data from numerous sources as well as indifferent styles [9].
Feature selection has shown its effectiveness in numerous applications by means of constructing modesties well as more

inclusive models, enlightening learning performance, and making clean, comprehensible data. In this research work, know about the challenges of feature selection for big data analytics. In order to choose features Improved Ant Colony Optimization based Feature Selection (IACO) technique is presented for resolving this issue. Previous to that firstly incomplete data, utilize a latent factor model to rebuild the missing data. IACO based feature selection for best selection of features from healthcare data. As well a novel Weighted Ensemble Based Neural Network based multimodal disease risk prediction (WENN-MDRP) technique is presented with the help of structured as well as unstructured data from hospital. No other previous method concentrated on both data types in the field of medical big data analytics. The remaining paper is organized in this manner. Section II provides the related work, and Section III elucidates the proposed method. SectionIV offers the experimentation design and analysis. Section V summaries the conclusion part.

2. Literature Review

The machine learning as well as data medical field have developed early detection systems. Subhapriyaet al [8] presented the improved data mining method for healthcare application. The system encompasses predominantly three steps those are as follows anomaly detection, clustering, and classification. Classification algorithm utilizes the ensemble technqiue known as a random forest. This system precisely foresees the patient outcome from a noteworthyvolume of data.

Kathleen et al [10] presented an ensemble machine learning technology, which uses an adaptive Boosting algorithm implemented for precise heart disease prediction results. The research techniques used to 4 diverse data sets for heart disease diagnosis. Datasets comprise heart disease from Hungarian Institute of Cardiology (HIC), Cleveland Clinic Foundation (CCF), Switzerland University Hospital (SUH) and Long Beach Medical Center (LBMC). The outcomes of the novel algorithm proved that the performance of novel ensemble techniques is superior to previoustechniques.

In [11], the disease risk prediction is expressed into a multilabel classification problem by Li et al. A new Ensemble Label Power-set Pruned datasets Joint Decomposition (ELPPJD) technique is presented in this research work. Initially, the multi label classification is transferred into a multiclass classification problem. As well recommended the pruned datasets as well as joint decomposition approaches to handle the imbalance learning problem. ELPPJD technique offers improved outcomes in the experimentationoutcomes.

Saxena and Sharma [12] illustrated a structure, which guesses the risk level of patients. In orderto make the accurate decision regarding the heart disease risk level, they concentrated on non-specialized Doctors. The research method focuses on rule generation parameters. The rules comprise Pruned Rules, Original Rules, Rules without duplicates, Sorted Rules, Classified Rules, andPolish.

In [13], Jaseena and Kovoor proposed a general idea of diverse deep learning methods for big data in biometrics and converses some problems and solutions. As wellcomplete survey of all the associated work in deep learning for biometric big data is presented. In [14], Farid et al presented an ensemble learning in which a novelmethod that imparted boosting as well as decision tree classifier together. AdaBoost algorithm is utilized as the algorithm for boosting ensemble. Separate decision tree technique is used for each sample,and those outcomes are brought up to date. The attained cases were again categorized with the help of the voting technique.

In [15], Chen et al introduced a novel machine learning technique for valuable prediction of disease. With the help of structured and unstructured data, anovel Convolutional Neural Network (CNN) based multimodal disease risk prediction technique is explained. The usesfor instance computer vision, speech recognition, and Natural language processing, its challenges, open research problems in addition to the future works are provided.

In [16], Chen et al presented design details, significant technologies and practical implementation techniques of smart clothing system. Distinctive applications powered by smart clothing as well as big data clouds are provided, for instance emotion care,medical emergency response, disease diagnosis, as well as real-time tactile interaction. Particularly, electrocardio graph signals gathered by smart clothing are utilized for mood monitoring as well as emotion identification. Lastly, as well highpoint certain design challenges as well as open problems, which must be stated to create smart clothing ubiquitous for an extensive range of applications.

