Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/11167
DOI URL: http://dx.doi.org/10.21474/IJAR01/11167

RESEARCH ARTICLE

Deep Learning for Health Informatics: A Secure Cellular Automata

Farrukh Arslan
School of Electrical and Computer Engineering, Purdue University, USA.

Manuscript Info

Manuscript History
Received: 10 April 2020
Final Accepted: 12 May 2020
Published: June 2020

Abstract

Health informatics has gained a greater focus as the data analytics role has become vital for the last two decades. Many machine learning-based models have evolved to process the huge data involved in this sector. Deep Learning (DL) augmented with Non-Linear Cellular Automata (NLCA) is becoming a powerful tool with great potential to process big data. This will help to develop a system that facilitates parallelization, rapid data storage, and computational power with improved security parameters. This paper provides a novel and robust mechanism with deep learning augmented with non-linear cellular automata with greater security, adaptability for health informatics. The proposed mechanism is adaptableandcanaddress many open problems in medical informatics, bioinformatics, and medical imaging. The security parameters considered in this model are Confidentiality, authorization, and integrity. This method is evaluated for performance, and it reports an average accuracy of 89.32%. The parameters precision, sensitivity, and specificity are considered to measure to measure the accuracy of the model.

Introduction:-

Cellular automata augmented with deep learning are one of the exiting trends in Machine Learning. The foundations of C.A. with complement and un-complemented rule transitions, together with convolution neural networks (CNN) have a strong mathematical foundation and architecture to address challenges in data evolved through health.Clinical imaging can create highlights that are progressively refined and hard to expound in graphic methods. Verifiable highlights could decide fibroids and polyps [1], and describe abnormalities in tissue morphology, for example, tumors [2]. In translational bioinformatics, such highlights may likewise decide nucleotide successions that could tie a DNA or RNA strand to a protein [3]. A fast flood of enthusiasm for profound learning as of late as far as the number of papers distributed in sub-fields in wellbeing informatics, including bioinformatics, clinical imaging, inescapable detecting, and clinical informatics.

Pradipta Maji [4], [5] has explored the use of C.A. in design grouping with certain esteemed information. A genetic algorithm is used to implement Fuzzy Cellular Automata, which is a special class of C.A. Pradipta Maji et al. has proposed a hypothesis and utilization of C.A. for design arrangement [6]. A genetic algorithm is used to develop fuzzy MACA. The same authors [7] have additionally proposed the mistake rectifying ability of cell automata dependent on cooperative memory. The ideal C.A. is advanced with the definition of a reenacted toughening program, which can be helpful in VLSI innovation**.** We have reviewed various types of CA[8],[9] that can be applied for this technique.

Corresponding Author:- Farrukh Arslan
Address**:-** School of Electrical and Computer Engineering, Purdue University, USA.

Deep Learning is productive when massive data is available for training, and these models have solved many complicated, dynamic real-time problems with higher accuracy with time. CNN is a unique class of neural networks [3] that processes known data, which has grid topology. CNN has many applications, and it operates on a mathematical operator, which is called convolution. It uses many linear operators, represented in matrix form, and then extracts the features of the samples. We propose a distinctive architecture that processes DNA sequence, and operates directly on the characters and uses simple pooling operations & convolutions, which is termed as CNN* augmented with the cellular automata rules to identify these diseases. The main challenge in this research is mapping of the Medical Informatics characteristics to CNN* and proceed to train /test the classifier[18].

Table 1:- Application of the Mechanism HI-DL-CA.

Field	Input	Application
Medical Informatics	Health Records Stored Electronically	Heart Diseases Human Behavior
Bioinformatics	Genomic Data	Promoter Prediction Gene Prediction
Medical Imaging	Clinical Images	Skin Cancer Diabetes

We have referred various mechanisms in the literature that addresses the open problems in medical informatics. After thorough literature, we found medical informatics, bioinformatics, and medical imaging is the most important areas in Health Informatics[16]. In medical informatics, the vital applications are heart diseases and analysis of human behavior from the health records stored electronically. In bioinformatics, the critical applications we identified are promoter prediction, gene prediction from genomic data. In medical imaging, we found skin cancer, diabetes prediction from clinical images are vital problems, as shown in table 1. An extensive literature survey was done on the problems cited above. After this step, we understood that DL with CA[17] could process both images and text to process input related to health informatics.

Design of HI-DL-CA (Health Informatics-Deep Learning-Cellular Automata)

The general architecture of HI-DL-CA is shown in fig 1. The input for the classifier is a set of datasets taken from The Uniform Hospital Discharge Data Set (UHDDS), Data, and Tools of the National Center for Health Statistics. C.A. rules initially process the input as per the application requirement. When HL-DL-CA is trained to treat genomes, the data is processed in the form of three, as the codons are in the multiples of three. The encoding, in general, will be done by a non-linear C.A. method, which was depicted in fig 2. The input is forwarded to CNN(Convolution Network), which was illustrated in fig 3 to predict the output.

