A Secure Cellular Automata Integrated Deep Learning Mechanism for Health Informatics

Kiran Sree Pokkuluri[1] and SSSN Usha Devi Nedunuri[2]

[1]Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women (A), India

[2]Department of Computer Science and Engineering, University College of Engineering-JNTU, India

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 adaptable and can address 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.*

1. 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, tumours [5]. In translational bioinformatics, such highlights may likewise decide nucleotide successions that could tie a Deoxyribo Nucleic Acid (DNA) or Ribonucleic Acid (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.

Mittal and Hasija [11], Alheejawi *et al*. [2] 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. Patil and Kumaraswamy [12] has proposed a hypothesis and utilization of C.A. for design arrangement. A genetic algorithm is used to develop fuzzy Multiple Attractor Cellular Automata (MACA). The same Jabbar *et al*. [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 re-enacted toughening program, which can be helpful in Very Large Scale Integration (VLSI) innovation. We have reviewed various types of CA [1, 4] that can be applied for this technique. Mishra *et al*. [10] has produced a standard promoter prediction system with 82.3% accuracy.

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 and 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 [16], Sree and Babu [17].

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 imaging, we found skin cancer, diabetes prediction from clinical images are vital problems, as shown in Table 1.

Table 1. Application of the mechanism HI-DL-CA.

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

An extensive literature survey was done on the problems cited above. After this step, we understood that DL with CA [19] could process both images and text to process input related to health informatics. After introduction section 2 provides the in-depth analysis of the design of HL-DL-CA, section 3 depicts various experimental results of the proposed classifier with comparisons with the baseline methods and section 4 provides the conclusion or our work.

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

The general architecture of Health Informatics-Deep Learning-Cellular Automata (HI-DL-CA) is shown in Figure 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 Centre 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 Figure 2. The input is forwarded to CNN (Convolution Network), which was illustrated in Figure 3 to predict the output.

The working of Non-Linear Cellular Automata (NLCA) is shown in Figure 2. NLCA operates with complemented and 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 neighbour 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.

The transitions will happen until it reaches a state termed as the attractor basin, which has the identified behaviour 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 Advanced Encryption System (AES) to achieve Confidentiality. Integrity protects our data from being tailored by the unauthorized and 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 Distributed Denial of Service (DDoS) attacks.

The working of CNN is shown in Figure 3. Initial convolution layers will process general characteristics, and when the iterations happen deeper go, they will treat more complex features very easity. C 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 4X4 kernel, followed by 3X3, followed by 2X2. 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.
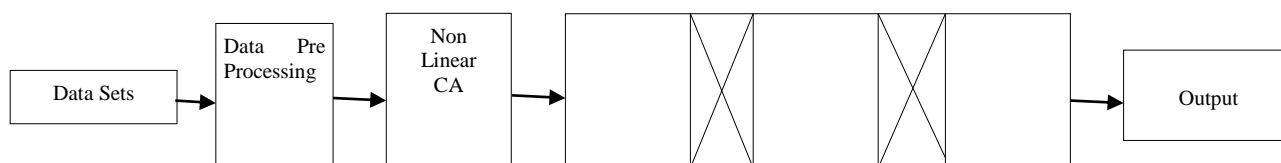
Convolution Neural Network

