

# Disease Prediction Using Novel Deep Learning Mechanisms

Sanjaya Kumar Sarangi<sup>1\*</sup>, Pallamravi<sup>2</sup>, Pallamravi<sup>2</sup>, Nilima Rani Das<sup>3</sup>, N. Bindu Madhavi<sup>4</sup>, Naveen P<sup>5</sup>, ATA. Kishore Kumar<sup>6</sup>

<sup>1</sup>Fellow and Academic Coordinator, Dept. of Computer Science, Utkal University, Bhubaneswar, INDIA  
[sanjaya.res.cs@utkaluniversity.ac.in](mailto:sanjaya.res.cs@utkaluniversity.ac.in) Corresponding Author

<sup>2</sup>Assistant professor, Department of CSE, Anurag University, Hyderabad, T.S, India [pallamravice@cvsr.ac.in](mailto:pallamravice@cvsr.ac.in)

<sup>3</sup>Assistant Professor, Dept. of Computer Applications, ITER, Siksha o Anusandhan, Deemed to be University, Bhubaneswar, INDIA  
[nilimadas@soa.ac.in](mailto:nilimadas@soa.ac.in)

<sup>4</sup>Associate Professor, KL Business School Koneru Lakshmaiah Education Foundation (Deemed to be University) Vaddeswaram, AP, India [dr.bindumadhavi@kluniversity.in](mailto:dr.bindumadhavi@kluniversity.in)

<sup>5</sup>Assistant professor, Department of CSE, St.joseph's College Of Engineering, Chennai, TN, India [navee4@gmail.com](mailto:navee4@gmail.com).

<sup>6</sup>Assistant professor Department of ECE Sree Vidyanikethan Engineering College

Mohan Babu University Tirupati, AP, India [drkishore1609@gmail.com](mailto:drkishore1609@gmail.com)

DOI: 10.47750/pnr.2022.13.S09.530

## Abstract

With the advent of cutting-edge computing tools, the health industry underwent a period of rapid change that resulted in an explosion of medical data and the emergence of new academic disciplines. There have been significant attempts made to both manage the influx of medical data and extract actionable insights from it. Digital records management, data integration, or computer-aided diagnosis and disease predictions are all areas where the healthcare business is struggling. Both goals, lowering healthcare costs and moving toward individualized care, require this. Scientists were thus motivated to employ cutting-edge tools such as big data analytics, data modeling, machine learning, and learning algorithms to mine data for actionable insights. To overcome these obstacles, deep learning provides numerous resources. Predictive analytics applied to health data is rapidly becoming a game-changing resource that makes it possible to provide patients with preventative care. This paper portrays the study of numerous deep learning methods and tools in practice, and it focuses on the deep learning in healthcare, with a particular emphasis on the framework for supervised neural data gathering to decision making.

**Keywords-** Healthcare data, Deep learning (DL), Predictive analytics.

## 1. Introduction:

As national healthcare costs continue to rise, healthcare organizations will require ongoing connections to identify cost-cutting opportunities without sacrificing quality of treatment. As these patients account for a large portion of administrative expenditures, focusing on them is one way to meet these goals [1]. According to the Government Accountability Office, only 5% of Medicaid recipients incurred 50% of the program's costs [2]. These individuals frequently experience intractable, persistent problems that defy easy analysis and management. Healthcare providers may be able to save costs and improve patients' ability to manage their ailments if they are able to identify these individuals. This research explores a potential method for distinguishing between patients who will need to be

