International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I3P231

Volume 69 Issue 3, 201-210, March 2021

A Survey on Artificial Intelligence in Cancer Medical and Nonmedical Datasets

Ajni K Ajai[1], A. Anitha[2]

Ph.D. Scholar[1], Associate Professor[2],
Department of Computer Science and Engineering,
Noorul Islam Centre for Higher Education, Tamil Nadu, India

[1]ajnikajai@gmail.com, [2]anithadathi@gmail.com

***Abstract** - Advances are the outcome of continually building on previous findings and surveillances. The study of cancer intends to a day when all cancers are cured by expanding efficient methods to prevent, detect, diagnose, treat cancer. This survey can accumulate extensive knowledge about solving meaningful, challenging, and neglected problems in cancer research. When the prognosis is worse and the treatment options are more critical, it leads the patients to the late stage of the disease, but if cancer diagnoses early, survival will be significantly improved. AI is changing our lives, and its work is detonating biomedical research and health care. The application potentials of AI are huge in all levels of cancer research. The integration of AI technology into cancer care is about saving a life through image analyzing, improving accuracy, speeding up the diagnosis, aid clinical decision-making, and patient triage with debility to reduce variation and duplicate testing. The subsets of AI are machine learning and deep learning. This review, pointing to the experimentally proved problem-solving for some challenging issues through their most effective methods.*

Keywords — *Clinical information, Deep learning, Early detection, Image analysis, Machine learning, Prevention.*

I. INTRODUCTION

Artificial means it can be anything that is made by human things that are not natural. Intelligence is the ability to understand, think, and learn. Combining these two, artificial intelligence is a vast field of computer science that makes machines seem to have human intelligence. This intelligence built using complex algorithms in mathematical functions. The goal of AI is to mimic the human brain and create systems that can function intelligently and independently. AI can manifest

Itself in many different ways. It uses Smart Speakers, Healthcare, E-Commerce, Smartphones, Smart Cars and Drones, Social Media Feeds, Smart Home Devices, Security and Surveillance, Smart Keyboard Apps, etc. [1]. Big data, Learning, Reasoning, and problem-solving are the key characteristics of an AI system [2]. In short, AI provides machines with the capabilities to adapt, reason, and provide solutions. Artificial Intelligence has three stages. It is based on its capability Artificial Narrow Intelligence is known as weak AI, which is the stage of AI that involves machines that can only perform a narrowly defined set of specific tasks. Known as the Artificial General Intelligence Strong AI, this is the stage of the evolution of AI, in which machines can think and act like humans. The capability of the artificial superintelligence computer is that humans overrun. Based on AI's functionalities, it has four types: Reactive machines, Limited memory, Theory of mind, and Self-awareness [3]. Reactive Machine AI includes machines that rely on existing data only to take into account the current data. Information and better decisions can be made by studying past data from the memory of a limited memory AI. An intelligence can better understand human beliefs and thoughts and is called the Theory of Mind AI. It is focusing on emotional intelligence. Some Machines have their consciousness and become self-aware, and that machines are said to be The Self-Aware AI. This kind of AI does not exist yet.

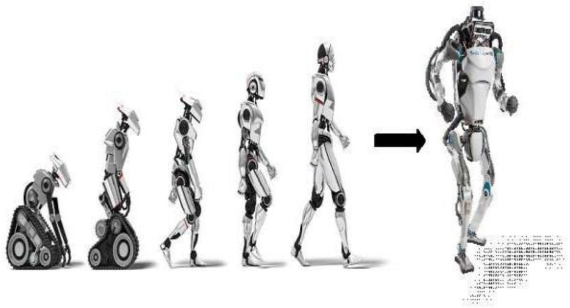

Fig. 1. Evolution and Future of AI Machines. Source: Adapted from [30]

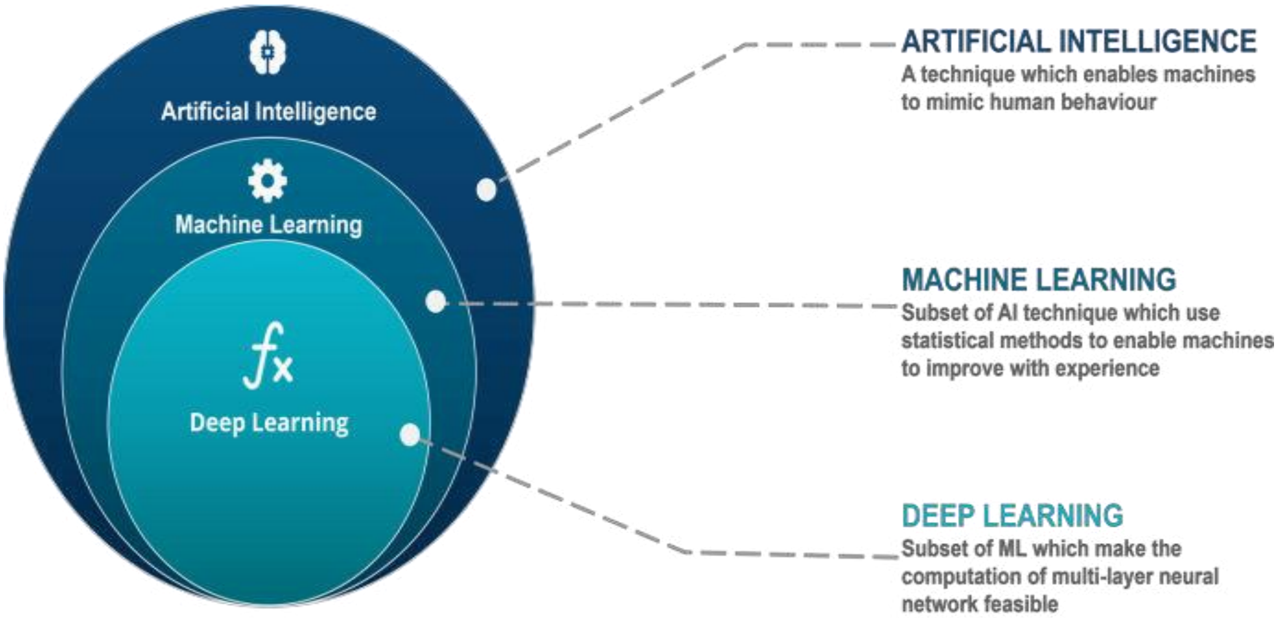

Fig. 3. AL vs. ML vs. DL. Source: Adapted from [32]

