Journal of
Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7642816

A NOVEL BIG DATA HANDLING APPROACH USING FUZZY RULE BASED CONVOLUTIONAL NEURAL NETWORK

Pasupulati. Sandhya and Dr. Nisarg Gandhewar
Department of Computer Science & Engineering, Dr. A.P.J. Abdul Kalam University, Indore (M.P.) - 452010, India
Corresponding Author Email : psandhya515@gmail.com

ABSTRACT: Due to the fact that healthcare applications produce a lot of data, it is different in terms of volume, diversity, velocity, veracity, and value. There will be issues with the classification of healthcare data as a result of the massive collection of medical data. Therefore, effective big data processing techniques are quickly needed for this type of data. Many of these issues have been effectively resolved using fuzzy systems. This method offers a new way for handling big data that uses a Fuzzy rule-based Convolutional Neural Network (FCNN) classifier to categorize the big data generated in this environment for healthcare services. In this analysis, a Fuzzy Convolutional Neural Network (FCNN) model is used to categorize healthcare data. The information from the big data is gathered by this model, which also does preprocessing, clustering, feature selection, and classification of the data. Using the principal component analysis algorithm, the attributes that are unrelated are eliminated. The presented technique utilizes a clustering strategy based on fuzzy rules. The classification of normal and disease-related data is then efficiently decided using an FCNN classifier. This approach is evaluated using a number of evaluation criteria, including precision, recall, accuracy, and F1-Score. The outcomes gained support the presented scheme's efficacy in relation to several performance evaluation parameters.
KEYWORDS: Healthcare, Big Data, Fuzzy Rule based classifier, Expectation-Maximization (EM).

I. INTRODUCTION

Big data in health informatics can be used to forecast the course of diseases and epidemics, enhance patient care and quality of life, and prevent the development of new diseases and early deaths [1]. Big data also provides information on diseases and warning signals so that appropriate treatment can be provided.

As a result, mortality and co-morbidities will go down, and the government will spend less on healthcare. Because of the huge amount of data that big data will provide, it is particularly helpful not only in clinical care for diagnosis or detection but also in epidemiological approach [2]. Patients are now asking for more information about their options or choices for healthcare and want to be involved in making those decisions [3]. Big data will help in providing patients with current information to help them make the best choice and follow the prescribed course of treatment. Cloud computing is one of the fastest developing technologies in modern technology because of its use in a wide range of applications such as remote healthcare, surveillance systems, weather forecasting, and effective medicine delivery methods [4].

Furthermore, because the data collected from various devices is spatiotemporal in nature, it may be challenging to handle large data sets using traditional database techniques on occasion. As a result, instead of using traditional storage and computer resources, effective centralized storage, such as the cloud, where all data is stored, processed, and available anytime, anywhere, is required [5]. The cloud services are utilized to store patient body sensor data in real time. This information is gathered through a variety of wearable sensors or body sensors that are implemented within the patient's body [6].

Since the development of big data analytics technology, disease prediction from the perspective of big data analysis has received considerable interest. Other studies have mechanically chosen qualities from a wide range of information to increase the accuracy of risk classification rather than pre-selecting features. Data that is part of big data analytics can be structured or unstructured [7]. To address vagueness, uncertainty, and inaccuracy in knowledge representation and reasoning, fuzzy logic offers an intelligent peripheral. Due to their capacity to incorporate human expert knowledge with granular computing, fuzzy systems have been successfully utilized in the healthcare industry to describe the conduct of complex systems without the need for a precise mathematical model. A method for producing data that is relevant for the ultimate decision-making process for the healthcare industry is the fuzzy approach to analysis of healthcare databases [8].

One of the fuzzy systems is the categorization system using fuzzy rules. Fuzzy classifier techniques have become widely used in medical diagnostics in recent years. It has been a challenge for decades to anticipate diseases using data from patient treatment records and health records and applying data mining and Machine Learning. Various implementations in this field used pathological data or medical profiles to mine data using data mining techniques in order to forecast certain diseases. Many strategies are offered by Soft Computing (SC), which places an emphasis on computational intelligence. Computational intelligence has developed intelligent algorithms for knowledge engineering, learning, searching, classification, and other activities. Different approaches, including Fuzzy Logic (FL), Neural Networks (NN), Evolutionary Computing, Probabilistic Reasoning (PR), Support Vector Machines (SV), and their hybrid variations, are included in the family of soft computing. These methods attempted to forecast the onset of disease.

