CHINESE JOURNAL OF MEDICAL GENETICS
ISSN: 1003-9406 Vol. 33 Iss. 1 2024

SECURED PREDICTION MODEL USING IOT BASED SMART DEVICES FOR HEALTHCARE ASSISTANCE

S Senthamarai [1] R Mala [2]

[1] Research Scholar, Department of Computer Applications, Alagappa University, Karaikudi, Tamilnadu

[2]Asst.Prof, Department of Computer Science, Government Arts and Science College for Women, Paramakudi, Tamilnadu

Abstract

Over the past three decades, the internet's expansion and the advancement of technology have altered the world. The Internet of Things (IoT) is a network of networked devices that can collect and share data using built-in sensors and communication protocols. There are many different device types included in the IoT, some of which could be useful for scientific research. The Internet of Things (IoT) enhances the Web by deploying ubiquitous devices with embedded identification, sensing, and data exchange capabilities. These intelligent objects provide the basis of adaptable cyber-physical networks that provide a platform for networked data transmission. In this work, we suggest a novel method for predicting and analyzing stroke severity in older adults over 65 that is based on the National Institutes of Health Stroke Scale (NIHSS). The input dataset is raw and may contain missing values and redundant packets when first gathered. After data has been acquired, normalize the data using Label Encoder Min-Max. Advanced Encryption Standard (AES) is a popular symmetric encryption technique that offers secure data encryption and decryption after test and training. With the use of the Linear Discriminant Analysis (LDA) method, we were able to extract features from the data and acquire different data properties. Next, we used Singular Value Decomposition to choose the features. Finally, we suggested that Genetic Decision Classifier (GDC) be used to forecast an individual's health condition based on the features chosen. Our method may produce 96% accuracy, 93% precision, 97% recall, and 95% f1-score.

Keywords:Internet of Things(IoT), smart devices,Linear Discriminant Analysis (LDA),Genetic Decision Classifier (GDC)

1. Introduction

The fast development of electronic devices like smart phones and tablets that allow for physical or wireless connection has made them an integral component of modern life [1]. The Internet of Things allows us to connect to anything, access everything at anytime from anywhere, and immediately get data on any item. IoT aims to maximize the advantages of the Internet by providing capabilities like remote control, data sharing, constant connection, and others [2]. IoT, commonly referred to as the Internet of Everything, was initially used by Kevin Ashton in 1999 to describe a scenario in which every physical item will be linked to the Internet through ubiquitous sensors. Digital oilfields, home and building automation, the intelligent grid, digital medical care, intelligent transportation, etc. are just a few examples of the domains of life where IoT technology is being used.

IoT in healthcare involves connecting medical devices, wearables, sensors, and other equipment to a network, allowing healthcare providers to gather real-time data, track patient vitals, monitor medication adherence, and manage chronic conditions more effectively [3]. These interconnected devices can communicate and collaborate, as well as with healthcare professionals, to deliver personalized and timely care. IoT also plays animportant role in the automation of healthcare processes [4]. Smart hospital systems can track and manage inventory, optimize resource allocation, and streamline workflows, leading to increased efficiency and cost savings. Real-time locations systems can help locate medical equipment and personnel quickly, ensuring a smooth workflow and reducing response times in emergencies [5].

Numerous types of research are being undertaken to identify strokes and utilize risk factors to prevent recurrence in light of a recent study that demonstrated the NIHSSobjectively validates the severity of stroke [6]. Since its first publication in 1972, the Mathew scale has been joined by additional stroke assessment tools, such as the European Stroke Scale, the Scandinavian Stroke Scale, and the NIHSSof the United States [7]. Researchers have shown that stroke predictions may be made quickly without imaging by using a scoring system based on the NIHSS, which accurately predicts cortical damage in an acute ischemic stroke. In this research, we propose a novel method for predicting and analyzing stroke severity in people over the age of 65 by combining NIHSS features with the GDC algorithm.