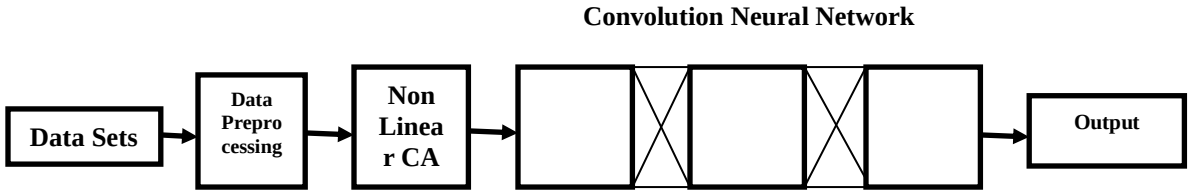


Figure 1:- General Architecture of HI-DL-CA.

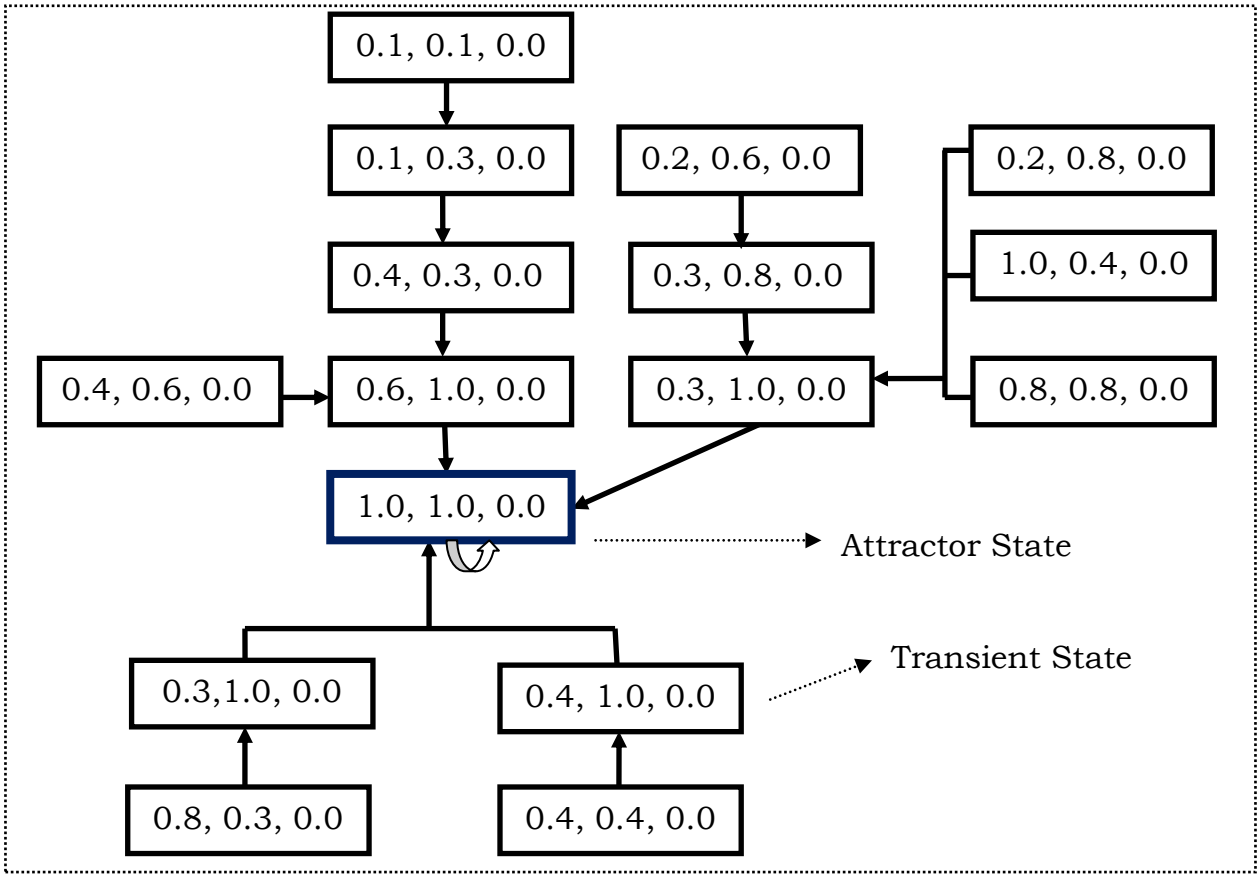


Figure 2:- Working of NL-CA(Non-Linear Cellular Automata).