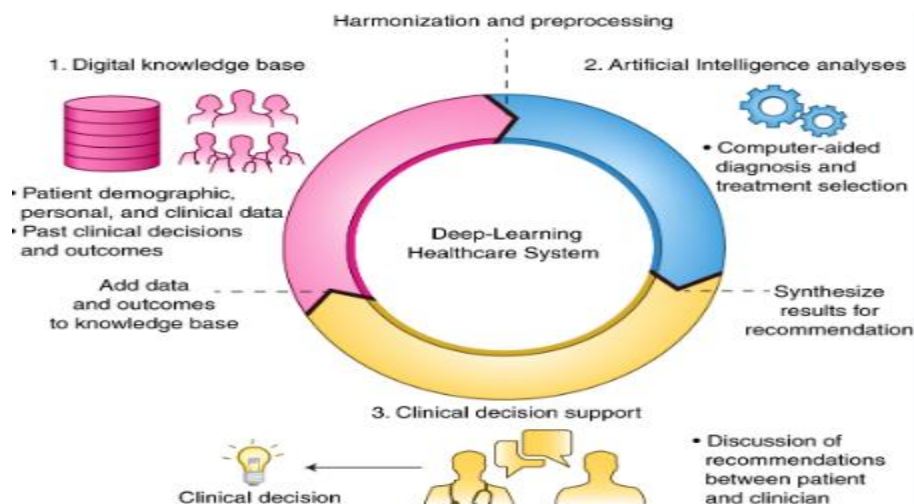
readmitted and those who will not. Individualized care can be given to these patients, and the information gleaned from them can be utilized to inform the development of better treatments for others in a similar situation. Top Healthcare system has started investigating how ML might improve company and customer loyalty as customers get increasingly particular about tailoring insurance purchases to their own needs. Effectively categorize their current tens of billions of items using machine learning techniques for data classification [3]. In this research, we present a framework for such cases that brings together two commonly used tools, Apache Spark and a Deep Learning driven stacked ensemble approach.

Artificial intelligence (AI) researchers in the healthcare sector are primarily concerned with creating algorithms and methods for evaluating the accuracy of a system's behavior when diagnosing sickness. Diagnosis is the process through which a medical professional determines the underlying cause(s) of a patient's symptoms and indicators. The physical examination and history of a patient are the usual starting points for a diagnosis [4]. It's not easy because many of the signs and symptoms are vague and require the expertise of medical professionals to identify. Therefore, it is difficult for nations like India and Bangladesh which have large but insufficient populations, to provide adequate diagnostic methods for all of their patients [5]. In addition, low-income persons may struggle to pay for necessary medical tests as part of the diagnostic process. Given the fallibility of humans, it's not shocking that a patient might experience overdiagnosis more frequently. Overdiagnosis can have detrimental effects on people's health and the economy if it leads to unneeded treatments like those that are often prescribed afterward [6]. At least one diagnostic error affects four out of five people at some point in their lives, according to a 2015 research by the National Academies of Science, Engineering, & Medicine [7].

The applications of machine learning (ML) are many, ranging from cutting-edge gadgets (such smart phones, computers, and robots) to emergency medicine. Disease diagnosis in medicine is only one area where ML is gaining traction. Machine learning-based disease diagnostics (MLBDD) has been shown to be promising by numerous academics and practitioners [8]. Common methods of diagnosis take a lot of time and money, not to mention the involvement of medical professionals. Traditional methods of diagnosis are limited by the capabilities of the individual, while ML-based systems are not, and robots do not tire like humans. As a result, an approach to disease diagnosis in settings where there are an unusually large number of patients may be created. X-ray and MRI scans, as well as tabular data on patients' ailments, ages, and sexes, are used to develop MLBDD systems [9].

When it comes to artificial intelligence (AI), machine learning (ML) is a subfield that relies heavily on data as an input [10]. Classification and regression are tasks that are notoriously challenging for humans, but can be easily accomplished by using specified mathematical functions. To give one concrete example, ML can be used to make the difficult task of identifying cancerous cells in a fluorescent microscope much easier. The most recent work [8] also demonstrates MLBDD efficiency of above 90% thanks to developments in machine learning (a type of machine learning). ML has been used to diagnose a wide variety of disorders, including Alzheimer's, heart failure, breast, and pneumonia. Deep learning (ML) algorithms are increasingly being used in the field of disease diagnosis, demonstrating the technology's potential benefits in the medical sector.

In a summary, there are many demanding fields to manage, but recent advancements in ML challenges, such as unbalanced data, ML interpretations, and ML morality in areas of medicine, are only a few examples. In order to better understand the current trend, techniques, and difficulties related to ML in illness diagnosis, we provide such a review that focuses on the unique applications of ML & DL in this area. We begin by providing an overview of many approaches to deep learning and machine learning techniques and specific architecture for recognizing and classifying different diseases. As can be seen in Figure1, Deep Learning is finding widespread use in the medical field.