2. Related works

The paper [8] suggested a new IoT architecture for keeping track of large amounts of sensor data (big data) for medical purposes. The suggested framework is split into two major parts: the Meta Fog Redirection (MFR) architecture and the Grouping and Choosing (GC) architecture.The study [9] provided a brief survey of the broad application of IoT solutions in healthcare, covering ground from the first wearable sensor-based health monitoring systems to the most current advances in edge computing for smart health.The paper thoroughly analyzed the most recent emerging technologies in personalized healthcare systems, emphasizing cloud computing, fog computing, big data analytics, the Internet of Things, and mobile apps.The research [10] inspired the need to present an Empirical Intelligent Agent (EIA) based on a novel Swarm-Neural Network (Swarm-NN) approach for detecting adversaries in an edge-centric Internet of Things (IoT) environment.The article [11] described the current state of smart healthcare and the technology that underpins it in several crucial contexts. After elaborating on the state of smart healthcare, they offer our best ideas for how to improve the system. Then, they conclude by discussing the potential of smart healthcare in the future.

The report [12] discussed the potential applications of the Internet of Things in healthcare and medicine and displayed a detailed design of an IoT health ecosystem. The soaring costs of caring for an ageing population plagued by chronic conditions are making healthcare administration increasingly challenging. The soaring costs of caring for an aging population plagued by chronic conditions are making healthcare administration increasingly challenging.The study [13] presented a literature assessment covering publications published up till 2018 on healthcare services based on the Internet of Things. The advantages and disadvantages of the studied mechanisms have also been discussed, and the primary difficulties in improving IoT strategies over health care services in the future have been underlined.The purpose of the study [14] created and empirically evaluated a theoretical framework for identifying the most important characteristics that influence the acceptability of smart home services for healthcare among the elderly. The researchers in four Asian nations surveyed 254 persons 55 and older by online questionnaire.The paper [15] presented a smart healthcare monitoring framework that is cloud-based and interacts with the surrounding smart devices, surroundings, and smart city stakeholders to make healthcare more inexpensive and accessible.

3. Methods

The proposed methodology will consist of the following sequence of tasks. Figure 1 shows the flow of the proposed approach.

