International Journal of Innovative Research in Medical Science (IJIRMS)
Volume 05, Issue 10, October 2020,
https://doi.org/10.23958/ijirms/vol05-i10/978

Review article

Clinical Applications of Artificial Intelligence on Accuracy of Cancer Prediction, Detection, and Diagnosis

William Huang [1,2], Chunli Zhao [2], Xiujun Fan [*2]

[1]Monta Vista High School, 21840 McClellan Road, Cupertino, CA 95014 USA; williamhuang.whh@gmail.com
[2]Biolife Valley Education Group, 1751 Fortune Drive Unit G, San Jose, CA 95131 USA; chunli.cl@gmail.com

*Corresponding author: Xiujun Fan; *xiujun.fan@gmail.com*

Received 06 October 2020; Accepted 15 October 2020; Published 18 October 2020

Abstract

Ever since the world entered the age of information, scientists have looked into the developments and applications of the growing prospects of machine learning and neural networks. In particular, the ability for deep learning machines to determine the risk, survivability, and prognosis of tumors based on medical cancer databases has intrigued healthcare researchers seeking to improve these algorithms in recent years. There are distinct aspects of medical procedures where artificial intelligence (AI) training can be applied; for example, the calculation of risk scores for patients based on mammographic screening, analysis of the presence of biomarkers like spermine and other polyamines in fluids surrounding tumors, genomic and epigenetic assessments to map genes that influence cancer expression, as well as the utilization of metabolomic data from FTIR spectroscopy of a patient's biofluids to help make a more reproducible and conclusive diagnosis. The goal of this review is to discuss the progress of AI and deep learning in clinical procedures and applications in recent years and evaluate the efficacy of certain AI methods for tumor diagnosis, prognosis, and prediction based on patient information from available medical databases.

Keywords: ***cancer diagnosis; artificial intelligence; machine learning; cancer prediction; microarray; imaging; mass spectroscopy***

1. Insight into Evaluations of Cancer Detection and Prediction with Artificial Intelligence Techniques Over Time

Before the widespread accessibility of large amounts of data in the early 21st century, cancer screening was mainly done with human practitioners, limiting the success of such techniques to the judgement and decision making abilities of the person in control. The current precision of cancer diagnosis and prediction is by no means flawless; however, the understanding and integration of artificial intelligence (AI) and its subsets like machine learning and deep learning (DL) into imaging, spectroscopy, and other medical procedures have already shown increases in specificity and quality of measurements as opposed to evaluations done by human examiners [1].

Initially (**Figure 1**), medical machine learning and deep learning were applied to imaging through the viewing of slides from liquid biopsies, analyzing fine-needle aspirates and cytological features of certain tumors [2]. As time progressed, the applications of computer-aided diagnosis (CAD) evolved to accompany mammography and endoscopic imaging, creating a density-based risk score for patients. For example, a convolutional neural network used to grade gliomas achieved a classification accuracy of 96% when classifying glioblastoma multiforme (GBM) vs low-grade glioma (LGG) data sets [3]. Another approach that developed to utilize the reliability and reproducibility of machine learning was the analysis of the biofluids surrounding tumors for deviations in the presence of certain biomarkers [4]. The final significant method that combines artificial intelligence and cancer datasets is the prediction power of DNA microarrays and genomics [5]. The genomic assessment of