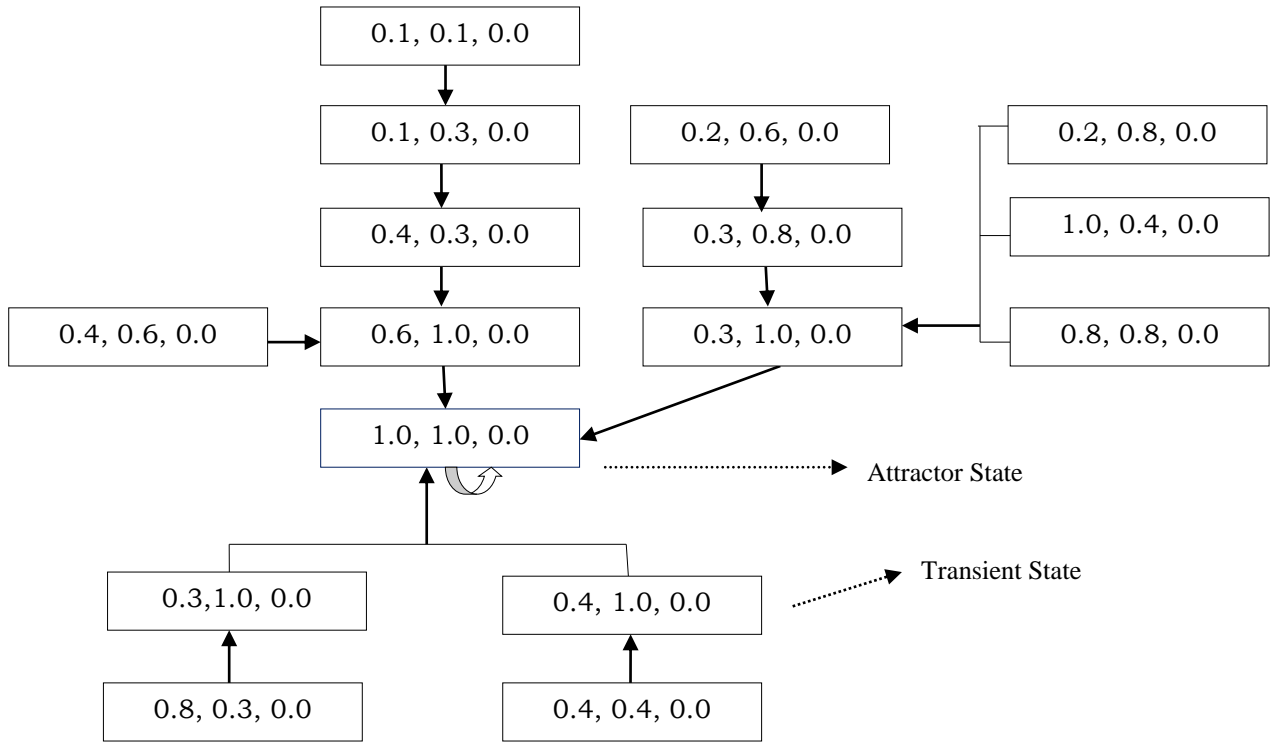Figure 1. General Architecture of HI-DL-CA.

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

3. 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 i.e., 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.