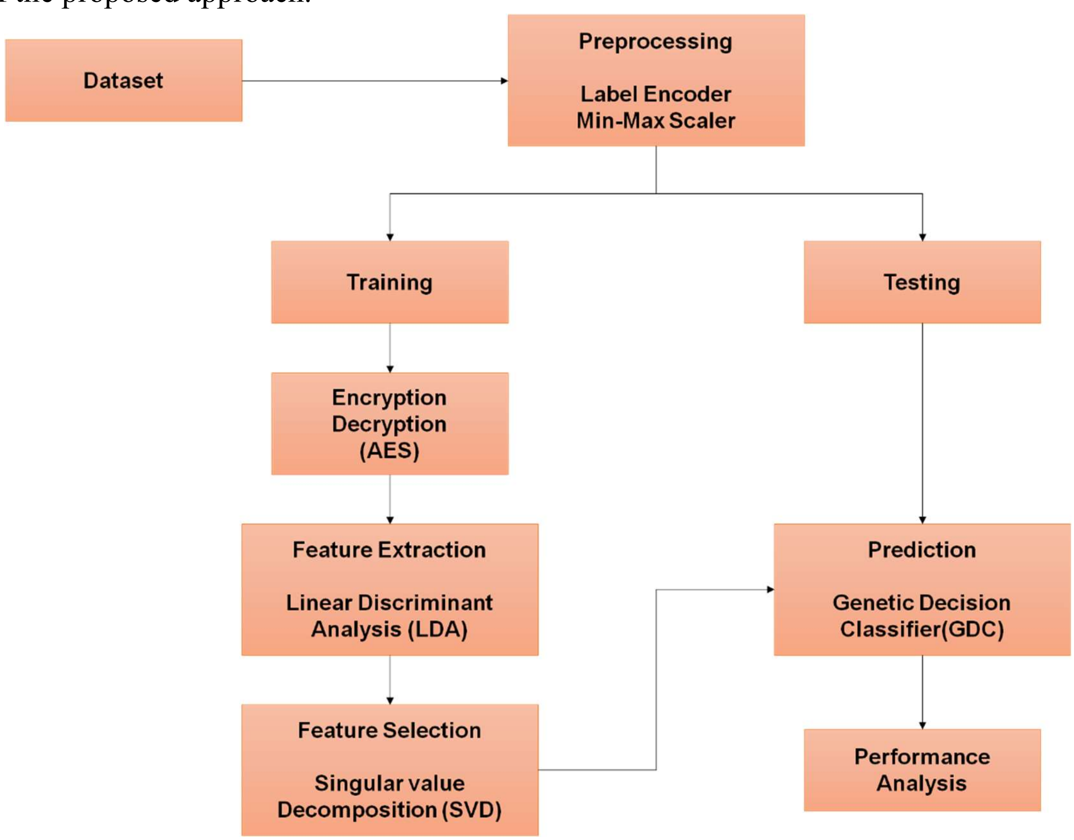

Figure 1: Flow of the suggested approach

Using machine learning methodology the dataset is trained and the prediction model will be generated. The module generated will be used to predict using the National Institutes of Health Stroke Scale, we conduct an in-depth examination of stroke severity in people aged 65 and above. We have collected the stroke data set from Kaggle.

3.1.Preprocessing using Min-Max normalization

After the data is collected using Min-Max normalization for preprocessing the data. Min-Max normalization is a technique of normalizing that uses linear modifications to the original data to provide a fair comparison of values before and after the procedure.

$$Y_{new}=\frac{Y-\min(Y)}{\max(Y)-\min(Y)} \quad (1)$$

Y_{new}=The adjusted value obtained after scaling the data
Y=outdated value
Max(Y) =Dataset's highest possible value
Min(Y) = Dataset's lowest possible value

3.2.Data encryption and decryption using AES

The Advanced Encryption Standard (AES) is a widely used symmetric encryption algorithm that provides secure encryption and decryption of data. AES operates on blocks of data, with a fixed block size of 128 bits (16 bytes). It supports three key sizes: 128 bits, 192 bits, and 256 bits. The key size determines the strength of the encryption, with longer key sizes providing greater security. To encrypt data using AES, the plaintext is separated into blocks and processed sequentially. Every block is XORed with the encryption key, and then a series of encryption rounds are performed to transform the block using the key. The resulting cipher text block is the encrypted version of the original plaintext block.

Decryption is the reverse process of encryption. The cipher text block is XORed with the decryption key, and a series of decryption rounds are performed to transform the block using the key. The resulting plaintext block is the decrypted version of the original ciphertext block.

3.3.Feature extraction using LDA

We utilized Linear Discriminant Analysis (LDA) to extract the data from the data after it had been collected. LDA, which Fisher first presented in 1936, is one of the first discriminant analysis techniques. This approach is predicated on the notion that each class's probability distribution follows a Gaussian (normal) distribution. Along with the normality assumption, the LDA also relies on the definition of a priori probabilities

$\pi_i$for each of the J classes. For example, the training set may estimate these probabilities as M_i/Mor as $1/I$ for all classes set equal. The second strategy will be used in this essay. Each sample is allocated to the group with the greatest posterior probability by the Bayes rule. This indicates that each sample is allocated to the class j that produces the least value of

$D_i$under the aforementioned suppositions.

$$D_i = (w_j - \mu_i)^s \sum^{-1}(w_j - \mu_i) + \log|\Sigma| - 2\log(\pi_j) \quad (2)$$

Where μ_is denotes the class means and Σdenotes the variance/covariance matrix shared by all classes. Be aware that this condition only applies to Mahalanobis distances when the prior probability for each class is identical. The data must be used to estimate the mean and covariance matrices. Typically, the group means w_j is used to determine the means. The following estimate $\Sigma$is often used for the common covariance matrix:

$$T = \sum_{i=1}^{i} \frac{(M_i - 1)T_i}{(M - I)} \quad (3)$$

where $T_i$is the class j empirical variance-covariance matrix. LDA's key drawback is that it needs a covariance matrix with good conditioning. This indicates that the approach is inapplicable not situations when several variables or samples are more than a few variables or when the variables are strongly linked.