Fig. 1 shows the artificial intelligence machines of evolution and the future. The Talos was an example giant animated bronze warrior under the Greek Mythology in classical ages, programmed to guard Crete's island created by Hephaestus. Alan Turing speculated and published a landmark paper about the possibility of creating machines that think in 1950. In the philosophy of AI, the Turing test is the first series proposal. 1951 this period is called Game AI because many computer scientists developed programs for checkers and chess. The vital year 1956, the birth of AI. John McCarthy first coined the term "Artificial Intelligence" at the Dartmouth Conference. Then, followed by the first AI laboratory, "MIT AI Lab," in 1959, the research began. In the 1960s, the first robot and AI chatbot were introduced. Later in 1997, IBM's Deep Blue World Champion defeated Gary Kasparov in a chess game. When 2005, an autonomous robotic car name Stanley wins the DARPA Grand Challenge. In 2011 the two greatest Jeopardy! Champions Brand Rutter and Ken Jennings defeated IBM's question answering system by Watson. In today, AI research is persistent and growing [4]. According to Technology writer Alice Bonasio, AI research has grown 12.9% annually worldwide over the past five years. China is estimated to become a comprehensive global source of artificial intelligence in the next four years. Europe is the most significant and most miscellaneous region with high levels of international coaction within artificial intelligence research. India stands in third place after China and the United States in terms of AI research results. In the future, AI is about to become a little less artificial and a lot more intelligent. As AI becomes more intelligent, it will take on the role of an intelligent worker. While most of the repetitive and largely time-oriented tasks can be automated using AI. We still need human experts who can build and maintain these synthetic BOTS and work on a job's strategic and creative aspects. AI is a boon that can help optimize a job. It does not mean AI versus human; instead, it is a combination of human and AI versus the actual problem statement.

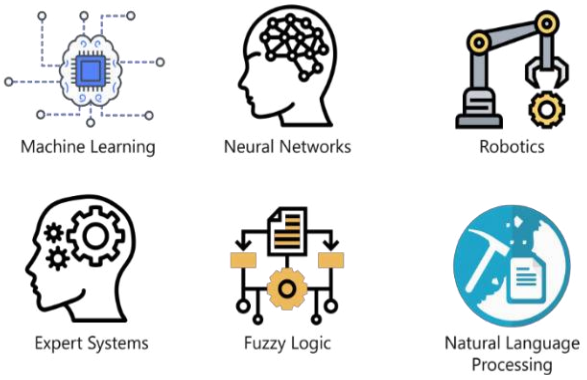

Fig. 2. The domain of AI. Source: Adapted from [31]

Fig. 2 shows the domain of artificial intelligence. There are several miscellaneous areas in AI ethics research; the most significant volume of research development shows Neural Networks, Machine Learning, Probabilistic Reasoning, and Computer vision. Artificial intelligence, Machine Learning, and Deep Learning are related fields (fig.3). ML & DL assists AL with a set of algorithms and neural networks to solve data-driven problems. People learn from bygone experiences, and machines follow up the instructions which will be given by people. It focuses mainly on designing the system, thereby allowing them to learn and make predictions based on some experiences.

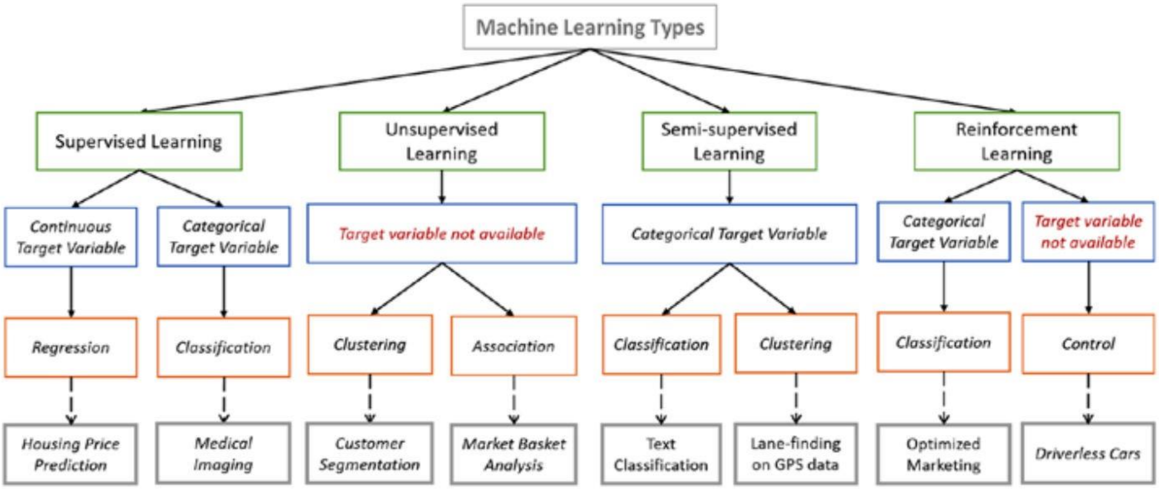

Fig. 4. Types of ML. Source: Adapted from [33]

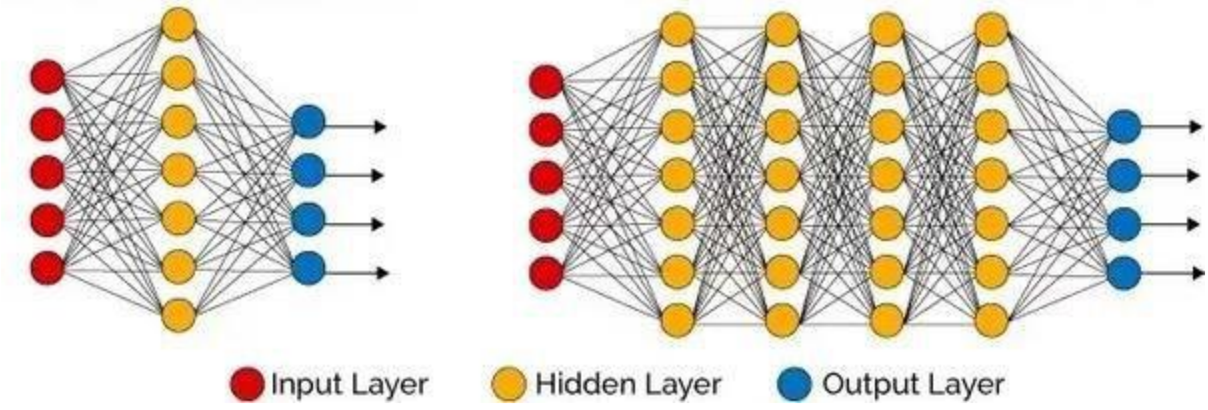

Fig. 6. Simple Neural Network and Deep Learning Neural Network. Source: Adapted from [35]