The working of Non-Linear Cellular Automata (NLCA) is shown in fig 2. NLCA operates with complemented & non-complemented rules. As shown in figure 2 the embedding is done in terms of three. The starting state is 0.1,0.2,0.0, which was applied the rule 256, which states that the transitions from one state to another state depending on its left neighbor and it state resulting in state 0.1,0.3,0.0. The same rule is applied to the first cell resulting in the next state 0.4,0.3,0.0.

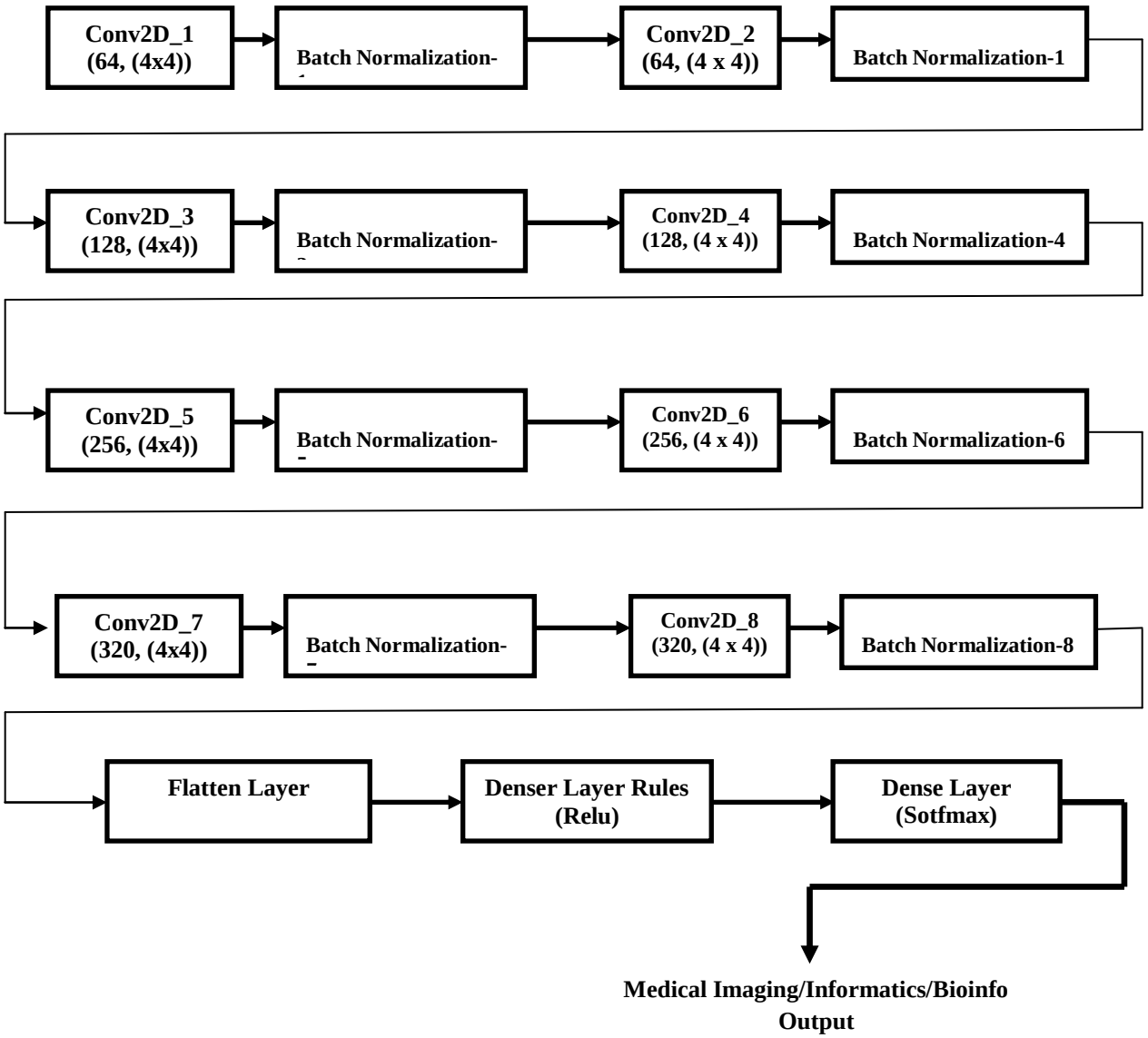


Figure 3:- Working of CNN(Convolution Neural Network).

The transitions will happen until it reaches a state termed as the attractor basin, which has the identified behavior depends on the application. Many rules such as 108,162,252,255,256 etc. can be applied based on the type of input and implementation.