**Figure 1. Deep Learning in Health Care**

## 2. Literature Survey:

The use of deep learning (DL) in a MI & medical natural language processing (NLP) has been a hot topic in the academic community for the past five years, prompting a slew of other reviews on the topic. The full layer-by-layer architecture of aims to equip has been reviewed by [11]. They brought attention to DL's challenges, such as its lengthy training period and high computational cost when using central processing units. Applications of DL in MI, clinical informatics, and medical technology have been the focus of research by [12]. Black box architecture, limited data size, over-fitting, generalization error, optimization of many hyper-parameters, and the possibility that even tiny changes to input data might have a huge effect on output were all mentioned as challenges to creating effective Deep models for health data. The uses of AI in the MI were discussed by [13]. Digital algorithms used on histopathology pictures were described by [14]. They talked about two different kinds of image processing pipelines: one where features are created by hand (the "handcrafted feature-based pipeline") and another where features are "learned" (the "learned feature-based pipeline"), with the latter involving the retrieval of high-level abstractions using DL techniques. In their research on DL techniques, [15] took a close look at both the techniques' optimization procedure and their use in MI. Different machine learning (ML) approaches used in the diagnosis of the degenerative condition were analyzed by [16]&[17].

He cited issues with medical records that impede DL implementation, including as erroneous and tiny data, nebulous language, acronyms, improper temporal representation, compartmentalized existence, lack of uniformity, lack of interpretability, and lack of explainability. The use of CAD, ML, and DL in radiography, pathology, drug development, lung nodules, molecular shuttling, obstetrics, & gynecology have all been studied recently. In their recent article, [18] surveyed the use of ML methods in the study of pathogenic variation or clinical genomics. They found that ML techniques aided in the comprehension of harmful variation, but that these methods' inherent complexity posed a hurdle that could restrict their potential impact on clinical genomics. There are a number of obstacles to deploying DL for EHRs, as noted by [19]&[20]. These include a lack of uniform benchmarks, data and labels, interpretability, and openness in the model. The application of DL methods to the evaluation of physiological signals was investigated by [21].

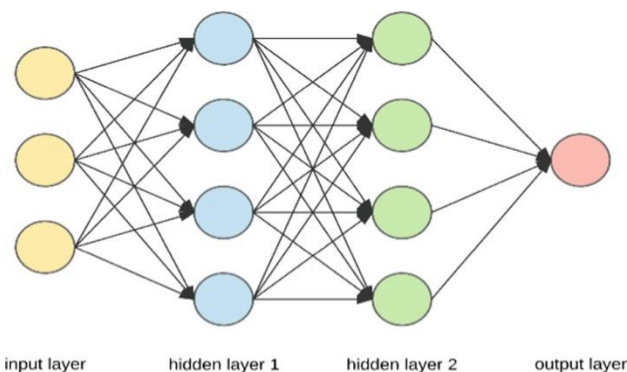
They pointed out problems with adopting DL, namely the inherent bias in medical records and cultural differences between people. Due to the findings of [22], the focus of DL research has shifted from data extraction to system enhancements. Endoscopic navigation is a broad field, and [23] analyzed a wide range of approaches to the topic. Clinical information extraction methods using DL were the subject of research by [24]. They observed that transfer learning is effective in dealing with issues such as sparse data, the need for professional expertise in medical textual

analysis, as recognition. Barriers to creating a DL-based model inside the health sector were studied by [25]. These included things like biased health records, age, gender, racism, etc. This was done in place of addressing biological causes and instead as risk factors. Recurrent neural neural net (MM-ResNet) has been proposed for Malaria detection by [26]. To determine the food value of soft drinks, [27] used a deep neural network (CNN). Cancer diagnosis has been implemented by [28] using a deep learning CNN architecture with a variety of layers. Using the Mean-shift-filter method, [29] have improved a medical image processing dataset with the aim of detecting cancer. When comparing Theano and tensor flow for cancer diagnosis, [29] found that Theano performed significantly better.

### 3. Proposed System:

To discover complex patterns in vast datasets, deep learning utilizes both increased processing power and multi-layered neural networks (See Fig. 2). In order to process complicated material with varying architectures, this variant of classical neural networks employs additional hidden layers.

Fig. 2. An illustration of deep learning with two hidden layers



AI technologies represented by deep learning topped Gartner's list of the top ten latest innovations for 2018 [30]. These days, deep learning is used for a wide variety of tasks in computer vision [31], including computer vision, robotic vision, NLP, neuroscience, text recognition, NLP, picture recognition and classification, audio and speech identification, and more. The best results can be achieved through the use of deep learning since it collects and analyzes a huge amount of data, such as patients' records, hospital records, and insurance information, and then applies its neuronal net-works. Because to its representational or recognition superiority, deep learning plays a crucial role in helping healthcare professionals determine, anticipate, evaluate, and put its ideas into practice when providing care.