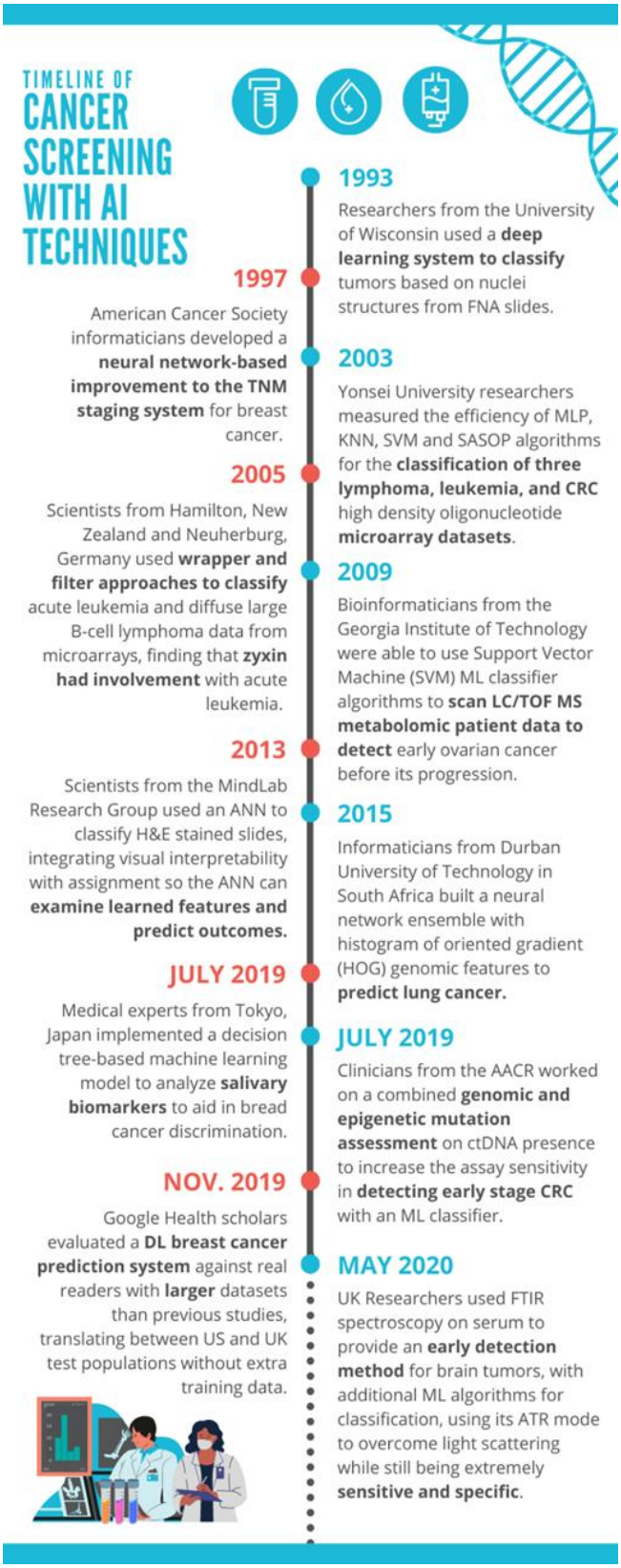

Figure 1: The development timeline of AI assistance in tumor prediction, detection, and diagnosis

ctDNA (circulating tumor DNA) allows for researchers to determine the survivability of the tumor and monitor any potential epigenetic developments [6]. In the microarray approach, neural networks will use vast amounts of patient data in the form of a DNA microarray, which displays different patterns of expressions of fluorescent markers on tumor cDNA (complementary DNA) and regular tissue cDNA. These deep learning algorithms will be able to make prognostic decisions that reflect on the degree of similarity between a patient's DNA microarray results and other DNA microarrays that suggest long survivability, allowing medical workers to avoid unnecessarily using chemotherapy. In addition, certain sections of genomes can be mapped numerically to look for and compare mutations in oncogenes and tumor suppressor genes, aiding clinicians in generating patient plans and predictions.

2. Machine Learning Advancements in Cancer Imaging and Computer Aided Diagnosis from 1990s to 2020s

As breast cancer screening programs began to increase in the 1980's, more possibilities for deep learning algorithms opened up in the field of tumor imaging and mophometrics. In 1993, researchers at the University of Wisconsin were able to use an interactive computer system that noted differences in cytologic features from a scan of Fine-needle aspiration (FNA) slides with the X Window System and the Athena WidgetSet on a DECStation 3100. Using slides from 569 patients, they were able to analyze the relative locations of the cell nuclei using an active contour model. With this setup, they were able to pinpoint grayscale discontinuities and evaluate nuclei radius, perimeter, area, compactness, smoothness, concavity, symmetry, and fractal dimension in a cancer diagnosing context. To predict the accuracy of the deep learning system, they performed a tenfold cross-validation, which divides the data set into ten equal sized, randomly selected parts and utilizes each part as a test set on a classifier from the remaining nine sets. This cross-validation yielded a separation accuracy of 97.3%. When testing the actual accuracy of the machine diagnosis on a sample of 54 slides (36 benign, 17 malignant, and 1 papilloma with atypia), the system was correct in every instance [2].

Furthermore, in 1997, a study done by American Cancer Society informaticians looked to neural networks that improved the existing TNM Classification of Malignant Tumors (TNM) staging system developed in the 1950s. TNM evaluates prognostic variables like tumor size, positive regional lymph nodes, radiographs, and distant metastasis. A Nevprop neural network was implemented for this study, using a gradient descent optimization function for backpropagation and adjusting weights. The data was divided into 5007 training cases, with a validation set made up of 3005 cases [7]. Using patient TNM variables from six different colorectal and breast cancer data sets, the neural network was able to give a more accurate 5-year specific prognosis than the TNM system by itself. For example, in the analysis of the breast carcinoma, the Artificial neural network (ANN) had a predictive accuracy of 0.770 as opposed to 0.720 for the TNM alone ($P<0.001$) [7]. In response to these results, the researchers were able to foresee how ANNs offer a better illness severity judgement and decrease interpatient variability, resulting in smaller and less expensive clinical trials with more homogenous groups. Following the trend of increasing clinical data types and availability over time, in 2013, scientists from Mindlab Research Group and Case Western Reserve University developed a deep learning network designed for analyzing digital images to detect and predict basal cell carcinoma. Using histopathology slides stained in hematoxylin and eosin (H&E), the auto-encoder algorithm is able to make out key patterns indicative of basal cell carcinoma in cell morphological structures rather than just observe common visual appearances in natural scene images. The scientists state how their approach differs from other older visual detection methods like discrete cosine transform and bag of features (BOF) by integrating

visual interpretability in the classification process, allowing the neural network to examine learned features and predict outcomes.

The cost function is represented with a logistic regression model to adjust the network's weights and biases during the training phase. Using a BCC dataset of 1417 images marked by a pathologist to show either a positive or negative presence of carcinoma, the network was able to significantly outperform BOF ($p>0.05$) in accuracy while differences among the deep learning networks were not significant ($p<0.01$), as shown by t test [8]. Another stage of the DL network was the addition of digital staining to highlight cancer features in red, explaining to pathologists why an automatic classifier is directing towards a certain classification. In the end of the study, they were able to showcase the ways that a learned network representation improves a canonical predefined representation, even aiding pathologists in interpretability with the automated classifier's digital staining output.