3.4.Feature selection by Singular value decomposition

To select features, we used singular value decomposition (SVD). Amxc data matrix with r linearly independent columns is given, and it is denoted as $W \in ℚ^{mxc}$. $W = V\sum U^S$Become the SVD for W. The singular vector matrices, however, are not singular, despite the singular values being unique. Both the Span (W) and the span of the first r columns of V are equal. The first r columns of V serve as another illustration of W.We only need one of the $\frac{C!}{(c-q)! \times q!}$bases that Column Space of X has to offer. This goal is accomplished by using a recently created feature selection method based on matrix factorization. Using this technique, the W is factored into WX G as follows:

$$\underset{X,G}{\operatorname{argmin}} ||W - WXG||_E^2 \quad (4)$$

Where J is an index set, and X is an indicator matrix obtained from J as follows:

$$x_{ji} = \begin{cases} 1 & \text{if i th element of J is j,} \\ & 0 \text{ otherwise} \end{cases} \quad (5)$$

The several rows in X equal the total number of features. For instance, if there are 5 features and

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (6)$$

Then, $W_j = WX$indicates that traits 1, 2, and 5 have been chosen (J = 1, 2, 5). The conditions $X \geq$ oand $X^S X = F$guarantee that X only contains entries that are zeros or ones, with a maximum of one nonzero element per row or column. Which of the $W_J$in Problem 1's matrix is the best, whether anything is considered the "best" depends on how we define it. Two factors are crucial in supervised feature selection algorithms: A maximum of one characteristic is redundant with the others. The highest possible correlation between the selected features and the target values. We want to choose r linearly independent columns of Wusing the suggested strategy. No linear combination of the other $q - 1$columns of W_j can create $e_j$for any$e_j \in W_j$; hence linear independence is taken into account as a redundancy metric in this article. The optimum approximation for $W_J$is V, according to the previous discussion of the SVD, to ensure the least amount of repetition. Thus, the following formulation of Problem (4) is possible:

$$\underset{J}{\operatorname{Argmin}} ||W - WXG||_E^2 \quad (7)$$

In which $W = \sum U^S$. This research aims to identify the most suitable replacement for V. The following expression is used to update V to new$V = W_J$.

ISSN: 1003-9406 Vol. 33 Iss. 1 2024

$$U^s \simeq (W_J \Sigma)^s W \quad (8)$$

Where (W_J) is the Moore-Penrose pseudo inverse of $W_J$and it is produced from $W \simeq W_J \Sigma^s$.
The selected characteristics should be able to forecast target values to meet maximum relevancy, hence this criterion is written as

$$\text{Arg}\max_x ||X^s - \ominus||_E^2 \quad (9)$$

Since the purpose of this term is to pick columns of Wthat most closely match the desired values, the IG values of the features that are cut are stored in the X^svector. The objective function, by the aforementioned criteria, reads as

$$\text{Arg}\min_x ||W - WXG||_E^2 - \left| |W^s - \ominus| \right|_E^2$$

$$X \geq 0 \text{ and } X^s X = F_m \quad (10)$$

Problem (10) simultaneously ensures the highest responsiveness of chosen features as well as their linear independence.

3.5.Genetic Decision Classifier (GDC)

An optimization algorithm called a genetic algorithm takes its cues from the principles of genetic evolution and natural selection. Natural selection is a method often employed to quantitatively assess technology choices in a building retrofit project. Because the fittest people have a superior chance of surviving and passing on their genes to the next generation.

A general description of a genetic algorithms operation is given below:

Step 1 (Initialization): Make a starting population of people (possible options). Each person is represented by an x number that falls between $[0, 10]$. The problem dictates the population size, which may be chosen depending on the preferred exploration-exploitation trade-off.

Step 2 (Evaluation): Calculate the value of f(x) for each x value to determine the fitness of each member of the population. Since we want to maximize this function, the fitness function in this situation is just $f(x) = x^2$.

Step 3 (Selection): Each person's selection likelihood varies in direct proportion to how fit they are. Users may choose from several selection techniques, including roulette wheel selection,rank-based selection, and tournament selection.