This method of clustering involves counting the members of each value that belongs to a specific cluster and updating the clusters as necessary. It is based on an expectation maximization method that has been updated. In order to categorize fuzzy data and extract expert-specified fuzzy queries, also suggests a CNN classifier based on fuzzy rules. Every time a new record is received after the cluster has been formed, the cloud transmits various parameters to each sub-clouds, where the data is kept in various clusters depending on the membership values. This method is innovative because when a new disease manifests, doctors can transmit its symptoms to the system, and the cloud generates a possible patient with those symptoms. The parameters can be easily quantified by a doctor based on the disease's symptoms. As a result, the parameters that are challenging to quantify exactly are processed and fuzzy-field.

II. LITERATURE SURVEY

Monika Fedorová1, Daniela Perduková , Zdenko Pirnik Viliam Fedák, Ondrej Sukel, And Padmanaban Sanjeevikumar et al. [9] For clinicians in daily practice, a fuzzy system approach has been presented for the analysis of healthcare databases. According to the findings of this system, the fuzzy system approach to healthcare data collected from insurance providers may be an effective technique for creating information helpful for comparable polymorbid patients in the medical field's ultimate decision-making process about drug selection.
Diny Melsye Nurul Fajri, Wayan Firdaus Mahmudy and Yusuf Priyo Anggodo et al. [10] introduced a Tsukamoto fuzzy inference system that improves decision-making for dental recommendation disease. The membership degree function, Particle Swarm Optimization (PSO) was used to achieve better results. From the average accuracy testing, the displayed accuracy point value reached 88 percent.

V.Kavitha and S. Kannudurai et. al [11] A Fuzzy C-means Clustering Algorithm that generates a centroid-based clustering starting with a predetermined set of samples. The methodology supports the low-cost maintenance of all types of health data and delivers the appropriate action to the right patient at the right time. The network model is profitable for patients, payers, providers, and management alike in a healthcare system.

Youjun Bao and Xiaohong Jiang [12] developed a framework for a global recommendation system for medical care that incorporates data mining techniques. The database system module, the data preparation module, the recommendation model module, the model assessment module, and the data visualization module make up the medicine recommendation system. Analyze the SVM (Support Vector Machine), BP Neural Network, and ID3 Decision Tree drug recommendation algorithms based on diagnostic data. To improve the performance of any algorithm, experiments are run to adjust its parameters. In order to achieve a suitable trade-off among model accuracy, model efficiency, and model scalability, an SVM recommendation model is chosen for the drug recommendation module in a given open dataset.

The amount of information that is now available is immense in its raw form; computers must chose and show context-relevant information proactively, yet this feature is difficult and time-consuming.

Paulo Roberto Massa Cereda João José Neto et. al [13] An adaptive automata-based recommendation engine has been offered as a portable, scalable replacement for traditional approaches to resource selection. The method used here focuses on frequency analysis rather than the more typical Machine Learning.

Mohammed et.al. [14] Summarized clinical big data analytics development strategies and raises awareness of the need to improve the output of these technologies. Described how the

Hadoop platform and the MapReduce programming framework could be used to process massive amounts of clinical data in medical health informatics.

Osden Jokonya et.al. [15] Presented a new big data platform to support the mining industry's efforts to prevent, control, and progress HIV (Human immunodeficiency virus) / AIDS (acquired immune deficiency syndrome), TB, and Silicosis (HATS). Pointed to big data integrated architecture that supports in the mining industry's efforts to prevent, advance, and regulate HATS. Outlined a simple big data architecture that supports in understanding how HATS relate to the mining industry. This framework may be able to satisfy the needs of predictive epidemiology, which are important in the mining industry's forecasting and disease control. This created a foundation for the use of big data and workable systems architecture to address the issues of HATS in the mining industry.