In 2015, two researchers from Stanford University created a similar modular approach to use a convolutional neural network (CNN) in the grading of brain gliomas. The treatment options for gliomas highly depend on the grade they receive. Thus, this modular pipeline classification method evaluates individual nuclei by assigning them a score based on nuclear morphometry, texture, intensity, and gradient statistic. Using a dataset from The Cancer Genome Atlas which contained H&E stained histopathology slides of GBM and the LGG variety, the researchers were able to select images where they segmented the nuclei in the preprocessing stage with top-hat filtering and hysteresis thresholding. The pipeline is implemented with Caffe software, first using a CNN to determine if the glioma is a GBM or LGG and then using another CNN to specify the grade of LGG if the glioma is not a GBM. When training the first ConvNet, the researchers used 875 samples of electron microbiopsy slides from 22 whole tissue slides from 4 distinct tissue source sites, with a training subset of 6998 and validation subset of 1752 samples. As for the second network, they used 766 e-microbiopsy samples from 22 whole tissue slides from 5 different tissue source sites, with a training subset of 5671 and validation subset of 1395 samples. During their testing of the modular pipeline on 10 independent test slides, the first CNN was accurate to 96% when classifying GBM vs LGG, while the second CNN was accurate to 71% when classifying Grade II vs Grade III for LGGs [3]. The lower success rate of the second CNN is explained by differences in slide preparation from institutions that provided data to the TCGA, as well as the subtler difference in morphology between a Grade II and Grade III LGG than for LGG vs GBM. Ultimately, the researchers noted that the ConvNet modular pipeline could help pathologists perform a second check while grading tumors and provide applications in teaching.

In the past five years, examiners have explored and expanded artificial intelligence to be applicable to a plethora of different early stage predictions and risk scores. For instance, in August 2019, medical scholars from Yonsei University, College of Medicine in Korea developed a lesion-based CNN that detected early gastric cancer (EGC) and predicted tumor depth from endoscopic imaging. In order to determine an appropriate treatment plan, staging must be accurate to endoscopy and biopsy findings. Since endoscopic resection and minimally invasive surgery are decided by the T stage, the plan is largely dependent on the tumor invasion depth. Endoscopic images are more subtle in differences than other medical imaging scenarios that implement deep learning. Therefore, an additional gradient-weighted class activation mapping (Grad-CAM) method was required to measure classification and localization errors. Grad-CAM allows the visualization of activation statuses over the training time, voiding the need for an additional module to generate visual explanations. A visual geometry group (VGG)-16 model was used to seperate endoscopic images into T1a, T1b, and non-EGC. In their study, 11,539 endoscopic images (896 T1a-EGC, 809 T1b-EGC, and 9834 non-EGC) were used from 800 patients (538 men and 262 women; age: 26–92 years; mean age: 62.6 years) from the Gangnam Severance Hospital in Seoul. In their portrayed results, blue lines indicate the activated regions at testing, while green lines encircle the actual EGC area. The EGC region was accurately classified despite a misclassification of depth or presence in some cases in EGC detection and EGC depth prediction. When evaluated on the same test image set, the sensitivity and specificity for EGC detection were 91.0% and 97.6%, respectively, with the overall area under the curve being 0.981. The sensitivity and specificity of the tumor depth prediction in the lesion-based VGG-16 were 79.2% and 77.8%, respectively, with the overall area under the curve being 0.851 [9]. In the end, the authors decided that a lesion based AI model was a promising tool in EGC prediction and diagnosis, despite a need for the expansion of AI that can better account for undifferentiated type histology and T1b-EGC.

Adding on to the expansion of neural network-based models into cancer prediction, Google Health scholars evaluated a DL cancer prediction system on data from the United Kingdom and the United States in November 2019. Looking past the limitations of previous computer-aided diagnosis methods, they decided to compare the results with readers in actual clinical practice rather than laboratory simulations, use larger datasets than the smaller, enriched ones in previous studies, and examine the ability of AI systems to translate between test populations without extra training data. The UK data contained screening mammograms that were collected in 2012-2015 from 25,856 women screened every three years at two screening centres in England. This set included 785 women who had a biopsy, and 414 women with cancer diagnosed within 39 months of imaging. As for the USA test set, women were screened every two years, with the data consisting of screening mammograms collected in 2001-2018 from 3,097 women at a single medical center. The images included 1,511 women who were biopsied during this time period (686 diagnosed with cancer within 27 months of imaging) and a random subset of women who have never received a biopsy. The AI system predictions were evaluated by biopsy-confirmed breast cancer outcomes along with the initial conclusion made by readers over the course of clinical practice. As for human performance, this was computed based on the decision of the clinician to recall the patient for additional diagnostics. Using a reader study that included six US-board certified radiologists interpreting 500 different mammograms with the breast imaging-reporting and data system (BI-RADS) scale, the scholars constructed a receiver operating characteristic (ROC) curve for each reader for AI system comparison. In the end, the AI model surpassed the average performance of radiologists by a notable margin, with a change in area under curve (AUC) of +0.115 (95% CI 0.055, 0.175; P = 0.0002). The ROC curve for the UK data has an AUC of 0.889 (95% CI 0.871, 0.907). For the US data curve, the AUC is 0.757 (95% CI 0.732, 0.780). Another purpose of the AI model is to specify certain areas of suspicion for malignant tumors. Similarly, the human readers in the study performed rectangular region-of-interest (ROI) annotations around findings. With multi-localization receiver operating characteristic (mLROC) analysis for reader and AI comparison, the scholars were able to summarize each mLROC plot by finding the partial area

under the curve (pAUC). The AI model outperformed the readers by +0.0192 (95% CI 0.0086, 0.0298; P= 0.0004) [10]. The final findings of the experiment reveal that the system is able to generalize from the UK to the US, surpassing the human readers and reducing the workload of second readers by 88%.