Step 4 (Reproduction): To produce offspring, we utilized genetic operators like crossover and mutation. Crossover is the procedure of producing a new person by fusing the genetic material of two parents. An individual's genes undergo random alterations as a result of mutation. These operators aid in the exploration of fresh areas of the search space while retaining certain traits of the fittest individuals.

Step 5 (Replacement): Replace the existing population with the next generation. Both parents and children will be part of the new population. By taking this measure, it is made sure that the fittest people have a better opportunity of passing on their genetic makeup to the next generation.

Step 6 (Termination): For a specified number of generations or until a termination requirement is satisfied, repeat steps 2 through 5. A maximum number of iterations, achieving a desirable degree of fitness, or a combination of criteria might serve as the termination criterion.

The GDCalgorithm searches the search space by repeatedly repeating these stages until it finds the value of x that maximizes the function $f(x) = x^2$. The most physically fit person in the ultimate population will be the answer.

The GDCalgorithm is a kind of supervised learning algorithm used in machine learning. This decision-making system is shown as a tree diagram. A decision tree takes as input a set of criteria and displays the result as true or false. This method is reportedly easier to implement and more effective. Values for the nodes are determined by a comparison of their respective attributes. The node is partitioned based on the importance of the data, and the result is shown in the leaf node. The entropy of the node is a measure of its significance. Algorithm 2 is a representation of the GDCalgorithm.

Algorithm 2: TheGDCalgorithm

Step 1:A training dataset is chosen to carry out the learning process.

Step 2: Create a diagram connecting each attribute to the appropriate classes.

Step 3: For each attribute, collect all reasonable values that correspond to reasonable classifications.

Step 4: Calculate the values of all the characteristics that correspond to different classes.

Step 5: The property with the least number of values found in distinct classes receives a root node.

Step 6: Choose another characteristic for the decision tree's next level based on the existing attributes that have the fewest values and the most different classes.

Step 7: End

The GDC method is suggested for predicting the health status of individuals. The dGDC uses nodes to reflect the qualities it uses. The speed and simplicity of this sort of algorithm make it accessible to human minds. The algorithmic GDCcan handle an abundance of highlighted data. These values are derived via an examination in which all qualities, both positive and negative, are considered. The root node's value is determined by examining all of the training data. The contrast classes are used to determine this attribute's value.

4. Result and discussion

Using the machine learning technique of the GDC algorithm, this experiment categorizes and predicts the severity of strokes. Four types of stroke severity were identified in this trial, and their effectiveness was confirmed. Here we compared some of the existing methods such as Convolutional Neural Network CNN [16], Bidirectional Long Short Term Memory and Bidirectional Encoder Representations from Transformers Bi-LSTM and BERT [17] and

Nonlinear Autoregressive Neural Network (NANN) [18] with our proposed methods using several metrics like Accuracy, Precision, Recall, and F1-score.

Accuracy is a frequently used parameter in machine learning to assess how well a GDC performs. It counts the number of successfully predicted cases among all the examples in the dataset.Figure 2 and Table 1 provide a comparison of accuracy between conventional and suggested methods, as well as how much more accurate our recommended strategy is than the other traditional approaches.