3.1. 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 behaviour 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.

```
┌─────────────────┐    ┌─────────────────────┐    ┌─────────────────┐    ┌─────────────────────┐
│ Conv2D_1        │───▶│ Batch Normalization-1│───▶│ Conv2D_2        │───▶│ Batch Normalization-1│
│ (64, (4x4))     │    │                     │    │ (64, (4 x 4))   │    │                     │
└─────────────────┘    └─────────────────────┘    └─────────────────┘    └─────────────────────┘

┌─────────────────┐    ┌─────────────────────┐    ┌─────────────────┐    ┌─────────────────────┐
│ Conv2D_3        │───▶│ Batch Normalization-3│───▶│ Conv2D_4        │───▶│ Batch Normalization-4│
│ (128, (4x4))    │    │                     │    │ (128, (4 x 4))  │    │                     │
└─────────────────┘    └─────────────────────┘    └─────────────────┘    └─────────────────────┘

┌─────────────────┐    ┌─────────────────────┐    ┌─────────────────┐    ┌─────────────────────┐
│ Conv2D_5        │───▶│ Batch Normalization-5│───▶│ Conv2D_6        │───▶│ Batch Normalization-6│
│ (256, (4x4))    │    │                     │    │ (256, (4 x 4))  │    │                     │
└─────────────────┘    └─────────────────────┘    └─────────────────┘    └─────────────────────┘

┌─────────────────┐    ┌─────────────────────┐    ┌─────────────────┐    ┌─────────────────────┐
│ Conv2D_7        │───▶│ Batch Normalization-7│───▶│ Conv2D_8        │───▶│ Batch Normalization-8│
│ (320, (4x4))    │    │                     │    │ (320, (4 x 4))  │    │                     │
└─────────────────┘    └─────────────────────┘    └─────────────────┘    └─────────────────────┘

┌─────────────────┐    ┌─────────────────────┐    ┌─────────────────┐
│ Flatten Layer   │───▶│ Denser Layer Rules  │───▶│ Dense Layer     │
│                 │    │ (Relu)              │    │ (Sotfmax)       │
└─────────────────┘    └─────────────────────┘    └─────────────────┘

                    Medical Imaging/Informatics
                    /Bio info
                    Output
```
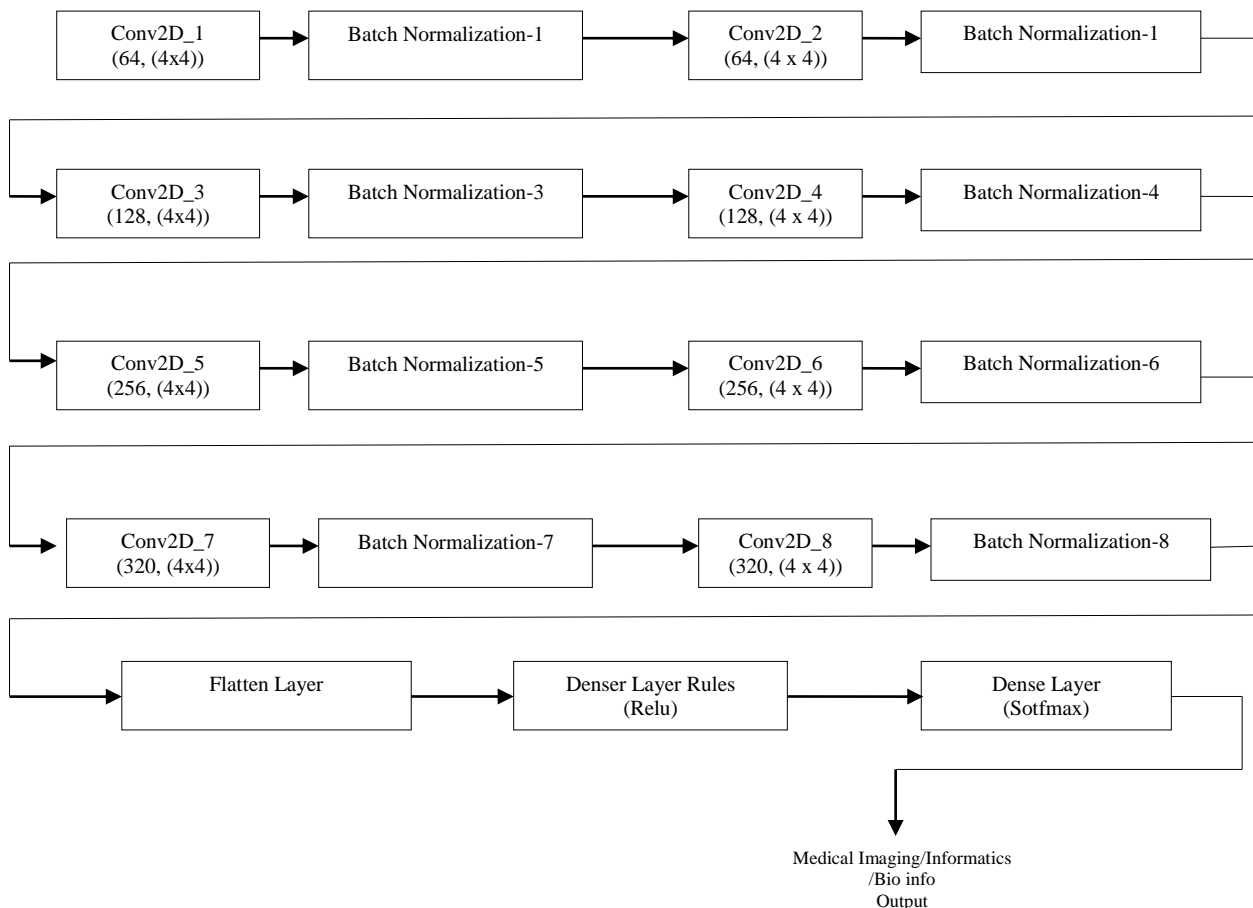
Figure 3. Working of CNN (Convolution Neural Network).

