

Automatika

Journal for Control, Measurement, Electronics, Computing and Communications



ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/taut20

Diabetes classification using MapReduce-based capsule network

G. Arun & C. N. Marimuthu

To cite this article: G. Arun & C. N. Marimuthu (2024) Diabetes classification using MapReduce-based capsule network, *Automatika*, 65:1, 73-81, DOI: [10.1080/00051144.2023.2284031](https://doi.org/10.1080/00051144.2023.2284031)

To link to this article: <https://doi.org/10.1080/00051144.2023.2284031>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 21 Nov 2023.



Submit your article to this journal [↗](#)



Article views: 660



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



Diabetes classification using MapReduce-based capsule network

G. Arun^a and C. N. Marimuthu^b

^aM.P.Nachimuthu M.Jaganathan Engineering College, Erode, India; ^bNandha Engineering College, Erode, India

ABSTRACT

Big data analytics is a complex exploratory process to uncover hidden data information from vast collections of data. It often provides enormous information from diverse sources and the use of analytics provides confined knowledge from the collected noisy data. In the case of diabetes data, there exist a massive collection of patient data that relates to significant information on patient health and its critical nature. In order of validating and analysing the data to get desired information about a patient and their health risk from the vast collection of data, the study uses bigdata based deep learning analytics. This study uses a Deep Learning Model namely capsule network (CapsNet) is executed on a MapReduce framework. The CapsNet present in the MapReduce framework enables the classification of instances via proper regulations. This model after suitable training with the training dataset enables optimal classification of instances to detect the nature of the risk of a patient. The validation conducted on the test dataset shows that the proposed CapsNets-based MapReduce model obtains increased accuracy, recall, and F-score than the conventional MapReduce and deep learning models.

ARTICLE HISTORY

Received 11 November 2021
Accepted 12 November 2023

KEYWORDS

Capsnets; MapReduce; classification; big data; network; framework

1. Introduction

Data analytics, clustering, healthcare systems, data mining, machine and deep learning, storage, and cloud computing are all seeing significant shifts in their respective directions recently. Health must be well cared for using all of these technologies. The development of statistics improves this since diabetes is a dangerous disease that, in some cases, may cause death. In other words, diabetes has a significant impact on the body's ability to produce insulin, which impairs carbohydrate metabolism and raises glucose levels in the blood. Diabetic patients are most commonly affected by high blood sugar. Those who have experienced high blood sugar have reported symptoms such as increased thirst, increased hunger, and frequent urination [1].

With increasing numbers of patients, social medical data, and sensing data are making medical data records larger today [2]. Additionally, the analysis process will need to be distributed across a distributed system to broaden the scope of the diabetes datasets. Additionally, ML techniques are needed to create models capable of accurately identifying patients according to certain characteristics like glucose, insulin, BMI, or diabetes. This requires building a model with machine learning techniques.

When this data is used correctly, it could help us solve many of our problems at little or no cost. Studies now investigate these topics from a scientific point of

view, based on analysis of historical records in various fields.

Data is now routinely created by businesses and research institutions, as well as governments, at an unprecedented scale and complexity. Organizations all over the world have realized that gathering useful information and advantages from a large quantity of data is crucial. Trying to quickly and easily obtain meaningful insights from this data is difficult. It has thus come to the point where analytics is now impossible to ignore to realize the full value of data and, therefore, improve the overall business performance and increase market share. Data volume, velocity, and variety have increased over the last few years. Technologies in general are not prohibitively expensive, and open-source software makes up the majority of the software [3].

Data analytics is one of the greatest challenges when it comes to missing data, which can be due to sensors, human errors, or the transmission of data from different systems locations, such as cloud servers. In very large datasets, if the number of missing items increases, the dataset needs to be fixed to preserve the power of the data [4,5].

Conventional MapReduce and deep learning models cannot effectively handle the complexity and diverse nature of diabetes patient data, resulting in suboptimal accuracy, recall, and F-score. Previous approaches

CONTACT G. Arun arungguru@gmail.com Department of Computer Science and Engineering, Erode Sengunthar Engineering College, Thudupathi, Perundurai, Erode, India; C. N. Marimuthu muthu_me2005@yahoo.co.in Department of ECE, Nandha Engineering College, Perundurai, Erode, India

failed to leverage the advantages of capsule networks (CapsNet) hindering the ability to uncover critical patient health information from massive collections of data.

The work's primary goal is to advance the process of classifying large datasets with smaller training datasets. In this paper, [6] CapsNet is developed on a MapReduce framework. The CapsNet present in the MapReduce framework enables the classification of instances via proper regulations. This model after suitable training with the training dataset enables optimal classification of instances to detect the nature of the risk of a patient.

The paper's key contribution is listed below:

- The authors used a map-reduce framework to adopt large data into the model for processing. The data volume is effectively reduced using the mapper and reducer function that simultaneously preserves the essential data.
- The study uses CapsNet for making the optimal decision with its simple formulation that focuses entirely on the creation of document illustrations. The CapsNet classifies the needed data to predict the risk of patients from big data with proper pre-processing and feature extraction procedures.

The following is the paper's outline: Review of the literature is presented in Section 2. The suggested big data categorization model is discussed in Section 3. The complete work is evaluated in part 4, and it is wrapped up in section 5.

2. Literature review

Big data analytics has been shown to have numerous issues within the healthcare industry, and there is an ongoing research project that aims to rectify these issues.

According to a study [7], massive data sets are no longer accessible to conventional applications. This research centres on understanding patients with diabetes by using data mining algorithms. Hadoop and MapReduce are used to conduct an in-depth examination.