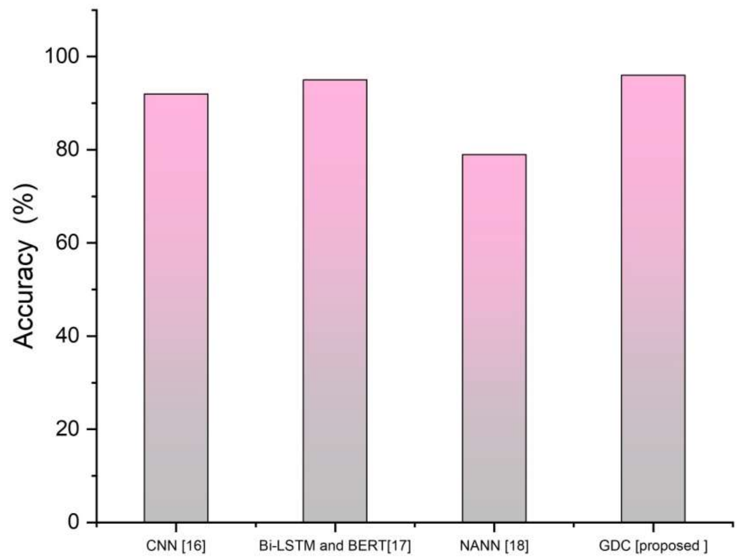

Figure 2: Comparison of the suggested and existing methods' accuracy

Table 1: Comparison of Accuracy

Methods	Accuracy (%)
CNN [16]	92
Bi-LSTM and BERT[17]	95
NANN [18]	79
GDC [proposed]	96

Precision is a performance metric that is widely used in machine learning and information retrieval to assess the accuracy of a GDC's positive predictions. It determines the proportion of correctly anticipated positive outcomes among all positive outcomes that were forecasted. Figure 3 and Table 2 provide a comparison of precision between the conventional and proposed approaches, as well as how much more efficient our suggested strategy is than the other traditional approaches.

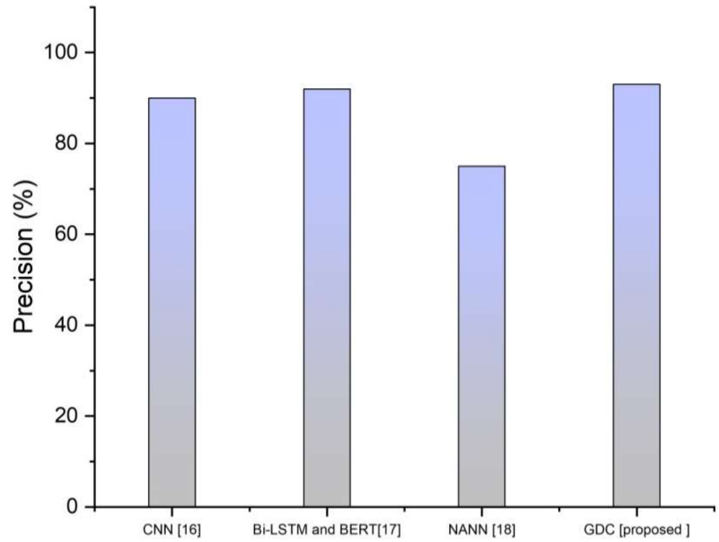

Figure 3: Comparison of the suggested and existing methods' precision

Table 2: Comparison of Precision

Methods	Precision (%)
CNN [16]	90
Bi-LSTM and BERT[17]	92
NANN [18]	75
GDC [proposed]	93

Recall, often referred to as sensitivity or true positive rate, is a performance statistic frequently used in information retrieval and machine learning to evaluate a GDC's capacity to accurately recognize all positive cases. It gauges the percentage of all genuine positive cases that the model accurately classified as true positives.Figure 4 and table 3 depict the comparison of recall between the existing and suggested approach and also it shows that our proposed method is higher than the other current methods

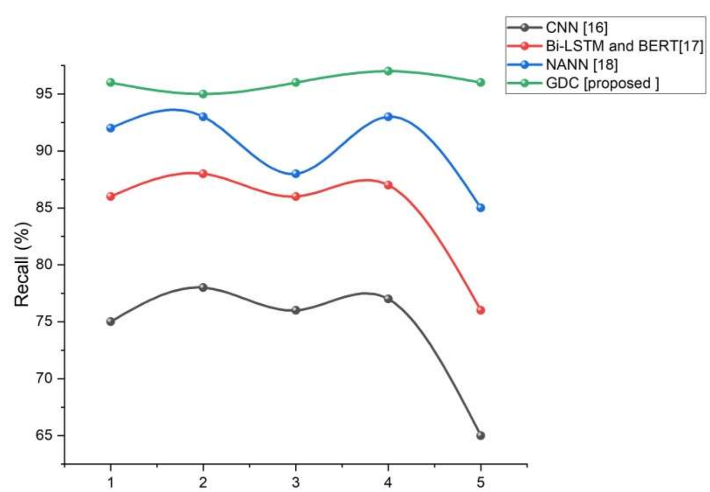

Figure 4: Comparison of the suggested and existing methods' recall

Table 3: Comparison of recall

S.NO	Recall (%)			
	CNN [16]	Bi-LSTM and BERT[17]	NANN [18]	GDC [proposed]
1	75	86	92	96
2	78	88	93	95
3	76	86	88	96
4	77	87	93	97
5	65	76	85	96