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

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

$$Se = \frac{Tp}{Tp + Fn} \qquad (1)$$

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

$$Sp = \frac{Tn}{Tn + Fp} \qquad (2)$$

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

$$Pr = \frac{Tp + Tn}{Total Testing Samples} \qquad (3)$$

The model accuracy prediction and error of heart attack is illustrated in Figure 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 Figure 5. We have identified four best mechanisms Significant Patterns (SP) [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.
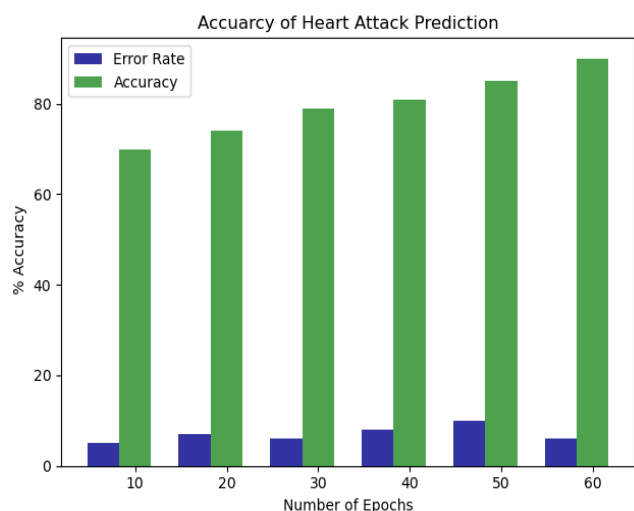
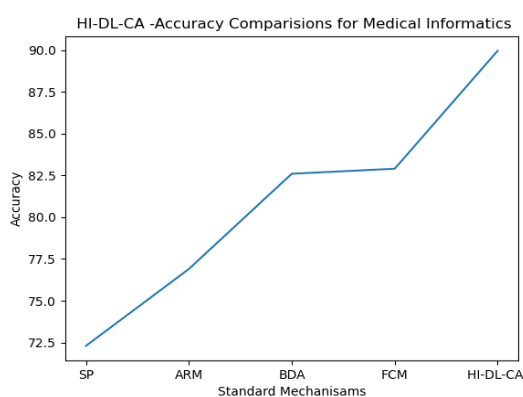Figure 4. Accuracy of heart attack prediction with HI-DL-CA.

Figure 5. Accuracy comparison of HI-DL-CA with standard mechanisms.

Table 2. Performance comparison of HI-DL-CA in medical informatics.

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

3.2. 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 identify the 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 Figure 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%.
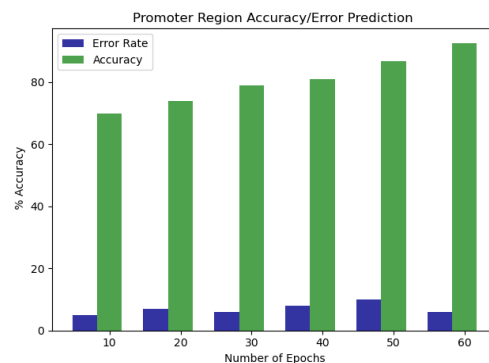

Figure 6. Accuracy of promoter region prediction with HI-DL-CA.

The performance of our classifier to predict promoter is compared with the existing literature, which was reported in Figure 7. We have identified three best mechanisms Neural Networks (NN) [6], DNA Energies (DNAE) [11], and Support Vector Machine (SVM) [9] 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.

Table 3. Performance comparison of HI-DL-CA in bioinformatics.

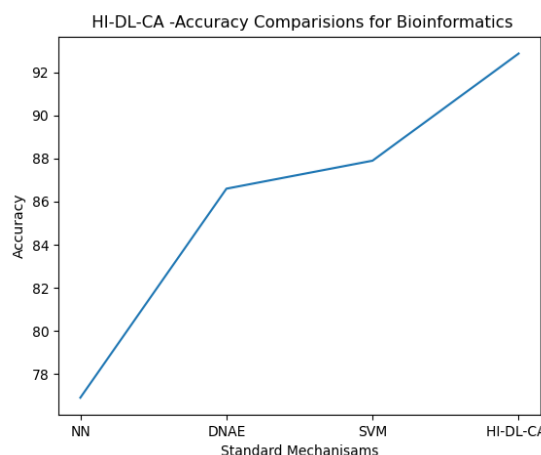| Mechanisms	Sensitivity	Specificity	Precision
Machine Learning (ML)	0.89	0.89	0.88
Data Mining (DM)	0.88	0.85	0.90
Image Processing (IP)	0.94	0.92	0.91
HL-DL-CA	0.97	0.96	0.93

Figure 7. Accuracy Comparison of HI-DL-CA with Standard Mechanisms.

3.3. HI-DL-CA for Medical Imaging

As discussed in section 3.1, HL-DL-CA was trained and tested to process clinical images to address problems in medical imaging. We have identified skin cancer and diabetes as potential problems in medical imaging. The architecture 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 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%.