In [17], a novel knowledge-based system was presented by Bates et al for diseases prediction with the help of clustering, noise removal, and prediction methods. As well, in order to produce the fuzzy rules to be utilized in the knowledge-based system, Classification and Regression Trees (CART) is utilized. The research technique was developed on numerous public medical datasets. Outcomes onMesothelioma, Pima Indian Diabetes, StatLog, WDBC, Cleveland and Parkinson's telemonitoring datasets prove that there search technique enhances the diseases prediction accurateness. The knowledge-based system couldaid medical practitioners in the healthcare practice as a medical analytical technique.

In [18], Qiu et al concentrated on the issue of data sharing problems in cloud computing as well aspresentamethod, which utilizes dynamic programming with the aim of producing thebest possible solutions to data sharing techniques. The research method is known as Optimal Telehealth Data Sharing Model (OTDSM) that takes transmission probabilities, increasing network capacities, and timing restraints. The experimentation outcomes have shown the flexibility as well as adoptability of the research technique.

In [19], Zhang et al presented a Cyber-Physical System (CPS) for patient-centric healthcare applications and services, known as Health-CPS, constructed on cloud and big data analytics technologies. This system encompasses a data collection layer with a unified standard, a data management layer for distributed storage as well as parallel computing, in addition to a data-oriented service layer. The outcomes of this research prove that the technologies of cloud as well as big data areutilized for improving the performance of the healthcare system with the intension that humans could then enjoy numerous smart health care applications as well as services. In [20], Gakwaya N. Joel et reviewed on role of big data in healthcare as the pivot key to ameliorate the health of nations, and the use of big data in order to discover the medication for fighting certain diseases. The suggested technlonogy such as hadoop fremarwok showed the advantage in analyzing huge data and storing those numerious data.

In [21], a new method was presented by Nori et al. efficiently incorporated medical domain knowledge associating with the similarity amongst diseases as well as the similarity amongst Electronic Health Records (EHRs) into a data-driven method by means of integrating graph Laplacians into the regularization term to encrypt these similarities. The experimentation outcomes on a real dataset from a hospital validated the efficiency of the research technique. In addition to some results for disease predictive feautures which were highly reliable as per medical techiniques additional analysis in the medical domain is still needed.In [22], introduced a machine learning technique by Bandyopadhyay et al dependent upon Bayesian networks trained on EHD to foresee the likelihood of containing a CV event within 5 years. In these data, event status might be unknown for certain individuals, since the event time is right-censored because of disenrollment as well as imperfect follow-up. Outcomes prove that the research technique could result in improved predictive performance compared to the Cox proportional hazards model (that is to say a regression-based method generally utilized for censored, time-to-event data) or a Bayesian network with ad hoc techniques to right-censoring. The research methods are inspired by as well as exemplified on data from a huge US Midwestern health care system.

3. Proposed System Methodology

Research technique focuses on feature selection as well as improved classification accurateness for healthcare data. In order to choose the important features, Improved Ant Colony Optimization based feature selection (IACO) method is presented in this section. After that a novel weighted ensemble based neural network disease prediction technique is presented for the structured as well as unstructured data from the hospital. Uniting ensemble as well as Convolutional Neural Networks (CNN) is probable by means of utilizing CNN as weak learners in the ensemble algorithm. Weights are attuned dependent upon the misclassification.

Improved Ant Colony Optimization Based Feature Selection (IACO) Algorithm

In any machine learning related task, Feature selection is an important process that identifies a subset of features, from the original feature set. ACO technique finds the best possible feature subset utilizing certain iterations. This research work uses an Improved Ant Colony Optimization Based Feature Selection (IACO) Algorithm for decreasing the redundancy.

```
Algorithm:
Initialize following
        Ci is the amount of change of pheromone trial
quantity for feature 'fe'
        Define the maximum number of iterations
        Define n, where the n-best subsets will influence
        the subsets of the next iteration
        Define s, where t=s is the number of features each
        ant will start within the second and following
        iterations
Initialize pheromone table
For i=1
        Randomly assign a subset of t features
        Select the features to construct a solution for each
ant
        Update the best ant table
        Update pheromone tables
        I=Maximum number of iterations
End For
```