Layers of infrastructure, data, analytics, and applications are required to construct an analytical environment. Several frameworks and applications can rely on the services provided by the infrastructure layer. Electronic health records (EHRs), pictures, genetic information, and eligible practitioner (EP) data are all included in the data layer's scope. For predictive analytics, the analytics layer should accommodate a variety of classification, clustering, regression, and other methods. Data analytic methods, such as natural language processing, can be integrated to examine subsets of healthcare data due to its unstructured nature. Predictive model findings can be displayed in an application layer. The suggested framework (shown in Fig. 3) begins with the creation of clinical data and concludes with clinical decisions; it is inspired by clinical challenges and can ultimately be used to improve clinical practice with the use of deep learning methods. The stages of the suggested framework are as follows:

- Data creation for clinical use;
- Natural language processing to improve unstructured data
- The use of deep learning for pattern discovery in large datasets
- Medical judgment making

A medical database is built by compiling information from several sources, including electronic medical records (EMRs) and clinical procedures. Patient information gathered over the course of diagnosis, testing, and treatment constitutes the clinical activities. All of the necessary data for DL algorithms to run directly exists in the form of machine-understandable images, EP information, and genetic information. Physical examination notes, clinical laboratory results, and similar narrative forms of clinical information are not organized in any particular way. Natural language (NLP) is helpful for unstructured data processing because it can help extract relevant information from narrative text to support clinical decision making. Two of the most important parts of an NLP pipeline are text processing and classification. Using the previously collected data, NLP is able to extract a list of disease-related keywords from the clinical notes. When determining which keywords to use, we look at how they influence the separation of normal and atypical situations. Ultimately, the goal of developing the NLP pipeline is to aid in medical decision making. Models similar to those described above are currently being implemented by the authors.

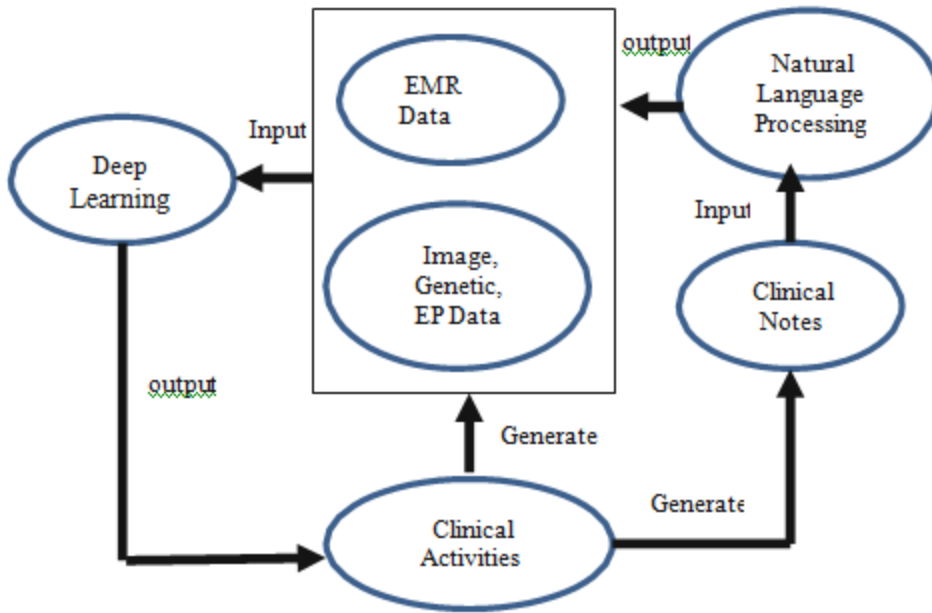


Figure 3. Proposed Architecture of the Deep Learning Model

#### 4. Evaluation of the Proposed Deep Learning Model:

The proposed methods are simulated using the MATLAB. The following variables, as shown in the equations, are used to evaluate the proposed system's performance by using the following parameters.

$$Sensitivity (Se) = TP / (TP + FN) \text{---(1)}$$

$$Specific(Sp) = TN / (TN + FP) \text{---(2)}$$

$$Accuracy(Acc) = (TP + TN) / (TP + FN + TN + FP) \text{---(3)}$$

However, these characteristics are determined by how many image pixels are correctly and incorrectly identified pictures.