Healthcare, which must grapple with vast amounts of data, was described by [8]. Important examples were mined using information mining techniques, eliminating incoherent information. Associative Rule Mining is a standard that shows how other things are connected, as well as how things work individually. The Apriori algorithm here debase due to the huge amount of data in the Hadoop MapReduce structure, an Apriori algorithm execution was provided.

[9] revealed that Effective diabetes detection and management strategies were developed. According to the evaluation in this study, the classifications for

diabetes types 1.0 and 2.0 have been finished. To improve diabetes screening and diagnosis, the 5G-Smart diabetes framework calls for the incorporation of modern technology.

[10] focused on big data analytics, which highlights how fast the healthcare division applies Big Data. Classification and clustering algorithms are employed to classify and group diabetes-related information. Moreover, the capability of the equivalent is examined.

This study noted that the recent development of smart DSSs that aim to recreate human characteristics has been mentioned previously [11]. The use of artificial neural networks (ANNs) enables one to handle a wide range of decision-making activities. This model utilizes an ANN, for gestational diabetes detection.

[12] created C4.5, an improved classification framework, and installed reverse planning, enabling the construction of a classification representation. DataSpeak had an accession to mining, classification, and conglomeration on the basis of the results of the study conducted by [13]. This method of overcoming the hindrance of k-nearest neighbour and getting quick access to information overcomes the issue.

[14] developed a classification plan that adheres to the MapReduce programming model and incorporates association rules. Because of the data inspection, which is present in MapReduce, [15] deduced the relationship connection the number of labels for the fuzzy components and the lack of information. It also looked at how the instance set was partitioned to see how fine-grained the results were.

The InfusedHeart Framework [16], employing knowledge-infused learning, outperforms various machine learning and deep learning approaches [17,18], achieving high accuracy in heartbeat acoustic event classification under different signal-to-noise ratio conditions. The FL-PMI leverages DRL and BiLSTM within the FL framework achieving high accuracy, reduced memory usage, and transmission data in smart healthcare applications [19].

[20] proposes a heart and diabetes disease prediction model using a combination of rough set theory for attribute reduction and a fuzzy logic system for classification. The experimental results demonstrate the superiority of the algorithm.

developed an effective Diabetes dataset classification and extraction process. Adaboost.M1 and Logit-Boost.M1 are two of the algorithms that [9] designed for building an apparatus representations for diabetes diagnosis utilizing medical test data. A model to forecast for the prevalence of DM was created using 150 machine learning classifiers and to make an assessment of the increase in clustering.

Several classification techniques were discussed for forecasting patient data, in addition to a general diabetic presentation, as well as for determining some relative diseases. To measure the sorting of diabetes

Table 1. Summary.

Author	Algorithm
Kumar et al. [7]	Hadoop and MapReduce
Bhattacharya [8]	Apriori algorithm
Mamatha et al. [10]	Big Data Classification
Moreira et al. [11]	ANN
Rawal et al. [12]	C4.5
Shankar et al. [13]	DataSpeak
Bechini et al. [14]	MapReduce programming model

chronic diseases, the researchers developed ML techniques like decision trees and random forests are used to create classification system-rooted forecast representation, and also decision tree and random forest to develop diagnostic rule-rooted forecast representation. In addition, an algorithm was created, which was given an RF signature to find difficult regions of type 2 diabetes.

LR and SVM are used to train classifiers. The formation of a relatively low microbial density for the development of infection. The biologic plausibility offered by relevant to the decision that antibiotic strategies aimed at defenseless anaerobic bacteria such as Bacteroidetes and Clostridiales, whose tryptophan supplies are not particularly plentiful, are causally linked to IFI risk in hematologic patients, and gives insight into IFI antimicrobial metabolic re-equilibrium Table 1.

3. Proposed model

Diabetes is a prevalent disease and this necessitates being aware of it as soon as possible [16–20], which ensures the patient's well-being. In the most recent healthcare plan, dealing with shapeless physical condition files will create even more complications. As a result, healthcare is faced with many difficult challenges when attempting to extract information from this kind of data, and this has led to a demand for more advanced data analytics. The goal of this research is to predict the occurrences of datasets to provide predictions that are as accurate as possible to whom the results apply.

The insulin dosage is determined by the patient's prior medical records, and the blood sugar stage is estimated using customary intervals. The platforms Hadoop and MapReduce are used in tandem with CapsNets in this case. Though methods that currently work can process MapReduce using this platform, they are only able to process a subset of attributes. This method is more effective because it incorporates many factors into consideration. Through the use of due to the later installation of the MapReduce platform and the multiclass outlier technique, a suggested hierarchical technique, the diabetes patient database will be successfully examined. Therefore, insulin dosage is determined in accordance with the findings, taking into consideration the patient records and the results from the experiments. An illustration of this is in Figure 1.

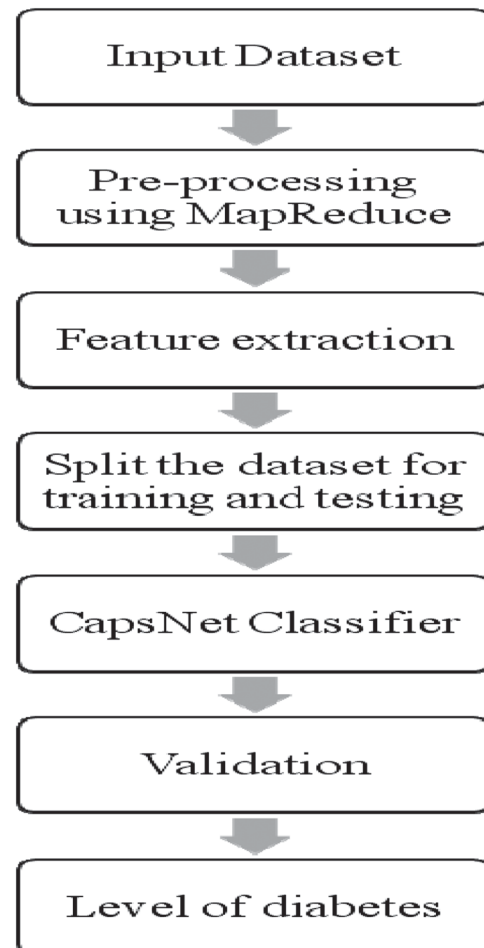
**Figure 1.** Architecture of suggested bigdata analytics.