CNN algorithm consists of 2 phases; Convolutional phase and Transfer Knowledge Phase.

```
Convolutional Neural Network
Phase I Conventional Phase
Initialize all weights and biases of the CNN to a small
value.
Set learning rate r such that 0<r<<1
n= 1
repeat
for m=1 to M do
propagate pattern x_m through the network
for k= 1 to the number of neurons in the output layer
Find error
End for
for layers L-1 to 1 do
for maps j = 1 to J do
find error factor to be back-propagated
end or
end for
for i=1 to L do
for j=1 to J do
for all weights of map, j do
Find W
Update weights and biases
W=W(Old)+W(New)
End for
end for
end for
n = n + 1
Find Mean Square Error (MSE1)
Until MSE1<s or n > maximum bounds
Phase II Transfer Knowledge Phase
repeat
for tk= 1 to TK (number of new training samples)
propagate pattern xtk through the network
for z= 1 to the number of neurons in the last Convolutional
layer(Z)
find output O^z oflastlayer of theconvolutional layer.
O_z = ( O_1, O_2, O_3 ........ O_Z)
Find O^zk using TSL framework (section III)
End for
end for
```

Ensemble Algorithm

Ensemble algorithms uniting various collection of individual models together to enhance the predictive accurateness.

Algorithm:

Consider *D* represent the original training data, *k* signify the number of base classifiers, and *T* is the test data

for $i = 1$ to k do

Create training set D_i from D.

Build a base classifier C_i from D.

End for

for each test record $x \in T$ do

$c^*(x)=\text{vote}(c_1(x),c_2(x),\ldots,c_k(x))$

end for

Weighted Ensemble Based Neural Network in Big Data

Combination ensemble algorithm with CNN is utilized in the research technique. Furthermore, this is developed in the big data environment using Hadoop – MapReduce.

Weighted Ensemble Based Neural Network Algorithm (WEBNN):
Input:
number of classes M, number of boosting, iterations N, dataset D
Algorithm:
Init: set f(x) = 0
for t = 1 to N do
compute weight w
train a network to optimize g
calculate optimal coefficient c
update the f(x)=f(x)+(c*g)
end for
Output: Predictor f(x)
MapReduce - Weighted Ensemble Based Neural Network Algorithm
Begin
MAPPER
Create training datasets D based on balanced bootstrapping from the original training dataset
Apply Weighted Ensemble Based Neural Network Algorithm
End for
REDUCER
for each training data record
Sum the results
endfor
end

4. Experiment Design and Analysis

In this part, define the hospital datasets utilize in this research. Also, offer disease risk prediction model as well as evaluation techniques. The data is gathered from UCI machine learning repository. It encompasses 100000 instances as well as 55 attributes. The dataset denotes10 years of clinical care at 130 US hospitals. It comprises over 50 features denoting patient as well as hospital results. Information was taken out from the database for encounters, which fulfilled the subsequent condition.

1. It is an inpatient encounter (a hospital admission).
2. It is a diabetic encounter, that is, one for the period of which any type of diabetes was entered to the system as a diagnosis.
3. The length of stay was at least 1 day as well as at most 14 days.
4. Laboratory tests were carried out for the period of the encounter.
5. Medications were administered in the course of the encounter.

The data encompasses attributes for instance patient number, age, gender, race, admission type, medical specialty of admitting physician, time in hospital, number of lab test carried out, diabetic medications, diagnosis, HbA1c test result, number of medication, and number of outpatient, inpatient, and emergency visits in the year beforehand the hospitalization, and so on.
For dataset assessment these traditional machine learning techniques, that is to say
Naive Bayesian (NB), Convolutional Neural Network based Multimodal Disease Risk Prediction (CNN-MDRP), K-nearest Neighbour (KNN), and Decision Tree (DT) algorithm [23-24] are utilized to foresee the risk of cerebral infarction disease. This is for the reason that these three machine learning techniques are extensively utilized [25].
For the performance assessment, initially, represent *TP*, *FP*, *TN* and *FN* as true positive (the amount of instances appropriately predicted as needed), false positive (the amount of instances inaccurately predicted as needed), true negative (the amount of instances appropriately predicted as not needed) and false negative (the amount of instances imperfectly foreseen as not needed), correspondingly. After that, could get four measurements: precision, accuracy, recall and F1-measure in this manner:

Accuracy = (TP+TN)/(TP+FP+TN+FN)
Precision= (TP)/(TP+FP)
Recall= (TP)/(TP+FN)
F1-measure= (2*Precision* Recall)/ (Precision+ Recall)

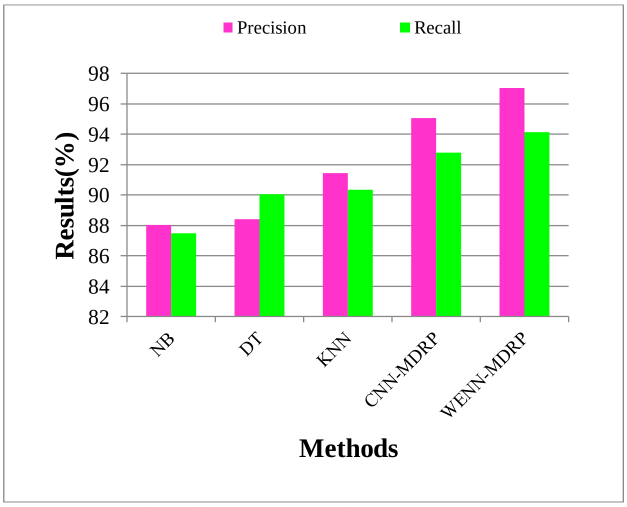
here the F1-Measure is known as the weighted harmonic mean of the precision as well as recall and denotes the complete performance. Along with the afore stated assessment condition, as well utilize Receiver Operating Characteristic (ROC) curve and the Area Under Curve (AUC) for assessing the advantages and disadvantages of the classifier. The ROC curve proves the trade-off amid the True Positive Rate (TPR) and the False Positive Rate (FPR), in which the TPR and FPR are defined in this manner:

TPR= (TP)/(TP+FN)
FPR= (FP)/(FP +FN)

The model is good when the ROC curve is nearer to the upper left corner of the graph. The AUC is the area under the curve. The model is better while the area is closer to 1. In medical data, we concentrate on the recall instead of accuracy. The greater the recall rate, the lesser the probability that a patient who would contain the risk of disease is identified to have no disease risk. In Table 1, the performance of numerous familiar techniques on disease prediction is stated, and it proves that presented WEBNN algorithm contain improved accuracy compared to other provided classifiers.

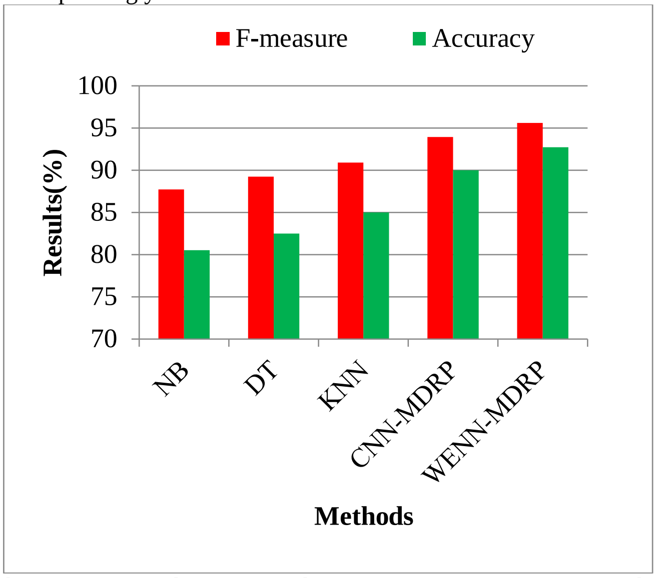
Table 1: Performance Comparison of Various Algorithms

Algorithm	NB	DT	KNN	CNN-MDRP	WENN-MDRP
Precision	88.05	88.41	91.46	95.09	97.05
Recall	87.50	90.06	90.36	92.81	94.16
F-measure	87.77	89.23	90.91	93.94	95.58
Accuracy	80.50	82.50	85.00	90.00	92.70
Error rate	19.50	17.50	15.00	10.00	7.30