In machine learning and information retrieval, the F1 score is a widely used statistic that combines precision and recalls into a single performance measure. By taking into account both the capacity to accurately identify positive cases (recall) and the capacity to prevent false positives (precision), it offers a fair assessment of GDC accuracy. Figure 5 and Table 4 depict the comparison of the F1 score between the existing and suggested approach and also shows that our proposed method is higher than the other current methods

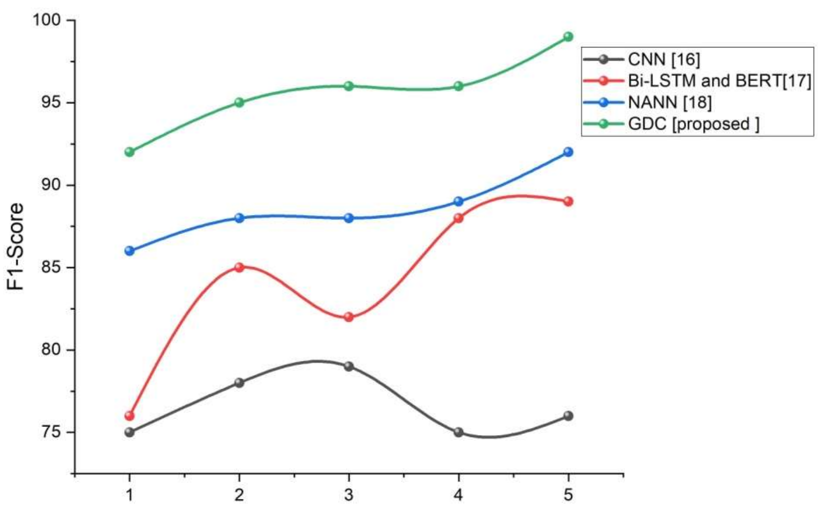

Figure 5: Comparison of the suggested and existing methods' F1-score

Table 4: Comparison of F1-score

S.NO	F1-Score			
	CNN [16]	Bi-LSTM and BERT[17]	NANN [18]	GDC [proposed]
1	75	76	86	92
2	78	85	88	95
3	79	82	88	96
4	75	88	89	96
5	76	89	92	99

5. Conclusion

In this study, we suggest using the GDC algorithm and the NIHSS characteristics to detect and analyze the severity of stroke in older people over 65. The following benefits apply to the proposed stroke severity prediction and in-depth analysis system. First, our technology uses real-time data from the NIHSS to automatically categorize and analyze stroke severity into four classifications. Second, the technology sends real-time alarm information about the severity of a stroke to patients and their relatives, enabling them to seek emergency care and visit a hospital. Third, semantic analysis was done on the more detailed rules that GDC provided. Finally, to deliver quicker and more precise service support, the suggested model uses only 13 of the 18 NIHSS features during real system operation, including age. The patient's NIHSS measurement time can be shortened in a manner that is supported by science, to briefly state the benefits of our approach. Additionally, it can help secure the ideal window of opportunity for a patient's emergency care and deliver exceptionally dependable services.