The following December, scientists from Karolinska University in Stockholm, Sweden developed a deep learning-based risk score for future breast cancer prediction, comparing it with a regular mammorgraphic density score. When screening for breast cancer, an MRI is often used for early detection, despite high costs and many more biopsies per cancer detected. By creating a score that performs risk stratification, MRI can be delivered to select women who need it most. Rather than a questionnaire based prediction method like the Gail or Tyrer-Cuzick model, their deep learning network can assess contributing factors outside of just age and mammographic density and extract more information. However, they do state a concern of the neural network being susceptible to learning irrelevant details from images with a flawed training setup. In the final study, 2238 women aged 40-74 were selected from the Stockholm county area, 278 of whom were diagnosed with breast cancer from 2013-2014. For developing the network, cases from 2008 to 2012 were used. The AUCs for the algorithm with DL risk score and the algorithm with age and dense area were 0.65 (95% CI: 0.63, 0.66), 0.60 (95% CI: 0.58, 0.61). For every DL based risk model, the AUC was higher than the models that were density-based. Since there was a modest correlation between the DL risk score and the density-based score, the scientists concluded that the DL risk score was not simply estimating density and assessed prediction scores without being limited by inter reader variability. They also noticed that the false-negative rate was lower for the DL based model than for the age-adjusted dense area model, with the effect being most apparent in women who were subsequently diagnosed with aggressive cancers like lymph node-positive cancers, where the febrile neutropenia rate was 31% for the DL risk score as opposed to 42% on the age-adjusted dense area model [11]. Although the AI risk score outperformed the density-based models, the scientists noted that the risk model should be trained with more cancer data from a variety of institutions, combined with another model to predict the risk of mammographic masking.

Backing the shift of machine learning models towards making predictions that aid clinicians in forming treatment plans and diagnoses, in early 2020, doctors from Chun Shan Medical University in Taichung, Taiwan constructed a machine learning algorithm that aided in the prediction of colorectal cancer (CRC) survival by identifying risk factors that influenced recurrence and secondary primary malignancies (SPMs), important indicators for treating CRC. They selected 4299 adult patients with primary CRC, 541 of which have had at least 1 SPM. Additionally, 1989 patients had recurrent CRC. In the past, risk factors for these two attributes were deemed to be tumor size, morphology, differentiation, prior radiation therapy, and smoking; however, no published rankings existed at the start of the study. Thus, the machine learning model filled the role and evaluated 20 different risk factors: patient age, primary site, histology, behavior code, differentiation, tumor size, pathologic stage (pStage), surgical margins, surgical procedures, radiation therapy, pre-operative radiation therapy, regional body order, dose levels of radiotherapy, extremum times of radiotherapy, BMI, smoking, areca consumption, and drinking. The researchers used the radial basis function kernel from the library for support vector machines (LIBSVM), building predictive models and optimizing both the C and γ parameters of each model. Furthermore, a Reduced Error Pruning Tree (REPTree) was implemented to produce a regression tree based on information gain or minimization of variance. As for the workflow of the model, the algorithm first split analyzed SPMs and recurrent CRC together, then four models for all 4 possible combinations of attributes were used. After that stage, feature selection through support vector machine (SVM)s and REPTrees was implemented and a 10-fold cross validation was employed to test the model's performance. The classifiers using only the top eight features (behavior code, differentiation, regional body order, age, areca, surgery, radiation therapy, and lowest dose) were evaluated with the classifiers using all 20 features. In the results, the REPTree model that used the top eight features possessed a Matthew's correlation coefficient (MCC) of 0.229, exceeding the MCC of the REPTree model without feature selection, which was 0.282. Despite the success in being able to weight selected features as predicting factors for SPM and CRC recurrence, the algorithm had limitations in which site-specific data like tumor markers were excluded and hereditary data could not be provided [12]. Ultimately, the researchers were able to develop a clean, feasible method with feature selection for both improving prognostic and diagnostic accuracy and identifying elements that contribute to SPM and CRC recurrence, distinguishing the top four factors as pStage, surgical margin, smoking, and drinking.

3. Progress of Machine Learning Modeling in Biofluid Analysis and Cancer Metabolomics from 2000s to 2020s

In contrast with cancer imaging, a process that requires a larger devotion of time and a human radiologist to make diagnoses based on experience, metabolomics is an emerging field that allows for changes in metabolic processes and micromolecule composition to be measured and accounted for in a cancer diagnosing context with mass spectroscopy devices, especially useful when combined with the data guided classification ability of machine learning.

For instance, in 2009, bioinformaticians from the Georgia Institute of Technology were able to use SVM ML classifier algorithms to scan liquid chromatography/ time of flight mass spectroscopy (LC/TOF MS) metabolomics data from ovarian cancer patients in order to detect the disease and circumvent the issue of its asymptomatic nature, often leading to a delayed diagnosis until its later progression (stage III/IV). LC/TOF MS was used due to its broadband metabolic profiling abilities and detection of metabolites that are distinct in chemical properties. The researchers obtained plasma serum from a total of 37 patients with papillary serous ovarian cancer and 35 controls. Using LC/TOF MS, mzMine was able to reveal 576 features in positive ion mode and 280 features in negative ion mode. Using leave-one-out cross validation to evaluate prediction performance without feature selection, the nonlinear SVM outperformed the linear SVM with a performance of 83.3%. However, using feature selection methods, the predictive performance soared to over 90% [13].

The bioinformaticians recognized that selection bias could be introduced if feature selection methods are applied to the whole dataset, affecting the prediction performances between feature selection methods. Statistical testing revealed that observed prediction performance between any ordered pair of the four feature selection methods were not statistically significant except for that between the SVMRFE_NL and the SVMRFE method. Despite this, they ultimately devised a way to aid in the diagnosis

of ovarian cancer in its early stages with metabolomic data as an indicator. In 2018, professors from Lanzhou University in Lanzhou, China were able to analyze patient plasma with known machine learning algorithms like SVMs, K-Nearest Neighbors (KNNs), Random Forest (RF), and back propagation (BP) to detect the presence of thyroid carcinoma (TC) and two kinds of tumors called pheochromocytoma and paraganglioma (PPGL). Using matrix-assisted laser desorption-ionization time of flight mass spectrometry (MALDI-TOF-MF), the scientists were able to analyze the mass spectrum of 150 normal plasma samples, 144 samples from PPGL patients, and 36 samples from TC patients. Conclusively, the SVM was the most accurate classifier for detecting TC and PPGL (92.86% and 99.15% respectively), slightly superior to BP (87.5% and 95.51% respectively) [14].