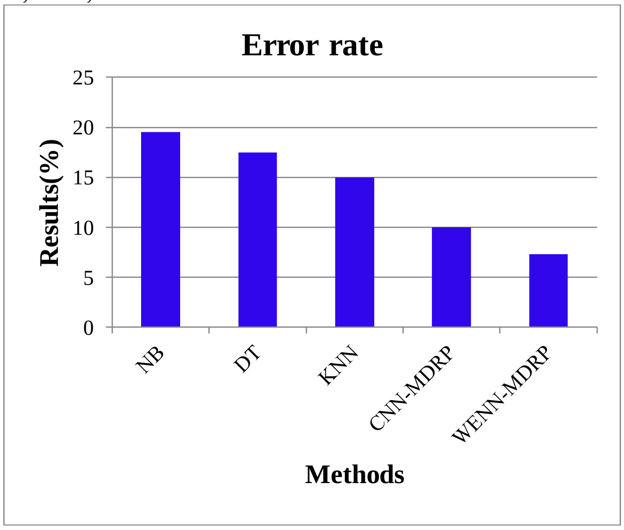
(a) Precision, and recall comparison under NB, DT, KNN, CNN-MDRP and WENN-MDRP

In the figure 1(a) the classifiers for instance NB, DT, KNN, CNN-MDRP and WENN-MDRP yields precision outcomes of 88.05 %, 88.41%, 91.46 %, 95.09 % and 97.05% correspondingly. It yields recall outcomes of 87.50%, 90.06%, 90.36%, 92.81% and 94.16% correspondingly for classifiers. According to the figure 1(a) presented WENN-MDRP classifier yields greater precision outcomes of 9%, 8.64%, 5.59% and 1.96% while matched up with NB, DT, KNN, and CNN-MDRP classifiers correspondingly. According to figure 1(a) presented WENN-MDRP classifier yields greater recall outcomes of 6.66%, 4.1%, 3.8% and 1.35% while matched up with NB, DT, KNN, and CNN-MDRP classifiers correspondingly.



(b) F1-Measure and accuracy under NB, DT, KNN, CNN-MDRP and WENN-MDRP

Figure 1 (a) – (c) exposes the performance comparison outcomes of five diverse classifiers under classification metrics. For example in figure 1(a) exposes the performance comparison outcomes of precision and recall metrics amid five diverse classifiers. In the figure 1(b) the classifiers for instance NB, DT, KNN, CNN-MDRP and WENN-MDRP yields f-measure outcomes of 87.77%, 89.23%, 90.91%, 93.94% and 95.58% correspondingly. It yields accuracy outcomes of 80.50%, 82.50%, 85.00%, 90.00% and 92.70% correspondingly for classifiers. According to the outcomes it proves that the presented classifier contains 7.81%, 6.35%, 4.67% and 1.64% greater f-measure outcomes while matched up with NB, DT, KNN, CNN-MDRP and WENN-MDRP classifiers. According to the outcomes it proves that the presented classifier contains 12.2%, 10.2%, 7.7% and 2.7% greater accuracy outcomes while matched up with NB, DT, KNN, CNN-MDRP and WENN-MDRP classifiers.



(c) Error rate comparison under NB, DT, KNN, CNN-MDRP and WENN-MDRP

Figure 1: Overall results of health care data

Figure 1(c) exposes the error rate outcomes of all classifiers for instance 19.50%, 17.50%, 15.00%, 10.00%and7.30% correspondingly. According to the outcomes it is clear that the presented WENN-MDRP classifier yields minimum error rate as well as greater accuracy outcomes while matched up with all classifiers.