Fig. 4 shows the types of machine learning with their examples. Supervised learning uses labeled data to trains the model. Here, the machines know the object's features, and the label associates with those features. Training the machine with classified or unlabeled information in Unsupervised learning allows the algorithms to work on that information without guidance. Reinforcement Learning is to generalize the ML model, and the input is giving to a model that sends the output according to the algorithm applied. If the output is correct, it will take as the final result. Otherwise, provide corrective action to the training model and interrogates it to predict until it learns. Deep learning makes use of artificial neural networks in our brains. Fig. 5 shows the structure of Biological Neural networks and Artificial Neural networks. A neural network works when providing some input data. This data is processed through the layers of perception to create the desired output. Fig. 6 shows the difference between the simple neural network and the deep learning neural network. A collection of statistical ML methods used to study feature sequences based on the ANN concept is called DL. There are lots of applications in Artificial Intelligence.

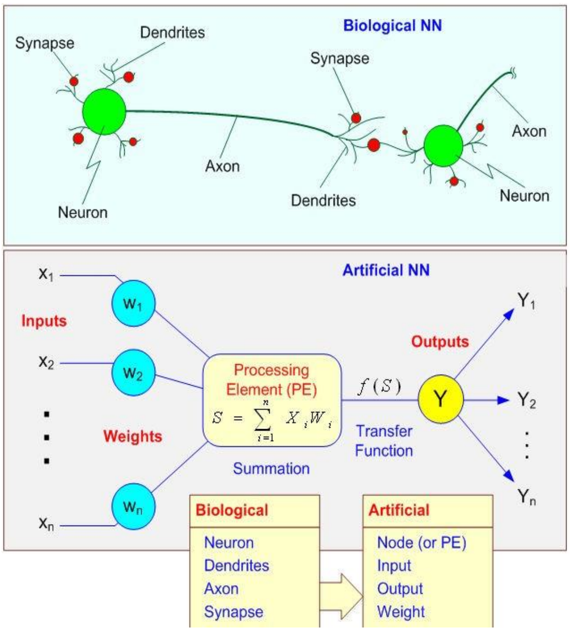

Fig. 5. Biological neuron vs. Artificial Neuron. Source: Adapted from [34]

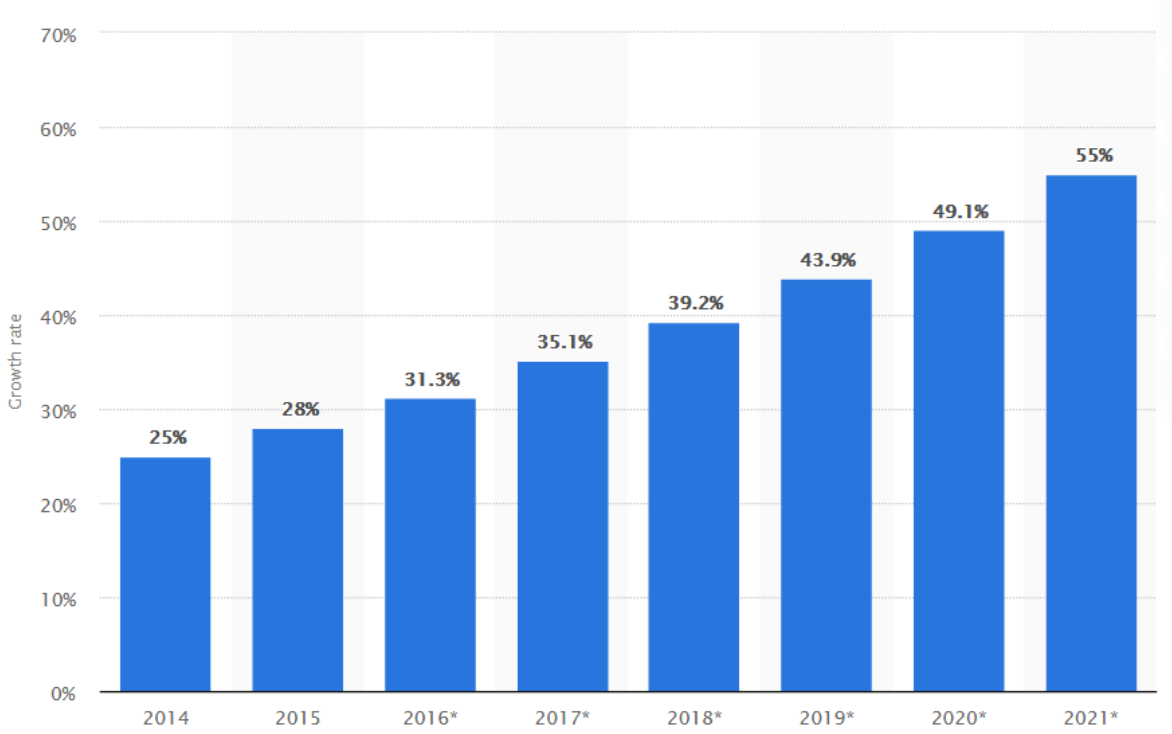

Fig. 7. Growth rate for the AI healthcare. **Source: Adapted from [36]**

AI is estimated that by 2030, AI will contribute more than $ 1 trillion to the world economy, and the most significant impact of AI will be on the health care sector. A couple of uses AI cases in healthcare, one of the uses is AI in managing medical data. Nuance provides AI-powered solutions to support doctors reduce documentation time and enhance reporting quality. Nuance prediction services use AI to predict customer intent and take the right action to meet their needs. By 2024 Global Market Insights shows that Artificial Intelligence should grow by more than 40 percent to surpass the US $ 2.5 billion powered for medical imaging and diagnosis. Voxel Morph is a Conventional NN that took seconds to perform MRI analysis, and the same analysis takes over 2 hours for a conventional MRI analysis program. In medical aid, virtual nurses monitor the patient's condition and follow up on treatment during doctor visits. In decision making, surgical robots can reduce case-to-case variations and even improve the best surgeons' efficiency. Fig. 7 shows the growth rate for the artificial intelligence healthcare market worldwide from 2014 to 2021.