Table 1 shows the evaluation of performance with respect to sensitivity, specificity and accuracy using Proposed Deep Learning Model

EMR Images	Sensitivity (%)	Specificity (%)	Accuracy (%)
10	98.3	94.5	94.5
20	97.6	93.4	93.2
30	96.2	92.7	92
40	94.5	92	91.5
50	94	91.6	91
60	93.5	91	90.5
70	93	90.89	90
80	92.7	90.4	89.4
90	92	89.5	88.6
100	91.6	88.9	87.5
<b>Average</b>	<b>94.34</b>	<b>91.4</b>	<b>90.8</b>

We have evaluated the three parameters i.e Sensitivity, Specificity and Accuracy of the EMR images which are given as input to the deep learning model. if there is any noise in the images the results will not be shown very well,as the images are clear the evaluated and average Sensitivity is 94.34%, Specificity is 91.4% and finally the Accuracy is 90.8% of our model which is reasonably good when compared to many of the existing deep learning Models and is shown in the Figure 4.

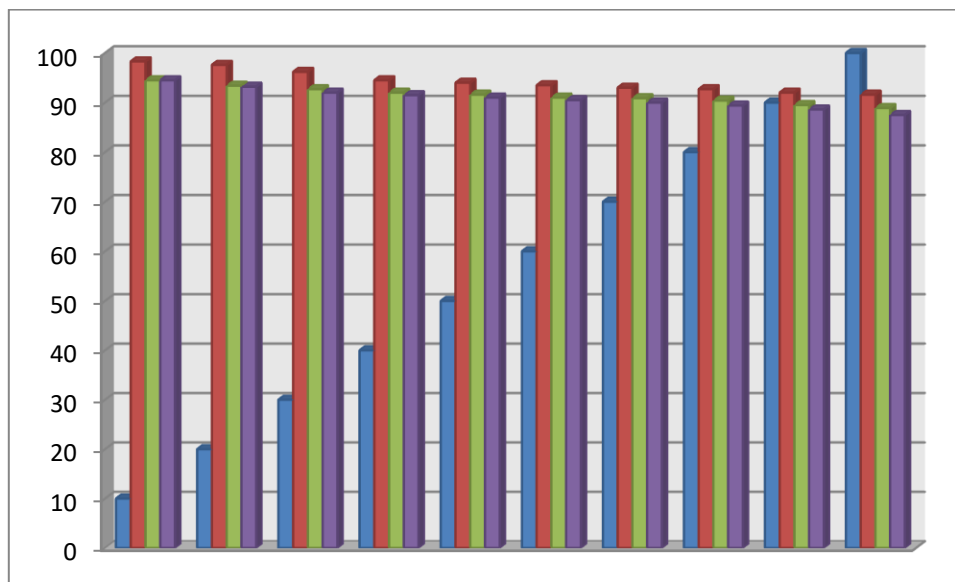


Figure 4. Evaluated Results of the Proposed Deep Learning Model

## 5. Predictions on Deep Learning's Future Impact on Healthcare:

Researchers and marketers in the healthcare industry are confident that deep learning management systems will replace humans in the near future, with the result that not only will diagnosis be conducted by intelligent machines, but they will also help prescribe medications, anticipate disease, and mentor in treatment. Since the introduction of digital imaging, deep learning methods have been increasingly used to analyze medical images. In order to process more patients' medical data across the five hospitals, Google Deep Brain Health is collaborating with the National Health Service of the U.k.. Recently, IBM Watson [32] expanded its capabilities by purchasing Merge's medical management solution, which includes coverage for radiology, cardiology, orthopedics, ophthalmology, and other healthcare disciplines, for an estimated \$1 billion.

They claim to employ deep learning to glean useful insights from millions of clinical cases, which in turn improves clinicians' speed and accuracy [33]. [34] use a deep learning approach to aid parents in keeping tabs on their children's health in real time via a mobile device, cutting down on the number of times the family must make doctor's appointments. In the future, equipment like the da Vinci will be able to do complex surgical procedures with the help of machine learning that combines visual data and movement patterns. There is a growing interest in using deep learning for the processing of medical imaging data, however the field is not without its difficulties. A fascinating and expanding field of study, the application of deep neural networks to diagnostic imaging data faces challenges such as the lack of a suitable dataset, the need for specialized medical experts, the use of nonstandard data machine-learning algorithms, concerns over patient privacy, and potential legal ramifications. Deep learning is used in personalized medicine to provide recommendations and diagnose and treat diseases based on a person's unique set of circumstances, including their health information, genetic makeup, current health, lifestyle habits, and stress levels, among other

factors.