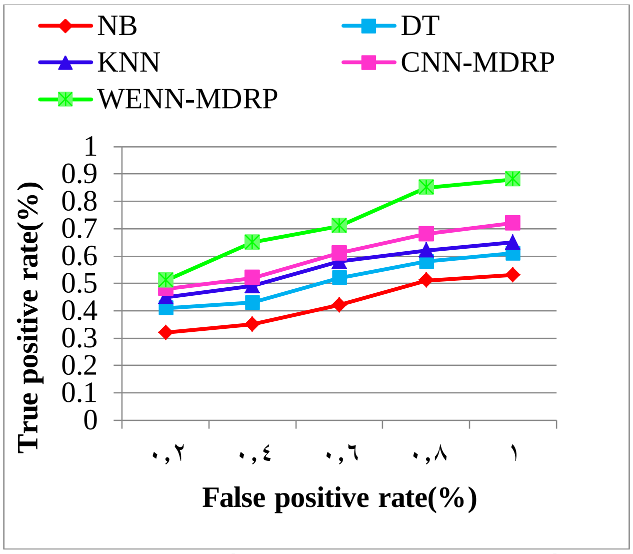


Figure 2: ROC curves under NB, DT, KNN, CNN-MDRP and WENN-MDRP classifiers

According to Figure 2, it is clear that the equivalent AUC of NB, DT, KNN, CNN-MDRP and WENN-MDRP are 0.51, 0.65, 0.71, 0.85 and 0.88, correspondingly. In summary, for data, the WENN-MDRP classification is the finest in experimentation[26]. On the other hand, it is as well noticed that presented classifier has precisely foreseen whether the patient is in a greater risk of diabetes in keeping with the patient's gender, age, clinical laboratory and other structured data. [27]

5. Conclusion and Future Work

Presently, people are often distressed by numerous diseases. Day by day, the number of patients distressed by certain infection is rising. In order to choose features, Improved Ant Colony Optimization based Feature Selection (IACO) algorithm is presented. Previous to that in the beginning incomplete data, a latent factor model was utilized in order to rebuild the missing data. These two techniques namely IACO based feature selection for best selection of features from healthcare data and Weighted Ensemble Based Neural Network for multimodal disease risk prediction (WENN-MDRP) methods are combined in the big data medical field to improve the health. The experimentation carried out on structured as well as unstructured data in order to choose the important features for prediction. This method is mainly concentrated on feature selection parameter. Outcomes of the novel method prove that it outdoes previous techniques. The improved outcomes of this method aid the specialists to get organized for a greater determine as well as offer the patient to contain initial determination outcomes. According to the outcomes the presented classifiers yields greater accuracy outcomes of 92.70%, while other classifiers for instance NB, DT, KNN, and CNN-MDRP yields accuracy outcomes of 80.50%, 82.50%, 85.00%, 90.00% and 92.70% respectively. It shows that the research technique is superior matched with other classifiers. When matched with other distinctive classifiers, the classification accuracy of presented WENN-MDRP algorithm attains 92.70% with a convergence speed that is sooner compared to that of the CNN- MDRP algorithm. In future, work will concentrate on using the same algorithm to other disease prediction. Numerous optimization approaches are utilized to choose the best possible features also expanding the scope of future work.

References

[1] Groves P, Kayyali B, Knott D & Van Kuiken S, "The 'big data'revolution in healthcare", *McKinsey Quarterly*, Vol.2, No.3, (2013), pp.1-22.

[2] Chen M, Mao S & Liu Y, "Big data: A survey", *Mobile Networks and Applications*, Vol.19, No.2, (2014), pp.171–209.

[3] Jensen PB, Jensen LJ & Brunak S, "Mining electronic health records: towards better research applications and clinical care," *Nature Reviews Genetics*, Vol.13, No.6, (2012), pp.395–405.

[4] Qian B, Wang X, Cao N, Li H & Jiang YG, "A relative similarity based method for interactive patient risk prediction", *Data Mining and Knowledge Discovery*, Vol.29, No.4, (2015), pp.1070–1093.

[5] Singh A, Nadkarni G, Gottesman O, Ellis SB, Bottinger EP & Guttag JV, "Incorporating temporal ehr data in predictive models for risk stratification of renal function deterioration", *Journal of biomedical informatics*, Vol.53, (2015), pp.220–228.

[6] Nasrabadi NM, "Pattern recognition and machine learning", *Journal of electronic imaging*, Vol.16, No.4, (2007), pp.049901.

[7] Chen XW & Lin X, "Big data deep learning: challenges and perspectives", *IEEE Access*, Vol.2, (2014), pp.514-525.

[8] Subhapriya P, Sujatha R & Meghana K, "Healthcare Prediction Analysis in Big Data using Random Forest Classifier", *International Journal of Advance Research, Ideas and Innovations in Technology*, (2017), pp.494-496.