II. LITERATURE SURVEY

The high dimensionality and complexity scrutinized cancer detection from the omic data is a challenge. The wordsmith (Singh, Verma et al. [5])

Pose a robust fuzzy PCA with interval type-2(IT-2) fuzzy membership functions for feature extraction of omics data problems. Here examined the system rendition of extensively used classifiers using the features extracted by proposed methods. The output indicates the newly implemented scheme may furnish a path of ameliorating molecular cancer diagnostics and prognostics.

In this paper, the authors Ellis et al. [6] discuss that when taking high-resolution histopathological images, the actual image is used as input to the deep learning model, while converting the original image to a lower resolution demands a lack of data. The Deep Selective Attention Design proposed selecting valuable areas in the original images for classification to address this issue. When analogized with the premature approach, the selection mechanism in work could be conserving the training time by 50%.

Onur Ozdemir et al. [7], the presented system is wholly premised on 3D convolutional neural networks. A full CADe/CADx system establishes the low dose CT scans to prognosis lung cancer. The coupling between detection and diagnosis components is said to be thought-out importance, and nodule detection systems are typically designed and optimized. The stage obliterates its need along with that utilizing this coupling allows elaborate a system that has forward like throughout first to last and too vigorous rendition. The degree of model uncertainty makes it possible for the system to supply valued subdivision probabilities. Research paper done by Chunyu Wang et al. [8] improves the prognosis performance of cancer survival. Genomic and clinical multiple data integrating a GCGCN were brought forward in the paper. With the use of similarity network fusion algorithm, multiple genomic data and clinical data were integrated, acquired sample similarity matrix, and using min-redundancy max-relevance mRMR feature selection algorithm, and cancer survival-related features were extracted. The influence of purposeless features diminished, and through the graph, convolutional network classification training and prediction were conducted. The GCGCN predicts better outcomes in cancer survival prediction methodology that combines data than current approaches. Sarfaraz Hussein et al. [9] analyzed tumor characterization through CAD tools, making it possible for the non-invasive cancer staging, prophesy, and foster personalized treatment to be available of precision medicine. Deep learning algorithms establish massive benefits, notably by utilizing a 3D CNN. The MTL and transfer learning is applied for 3D deep networks to ameliorate lung nodules' peril stratification. The noteworthiness of miscellaneous imaging attributes resembling lung nodules is also analyzed, including that speculation, texture, classification, and others for risk assessment. The second stage is an unsupervised learning algorithm for solving and exploring the limits.

Jianhua et al. [10] author design the model with the support of information technology for breast cancer prevention and control. This paper adopts cloud computing to showoff association rule-based breast cancer. There are two phases, phase1 and phase2, in phase1, titled prevention and control and the recommend algorithm are IAR and N-IAR. In phase2, they introduce a new frame based on machine learning as real data and more risk elements. Moreover, the author (Sharma et al. [11]) reports no licensed vaccination or HTLV-1 infection, and rapidly the diagnostic procedures prescribed to detect HTLV-1 need to elaborate and well-organized. The first study with 64 hybrid machine learning techniques proposes predicting different types of HTLVs by Penman. The hybrid techniques are composed of four classification methods it considered filtering functions that enhance the STA performance, four feature weighting, and four feature selection techniques. The dataset is labeled here using the K-Media clustering algorithm, which then performs feature weighting to identify key features for training machine learning models.

Tianshu Feng et al. [12] examines that TDA is a powerful method for consuming data dimensionality, mining data relationships, and naturally representing data structure. By harnessing a filter function used to reconstruct, a mapper algorithm is a tool for integrating large amounts of data into 1-dimensional space. STA's new framework is demonstrated using the simulation data and then applies to samples. The data's topological structure is the ordination/manifold methods of iso map and T-SNE; Whole image classification of both 2D mammogram and 3D tomosynthesis images developing and optimizing, the writers Jinze Liu et al. [13]have been done this in the preliminary work on machine learning models. Ten various CNN architectures are learned here and conclude that the ameliorating classification performs combining both data augmentation and transfer learning methods in a convolutional neural network.

Can Taylan Sari et al. [14] paper present a semi-supervised classification method for histopathological tissue images? There are three benefactions; at first, the scheme executes to establishes salient subregions in an image along with that the scheme uses only the size of these key sub-areas for image presentation and classification. As the second benefactions a new unsupervised technique to apprehend the sub-region quantization brings up, and it accomplishes building a more profound trust network of continuous RBMs. The approach then provisions these in-depth features to a clustering algorithm to investigate the salient sub-areas dimensions in no guidance way. As its last benefactions, the acme is a triumphant demonstration of manipulating RBM. Among the breast cancer molecular features, the advent of deep learning methods and multi-dimensional data trots out occasions for a perfect understanding and, therefore, can amend diagnosis, treatment, and prevention. In this work, Dongdong Sun et al. [15] proposed an MDNNMD to predict breast cancer prognosis. Three independent DNN models to efficiently incorporate data are constructed, considering the heterogeneity of different types of data to generate a final multimodal DNN model.

Hsueh-Chien Cheng et al. [16] presents deep learning assisted volume visualization to illustrate complex structures, which are otherwise demanding for conventional approaches. The authors bring a new system to spectral methods free from obstruction user interactions with high-dimensional characteristics. Multi-view Knowledge-Based Collaborative (MV-KBC) Model Yutong et al. [17] examined to separate benign from malignant lung nodules on chest CT by considering the nodule appearance of nine view planes and the nodule heterogeneity.