Confidentiality is termed as protection of our information from unauthorized people, which can be guaranteed through proper encoding and encryption mechanisms, which was taken care of in the design. We have used AES (Advanced Encryption System) to achieve Confidentiality. Integrity protects our data from being tailored by the unauthorized & untrusted parties. Repeated Hashing is implemented to provide integrity. Availability is termed as guaranteeing the authorized parties to access the system and information when required. This is addressed by building a robust architecture that can resist DDoS attacks.

The working of CNN is shown in fig 3. Initial convolution layers will process general characteristics, and when the iterations happen deeper go, they will treat more complex features very easity.CA strengthens the filters we used during training and testing—batch normalization aim at improving the stability, speed, the performance of CNN. Activation functions augmented with C.A. rules are used to induce non-linearity into the system, and these are located in dense layers. This is mainly used to standardize the batch of inputs and reduces the number of epochs for

training. The minimum, mean and maximum values of the above parameters are extracted from the dataset and processed them for the prediction. Each convolution uses 4 X 4 kernel, followed by 3X3, followed by 2 X2. After processing, the datasets collected are classified as per the application requirement.With the above discussion, we are confident HI-DL-CA provides a secure mechanism for health informatics. The implementation of the proposed mechanism is discussed in the next section.

Implementation and Comparison of HI-DL-CA

The input for the classifier is a set of datasets taken from The Uniform Hospital Discharge Data Set (UHDDS), Data, and Tools of the National Center for Health Statistics, as discussed in the earlier sections. The genomic data, clinical images, digitalized health records of patients are extracted and processed to verify the validity of our developed system. The one advantage of HL-DL-CA is it is trained to process text ie. genomic input in terms of DNA sequence or Amino Acid sequence, and also, the second version can process images like health records, X rays, etc.

HI-DL-CA for Medical Informatics

As discussed earlier, we have identified two potential problems in medical informatics, i.e.,heart diseases and analysis of human behavior from the health records stored electronically. We have applied our developed classifier on these two problems identified. The classifier has processed the images of people that are suffering from the heart attack and health records stored electronically. For evaluating the developed classifier accuracy, sensitivity, specificity, and precision are considered.

1. Let Tp(True Positives) represent the number of sequences correctly identified related to a heart attack.
2. Let Fn(False Negatives) represent the number of sequences incorrectly identified related to a heart attack.
3. Let Tn(True Negatives) represent the number of non-courses correctly identified not related to a heart attack.
4. Let Fp(False Positives) represent the number of non-sequences incorrectly identified related to a heart attack.

For explaining the metrics, we have considered heart attack prediction.

1. Sensitivity is the ratio of sequences that are correctly predicted

$$Se=\frac{Tp}{Tp+Fn}$$

2. Specificity is the ratio of non-sequences that are correctly predicted.

$$Sp=\frac{Tn}{Tn+Fp}$$

3. Precision (Pr) gives the ratio of correctly identified test samples with total tested samples.

$$Pr=\frac{Tp+Tn}{TotalTestingSamples}$$

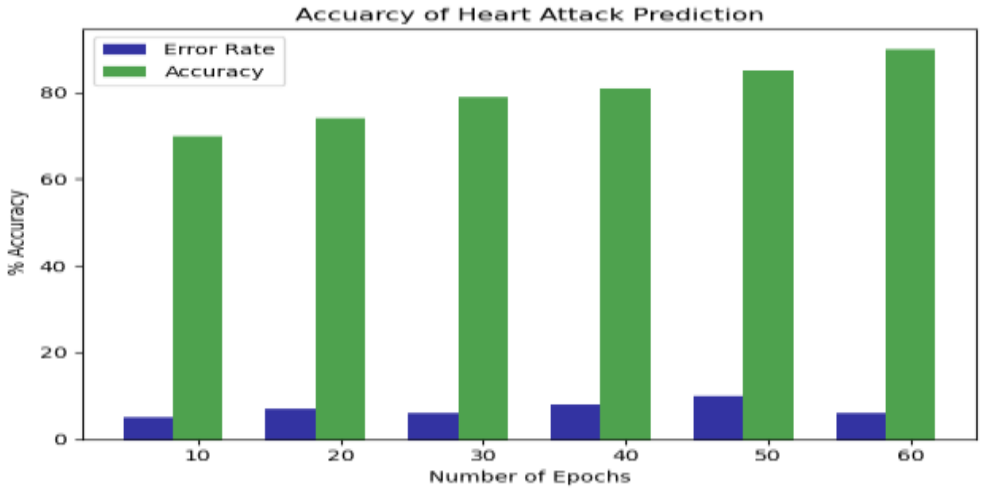


Figure 4:- Accuracy of Heart Attack Prediction with HI-DL-CA.