Figure 1 depicts the architecture of the proposed model, which involves pre-processing steps using map reduction, dataset splitting, and training, followed by CapsNet-based classification and validation. To reduce the rough set theory and a hybrid firefly and BAT optimization technique are used in the model's feature extraction block to reduce the dimensionality of the input data and identify relevant characteristics for further processing and classification using CapsNet.

3.1. Hadoop and map-reduce

The Hadoop ecosystem uses the platform to store, access, and analyse massive quantities of data. The MapReduce Framework is utilized for the parallel processing of data. It is a parallel processing programming technique that is commonly applied to parallel processing. A Hadoop MapReduce instance has two phases: Mapper and Reducer. The results from the map phase are fed into the reduce phase as input.

To reduce the mined data into key-value pairs, the output of the mapper is applied to the input of the MapReduce Framework. The initial step involves extracting a certain number of field values from every record. Also, the value that is no longer there in the dataset is noted. Pairs of key-value pairs are created by

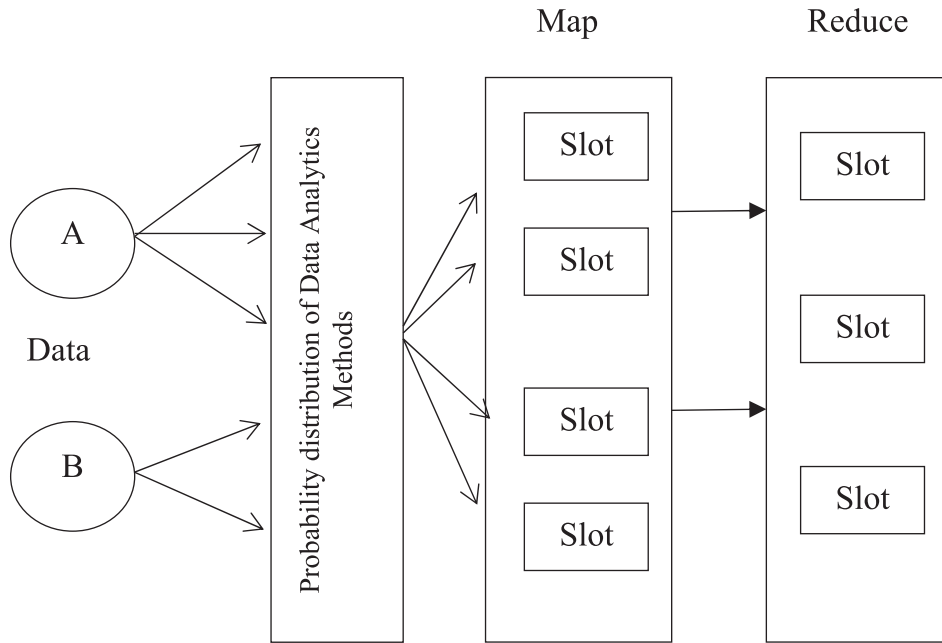


Figure 2. Map reduce framework.

processing all the documentation sequentially and in parallel, and these are equated to pairs of the form as in Equation (1)

$$\text{Map_Phase}(\text{key}_1, \text{value}_1) \rightarrow \text{list}(\text{key}_2, \text{value}_2) \quad (1)$$

Alternatively, you can obtain the key-value pairs of the mapper phase in the text as the input, but instead of reducing the phase. After all the intermediate values have been calculated, the last stage's output is the unified structure of all those values, as in Eq.(2)

$$\begin{aligned} &\text{Reduce_Phase}(\text{key}_2, \text{list}(\text{value}_2)) \\ &\rightarrow \text{list}(\text{key}_3, \text{value}_3) \end{aligned} \quad (2)$$

MapReduce Framework processing is visualized in Figure 2. After reduction by Map-Reduce, the big dataset is smaller and acts as a prerequisite for further processing.

Algorithm 1: MapReduce Framework

```

Input: Diabetic Dataset
Output: Classified Results
Start
Load the Pima Indian dataset
Apply MapReduce to reduce the data dimensionality
Apply the mapping function on the cleaned dataset
Eliminate the PIMA dataset with reduced records and attributes
Start the CapsNet Algorithm
  Enable optimal decisions on finding the critical level of a patient
  Classify the result
  IF the patient has diabetes
    Then check the risk level of diabetes
  IF the patient is not diabetic
    Then do not check the risk levels
End
End

```

3.2. Capsnet

The CapsNet architecture is presented in this section, where we describe how both malignant and benign cancer cells are distinguished in MR imaging. In the CapsNet, capsules are viewed as lot of neurons that represent the activity of parameter neurons. The specific entity is likely present because the vector length of these neurons indicates that.

CapsNet overcame the CNN problems with the connected layers that are subject to refinement of connected layers. It may, however, vary with increased iterations, and the coupling coefficients could be different in the child capsules. Therefore, the parent capsule counts how many more capsules there are after the prediction compared to before the prediction. As illustrated in Figure 3, a CapsNet model was developed.