TABLE 1. AI APPLIED TO VARIOUS KINDS OF CANCER PROGNOSIS BY THE TYPES OF DATA IN DIFFERENT STUDY

Reference	Year	Type Of Cancer	Type of Data	Methods	Results
[5]	2019	DLBCL-Harvard, Colon, Arcene, Leukemia, DLBCL-NIH, Ovarian, Lung, Breast, WDBC, WPB C, Childhood Tumor, DLBCL- Stanford	Transcriptome And Proteome Datasets	IT-2FPCA(Interval type 2 fuzzy Principal component analysis), FPCA, LDA and SVM-RFE	The difference in average ranks :IT-2FPCA(1.87), FPCA(1.62), LDA(2.08) and SVM-RFE(3.37) CD0.1 = 1.45
[8]	2020	BRCA. , LUSC	Cancer genomic data and clinical data	Genomic and Clinical data based on Graph Convolutional Network (GCGCN), similar network fusion algorithm, mRMR	BRCA:Pre(0.769), Acc(0.7933), recall(0.9850), AUC(0.9280) LUSC: Pre(0.5539), Acc(0.7200), recall(0.8400) AUC(0.8050)
[10]	2019	Breast Cancer	Questionnaire attributes data	IAR ,N-IAR, machine learning classification algorithms(Logistic Regression,Decision Tree, Random Forest, XGBoost,Light GBM,MLP,Gail)	AUC(0.9655), Pre(0.9164), Recall(0.9248), F-Measure(0.9195), Acc(0.9191)
[11]	2019	Human T-cell Lymphotropic virus	Fasta format	64 hybrid machine learning methods	(TPR,TNR,PPV,NPV,F1Score, AUROC):0.99,Acc(99.85)
[12]	2019	Brain cancer, ovarian cancer	Genotype-Tissue Expression and transcriptome data	Semi-supervised topological analysis (STA) : Isomap and t-SNE	HR(1.02), CI(95%), P values(1.73E-06)
[15]	2019	Breast cancer	cancer patients'data	mRMR ,Deep neural network (DNN), MDNNMD	AUC(0.845), Sp=99.0%(threshold of 0.591): Acc(0.794),Pre(0.875), Sn(0.200),Mcc(0.356) Sp= 95.0% (threshold of 0.443) : Acc(0.826),Pre(0.749), Sn(0.450),Mcc(0.486)
[21]	2019	Breast cancer	EHRs	Statistical Feature Selection (SFS): statistical methods and XGBoost algorithm	AUC (0.8451), Youden index(0.535), cutoff(0.235), Sn(0.742),Sp(0.794), F-measure(0.605)
[22]	2018	Prostate, Leukemia, and Colon	Microarray data	Partial least squares (PLS)	Acc(89.65%)
[24]	2019	Breast, colorectal, glioblastoma, Hepatobiliary,lung, pancreatic	Gene Expression Omnibus (GEO)	SVM, Naïve Bayes (NB) and Random Forest (RF)	Acc(85%)
[27]	2020	Ovarian cancer	Clinical Proteomics, PPGL,Mass spectrometry data.	DAE,DCAE,PSO,ELMBPNN, SVM, KNN and RF	Sn: Clinical Proteomics(96.55%),PPGL(100%),mass spectrometry data(100%)

Full form of Table 1 & Table 2:Diffuse large B-cell lymphoma (DLBCL)-National Institute of Health(NIH),Wisconsin Diagnostic Breast Cancer (WDBC),Wisconsin Prognostic Breast Cancer (WPBC),Critical Difference(CD, Area Under the Curve (AUC),Accuracy (Acc),Precision(Pre),Item Association Rule (IAR), N-item IAR(N-IAR), Multi-Layer Perception(MLP),True Positive Rate(TPR),True Negative Rate(TNR),Positive Predicted Value(PPV), Negative Predicted Value (NPV), Area Under Roc Curve (AUROC), Hazard Ratios (HRs), Confidence Intervals(CIs),Sensitivity(Sn),Specificity(Sp),Matthew's correlation coefficient (Mcc), Phemotocytoma And Paraganglioma (PPGL),Particle Swarm Optimization (PSO),Extreme Learning, Mammography(MG),Tomosynthesis(TS), Computed Tomography(CT), Immunohistochemistry(IHC),Ultra Sound(US), Computed tomography laser mammography (CTLM),Magnetic Resonance imaging(MRI), Convolutional Neural Networks(CNN), Receiver Operating Characteristic(ROC)curve, Area Under the ROC curve (auROC), Standard Definitions(SD),Relaxed Definitions(RD),Volatile Organic Compounds(VOCs), Deep Neural Network (DNN).Multi-View Knowledge-Based Collaborative (MV-KBC), Fully Convolutional Networks (FCN), Mask- Guided Hierarchical Learning (MHL)Dice Similarity Coefficient(DSC), Region Proposal Network (RPN), Correct Localization (CorLoc), Electronic Health Records (EHRs), Mean Detection Error Rate (Mean eD), Multi-Layer Perceptron Neural Network(MLPNN), Gaussian Mixture Models (GMMs), Position Adjustment(PA), Intensity Adjustment (IA), Size Adjustment (SA),Dice's coefficient(DC), Mask Region-Convolutional Neural Network (Mask R-CNN), Feature Pyramid Network (FPN),Beetle Antennae Search(BAS