The performance of our classifier to predict promoter is compared with the existing literature, which was reported in Figure 9. We have identified three best mechanisms Machine Learning (ML) [13], Data Mining (DM) [14], and Image Processing (IP) [18] to compare the performance. We found IP report an accuracy of 89.99, which is better among the existing literature, and HI-DL-CA indicates an accuracy of 94.78%. The specificity, sensitivity, and precision of our approach to standard approaches were reported in Table 4.

After the set of experiments conducted, we strongly conclude HI-DL-CA architecture offers a common mechanism to address problems in health informatics, providing high accuracy, availability, integrity, and Confidentiality.
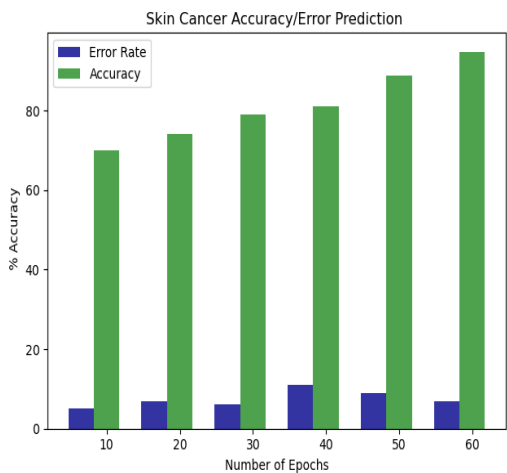

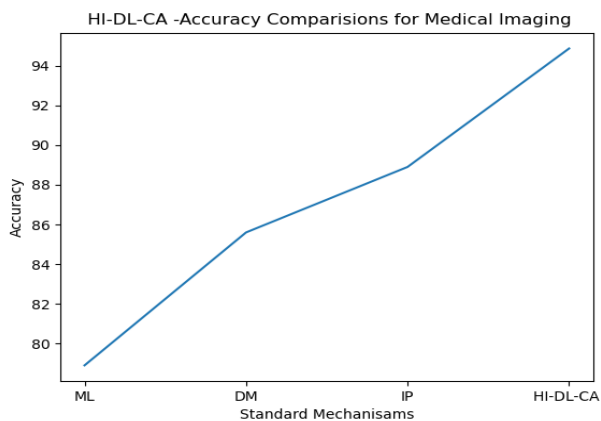Figure 8. Accuracy of skin cancer prediction with HI-DL-CA.

Figure 9. Accuracy Comparisons of HI-DL-CA with standard mechanisms.

Table 4. Performance comparison of HI-DL-CA in medical imaging.

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

4. Conclusions

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. The future extensions of the paper can be augmenting a ternary neural network for with high dimensional Cellular Automata rules which will act as potential classifier with rough fuzzy parameters.

References

[1] Alexander C. and Wang L., "Big Data Analytics in Heart Attack Prediction," *Journal of Nursing and Care*, vol. 6, no. 393, pp. 2167-1168, 2017.

[2] Alheejawi S., Mandal M., Xu H., Lu C., Berendt R., and Jha N., "Deep Learning-Based Histopathological Image Analysis for Automated Detection and Staging of Melanoma," *Deep Learning Techniques for Biomedical and Health Informatics*, pp. 237-265, 2020.

[3] Arslan H., "A New Promoter Prediction Method Using Support Vector Machines," *in Proceedings 27th Signal Processing and Communications Applications Conference*, Sivas, pp. 1-4, 2019.

[4] Chitra R. and Seenivasagam V., "Heart Attack Prediction System using Fuzzy C Means Classifier," *IOSR Journal of Computer Engineering*, vol. 14, no. 2, pp. 23-31, 2013.

[5] Dash S., Acharya B., Mittal M., Abraham A., and Kelemen A., *Deep Learning Techniques for Biomedical and Health Informatics*, Springer, 2020.

[6] Demeler B. and Zhou G., "Neural Network Optimization for E. Coli Promoter Prediction,"

Nucleic Acids Research, vol. 19, no. 7, pp. 1593-1599, 1991.

[7] Jabbar M., Chandra P., and Deekshatulu B., "Cluster Based Association Rule Mining for Heart Attack Prediction," *Journal of Theoretical and Applied Information Technology*, vol. 32, no. 2, pp.196-201, 2011.