Considering the u_i of the output as the capsule i , the predicted value for the parent capsule j is as evaluated.

$$u_{j|i} = W_{ij}u_i \quad (1)$$

where $u_{j|i}$ is considered as the forecast vector obtained from capsule output j , i is considered as the capsule that estimates the prediction vector, and W_{ij} is considered a weighting matrix that is applied on a backward pass.

Another layer is computed using the softmax function to calculate the c_{ij} coefficients, using how well the capsules in previous layers matched the ground truth.

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (2)$$

Where b_{ij} is considered as the log-likelihood function, where i functions if capsule i is interconnected with capsule j and which has its setting equal to zero at the start of each iteration.

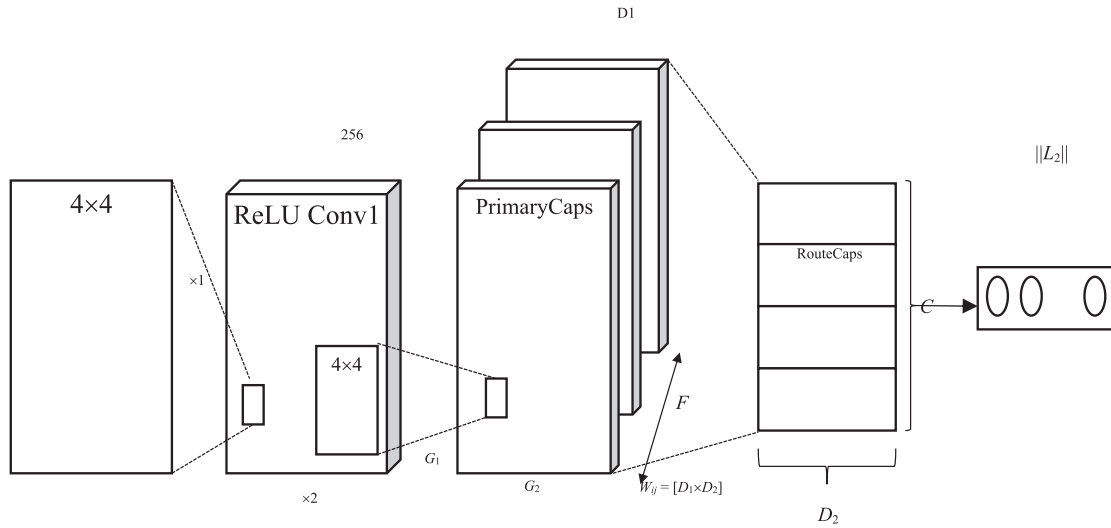


Figure 3. CapsNet model.

To estimate the input parent capsule vector j , we first use the following formula:

$$s_j = \sum_i c_{ij} u_{ji} \quad (3)$$

The output vector of a function goes beyond each output vector of a capsule, and it then uses an initial vector to yield the final result.

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (4)$$

where, s_j is considered the input vector of j^{th} capsule and v_j is considered the output vector of j^{th} capsule.

The internal product of the two vectors is used to update the log probability during the classification in terms of agreement between the output capsules vector produced by the method and the output capsules vector produced without the use of a non-linear function, u_j .

The study uses an agreement function to estimate the estimated value of the log likelihood coupling coefficient is as follows:

$$a_{ij} = v_j \cdot u_{ji} \quad (5)$$

A loss function is present in the final capsule layer k . l_k that calculates how inaccurate its prediction increases the prediction errors to meet the specified level of accuracy. l_k is estimated as follows:

$$l_k = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k) \max(0, \|v_k\| - m^-)^2 \quad (6)$$

Here

$$T_k = \begin{cases} 1 & \text{if } k \text{ exist} \\ 0 & \text{if } k \text{ does not exist} \end{cases}$$

Due to the process of training, the hyperparameters (m^+ , m^- , and λ) are estimated. Thus, true labels

are exceptionally tough to obtain, CapsNet layers are required one for convolutional operations and two for capsule operations.

4. Results and discussions

This part compares the proposed MapReduce-CapsNet model with traditional MapReduce and traditional deep learning models such recurrent neural network (RNN) models to classify the nature of the risk of diabetic disease in a patient.

The entire simulation is conducted in a high-end computing engine with a primary memory of 8 gigabytes of random access memory running on an i5 core generation processor. The study uses the Pima Indian dataset to train and test the entire dataset. Table 2 shows the parameters considered for training and testing the classifier, where The UCI repository is where the information is gathered. The data pieces of 2500 with 15 attributes are used for validation of the classifier after training. The necessity of reducing records and attributes in a small dataset like PIMA that acts as a properly complied training dataset to improve the process of training in this framework.

Table 2. Attribute considered for the study.

S.No	Attribute
1	Age of the patient
2	Gender of the patient
3	Plasma glucose postprandial level
4	Plasma glucose fasting level
5	Pregnancy in women
6	Blood glucose level
7	Blood pressure level
8	Skin thickness value
9	Insulin level
10	Body Mass Index level
11	Diabetes pedigree function
12	Serum creatinine level
13	Serum sodium level
14	Serum potassium level
15	Hemoglobin A1c, a blood pigment that carries oxygen