TABLE 2. AI APPLIED TO VARIOUS KINDS OF CANCER PROGNOSIS BY STAINING\ MODALITY IN DIFFERENT STUDY

Reference	Year	Type Of Cancer	Staining \ Modality	Methods	Results
[6]	2019	Breast Cancer	Microscopy	Deep selective attention network model	AUG in patient level (98.1%) and image level(97.9%) RAW in patient level (97.5% to 98.1%) and image level(96.6% to 97.9%),Acc(98%)
[7]	2019	Lung Cancer	CT	3D CNNs	CADe: Sn: (96.5%),CADx: ROC(95%),AUC(0.87) F1 score(0.40),Pre (0.25),Recall(0.93)
[9]	2019	Lung and pancreatic cancer	MRI, CT	3DCNN, Transfer Learning, proportion-SVM	Supervised:Acc(91.26%) Mean score(0.459) Unsupervised:Lung-Acc(78.06%) Sn(77.85%)Sp(78.28%) Pancreatic:Acc(58.04%) Sn(58.61%) Sp(41.67%)
[13]	2018	Breast cancer	2D MG,3D TS	Deep CNN	2D MG: auROC(0.73)
[14]	2019	Colorectal cancer	Microscopy	Restricted Boltzmann machines (RBMs)	Large Image:Normal SD:Pre(92.96), Recall(79.71), F-Measure(85.83) RD:Pre(99.48), Recall(88.37), F-Measure(93.60) Low grade SD:Pre(83.01), Recall(91.30), F-Measure(86.96) RD:Pre(91.03), Recall(93.32), F-Measure(92.16) High grade SD:Pre(70.82), Recall(98.61), F-Measure(82.44) RD:Pren(87.00), Recall(99.93), F-Measure(93.02)
[16]	2019	Brain tumor	Microscopy, MRI	The deep-learning-assisted technique, spectral methodsCNN	Dataset 1: VOC score(0:733 to 0:690) Dataset 2: VOC score(0:566 to 0:752) Time=3.4 minutes per million voxels
[17]	2019	Lung cancer	CT	DNN,MV-KBC	Acc(91.60), Sn(86.52), Sp (94.00),AUC (95.70)
[18]	2019	Breast cancer	MRI	FCN,MHL	AUC(69.01%), DSC(78:51±26:25) Se(73.31± 30.38), PPV(87.99± 20.05)
[19]	2018	Breast cancer	IHC	HER2 deep neural network (Her2Net)	Pre(96.64%), Rec(96.79%), NPV (93.08%),F-score(96.71%), Acc(98.33%) ,FPR(6.84%)
[20]	2019	Breast cancer	US	Faster R-CNN,RPN	CI(95%),CorLoc(80.00% to 84.50%)
[23]	2020	Lung cancer	CT	Median intensity projection, mRPN, 3D Deep Convolutional Neural Network (3DDCNN)	Acc(98.51%),S(98.4%), Sp(92%), FP(2.1), AUROC(96%), Time (sec)=0.025, Mean eD(3.30 ±0.58), MeaneC(1.28 ±0.19), Var eD(1.4 ±2.48), VareC(1.10±0.26), Std eD(2.88 ±3.61), StdeC(4.5±3.8), m-AP (59.7 ±3.2), Time (min)=2-4
[25]	2018	Breast cancer	CTLM	SVM, MLPNN	SVM:AUROC(0.98),Acc(97.9%), Sn(95.5%),Sp(96.1%)
[26]	2019	Lung cancer	CT	Mask R-CNN, Bayes,SVM, K-means,GMMs	Mask R-CNN+Kmeans: PA(99.21%), IA(99.40%), SA (97.62%),DC(97.33%), Acc(97.11%)Sn(96.58%), Mask R-CNN+SVM (96.69%).
[28]	2020	Lung cancer	CT	FPN,RPN,Mask R CNN,MNN	Recall(0.887) Pre(0.486) F-score(0.628),AP@50(0.882)
[29]	2020	Liver cancer	CT	CNN,BAS	Acc(92%)

Breast tumor segmentation is a thought-provoking problem based on DCE-MRI. The wordsmith Jun Zhang et al. [18] recommended a mask-guided hierarchical learning (MHL) framework for this problem. To precisely identified tumor regions, the writer creates two cascaded FCNs with breast masks as guidance. From (HER2)-stained breast cancer images with minimal user intervention, the authors Monjoy Saha, Chandan Chakraborty [19] present an efficient deep learning framework. A deep neural network (Her2Net) is proposing to solve this problem, and also it has convolutional and deconvolutional parts. For classification and error estimation, a fully connected layer and a softmax layer are using, and finally, method scores are calculating. Seung Yeon Shin et al. [20], based on observation, found the method conventional neural network-based mass detectors that train with big and feeble interpreted datasets is a significant issue, whereas over-fitting may happen with those trained in small and strongly interpreted datasets. In a hybrid manner, the penman uses a weakly interpreted dataset and a smaller strongly interpreted dataset to overcome these issues. Upon breast cancer, clinical information, the author's Bo Fu et al. [21] analyzed and cleaned the data. Based on statistical theory and gradient boosting decision tree framework have been constructed MP4Ei. It accepts stratified feature selection, including statistics and synthesis techniques on patient demographics, diagnosis, pathology, and therapy to select 23 of the patient's 89 features. To find an easy model, the variables imported the XGBoost algorithm with tuning parameters for 5-year prediction. Ho-Chun Wu, Si-Guang Wei, and Shing-Cha Chan [22] presented a criterion called novel consensus gene selection for analysis of PLS-based gene expression. A better classification rendition can be acquired to describe as similar or dissimilar with conventional techniques based on observational outcomes with micro-array datasets.

Anum Masood et al. [23] have recommended lung cancer diagnosis, and CAD systems develop to assist the radiologist nodule discovery process with a reference comment. To utilize .three-dimensional data from CT scans, the writers applied the Median Intensity Projection and the mRPN for self-operating the selection of a latency area of interest. The penman has integrated the cloud computing, trained, and tested to their cloud-based (3DDCNN). It demonstrates an in-depth study in conjunction with cloud computing to identify precise and fruitful lung nodules through CT imaging. In this study, Alkan et al.[24]presents the classification of cancer types and implemented healthy samples using a binary classifier. The selected features(genes) to be the cause of the correlation coefficient and data gain analysis would be trained SVM, NB, and RF algorithms. The interpretation of each machine learning model was experimentally using 10-fold cross-validation, and when the machine learning models estimate in terms of accuracy, in many cases NB robust than RF and SVM. Afsane Jalalian and Babak Karasfi [25] projected that CAD systems developed to automate carcinoma findings and classify unhealthy and malignant lesions in various ways such as ultrasound, mammography, and MRI. This paper includes two supervised machine learning methods: Support Vector Machine and Multi-layer Neural Network for Determination of Angiogenesis in the system. Overall quantitative results have shown that the SVM classifier provides better results when compared with MLPN in determining breast cancer in CTLM images.