The model accuracy prediction and error of heart attack is illustrated in fig 4. The accuracy of the model tends to increase with the number of epochs. After reaching 60 epochs, our proposed classifier reports the highest accuracy of 89.95% with an error rate of less than 6%. The accuracy of the model to predict the nature of employees from health records also tends to increase with the number of epochs. After reaching 60 epochs, our proposed classifier reports the highest accuracy of 84.95% with an error rate of less than 12.3%. The performance of our classifier to predict heart attack is compared with the existing literature, which was reported in fig 5. We have identified four best mechanisms Significant Patterns(S.P.)[6],

Association Rule Mining(ARM)[7], Big Data Analytics(BDA)[8], and Fuzzy C Means(FCM)[9] to compare the performance. We found FCM report an accuracy of 83.6, which is better among the existing literature, and HI-DL-CA indicates an accuracy of 89.69%. The specificity, sensitivity, and precision of our approach to standard approaches were reported in table 2.

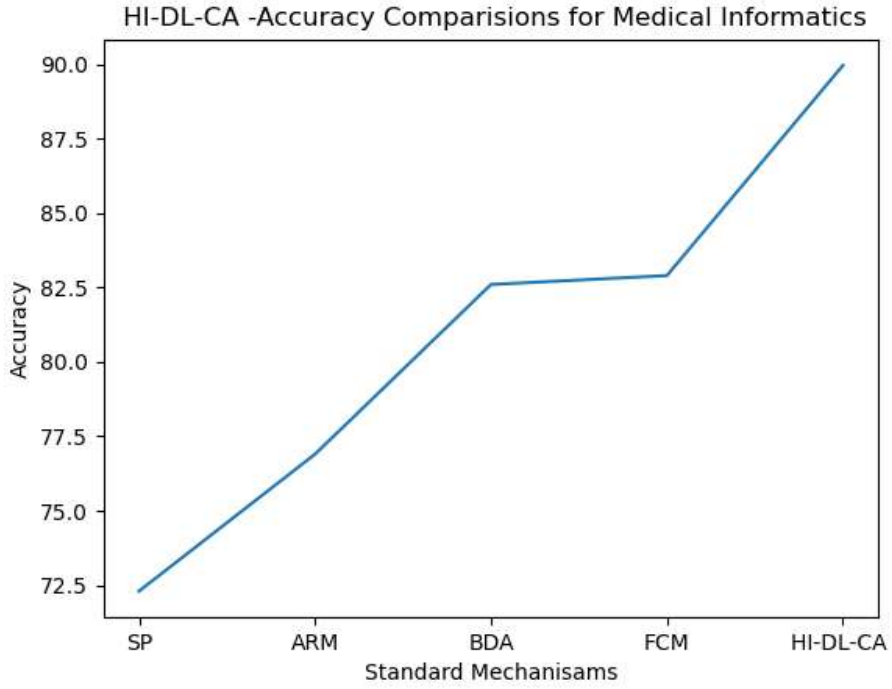


Figure 5:- Accuracy Comparisionof HI-DL-CA with Standard Mechanisms.

HI-DL-CA for Bioinformatics

In continuation of the earlier discussion, HL-DL-CA was trained and tested to process genomic data to address problems in bioinformatics. For example, when protein-coding regions are to identified th input is a DNA sequence, when the protein structure is to be identified, the input is an Amino Acid sequence and so on. The architecture of HL-DL-CA is so versatile and robust to process any information for accurate prediction.

The model accuracy prediction and error promoter prediction is illustrated in fig 6. The accuracy of the model tends to increase with the number of epochs. After reaching 60 epochs, our proposed classifier reports the highest accuracy of 92.36% with an error rate of less than 7.2%. The accuracy of the model to predict genes also tends to increase with the number of epochs. After reaching 60 epochs, our proposed classifier reports the highest accuracy of 89.27 with an error rate of less than 14.6%.

The performance of our classifier to predict promoter is compared with the existing literature, which was reported in fig 7. We have identified three best mechanisms Neural Networks(N.N.)[10], DNA energies(DNAE)[11], and Support Vector Machin(SVM)[12] to compare the performance. We found SVM report an accuracy of 87.89, which is better among the existing literature, and HI-DL-CA indicates an accuracy of 92.36%. The specificity, sensitivity, and precision of our approach to standard approaches were reported in table 3.