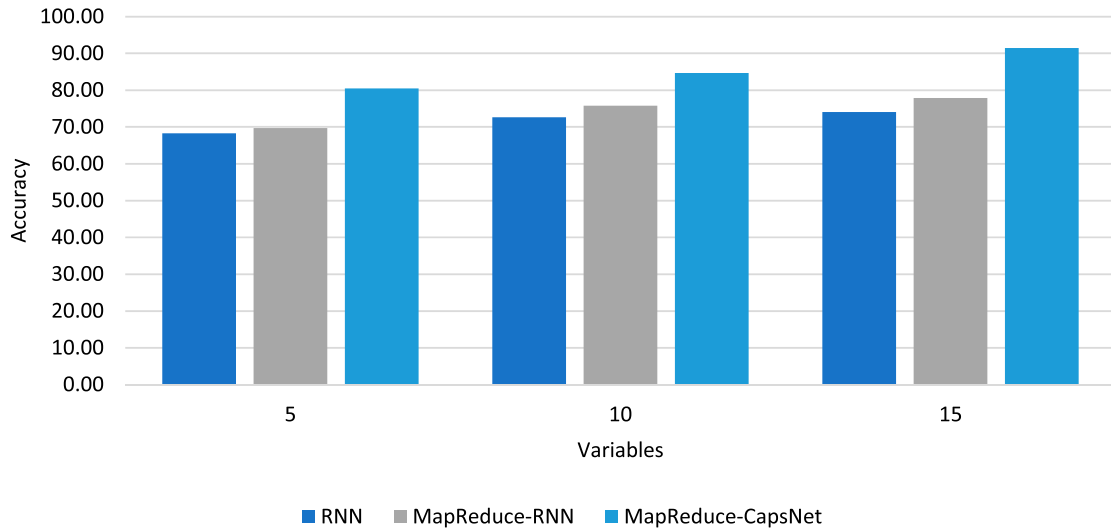


Figure 4. Accuracy.

The study does not accommodate the same dataset for training and testing, 786 valid items are employed for training, and 70% of the remaining 30% are utilized to test the classifier. A 10-fold cross-validation is conducted on the high-end computing system using the MapReduce – CapsNet model.

The MapReduce-CapsNet for the diabetic classification is compared with conventional MapReduce and RNN. The validation helps to detect the risk of the patient diabetic level using the given classifier. The measures used to evaluate the effectiveness of the following are a list of suggested classifiers:

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

$$\text{F-measure} = \frac{TP}{(TP + 0.5(FP + FN))}$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

$$\text{Sensitivity} = \frac{TP}{(TP + FN)}$$

$$\text{Specificity} = \frac{TN}{(TN + TP)}$$

where, represents the Total number of iterations; A_t represents the Actual value; F_t represents the Forecast value; TP stands for True Positive and FP for False Positive; TN stands for True Negative; FN stands for False Negative.

Figure 4 shows comparison of the MapReduce-CapsNet classifier's accuracy results with those of the traditional MapReduce-RNN and RNN classifier. The MapReduce-CapsNet provides a multi-variable classification of input diabetic datasets that enables optimal classification with supervised learning in a 10-fold cross-validation. The results of testing in Figure 4 show that the MapReduce-CapsNet classifier achieves a higher accuracy level than the MapReduce-RNN and RNN classifier.

The MapReduce-CapsNet classifier's F-measure findings are displayed in Figure 5 with the conventional MapReduce-RNN and RNN classifier. The MapReduce-CapsNet provides a multi-variable classification of input diabetic datasets that enables optimal A 10-fold cross-validation that classifies a greater rate of F-measure as sensitivity and higher specificity. The results of testing in Figure 6 show that the MapReduce-CapsNet classifier achieves a higher F-measure level than the MapReduce-RNN and RNN classifier.

Figure 6 displays the MapReduce-CapsNet classifier's G-mean results using the traditional MapReduce-RNN and RNN classifier. The MapReduce-CapsNet provides a multi-variable classification of input diabetic datasets that enables optimal classification with reduced errors in a 10-fold cross-validation. The results of testing in Figure 6 show that the MapReduce-CapsNet classifier achieves a higher G-mean level than the MapReduce-RNN and RNN classifier.

Figure 7 shows the MAPE results of the MapReduce-CapsNet classifier with the conventional MapReduce-RNN and RNN classifier. The MapReduce-CapsNet provides a multi-variable classification of input diabetic datasets that enables optimal classification with a sigmoid loss function in a 10-fold cross-validation enabling reduced MAPE. The results of testing in Figure 7 show that the MapReduce-CapsNet classifier achieves reduced MAPE than the MapReduce-RNN and RNN classifier.

Figure 8 shows comparison of the MapReduce-CapsNet classifier's sensitivity findings with those of the traditional MapReduce-RNN and RNN classifier. The MapReduce-CapsNet provides the multi-variable classification of input diabetic datasets that enables positive samples in a 10-fold cross-validation. The test findings shown in Figure 8 demonstrate that the MapReduce-CapsNet classifier achieves a higher sensitivity level than the MapReduce-RNN and RNN classifier.