III. A NOVEL BIG DATA HANDLING USING FUZZY RULE BASED CNN

The architecture of big Data handling system using fuzzy rule based CNN is shown in Fig. 1. This architecture consists of three major steps namely data acquisition, Fuzzy based clustering and FCNN classification.

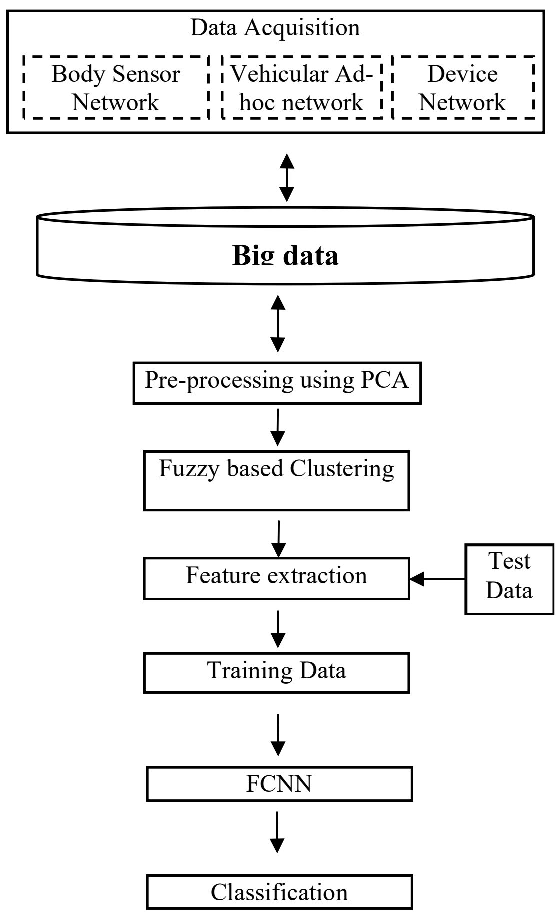

Fig. 1: ARCHITECTURE OF BIG DATA HANDLING SYSTEM USING FUZZY RULE BASED CNN

3.1 Data Acquisition:

The stage of data acquisition is in charge of compiling patient data from various geographic regions. This information may come from a network of body sensors, a specific automotive network, a network of home devices, or a network of hospitals. Either body sensors or sensors installed in the vehicles are used to generate the data from the vehicular network. The majority of these vehicles are ambulances and hospital vans, but if the patients' own cars have the right sensors, they may also submit data from family members. These networks devices either send data separately or can select a network head who is in charge of managing data transmission for each network.

3.2 Principal Component Analysis (PCA):

Using PCA, the data estimation can be minimized. Dimensionality reduction is thought to be a better method for handling high-dimensional data sets. The 2D data set will specifically be changed to 1D. It might also be helpful for dealing with the noisy parts of the data sets. After performing dimension reduction, the grouping calculation is completed. Pre-processing procedures are part of PCA. Prior to running PCA, pre-processing is determined. Calculate the covariance matrix. Correct the eigenvectors. Constructing segments and a module vector. Create the main modules.

The data is initially reduced from n-aspect to k-aspect. The widely recognized are x 1, x 2,..., x n, which are then used to compute the mean.

$$\mu = \left(\frac{1}{n}\right)\sum_{i=1}^{n} x_i \qquad \text{------- (1)}$$

Once the mean has been determined, use the resultant equation to calculate the covariance.

$$C = \left(\frac{1}{n}\right)\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T \qquad \text{---- (2)}$$

In this instance, the number of data elements is n.
The order of the eigenvectors is determined by the following set of criteria:

$$ce_i = \lambda_i e_i \qquad \text{------ (3)}$$

In this instance, n is the total number of data elements.
The eigenvectors are arranged according to the following set of criteria:

$$\det\,(c - \lambda I) = 0 \qquad \text{------ (4)}$$

The eigenvector with the most significant Eigen values is used to deal with the primary components.