## 6. Conclusion:

With the intent of employing deep learning in healthcare, we have presented a quick review of deep learning as it relates to predictive analysis of healthcare data. We demonstrated the many applications of DL in healthcare and conducted a comprehensive study of the most common disease categories for which DL has been used. The primary goal of this research was to suggest a paradigm for healthcare data monitoring based on DL with predictive modeling. As a result, the developed framework can be put into practice to confirm empirical results and demonstrate its value. By analyzing the latest tendencies in deep learning, we've pinpointed an important subfield that has the potential to be highly beneficial for medical imagery and is drawing attention in unsupervised learning. We also anticipate that deep learning methods will be used to the health industry's varying domains. As a result, deep learning will have a significant effect on the analysis of healthcare data.

## References:

- [1] M. A. Cucciare and W. O'Donohue. "Predicting future healthcare costs: how well does risk adjustment work?", *Journal of Health Organization and Management*, vol. 20, no. 2, pp. 150–162, 2020, PMID:16869351.
- [2] Medicaid: A small share of enrollees consistently accounted for a large share of expenditures, no. 15-460. United States Government Accountability Office .
- [3] Pandey S., M S., Shrivastava A, "Data Classification Using Machine Learning Approach", Thampi S., Mitra S., Mukhopadhyay J., Li KC., James A., Berretti S. (eds) *Intelligent Systems Technologies and Applications. ISTA 2021. Advances in Intelligent Systems and Computing*, vol 683. Springer.
- [4] McPhee, S.J.; Papadakis, M.A.; Rabow, M.W. (Eds.) *Current Medical Diagnosis & Treatment*; McGraw-Hill Medical: New York, NY, USA, 2010.
- [5] Ahsan, M.M.; Ahad, M.T.; Soma, F.A.; Paul, S.; Chowdhury, A.; Luna, S.A.; Yazdan, M.M.S.; Rahman, A.; Siddique, Z.; Huebner, P. Detecting SARS-CoV-2 From Chest X-ray Using Artificial Intelligence. *IEEE Access* 2021, 9, 35501–35513.
- [6] Macha Babitha. "Machine Learning Based Identification of Covid-19 From Lung Segmented CT Images Using Radiomics Features." *Biosc. Biotech. Res. Comm. Special Issue* 14, no. 07: 350-355.
- [7] GT. Pavan Kumar. "A Dynamic and adaptive learning mechanism to reduce cross layer attacks in cognitive networks." *Materials Today: Proceedings* (2020).
- [8] K. Jyothi Prakash. "Internet of things: IETF protocols, algorithms and applications." *Int. J. Innov. Technol. Explor. Eng* 8.11 (2019): 2853-2857.
- [9] Sangamithra, B. "A memetic algorithm for multi objective vehicle routing problem with time windows." 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE). IEEE, 2017.
- [10] Harika, A., et al. "Business process reengineering: issues and challenges." *Proceedings of Second International Conference on Smart Energy and Communication*. Springer, Singapore, 2021.
- [11] P. Sai Kiran,. "Resource aware virtual machine placement in IaaS cloud using bio-inspired firefly algorithm." *Journal of Green Engineering* 10 (2020): 9315-9327.
- [12] Natarajan, V. Anantha, et al. "Prediction Of Soil Ph From Remote Sensing Data Using Gradient Boosted Regression Analysis." *Journal of Pharmaceutical Negative Results* (2022): 29-36.
- [13] Kumar, M. Sunil, et al. "Automated Extraction of Non-Functional Requirements From Text Files: A Supervised Learning Approach." *Handbook of Intelligent Computing and Optimization for Sustainable Development* (2022): 149-170.
- [14] V. Tamizhazhagan. "Forecasting of Wind Power using LSTM Recurrent Neural Network." *Journal of Green Engineering* 10 (2020).
- [15] Prasad, Tvs Gowtham, et al. "Cnn Based Pathway Control To Prevent Covid Spread Using Face Mask And Body Temperature Detection." *Journal of Pharmaceutical Negative Results* (2022): 1374-1381.
- [16] Burada, Sreedhar. "Computer-Aided Diagnosis Mechanism for Melanoma Skin Cancer Detection Using Radial Basis Function Network." *Proceedings of the International Conference on Cognitive and Intelligent Computing*. Springer, Singapore, 2022.
- [17] Kumar, M. Sunil, et al. "APPLYING THE MODULAR ENCRYPTION STANDARD TO MOBILE CLOUD COMPUTING TO IMPROVE THE SAFETY OF HEALTH DATA." *Journal of Pharmaceutical Negative Results* (2022): 1911-1917.
- [18] B. E. Manjunath Swamy. "Personalized Ranking Mechanism Using Yandex Dataset on Machine Learning Approaches." *Proceedings of the International Conference on Cognitive and Intelligent Computing*. Springer, Singapore, 2022.
- [19] Ganesh, D., et al. "Improving Security in Edge Computing by using Cognitive Trust Management Model." *2022 International Conference on Edge Computing and Applications (ICECAA)*. IEEE, 2022.
- [20] Anantha Natarajan, V., Kumar, M. S., & Tamizhazhagan, V. (2020). Forecasting of Wind Power using LSTM Recurrent Neural Network. *Journal of Green Engineering* 10.
- [21] Faust, O., Hagiwara, Y., Hong, T.J., Acharya, U.R., 2018. Deep learning for healthcare applications based on physiological signals: A review. *Comput. Methods Programs Biomed.* 161, 1–13.