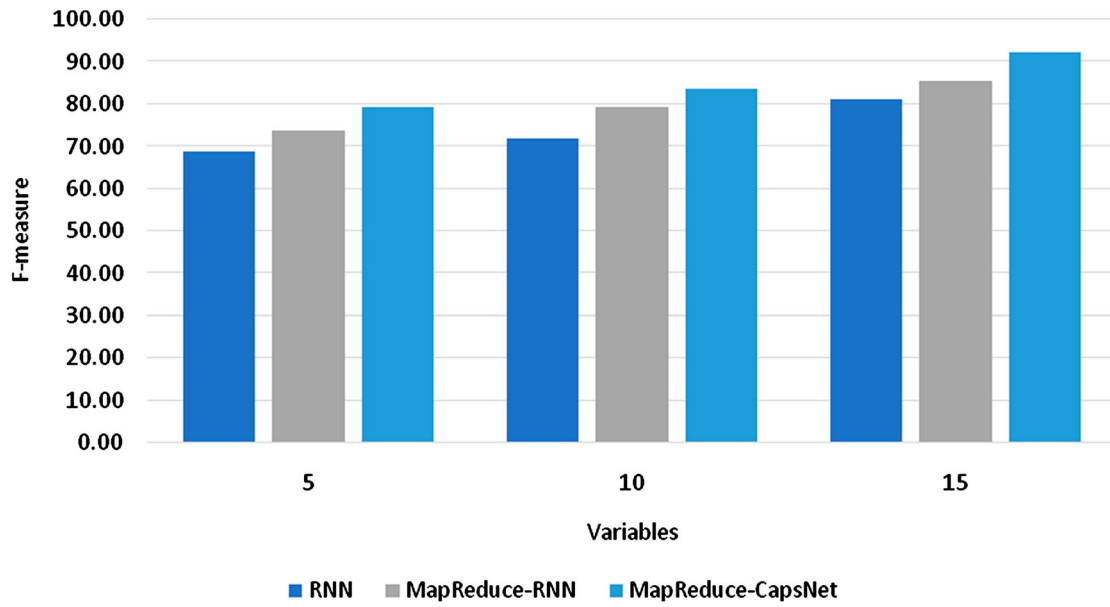


Figure 5. F-measure.

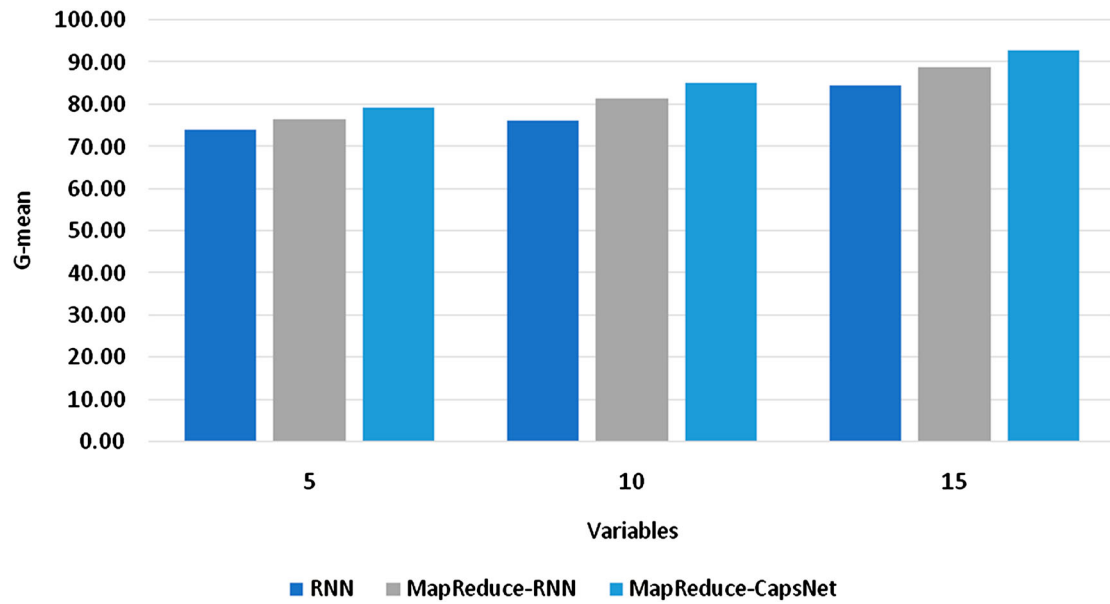


Figure 6. G-mean.

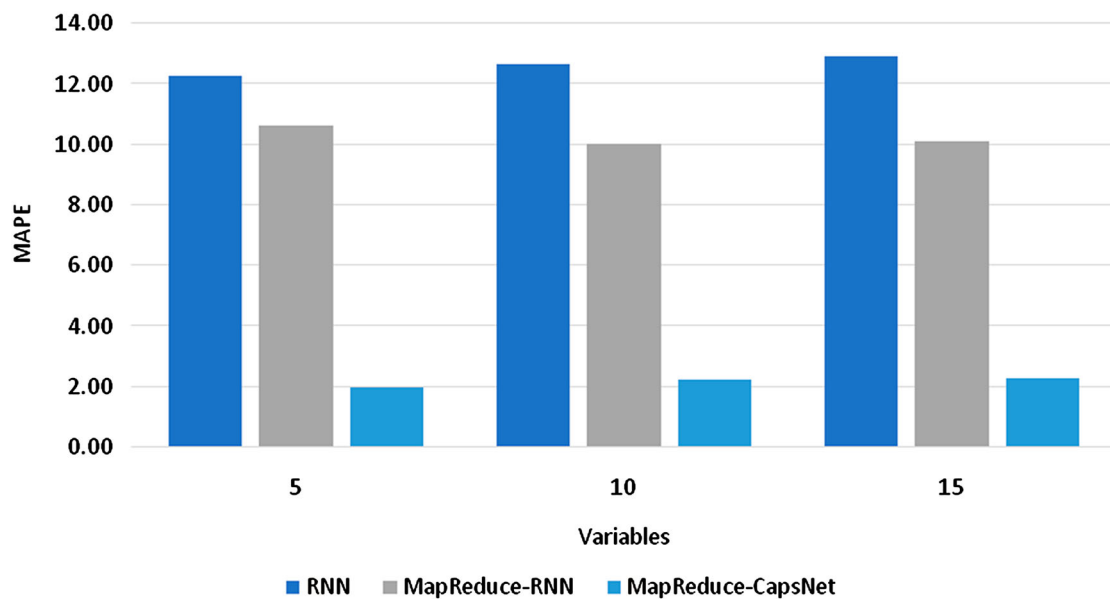


Figure 7. MAPE.