3.3 Fuzzy Based Cluster:

The process of clustering is the division of a set of data into a number of meaningful subclasses known as clusters, where elements within a cluster share similar attributes while having different properties from objects in other clusters. The three main areas of clustering in healthcare are cancer cells, genes, and pictures. The rule base contains fuzzy if-then rules, and the database defines membership functions of fuzzy sets that are utilized in fuzzy rules. The fuzzy interface unit generates fuzzy quantities. Crisp quantities are transformed into fuzzy quantities by the fuzzification interface unit. The interface unit for defuzzification transforms the fuzzy quantities into crisp quantities. When the doctor's rules are not precise, fuzzy cluster creation is applied. For instance, it says that the Blood Pressure is high. The fuzzy membership function is used to classify all types of factors to store the initial grouping and updated values of the patient data.

The declining step is extremely sensitive to the structure of the membership function. For faster calculations, the triangular membership function is typically used. Only the membership functions for the input parameters of alcohol, Sp02 level, blood pressure, and heart rate are evaluated. Utilizing a predetermined language rule-base, these inputs are merged to create a fuzzy result. The Centre Of Gravity (COG) approach is used in the defuzzification process to determine the value of the fuzzy output. It is the approach that is most frequently employed and produces accurate defuzzification results.

Following the division of each cloud into clusters and the storage of the original data values in each cluster, any new values entering the cloud are first stored in the cluster in conjunction with an algorithm. Each cluster is a distinct membership function. The most extreme worth of the enrollment capability is put away in the variable max and the comparing group number is put away in the variable. This new data value will eventually be stored in the cluster with the highest membership function.

The modified Expectation-Maximization (EM) method is used to group the quantifiable patient data into clusters. According to Eq.(5), for each data value V, the intermediate function is computed with the kth cluster from the cluster set C = C1, C2,..., Cn . In Eq. (2), this intermediate cluster function is utilized to compute V's membership in the various clusters.

$$\mu_k = (C_k - V)^2 \text{ -------- (5)}$$

The intermediate cluster value of the data value V with the K^{th} cluster in this case is μ_k. For each cluster, the final membership value (Ev) of the data item V stored in the membership matrix M is computed in conjunction with Eq (6). The number of clusters created in this case is n. This is modified EM's E-phase.

$$E_v = \left(\frac{\sum_{j=1}^{n}(\mu_j) - \mu_i}{(n-1) \times \sum_{j=1}^{n}(\mu_j)}\right) ; \forall\{i = 1 \; to \; n\} \text{ -- (6)}$$

The membership matrix, M, comprises these values for each cluster after computing the membership values for each data value using Eq. (6). The cluster centroids are updated using Equation (7) after this membership matrix has been constructed. This process continues until the difference between the previous computed value and the new value is negligible. This is the modified EM M-step.

$$C_j = \left\{\sum_{i,j=1}^{i=n,j=d} (M_{i,j})^2 \cdot V_j\right\} \Big/ \left\{\sum_{i,j=1}^{i=n,j=d} (M_{i,j})^2\right\} \text{ - (7)}$$

Here, $M_{i,j}$ is the membership value for the ith cluster and the jth data value, and C_j is the updated centroid value for the jth cluster. V_j is the jth data value from V1 to Vd.

3.4 Feature selection:

A subset of functions that can be utilized to generate mapping functions from a sample of a particular size to a class of the "same imaginable class" are the focus of this scheme's problem-solving approach. A feature that flows sequentially in time while the quantity of training samples stays constant referred as a stream feature. In this task, qualities are chosen using a Genetic Algorithm based on criteria.

3.5 Fuzzy hybridized CNN (FCNN):

A fusion layer combines data from the fuzzy and CNN models to create an overall representation of the information sequence. Fuzzy and neural statements will both reduce noise and minimize uncertainty on the original data. FCNN is an innovative method for creating a fused representation for final classification that combines fuzzy and neural representations. For challenging classification applications, such as those involving noisy and fuzzy data, FCNN is used. The FCNN experiments are evaluated with a variety of challenging data categorization tasks. The elimination of uncertainties and noises from the original data is made easier by the FCNN method that is being discussed. Finally, the categorization is done using the two alternative representations. The gathered raw data is used as the input.