- [22] Alawad, M., Gao, S., Qiu, J.X., Yoon, H.J., Christian, J.B., Penberthy, L., Mumphrey, B., Wu, X.-C., Coyle, L., Tourassi, G., 2020. Automatic extraction of cancer registry reportable information from free-text pathology reports using multitask convolutional neural networks. *J. Am. Med. Inform. Assoc.* 27 (1), 89–98.
- [23] Liu, Z., Wang, X., Chen, Q., Tang, B., Xu, H. 2019. Temporal indexing of medical entity in Chinese clinical notes. *BMC Med. Inform. Decision Making*, 19, art. no. 17. <https://doi.org/10.1186/s12911-019-0735-x>.
- [24] Zhou, L., Suominen, H., Gedeon, T., 2019. Adapting state-of-the-art deep language models to clinical information extraction systems: Potentials, challenges, and solutions. *J. Med. Internet Res.* 21 (4).
- [25] Prospero, M., Min, J.S., Bian, J., Modave, F., 2018. Big data hurdles in precision medicine and precision public health. *BMC Med. Inform. Decision Making* 18 (1).
- [26] Pattanaik, A.P., Mittal, M., Khan, M.Z., Panda, S.N., 2020. Malaria detection using deep residual networks with mobile microscopy. *J. King Saud Univ. Comput. Inform.*
- [27] Hafiz, R., Haque, M.R., Rakshit, A., Uddin, M.S., 2020. Image-based soft drink type classification and dietary assessment system using deep convolutional neural network with transfer learning. *J. King Saud Univ. Comput. Inform.*
- [28] Prospero, M., Min, J.S., Bian, J., Modave, F., 2018. Big data hurdles in precision medicine and precision public health. *BMC Med. Inform. Decision Making* 18 (1).
- [29] Pesaraghader, A., Matwin, S., Sokolova, M., Pesaraghader, A., 2019. A. DeepBioWSD: Effective deep neural word sense disambiguation of biomedical text data. *J. Am. Med. Inform. Assoc.* 26 (5), 438–446.
- [30] Top Strategic Technology Trends for 2018: <http://www.gartner.com/technology/research/top-10-technology-trends/>. Accessed 15 Nov 2018
- [31] Howard, J.: The business impact of deep learning. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 1135 (2013)
- [32] Lohr, S.: IBM is counting on its bet on Watson, and Paying Big Money for It: <https://www.nytimes.com/2016/10/17/technology/ibm-is-counting-on-its-bet-on-watson-and-paying-big-money-for-it.html>. Accessed 15 Nov 2018
- [33] Enlitic: <https://www.enlitic.com/>. Accessed 15 Nov 2018
- [34] Cellscope: <https://www.cellscope.com/>. Accessed 15 Nov 2018