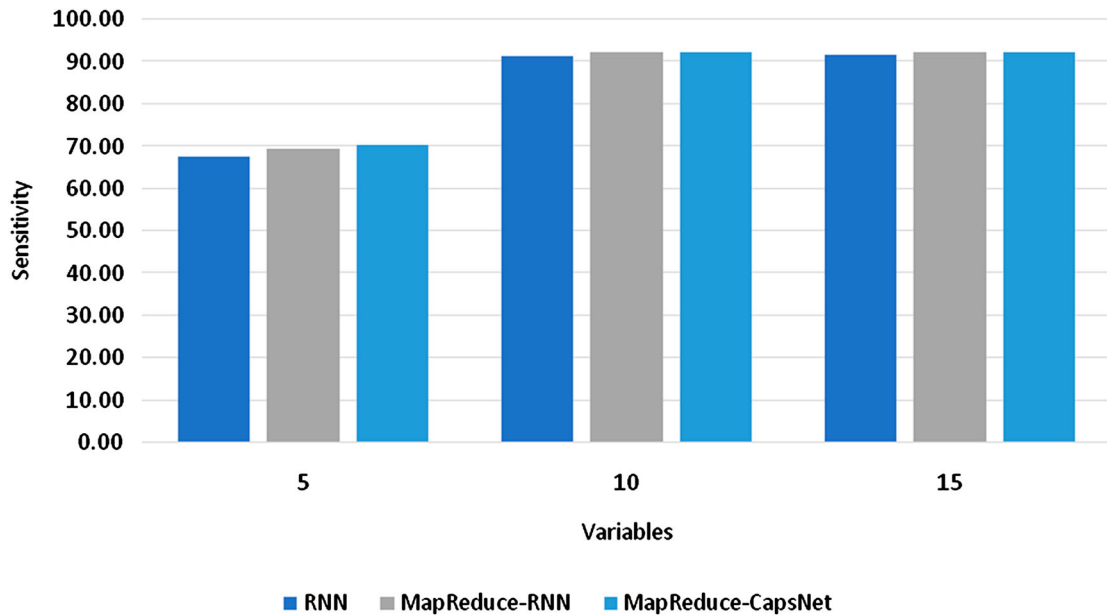


Figure 8. Sensitivity.

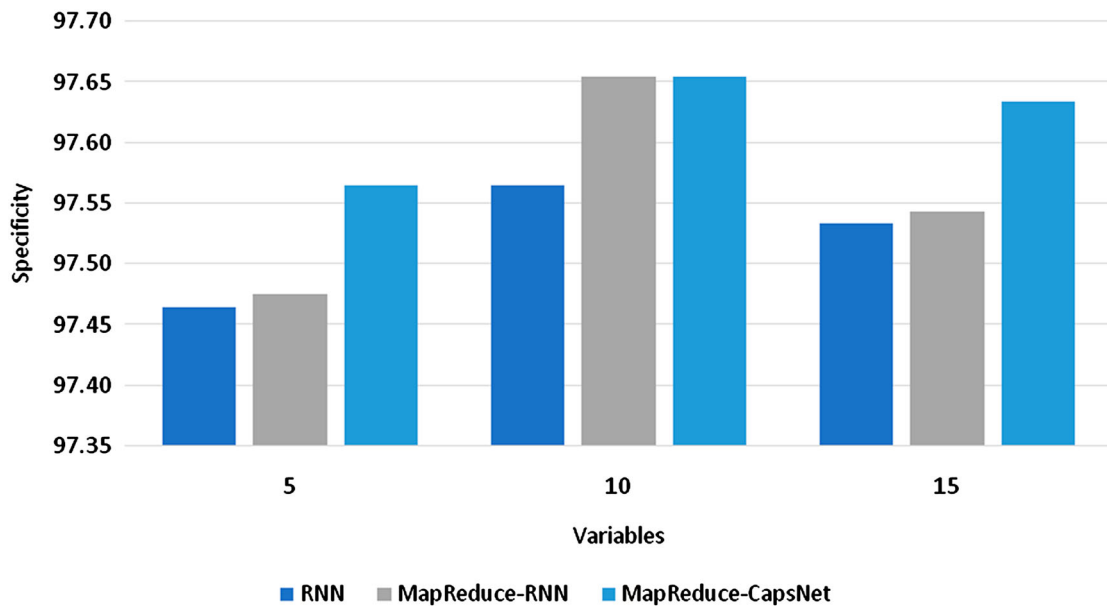


Figure 9. Specificity.

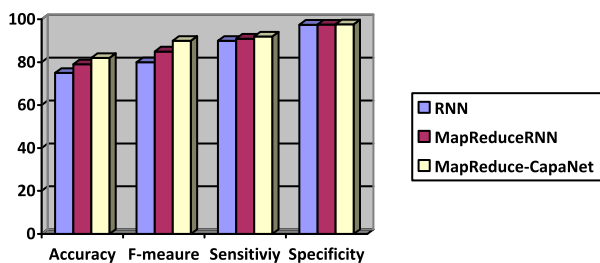


Figure 10. Comparative analysis.

Figure 9 shows comparison of the MapReduce-CapsNet classifier’s particular results with those of the traditional MapReduce-RNN and RNN classifier. The MapReduce-CapsNet provides a multi-variable classification of input diabetic datasets that enables optimal tenfold cross-validation classification for true samples

that were negative. The results of testing in Figure 9 show that the MapReduce-CapsNet classifier achieves a higher specificity level than the MapReduce-RNN and RNN classifier Figure 10.