The applied input data is used as the input for the first layer of training, after which the raw data is taken from it and transformed into a multidimensional matrix. The pooling layer next to the convolution layer performs a maximum-pooling technique to minimize the dimensions by picking coefficients on each sample cell. The other pooling and convolution layers were constructed in the same way. The pooling layer sends its output to the fully connected layer as an input. Features are extracted from the input raw data set using the output from the fully connected layer. The input raw data set is utilized to extract features using the output from the fully connected layer. Depending on the needs, this layer-by-layer procedure might be repeated multiple times. In the final CNN step of the presenting FCNN technique, a sparse regularization penalty is additionally introduced to reduce the quantity and complexity of parameters while processing on layers. The distribution of the weights is made sparse by the sparse regularization penalty. The feature map will then be provided as a on the support vector regression classifier in that it contains a lot of useful data.

3.6 FCNN training:

The FCNN method's training process contains two major steps: parameter initialization and fine-tuning. Due to the symmetry of the entire learning system, the early stages of deep learning are particularly difficult. Effective neural network coverage is the outcome of enabling superior approach. Initialization is carried out for elements like convolutional and fuzzy neural networks. The weights of each layer are located in the deepest area. The processes of categorization and hybridization will be completed. Every node's bias b is set to zero. After that, the weight between the layers is initialized and supplied as follows:

$$Un\left[-\frac{1}{\sqrt{m^{(l-1)}}},\frac{1}{\sqrt{m^{(l-1)}}}\right] \qquad \text{-------- (8)}$$

Un is the even amplitude, and level (l-1) is represented by m^((l-1)). The number of nodules in the last levels of fuzzy and twisted features is counted for the hybridization level using the formula m^((l-1)).

3.7 Evaluation Method:

To begin, refer to TP, FP, TN, and FN as true positive (the number of instances correctly predicted as required), false positive (the number of instances incorrectly predicted as required), true negative (the range of instances properly predicted as not required), and false negative (the range of instances incorrectly predicted as not required), respectively. Along with predetermined evaluation criteria, Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) are used to analyze the benefits and drawbacks of classification. On the ROC curve, the True Positive Rate (TPR) and False Positive Rate (FPR) are equal. If the ROC curve is more closely located near the graph's upper left corner, the model will perform better. The area under the curve is referred to as the AUC. When the area is closer to one, the model performs better.

IV. RESULT ANALYSIS

The entire programme was created using the MATLAB R2013a programming language. On a benchmark dataset for heart disease, assess the presented FCNN's clustering and classification abilities. the dataset was chosen from the UCI (**University of California Irvine**) Machine Learning, Repository at the University of California, Irvine. For the experiments, an Intel Core i7 processor with a CPU utilization rate of 2 GHz and 48 GB of RAM (Random Access Memory) is utilized. Data from training and test sets were compared by the classifier. 60% of the data is used for training and 40% for testing. With 16 convolutional layers constituting the network architecture, the prediction aim for FDCN is set to one-step prediction. To improve the training processes, a mini-batch learning technique is used, using 13 samples in each batch. For the training data on heart disease, the prediction made in accordance with the confusion matrix is highly accurate.

Actual class instances are displayed in columns, while projected class instances are displayed in rows. The system's two confusing classes make it simple for it to categorize people incorrectly. Actual and forecasted two-dimensional data are expressed in the confusion matrix format of Fig. 2.

	Positive	Negative
Positive	TP	FP
Negative	FN	TN

Fig. 2: CONFUSION MATRIX FORMAT

Accuracy, precision, F-measure, and recall are the performance measures that are calculated.