Reference

1. Ullah, K., Shah, M.A. and Zhang, S., 2016, January. Effective ways to use the Internet of Things in the field of medical and smart health care. In 2016 international conference on intelligent systems engineering (ICISE) (pp. 372-379). IEEE.
2. Gurjar, A. and Sarnaik, N.A., 2018. Heart attack detection by heartbeat sensing using Internet of Things: IoT. Heart, 5(03).
3. Fernandez, F. and Pallis, G.C., 2014, November. Opportunities and challenges of the Internet of Things for healthcare: A systems engineering perspective. In 2014 4th international conference on wireless mobile communication and healthcare-transforming Healthcare through innovations in Mobile and wireless technologies (MOBIHEALTH) (pp. 263-266). IEEE.
4. Mamdiwar, S.D., Shakruwala, Z., Chadha, U., Srinivasan, K. and Chang, C.Y., 2021. Recent advances in IoT-assisted wearable sensor systems for healthcare monitoring. Biosensors, 11(10), p.372.
5. Evjen, T.Å., Hosseini Raviz, S.R., Petersen, S.A. and Krogstie, J., 2020. Smart facility management: Future healthcare organization through indoor positioning systems in the light of enterprise BIM. Smart Cities, 3(3), pp.793-805.
6. Zhuo, Y., Qu, Y., Wu, J., Huang, X., Yuan, W., Lee, J., Yang, Z. and Zee, B., 2021. Estimation of stroke severity with National Institutes of Health Stroke Scale grading and retinal features: A cross-sectional study. Medicine, 100(31).
7. Saber, H. and Saver, J.L., 2020. Distributional validity and prognostic power of the national institutes of health stroke scale in US administrative claims data. JAMA Neurology, 77(5), pp.606-612.
8. Manogaran, G., Varatharajan, R., Lopez, D., Kumar, P.M., Sundarasekar, R. and Thota, C., 2018. A new architecture of the Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. Future Generation Computer Systems, 82, pp.375-387.

9. Greco, L., Percannella, G., Ritrovato, P., Tortorella, F. and Vento, M., 2020. Trends in IoT-based solutions for health care: Moving AI to the edge. Pattern recognition letters, 135, pp.346-353.
10. Jagadeeswari, V., Subramaniyaswamy, V., Logesh, R. and Vijayakumar, V., 2018. A study on medical Internet of Things and Big Data in the personalized healthcare system. Health information science and systems, 6, pp.1-20.
11. Nandy, S., Adhikari, M., Khan, M.A., Menon, V.G. and Verma, S., 2021. An intrusion detection mechanism for secured IoMT framework based on swarm-neural network. IEEE Journal of Biomedical and Health Informatics, 26(5), pp.1969-1976.
12. Tian, S., Yang, W., Le Grange, J.M., Wang, P., Huang, W. and Ye, Z., 2019. Smart healthcare: making medical care more intelligent. Global Health Journal, 3(3), pp.62-65.
13. Farahani, B., Firouzi, F., Chang, V., Badaroglu, M., Constant, N. and Mankodiya, K., 2018. Towards fog-driven IoT eHealth: Promises and challenges of IoT in medicine and healthcare. Future Generation Computer Systems, 78, pp.659-676.
14. Usak, M., Kubiatko, M., Shabbir, M.S., Viktorovna Dudnik, O., Jermsittiparsert, K. and Rajabion, L., 2020. Health care service delivery based on the Internet of things: A systematic and comprehensive study. International Journal of Communication Systems, 33(2), p.e4179.
15. Pal, D., Funilkul, S., Charoenkitkarn, N. and Kanthamanon, P., 2018. Internet-of-things and smart homes for elderly healthcare: An end user perspective. IEEE Access, 6, pp.10483-10496.
16. Opoku Agyeman, M., Felipe Guerrero, A. and Vien, Q.T., 2022. Classification Techniques for Arrhythmia Patterns Using Convolutional Neural Networks and Internet of Things (IoT) Devices. IEEE Access, 10, pp.87387-87403.
17. Zeberga, K., Attique, M., Shah, B., Ali, F., Jembre, Y.Z. and Chung, T.S., 2022. A novel text mining approach for mental health prediction using Bi-LSTM and BERT model. Computational Intelligence and Neuroscience, 2022.
18. An, Q., Szewczyk, P., Johnstone, M.N. and Kang, J.J., 2021, November. Enhancement of Healthcare Data Performance Metrics using Neural Network Machine Learning Algorithms. In 2021 31st International Telecommunication Networks and Applications Conference (ITNAC) (pp. 172-179). IEEE.