From the results, it is found that with increasing variables, the accuracy tends to get increased and the increase of training data improves the rate in which the instances are classified.

5. Conclusion

In this paper, MapReduce-based CapsNet enables optimal classification of diabetes conditions from big data. The MapReduce-based CapsNet is undergone with suiTab training that enables improved diabetes data classification and detects the nature of the risk of a

patient. The simulation on the test datasets shows that the MapReduce-based CapsNets framework obtains increased classification accuracy, recall rate, and F-score than the conventional MapReduce and deep learning algorithms like RNN and DenseNet. In the future, the application of large datasets can inclusively be added to the dataset to improve the training rate.

Ethics approval and consent to participate

No participation of humans takes place in this implementation process.

Human and animal rights

No violation of Human and Animal Rights is involved.

Data availability statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Authorship contributions

All authors are contributed equally to this work.

Acknowledgement

There is no acknowledgement involved in this work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

No funding is involved in this work.

References

- [1] Nibareke T, Laassiri J. Using big data-machine learning models for diabetes prediction and flight delays analytics. *J Big Data*. 2020;7(1):1–18. doi:10.1186/s40537-020-00355-0
- [2] Hosseini MM, Zargoush M, Alemi F, et al. Leveraging machine learning and big data for optimizing medication prescriptions in complex diseases: a case study in diabetes management. *J Big Data*. 2020;7(1):1–24. doi:10.1186/s40537-020-00302-z
- [3] Lin HC, Kuo YC, Liu MY. A health informatics transformation model based on intelligent cloud computing – exemplified by type 2 diabetes mellitus with related cardiovascular diseases. *Comput Methods Programs Biomed*. 2020;191:105409. doi:10.1016/j.cmpb.2020.105409
- [4] Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discovery*. 2016;15(7):473–484. doi:10.1038/nrd.2016.32
- [5] Kousik NV, Sivaram M, Mahaveerakannan R. Improved density-based learning to cluster for user web log in data mining. *Inv Comput Inform Technol*. 2021: 813–830. doi:10.1007/978-981-33-4305-4_59
- [6] Sara SBV, Anand M, Priscila SS, et al. Design of autonomous production using deep neural network for complex job. *Mater Today Proc*. 2021.
- [7] Kumar S, Singh M. Diabetes data analysis using MapReduce with Hadoop. *Eng Vib Commun Inform Process*. 2019: 161–176. doi:10.1007/978-981-13-1642-5_15
- [8] Bhattacharya N, Mondal S, Khatua S. A MapReduce-based association rule mining using Hadoop cluster—an application of disease analysis. *Innov Comp Sci Eng*. 2019: 533–541. doi:10.1007/978-981-13-7082-3_61
- [9] Chen M, Yang J, Zhou J, et al. 5G-smart diabetes: toward personalized diabetes diagnosis with healthcare big data clouds. *IEEE Commun Mag*. 2018;56(4):16–23. doi:10.1109/MCOM.2018.1700788
- [10] Mamatha N. Analysis and detection of diabetes using data mining techniques—a big data application in healthcare. In *Proceedings of Fifth International Conference on Emerging Research in Computing, Information, Communication and Applications (ERCICA)*. 2019.
- [11] Moreira MWL, Rodrigues JPC, Kumar N, et al. Evolutionary radial basis function network for gestational diabetes data analytics. *J Comput Sci*. 2018;27:410–417. doi:10.1016/j.jocs.2017.07.015
- [12] Rawal B, Agarwal R. Improving accuracy of classification based on c4. 5 decision tree algorithm using big data analytics. *Comput Intell Data Min*. 2019: 203–211. doi:10.1007/978-981-10-8055-5_19
- [13] Shankar VG, Devi B, Srivastava S. DataSpeak: data extraction, aggregation, and classification using big data novel algorithm. In *Computing, communication and signal processing: Proceedings of ICCASP 2018*. Singapore: Springer; 2019. p. 143–155.
- [14] Bechini A, Marcelloni F, Segatori A. A MapReduce solution for associative classification of big data. *Inf Sci (Ny)*. 2016;332:33–55. doi:10.1016/j.ins.2015.10.041
- [15] Fernández A, del S, Bawakid A, et al. Fuzzy rule based classification systems for big data with MapReduce: granularity analysis. *Adv Data Anal Classif*. 2017;11(4): 711–730. doi:10.1007/s11634-016-0260-z
- [16] Pandya S, Gadekallu TR, Reddy PK, et al. InfusedHeart: a novel knowledge-infused learning framework for diagnosis of cardiovascular events. *IEEE Trans Comput Soc Syst*. 2022.
- [17] Arikumar KS, Prathiba SB, Alazab M, et al. FL-PMI: federated learning-based person movement identification through wearable devices in smart healthcare systems. *Sensors*. 2022;22(4):1377. doi:10.3390/s22041377
- [18] Gadekallu TR, Gao XZ. An efficient attribute reduction and fuzzy logic classifier for heart disease and diabetes prediction. *Recent Adv Comput Sci Commun (Formerly: Recent Patents on Computer Science)*. 2021;14(1):158–165. doi:10.2174/2213275911666181030124333
- [19] Kumar Das S, Kumar Mishra A, Roy P. Automatic diabetes prediction using tree based ensemble learners. *Int J Comput Intell & IoT*. 2019;2(2).
- [20] Das SK, Roy P, Mishra AK. Oversample-select-tune: a machine learning pipeline for improving diabetes identification. *Concurr Comput Pract Exp*. 2022;34(5):e6741. doi:10.1002/cpe.6741