Furthermore, in July 2019, medical experts from Keio University, Teikyo University, and Kitasato University in Tokyo, Japan implemented a decision tree-based machine learning model to analyze salivary biomarkers to aid in breast cancer discrimination. Reflecting the potential overdiagnosis, high costs, and radiation exposure found in screening mammography, the medical experts decided to screen using saliva as an informative biofluid due to the presence of hydrophilic metabolites like amino acids and polyamines that can differentiate healthy controls and patients with breast cancer, as well as its practical and cost-effective collection. From 101 patients with invasive breast carcinoma, 23 patients with ductal carcinoma in situ (DCIS), and 42 healthy controls, the researchers were able to collect 166 unstimulated saliva samples. For the non-targeted examination for the presence of hydrophilic metabolites in samples, CE-time-of-flight-MS (CE MS) was used. Meanwhile, LC-triple quadrupole MS (QQQMS) was implemented for the quantification of polyamines. In an effort to remove unreliable data, metabolites that were either detected in less than 50% of IC samples or below the quantification limit in 20+ samples were eliminated from the data set. The Mann–Whitney test was used for comparisons between controls versus IC and Q-values were obtained through adjusting P-values using a false discovery rate (FDR) that considered multiple independence tests. For between all three groups, the researchers decided on the Kruskal–Wallis test and Dunn's post test. After refining the data, multiple machine learning improved alternative decision trees (ADTrees) were implemented in an ensemble approach with information collected after bias controlled resampling. The number of nodes per tree (boosting number) and the total number of trees (bagging number), were determined by two-fold cross-validation. The resampling involved a process where individual data was randomly selected with redundant selection, repeated a total of 200 times with both differing and random values. The polyamine spermine displayed the greatest AUC values for comparing IC to C and the MLR model containing spermine and Ru5P together showed higher AUC values than each component model alone. Since only these two metabolites remained after features were selected using P = 0.05, this suggested a positive correlation between other metabolites and spermine or Ru5P [15]. When compared together, both models showed no significant difference between ROC curves. The ADTree model showed better AUC values than the spermine and MLR model, while ADTree+Bagging showed the highest AUC values as this was the only model that showed significant differences in ROC curves compared to the other models. The researchers named several limitations in their study, such as polyamine concentrations in biofluids being influenced by diet, environmental factors, and various diseases. The researchers state that the final discrimination model should be compared with other cancer data to evaluate specificity. Ultimately, they concluded that salivary metabolomics combined with a machine learning based approach to classification can provide a non-invasive screening procedure and can be conducted before a mammography to recommend a biopsy.

Among the popular forms of mass spectroscopy for cost-effective biofluid data analysis is Fourier Transform Infrared spectroscopy (FTIR), which can be run on human blood serum to discriminate between cancer patients and controls. In May 2020, clinicians from the University of Strathclyde and University of Liverpool in the UK used FTIR spectroscopy on serum to provide an early detection method for brain tumors, with additional machine learning algorithms for classification. In their report, they first discuss how cancer antigen tests like CA 19-9 and CA 125 have high specificity values but low sensitivity values, yielding a low Positive Predictive Value (PPV). Since this value determines the probability that a test will predict a true cancer diagnosis, current screening tests may lead to unnecessary diagnosis and treatment. However, FTIR spectroscopy's new attenuated total reflection mode allows for scientists to measure the change in infrared radiation from an internally reflected beam touching the sample, overcoming light scattering while still being highly sensitive and specific. The researchers made note of a study done in 2010 where 31-96 year old patients with breast carcinoma in situ and 98 healthy controls had their serum extracted and analyzed by FTIR spectroscopy. They then used cluster analysis and an ANN (unsupervised and supervised respectively) to classify the patients, yielding a sensitivity and specificity above 95%. Furthermore, the difficulty of brain cancer diagnosis through imaging encouraged the use of non-invasive alternatives to biopsies. In another described experiment, 41 lymphoma and GBM serum samples were analyzed with ATR-FTIR spectroscopy and classified with machine learning techniques like RF, partial least square discriminant analysis (PLS-DA), and SVM. The PLS-DA algorithm outperformed the others, with a sensitivity of 86.3% and a specificity of 90.1% [4]. Therefore, despite the current obstacles of clinical workflow disruption and increased economic burden placed on healthcare workers, the researchers made it clear that FTIR spectroscopy in combination with a machine learning-based classifier allows for a reagent-free, noninvasive, and inexpensive platform that allows for clinicians to discover new tumor biomarkers and obtain an early diagnosis.