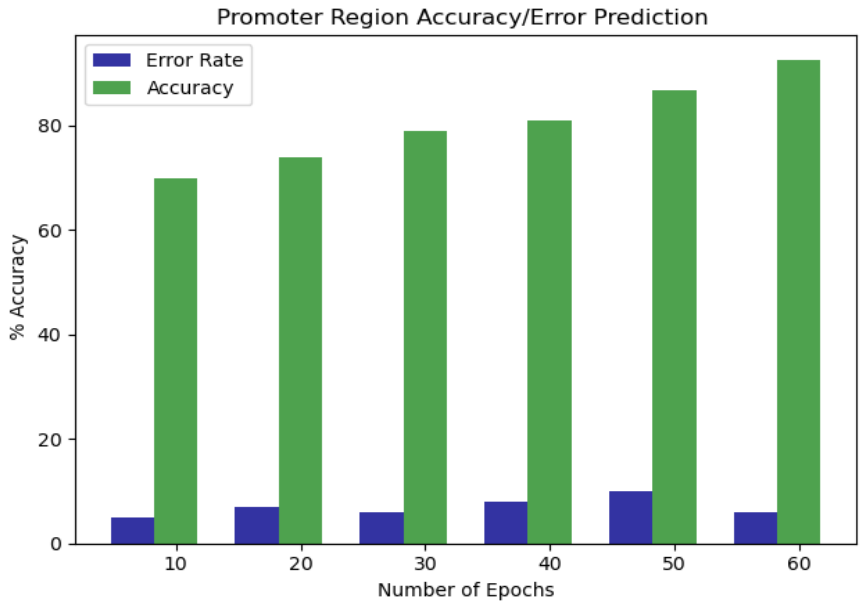


Figure 6:- Accuracy of Promoter Region Prediction with HI-DL-CA.

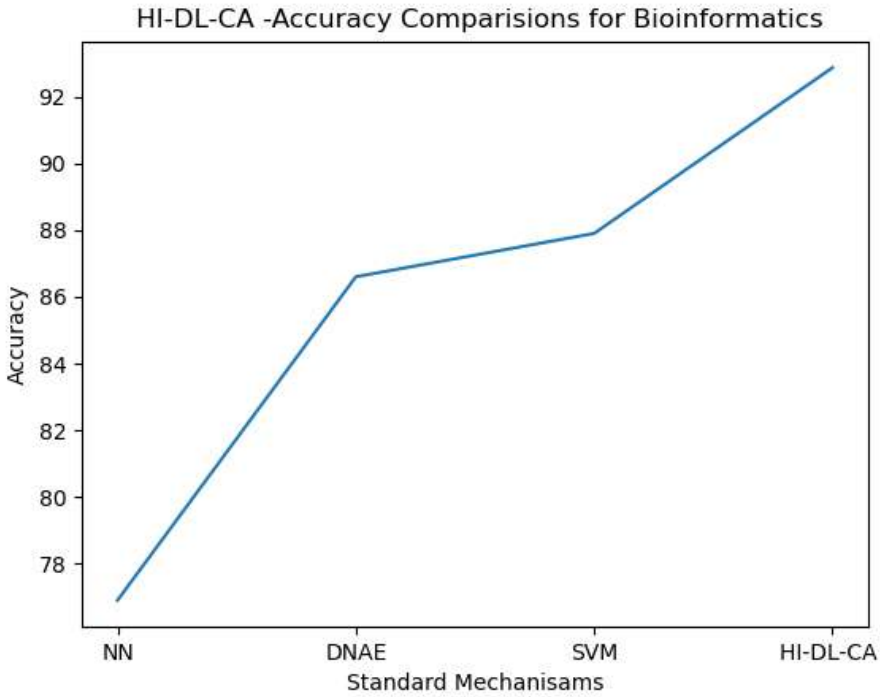


Figure 7:- Accuracy Comparisionof HI-DL-CA with Standard Mechanisms.