Precision: Precision, which is the proportion of pertinent examples to the retrieved instances, is often referred to as positive predictive value. As indicated in the Eq, the precision is expressed (9).

$$Precision = \frac{TP}{TP+} \text{------- (9)}$$

Recall: The recall rate is calculated by dividing the total number of relevant information discovered by the number of official information actually discovered. Equation (10) expresses the accuracy.

$$Recall = \frac{TP}{TP+FN} \text{--------- (10)}$$

F1-score: The statistical variability known as F-measures performs representation of random errors. As indicated in the Eq, the precision is expressed (11).

$$F - measure = \frac{2TP}{(2TP+FP+FN)} \text{--- (11)}$$

Figure 2 displays a graph plot for the performance measures precision, recall and the FCCN for F1-Socre. It was shown that the FCCN model had an F1-score of 93.5, 92.3% precision, and 97.9% recall. Figure 3 represents a graph plot for FCCN that shows a performance measurement accuracy of 93.2%.

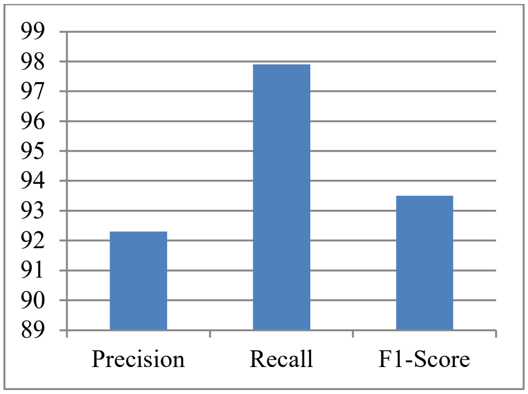

Fig. 3: COMPARATIVE ANALYSIS OF PERFORMANCE EVALUTION

Accuracy is defined as the proportion of predictions that came true to all observations. Using the Eq (12), the accuracy is determined.

$$Accuracy = \frac{(TP+T\ \ )}{(TP+TN+FP+FN)} \quad \text{--- (12)}$$

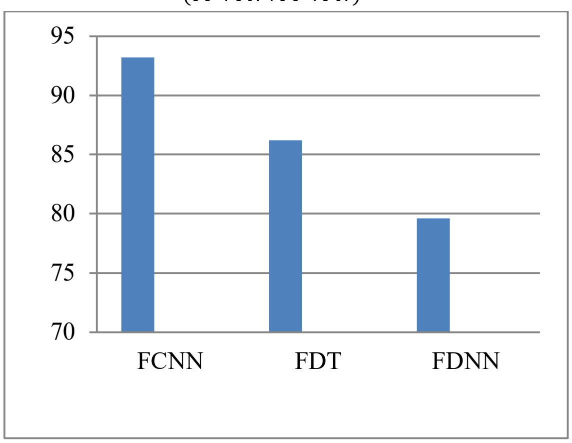

Fig. 4: ACCURACY ANALYSIS

The FCNN analytics performance metrics are presented in Table 1 along with comparisons to other existing approaches such as Fuzzy Decision Trees (FDT) and fuzzy Deep Neural Networks (DNN). Based on the testing data's confusion matrix, these metrics were developed. The FCNN approach looked at the actual class utilized to forecast the existence of heart disease based on features and attributes enhanced throughout classification to the predicted class.

Table 1: PERFORMANCE MEASURES OF FCNN USING CONFUSION MATRIX

Parameter	FCNN	FDNN	FDT
Precision	92.3%	75.8%	81.5%
Recall	97.9%	79.4%	82.6%

F1-Score	93.5%	81.3%	86.6%

Performance metrics including recall, F-score, accuracy, and precision are used to evaluate the findings. When compared to the current FDT (Fuzzy Decision Tree), which obtained accuracy, F1-score of 86.2%, 86.6%, and FDNN (Fuzzy Deep Neural Network), of 79.6%, 81.3%, the FCCN demonstrated superior accuracy, F-score of 93.2% and 93.5%. When compared to the current FDT and FDNN classification approaches, the evaluation findings demonstrated that the FCNN model produced better classification results for heart disease prediction.

V. CONCLUSION

The approach used in this analysis combines CNN and fuzzy clustering to handle healthcare information and support physicians in diagnosing disorders. In order to deliver healthcare services, this analysis presents a brand-new fuzzy rule-based data analyzing method. The presented method uses FCNN to categorize the massive data produced in this setting. The performance of new cluster formation and data retrieval algorithms is evaluated using a variety of assessment criteria, including accuracy, precision, recall, and F1-Score. This approach has demonstrated great accuracy in the diagnosis of cardiac disease in experiments. This method's accuracy was 93.2%; it was higher than that of other antiquated systems. The acquired results confirm the FCNN scheme's efficacy in relation to these evaluation measures. Additionally, this method outperformed its competitors, the Fuzzy Deep Neural Network and the Fuzzy Decision Tree.