4. Developments in DNA Microarrays and Genomics with Machine Learning Classification from 2000s to 2020s

As the use of DNA microarrays to measure gene expression from mRNA transcripts expanded to include clinical biocomputation in the past few decades, new applications were uncovered for selecting cDNA fluorescent markers that combined with AI classifiers, could detect and predict the early onset of cancer for many patients. In early 2003, shortly before the Human Genome Project was completed, researchers Sung-Bae Cho and Hong-Hee Won from Yonsei University in Seoul, South Korea were able to measure the efficiency of a multi-layer perceptron (MLP), KNN, SVM and structure adaptive self–organizing map (SASOP) algorithm for the classification of three lymphoma, leukemia, and CRC high density oligonucleotide microarray datasets. For feature selection, they decided on 7 different methods: Pearson's correlation coefficient (PC), Spearman's correlation coefficient (SC), Euclidean distance (ED), cosine coefficient (CC),

information gain (IG), mutual information (MI), and signal to noise ratio (SN). The informative genes that were chosen from gene/feature selection were visualized through statistical correlation analysis with a linear relationship. For the leukemia dataset, 25 samples were acute myeloid leukemia (AML) and 47 samples were acute lymphoblastic leukemia (ALL), taken from 63 bone marrow samples and 9 peripheral blood samples. In the colon dataset, there are 62 samples of colon epithelial cells from CRC patients, 40 of which are positive CRC samples, with every sample containing 2000 gene expression levels. The lymphoma dataset had 24 samples of GC B-like and 23 samples of activated B-like. When comparing the average performance of each feature selection method, IG and PC had the highest recognition rate. As for the classifiers, KNN (with PC feature selection) and MLP produced a better average recognition rate of 85.3% for both algorithms in the leukemia dataset [16]. In the end, for the ensemble method of cancer classification, the correct combination of KNN, MLP, IG, and PC together allowed for a simple approach to rank genes and categorize from DNA chips.

Likewise, in 2005, a similar study was conducted by bioinformaticians from Hamilton, New Zealand and Neuherburg, Germany, using wrapper and filter approaches to classify acute leukemia and diffuse large B-cell lymphoma data from microarrays. Filter approaches, which are faster due to less of a computational requirement, remove irrelevant characteristics according to the overall composition of the data's features, while wrapper approaches, generally being more accurate based on the use of machine learning algorithms, use selected feature subsets and cross-validation to evaluate said feature subsets. For their filter approach, they mainly used correlation-based feature selection (CFS), with the classifiers being the decision tree learner C4.5, Naïve Bayes, and a SVM. The leukemia data set contains ALL and AML, with the training set having 38 bone marrow samples (27 ALL and 11 AML). When ranking genes using two classifiers, C4.5 and Naïve Bayes, and the wrapper, a single gene is selected, zyxin, which is also the only gene selected by CFS. In 38 runs, When run 38 times, zyxin was selected 34 times (92%) by CFS, 34 times (92%) by the C4.5 wrapper and 28 times (74%) by the Naïve Bayes wrapper. Boxplots of zyxin expression levels for each training data set indicates that zyxin expression can differentiate ALL from AML in the training set, as the medians are 360.0 and 2947, respectively [5]. The researchers then note that zyxin has been demonstrated to enter the nucleus with other proteins, exiting with leucine-rich nuclear export sequences and potentially regulating transcription activity by interactions with transcription factors. Because of these results alone, the researchers were able to conclude that zyxin had at least indirect involvement with acute leukemia, and that a combination of classifiers and wrappers is an accurate tool for pulling relevant features out of microarray data. Nevertheless, the researchers state that one computational drawback of some ranking filter algorithms is that each gene is scored individually, while in reality a combination of expression levels for multiple genes might be responsible for cancer. By not recognizing that several genes may contribute to the onset of some cancers, these filters might look past these genes if their individual expression levels are not informative enough for classification. As the proliferation of big data allowed clinical applications of machine learning on cancer classification and detection to scale up in the last decade, more clinicians have been looking at approaches to cancer prediction and treatment plans.

One such example was in 2015, where scientists Emmanuel Adetiba and Oludayo O. Olugbara from Durban University of Technology in Souh Africa built a neural network ensemble with histogram of oriented gradient (HOG) genomic features to predict lung cancer. They first identified two major groups of genes that are mutated in lung cancer cells: oncogenes and tumor suppressor genes. Then, they discussed current methods of genetic screening for cancer, such as direct tumor sequencing (limited by low sensitivity), quantitative polymerase chain reaction (PCR), fluorescence in situ hybridization (FISH), immunohistochemistry, and microarray technology (all limited by degree of mutation coverage). In the end, the authors decided to evaluate the targeted sequencing (TS) capability of next generation sequencing (NGS) to predict non-small cell lung cancer (NSCLC). Using 6,406 samples from the National Center for Biotechnology Information (NCBI), their experimental dataset contained normal, EGFR deletion, EGFR substitution, KRAS substitution, TP53 deletion, and TP53 substitution mutations. They mapped nucleotides numerically with the Voss mapping method, which allows for various digital image processing (DIP) techniques to acquire feature descriptors like HOG and LBP for genomic sequences. The Voss map is constructed for the EGFR nucleotides, providing a unique feature representation for ANN and SVM use. When running the ensemble and nonensemble MLP-ANNs, the researchers partitioned the HOG data into 70% training, 15% validation, and 15% testing. In addition, they chose to vary the number of neurons in the hidden layer from 10 in steps of 10 to 100, recording the mean square errors. The 8th MLP-ANN had the highest accuracy of 87.6%, with a validation performance of 0.0584 at 490 epochs [17]. Ultimately, the researchers concluded that their approach with a HOG descriptor, MLP-ANN, and Voss numerical mapper was versatile in automated cancer prediction, with several biomarkers for lung cancer on a single platform. Another benefit was cooperation with NGS genomic-based technology, allowing high prediction accuracies.

In more recent years, machine learning research has branched DNA based assessments to include epigenetic factors as well. In particular, in July 2019, clinicians from the AACR developed a combined genomic and epigenetic mutation assessment on ctDNA presence to increase the assay sensitivity in early stage CRC. This model was based on a machine learning classifier trained on 111 ctDNA samples (38 late stage, 10 early stage, and 63 healthy controls). With their plasma-only assay that included somatic genomic variant detection and epigenomic analysis in conjunction with the ML classifier for filtering, the ctDNA detection rate for early stage CRC was 94%, with 94% specificity, exceeding the detection rate of somatic genomic variant detection alone [6]. Consequently, the clinicians emphasized the implications of patient ctDNA detection rate in medical utility and early stage management, working as a significant indicator of stage I-III CRC.