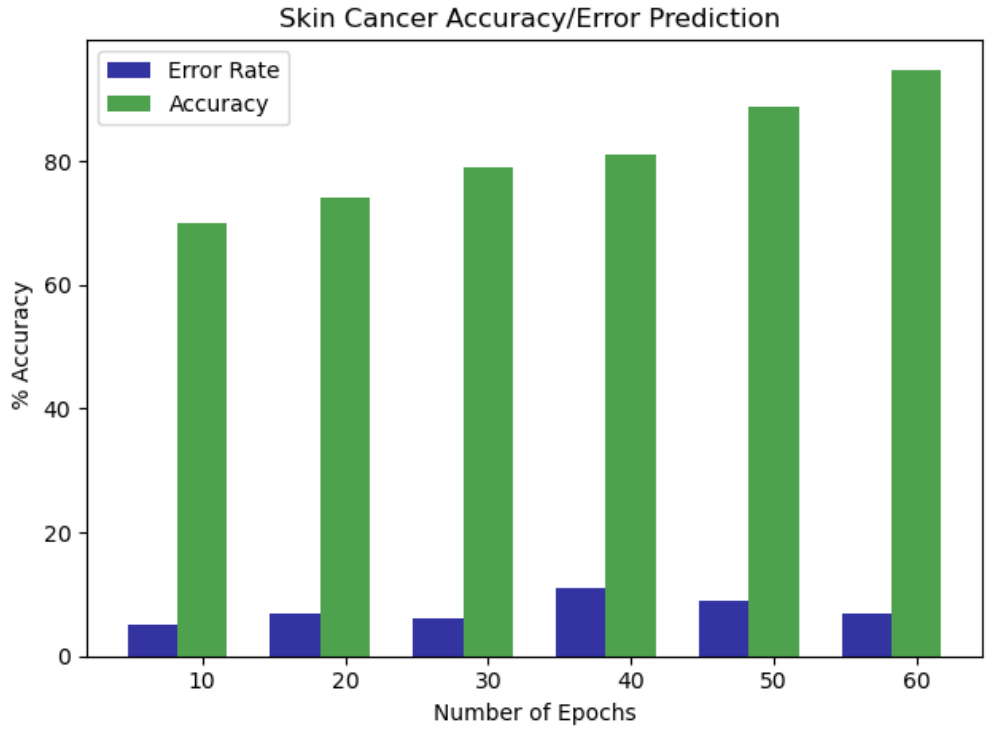
HI-DL-CA for Medical Imaging

As discussed in secion 3.1, HL-DL-CA was trained and tested to process clinical images to address problems in medical imaging. We hav identified skin cancer and diabetes as potential prolemsin medical imaging.Thearchitectur of HL-DL-CA is so versatile and robust to process any number of images for accurate prediction.

The model accuracy prediction and error skin cancer prediction is illustrated in fig 8. The accuracy of the model tends to increase with the number of epochs. After reaching 60 epochs, our proposed classifier reports the highest

Mechanisms	Sensitivity	Specificity	Precision
Significant Patterns(SP)	0.85	0.81	0.86
Association Rule Mining(ARM)	0.82	0.84	0.84
Big Data Analytics(BDA),	0.86	0.86	0.87
Significant Patterns(SP	0.89	0.85	0.90
Fuzzy C Means(FCM)	0.85	0.901	0.91
HL-DL-CA	0.92	0.912	0.93

accuracy of 94.78% with an error rate of less than 5.2%. The accuracy of the model to predict diabates also tends to increase with the number of epochs. After reaching 60 epochs, our proposed classifier reports the highest accuracy of 96.7 with an error rate of less than 10.6%.



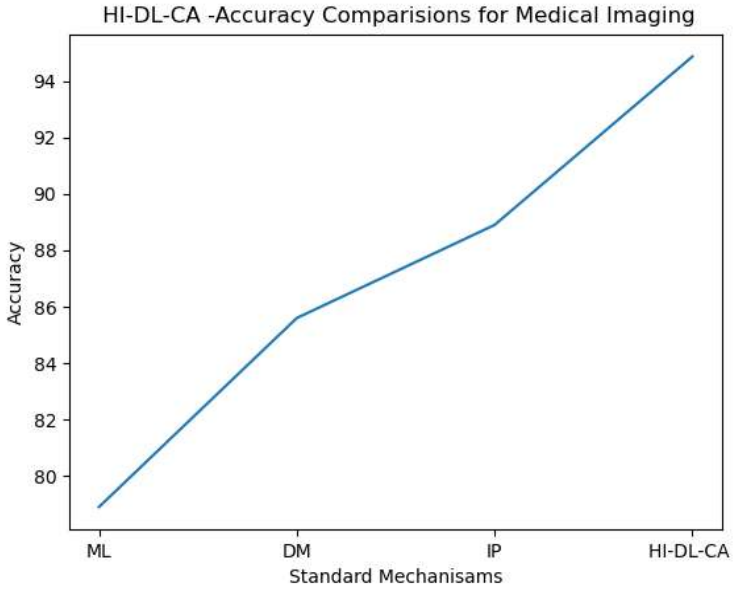


Figure 8: Accuracy of Skin Cancer Prediction with HI-DL-CA

Mechanisms	Sensitivity	Specificity	Precision
Neural Networks(NN)	0.86	0.88	0.89
DNA Energey(DNE)	0.87	0.87	0.90
Support Vector Machine(SVM)	0.91	0.90	0.92
HL-DL-CA	0.94	0.92	0.94

Figure 9:- Accuracy Comparisionof HI-DL-CA with Standard Mechanisms.

Table 2: Performance comparison of HI-DL-CA in Medical Informatics.

Mechanisms	Sensitivity	Specificity	Precision
Machine Learning(ML)	0.89	0.89	0.88
Data Mining(DM)	0.88	0.85	0.90
ImagProcessing(IP)	0.94	0.92	0.91
HL-DL-CA	0.97	0.96	0.93

Table 3:- Performance comparison of HI-DL-CA in Bioinformatics.