VI. REFERENCES

[1] Ramsay JO, Wickham H, Graves S, Hooker G. fda: Functional Data Analysis. R package version 2.4.8.1; 2020

[2] M. Zhang, Y. Chen, and W. Susilo, “PPO-CPQ: a privacy preserving optimization of clinical pathway query for E-healthcare systems,” IEEE Internet of Dings Journal, vol. 7, no. 10, pp. 10660–10672, 2020

[3] F. Ozyurt, E. Sert, and D. Avcı, “An expert system for brain ¨ tumor detection: fuzzy C-means with super resolution and convolutional neural network with extreme learning machine,” Medical Hypotheses, vol. 134, Article ID 109433, 2020.

[4] H. Das, B. Naik and H. S. Behera, "Medical disease analysis using neuro-fuzzy with feature extraction model for classification", Inform. Med. Unlocked, vol. 18, pp. 1-14, 2020.

[5] F. Ozyurt, E. Sert, E. Avci, and E. Dogantekin, “Brain tumor ¨ detection based on Convolutional Neural Network with neutrosophic expert maximum fuzzy sure entropy,” Measurement, vol. 147, Article ID 106830, 2019

[6] R. B. Lukmanto, A. Nugroho and H. Akbar, "Early detection of diabetes mellitus using feature selection and fuzzy support vector machine", Procedia Comput. Sci., vol. 157, pp. 46-54, 2019

[7] R. Venkatesh, C. Balasubramanian, and M. Kaliappan, “Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique”, Journal of Medical Systems, Vol. 43, No. 8, pp.272, 2019

[8] W. Chen, J. An, R. Li, L. Fu, G. Xie, M. Z. A. Bhuiyan, and K. Li, “A novel fuzzy deeplearning approach to traffic flow prediction with uncertain spatial–temporal data features”, Future Generation Computer Systems, Vol. 89, pp.78-88, 2018.

[9] Monika Fedorová1, Daniela Perduková , Zdenko Pirnik Viliam Fedák, Ondrej Sukel, And Padmanaban Sanjeevikumar, (Senior Member, IEEE) ,” The Fuzzy System as a Promising Tool for Drugs Selection in Medical Practice”, IEEE Access, pp. 27294-27301, June 5, 2018.

[10] Diny Melsye Nurul Fajri, Wayan Firdaus Mahmudy and Yusuf Priyo Anggodo (Brawijaya University Faculty of Computer Science Malang, Indonesia), “Optimization of FIS Tsukamoto using Particle Swarm Optimization for Dental Disease Identification”, International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 261 – 268, October 2017.

[11] V.Kavitha and S. Kannudurai, “Health Care Analytics With Hadoop Big Data Processing”, in International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST), Vol.2, Special Issue 15, March 2016, ISSN 2395-695X

[12] Youjun Bao and Xiaohong Jiang (College of computer science, Zhejiang University Hangzhou 310027, China),“An Intelligent Medicine Recommender System Framework”, IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), pp. 1383 – 1388, 2016.

[13] Paulo Roberto Massa Cereda João José Neto ,” A recommendation engine based on adaptive automata”, 17th International Conference on Enterprise Information Systems, pp. 1-8, April 2015.

[14] Mohammed., “Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends”, BioData Mining 2014, 7:22, doi:10.1186/1756-0381-7-22.

[15] Osden Jokonya, “Towards a Big Data Framework for the prevention and control of HIV/AIDS, TB and Silicosis in the mining industry”, in CENTERIS 2014 – Conference on ENTERprise Information Systems/ ProjMAN 2014 – International Conference on Project MANagement/ HCIST 2014 – International Conference on Health and Social Care Information Systems and Technologies, Procedia Technology 16(2014) 1533 – 1541,