5. Conclusions and Future Perspectives

The impacts of machine learning and later, deep learning on clinical data and diagnostics were relatively new, having expanded in generalizability and predictive power in the last few decades. The advantages and drawbacks of AI-based cancer screening are listed in table 1 and divided into three subcategories: image analysis, biofluid analysis, and genetic analysis. When comparing the general trends of the advancements regarding each method over time, the contrast of reports from the 1990s to the 2020s establish an increase in the clinical translation of deep learning and machine learning tools, decreasing costs, healthcare worker burdens, human error, and biopsies per cancer detected. Newer methods like FTIR

spectroscopy, salivary metabolomics, and combined genomic/epigenetic assessments work with alternative tree-based classifiers and convolutional ANNs to elevate predictive performance, sensitivity, and specificity beyond predecessors that worked with smaller and less homogenous data sets. As patient databases multiply in size and availability in the future, the clinical implications may reveal a similar trend in expanding predictive accuracy in cancer screening methods, decreasing both clinical trial and patient healthcare costs, and increasing the rate at which patients are diagnosed in early stage cancers.

Currently, more research needs to be conducted in the field of proper, effective initial neural network training and multi-gene dependent cancer expression factors in order to deliver more pinpoint classification and allow AI models to learn specific features to select in genomic assessments, respectively.

Table 1: The advantages and limitations of available AI assisted tumor analysis.

ML Analysis Method	Applications	Advantages	Limitations
Image Analysis	FNA slides, histopathology slides, MRIs, endoscopies	• Imaging analysis allows clinicians to reduce false positives with higher sensitivity in classification. • The characterization and classification of tumors influence several kinds of risk predictors, such as a mammographic risk score. • AI based CAD models were able to more accurately read mammograms than human readers with a change in AUC of +0.115 (95% CI 0.055, 0.175; P = 0.0002) • When combining the TNM staging system with ANNs, the combined system provides a better illness severity judgement and decreases interpatient variability, decreasing the cost of clinical trials. • The early detection and characterization of breast cancer allows for the conservation of MRIs for patients who need it most, further decreasing biopsies per cancer detected [11].	• During classification, neural networks are susceptible to learning irrelevant features if there are flaws in training setup [11]. • CRC risk evaluating ML models did not account for other site-specific tumor markers and hereditary data. • Some recurring instances of neoplastic lesions contain molecular aberrations different from the main tumor, increasing resistance to radiation and imaging therapies [18]. • Cancer imaging requires a larger devotion of time than other cancer detection and prediction technologies, as well as a human radiologist to make experience guided diagnoses.
Biofluid Analysis	Plasma fractions, blood serum, saliva	• Combined with LC/TOF MS, which revealed 576 features in positive ion mode and 280 features in negative ion mode, feature selection methods allowed the predictive performance of the model to exceed 90% when using leave-one-out cross validation to evaluate prediction. • The analysis of FTIR mass spectroscopy allows for a wide range of biomolecules with differing charges, concentrations, and chemical properties to be evaluated [4]. • Metabolomic data from mass spectroscopy does not need a large commitment of time to analyze and can be automated to a larger extent than radiological machine learning methods. • Screening with saliva as an informative biofluid eliminates radiation exposure, promotes a cost-effective collection method, and differentiates healthy controls and patients with breast cancer due to the presence of hydrophilic metabolites like amino acids and polyamines. • The attenuated total reflection mode for FTIR spectroscopy allows change in infrared radiation to be calculated from an internally reflected beam in contact with the sample, conserving high sensitivity and specificity while also overcoming light scattering.	• Clinical translation of medical spectroscopy technology has yet to be achieved due to current obstacles of clinical workflow disruption and a larger economic burden placed on healthcare employees. • Selection bias can be introduced in prediction models when feature selection methods are applied to the whole dataset, changing the prediction performances between feature selection algorithms.

Genetic Analysis	ctDNA, DNA microarrays, epigenetic and genomic biomarkers	• DNA microarrays provide a large volume of data to allow neural networks to train and increase predictive performance. • ctDNA allows clinicians to view the current and dynamic stages of a cancer, tracking progression and monitoring for genetic or epigenetic mutations in short time periods [6]. • Using a HOG descriptor, MLP-ANN, and Voss numerical mapper allowed several biomarkers for lung cancer on a single platform for automated cancer prediction that returned high accuracy due to cooperation with NGS genomic-based technology [17]. • When combining an ML classifier for filtering with an epigenetic and genomic plasma assay, the ctDNA detection rate for early stage CRC improves to 94%, with 94% specificity [6].	• Some ranking filter algorithms rank genes individually; however, the reality is that combinations of several genes and expression levels may account for cancerous traits. Without considering the possibility of several genes contributing to cancer, filters can potentially look past these genes if they are not expressive and informative enough individually [5]. • Some metabolites like polyamines and amino acids fluctuate according to diet, diseases, lifestyle, and environmental factors, meaning that cancer may not be the root cause of some polyamine abnormalities. Thus, when combining multiple markers, extraneous factors should be minimized for accurate determination [15].

Supplementary Materials

Author Contributions: Conceptualization, W.H., C.Z. and X.F.; Literature search, W.H. and X.F.; Writing – Original Draft Preparation, W.H.; Writing – Review & Editing, W.H., C.Z. and X.F.; Supervision, C.Z. and X.F.; Project Administration, C.Z.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Cruz, J.A.; Wishart, D.S. Applications of machine learning in cancer prediction and prognosis. Cancer Inform. 2007, 2, 59–77.