[8] Javaid A., Niyaz Q., Sun W., and Alam M., "A "Deep Learning Approach for Network Intrusion Detection System," *Eai Endorsed Transactions on Security and Safety*, vol. 3, no. 9, 2016.

[9] Mulani J., Heda S., Tumdi K., Patel J., Chhinkaniwala H., and Patel J., *Deep Learning Techniques for Biomedical and Health Informatics*, Springer, 2020.

[10] Mishra A., Dhanda S., Siwach P., Aggarwal S., and Jayaram B., "A Novel Method Seprom for Prokaryotic Promoter Prediction Based on DNA Structure and Energetics," *Bioinformatics*, vol. 36, no. 8, pp. 2375-2384, 2020.

[11] Mittal S. and Hasija Y., *Deep Learning Techniques for Biomedical and Health Informatics*, Springer, 2020.

[12] Patil S. and Kumaraswamy Y., "Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction," *International Journal of Computer Science and Network Security*, vol. 9, no. 2, pp. 228-235, 2009.

[13] Putra T., Rufaida S., and Leu J., "Enhanced Skin Condition Prediction Through Machine Learning Using Dynamic Training and Testing Augmentation," *IEEE Access*, vol. 8, pp. 40536-40546, 2020.

[14] Pokkuluri K. and Nedunuri S., "A Novel Cellular Automata Classifier for COVID-19 Prediction," *Journal of Health Sciences*, vol. 10, no. 1, pp. 34-38, 2020.

[15] Sree P. and Nedunuri S., "Deep Learning Supported Food Security in Developing Countries," *International Journal of Recent Development in Computer Technology and Software Applications*, vol. 4, no. 1, pp. 2581-6276, 2020.

[16] Sree P., Babu I., and Devi N., "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*, vol. 5, no. 6, pp. 647-662, 2009.

[17] Sree K. and Babu R., "Identification of Promoter Region in Genomic DNA Using Cellular Automata Based Text Clustering," *The International Arab Journal of Information Technology*, vol. 7, no. 1, pp. 75-78, 2010.

[18] Tamošiūnas M., Plorina E., Lange M., Derjabo, A., Kuzmina I., Bļizņuks D., and Spigulis J., "Autofluorescence Imaging for Recurrence Detection in Skin Cancer Post-Operative Scars," *Journal of Biophotonics*, vol. 13, no. 3, pp. e201900162, 2020.

[19] Verma A., Pal S., and Kumar S., "Prediction of Skin Disease Using Ensemble Data Mining Techniques and Feature Selection Method-A Comparative Study," *Applied Biochemistry and Biotechnology*, vol. 190, no. 2, pp.341-359, 2020.

Kiran Sree Pokkuluri has received his B.Tech and M.E in Computer Science and Engineering from JNTU and Anna University, respectively. He has obtained his Ph.D. degree in the area of Artificial Intelligence from JNTU-Hyderabad. He has authored Six textbooks for UG and PG students of engineering in AI and published more than 96 research articles in various international journals and conferences. He has filed and published six patents in the area of Deep Learning. His biography was listed in Marquis Who's Who in the World, 29th Edition (2012), USA. Prof Kiran is the Recipient of Bharat Excellence Award from Dr. G.V. Krishna Murthy, Former Election Commissioner of India for two times and recipient of RashtryaRatan Award. He was the BOS member of CSE&IT in some universities and autonomous colleges. He also worked as Principal of the N.B.K.R. Institute of Science & Technology (Second Oldest Private Engg College), Vidyanagar, for two years. He has got 18+ years of teaching experience and working as Head &Professor in the department of CSE at Shri Vishnu Engineering College for Women(A), Bhimavaram. He has delivered many technical talks on Deep Learning and AI in various International Conferences, FDP'S, Webinars. His research interests include Deep Learning, Big Data Analytics, Bioinformatics, and Cloud Computing. He is associated with various journals& conferences in various capacities as Editor in Chief, Editorial Member, and Reviewer. His the Global Vice President of WSA: World Statistical Data Analysis Research Association.

SSSN Usha Devi Nedunuri has received her B.Tech degree from JNTU Hyderabad and M.Tech from JNTU Kakinada. She is pursuing her Ph.D from National Institute of Technology, Trichy in the area of Deep Learning. She has published 52 papers in various journals and conferences. She has filed a patent on Deep Learning integrated with IOT. She has acted as resource person for many AICTE sponsored FDP'S and Conferences.