[9] Jaseena KU & Binsu CK, "A Survey on Deep Learning Techniques for Big Data in Biometrics", *International Journal of Advanced Research in Computer Science*, Vol.9, No.1, (2018), pp.12-17.

[10] Kathleen HM, Julia HM & George JM, "Diagnosing Coronary Heart Disease Using Ensemble Machine Learning", *International Journal of Advanced Computer Science and Applications*, Vol.7, No.10, (2016), pp.30-39.

[11] Li, R., Liu, W., Lin, Y., Zhao, H & Zhang, C., 2017. An Ensemble Multi label Classification for Disease Risk Prediction. Journal of healthcare engineering, (2017), pp.1-11.

[12] Saxena K & Sharma R, "Efficient Heart Disease Prediction System", *Procedia Computer Science*, Vol.85, (2016), pp.962-969.

[13] Jaseena KU & Kovoor BC, "A Survey on Deep Learning Techniques for Big Data in Biometrics", *International Journal of Advanced Research in Computer Science*, Vol.9, No.1, (2018), pp.12-17.

[14] Farid DM, Marufand GM & Rahman CM., "A new approach of Boosting using decision tree classifier for classifying noisy data", *IEEE International Conference on Informatics, Electronics & Vision*, (2013), pp.1-4.

[15] Chen M, Hao Y, Hwang K, Wang L & Wang L, "Disease prediction by machine learning over big data from healthcare communities", *IEEE Access*, Vol.5, (2017), pp.8869-8879.

[16] Chen M, Ma Y, Song J, Lai CF & Hu B, "Smart clothing: Connecting human with clouds and big data for sustainable health monitoring", *Mobile Networks and Applications*, Vol.21, No.5, (2016), pp.825-845.

[17] Bates DW, Saria S, Ohno-Machado L, Shah A & Escobar G, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients", *Health Affairs*, Vol.33, No.7, (2014), pp.1123–1131.

[18] Qiu L, Gai K & Qiu M, "Optimal big data sharing approach for tele-health in cloud computing", *IEEE International Conference on Smart Cloud (SmartCloud)*, (2016), pp.184–189.

[19] Zhang Y, Qiu M, Tsai CW, Hassan MM & Alamri A, "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data", *IEEE Systems Journal*, Vol.11, No.1, (2017), pp.88-95.

[20] Gakwaya NJ & Manju PS, "Big Data Analytics in Healthcare and Delve Bioinformatics Data Space for Health Amelioration", *International Journal of Computer Applications*, Vol.180, No.9, (2018), pp.43-45.

[21] Kashima HN, Yamashita K, Ikai H & Imanaka Y, "Simultaneous modeling of multiple diseases for mortality prediction in acute hospital care", *21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2015), pp.855–864.

[22] Bandyopadhyay S, Wolfson J, Vock DM, Vazquez-Benitez G, Adomavicius G, Elidrisi M, Johnson PE & O'Connor PJ, "Data mining for censored time-to-event data: a bayesian network model for predicting cardiovascular risk from electronic health record data", *Data Mining and Knowledge Discovery*, Vol.29, No.4, (2015), pp.1033–1069.

[23] Hwang K and Chen M. *Big-data analytics for cloud, IoT and cognitive computing*. John Wiley & Sons, (2017).

[24] Chen H, Chiang RH & Storey VC, "Business intelligence and analytics: From big data to big impact", *MIS quarterly*, Vol.36, No.4, (2012), pp.1165–1188.

[25] Basu Roy S, Teredesai A, Zolfaghar K, Liu R, Hazel D, Newman S & Marinez A, "Dynamic hierarchical classification for patient risk-of readmission", *21th ACM SIGKDD international conference on knowledge discovery and data mining*, (2015), pp. 1691–1700.

[26] Villalobos Antúnez, JV (2017). Karl R. Popper, Heráclito y la invención del logos. Un contexto para la Filosofía de las Ciencias Sociales. Opción Vol. 33, Núm. 84. 5-11

[27] M Pallarès Piquer and O Chiva Bartoll (2017). La teoría de la educación desde la filosofía de Xavier Zubiri. Opción, Año 33, No. 82 (2017): 91-113