Conclusion:-

We have successfully developed a robust, secure, and adaptable mechanism that provides high accuracy, specificity, precision, sensitivity, availability, integrity, and Confidentiality for majority applications of health informatics. HI-DL-CA reports an average accuracy of 89.95% while addressing the problems in Medical Informatics. The proposed classifier indicates an average accuracy of 92.36%,94.78%, while solving the problems in Bioinformatics and Medical Imaging, respectively. The analysis of X-rays pertaining to the patients affected by COVID-19 can be done by using this framework that can predict the death rate variations. This framework can be improved by considering more robust and secure parameters that can attract more people to use these systems.

References:-

1. Saha, J., Chowdhury, C. and Biswas, S., 2020. Review of Machine Learning and Deep Learning Based Recommender Systems for Health Informatics. In Deep Learning Techniques for Biomedical and Health Informatics (pp. 101-126). Springer, Cham.

2. Dash, S., Acharya, B.R., Mittal, M., Abraham, A. and Kelemen, A., 2020. Deep Learning Techniques for Biomedical and Health Informatics. Springer.
3. Mulani, J., Heda, S., Tumdi, K., Patel, J., Chhinkaniwala, H. and Patel, J., 2020. Deep Reinforcement Learning Based Personalized Health Recommendations. In Deep Learning Techniques for Biomedical and Health Informatics (pp. 231-255). Springer, Cham.
4. Mittal, S. and Hasija, Y., 2020. Applications of Deep Learning in Healthcare and Biomedicine. In Deep Learning Techniques for Biomedical and Health Informatics (pp. 57-77). Springer, Cham.
5. Alheejawi, S., Mandal, M., Xu, H., Lu, C., Berendt, R. and Jha, N., 2020. Deep learning-based histopathological image analysis for automated detection and staging of melanoma. In Deep Learning Techniques for Biomedical and Health Informatics (pp. 237-265). Academic Press.
6. Patil, S.B. and Kumaraswamy, Y.S., 2009. Extraction of significant patterns from heart disease warehouses for heart attack prediction. IJCSNS, 9(2), pp.228-235.
7. Jabbar, M.A., Chandra, P. and Deekshatulu, B.L., 2011. Cluster based association rule mining for heart attack prediction. Journal of Theoretical and Applied Information Technology, 32(2), pp.196-201.
8. Alexander, C.A. and Wang, L., 2017. Big data analytics in heart attack prediction. J Nurs Care, 6(393), pp.2167-1168.
9. Chitra, R. and Seenivasagam, V., 2013. Heart attack prediction system using Fuzzy C Means classifier. IOSR Journal of Computer Engineering, 14, pp.23-31.
10. Demeler, B. and Zhou, G., 1991. Neural network optimization for E. coli promoter prediction. Nucleic acids research, 19(7), pp.1593-1599.
11. Mishra, A., Dhanda, S., Siwach, P., Aggarwal, S. and Jayaram, B., 2020. A novel method SEProm for prokaryotic promoter prediction based on DNA structure and energetics. Bioinformatics.
12. Arslan, H., 2019, April. A New Promoter Prediction Method using Support Vector Machines. In 2019 27th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.
13. Putra, T.A., Rufaida, S.I. and Leu, J.S., 2020. Enhanced Skin Condition Prediction Through Machine Learning Using Dynamic Training and Testing Augmentation. IEEE Access, 8, pp.40536-40546.
14. Verma, A.K., Pal, S. and Kumar, S., 2020. Prediction of skin disease using ensemble data mining techniques and feature selection method—a comparative study. Applied biochemistry and biotechnology, 190(2), pp.341-359.
15. Tamošiūnas, M., Plorina, E.V., Lange, M., Derjabo, A., Kuzmina, I., Bļizņuks, D. and Spigulis, J., 2020. Autofluorescence imaging for recurrence detection in skin cancer post- operative scars. Journal of Biophotonics, p.e201900162.
16. Sree, P.K., 2020. Deep Learning Supported Food Security in Developing Countries. International Journal of Recent Development in Computer Technology & Software Applications [ISSN: 2581-6276 (online)], 4(1).
17. Pokkuluri, K.S. and Nedunuri, S.U.D., 2020. A Novel Cellular Automata Classifier for COVID-19 Prediction. Journal of Health Sciences, 10(1), pp.34-38.
18. Sree, P.K., Babu, I.R. and Devi, N.U., 2009. Investigating an Artificial Immune System to strengthen protein structure prediction and protein coding region identification using the Cellular Automata classifier. International Journal of Bioinformatics Research and Applications, 5(6), pp.647-662.