Qinhua Hua et al. [26] suggested an automatic segmentation of the lungs in CT images, using the (CNN) Mask R-CNN, to uniquely model the lung area mapping model, integrating with supervised and unsupervised machine learning methods. Transfer learning is the erosion and dilation method, and the authors applied transfer learning to the final result of the Mask R-CNN network. Qingguo Zhou et al. [27] propose to use deep autoencoder to train latent illustration of high-dimensional mass spectrometry information. Traditional particle swarm optimization (PSO) algorithms manipulate to select optimized characteristics from mass spectrometry data. The results show that the DCAE model is better suited for high-dimension feature learning than the traditional PSO method, and the ELM and SVM models can be better used to detect cancers. In this paper, Linqin Cai et al. [28] examined the methods for supporting the medical experts to detect and diagnose the pulmonary nodules using mask R-CNN and ray-casting volume rendering algorithm. To test and evaluate the proposed method in this paper, they performed experiments, then the evaluation of the independent dataset shows that the scheme generalizes to the missing data. In this paper (Yang Han et al. [29]), the CNN algorithm and the BAS algorithm are combined and integrated into the Intelligent Recommendation System. The model can accurately propose the nutritional support plan for necessary cancer patients for postoperative recovery, and this has some guidance for cancer rehabilitation.

III. DISCUSSION

In this paper [10], the present review lacks real-time dynamic probabilistic methods and non-equilibrium data algorithms for early breast cancer detection and risk assessment. Although the research model [11] showboats better performance using 64 hybrid machine learning methods, some algorithms have their drawbacks. The authors [12] not disclosed novel subtypes to represent biologically relevant relations of the various subtypes of the generated graph based on the geometric structure. In this paper's treatment process [21], effective, accurate scientific methods are deficient, and there is a vast paucity amount of clinical or omics data. An algorithm for evaluating the p-values of genes related to cancer the wordsmith [22] has not been determined. Biomarker genes are required to classify different cancer types and analyze their biological functions in cancer diagnosis for this article [24]. It looks vague the model optimization in the treatise [27] and in no wise spectrometry data to improve and verify the classification models. The penman [9] omitted the visualization of activation map concepts, but it was advantageous for clinicians to identify new imaging biomarkers.

Nevertheless, the author could have used more unsupervised learning approaches to cope with unique medical imaging challenges. Meanwhile, the model [17] is computationally less efficient and diminishes the benign-malignant lung nodule classification. The unlabeled nodules could be used as training samples to dwindle the data annotation. The average segmentation time consuming has been done in this design [26], and according to this paper [28], the accuracy of detection and segmentation are minor. In this study, specified problems are there, the images in BreakHis dataset from raw histopathological data are cropped, resizing the image in the dataset of exiting model has not been assayed on the entire slide dataset, the small dimension of the testing dataset because outputs are biased, and only one dataset is evaluating [6]. When the CADe model misses a malignant lesion, the system's output shall be the hugest source of false negatives the subsequent CADx model has no way to classify as malignant. The writer himself states the work still has some limitations, such as the lack of extensive nodule annotations, and they tried to solve these problems, but the solutions were flawed. The patient referral and informative biomarkers are have been omitted in the acme [7]. Furthermore, the method has a lack that denotes the current sample size have limited by the availability of multiple data. To ameliorate the research direction, the authors [8] saying to amalgamate genomic data and take into account pathological image features and subtypes of patients.

To achieve classic rendition, the litterateur [13] does not set forth an assembled classifier, and it could have amalgamated with the 2D mammogram and 3D tomosynthesis data. There are no ways to probe [14] on the global feature set to dictate texture measures on the quantized subregions. The design could improve more if there were sophisticated algorithms and a new set of in-depth features. The research [15] has no intermingled of more omics data, and there is in absentia cannibalizing features from pathology images of cancer patients. In the delving [16], there are no possibilities of showing the substructures within a structure, but there is a proliferation of time spending on voxel-wise prediction. Some methods are malfunctions so that the acme [18] is inexact and awfully, and for improving the performance better, the structure information is related to the combined-study of molecular subtype classification and breast tumor segmentation. Obscurity allows pathologists to manually annotate the cell membrane because the design [19] contains images where the cytoplasm colored with HER2 monoclonal antibodies. In work [20], the image interpretation may be more running out of time due to its higher dimensionality. A problem that gave rise to the local optimal is the overfitting problem; likewise, the performance [23] could be less accurate and sluggish. In the research [29], there is no cancer rehabilitation recommendation scheme and has been inferior mathematical models.

IV. CONCLUSION

This survey paper overviewed the most recent studies on applying machine learning and deep learning methods in cancer prevention, detection, and diagnosis based on images and certain types of data. Specific types of cancer such as DLBCL-Harvard, Colon, Arcene, Leukemia, DLBCL-NIH, Ovarian, Lung, Breast, WDBC, WPBC, Childhood Tumor, DLBCL-Stanford, Lung squamous cell cancer(LUSC), Human T-cell Lymphotropic Virus, Prostate, Colorectal, Glioblastoma, Hepatobiliary, Pancreatic, Liver, Brain Tumor analyzed in the review with a variety of modalities /strainings and different types of medical or nonmedical data. ML algorithms work with a massive amount of structured data, whereas DL algorithms can work with an enormous amount of structured and unstructured data. DL algorithm is most required learning to perform complex operations since the ML algorithm has limitations. To ensure that a model's performance good, it is enough to consider the accuracy of that design. Under review, the study shows when the number of data increases, the ML algorithm's performance decreases. The DL algorithm takes large volumes of data as input, analyzes the input to extract features out of an object, and identifies similar objects.

REFERENCES

[1] R. Agarwal, 15 Examples of Artificial Intelligence You're Using in Daily Life, Beebom publishing, https://beebom.com/examples-of-artificial-intelligence/.,(2020).

[2] Stefan van Duin and Naser Bakhshi, Artificial Intelligence: Key characteristics of an AI, Deloitte publishing,(2018). https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/deloitte-analytics/deloitte-nl-data-analytics-artificial-intelligence-whitepaper-eng.pdf.

[3] R. Reynoso, 4 Main Types of AI, Learning hub, (2019). https://learn.g2.com/types-of-artificial-intelligence.

[4] C. Aguis, Evolution of AI: Past, Present, Future, Data-Driven Investor, (2019). https://medium.com/datadriveninvestor/evolutionof-ai-past-present-future-6f995d5f964a .