[2] Wolberg, W.H.; Street, W.N.; Mangasarian, O.L. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. Cancer Lett. 1994, 77, 163–171, doi:10.1016/0304-3835(94)90099-x.

[3] Ertosun, M.G.; Rubin, D.L. Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks. In Proceedings of the AMIA Annual Symposium Proceedings; American Medical Informatics Association, 2015; Vol. 2015, p. 1899.

[4] Sala, A.; Anderson, D.J.; Brennan, P.M.; Butler, H.J.; Cameron, J.M.; Jenkinson, M.D.; Rinaldi, C.; Theakstone, A.G.; Baker, M.J. Biofluid diagnostics by FTIR spectroscopy: A platform technology for cancer detection. Cancer Lett. 2020, 477, 122–130, doi:10.1016/j.canlet.2020.02.020.

[5] Wang, Y.; Tetko, I.V.; Hall, M.A.; Frank, E.; Facius, A.; Mayer, K.F.X.; Mewes, H.W. Gene selection from microarray data for cancer classification—a machine learning approach. Comput. Biol. Chem. 2005, 29, 37–46, doi:10.1016/j.compbiolchem.2004.11.001.

[6] Kim, S.-T.; Raymond, V.M.; Park, J.O.; Zotenko, E.; Park, Y.S.; Schultz, M.; Kang, W.K.; Westesson, O.; Kim, H.-C.; He, Y.; et al. Abstract 916: Combined genomic and epigenomic assessment of cell-free circulating tumor DNA (ctDNA) improves assay sensitivity in early-stage colorectal cancer (CRC). Clinical Research (Excluding Clinical Trials) 2019.

[7] Burke, H.B.; Goodman, P.H.; Rosen, D.B.; Henson, D.E.; Weinstein, J.N.; Harrell, F.E., Jr; Marks, J.R.; Winchester, D.P.; Bostwick, D.G. Artificial neural networks improve the accuracy of cancer survival prediction. Cancer 1997, 79, 857–862, doi:3.0.co;2-y">10.1002/(sici)1097-0142(19970215)79:4<857::aid-cncr24>3.0.co;2-y.

[8] Cruz-Roa, A.A.; Arevalo Ovalle, J.E.; Madabhushi, A.; González Osorio, F.A. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. Med. Image Comput. Comput. Assist. Interv. 2013, 16, 403–410, doi:10.1007/978-3-642-40763-5_50.

[9] Yoon, H.J.; Kim, S.; Kim, J.-H.; Keum, J.-S.; Oh, S.-I.; Jo, J.; Chun, J.; Youn, Y.H.; Park, H.; Kwon, I.G.; et al. A Lesion-Based Convolutional Neural Network Improves Endoscopic Detection and Depth Prediction of Early Gastric Cancer. J. Clin. Med. Res. 2019, 8, doi:10.3390/jcm8091310.

[10] McKinney, S.M.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Back, T.; Chesus, M.; Corrado, G.C.; Darzi, A.; et al. International evaluation of an AI system for breast cancer screening. Nature 2020, 577, 89–94, doi:10.1038/s41586-019-1799-6.

[11] Dembrower, K.; Liu, Y.; Azizpour, H.; Eklund, M.; Smith, K.; Lindholm, P.; Strand, F. Comparison of a Deep Learning Risk Score and Standard Mammographic Density Score for Breast Cancer Risk Prediction. Radiology 2020, 294, 265–272, doi:10.1148/radiol.2019190872.

[12] Ting, W.-C.; Lu, Y.-C.A.; Ho, W.-C.; Cheewakriangkrai, C.; Chang, H.-R.; Lin, C.-L. Machine Learning in

Prediction of Second Primary Cancer and Recurrence in Colorectal Cancer. Int. J. Med. Sci. 2020, 17, 280–291, doi:10.7150/ijms.37134.

[13] Guan, W.; Zhou, M.; Hampton, C.Y.; Benigno, B.B.; Walker, L.D.; Gray, A.; McDonald, J.F.; Fernández, F.M. Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. BMC Bioinformatics 2009, 10, 259, doi:10.1186/1471-2105-10-259.

[14] Wang, X.; Yong, B.; Cao, Y.; Zhang, G.; Zhou, Q. TC and PPGL Detection Based on Machine Learning Models. In Proceedings of the 2018 Sixth International Conference on Advanced Cloud and Big Data (CBD); 2018; pp. 349–353.

[15] Murata, T.; Yanagisawa, T.; Kurihara, T.; Kaneko, M.; Ota, S.; Enomoto, A.; Tomita, M.; Sugimoto, M.; Sunamura, M.; Hayashida, T.; et al. Salivary metabolomics with alternative decision tree-based machine learning methods for breast cancer discrimination. Breast Cancer Res. Treat. 2019, 177, 591–601, doi:10.1007/s10549-019-05330-9.

[16] Cho, S.-B.; Won, H.-H. Machine learning in DNA microarray analysis for cancer classification. In Proceedings of the Proceedings of the First Asia-Pacific bioinformatics conference on Bioinformatics 2003-Volume 19; 2003; pp. 189–198.

[17] Adetiba, E.; Olugbara, O.O. Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features. ScientificWorldJournal 2015, 2015, 786013, doi:10.1155/2015/786013.

[18] Bi, W.L.; Hosny, A.; Schabath, M.B.; Giger, M.L.; Birkbak, N.J.; Mehrtash, A.; Allison, T.; Arnaout, O.; Abbosh, C.; Dunn, I.F.; et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. CA Cancer J. Clin. 2019, 69, 127–157, doi:10.3322/caac.21552.