[5] V. Singh, N. K. Verma, and Y. Cui, Type-2 Fuzzy PCA Approach in Extracting Salient Features for Molecular Cancer Diagnostics and Prognostics, IEEE Trans. Nanobioscience, 18(3)(2019) 482–489.

[6] B. Xu et al., Attention by Selection: A Deep Selective Attention Approach to Breast Cancer Classification, IEEE Trans. Med. Imaging, 39(6)(2020) 1930–1941.

[7] O. Ozdemir, R. L. Russell, and A. A. Berlin, A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans, IEEE Trans. Med. Imaging, 39(5)(2020) 1419–1429.

[8] C. Wang et al., A Cancer Survival Prediction Method Based on Graph Convolutional Network, IEEE Trans. Nanobioscience, 19(1)(2020) 117–126.

[9] S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace, and U. Bagci., Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches, IEEE Trans. Med. Imaging,38(8) (2019) 1777–1787.

[10] J. Ma et al., Association Rule-Based Breast Cancer Prevention and Control System, IEEE Trans. Comput. Soc. Syst., 6(5) (2019) 1106–1114.

11] G. Sharma, P. S. Rana, and S. Bawa., Hybrid machine learning models for predicting types of Human T-cell Lymphotropic Virus, IEEE/ACM Trans. Comput. Biol. Bioinforma.,(2019) 1–1.

[12] T. Feng, J. I. Davila, Y. Liu, S. Lin, S. Huang, and C. Wang., Semi-supervised Topological Analysis for Elucidating Hidden Structures in High-Dimensional Transcriptome Datasets, IEEE/ACM Trans. Comput. Biol. Bioinforma., (2019) 1–1.

[13] X. Zhang et al., Classification of Whole Mammogram and Tomosynthesis Images Using Deep Convolutional Neural Networks, IEEE Trans. Nanobioscience, 17(3)(2018) 237–242,

[14] C. T. Sari and C. Gunduz-Demir., Unsupervised Feature Extraction via Deep Learning for Histopathological Classification of Colon Tissue Images, IEEE Trans. Med. Imaging, 38(5) (2019) 1139–1149.

[15] D. Sun, M. Wang, and A. Li, A Multimodal Deep Neural Network for Human Breast Cancer Prognosis Prediction by Integrating Multi-Dimensional Data, IEEE/ACM Trans. Comput. Biol. Bioinforma., 16(3)(2019) 841–850.

[16] H.-C. Cheng et al., Deep-Learning-Assisted Volume Visualization, IEEE Trans. Vis. Comput. Graph., 25(2)(2019) 1378–1391.

[17] Y. Xie et al., Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT, IEEE Trans. Med. Imaging, 38(4)(2019) 991–1004.

[18] J. Zhang, A. Saha, Z. Zhu, and M. A. Mazurowski., Hierarchical Convolutional Neural Networks for Segmentation of Breast Tumors in MRI With Application to Radiogenomics, IEEE Trans. Med. Imaging, 38(2)(2019) 435–447.

19] M. Saha and C. Chakraborty., <italic>Her2Net</italic>: A Deep Framework for Semantic Segmentation and Classification of Cell Membranes and Nuclei in Breast Cancer Evaluation, IEEE Trans. Image Process., 27(5)(2018) 2189–2200.

[20] S. Y. Shin, S. Lee, I. D. Yun, S. M. Kim, and K. M. Lee., Joint Weakly and Semi-Supervised Deep Learning for Localization and Classification of Masses in Breast Ultrasound Images, IEEE Trans. Med. Imaging, 38(3)(2019) 762–774.

[21] B. Fu, P. Liu, J. Lin, L. Deng, K. Hu, and H. Zheng., Predicting Invasive Disease-Free Survival for Early Stage Breast Cancer Patients Using Follow-Up Clinical Data, IEEE Trans. Biomed. Eng., 66(7)(2019) 2053–2064.

[22] H.-C. Wu, X.-G. Wei, and S.-C. Chan., Novel Consensus Gene Selection Criteria for Distributed GPU Partial Least Squares-Based Gene Microarray Analysis in Diffused Large B Cell Lymphoma (DLBCL) and Related Findings, IEEE/ACM Trans. Comput. Biol. Bioinforma., 156 (2018) 2039–2052.

[23] A. Masood et al., Cloud-Based Automated Clinical Decision Support System for Detection and Diagnosis of Lung Cancer in Chest CT, IEEE J. Transl. Eng. Heal. Med., 8 (2019) 1-1.

[24] C. B. Alkan and Z. Isik., Characterization of Cancer Types by Applying Machine Learning Methods on Blood RNA-Sequencing Data, in 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), (2019) 1–4.

[25] A. Jalalian and B. Karasfi., Machine Learning Techniques for Challenging Tumor Detection and Classification in Breast Cancer, in 9th Conference on Artificial Intelligence and Robotics and 2nd Asia-Pacific International Symposium, (2018) 19–24.

[26] L. Cai, T. Long, Y. Dai, and Y. Huang., Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis, IEEE Access, 8 (2020) 44400–44409.

[27] Q. Zhou, B. Yong, Q. Lv, J. Shen, and X. Wang., Deep Autoencoder for Mass Spectrometry Feature Learning and Cancer Detection, IEEE Access, 8 (2020) 45156–45166.

[28] L. Cai, T. Long, Y. Dai, and Y. Huang., Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis, IEEE Access, 8 (2020) 44400–44409.

[29] Y. Han et al., Artificial Intelligence Recommendation System of Cancer Rehabilitation Scheme Based on IoT Technology, IEEE Access, 8 (2020) 44924–44935.

[30] https://hackernoon.com/hn-images/1*RJ7KdSh-3wQvNy_hQZApzA.png

[31] https://www.edureka.co/blog/types-of-artificial-intelligence/

[32] https://www.youtube.com/watch?v=X-FKcenZ-jo

[33] https://en.proft.me/media/science/ml_types2.png

[34] http://3.bp.blogspot.com/-X81SfTlAm5Y/T4y7VyIo7cI/AAAAAAAAAIA/sLAzr5qYNYs/s1600/biological_neuron_vs_ANN.png

[35] https://cdn-images-1.medium.com/max/1600/1*dnvGC-PORSoCo7VXT3PV_A.png

[36] https://www.statista.com/statistics/938794/global-healthcare-artificial-intelligence-market-growth-rate/